Site-Specific Recognition by an Isolated DNA-Binding Domain of the Sine Oculis Protein†

Tony R. Hazbun, Florence L. Stahura,‡ and Michael C. Mossing*

Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556

Received October 8, 1996⊗

ABSTRACT: The *sine oculis* (*so*) gene is required for the development of the *Drosophila* visual system. The 416 amino acid SO protein contains a 40 amino acid region homologous to the helix−turn−helix (HtH) region of the homeodomain. Three HtH-containing peptides ranging in size from 63 to 93 amino acids ($SO_{218-279}$, $SO_{204-279}$, and $SO_{188-279}$) were expressed in *Escherichia coli* and characterized *in vitro*. These fragments show circular dichroism spectra characteristic of helical proteins and cooperative unfolding transitions. Derivatization of these three peptides with the chemical nuclease 1,10-phenanthroline:copper (OP-Cu) allowed the identification of specific DNA-binding sites within the 3.1 kb pUC119 plasmid. Similar cleavage patterns with similar relative affinities were obtained for all three peptides. Nucleotide resolution mapping of the predominant cleavage area identified two primary cleavage sites with a similar core sequence. The DNA cleavage sites were confirmed by DNase I footprinting with both native and OP-Cu-conjugated SO HtH peptides. This study identifies a 63 amino acid peptide as sufficient for specific DNA binding.

The *sine oculis* (*so*) gene is involved in the development of the *Drosophila* visual system (Cheyette et al., 1994; Serikaku & O'Tousa, 1994). The *so* gene encodes a 416 amino acid protein that contains a sequence similar to the helix−turn−helix (HtH)[1] motif of the homeodomain. The homeodomain consists of three α helices and an N-terminal arm with the α2 and α3 helices forming the HtH motif (Figure 1A). The α3 helix is involved in specific DNA contacts with the major groove, while the basic N-terminal arm is involved in specific minor groove contacts (Kornberg, 1993). SO residues 240−279 are homologous to the homeodomain HtH region. However, residues 218−240 show no significant homology to the α1 helix and completely lack the conserved basic residues of the N-terminal arm (Figure 1A).

SO exhibits a higher degree of homology with the PBC homeodomain family, which includes Exd, Pbx, and ceh-20 (Mann & Chan, 1996), than the Hom-C homeodomain family (McGinnis & Krumlauf, 1992). Analysis of alignments between SO and Exd suggests additional homology N-terminal to the HtH, if an insertion is postulated in the SO sequence. N-Terminal basic residues 204−208 of SO align with the basic residues of the N-terminal arm of Exd (Figure 1B).

SO is representative of a new class of proteins that includes four of murine origin (Six1, Six2, Six3, and AREC3) and a putative *C. elegans* protein (Accession No. 1469143). These proteins exhibit a region of strong homology (46% identical) from residues 112−274 of SO which includes the HtH region (Figure 1C). The Six proteins are expressed in visual, neural, and connective tissue during development of the mouse (Oliver et al., 1995a,b). AREC3 is involved in regulating the Na,K-ATPase α1 subunit gene (Kawakami et al., 1996). The *C. elegans* protein has unknown function and was discovered through the Nematode Sequencing Project (Wilson et al., 1994).

Here we describe the identification of a minimal, autonomous HtH-containing domain capable of specific DNA recognition *in vitro*. Since sequence homologies suggested several potential N-terminal end points for the HtH domain, we investigated the structural and DNA-binding properties of three HtH-containing peptides with different N-termini (Figure 1C).

The SO HtH DNA-binding specificity was determined by covalently linking the chemical nuclease 1,10-phenanthroline−Cu (OP−Cu) to a unique native cysteine present at amino acid 224 (Figure 1C). The OP moiety has been used to investigate and identify DNA-binding sites of various transcription factors and domains (Sigman et al., 1991). We use SO−OP peptides to locate preferred binding sites in a 3.1 kb plasmid.

MATERIALS AND METHODS

Chemicals and Reagents. Neocuproine and 3-mercaptopropionic acid were obtained from Sigma. IAAOP was a

† Supported by NIH Grant GM46513.

* To whom correspondence should be addressed. Phone: (219) 631-8412. Email: mossing.1@nd.edu.

‡ F.L.S. was supported by a Walther Cancer Institute postdoctoral fellowship.

⊗ Abstract published in *Advance ACS Abstracts,* March 15, 1997.

