Biochemical Characterization of the Oct-2 POU Domain with Implications for Bipartite DNA Recognition†

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 23, 1992

ABSTRACT: B-cell specific regulation of immunoglobulin gene expression provides a model for the interaction of promoter and enhancer elements with eukaryotic sequence-specific DNA binding proteins. A critical element of this system, the octamer site (5′-ATGCAAAT-3′), is recognized by the B-cell transcription factor Oct-2. Octamer recognition is mediated by the POU domain, a conserved structural motif which—like the zinc finger and leucine zipper—defines a family of related transcription factors. Homologies among POU sequences suggest a bipartite structure, consisting of an N-terminal POU-specific subdomain and C-terminal variant homeodomain connected by a linker of variable length and sequence. As a first step toward a molecular understanding of the Oct-2 POU domain and its mechanism of DNA recognition, we have overexpressed in *Escherichia coli* the intact POU domain and subdomains as thrombin-cleavable fusion proteins and have purified these fragments to homogeneity following digestion with thrombin. Biochemical and biophysical characterization yields the following results. (i) The intact POU domain (166 residues) is monomeric and exhibits high-affinity octamer-specific DNA-binding activity. (ii) Limited proteolytic digestion demonstrates that the POU domain contains two proteolytically stable subdomains (the POU-specific subdomain and the variant homeodomain) connected by a proteolytically sensitive linker. (iii) The isolated subdomains are each monomeric and do not interact to form noncovalent heterodimers. (iv) Unlike the intact POU domain, the isolated subdomains do not exhibit high-affinity binding to the octamer DNA site, nor is DNA binding restored by mixing the two isolated subdomains. (v) Circular dichroism studies demonstrate that the intact POU-specific region and homeodomain each contain substantial α-helix and exhibit cooperative folding/unfolding transitions. The α-helix content of the intact POU region is equal to the sum of the α-helix content of the two subdomains. (vi) ^{1}H-NMR spectra of the isolated subdomains are additive to yield a spectra that is nearly identical to that of the parent POU domain. Together, these results confirm the bipartite structure of the POU domain and the existence of the POU-specific and variant homeo-subdomains as autonomous structural units.

Sequence-specific protein–DNA recognition is mediated by families of related structural motifs, which include the helix–turn–helix (Pabo & Sauer, 1984), helix–loop–helix (Murre et al., 1989), Zn finger (Klug & Rhodes, 1987), leucine zipper/bZIP (Lanschultz et al., 1988), and POU motifs (Herr et al., 1988). Although the structure of the POU motif has not been determined, it is predicted to consist of a conserved N-terminal POU-specific subdomain of 76–78 amino acids, a poorly conserved intervening linker of 14–27 amino acids, and a variant C-terminal homeodomain of approximately 60 amino acids. The variant POU-specific homeodomains differ from canonical homeodomains (as exemplified by the *Drosophila* antennapedia and engrailed proteins; Scott et al., 1988) by four highly conserved substitutions in the first and third DNA recognition helices (Herr et al., 1988). The POU-specific subdomain is required for high-affinity sequence-specific DNA binding, but it is incapable of DNA binding in the absence of the homeodomain. In contrast, the isolated homeodomain is sufficient for low-affinity DNA binding and specific protein–protein interactions. A large number of POU-motif family members representing both positive (He et al., 1989; Suzuki et al., 1990; Rosner et al., 1990; Schöler et al., 1990; Okamoto et al., 1990; Burglin et al., 1989, Johnson & Hirsch, 1990; Petryniak et al., 1990) and negative (Treacy et al., 1991; Okazawa, et al., 1991) regulators of gene expression have been cloned from vertebrate, *Caenorhabditis elegans*, and *Drosophila* libraries. Like homeodomain-containing proteins in *Drosophila*, POU-motif proteins are involved in cell-specific gene expression and developmental determination of cell fate. For example, *unc*-86 controls cell lineage determination and sensory neuron differentiation in *C. elegans* (Finney et al., 1988), and Pit-1 controls the development of lactotrophic and somatotrophic cells in the anterior pituitary of placental mammals (Bodner et al., 1988; Ingraham et al., 1988).

The octamer-binding proteins represent an extended subfamily of POU-motif proteins with related protein sequence and DNA specificity. The members of this subfamily (designated Oct-1, Oct-2, Oct-3, etc.) recognize an evolutionarily conserved octanucleotide component of vertebrate promoter and enhancer elements (5′-ATGCAAAT-3′). Oct-1 is a ubiquitous transcription factor involved in the expression of certain "housekeeping" genes, such as those encoding histones and small nuclear RNAs (Fletcher et al., 1987; Sturm et al., 1987, 1988). Oct-3/4 is expressed exclusively in totipotent and pluripotent stem cells prior to gastrulation and is essential

† This work was supported by the Lucille Markey Charitable Trust and by grants to M.A.W. from the NIH (GM45290) and the Tobacco Research Institute. M.C.B. is supported by a Cancer Research Institute Postdoctoral Fellowship. M.A.W. is supported in part by the American Cancer Society and the Pfizer Scholars Program for New Faculty.

* Address correspondence to this author at the Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School.

‡ Harvard Medical School.

§ Massachusetts General Hospital.

0006-2960/92/0431-5841$03.00/0

for mouse development beyond the one-cell stage (Rosner et al., 1991, 1990; Okamoto et al., 1990; Schöler, et al., 1990; Okazawa, 1991). Oct-6 is expressed predominantly in embryonic neural tissue and may be involved in neurogenic differentiation (Suzuki et al., 1990). Oct-2, the object of the present study, is a B-cell-specific transcription factor which is proposed to play a key role in regulating the expression of immunoglobulin genes (Singh et al., 1986; Clerc et al., 1988; Ko et al., 1988; Müller-Immerglück et al., 1988; Scheidereit et al., 1988; Miller et al., 1991).

As a first step toward the biochemical and biophysical characterization of the POU-motif, we describe in this paper the overexpression of a representative POU region (Oct-2) as a thrombin-cleavable fusion protein in *Escherichia coli*, its purification following thrombin digestion, and initial characterization of its domain structure and DNA-binding properties. In accord with previous observations (Sturm & Herr, 1988), it is shown that high-affinity sequence-specific DNA binding requires an intact POU domain. Biochemical and biophysical characterization of these fragments by limited proteolysis, circular dichroism (CD), and 1D ^{1}H-NMR[1] spectroscopy demonstrates that the Oct-2 POU domain consists of two autonomously folded protein subdomains. These correspond to the POU-specific subdomain and variant homeodomain as previously defined by sequence homologies. The ^{1}H-NMR and CD spectra of the intact POU domain are nearly identical to the sum of the spectra of the constituent subdomains. Together, these results establish (i) the bipartite structure of the POU domain, (ii) the existence of the POU-specific and homeo-subdomains as autonomously folding structural units, and (iii) the validity of the isolated POU-specific domain and homeodomain as structural models for the corresponding regions of the intact POU motif.

