Solution Structure of a POU-Specific Homeodomain: 3D-NMR Studies of Human B-Cell Transcription Factor Oct-2[†,‡]

Mohanram Sivaraja,[§,‖] Martyn C. Botfield,[§,⊥] Mark Mueller,[§,#] 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 February 8, 1994; Revised Manuscript Received April 28, 1994[⊛]

ABSTRACT: The POU DNA-binding motif defines a conserved family of eukaryotic transcription factors involved in regulation of gene expression. This bipartite motif consists of an N-terminal POU-specific domain (POU_S), a flexible linker, and a C-terminal POU-specific homeodomain (POU_{HD}). Here we describe the solution structure of a POU-specific homeodomain. An NMR model is obtained from Oct-2, a human B-cell specific transcription factor which participates in the regulation of immunoglobulin genes. A fragment of Oct-2 containing POU_{HD} and an adjoining linker was expressed in *Escherichia coli* and characterized by three-dimensional nuclear magnetic resonance (3D-NMR) spectroscopy. Complete ^1H and ^{15}N resonance assignment of the POU_{HD} moiety is presented. The POU_{HD} solution structure, as calculated by distance geometry and simulated annealing (DG/SA), is similar to that of canonical homeodomains. A salient difference between solution and crystal structures is observed in the C-terminal segment of α-helix 3 (the HTH recognition helix), which is not well ordered in solution. Because this segment presumably folds upon specific DNA binding, its flexibility in solution may reduce the intrinsic DNA affinity of POU_{HD} in the absence of POU_S.

The bipartite POU[1] motif defines a family of homeodomain-containing transcription factors (Bodner *et al.*, 1988; Ingraham et al., 1988; Strum *et al.*, 1988; Clerc *et al.*, 1988; Ko *et al.*, 1988; Scheidereit *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 divergent homeodomain (POU_{HD}). Although jointly required for high-affinity DNA recognition, the two domains fold independently and are flexibly linked (Botfield *et al.*, 1992). The structure of a POU_S fragment has recently been shown to be similar to the DNA-binding domains of phage repressor and Cro proteins (Assa-Munt *et al.*, 1993; Dekker *et al.*, 1993). In this paper we describe the ^1H and ^{15}N resonance assignment and solution structure of a POU-specific homeodomain. Comparison with canonical homeodomains demonstrates global similarities in structure with local differences in the stability of the putative HTH recognition α-helix. Taken together, NMR studies of POU_S and POU_{HD} fragments demonstrate that the POU motif contains two distinct HTH elements (Assa-Munt *et al.*, 1993; Dekker *et al.*, 1993; Morita *et al.*, 1993). Biochemical studies indicate that these elements are aligned in successive DNA major grooves (Verrijzer *et al.*, 1992; Jancso *et al.*, 1994; Botfield & Weiss, 1994). A crystal structure of a specific POU–DNA complex has recently been described (Klemm *et al.*, 1994; see Added in Proof).

The present studies focus on Oct-2, a human B-cell specific transcription factor which is also expressed in certain neuronal lineages (Singh *et al.*, 1986; Clerc *et al.*, 1988; Ko *et al.*, 1988; Müller-Immerglück *et al.*, 1988; Scheidereit *et al.*, 1988). Oct-2 belongs to a subfamily of POU proteins whose members (designated Oct-1, Oct-2, Oct-3, etc.) recognize a conserved octanucleotide sequence (5′-ATGCAAAT-3′)[2] in vertebrate promoter and enhancer elements. This laboratory has previously described the overexpression of Oct-2 POU and POU_{HD} fragments in *E. coli* and initial characterization of their domain structure and DNA-binding properties (Botfield *et al.*, 1992, 1994; Jancso et al., 1994). Alignment of the POU_{HD} HTH against a consensus octamer site has been established by study of a mutant POU domain with relaxed sequence specificity (Botfield *et al.*, 1994). Unlike canonical homeodomains, the isolated Oct-2 POU_{HD} and other POU-specific homeodomains (Verrijzer & Van der Vliet, 1993) exhibit only weak specific DNA binding ($K_d > 1$ μM). Although its structural basis is not understood, such weak binding enforces a functional requirement for coordinate POU_S binding. Bipartite DNA recognition permits combinatorial binding to longer DNA sites (Verrijzer *et al.*, 1990, 1992; Li *et al.*, 1993; Botfield *et al.*, 1994); flexibility in subdomain orientation also permits the POU motif to recognize adjoining DNA subsites with

[†] This work was supported in part by grants from the NIH (GM45290) and the Council for Tobacco Research to M.A.W., a Cancer Research Institute Postdoctoral Fellowship (M.C.B.), and an NIH Postdoctoral Training Grant in Endocrinology to Massachusetts General Hospital (M.M.).

[‡] Molecular coordinates have been deposited in the Brookhaven Protein Data Bank (Accession Number 1HDP).

* Author to whom correspondence should be addressed.

§ Harvard Medical School.

‖ Present address: Department of Surgery, Brigham and Women's Hospital, Boston, MA 02146.

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

Massachusetts General Hospital.

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

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

⊛ Abstract published in *Advance ACS Abstracts*, July 15, 1994.

[1] Abbreviations: DG, distance geometry; DQF-COSY, double-quantum-filtered correlated spectroscopy; HD, homeodomain; HMQC, heteronuclear multiple-quantum coherence; HSQC, heteronuclear single-quantum coherence; HTH, helix–turn–helix; l-POU_{HD}, Oct-2 fragment containing the linker sequence and adjoining homeodomain; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; NOESY, NOE 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; see Figure 1); POU_S, POU-specific subdomain of Oct-2 (amino acids 194–270); RMS, root mean square; SA, simulated annealing; SDS, sodium dodecyl sulfate; TOCSY, total correlation spectroscopy; 2D-NMR, two-dimensional NMR; 3D-NMR, three-dimensional NMR.

[2] Amino acids and DNA bases are designated by standard single-letter code.

0006-2960/94/0433-9845$04.50/0 © 1994 American Chemical Society

altered internal spacing or polarity (Li *et al.*, 1993). Finally, modular design of DNA-binding motifs permits regulation of protein–protein interactions in the assembly of higher-order transcription complexes (Kristie *et al.*, 1989; Stern *et al.*, 1989; Kristie & Sharp, 1990).

MATERIALS AND METHODS

Protein Expression and Purification. The isolated homeodomain (designated POU_{HD}; sequence given in Figure 1) and the linker–homeodomain fragment (l-POU_{HD}) were initially expressed in pGEX-2T (Pharmacia, Inc.) as fusion proteins with glutathione *S*-transferase and purified as described (Botfield *et al.*, 1992). To facilitate ^{15}N labeling, a second expression plasmid was constructed in pT7-7 (Tabor, 1989). This plasmid (designated pMB-H_6IH) contains a T7 promoter and His_6 N-terminal tag, permitting more efficient expression and purification by Ni^{2+}-affinity chromatography; the protocol is described in Qian *et al.* (1993). The final yield was 1 mg/L minimal medium. For this construction, the Oct-2 linker–POU_{HD} sequence (nucleotides 877–1143, encoding residues 271–359; LeBowitz *et al.*, 1988) was modified by PCR mutagenesis to contain a 5′ (N-terminal) initiation codon embedded in a unique *Nde*I restriction site, a polyhistidine linker (encoding GSSHHHHHH), a 3′ (C-terminal) ohcre stop codon (5′-TAA), and a unique *Bam*HI restriction site. The *Nde*I–*Bam*HI expression cassette was subcloned into corresponding restriction sites in pT7-7; the construction was verified by double-stranded DNA sequencing. The expressed fragment contains 99 residues with predicted molecular mass 11 303 Da. The N-terminal His_6 tag was observed not to perturb POU_{HD}-specific resonances. Relative to the longer fragment, pT7-7-based expression of POU_{HD} (65 residues) was reduced by 10-fold. Such inefficient expression has not been observed in other HD systems [for example, see Muller *et al.* (1988)] and may reflect the limited solubility and thermal stability of the Oct-2 POU_{HD} (part I of Results).

DNA-Binding Assays. Specific DNA-binding activities of expressed POU_{HD} fragments were evaluated using gel-retardation assays (GRA; Fried & Crothers, 1981) at 4 °C under low-salt conditions as described (Botfield *et al.*, 1992, 1994). The DNA probe (5′-GTATGCAAATGG-3′) contains a consensus octamer site (underlined). The POU_{HD} dissociation constant for this site is approximately 6 μM (Botfield *et al.*, 1992) and is not affected by inclusion of N-terminal His_6 and linker sequences.