[1] Abbreviations: SO, Sine Oculis protein; OP−Cu, 1,10-phenanthroline:copper(I); IAAOP, iodoacetamido-OP; neocuproine, 2,9-dimethyl-1,10-phenanthroline; HtH, helix−turn−helix; HOM-C, homeotic complex; PBC, Exd/Pbx/ceh-20 family; OD, optical density; BSA, bovine serum albumin; DTT, dithiothreitol; NP40, Nonidet P-40; bp, base pair(s); Tris, tris(hydroxymethyl)aminoethane; IPTG, isopropyl thio-β-galactoside; dNTP, deoxynucleotide triphosphate; EDTA, ethylenediaminetetraacetic acid; BME, β-mercaptoethanol; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; LB, Luria−Bertani broth; PEI, poly(ethylenimine); CD, circular dichroism spectropolarimetry; PCR, polymerase chain reaction; F_u, fraction unfolded; T_m, temperature at which half of the protein has undergone the unfolding transition. Specific SO peptides are denoted by residue numbers from the native sequence, for example, $SO_{218-279}$ represents residues 218−279 of SO plus an N-terminal methionine. Peptides conjugated to OP-Cu via Cys 224 are denoted by -OP, for example, $SO_{218-279}$-OP. Cysteine linked to OP-Cu was denoted by Cys-OP.

S0006-2960(96)02520-2 CCC: $14.00

gift from D. S. Sigman. All chemicals were reagent grade. All chromatographic steps were performed on a Pharmacia FPLC system.

Cloning. Expression plasmids for the various SO HtH peptides were constructed by inserting PCR products into a pET11a vector (Studier et al., 1990). Primers were designed to generate HtH peptides with a common C-terminus but varying N-termini. The N-terminal primers were:

SO N-188 5′ GCGCATCATATGGAAGCCGAAAAACTG 3′

SO N-204 5′ TTGGCACATATGCGTGTTCGCCGTAA 3′

SO N-218 5′ GGGCGGCATATGGGCGAGGAGACGAGCTAC 3′

Primers corresponding to N-terminal codons added an initiating methionine codon within a *Nde*I site (underlined) at the 5′ end of the *so* gene fragment.

The single C-terminal primer contained C-terminal *so* codons followed by a termination codon and a *Bam*HI restriction site (underlined):

SO C-279
5′ GGGGGATCCTTATTTGTGTTCGGCAGCTCG 3′

These primers generated PCR products that had an initiating methionine codon followed by native SO codons which encoded residues 188−279, 204−279, and 218−279. The template used in the PCR reaction was 100 pg of plasmid *so*^{mda} cDNA6 (Serikaku & O'Tousa, 1994), and PCR conditions were 10 mM Tris (pH 9 at 25 °C), 50 mM KCl, 1% Triton X-100, 0.2 mM dNTP's, and 5 μM primers. The PCR reaction was cycled 30 times at 94 °C for 1 min, 57 °C for 2 min, and 72 °C for 30 s. The PCR products were concentrated and desalted on a Sephadex G-50 spin column before digestion with *Nde*I and *Bam*HI. Digested PCR fragments and digested pET11c vector were excised from a low melting point agarose gel after staining with ethidium bromide. Excised gel slices were melted and used directly in ligation reactions. The ligation products were transformed into *E. coli* strain JM109 (DE3). Plasmid DNA from the transformants was screened by PCR, using a T7 promoter primer and the C-terminal primer. The three constructs carry genes encoding 63, 77, and 93 amino acid residues corresponding to plasmids pET11c $SO_{218-279}$, pET11c $SO_{204-279}$, and pET11c $SO_{188-279}$, respectively.

Expression and Purification of SO Peptides. The peptides were purified from JM109 (DE3) containing the plasmids pET11c $SO_{218-279}$, pET11c $SO_{204-279}$, and pET11c $SO_{188-279}$. Cultures were grown in 2 L of LB broth containing 100 mg/mL ampicillin in 6 L flasks, at 37 °C with shaking at 300 rpm. Protein expression was induced with 1 mM IPTG when the cultures reached an absorbance of 1.5 at 600 nm, and were grown 3 h postinduction (Studier et al., 1990). SO polypeptides typically constituted 5−10% of total cellular protein as monitored by SDS−PAGE.

Cultures were centrifuged, and approximately 5 g (wet weight) of bacterial pellet was resuspended in 40 mL of sonication buffer (20 mM K_2HPO_4, pH 7,500 mM NaCl, 2 mM EDTA, 5% glycerol, 5 mM BME, and 1 mM PMSF). Samples were sonicated until the initial OD_{600} was reduced by 80%, followed by addition of PEI to a final concentration of 0.6% with stirring at 4 °C for 15 min. Subsequent centrifugation at 18000*g* removed cell debris and precipitated nucleic acids. Protein in the PEI supernatant was precipitated with 50 g of ammonium sulfate/100 mL of culture. The resulting pellet was resuspended in 25 mL of resuspension buffer (20 mM K_2HPO_4, pH 7, 200 mM NaCl, 2 mM EDTA, 5% glycerol, 5 mM BME, and 1 mM PMSF) and dialyzed against 20 mM K_2HPO_4, pH 7, 1 mM EDTA, and 2.5 mM BME.