Materials and Methods

Construction of Expression Plasmids. Expression systems for the following five overlapping POU-motif subdomains were constructed: (i) the full-length POU motif consisting of the POU-specific subdomain, intervening linker region, and homeo-subdomains [pGEX-PIH; nucleotides 646–1143 as defined by Clerc et al. (1988), 166 aa]; (ii) the isolated POU-specific subdomain (pGEX-P; nucleotides 646–876, 77 aa); (iii) the isolated homeodomain (pGEX-H; nucleotides 949–1143, 65 aa); (iv) POU-specific subdomain plus linker (pGEX-PI; nucleotides 646–948, 101 aa); and (v) linker plus homeodomain (pGEX-IH; nucleotides 877–1143, 89 aa). In each case, cDNA encoding the human Oct-2 protein (LeBowitz et al., 1988) was modified by PCR mutagenesis to contain the appropriate N-terminal *Bam*HI restriction site, ochre (TAA) stop codon, and C-terminal *Bam*HI restriction site. Fragments were subcloned into the unique *Bam*HI site of pGEX-2T (Pharmacia, Inc.) resulting in C-terminal fusions to glutathione S-transferase of *Schistosoma japonicum* (amino acids 1–226; Smith et al., 1988) under the control of a *tac* promoter. Fidelity of the PCR mutagenesis and subcloning was directly verified by double-stranded DNA sequencing.

Growth and Affinity Purification. E. coli strain DH5α (F$^-$ ϕ80d Δ*lacZ*ΔM15 Δ(*lacZYA–argF*)U169 *recA1 endA1 hsdR17* (r_k^-, m_k^+) *supE44* λ^- *thi-1 gyrA relA1*) containing the appropriate plasmid was in each case grown at 37 °C to mid log phase in STG medium, induced with 0.1 mM IPTG, and grown 3 h postinduction. The cells were harvested by centrifugation and stored at −80 °C until processed. Cell pellets were rapidly thawed at 37 °C and resuspended in 3 mL of ice-cold lysis buffer per gram wet weight. PMSF (8 μL/g of a 50 mM solution) and lysozyme (80 μL/g of a 10 mg/mL solution) were added, and the cells were gently mixed at 4 °C for 30–60 min. Final lysis was achieved by brief sonication (3 min on a 30% duty cycle with a peak power of 50 W) using a Branson 250 sonifier fitted with a 0.5-inch diameter horn. To reduce the viscosity of the lysate, 10 mM $MgCl_2$ and DNase I were added (20 μL/g of the original wet weight of a 1 mg/mL solution), and the preparation was incubated at room temperature for 20–30 min. Protease inhibitors leupeptin and pepstatin (Boehringer-Mannheim) were added to 1 μg/mL, and cell debris was removed by centrifugation at 12000*g* at 4 °C. The cell pellet was back-extracted twice with 9 volumes of lysis buffer containing protease inhibitors, centrifuged as above, and the clarified lysates from the three centrifugations pooled. The pooled lysate was bound to an equal volume of preswollen glutathione-agarose (Sigma Cat. No. G4510) for 15 min at 4 °C with gentle end-over-end mixing. The eluent, which still contained a significant amount of fusion protein, was extracted twice again with equal volumes of preswollen glutathione agarose as above. The pooled glutathione agarose was washed with 5 volumes of thrombin buffer to remove nonspecifically bound protein and resuspended in an equal volume of thrombin buffer containing 4 mM $CaCl_2$. Due to the presence of a unique thrombin cleavage site between the N-terminal glutathione S-transferase and fused C-terminal POU fragments, the POU-motif subdomains were separated from the immobilized glutathione S-transferase by digestion with thrombin. Complete digestion was achieved following incubation for 8–12 h at room temperature in the presence of 3–5 NIH units of thrombin (Sigma Cat. No. T-6634)/mL of glutathione agarose. The released POU-motif fragments were collected by filtration (Whatman No. 3) while washing with several volumes of high salt buffer. The protein was dialysed against several changes of 50 mM NH_4HCO_3, 1 mM DTT at 4 °C, and lyophilized in preparation for RP-HPLC purification.

RP-HPLC Purification and Protein Renaturation. Final purification of the PIH, P, and H proteins was achieved by RP-HPLC (25 mm × 10 cm, Waters DeltaPak C18 column), using a 10%–70% acetonitrile gradient in 0.1% TFA. Between 43% and 55% acetonitrile, the acetonitrile was increased by 0.1%/mL for isolating PIH, by 0.03%/mL for P, or 0.3%/mL for the H. Peak fractions were lyophilized, dissolved in 5 M guanidine hydrochloride, 20 mM MOPS (pH 7.0), 10 mM DTT and exhaustively dialysed against 20 mM MOPS (pH

[1] Abbreviations: 1D ^{1}H-NMR, one-dimensional nuclear magnetic resonance spectroscopy; P, POU-specific subdomain of Oct-2 (amino acids 194–270); BSA, bovine serum albumin; CD, circular dichroism; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; FPLC, fast-protein liquid chromatography; GRA, gel-retardation assay; GST, glutathione S-transferase of *Schistosoma japonica* (amino acids 1–226); GST-H, fusion protein of glutathione S-transferase and the Oct-2 homeodomain; GST-IH, fusion protein of glutathione S-transferase and the Oct-2 intervening linker region and homeodomain; GST-P, fusion protein of glutathione S-transferase and the Oct-2 POU-specific subdomain; GST-PI, fusion protein of glutathione S-transferase and the Oct-2 POU-specific subdomain and intervening linker region; GST-PIH, fusion protein of glutathione S-transferase and the intact Oct-2 POU motif; H, homeodomain of Oct-2 (amino acids 295–359); HEPES, *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid; IH, intervening linker region and homeodomain of Oct-2 (amino acids 271–359); IPTG, isopropyl β-D-thiogalactopyranoside; MOPS, 3-(*N*-morpholino)propanesulfonic acid; PCR, polymerase chain reaction; PI, POU-specific subdomain with intervening linker region of Oct-2 (amino acids 194–294); PIH, POU-specific subdomain, intervening linker region, and homeodomain of Oct-2 (amino acids 194–359); PAGE, polyacrylamide gel electrophoresis; PMSF, phenylmethanesulfonyl fluoride; RP-HPLC, reverse-phase high-performance liquid chromatography; SDS, sodium dodecyl sulfate; TFA, trifluoroacetic acid.

7.0), 50 mM KCl, 1 mM DTT. Low molecular weight contaminants were removed from the PIH preparation by gel filtration on a Superose 12 FPLC column. The isolated proteins were judged to be >98% pure by SDS–PAGE (Figure 1B). Yields were approximately 12, 9, and 2 mg L^{-1} of original bacterial culture for PIH, P, and H, respectively.

Gel-Retardation Assay. Specific DNA-binding activity was evaluated using a gel-retardation assay (GRA; Fried & Crothers, 1981). Specific and control duplex oligonucleotides were end-labeled on one strand with [γ-^{32}P]ATP (New England Nuclear) using T4 polynucleotide kinase (Boehringer-Mannheim, Ltd.), and unincorporated label was removed by gel filtration. Specific activity of the resulting probe was approximately 10^6 dpm pm^{-1}. Reactions (typically in 10 μL) contained the following: 50 mM HEPES (pH 8.0), 5 mM $MgCl_2$, 1 mM DTT, 10% (w/v) glycerol, 0.1 mg/mL BSA, 0.1 mg/mL double-stranded poly(dI-dC) (Pharmacia Cat. No. 27-7880), 1 nM labeled specific DNA, and the indicated protein. Reactions were incubated 2–4 h at 4 °C, and bound versus unbound DNA resolved by gel electrophoresis at 4 °C (7% acrylamide, 38:1 ratio, in 66 mM Tris (pH 8.0), 33 mM sodium acetate, 1 mM EDTA).