NMR Methods. The POU_{HD} fragments (1.5 mM) were dissolved in 0.7 mL of NMR buffer containing 50 mM potassium phosphate (pH or pD 6.0; direct meter reading) and 5 mM deuterated dithiothreitol (DTT; Merck Isotopes, Inc., St. Louis, MO). Spectra were obtained in 99.98% D_2O and 90% H_2O/10% D_2O at 500 MHz. Three-dimensional NOESY–HMQC spectra (Marion *et al.*, 1989) were obtained as described (Qian *et al.*, 1993). Double-quantum-filtered correlation spectroscopy (DQF-COSY), total correlation spectroscopy (TOCSY mixing times 30 and 55 ms), and nuclear Overhauser enhancement spectroscopy (NOESY mixing times 100 and 200 ms) experiments were performed by the pure-phase method of States *et al.* (1982). $^3J_{\alpha N}$ coupling constants were measured on the basis of (i) the structure of "fingerprint" cross-peaks in the DQF-COSY spectrum in H_2O with distortions arising from finite line widths corrected by spectral simulation (Redfield & Dobson, 1990; Smith *et al.*, 1991) and (ii) analysis of cross-peak displacement in TOCSY spectra of uniform ^{15}N-labeled protein recorded in the absence of ^{15}N decoupling (Montelione & Wagner, 1989). Experiments were performed with Varian Unity and VXR 500-MHz spectrometers. Quadrature in both dimensions was achieved using the States–TPPI method. Water suppression was achieved by solvent presaturation. Data were processed with a combination of exponential and shifted sine-bell window functions in each dimension. Chemical shifts are measured in parts per million (ppm) relative to H_2O (presumed to be 4.78 ppm at 25 °C and 4.85 ppm at 18 °C). Selected stereospecific assignments were obtained as described for seven β-methylene pairs (Wagner *et al.*, 1987), three valine γ-CH_3 groups (Zuiderweg *et al.*, 1985), and one leucine δ-CH_3 group (Kochoyan *et al.*, 1991) as underlined in Table 1.

Heteronuclear ^{15}N-NMR. Experiments were performed with His_6 l-POU_{HD} containing uniform ^{15}N enrichment. A 2D HSQC spectrum was obtained with spectral widths of 2000 Hz (ω_1) and 7000 Hz (ω_2) (Figure 4; part II of Results). Experimental data consisted of 512 × 128 complex points and were zero-filled to 2048 × 1024 complex points. A 3D NOESY–HMQC spectrum (supplementary material) was collected with a spectral width of 7000 Hz in both ω_1 and ω_3 and with 1500 Hz in ω_2. A total of 32 × 128 × 512 data points were obtained. The data were extended by linear prediction to 128 × 256 × 512. In each case the NOESY mixing time was 200 ms. Quadrature detection in ω_1 and ω_2 was obtained using the TPPI–States method (Marion *et al.*, 1989).

NMR-Derived Restraints. NOE and *J*-coupling (dihedral angle) restraints were used for molecular modeling as previously described (Qian *et al.*, 1993). NOEs were classified as strong (<2.7 Å), medium (<3.4 Å), or weak (<4.3 Å) on the basis of amplitude of NOESY cross-peaks relative to the intraresidue ortho–meta cross-peak of F49. Distance-bound corrections were applied for methyl and methylene protons for which stereospecific assignments could not be obtained (Wuthrich, 1986). Dihedral-angle restraints were introduced on the basis of *J*-coupling constants. ϕ dihedral angles of residues with small coupling constants ($^3J_{\alpha NH} < 6.0$ Hz) were constrained between −100° and −20°; those with large coupling constants ($^3J_{\alpha NH} > 8.0$ Hz) were constrained between −160° and −80°; restraints were not introduced for intermediate *J* values. χ_1 dihedral angles, when obtained (Figure 5; part II of Results) were restrained about the preferred rotamer with a tolerance of ±40°. Experimental restraints are provided as supplementary material.

Structure Calculations. Distance-geometry/simulated annealing (DG/SA) calculations were performed with the program DG-II (Havel, 1991). Empirical energies and deviations from ideal covalent geometry (Table 2) were analyzed using the programs X-PLOR (A. T. Brunger, Yale University) and Quanta (Molecular Simulations, Inc., Waltham, MA). Ring-current shifts were calculated as described (Weiss & Hoch, 1987).

Hydrogen-Bond Restraints. Hydrogen bonds were assigned to seven slowly exchanging amide protons (residues 16, 20, 33, 35, 38, 48, and 52) on the basis of preliminary DG structures (Qian *et al.*, 1992, 1993). These sites occur within α-helices; in each case a unique (*i*,*i*+4) C=O acceptor was identified in 20/20 structures. Although slowly exchanging amide protons were also observed at the termini of helices 1 and 2 (residues 15 and 38, respectively), in preliminary DG calculations these participate in either (*i*,*i*+3) or (*i*,*i*+4) hydrogen bonds and were hence not restrained. To avoid systematic DG errors at these and other α-helical sites, lower bound corrections to default hard-sphere radii were introduced as proposed by Chazin and co-workers (Akke *et al.*, 1992).

```
                    helix 1          helix 2          helix 3
Antp  RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN
en    EKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFQNKRAKIKKST
Oct1  RRKKRTSIETNIRVALEKSFLENQKPTSEEITMIADQLNMEKEVIRVWFCNRRQKEKRIN
Oct2  RRKKRTSIETNVRFALEKSFLANQKPTSEEILLIAEQLHMEKEVIRVWFCNRRQKEKRIN
                  10        20        30        40        50        60
```

FIGURE 1: Comparison of amino acid sequences of canonical homeodomains Antennapedia (Antp) and Engrailed (en) with Oct-1 and Oct-2 POU_{HD} sequences. Positions of three α-helices in NMR structures of Antennapedia (Otting *et al.*, 1988, 1990) and the X-ray structure of the Engrailed complex (Kissinger *et al.*, 1990) are indicated; the dashed line in the C-terminal region of α-helix 3 indicates the position of helix 4 in the free Antennapedia fragment (Otting *et al.*, 1988) and the partially ordered structure in the Oct-3 POU_{HD} (Morita *et al.*, 1993).

RESULTS

(I) Characterization of Conditions. Initial NMR studies were conducted using isolated POU_{HD} (Oct-2 nucleotides 949–1143; 65 residues) and linker–POU_{HD} (Oct-2 nucleotides 877–1143; 89 residues) fragments as described (Botfield *et al.*, 1992). Their ^{1}H-NMR spectra exhibited similar dispersion of chemical shifts at pH 6.0, differing only by the absence (POU_{HD}) or presence (linker–POU_{HD}) of narrow resonances at random-coil frequencies. The narrow line widths and limited chemical-shift dispersion of non-POU_{HD} sequences suggest that the linker is disordered in solution. In each case, however, progressive line broadening was observed over 12–24 h. These changes were reversible with addition of 5 mM deuterated dithiothreitol (DTT) and so ascribed to intermolecular disulfide bond formation; POU_{HD} contains one cysteine (corresponding to residue 50 in a consensus homeodomain; Kissinger *et al.*, 1990). Analogous disulfide bond formation has been observed in the Antennapedia HD via a solvent-exposed cysteine (Muller *et al.*, 1988; Otting *et al.*, 1988).

Acquisition of NMR data was further limited by thermal instability. Samples underwent macroscopic precipitation at temperatures >30 °C; partial precipitation is seen after 2–3 days at 25 °C and after 7–10 days at 18 °C. Precipitation is not reversible by DTT and is presumably mediated by aggregation of the unfolded state. Circular dichroism (CD) spectra exhibit progressive loss of α-helix content ($[\theta]_{222}$) at temperatures greater than 25 °C (data not shown). The linker–POU_{HD} fragment was observed to be more soluble than the isolated homeodomain, presumably due to polar and charged residues in the linker. Expression and purification of this fragment was thus optimized as described in Materials and Methods. Inclusion of an N-terminal His_6 tag[3] (MGSSH-HHHHH) was observed not to alter protein solubility, thermal stability, specific DNA binding, or chemical shifts of POU_{HD}-specific resonances (as assigned below). Line widths of the latter increase, however, with increased fragment length, presumably due to an increase in overall correlation time.

The present NMR analysis is restricted to the ordered POU_{HD} moiety (60 residues; Figure 1) of the 99-residue fragment. The one-dimensional ^{1}H-NMR spectrum of this fragment at 25 °C and 500 Mhz is shown in Figure 2. The marked dispersion of chemical shifts, including prominent upfield-shifted aliphatic resonances, provides evidence of ordered structure. Conversely, evidence of flexibility is provided by rapid amide proton exchange in freshly prepared D_2O solution (complete exchange is observed within 1 h at pH 6.0 and 18 °C) and equivalent phenylalanine aromatic resonances (fast ring rotation on the NMR time scale).

[3] The His_6 linker–POU_{HD} polypeptide (99 residues) is numbered according to consensus homeodomain nomenclature (Kissinger *et al.*, 1990). Thus, the C-terminal POU_{HD} moiety is numbered 1–63 (see Figures 1 and 5); the N-terminal non-POU_{HD} residues (not analyzed) are numbered −35 to 0.

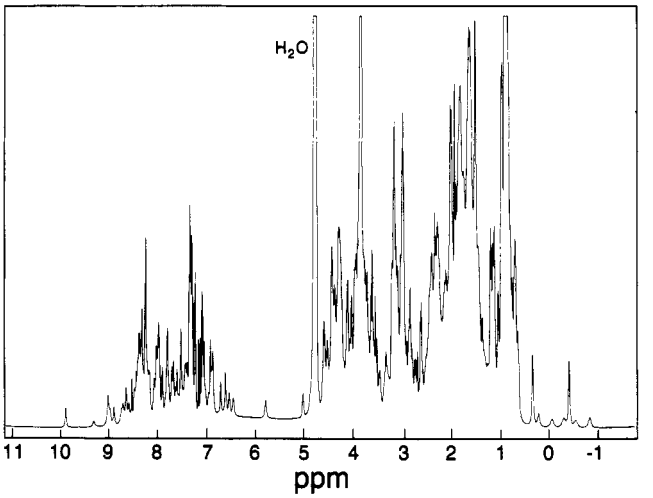

FIGURE 2: One-dimensional ^{1}H-NMR spectrum of the Oct-2 POU_{HD} fragment in H_2O (pH 6.0 and 18 °C). Resonances with significant secondary shifts are assigned to POU_{HD} (Table 1); many additional resonances are observed with random-coil chemical shifts and assigned in adjoining linker residues.