The dialysis sample (25 mL) was loaded onto a 55 mL Q-Sepharose Fast Flow (SFF) column which was connected in series to a 20 mL S-SFF column. Both columns were washed with 140 mL of buffer A (20 mM K_2HPO_4, pH 7, 1 mM EDTA, and 2.5 mM BME) until the absorbance at 280 nm returned to base line from a peak value of 0.5. The bound SO peptide was eluted from the S-SFF column (approximately 0.5−0.7 M NaCl) with 10 column volumes (200 mL) of a 0−1 M NaCl gradient containing buffer A. SO peptide-containing fractions were detected by SDS−PAGE, and fractions were pooled and concentrated by ammonium sulfate precipitation and separated by gel filtration chromatography on a 120 mL Superdex 75 column. The concentration of the purified peptide was determined using extinction coefficients (Gill & von Hippel, 1989) at 280 nm of 15 720, 21 520, and 22 920 M^{-1} for $SO_{218-279}$, $SO_{204-279}$, and $SO_{188-279}$ respectively.

CD of SO Peptides. Circular dichroism experiments were performed on an AVIV 62DS spectrometer. Spectra were measured in a 2 mm path length cell at 25 °C. The concentration of each peptide was 20 μM in 20 mM potassium phosphate buffer (pH 7.0) containing 0.1 mM EDTA, 5 mM BME, and 0.2 M NaCl. The wavelengths scanned were in the range of 200−250 nm, and a series of 4 scans of each spectrum were averaged and base line corrected. Thermal denaturation was monitored at 222 nm while the temperature was varied from 15 to 85 °C in 2.5 °C increments with a 2 min equilibration time. CD signals characteristic of the folded and unfolded species at a given temperature were obtained by linear extrapolation (Mollah et al., 1996). The following equation was used to obtain the fraction of protein unfolded (F_u) at a given temperature:

$$F_u = (Y_f - Y)/(Y_f - Y_u)$$

where Y is the CD signal at a particular temperature, Y_f is the extrapolated signal of the folded base line, and Y_u is the extrapolated signal of the unfolded base line. T_m is the temperature at which F_u equals 0.5.

Modification of SO Peptides. SO peptides at 100−170 μM were incubated in 20 mM Tris (pH 8.4), 50 mM DTT for 60 min at 4 °C to reduce any disulfide bonds. The reduced peptides were extensively dialyzed against 0.1 M acetic acid to remove excess DTT and prevent oxidation of free sulfhydryl groups. Derivatization was initiated by adjusting the pH of 620 μL of dialyzed peptide solution with 25 μL of 1 M Tris (pH 8.4), 7.5 μL of 4 M NaCl, 1 μL of 500 mM EDTA, and 15.5 μL of 4 M NaOH followed by immediate treatment with 60 μL of 23 mM IAAOP (Sigman et al., 1991). The reaction was incubated overnight at 4 °C and quenched by adding DTT to 20 mM DTT. Excess unreacted IAAOP was removed by dialysis against a solution containing 10 mM Tris (pH 7.5), 50 mM NaCl, and 1 mM EDTA. Cys-OP was generated by the addition of a 5-fold excess of cysteine over IAAOP. The extent of derivatization was determined spectrophotometrically. Derivatized protein spectra have a peak at 270 nm and increased absorbance due to the attached phenathroline group compared to the underivatized protein. The derivatized proteins generated spectra equivalent to underivatized protein with an equal

molar ratio of IAAOP, indicating the proteins are greater than 95% derivatized. Concentrations of the modified peptides were determined by Bradford assay (Bradford, 1976) and stored in aliquots at −20 °C.

DNA Cleavage. Screening of binding sites was performed on the plasmid pUC119 (Vieira & Messing, 1987). The numbering for pUC119 starts with the first base derived from the lac operon (NCBI sequence ID g464018). Procedures for OP-Cu conjugation and DNA cleavage generally followed those of Sigman or Ebright and colleagues (Sigman et al., 1991; Shang et al., 1994a). Derivatized HtH peptides (500 nM) in a 17.5 μL solution containing 20 mM HEPES (pH 7.5), 100 mM NaCl, 50 μg/mL BSA, 0.02% NP40, 5% glycerol, 2.5 mM 3-mercaptopropionic acid, and 10 μM $CuSO_4$ were incubated with 2.5 μL (500 ng) of *Eco*RI linearized plasmid. The reaction was stopped by addition of neocuproine to 3 mM. The samples were extracted once with phenol/chloroform before ethanol precipitation. Double-stranded cleavage was visualized after electrophoresis on agarose gels and staining with ethidium bromide.