FPLC Gel Filtration. Aggregation states of the purified proteins were assessed by FPLC gel filtration on a Superose 12 HR 10/30 column (Pharmacia) in gel-filtration buffer. Bovine serum albumin (molecular weight 66 000), carbonic anhydrase (29 000), cytochrome *c* (12 400), and aprotinin (6500) were used as molecular weight standards.

Protease Digestion. Thermolysin and trypsin were obtained from Boehringer-Mannheim. Digestions were conducted at two or more different protein-to-protease stoichiometries in proteolysis buffer at a protein concentration of 10–40 μM and followed as a function of time and temperature by SDS–PAGE. Digestion was stopped by the addition of EDTA to a final concentration of 10 mM. Proteolytic products were separated by SDS–PAGE and blotted onto ProBlott membrane (Applied Biosystems), and the cleavage sites were identified by automated N-terminal sequencing at the Harvard Microchemistry Facility.

Circular Dichroism. Circular dichroism (CD) measurements were made with an Aviv DS 60 spectropolarimeter (Lakewood, NJ) fitted with a Hewlett-Packard 89100A temperature controller. Protein samples (10–20 μM) were dissolved in 25 mM phosphate buffer (pH 7.0) and 50 mM KCl. Protein concentrations were determined directly from the experimental samples by UV absorbance at 280 nm using the following molar extinction coefficients: PIH = 16 807, P = 9200, H = 6520.

NMR Methods. ^{1}H-NMR spectra were recorded at 500 MHz at the Harvard Medical School NMR Facility in 25 mM phosphate buffer (pH 6.0, 10% D_2O), 50 mM KCl, 2 mM deuterated DTT (Merck Isotopes, St. Louis, MO) at 25 °C. Protein concentrations were 68 μM for PIH and 300 μM for P and H. The solvent resonance was attenuated by presaturation.

SDS–PAGE. Low molecular weight proteins were separated on polyacrylamide gels using either the Laemmli buffer system (Laemmli, 1970) with 10% glycerol or a tricine buffer system (Schagger & von Jagow, 1987) as indicated.

Buffers. STG medium: 2% Bactotryptone, 1% yeast extract, 0.5% NaCl, 0.2% glycerol, 50 mM KPO_4 (pH 7.2), 50 μg/mL ampicillin. Lysis buffer: 50 mM Tris (pH 8.0), 100 mM NaCl, 1 mM EDTA, 1% (v/v) Triton-X 100. Thrombin buffer: 50 mM Tris (pH 8.0), 150 mM NaCl, 0.5 mM DTT. High salt buffer; 50 mM Tris (pH 8.0), 1000 mM NaCl, 1 mM EDTA, 1% (v/v) Triton-X 100. Gel filtration buffer: 150 mM NaCl, 25 mM MOPS (pH 7.5), 0.1 mM EDTA, 0.5 mM DTT. Proteolysis buffer: 150 mM NaCl, 20 mM HEPES (pH 7.6), 2.5 mM $CaCl_2$, 0.5 mM DTT.

Table I: N-Terminal and C-Terminal Amino Acid Sequences of the Expressed Oct-2 POU-Motif Subdomains and Expected and Observed Amino Acid Compositions of the Purified Subdomains[a]

A. Predicted Amino Acid Sequences

PIH (168aa)		194	359
	NH_2-GS	**PEEPSDLE.....EKRINPCS**	-COOH
P (79aa)		194	270
	NH_2-GS	**PEEPSDLE.....KWLNDAET**	-COOH
H (67aa)		295	359
	NH_2-GS	**GRRRKKRT.....EKRINPCS**	-COOH

B. Amino Acid Compositions

	PIH		P		H	
	exptd	obsd	exptd	obsd	exptd	obsd
Ala	7	7.6	4	4.1	3	3.0
Arg	13	13.4	4	4.1	9	9.1
Asx	15	16.0	8	7.9	4	4.1
Cys	3	0.0	1	0.0	2	0.0
Glx	24	23.4	12	11.9	11	10.9
Gly	9	10.0	6	6.4	2	2.2
His	1	1.3	0	0.0	1	1.0
Ile	7	6.1	2	1.8	5	4.3
Leu	20	20.9	11	11.4	5	5.1
Lys	14	14.9	7	7.2	7	7.4
Met	4	3.1	2	2.2	1	0.6
Phe	10	8.9	6	5.4	3	2.7
Pro	9	9.6	3	3.2	2	2.4
Ser	16	14.0	5	4.5	5	4.5
Thr	8	9.3	5	4.2	3	3.3
Trp	2		1		1	
Tyr	1	0.4	1	0.6	0	0.0
Val	5	4.8	1	1.1	3	2.4

[a] Residues corresponding to Oct-2 sequences are shown in boldface; residues corresponding to nonnative additions resulting from expression as fusion proteins are shown in lightface (see Materials and Methods). Residue numbering corresponds to that defined by Clerc et al. (1988). Amino acid compositions were determined by quantitative acid hydrolysis followed by HPLC. Values for Cys and Trp were not determined. Asx denotes the sum of Asn and Asp; Glx denotes the sum of Gln and Glu.

Results

(I) Bacterial Expression and Purification

Overexpression and Purification of Fusion Protein. Initial attempts to express the POU motif and subdomains directly in a T7-based system (Studier & Moffat, 1986; Tabor & Richardson, 1985; Tabor, 1990) yielded no observable inducible proteins of the appropriate sizes, as monitored by SDS–PAGE of total cell lysates. We therefore elected to express this region as a fusion protein with glutathione S-transferase (Smith et al., 1988). Expression systems for five fragments respresenting the following overlapping POU subdomains were constructed: (i) the full-length POU region consisting of the POU-specific subdomain, intervening linker region, and variant homeodomains (PIH; nucleotides 646–1143, 166 aa); (ii) the isolated POU-specific subdomain (P; nucleotides 646–876, 77 aa); (iii) the isolated homeodomain (H; nucleotides 949–1143, 65 aa); (iv) POU-specific subdomain plus linker (PI; nucleotides 646–948, 101 aa); (v) and linker plus homeodomain (IH; nucleotides 877–1143; 89 aa). Three of the five fusion proteins (PIH, P, and H; Table IA) were purified to homogeneity by a combination of affinity chromatography and RP-HPLC