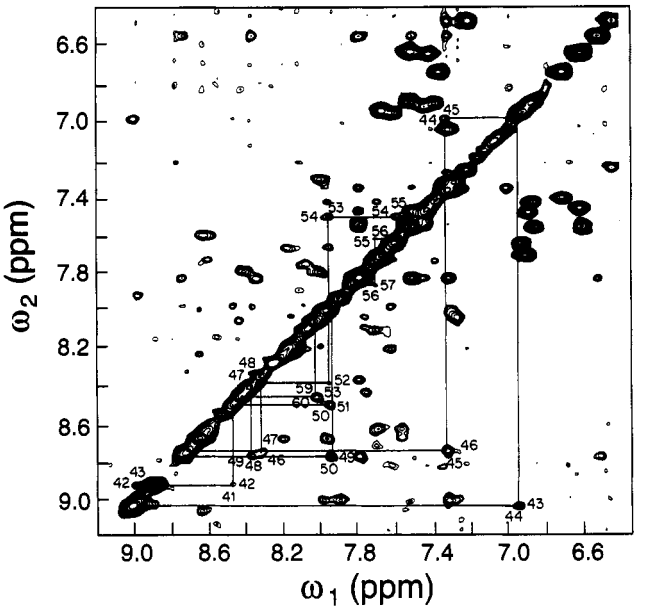

FIGURE 3: Portion of the NOESY spectrum showing d_{NN} NOEs (see Figure 6). Connectivities in α-helix 3 are outlined. The protein was made 1.5 mM in 50 mM potassium phosphate and 5 mM deuterated DTT at 18 °C and pH 6; the mixing time was 200 ms.

(II) NMR Resonance Assignment. 2D NMR spectra were acquired at 18° and 25 °C. Partial assignment of $d_{\alpha N}$ (fingerprint) and d_{NN} regions (Figure 3) of the NOESY spectrum was obtained by standard sequential methods (Wuthrich, 1986). Complete analysis was not possible due to resonance overlap. Assignment of the POU_{HD} moiety was completed by analysis of the ^{15}N-edited 3D NOESY–HMQC

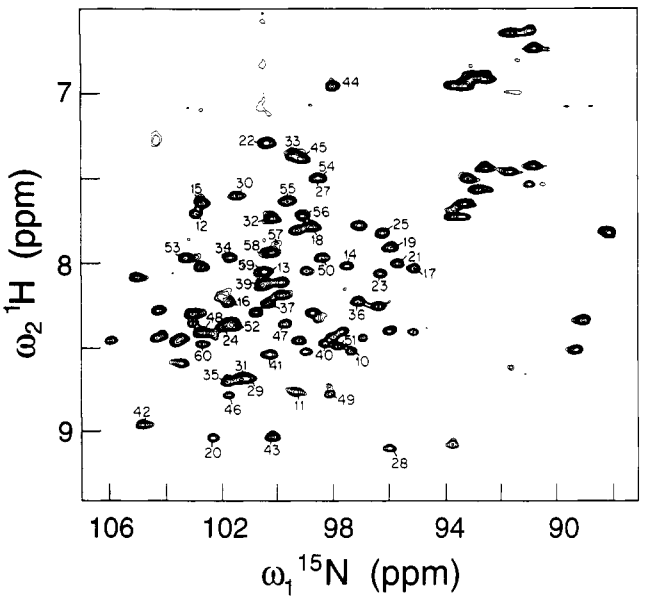

FIGURE 4: ^{1}H–^{15}N HSQC spectrum of the uniformly labeled sample. Solvent suppression was achieved by presaturation. Assignments are as indicated; experimental conditions were as described in Figure 3.

spectrum obtained at 18 °C. In Figure 4 is shown the ^{15}N HSQC spectrum of the uniformly labeled fragment. As observed in other systems, there is no apparent correlation between H_N and ^{15}N chemical shifts, facilitating resolution of NOEs in a 3D NOESY–HMQC experiment. Representative ω_1–ω_3 strips, arranged in order of the amino acid sequence, are provided as supplementary material. ^{1}H and ^{15}N chemical shifts of the ordered POU_{HD} moiety are given in Table 1. Resonance assignment of the Oct-1 homeodomain was obtained in parallel and will be described elsewhere (G. Wagner, M. Mueller, and M. A. Weiss, manuscript in preparation). Homologous ^{1}H-NMR resonances in POU_{HD} and Antennapedia are shifted to high field (asterisks in Table 1), presumably due to corresponding ring-current interactions (see Discussion).

(i) Secondary Structure. The Oct-2 POU_{HD} N-terminal arm (HD residues 1–7) is disordered (as indicated by narrow line widths and absence of nonlocal NOEs), consistent with the solution structure of the Antennapedia HD (Otting *et al.*, 1988; Qian *et al.*, 1989; Billeter *et al.*, 1990). α-Helical segments in POU_{HD}, as identified by short- and medium-range NOEs, span residues 10–21, 28–38, and 42–53. Helix-related d_{NN} NOEs in α-helix 3 (the putative recognition α-helix) are outlined in Figure 3. In Figure 5 are summarized sequential and medium-range NOEs, $^3J_{HN}$ coupling constants, and selected χ_1 dihedral angles based on stereospecific assignment of β protons. The α-helical segments are similar to those of canonical homeodomains. α-Helix 3 is foreshortened; there is no evidence for its continuous extension to residue 60 (as observed in crystal and NMR structures of specific homeodomain–DNA complexes; Otting *et al.*, 1990; Kissinger *et al.*, 1990; Wolberger *et al.*, 1991). In the 54–60 region only two $(i,i+3)$ contacts are observed: weak $d_{\alpha N(i,i+3)}$ NOEs between residues 52 and 55 and between residues 54 and 57. In addition, sequential $d_{\alpha N(i,i+1)}$ NOEs are unusually large at the 54–55, 55–56, and 59–60 steps.

The 54–60 region, although not well ordered, exhibits nonrandom structure. Long-range NOEs are observed between one side chain and α-helix 1 (E56–F20). Moreover, small $^3J_{\alpha N}$ coupling constants (<6 Hz) are observed at residues 54, 56, and 57, suggesting a conformational preference for local α-helical values. Taken together, these data strongly suggest that the 54–60 region is partially ordered with fluctuating secondary structure. Similar observations have been reported in NMR studies of Antennapedia (Otting *et al.*, 1988; Billeter *et al.*, 1990) and Oct-3 POU_{HD} fragment (Morita *et al.*, 1993) as considered further in the Discussion.

(ii) Tertiary Structure. In Figure 6A is shown a diagonal plot of interresidue NOEs; main-chain contacts are summarized at the upper left and side-chain contacts at the lower right. In this format the locations of α-helices are defined by strings of $(i,i+3)$ contacts parallel to the diagonal. Long-range NOEs constrain the interhelical orientation. NOEs are observed between α-helices 1 and 3 (L16–M40, L16–I45, L16–W48, L16–F49, E17–R52, F20–W48, F20–F49, and F20–R52) and between α-helices 2 and 3 (I31–I45, A35–K42). The loop between α-helices 1 and 2 (residues 22–27) also packs against helix 2 (P26–I31) and helix 3 (E24–F49 and P26–F49). Homologous interactions define similar hydrophobic cores in NMR and crystal structures of canonical homeodomains (Qian *et al.*, 1989; Kissinger *et al.*, 1990; Wolberger *et al.*, 1991). Because the residues involved in these interactions are usually conserved (Scott *et al.*, 1989), observation of analogous long-range NOEs provides qualitative evidence of an homologous tertiary structure.

(III) Distance-Geometry Calculations. The program DG-II (Havel, 1991) was used to generate DG/SA models of the POU_{HD} moiety. A total of 75 structures were obtained on the basis of 525 restraints as summarized in Table 2 (12 restraints/residue in the ordered POU_{HD} moiety). Restraints were not introduced for non-POU_{HD} residues nor for HD arm residues 1–7. Of the 75 structures, 50 had a residual error function of less than 1.0 kcal/mol. In this set the mean NOE restraint violation was 0.064 Å; there were no violations >0.37 Å. The final restraint list, provided as supplementary material, contains hydrogen-bond restraints at seven α-helical sites (see Materials and Methods). These sites (residues 16 and 20 in α-helix 1; residues 33, 35, and 38 in α-helix 2; and residues 48 and 52 in α-helix 3) (i) correspond to slowly exchanging amide resonances as identified in rapidly recorded TOCSY spectra and (ii) exhibit unique $(i,i+4)$ hydrogen bonds in preliminary DG models. Statistical parameters, including deviations from ideal covalent geometry and empirical energies, are provided in Table 2.

An ensemble of refined backbone structures is shown in red in Figure 7A (right-hand panel); a corresponding ribbon model is shown at the left. Main-chain ϕ and ψ dihedral angles in the ensemble (Ramachandran plot) are as expected (Figure 8). The precision of the POU_{HD} DG ensemble is shown by residue in Figure 9. The structure is most precisely defined within the three α-helices (Table 2). As illustrated in Figure 7B (above), internal side chains are more precisely defined than external side chains. The loop between α-helices 1 and 2, the turn between HTH helices 2 and 3, and C-terminal residues 54–60 are not as well-defined, reflecting the fewer long-range restraints in these regions. Residues 1–7 are disordered. We emphasize that region-specific RMS deviations reflect the relative numbers of restraints and do not directly provide dynamic information.