Nucleotide resolution mapping of cleavage sites was performed on the plasmid pUC119 Alu521. A 521 bp *Alu*I fragment was subcloned into the *Sma*I site of the pUC119 to generate the plasmid pUC119 Alu521. This *Alu*I fragment consisted of nucleotides 1959−2480 from pUC119. Approximately 12 μg of pUC119 Alu521 was linearized at the *Eco*RI site and radiolabeled with α-^{32}P, phenol/chloroform extracted, ethanol precipitated, and digested with *Pvu*II, resulting in 751 and 92 bp fragments that are singly end-labeled. The DNA was desalted by a Sephadex G-50 spin column, giving a final DNA concentration of less than 80 nM. Reaction mixtures (40 μL) containing various concentrations (30−120 nM) of SO peptides were incubated with 1 nM pUC119 Alu521 (labeled at the *Eco*RI site, digested with *Pvu*II), 20 mM HEPES (pH 7.5), 100 mM NaCl, 50 μg/mL BSA, 0.02% NP40, 5% glycerol, 2.5 mM 3-mercaptopropionic acid, and 10 μM $CuSO_4$. The reactions were incubated at room temperature for 4 h and terminated by adding neocuproine to a final concentration of 3 mM. The samples were phenol/chloroform extracted, and 10 μg of tRNA was added before ethanol precipitation. Samples were resuspended in formamide buffer (95% formamide, 10 mM EDTA, 0.1% Bromophenol Blue and Xylene Cyanol dyes) for electrophoresis on a 6% acrylamide [19:1 acrylamide: bisacrylamide) ratio] sequencing gel. Chemical sequencing ladders were generated according to standard methods (Maxam & Gilbert, 1980). The 92 bp fragment is electrophoresed to the end of the gel and does not interfere with the analysis.

DNase I Footprinting. SO DNA-binding sites were localized on the Alu521 plasmid by DNase I protection (Hudson & Fried, 1989). The same batch of DNA used in the nucleotide resolution cleavage experiments was used as the substrate for footprinting. Protein−DNA complexes were formed by incubating SO peptide (31−250 nM) with 10 nM pUC119 Alu521 (labeled at the *Eco*RI site and digested with *Pvu*II) in a 9 μL solution containing 10 mM HEPES (pH 7.5), 200 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 2 mM DTT, 50 μg/mL BSA, and 2 μg/mL sonicated salmon sperm DNA. The protein−DNA mixture was allowed to equilibrate for 30 min before addition of 1 μL of 3.2 μg/mL DNase I. The solution was mixed and incubated at room temperature for 1 min, and the reaction was stopped by the addition of 20 μL of formamide buffer. Samples were heated at 95 °C for 5 min before 5 μL was loaded onto a 6% acrylamide sequencing gel.

RESULTS

Expression and Purification of SO HtH-Containing Peptides. Three SO polypeptides were expressed and purified from *E. coli*. The 63, 77, and 93 residue peptides corresponded to SO residues 218−279, 204−279, and 188−279, respectively. Immediately after sonication, proteolytic products consisting of up to 20% of the total recombinant peptide were evident in the crude extracts of $SO_{204-279}$ and $SO_{188-279}$. These were removed in subsequent chromatographic steps. The final purity of each peptide was greater than 95% as estimated by SDS−PAGE. The molecular mass of the purified peptides was confirmed by electrospray mass spectrometry (University of Notre Dame Mass Spectrometry Facility) to be within 3 daltons for $SO_{204-279}$ and $SO_{188-279}$ of their expected values. $SO_{218-279}$ had a molecular mass which corresponded to a peptide that has the initiator methionine removed. Hydrolysis of the methionine residue is likely since the penultimate amino acid is a glycine (Kendall et al., 1990). The yield of SO protein was 5 mg/L of *E. coli* culture.

The three peptides have a common end point at the C-terminus of the HtH but varied N-termini. $SO_{218-279}$ contains 63 amino acids, the characteristic size of the homeodomain. $SO_{204-279}$ contains an additional 14 residues at the N-terminus and includes basic residues that have homology to Exd. $SO_{188-279}$ is N-terminally elongated by 30 residues from $SO_{218-279}$ for a total of 93 residues (Figure 1C).

Expression of SO peptides from other clones in *E. coli* was also investigated. Full-length *so* in various expression vectors failed to express detectable levels of protein. A series of N-terminal deletions was generated with Bal 31 nuclease. These genes were expressed as fusions to the T7 gene 10 leader peptide. SO_{71-416} and $SO_{102-416}$ were proteolytically degraded and failed to accumulate to high levels. Other constructs (SO_{71-279}, $SO_{102-279}$, $SO_{134-279}$, and $SO_{171-279}$) exhibited strong expression but formed inclusion bodies and were resistant to initial refolding attempts.