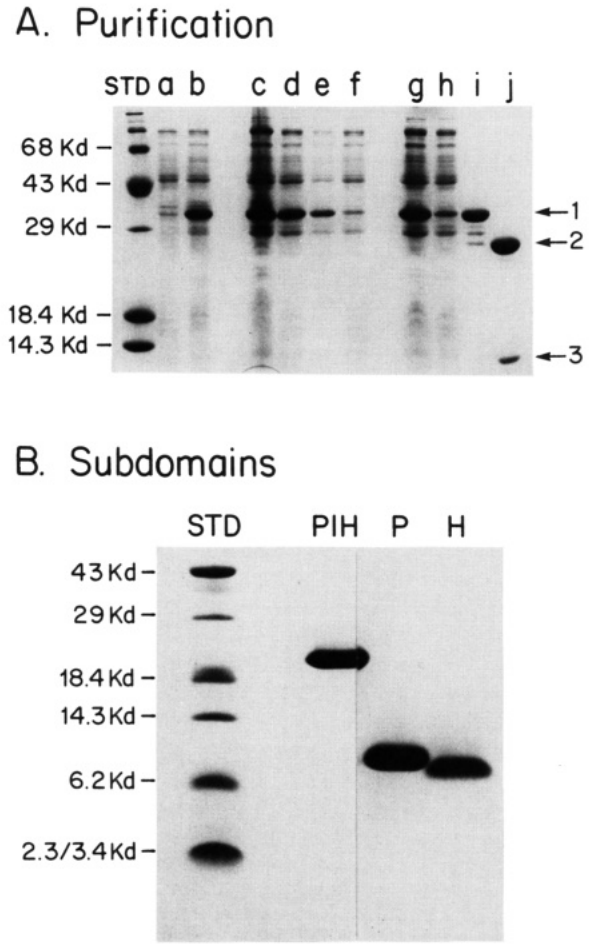

FIGURE 1: SDS–PAGE of purification intermediates and final protein preparations. (A) Purification intermediates as described in Materials and Methods. All lanes are from a single preparation and represent (except where noted) comparative yields of each step. Proteins were separated on a 10% Laemmli gel (Laemmli, 1970) and stained with Coomassie Brilliant Blue G250. Lanes a–b: preinduction and postinduction whole cell extracts of *E. coli* DH5α/pGEX-P equivalent to 30 μL of culture. Lanes c–f: clarified lysate, first Triton extraction, second Triton extraction, and postextraction pellet, respectively. Lanes g–h: pooled lysate before and after affinity extraction of the fusion protein. Lanes i–j: affinity-purified GST-P before and after cleavage with thrombin (lane h contains twice as much protein as lane i). Arrows indicate the positions of the GST-P, GST, and P proteins, respectively. (b) SDS–PAGE of RP-HPLC-purified proteins illustrating protein homogeneity. Proteins were separated on 15% SDS–tricine gels (Schagger & von Jagow, 1987) and stained as above. The image is a composite of two gels aligned at the PIH band.

followed by quantitative renaturation from guanidine hydrochloride (see Figure 1 and Materials and Methods). The presence of a unique thrombin cleavage site at the fusion joint permitted separation of the glutathione S-transferase from the POU motif subdomains following selective digestion with thrombin. This process resulted in the addition of two nonnative amino-terminal amino acids (NH_2-Gly-Ser-) to each fragment. In each case, characterization of the purified products by quantitative amino acid analysis and amino-terminal sequencing was consistent with specific cleavage at the designated thrombin site; spurious cleavage elsewhere was not detected. The predicted and observed amino acid compositions of the isolated POU subdomains are given in Table IB.

Specific DNA-Binding Activity. The DNA-binding properties of the fusion proteins and isolated Oct-2 fragments were evaluated using a gel-retardation assay (Figures 2 and 3). In this assay, specific binding is detected on the basis of the different electrophoretic mobilities of free and protein-bound

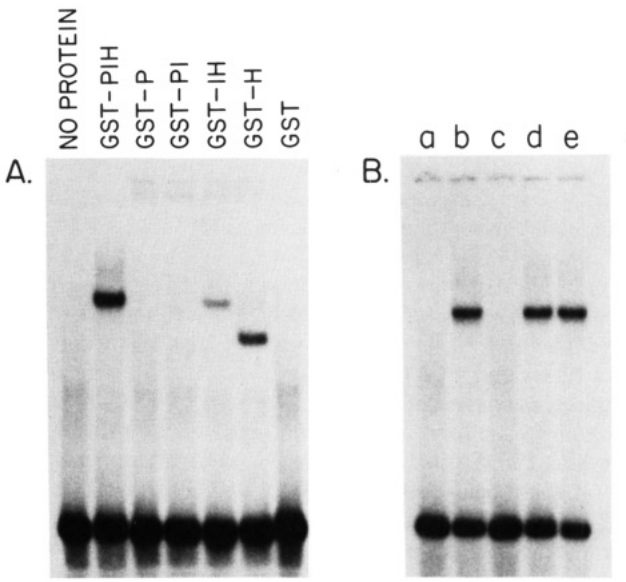

FIGURE 2: In vitro binding of GST/Oct-2 fusion proteins to octamer-containing oligonucleotide probes. (A) Gel-retardation of octamer-containing probe by affinity-purified fusion proteins. All lanes contain 1 nM ^{32}P-labeled duplex oligonucleotide (5′-TTCAGCCA-GACGACCAAAAAAAAA-AAAA**ATGCAAAT**AAAAAAAA-AAAAAGTCCGCTTCGCCTTA-3′) containing a central octamer (bold face type). Lane 1: no protein (negative control). Lanes 2–6: 0.1 μM GST-PIH, 200 μM GST-P, 200 μM GST-PI, 100 μM GST-IH, 100 μM GST-H, or 200 μM GST, respectively. Qualitatively similar results were obtained with RP-HPLC-purified PIH, P, and H proteins (data not shown). PI and IH were not tested. (B) Competition with excess unlabeled specific and nonspecific oligonucleotides. All lanes contain 1 nM ^{32}P-labeled duplex octamer-specific oligonucleotide (same as in panel A). Lanes b–e contain 0.1 μM affinity-purified GST-PIH. Lane a: no protein (negative control). Lane b: GST-PIH (positive control). Lanes c–e: GST-PIH plus 100 nm unlabeled octamer-specific oligonucleotide; GST-PIH plus 1000 nM unlabeled duplex oligonucleotide containing the λ operator sequence (5′-CAGTCGTCTGAGCTCAAAAAAAAAAAAT**AC-CTCTGGCGGTGATA**AAAAAAAAAA-AAAAGAGCTC-CAGTCGTCT-3′); and GST-PIH plus 1000 nM unlabeled duplex oligonucleotide containing the flanking regions of the octamer-specific oligonuclotide (5′-TTCAGCCAGACGACCGTCCGCTTCGC-CTTA-3′), respectively. Qualitatively similar results were obtained with GST-H and purified PIH and H proteins (data not shown).