The global structure is essentially identical to that of the Engrailed HD in the cocrystal of a specific DNA complex (shown in blue in Figure 7A; Kissinger *et al.*, 1990). Structural conservation extends to the organization of the hydrophobic core; positions of selected internal side chains are shown in Figure 10B (canonical HDs) and Figure 10C (POU_{HD}). RMS differences between the average POU_{HD} structure and this

Table 1: Chemical Shifts[a] for ^{15}N- and ^{1}H-NMR Resonances of the Oct-2 Homeodomain

residue	^{15}N	HN	Hα	Hβ	others
1 Arg					
2 Arg					
3 Lys					
4 Lys					
5 Arg					
6 Thr			4.42	4.31	γCH_3 1.24
7 Ser					
8 Ile		8.59	3.98	1.86	γCH_3 0.78
9 Glu		7.47			
10 Thr	97.5	8.49	3.88	4.16	γCH_3 1.17
11 Asn	99.5	8.71	4.57	2.88	δNH_2 6.91, 7.7
12 Val	103.0	7.74	3.71	2.16	γCH_3 1.07, 0.90
13 Arg	100.3	8.09	3.60	1.89, 1.61	δCH_2 3.15; ϵNH 7.44
14 Phe	97.6	7.98	4.32	3.17, 3.17	δH 7.35; ϵH 7.38; ζH 7.35
15 Ala	102.9	7.66	4.03	1.51	
16 Leu	101.9	8.19	3.61	0.69, -0.91*	γH 1.14; δCH_3 0.38, -0.39*
17 Glu	95.2	8.09	4.32		
18 Lys	98.8	7.82	4.02	1.82, 2.05	
19 Ser	95.9	7.92	4.28	4.17	
20 Phe	102.2	8.99	4.49	3.38, 3.32	δH 7.32; ϵH 7.32; ζH 7.02
21 Leu	95.8	8.05	4.05	1.82, 1.96	
22 Ala	100.5	7.23	4.29	1.52	
23 Asn	96.5	8.05	4.43	2.81, 2.81	δNH_2 6.63, 7.55
24 Gln	102.0	8.36	4.44		γCH_2 2.62
25 Lys	96.3	7.84	4.55	1.31, 1.50	
26 Pro			4.59	1.80, 1.60	γCH_2 0.99, 0.24*; δCH_2 3.02, 3.00
27 Thr	98.5	8.66	4.05	3.98	γCH_3 1.19
28 Ser	96.0	9.05	4.30	4.05, 3.93	
29 Glu	101.6	8.62	4.05		
30 Glu	101.8	7.60	3.93		
31 Ile	101.7	8.62	3.50	1.95	γCH_2 1.82, 1.50; γCH_3 0.89; δCH_3 0.66
32 Leu	100.2	7.70	3.92	1.89, 1.72	
33 Leu	99.6	7.41	4.13	1.83, 1.65	
34 Ile	101.9	7.99	3.69	1.85	γH 1.65, 0.91; γCH_3 0.81; δCH_3 0.72
35 Ala	101.9	8.66	3.70	1.52	
36 Glu	97.3	8.22	4.09		
37 Glu	100.4	8.18	4.01	2.07, 2.28	γH 2.57; δNH_2 6.72, 7.39
38 Leu	102.0	8.03	4.28	1.45, –	
39 His	100.7	8.19	4.60	3.02, 3.16	δ^2H 7.25; ϵ^1H 8.35
40 Met	98.3	8.42	4.91	1.50, 1.50	γCH_2 2.40, 2.47
41 Glu	100.3	8.51	4.31	1.97	
42 Lys	104.8	8.90	3.80	1.81, 1.91	
43 Glu	100.2	9.01	4.39	2.15, 1.98	γCH_2 2.41, –
44 Val	98.0	6.97	3.73	2.31	γCH_3 1.12, 1.00
45 Ile	99.2	7.34	3.89	2.34	γCH_2 1.22, 1.52; γCH_3 0.82; δCH_3 0.71
46 Arg	101.9	8.72	4.02	2.15, –	γCH_2 1.99, –; δCH_2 2.32
47 Val	99.8	8.32	3.62	2.18	γCH_3 1.09, 0.95
48 Trp	102.9	8.38	3.98	3.28, 3.53	δ^1H 7.10; ϵ^3H 6.53; ϵ^1NH 9.9; ζ^2H 7.22; ζ^3H 5.8; η^2H 6.45
49 Phe	98.2	8.73	3.90	3.36, 3.36	δH 7.43; ϵH 7.79; ζH 7.49
50 Cys	98.4	7.98	4.12	3.18, 3.05	
51 Asn	97.9	8.49	4.35	2.6, 2.45	δNH_2 6.92, 7.69
52 Arg	101.9	8.37	3.56	-0.3, 0.78*	γCH_2 –0.02, –0.51*; δCH_2 2.45, 2.12*; ϵNH 9.40
53 Arg	103.7	7.99	4.20		
54 Gln	98.5	7.50	4.13	2.40, 2.15	
55 Lys	99.7	7.62	4.30	1.77, 1.77	
56 Glu	99.0	7.72	4.20	2.22, 2.05	γCH_2 2.40, –
57 Lys	100.1	7.86	4.22	1.85, 1.85	γCH_2 1.50, –
58 Arg	100.2				
59 Ile	100.7	8.05	4.14	1.84	γCH_2 1.20, –; γCH_3 0.89
60 Asn	102.8	8.47	5.05	2.86, 2.68	
61 Pro			4.45	2.27, 2.27	γCH_2 1.94, –; δCH_2 3.93, 3.85
62 Cys					
63 Ser					

[a] ^{15}N chemical shifts (18 °C) were measured relative to $^{14}NH_4{}^{15}NO_3$ in D_2O (376.25 ppm); ^{1}H chemical shifts (25 °C) were measured relative to the H_2O resonance, presumed to be 4.78 ppm at pH 6.0 and 25 °C. Underlined values indicate stereospecifically assigned methylene resonances (listed in order of $\beta 2$ and $\beta 3$) and methyl resonances [listed in order of $\gamma 1$ and $\gamma 2$ (Val) or $\delta 1$ and $\delta 2$ (Leu)]. Values with an asterisk indicate sites of significant secondary shift (Wuthrich, 1986).

and other canonical HDs are provided as supplementary material. The overall pattern of *buried nonpolar groups* and *exposed polar or charged groups* is similar to that of the Engrailed HD (calculated with DNA removed from the cocrystal structure; Kissinger *et al.*, 1990). This pattern leads to a favorable empirical solvation free energy change (–31 ± 1 kcal/mol) for the POU_{HD} ensemble as calculated by the method of Eisenberg and McLachlan (1986). This value is similar to that of Engrailed (–29 kcal/mol), Antennapedia (–39 kcal/mol), and MATα2 (–34 kcal/mol).

As anticipated by analysis of secondary structure (part II above), POU_{HD} and canonical HDs differ in C-terminal