SO HtH Peptides Are Folded. CD experiments showed that the peptides had significant amounts of nonrandom structure and exhibited cooperative unfolding transitions. CD spectra for all three peptides had minima at 208 and 222 nm indicative of α-helix-containing proteins. Molar ellipticities plotted in Figure 2A show that $SO_{204-279}$ has additional CD intensity at 222 nm compared to $SO_{218-279}$. This indicates that the N-terminal residues of $SO_{204-279}$ may contribute some secondary structure. In contrast, the similarity of the CD spectra for $SO_{188-279}$ and $SO_{204-279}$ is consistent with minimal secondary structure in the additional residues (188−204) (Figure 2). The mean residue ellipticity at 222 nm for $SO_{218-279}$ was 12 500, which predicts ∼38% α-helical content if a mean residue ellipticity of 33 000 is assumed for 100% helix (Yang et al., 1986). This should only be taken as a rough estimate, however, as the mean residue ellipticities measured for several homeodomains of known structure underrepresented their true helical content (Ades & Sauer, 1994).

Thermal denaturation monitored at 222 nm indicated that all three peptides exhibit cooperative unfolding curves. Thermal denaturation was reversible. $SO_{218-279}$, $SO_{204-279}$,

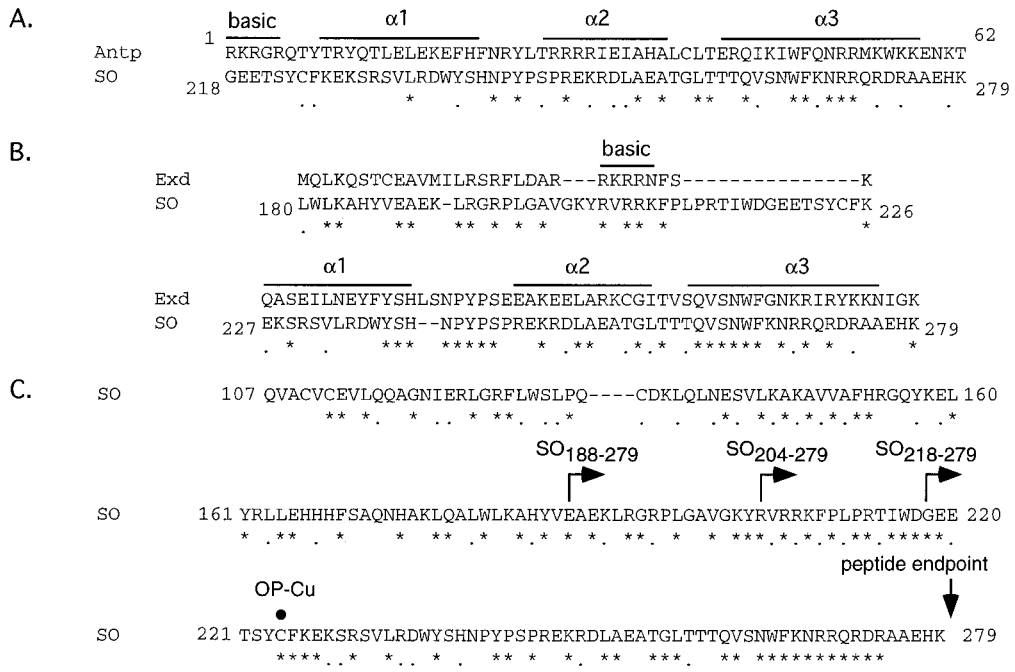

FIGURE 1: SO sequence homologies. (A) Alignment of SO with homologous regions of the Hom-C homeodomain Antennapedia (Antp). There is extensive homology with helices α2 and α3 but little homology with the basic N-terminal arm and α1. (B) SO homology with Exd. There is strong homology within α2 and α3 but weak homology through α1 of Exd. Additional homology in the N-terminus allows the alignment of the N-terminal arm residues with basic residues of SO and requires the insertion of a 15 amino acid gap in the Exd sequence. (C) SO family homologies and peptide end points. SO homology of residues 107−279 with consensus residues based on the SO class of proteins which includes Six1, Six2, Six3, AREC3, and a *C. elegans* protein (Accession No. 1469143). Residues which are identical among all six proteins are marked by an asterisk (∗) while similar residues were marked by a period (.). Gaps in the amino acid sequence were marked by a hyphen (-). Sequence alignments were done by the Clustalw (1.5) program (Thompson et al., 1994). The peptides used in this study are indicated by arrows marking the three start points and the single end point. The site of OP-Cu derivatization of Cys 224 is marked by a closed circle. $SO_{218-279}$ is equivalent to the minimal HtH domain while $SO_{204-279}$ and $SO_{188-279}$ are N-terminal elongations which include regions homologous to Exd.

and $SO_{188-279}$ have T_ms of 52.5, 52.5, and 55 °C, respectively. These T_ms are in the upper range of values (41−55 °C) reported for other homeodomains (Ades & Sauer, 1994; Clarke et al., 1994; Damante et al., 1994; Otting et al., 1988).