^{32}P-labeled DNA probes. Both the fusion proteins (Figure 2A) and isolated fragments (data not shown) exhibit an identical pattern of DNA affinities: high-affinity octamer recognition by PIH, low-affinity recognition by IH and H, and no detectable recognition by P or PI. The sequence specificity of DNA binding of PIH, H, and their fusion proteins (Figure 2B) was confirmed by GRA in the presence of specific and nonspecific competitor DNA. Inclusion of a 100-fold excess of nonradioactive competitor DNA containing an octamer recognition site in the GSA reaction mixture completely abolished the PIH shift (Figure 2B, lane c), while inclusion of a 1000-fold excess of competitor DNA containing a GCN4 (data not shown) or λ recognition site (Figure 2B, lane e) had no effect. Similar behavior was exhibited by H, GST-PIH, and GST-H proteins (data not shown). Binding affinities of the RP-HPLC-purified PIH and H domains for the octamer site were determined by quantitative GSA as described in Figure 3. The intact POU domain exhibits an affinity at least 1000-fold greater than that of the isolated homeodomain (K_d = 60 nM^{-1} vs $K_d > 6 \times 10^4$ nM^{-1}). These results are in accord with previous studies (Sturm & Herr, 1988). The DNA-binding isopleth of the intact POU domain is sigmoidal, suggesting cooperative binding. Cooperative binding of POU proteins to multiple sites has previously been described (Le-

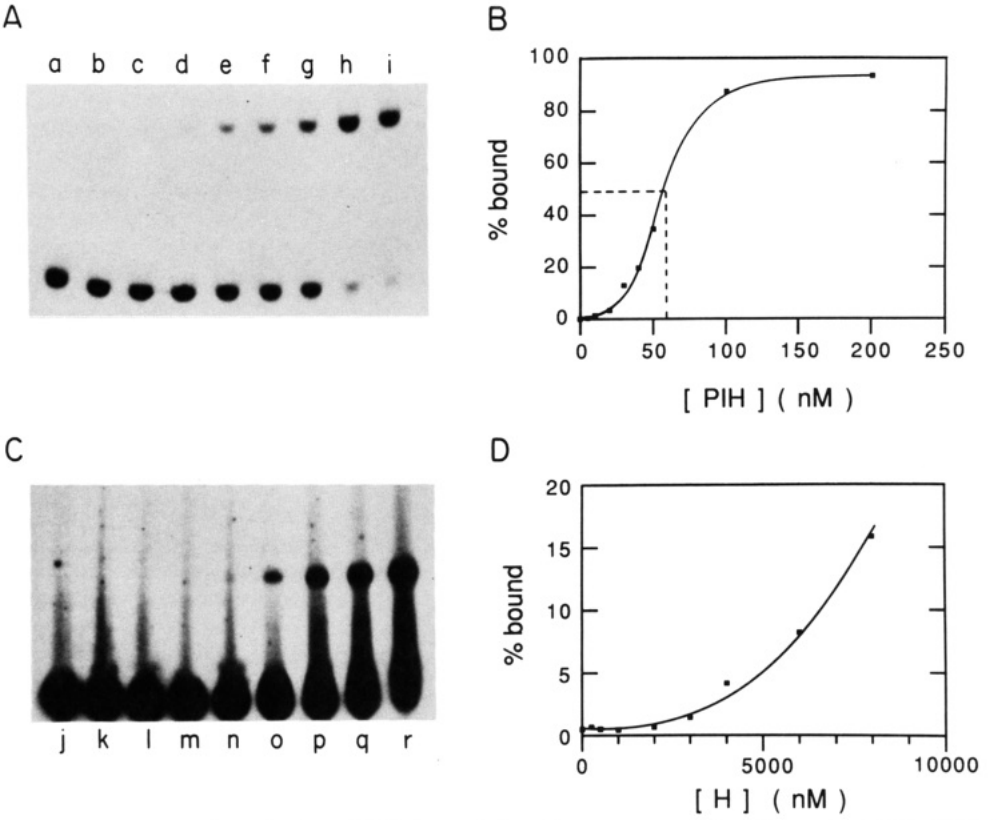

FIGURE 3: Determination of octamer-binding affinity. (A) Gel retardation of octamer-containing oligonucleotide by PIH. All lanes contain 1 nM ^{32}P-labeled duplex oligonucleotide (5′-GTATGC-AAATGG-3′) containing a central octamer (boldface type). Lane a: no protein (negative control). Lanes b–i, increasing concentrations (5, 10, 20, 30, 40, 50, 100, or 200 nM) of RP-HPLC-purified and guanidine hydrochloride refolded PIH. Shifted and unshifted counts were quantified by Cerenkov emission and plotted against protein concentration (panel B) to facilitate the estimation of binding affinity. (C) Gel retardation of octamer-containing oligonucleotide by H. This panel is as in panel A, except lanes b–i contain increasing concentrations (250, 500, 1000, 2000, 3000, 4000, 6000, or 8000 nM) or RP-HPLC-purified and guanidine hydrochloride refolded H. Shifted and unshifted counts were determined by Cerenkov counting and plotted against protein concentration (panel D).

Table II: N-Terminal Amino Acid Sequences of the 7-kDa and 11-kDa Products of Thermolysin Digestion[a] (See Results)

11-kDa band	NH_2 – *GS*PEEPSDL.... (194)
7-kDa band	NH_2 – *S*PEEPSDL.... (194)
	NH_2 – LGFDGLP.... (288)
	NH_2 – FDGLPGR.... (290)

[a] Residues corresponding to Oct-2 sequences are shown in boldface; residues corresponding to the nonnative amino-terminal additions resulting from expression of PIH as a thrombin-cleavage fusion protein are shown in italics (see Materials and Methods). Residue numbering corresponds to that defined by Clerc et al. (1988).

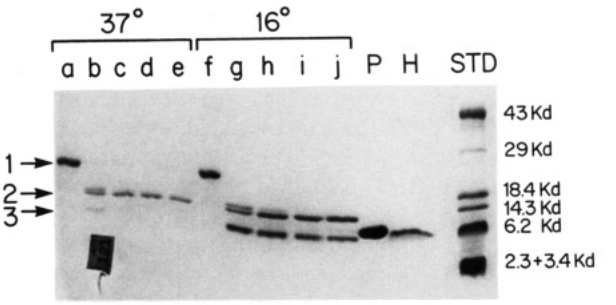

FIGURE 4: Thermolysin digestion of PIH. RP-HPLC purified and guanidine hydrochloride refolded PIH (40 μM) was digested with thermolysin at 1:100 enzyme to substrate ratio at 37 °C (lanes a–e) and at 16 °C (lanes f–j) as described in Materials and Methods. Aliquots were removed before digestion (a, f), after 5 min (b, g), 15 min (c, h), 30 min (d, i), and 60 min (e, j) of digestion and run on 15% SDS–polyacrylamide gel containing 10% glycerol. Positions of the undigested PIH and the 11-kDa and the 7-kDa thermolysin fragments are incidated by arrows 1–3, respectively. For comparison, RP-HPLC purified POU-specific subdomain (lane P) and variant homeodomain (lane H) are included. The number in the lower left corner of the gel (121) is an index code incorporated into the gel.

bowitz et al., 1989). The mechanistic implications of a sigmoidal isopleth when the target is a small oligonucleotide are under investigation.

Aggregation State. The oligomeric states of the intact POU-domain and constituent subdomains were evaluated by FPLC gel-filtration chromatography in reference to known standards (see Materials and Methods). The Oct-2 POU domain, the POU-specific subdomain, and the variant homeodomain were each observed to be monomeric at 10–20 μM protein concentration in gel-filtration buffer. No stable noncovalent interactions were observed between the isolated subdomains when mixed and coloaded onto the column.