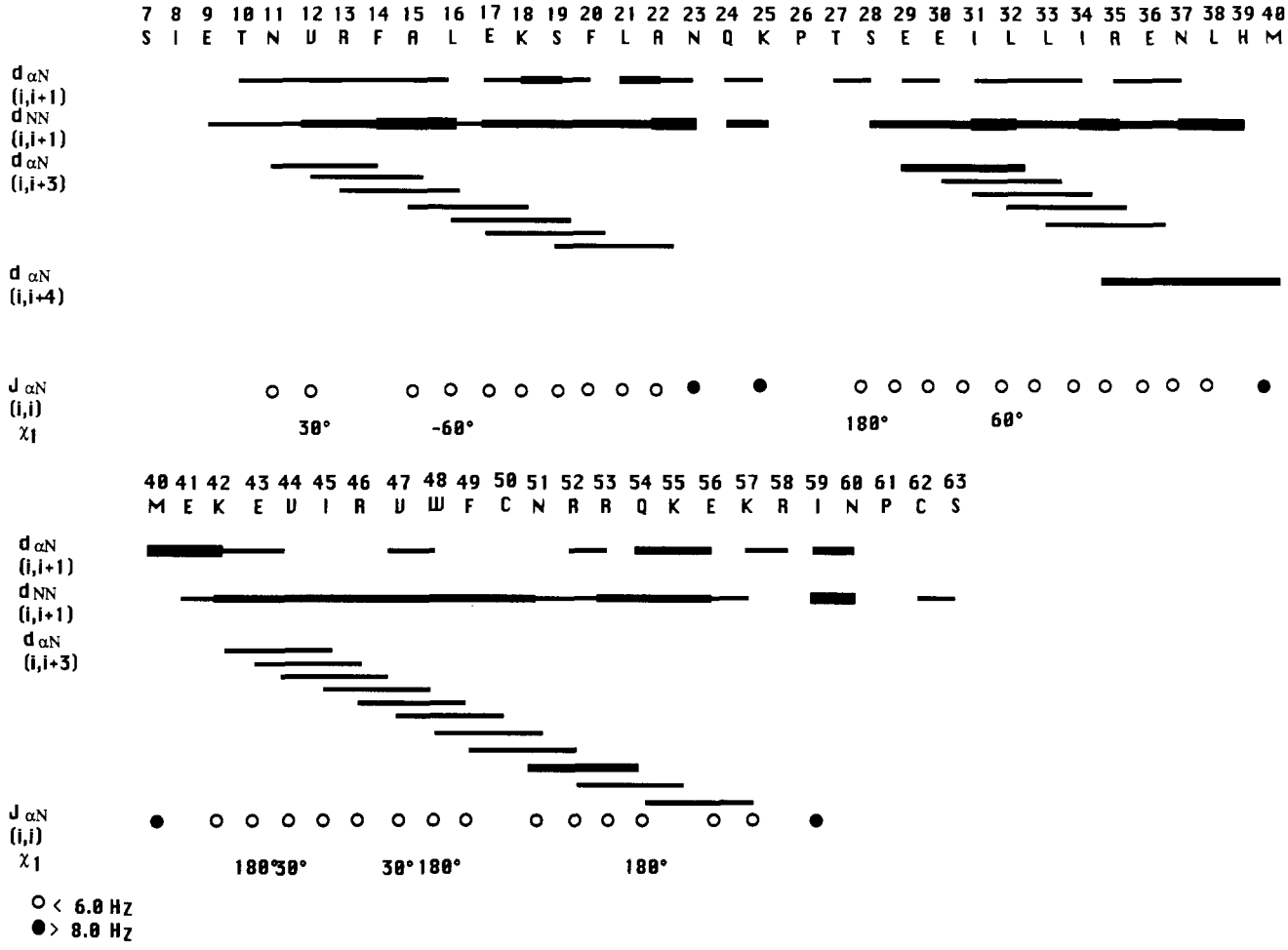

FIGURE 5: Summary of sequential and medium-range NOEs, $^3J_{\alpha N}$ coupling constants, and selected χ_1 angles. Within each NOE class, relative intensity is schematically indicated by the thickness of the line. Large (>8 Hz) and small (<6 Hz) $^3J_{\alpha N}$ are as indicated (● and ○, respectively).

residues 54–60 as shown in expanded form in Figure 7B. Internal side chains I45, W48, and F49, which are invariant among homeodomain sequences (Scott *et al.*, 1989), adopt similar configurations in POU_{HD} and Engrailed. External side chains C50 and N51 are not well-defined in the NMR ensemble (green) but are near the positions of corresponding side chains (Q50 and N51) in the Engrailed crystal structure (white). The latter side chains make base-specific hydrogen bonds in the DNA complex (Kissinger *et al.*, 1990). We have recently demonstrated that substitutions of N51 in POU_{HD} lead to relaxed specificity of the Oct-2 POU domain at one position (underlined) in the consensus octamer (5′-ATG-CAAAT), strongly suggesting that POU_{HD} and Engrailed HD exhibit similar mechanisms of DNA binding (Botfield *et al.*, 1994). Although residue 50 contributes to the specificity of canonical homeodomains (Desplan *et al.*, 1988), the role of C50 in POU-DNA recognition is not well understood.

DISCUSSION

The HTH motif was first described in prokaryotic gene-regulatory proteins (Anderson *et al.*, 1981; McKay & Steitz, 1981; Pabo & Lewis, 1982). Its extension to eukaryotic proteins was inferred in sequences of homeodomains (Laughon & Scott, 1984; Shepherd *et al.*, 1984), which define a broad class of gene-regulatory proteins [for review, see Scott *et al.* (1989)]. Structures have been determined of two isolated HDs (Antennapedia and variant HD LFB1/HNF1; Qian *et al.*, 1989; Leiting *et al.*, 1993; Ceska *et al.*, 1993) and threespecific HD–DNA complexes (Engrailed, Antennapedia, and MATα2; Kissinger *et al.*, 1990; Otting *et al.*, 1990; Wolberger *et al.*, 1990). These structures exhibit a novel architecture consisting of three α-helices and a flexible N-terminal arm. α-Helices 2 and 3 define an HTH motif. The HD HTH binds in the major groove but with altered orientation relative to that of prokaryotic DNA-binding proteins (Anderson *et al.*, 1985, 1987; Jordan & Pabo, 1988; Aggarwal *et al.*, 1988; Wolberger *et al.*, 1988). The N-terminal arm binds in the minor groove. Canonical homeodomains constitute autonomous DNA-binding elements that recognize specific sites with nanomolar affinity [Desplan *et al.*, 1988; see Scott *et al.* (1989) for review]. Their specificity and affinity in biological control mechanisms may nevertheless be influenced by interactions with other transcription factors in the preinitiation complex; an outstanding example is the cooperative assembly of Oct-1 with viral and host factors at regulatory sites of herpes simplex virus (Kristie *et al.*, 1989; Stern *et al.*, 1989; Kristie & Sharp, 1990). A crystal structure of a specific Oct-1 POU–DNA complex has recently been described (Klemm *et al.*, 1994; see Added in Proof).

Oct-2, a B-cell and neural transcription factor (Singh *et al.*, 1986; Clerc *et al.*, 1988; Ko *et al.*, 1988; Müller-Immerglück *et al.*, 1988; Scheidereit *et al.*, 1988), belongs to the POU family [for review, see Verrijzer and Van der Vliet (1993)]. Unlike a canonical HD, the POU-specific homeodomain is part of a bipartite DNA-binding motif [for review, see Verrijzer and Van der Vliet (1993)]: two distinct domains, POU_{HD} and POU_S, are jointly required for high-affinity DNA recognition. Isolated POU_{HD} or POU_S fragments exhibit weak or absent DNA binding, respectively. The solution structure

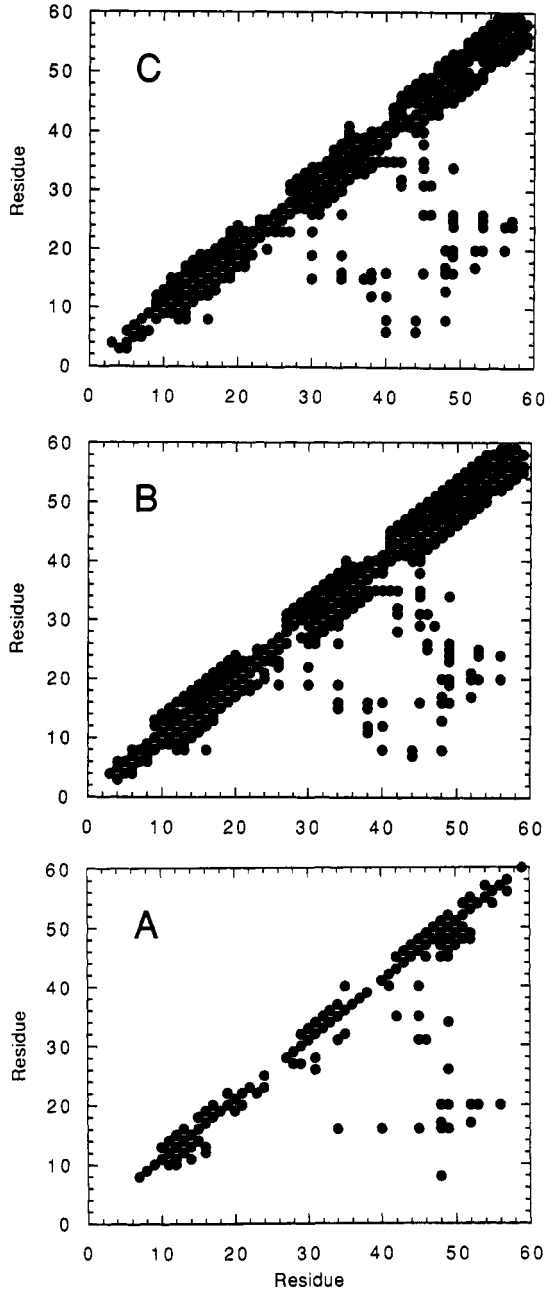

FIGURE 6: (A) Diagonal plot of interresidue NOEs assigned in the spectrum of Oct-2 POU_{HD}. Backbone–backbone NOEs are shown above the diagonal; backbone–side-chain and side-chain–side-chain NOEs are shown below the diagonal. (B) Diagonal plot of interresidue contacts observed in the DG/SA ensemble. Contacts are shown if present in each structure. (C) Diagonal plot of interproton contacts inferred from the crystal structure of the Engrailed homeodomain–DNA complex (Kissinger *et al.*, 1990).

of a POU_S fragment 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 solution structure of a POU_{HD} fragment and its comparison to canonical homeodomains (Qian *et al.*, 1988; Otting *et al.*, 1990; Kissinger *et al.*, 1990; Wolberger *et al.*, 1991). A related study of POU_{HD} DNA-binding properties has recently been published (Botfield *et al.*, 1994).