Screening of pUC119 for DNA-Binding Sites. The SO peptides were derivatized at amino acid residue 224, the single native cysteine in each of the peptides. Cys 224 was alkylated with an iodoacetamide derivative of 1,10-phenanthroline (OP). The three phenanthroline conjugate peptides were designated $SO_{218-279}$-OP, $SO_{204-279}$-OP, and $SO_{188-279}$-OP.

The plasmid pUC119 was screened for specific DNA-binding sites with the three peptides $SO_{218-279}$-OP, $SO_{204-279}$-OP, and $SO_{188-279}$-OP. The plasmid was linearized with *Eco*RI before exposure to SO-OP peptides at 500 nM. The three SO peptides generated similar cleavage patterns that included two predominant bands that were 1.3 and 1.9 kb in size (Figure 3). Further mapping indicated that the site was near nucleotide 2150 in the β-lactamase gene region (Figure 4A).

Numerous double-stranded cleavage products of varying intensity are observed, suggesting that there are many binding sites throughout the substrate but none equal the intensity of the site near 2150. The overall pattern of cleavage was similar between the three SO-OP peptides. Exposure of pUC119 with Cys-OP indicated no detectable cleavage at equivalent concentrations and a uniform background of cleavage at a higher concentration (10 μM) (data not shown).

Nucleotide Resolution Mapping of DNA-Binding Sites. Nucleotide resolution mapping of the binding site observed in the screen was carried out with $SO_{204-279}$-OP on the

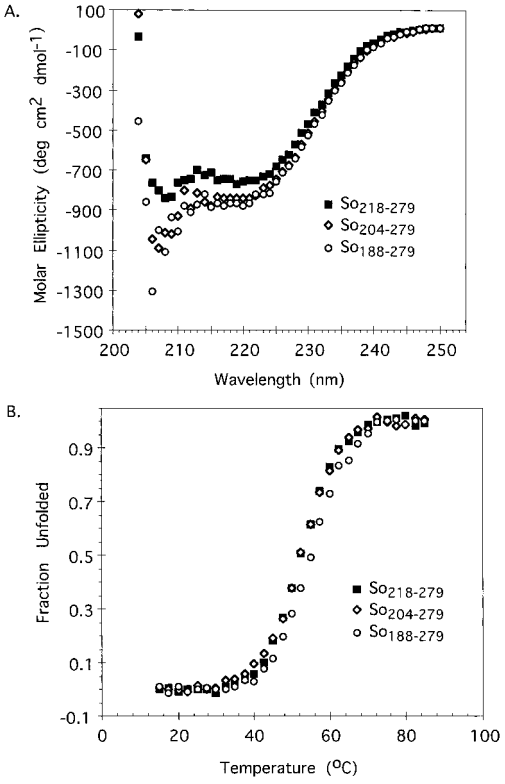

FIGURE 2: CD studies of SO HtH peptide fragments. (A) CD spectrum of the indicated SO HtH peptides measured from 200 to 250 nm represented as molar ellipticity. The concentrations of the peptides were 20 μM. (B) Thermal denaturation of the indicated SO HtH peptides as monitored by the CD signal at 222 nm. Fractions unfolded are plotted as a function of temperature.

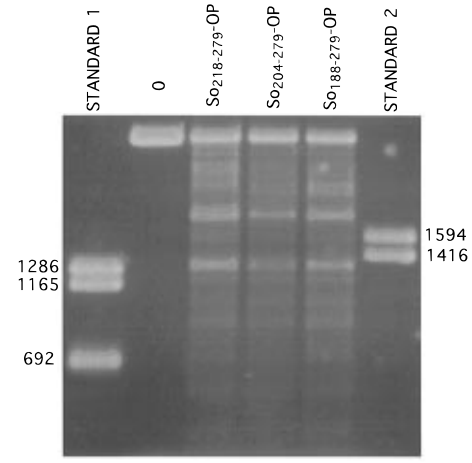

FIGURE 3: DNA cleavage pattern generated by SO-OP derivatives on pUC119. The 3.16 kb plasmid pUC119 was linearized with *Eco*RI and exposed to SO-OP peptides (500 nM) in solution. The DNA was then electrophoresed on an agarose gel. The cleavage pattern for all three SO-OP peptides was similar with two predominant bands of 1.3 and 1.9 kb. The control lane is *Eco*RI-linearized pUC119 which is exposed to the cleavage buffer in the absence of SO-OP. Standard 1 is a double digest of pUC119 with *Eco*RI and *Dra*I while standard 2 is a double digest of pUC119 with *Eco*RI and *Bgl*I.