(II) Domain Structure and Stability

Limited Proteolysis Defines Two Domains. The domain organization of the intact POU region was probed by limited proteolysis with thermolysin and trypsin. Limited proteolysis with thermolysin at temperatures between 4 and 25 °C produced two stable fragments of apparent molecular masses 7 and 11 kDa, with formation of the 11-kDa fragment proceeding through a short-lived higher molecular mass intermediate (Figure 4, lanes 2 and 7). In order to deduce their identities, the amino termini of the fragments were sequenced (Table II). The 11-kDa band exhibited an amino-terminal sequence identical to that of the intact PIH domain. Given its apparent molecular mass of 11 kDa, the fragment must

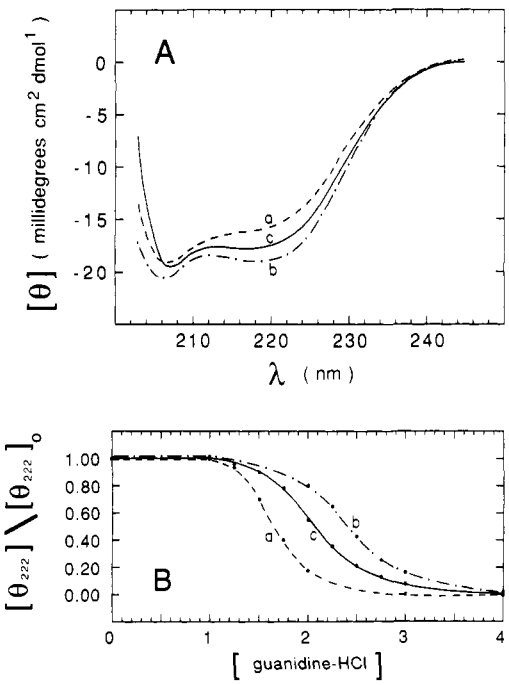

FIGURE 5: Circular dichroism spectra and guanidine hydrochloride denaturation curves of purified POU-motif subdomains. (A) Circular dichroism spectra of PIH (—), P (- - -), and H (-·-) proteins. (B) Guanidine hydrochloride denaturation monitored by CD at 222 nm. Labels are as in panel A. Data for both panels were obtained at 4 °C with 12–15 μM protein in 25 mM potassium phosphate (pH 6.0), 50 mM KCl. Aliquots of a protein stock were diluted with an equal volume of 2× guanidine buffer (25 mM KP_i (pH 6.0), 50 mM KCl, guanidine hydrochloride) and incubated at 4 °C for approximately 1 h prior to CD measurements.

contain the POU-specific subdomain (nominal MW = 9023) and much of the linker (nominal MW = 2316). The 7-kDa fragment contained a 1:1:1 mixture of three amino-terminal sequences (Table II). Given its apparent molecular weight and the N-terminal sequences, the 7-kDa band must contains three protein species, two of which represent the homeodomain plus a short (5 or 7 amino acids) segment of the linker region; the third is a truncated POU-specific subdomain. A qualitatively similar pattern is obtained by limited proteolysis with trypsin (not shown). The results suggest that the POU domain consists of two autonomous folding units connected by a flexible linker. The lambdoid phage repressors exhibit a similar pattern of proteolytic susceptibilities and domain organization (Pabo et al., 1979; DeAnda et al., 1984; Weiss et al., 1983; Hecht et al., 1983)

Temperature Dependence of Thermolysin Cleavage. Further evidence for domain structure may be inferred from the temperature dependence of proteolytic digestion (Hecht et al., 1983). The susceptibility of the Oct-2 POU domain to thermolysin digestion was determined at 4, 16, 25, 37, and 45 °C. The time courses of representative digestions (16 and 37 °C) are shown in Figure 4. At temperatures between 4 and 25 °C, both the 7-kDa and 11-kDa bands exhibit similar resistance to thermolysin digestion. Between 37 and 45 °C, however, the 7-kDa fragment becomes increasingly susceptible to proteolysis, whereas the 11-kDa fragment maintains its resistance. These data suggest that POU-specific subdomain plus linker region (i.e., the 11-kDa fragment) is significantly more stable than either the POU-specific subdomain alone or the homeo-subdomain (i.e., the components of the 7-kDa fragment). To confirm these findings, the temperature dependence of thermolysin digestion was determined for the cloned and purified P and H fragments. As expected, the purified H protein exhibited thermostability similar to those of the 7-kDa thermolytic fragments of PIH, while the purified P protein was more stable than the H, but less stable than the 11-kDa thermolytic fragment of PIH (data not shown).

Circular Dichroism. In Figure 5A are shown CD spectra of the intact POU domain (labeled a), the purified POU-specific subdomain (labeled b), and the purified homeodomain (labeled c). The POU-specific subdomain is ~40% α-helix, the homeodomain is ~55% α-helix, and the intact POU domain is ~48% α-helix. In the case of the homeo-subdomain, this estimate is in accord with NMR and crystallographic structures of canonical homeodomains (Otting et al., 1989; Qian et al., 1990; Kissinger et al., 1990; Wolberger et al., 1991). In addition, the sum of the constituent subspectra (P + H) is essentially identical to the spectrum of the intact PIH. These results provide further evidence that (a) the domains are correctly folded and (b) the isolated subdomains are autonomously folding units. These results also suggest that the linker region (which is absent in the isolated P and H fragments) does not fold as an α-helix, in accord with sequence-based predictions of secondary structure (Chou & Fasman, 1974).

Mean residue ellipticity at an α-helix-associated wavelength (222 nm) was used to monitor the unfolding of the POU domain and its subdomains as a function of denaturant concentration (guanidine hydrochloride). These data are shown in Figure 5B. In each case, a cooperative unfolding transition is observed. $\Delta G_{unfolding}$ was determined for the POU-specific subdomain and homeodomain according to the equation

$$\Delta G_{unfolding} = -RT \ln K_{unfolding}$$

using the model

$$P_{folded} \overset{K_{unfolding}}{\rightleftharpoons} P_{unfolded}$$

where

$$K_{unfolding} = \frac{P_{unfolded}}{P_{folded}} = \frac{P_{total}(f_{unfolded})}{P_{total}(1 - f_{unfolded})} = P_{total}\frac{(f_{unfolded})}{(1 - f_{unfolded})}$$

and P_{total} is the total protein concentration, P_{folded} and $P_{unfolded}$ are the concentrations of folded and unfolded protein, respectively, and $f_{unfolded}$ is the fraction of total protein that is unfolded. The free energies of unfolding were calculated at those denaturant concentrations resulting in at least 5% but less than 95% subdomain unfolding, and fitted by least-squares linear regression to the equation

$$\Delta G_{unfolding} = m[\text{guanidine hydrochloride}] + \Delta G^{H_2O}_{unfolding}$$

where $\Delta G^{H_2O}_{unfolding}$ is the extrapolated free energy of unfolding in the absence of denaturant. In 25 mM phosphate buffer (pH 7.0), 50 mM KCl at 4 °C, the isolated POU-specific subdomain gave a $\Delta G^{H_2O}_{unfolding}$ ~ 11–11.5 kcal mol^{-1} ($K_{unfolding} \sim 10^{-9}$), and the isolated homeodomain 10–10.5 kcal mol^{-1} ($K_{unfolding} \sim 10^{-8}$). These values are in accord with the relative stabilities of the subdomains to thermolytic digestion described above.