(I) POU_{HD} Exhibits a Canonical Fold with Shared ^{1}H-NMR Chemical Shifts. The solution structure of the Oct-2 POU_{HD} is remarkably similar to that of canonical HDs Engrailed and Antennapedia (Figure 7A; RMS values provided as supplementary material). In the well-defined-structural moiety (residues 8–52) the Oct-2 POU_{HD} and the

Table 2: NMR Restraints and Statistical Parameters

(A) Summary of Restraints

intraresidue NOE	174	total NOE restraints	457
interresidue NOE	283		
strong	13	dihedral angle ϕ	44
medium	35	dihedral angle χ_1	10
weak	235	hydrogen bonds	14
		total restraints	525

(B) RMS Deviations[a]

main chain	helix 1	0.92	side chain	helix 1	1.60
	helix 2	0.98		helix 2	1.58
	helix 3	0.76		helix 3	1.40
	core	0.79		core	1.09
	surface	1.28		surface	2.26
	total	1.15		total	2.09

(C) Average Restraint Violations

NOE	0.064 Å	dihedral angles	0.34°

(D) Deviation from Ideal Covalent Geometry

bond lengths	0.018 Å	bond angles	2.75°

(E) Restrained Emperical Energy Function[b]

constrained dihedral angle	1.05 kcal/mol
NOE restraint energy	24.21 kcal/mol
van der Waals	−141.94 kcal/mol
hydrogen bonds	−25.96 kcal/mol
improper dihedral angles	197.77 kcal/mol
dihedral angles	540.60 kcal/mol
covalent bond lengths	90.11 kcal/mol
bond angles	255.59 kcal/mol
electrostatics	3.81 kcal/mol

[a] DG structures were aligned on the backbone atoms of the three α-helices. Helix 1 is defined as residues 11–20; helix 2, residues 28–38; and helix 3, residues 42–53. Core is defined as residues with a solvent accessibility of less than 30% with respect to the random-coil value. Total is defined as residues 10–59. [b] The restrained NOE force constant was 30 kcal/Å^2; the restrained dihedral force constant was 30 kcal/rad^2.

Engrailed HD exhibit 20% sequence identity; POU_{HD} and the Antennapedia HD exhibit 30% sequence identity (Figure 1). Invariant residues contribute to the key helix–helix contacts, including L16, W48, and F49 (Figure 10). In Figure 6C is shown a diagonal plot of inferred interproton contacts obtained from the Engrailed cocrystal structure (Kissinger *et al.*, 1990). Comparison with resolved NOEs (Figure 6A) reveals a similar overall pattern. The relative paucity of NOEs suggests, however, that a large number of potential contacts in POU_{HD} have not been identified at the present stage of analysis. Nevertheless, the majority of such "missing" contacts are geometrically required by the present restraints. In Figure 6B is shown the complete set of interproton contacts predicted by the POU_{HD} ensemble (i.e., interproton distances less than 4.5 Å in each DG/SA structure). This set exhibits more complete agreement with inferred crystal contacts (panel B).

This common HD framework leads to similarities in the relative orientations of aromatic rings in the hydrophobic core. Conservation of ring orientations among homologous proteins predicts in turn similar patterns of aromatic ring currents and hence patterns of anomalous ^{1}H-NMR chemical shifts (Lee *et al.*, 1992). This is indeed observed. In Table 3 are given secondary shifts (defined as the difference between observed chemical shift and tabulated random-coil shift; Wuthrich, 1986) of three residues which exhibit unusual upfield resonances (asterisks in Table 1). The correspondence between such shifts in Oct-2 POU_{HD} (left-hand column) and Antennapedia (right-hand column) is striking. In Table 4 are given predicted ring-current shifts based on the POU_{HD} ensemble. In each case the aromatic ring responsible for the upfield resonance frequency is either W48 or F49 (or both). The chemical shift of L16 provides a sensitive marker for the

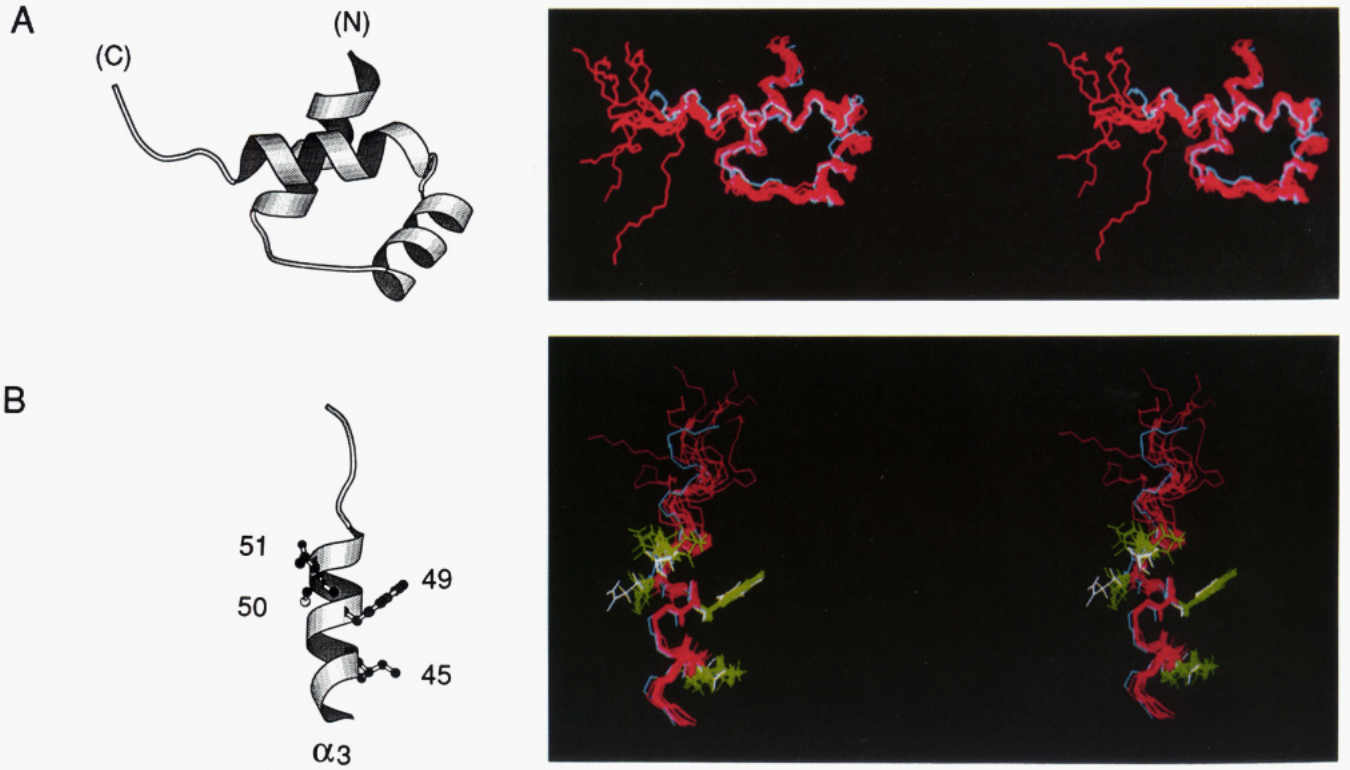

FIGURE 7: (A) Right: DG/SA ensemble of Oct-2 POU_{HD} (red) and the crystal structure of the Engrailed (blue; the blue line appears white when overlapping red). Left: corresponding ribbon model in a similar orientation; N- and C-termini are as indicated. (B) Right: DG/SA models of α-helix 3 of Oct-2 POU_{HD} (red) and Engrailed (blue). POU_{HD} side chains I45, F49, C50, and N51 are shown in green; Engrailed side chains I45, F49, Q50, and N51 are shown in white. Left: corresponding ribbon model with side chains labeled.

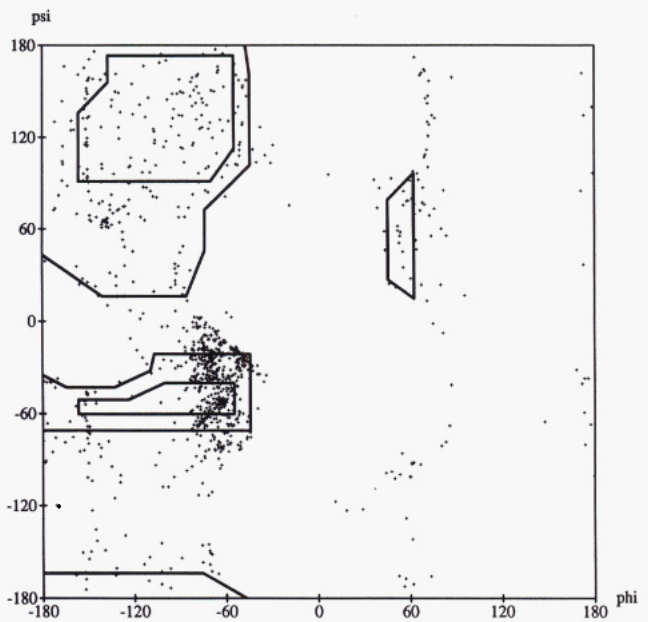

FIGURE 8: Ramachandran plot of the POU_{HD} DG/SA ensemble; favorable regions are outlined.

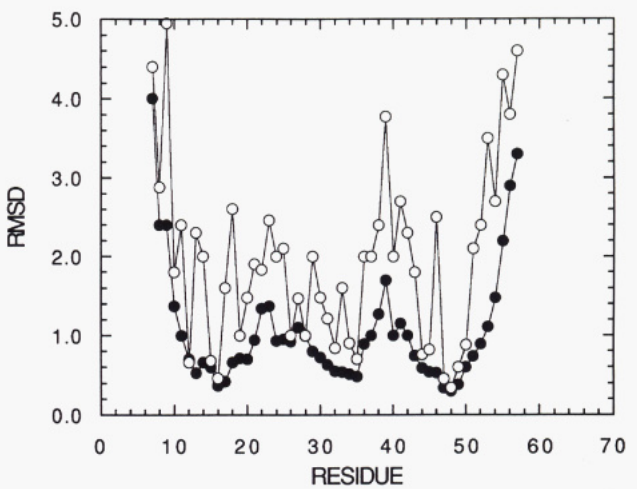

FIGURE 9: RMS deviations in the DG/SA ensemble shown by residue for backbone (●) and side-chain (○) atoms.

packing of α-helix 1 against α-helix 3 (Figure 10); the chemical shift of residue 26 reflects the orientation of the loop between helices 1 and 2 against helix 3, and the chemical shift of R52 reflects the stable orientation of the W48 and R52 side chains in helix 3. It is of interest to note that imprecision of the present POU_{HD} DG ensemble leads to significant variations in predicted ring-current shifts (RMS values are given in column 4 of Table 4). This is particularly apparent for P26 (located in the loop between α-helices 1 and 2), whose actual secondary shifts are significantly larger than the ensemble average. It is likely that the configuration of this loop is better ordered in solution than is suggested by the range of DG structures; i.e., at the present level of resolution, imprecision is due in large measure to insufficient identification of restraints. A correlation between invariant framework residues and ^{1}H-NMR chemical shifts has previously been noted among classical Zn fingers (Lee *et al.*, 1992). Such correlations provide a useful guide for assignment of homologous proteins and a convenient assay for structural perturbations among mutant proteins.