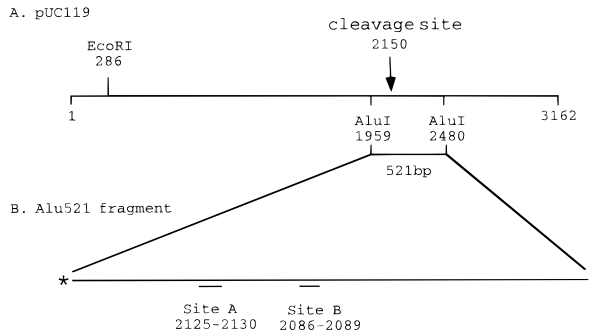

FIGURE 4: Schematic map of pUC119 with SO HtH-binding sites. (A) The approximate site of cleavage by SO-OP peptides in pUC119 determined by agarose gel electrophoresis. The cleavage site is approximately at nucleotide 2150 in the β-lactamase gene. The 521 bp *Alu*I fragment containing the cleavage area was subcloned into the multiple cloning site of pUC119 to facilitate end-labeling. (B) Alu521 fragment within pUC119 was end-labeled at the *Eco*RI site for nucleotide resolution mapping of SO-OP cleavage and DNase I footprint analysis. The two sites identified are indicated and numbered based on their nucleotide positions in pUC119.

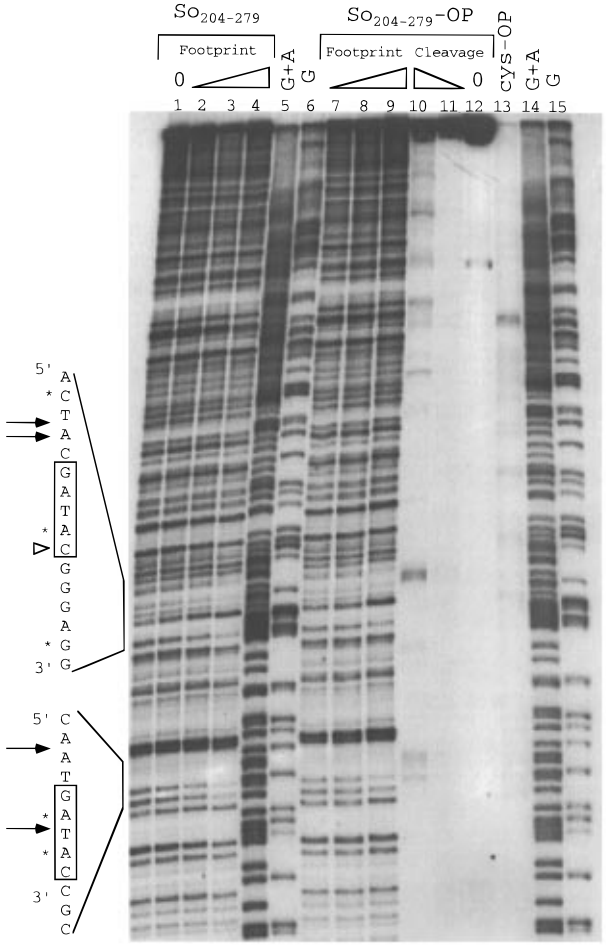

FIGURE 5: Nucleotide resolution mapping of the cleavage area identified in the screen for binding sites in pUC119. The 751 bp end-labeled fragment described in Figure 4B was used for footprinting and nucleotide resolution cleavage analysis. Lanes 1−4 are a DNase I footprint of $SO_{204-279}$ at 0 nM, 31 nM, 125 nM, and 500 nM. Lanes 5 and 6 are a chemical sequencing ladder. Lanes 7−9 are a DNase I footprint of $SO_{204-279}$-OP at 31 nM, 125 nM, and 500 nM. Lanes 10−12 contain 62.5 nM, 15.6 nM, and no $SO_{204-279}$-OP, respectively. Lane 13 contains Cys-OP at 2.5 μM. Lanes 14 and 15 are a chemical sequencing ladder. The sequences of the protected areas are indicated up to the boundaries where bases are not protected. The actual bases showing protection are marked by an asterisk (*). The hypersensitive site for site A is marked by an open arrowhead. The approximate area of SO-OP cleavage within the footprint sequence is indicated by filled arrows.