1D ^{1}H-NMR Spectroscopy. In Figure 6 are shown ^{1}H-NMR spectra of the intact POU domain (panel A), purified POU-specific subdomain (panel B), and homeodomain (panel C). Panels D–F show in expanded form the upfield methyl region; corresponding resonances in this region emphasize the additive nature of the constituent subspectra. The intact POU domain exhibits five methyl shifts in this region (labeled a–e). The POU-specific subdomain clearly contributes three of the

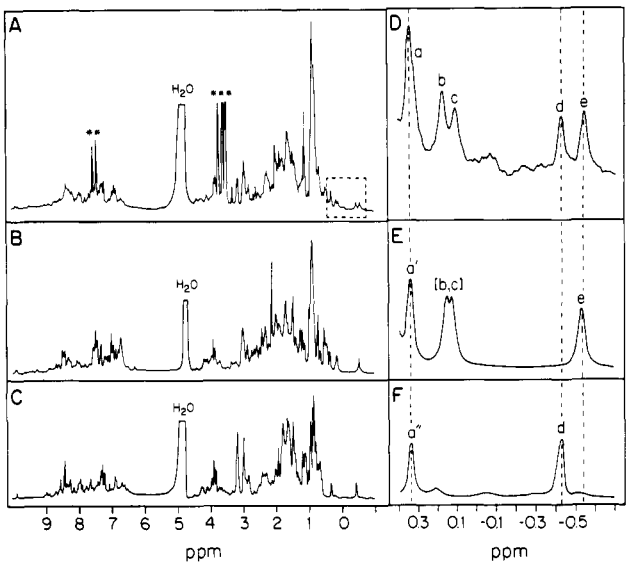

FIGURE 6: 1D ^{1}H-NMR of purified proteins. Panel A: 1D ^{1}H spectra of 68 μM PIH at 25 °C. Panels B and C: corresponding spectra of P (300 μM) and H (300 μM). Residual water is indicated. The PIH spectrum exhibits several sharp lines [marked by asterisks (*)] resulting from small molecule contaminants that were inadvertently introduced in preparation of the sample. Panels D, E, and F: expansion of the corresponding near methyl regions indicated by the dashed box in panel A. Lowercase letters mark corresponding methyl groups in the three spectra and emphasize their additive nature. The detailed correspondence between resonances b and c of PIH (panel D) and overlapping resonances [b,c] of P (panel E) is not clear. The integrated area of shift α is three protons greater than the sum of the constituent subspectra (data not shown) indicating either displacement of resonances from other regions of the P or H spectra or additional contributions by resonances of the linker region. All samples were prepared by RP-HPLC and refolded from guanidine hydrochloride as described in Materials and Methods. Samples were extensively dialysed against 25 mM potassium phosphate (pH 6.0 and 50 mM KCl prior to the addition of 10% D_2O and 2 mM deuterated DTT. Spectra were acquired at 500 MHz.

five shifts (b, c, and e), while the homeodomain contributes shift d. Both subdomains contribute methyl protons (a′ and a″) to shift a. We emphasize that such correspondences in the 1D spectra are suggestive, but not rigorous; detailed comparison of these resonances will require independent sequence-specific assignment in each case. Nevertheless, the present results strongly support the validity of the isolated POU-specific domain and homeodomain as structural models for the corresponding regions of the intact POU motif. A similar correspondence of NMR resonances has previously been observed in studies of the intact λ repressor and its constituent N- and C-terminal domains (Weiss et al., 1987).

Discussion

Eukaryotic transcription factors provide striking examples of the modular construction of proteins, as originally envisaged by the exon-shuffling hypothesis (Gilbert, 1978, 1979). Distinctive sequence motifs appear to encode structural domains with different functions, such as DNA recognition, specific protein–protein interactions, transcriptional activation, and ligand binding. In accord with this general scheme, the sequence of Oct-2 consists of several distinct motifs, including the POU domain, putative leucine zipper, and glutamine-rich transactivating region (Clerc et al., 1988). In this paper, we describe the biochemical characterization of one of these motifs, the POU domain.

The bipartite character of the POU motif was originally inferred from sequence homologies (Herr et al., 1988). Although the three-dimensional structure of a representative POU domain has not been determined, evidence supporting a bipartite structure has emerged from several lines of investigation. These include the following findings: (i) The homeodomain is sufficient for low-affinity sequence-specific DNA binding, but it requires the POU-specific subdomain for high-affinity binding (Sturm & Herr, 1988). (ii) Whereas sequences within the intervening linker can be be deleted with little effect (Theill et al., 1989), deletions within the POU-specific subdomain reduce DNA binding, and deletions within the homeodomain abolish it (Sturm & Herr, 1988; Theill et al., 1989; Verrijzer et al., 1990a,b; Müller-Immerglück et al., 1990). (iii) The Oct-1 homeodomain alone is sufficient to direct the cooperative binding of the herpes simplex virus transactivator VP16 and to assemble a specific protein–DNA complex containing Oct-1, VP16, and an additional cellular protein (Kristie & Sharp, 1990). (iv) Pit-1, which exists as a monomer in solution but binds as a dimer to its DNA response element, requires the POU-specific subdomain to dimerize (Ingraham et al., 1990). (v) POU-specific subdomain protein–protein interactions are responsible for DNA bending by Oct-1, Oct-2, Oct-6, and Pit-1 homodimers (Verrijzer et al., 1991).

The above investigations have provided strong, but indirect, evidence for a bipartite domain organization of the POU motif. However, no direct structural method has previously been used to verify this model. In this paper, we have presented a complementary biochemical characterization of the Oct-2 POU domain and its constituent subdomains. In accord with previous observations (Sturm & Herr, 1988), it is shown that high-affinity specific DNA binding requires the intact POU domain and is not exhibited by the isolated fragments. Limited proteolysis with thermolysin and trypsin indicates that the POU motif consists of two domains connected by a proteolytically sensitive linker. The results are consistent with the conservation of two blocks of homology (the POU-specific domain and variant homeodomain, respectively) connected by a linker of variable length and sequence. Circular dichroism (CD) studies demonstrate that the POU-specific and homeodomains contain 40%–55% α-helix and exhibit cooperative folding/unfolding transitions with significant free energies ($\Delta G_{unfolding} > 10$ kcal mol^{-1}). The isolated subdomains are monomeric, and they do not interact to form noncovalent heterodimers. Furthermore, the CD and ^{1}H NMR spectra of the intact POU domain are similar to the sum of the constituent subspectra, indicating that the isolated fragments retain native structure as independent folding units. Their future characterization by mutagenesis and multidimensional NMR spectroscopy will provide insight into a novel structural motif and have broad implications for bipartite DNA recognition.

Acknowledgments

We thank J. LeBowitz and P. A. Sharp for Oct-2 cDNA; R. Nazarian and J. Denninger for technical assistance; A. Nussbaum for oligonucleotide synthesis; M. Lachenmann for help with the construction of pGEX-H; S. Tabor and K. Struhl for helpful discussion; J. Lee for assistance with NMR measurements; H. T. Keutmann (Massachusetts General Hospital) for amino acid analysis; and W. Lane (Harvard Microchemistry Facility) for protein sequencing. The NMR Facility at HMS is supported by an NIH Shared Instrumentation Grant (1 S10 RR04862-01) and an award from the National Health Resource Foundation.