(II) The HTH Recognition α-Helix Exhibits Differential C-Terminal Stability. In NMR and X-ray structures of specific HD–DNA complexes the HTH recognition α-helix is continuous (residues 42–60; Otting *et al.*, 1990; Kissinger *et al.*, 1990; Wolberger *et al.*, 1991; Billeter *et al.*, 1993).

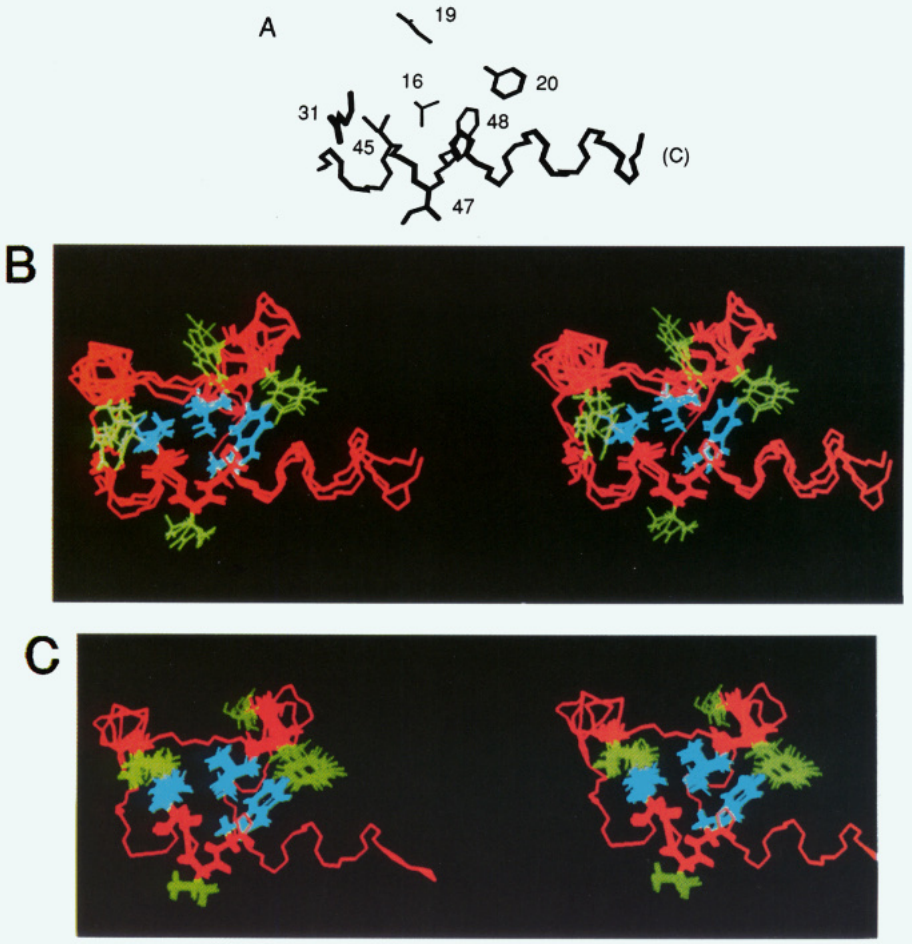

FIGURE 10: (A) Selected internal side chains and backbone (residues 40–59) of the Engrailed HD. (B) Stereo representation of canonical homeodomains (Engrailed, Antennapedia, and MATα2). The backbone (residues 10–59) is shown in red; side chains 16, 45, and 48 are in blue; and side chains 19, 20, 31, and 47 are in green. (C) Corresponding stereoview of Oct-2 POU_{HD} with average backbone (red) and an ensemble of well-ordered side chains in hydrophobic core (blue and green).

Table 3: Secondary Shifts in POU_{HD} and Antennapedia[a]

POU_{HD}			Antennapedia		
residue	proton	$\Delta\sigma$	residue	proton	$\Delta\sigma$
L16	H_α	0.76	L16	H_α	0.92
L16	$H_{\beta 1,2}$	0.96, 2.56	L16	$H_{\beta 1,2}$	1.22, 2.86
L16	H_γ	0.50	L16	H_γ	0.61
L16	$\delta_{1,2}CH_3$	0.56, 1.29	L16	$\delta_{1,2}CH_3$	0.53, 1.59
P26	H_α	–0.15	L26	H_α	–0.12
P26	$H_{\beta 1,2}$	0.42, 0.48	L26	$H_{\beta 1,2}$	0.21, 0.50
P26	$H_{\gamma 1,}^{2}$	1.04, 1.79	L26	H_γ	1.04
P26	$H_{\delta 1,2}$	0.66, 0.65	L26	$\delta_{1,2}CH_3$	0.80, 0.53
R52	H_α	0.82	R52	H_α	0.88
R52	$H_{\beta 1,2}$	1.11, 2.09	R52	$H_{\beta 1,2}$	1.14, 2.27
R52	$H_{\gamma 1,2}$	1.72, 2.21	R52	$H_{\gamma 1,2}$	1.92, 2.34
R52	$H_{\delta 1,2}$	0.87, 1.10	R52	$H_{\delta 1,2}$	0.97, 1.37

[a] Secondary shifts ($\Delta\sigma$) are defined as the difference between observed chemical shifts and random-coil values given by Wuthrich (1986). Positive (negative) difference values indicate upfield (downfield) resonance frequencies. POU_{HD} chemical shifts at pH 6.0 and 25 °C are given in Table 1; Antennapedia chemical shifts are given at pH 4.3 and 20 °C in Billeter *et al.* (1990).

Table 4: Calculated Ring-Current Shifts[a] in the POU_{HD} Ensemble

residue	proton	RCS	RMSD	ring current
L16	H_α	0.23	0.11	W48
L16	$H_{\beta 1}$	**1.71**	0.10	W48
L16	$H_{\beta 2}$	**1.94**	0.59	W48 and F49
L16	H_γ	**1.76**	0.21	W48 and F49
L16	δ_1CH_3	**0.51**	0.21	W48 and F49
L16	δ_2CH_3	**1.10**	0.11	W48 and F49
P26	H_α	0.06	0.08	F49
P26	$H_{\beta 1}$	0.08	0.18	F49
P26	$H_{\beta 2}$	0.12	0.32	F49
P26	$H_{\gamma 1}$	**0.60**	0.45	F49
P26	$H_{\gamma 2}$	0.22	0.23	F49
P26	$H_{\delta 1}$	0.11	0.08	F49
P26	$H_{\delta 2}$	0.19	0.12	F49
R52	H_α	**0.69**	0.18	W48
R52	$H_{\beta 1}$	**0.72**	0.51	W48
R52	$H_{\beta 2}$	**0.88**	0.61	W48
R52	$H_{\gamma 2}$	**1.18**	0.76	W48
R52	$H_{\gamma 2}$	**0.88**	0.69	W48
R52	$H_{\delta 1}$	**1.50**	1.24	W48
R52	$H_{\delta 2}$	**0.83**	0.76	W48

[a] Predicted ring-current shifts (RCS) >0.5 ppm are shown in boldface; the predominant source is given in column 5.

NMR analysis of the free Antennapedia HD suggested, however, that α-helix 3 consists of two parts: a well-ordered α-helix (residues 42–52) and an adjoining "α-helix 4" of intermediate stability (residues 55–60). As observed at 10 °C and pH 4.5, Antennapedia residues 55–60 exhibit strong d_{NN} and $d_{\alpha N(i,i+3)}$ NOEs characteristic of α-helix but are distinguished from α-helix 3 (residues 42–52) by (i) the simultaneous presence of strong $d_{\alpha N}$ NOEs and (ii) the absence of slowly exchanging amide resonances in D_2O. Antennapedia residues 52 and 53 exhibit an interruption in (i,i+3) NOEs (Otting *et al.*, 1988) and so were originally interpreted as a "kink". Absence of stable helical structure between residues 53 and 59 has also been reported for the Oct-3 POU_{HD} studied

at 30 °C (Morita *et al.*, 1993). Although solution structures were not obtained, analysis of secondary structure revealed only two weak $d_{\alpha N(i,i+3)}$ NOEs in the C-terminal region (52–55 and 56–59). Moreover, selected $^3J_{\alpha N}$ couplings were found to be anomalously large: >6 Hz (residue 56) or >7 Hz (residues 54 and 57). These data indicate that the C-terminal region of Oct-3 POU_{HD} (at 30 °C and pH 5.2) is substantially less than that of the Antennapedia HD (at 10 °C and pH 4.5).