References

Bodner, M., & Karin, M. (1987) *Cell 50*, 267–275.

Bürglin, T. R., Finney, M., Coulson, A., & Ruvkun, G. (1989) *Nature (London)* 341, 239–243.
Chou, P. Y., & Fasman, G. D. (1974) *Biochemistry 13*, 211–222.
Clerc, R. G., Corcoran, L. M., LeBowitz, J. H., Baltimore, D., & Sharp, P. A. (1988) *Genes Dev. 2*, 1570–1581.
DeAnda, J., Poteete, A. R., & Sauer, R. T. (1983) *J. Biol. Chem. 258*, 10536–10542.
Finney, M., Ruvkun, G., & Horvitz, H. R. (1988) *Cell 55*, 757–769.
Fletcher, C., Heintz, N., & Roeder, R. G. (1987) *Cell 51*, 773–781.
Fried, M., & Crothers, D. M. (1981) *Nucleic Acids Res. 9*, 6505–6525.
Gilbert, W. (1978) *Nature (London) 271*, 501.
Gilbert, W. (1979) in *Eucaryotic Gene Regulation: ICN–UCLA Symposia on Molecular and Cellular Biology* (Axel, R., Maniatis, T., & Fox, C. F., Eds.) Vol. 14, pp 1–12, Academic Press, New York.
He, X., Treacy, M. N., Simmons, D. M., Ingraham, H. A., Swanson, L. W., & Rosenfeld, M. G. (1989) *Nature (London) 340*, 35–42.
Hecht, M. H., Nelson, H. C. M., & Sauer, R. T. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2676–2680.
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.
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.
Ingraham, H. A., Flynn, S. E., Voss, J. W., Albert, V. R., Kapiloff, M. S., Wilson, L., & Rosenfeld, M. G. (1990) *Cell 61*, 1021–1033.
Johnson, W. A., & Hirsh, J. (1990) *Nature (London) 343*, 467–470.
Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., & Pabo, C. O. (1990) *Cell 63*, 579–590.
Klug, A., & Rhodes, D. (1987) *Trends Biochem. Sci. 12*, 464–468.
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.
Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.
Landschultz, W. H., Johnson, P. F., & McKnight, S. L. (1988) *Science 240*, 1759–1764.
LeBowitz, J. H., Kobayashi, T., Staudt, L., Baltimore, D., & Sharp, P. A. (1988) *Genes Dev. 2*, 1227–1237.
Miller, C. L., Feldhaus, A. L., Rooney, J. W., Rhodes, L. D., Sibley, C. H., & Singh, H. (1991) *Mol. Cell. Biol. 11*, 4885–4894.
Müller, M. M., Ruppert, S., Schaffner, W., & Matthias, P. (1988) *Nature (London) 336*, 544–551.
Müller-Immerglück, M. M., Schaffner, W., & Matthias, P. (1990) *EMBO J. 9*, 1625–1634.
Murre, C., McCaw, P. S., & Baltimore, D. (1989) *Cell 56*, 777–783.
Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M., & Hamada, H. (1990) *Cell 60*, 461–472.
Okazawa, H., Okamoto, K., Ishino, F., Ishino-Kaneko, T., Takeda, S., Toyoda, Y., Muramatsu, M., & Hamada, H. (1991) *EMBO J. 10*, 2997–3005.
Otting, G., Qian, Y.-q., Müller, M., Affolter, M., Gehring, W., & Wüthrich, K. (1988) *EMBO J. 7*, 4305–4309.
Pabo, C. O., & Sauer, R. T. (1984) *Annu. Rev. Biochem. 53*, 293–321.
Pabo, C. O., Sauer, R. T., Sturtevant, J. M., & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 59*, 1259–1263.
Petryniak, B., Staudt, L. M., Postema, C. E., McCormack, W. T., & Thompson, C. B. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 1099–1103.
Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J., & Wüthrich, K. (1989) *Cell 59*, 573–580.
Rosner, M. H., Vigano, M. A., Ozato, K., Timmons, P. M., Poirier, F., Rigby, P. W. J., & Staudt, L. M. (1990) *Nature (London) 345*, 686–692.
Rosner, M. H., Vigano, M. A., Rigby, P. W. J., Arnheiter, H., Staudt, L. M. (1991) *Science 253*, 144–145.
Scheidereit, C., Heguy, A., & Roeder, R. G. (1987) *Cell 51*, 783–793.
Scheidereit, C., Cromlish, J. A., Gerster, T., Kawakami, K., Balmaceda, C.-G., Currie, R. A., & Roeder, R. G. (1988) *Nature (London) 336*, 551–557.
Schöler, H. R., Ruppert, S., Suzuki, N., Chowdhury, K., & Gruss, P. (1990) *Nature (London) 344*, 435–439.
Scott, M. P., Tamkun, J., & Hartzell, G. W., III (1989) *Biochim. Biophys. Acta 989*, 25–48.
Singh, H., Sen, R., Baltimore, D., & Sharp, P. A. (1986) *Nature (London) 319*, 154–158.
Smith, D. B., Rubira, M. R., Simpson, R. J., Davern, K. M., Tiu, W. U., Board, P. G., & Mitchell, G. F. (1988) *Mol. Biochem. Parasitol. 27*, 249–256.
Staudt, L. M., Clerc, R. G., Singh, H., LeBowitz, J. H., Sharp, P. A., & Baltimore, D. (1988) *Science 241*, 577–580.
Studier, F. W., & Moffatt, B. A. (1986) *J. Mol. Biol. 189*, 113–130.
Sturm, R. A., & Herr, W. (1988) *Nature (London) 336*, 601–604.
Sturm, R., Baumruker, T., Franza, B. R., Jr., & Herr, W. (1987) *Genes Dev. 1*, 1147–1160.
Sturm, R. A., Das, G., & Herr, W. (1988) *Genes Dev. 2*, 1582–1599.
Suzuki, N., Rohdewohld, H., Neumann, T., Gruss, P., & Schöler, H. R. (1990) *EMBO J. 9*, 3723–3732.
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.
Tabor, S., & Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1074–1078.
Theill, L. E., Castrillo, J.-L., Wu, D., & Karin, M. (1989) *Nature (London) 342*, 945–948.
Treacy, M. N., He, X., & Rosenfeld, M. G. (1991) *Nature (London) 350*, 577–584.
Verrijzer, C. P., Kal, A. J., & van der Vliet, P. C. (1990a) *EMBO J. 9*, 1883–1888.
Verrijzer, C. P., Kal, A. J., & van der Vliet, P. C. (1990b) *Genes Dev. 4*, 1964–1974.
Verrijzer, C. P., van Oosterhout, J. A. W. M., van Weperen, W. W., & van der Vliet, P. C. (1991) *EMBO J. 10*, 3007–3014.
Weiss, M. A., Karplus, M., Patel, D. J., & Sauer, R. T. (1983) *J. Biomol. Struct. Dyn. 1*, 151–157.
Weiss, M. A., Karplus, M., & Sauer, R. T. (1987) *J. Biomol. Struct. Dyn. 5*, 539–556.
Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., & Pabo, C. O. (1991) *Cell 67*, 517–528.