The present study of Oct-2 POU_{HD} (18 and 25 °C and pH 6.0) is intermediate between the previous studies. Like Oct-3, only two weak $d_{\alpha N(i,i+3)}$ NOEs are observed. Unlike Oct-3, $^3J_{\alpha N}$ couplings are predominantly less than 6 Hz, implying that α-helical ϕ angles are significantly populated under the conditions of this study. Nevertheless, there is no evidence for a well-defined kink or discrete helix 4 suggested for Antennapedia (Qian *et al.*, 1989; Billeter *et al.*, 1990). The extent to which these structural features reflect differences in solution conditions has not been investigated. The structure and dynamics of the homeodomain recognition α-helix under physiological conditions (37 °C and pH 7.4) would be of central biological interest. Since the C-terminal region is well ordered in a specific DNA complex (Klemm *et al.*, 1994; see Added in Proof), there must be an entropy penalty (ΔS) incurred by its DNA-dependent stabilization. If POU-specific homeodomains exhibit less stable recognition helices, this penalty could contribute to the reduced specific DNA affinity (ΔG) of isolated POU_{HD} fragments relative to canonical homeodomains (Verrijzer & Van der Vliet, 1993). Design of a "flawed" POU-specific homeodomain may contribute to gene regulation by enforcing a requirement for coordinate POU_S–DNA recognition.

ADDED IN PROOF

The crystal structure of a specific complex between the Oct-1 POU domain and a consensus octamer site (5′-GTATGCAAATAAGG) has recently been determined by Pabo and colleagues (Klemm *et al.*, 1994). The bound-state structure of Oct-1 POU_{HD} is similar to that of the Engrailed homeodomain (Kissinger *et al.*, 1990) in overall accord with the present NMR study. Unlike the solution structure of Oct-2 POU_{HD}, the crystal structure exhibits at 3-Å resolution a well-ordered HTH recognition α-helix. Asn51 (residue 10 in POU_{HD} helix 3) contacts an adenine (5′-ATGCAAAT) in accord with the results of site-directed mutagenesis (Botfield *et al.*, 1994). The sequence specificity of POU_{HD} is most stringent at this position. *This and other DNA contacts are mediated by that portion of helix 3 which is well ordered in the POU_{HD} solution structure.* Strikingly, contacts by side chains in the C-terminal portion (Q54, K55, and K57) are not well visualized; corresponding contacts in cocrystal structures of canonical homeodomains (Engrailed and MATα2) are by contrast well-defined (Kissinger *et al.*, 1990; Wolberger *et al.*, 1991). There thus appears to be a correspondence between segmental flexibility in the isolated fragment and the definition of protein–DNA contacts in the electron-density map.

ACKNOWLEDGMENT

We thank Prof. G. Wagner and M. Mueller for generously providing Oct-1 POU_{HD} resonance assignments prior to publication; A. S. Stern and J. C. Hoch for ring-current shift calculations; P. A. Sharp and J. LeBowitz (MIT, Cambridge, MA) for Oct-2 cDNA; E. Collins and W. Jia for technical assistance; Q. X. Hua, J. P. Lee, and X. Qian for assistance with NMR measurements; T. F. Havel for the program DG-II; A. T. Brunger for the program X-PLOR; and Y. Kyogoku, C. O. Pabo, H. Singh, K. Struhl, and S. Tabor for helpful discussions.

SUPPLEMENTARY MATERIAL AVAILABLE

One figure showing the 3D NOESY–HMQC spectrum and two tables providing the DG restraint list and RMS differences between Oct-2 POU_{HD} and canonical homeodomains (12 pages). Ordering information is given on any current masthead page.

REFERENCES

Aggarwal, A. K., Rogers, D. W., Drottar, M., Ptashne, M., & Harrison, S. C. (1988) *Science 242*, 899–907.
Akke, M., Drakenberg, T., & Chazin, W. (1992) *Biochemistry 31*, 1011–1020.
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.
Anderson, W. F., Ohlendorf, D. H., Takeda, Y., & Matthews, B. W. (1981) *Nature 298*, 754–758.
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. (1990) *J. Mol. Biol. 155*, 183–197.
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*, 8113–8121.
Ceska, T. A., Lamers, M., Monaci, P., Nicosia, A., Cortese, R., & Suck, D. (1993) *EMBO J. 12*, 1805–1810.
Clerc, R. G., Corcoran, L. M., LeBowitz, J. H., Baltimore, D., & Sharp, P. A. (1988) *Genes Dev. 2*, 1570–1581.
Clore, G. M., & Gronenborn, A. M. (1991) *Prog. NMR Spectrosc. 23*, 43–92.
Dekker, N., Cox, M., Boelens, R., Verrijer, C. P., van der Vliet, P. C., & Kaptein, R. (1993) *Nature 362*, 852–855.
Desplan, C., Theis, J., & O'Farrel, P. H. (1988) *Cell 54*, 1081–1090.
Eisenberg, D., & McLachlan, A. D. (1986) *Nature 319*, 199–203.
Finney, M., Ruvkun, G., & Horvitz, H. R. (1988) *Cell 55*, 757–769.
Fried, M., & Crothers, D. M. (1981) *Nucleic Acids Res. 9*, 6505–6525.
Havel, T. F. (1991) *Prog. Biophys. Mol. Biol. 56*, 43–78.
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.* (in press).
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., 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.
Kochoyan, M., Havel, T. F., Nguyen, D., Dahl, C. E., Keutmann, H. T., & Weiss, M. A. (1991) *Biochemistry 30*, 3371–3386.
Kristie, T. M., & Sharp, P. A. (1990) *Genes Dev. 4*, 2383–2396.
Kristie, T. M., LeBowitz, J. H., & Sharp, P. A. (1989) *EMBO J. 8*, 4229–4238.

Laughon, A., & Scott, M. P. (1984) *Nature 310*, 25–31.
Lee, M. S., Palmer, A. G., III, & Wright, P. E. (1992) *J. Biomol. NMR 2*, 307–322.
Leiting, B., De Francesco, R., Tomei, L., Cortese, R., Otting, G., & Wuthrich, K. (1993) *EMBO J. 12*, 1797–1803.
Li, P., He, X., Gerrero, M. R., Mok, M., Aggarwal, A., & Rosenfeld, M. G. (1993) *Genes Dev. 7*, 2483–2496.
Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989) *J. Magn. Reson. 85*, 393–399.
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.
Montelione, G. T., & Wagner, G. (1989) *J. Am. Chem. Soc. 111*, 5484–5475.
Morita, E. H., Shirakawa, M., Hayashi, F., Imagawa, M., & Kyogoku, Y. (1993) *FEBS Lett. 321*, 107–110.
Muller, M., Affolter, M., Leupin, W., Otting, G., Wuthrich, K., & Gehring, W. J. (1988) *EMBO J. 7*, 4299–4304.
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., Muller, M., Affolter, M., Gehring, W., & Wuthrich, K. (1988) *EMBO J. 7*, 4305–4309.
Otting, G., Qian, Y.-Q., Billeter, M., Müller, M., Affolter, M., Gehring, W., & Wuthrich, K. (1990) *EMBO J. 9*, 3085–3092.
Qian, X., & Weiss, M. A. (1992) *Biochemistry 31*, 7463–7476.
Qian, X., Gozani, S., Yoon, H.-S., Jeon, C., Agarwal, K., & Weiss, M. A. (1993) *Biochemistry 32*,
4–9959.
Qian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W. J., & Wuthrich, K. (1989) *Cell 59*, 573–580.
Redfield, C., & Dobson, C. M. (1990) *Biochemistry 29*, 7201–7214.
Scheidereit, C., Cromlish, J. A., Gerster, T., Kawakami, K., Balmaceda, C.-G., Currie, R. A., & Roeder, R. G. (1988) *Nature 336*, 551–557.
Scott, M. P., Tamkun, J., & Hartzell, G. W., III (1989) *Biochim. Biophys. Acta 989*, 25–48.
Shepherd, J. C. W., McGinnis, W., Carrasco, A. E., DeRobertis, E. M., & Gerhing, W. J. (1984) *Nature 310*, 70–71.
Singh, H., Sen, R., Baltimore, D., & Sharp, P. A. (1986) *Nature 319*, 154–158.
Smith, L. J., Sutcliffe, M. J., Redfield, C., & Dobson, C. M. (1991) *Biochemistry 30*, 986–996.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
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.
Stern, S., Tanaka, M., & Herr, W. (1989) *Nature 341*, 624–630.
Sturm, R. A., & Herr, W. (1988) *Nature 336*, 601–604.
Tabor, S. (1990) in *Current Protocols in Molecular Biology* (Ausubel, F. A., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., & Struhl, K., Eds.) pp 16.2.1–16.2.11, Greene Publishing Associates and Wiley-Interscience, New York.
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.
Wagner, G., Braun, W., Havel, T. F., Shaumann, T., Go, N., & Wuthrich, K. (1987) *J. Mol. Biol. 196*, 611–633.
Weiss, M. A., & Hoch, J. C. (1987) *J. Magn. Reson. 72*, 324–333.
Weiss, M. A., Wagner, G., Mueller, M., Stern, A. S., Botfield, M. C., Jancso, A., Sivaraja, M., Hyberts, S., & Hoch, J. C. (1994) *J. Biomol. NMR* (submitted for publication).
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.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Zuiderweg, E. R. P., Boelens, R., & Kaptein, R. (1985) *Biopolymers 24*, 601.