Alu521 fragment. Two predominant cleavage areas were observed and designated site A and site B (Figure 4B). Site A exhibited multiple cleavage bands at nucleotides 2131−2133 of pUC119. Site B exhibited a predominant cluster of bands at nucleotides 2092−2093. Secondary cleavage of site B was observed at nucleotides 2088−2089 (Figure 5). A structural explanation of these observed cleavage patterns is beyond the scope of the present study. However, we can propose two alternatives to explain the additional cleavage observed in site B. It may be either that the SO peptide occupies alternative binding sites on the DNA or that the region of the protein including residue 224, the site of OP-Cu attachment, can adopt alternative conformations.

DNase I footprint analysis with $SO_{204-279}$ showed protection of these same cleavage sites with both the native and OP-Cu-conjugated SO HtH peptide (Figure 5). The two protected areas, corresponding to cleavage site A and site B, contained an identical 5 bp core sequence of 5′-GATAC-3′ although site A was the only site that exhibited DNase I hypersensitivity. The strongest cleavage was at the 5′ end when compared to the protected area and the 5 bp core sequence. Site A exhibited the greatest intensity of cleavage. Its core sequence of CGATAC is unique to the pUC119 plasmid. The site B sequence, TGATAC, is present 3 times in the pUC119 plasmid. A computer search of pUC119 indicates that many related sites are present throughout the plasmid. Other SO-OP scission sites of lesser intensity are observed which do not have corresponding footprints at the same concentrations of SO. The cleavage pattern observed by Cys-OP was different from SO-OP. In order to achieve a similar amount of cleavage, a 40-fold higher concentration of Cys-OP was required. Even at these elevated concentrations, site selectivity was reduced with respect to the SO-OP conjugate (lane 13). Similar results for all three peptides were obtained although only $SO_{204-279}$ is shown in Figure 5.

DISCUSSION

Genetic studies indicate the involvement of SO in the development of the larval and adult visual system of *Drosophila melanogaster*. The nuclear localization, expression pattern, and developmental role of SO implicate it in regulating genes controlling the morphogenesis of the *Drosophila* visual system (Cheyette et al., 1994; Serikaku & O'Tousa, 1994). Further evidence of a role as a transcription factor was the initial finding that SO contained a region homologous to homeodomains. Although originally described as a homeodomain, the homology is limited to the HtH motif of the homeodomain and lacks the conserved N-terminal arm basic residues. Further N-terminal homologies with the PBC homeodomains suggest that an insertion may have shifted the position of the basic region relative to the HtH motif.

On the basis of the PBC family sequence homologies, we designed, expressed, and purified three SO polypeptide fragments containing the HtH homology region but different N-terminal sequences. The addition of residues to the N-termini was potentially important structurally and functionally. Structurally, the omission of residues at the termini of a peptide could prevent it from adopting a fully native structure (Darby et al., 1996; Taniuchi, 1973). Functionally, in the case of homeodomains, the deletion of the basic N-terminal arm can reduce DNA-binding affinity (Qian et al., 1994; Shang et al., 1994b).

The SO HtH Region Is an Autonomously Folded Region with Nonrandom Secondary Structure. CD spectra of the 63 amino acid peptide $SO_{218-279}$ showed evidence of modest α-helix content consistent with the HtH homology. Thermal denaturation of $SO_{218-279}$ produced a cooperative unfolding curve, indicating the peptide to be autonomously folded. The CD spectra and thermal denaturation profiles of the N-terminally elongated peptides $SO_{188-279}$ and $SO_{204-279}$ are qualitatively similar but indicate that additional N-terminal residues may contribute to the absolute helix content and the thermal stability of the 63 amino acid helical domain. The addition of N-terminal residues in $SO_{204-279}$ was found to contribute additional ellipticity at 222 and 208 nm when compared to $SO_{218-279}$ although there was no thermal stabilization. However, further N-terminal elongation in $SO_{188-279}$ resulted in a thermal stabilization by 2.5 °C with no additional contribution to the helical CD signal. The 93 amino acid peptide is the longest that we have characterized. Although the 63 amino acid peptide is autonomously folded, upon addition of N-terminal residues the CD spectra and thermal stabilization suggest an interaction with other parts of the SO protein. In order to characterize the atypical N-terminal region, $SO_{204-279}$ is being studied by NMR. Preliminary results confirm the presence of two helices forming the HtH region common with other homeodomains, but also show that the extra N-terminal residues excluding the first four are ordered which differs from the typical homeodomain (F. L. Stahura, unpublished results). The refinement of the solution structure is in progress.

SO HtH Peptides Bind DNA with Moderate Specificity. DNA scission patterns by SO-OP indicated two predominant scission sites which correlated directly to the DNase I footprint of both derivatized and native peptides. DNase I protection of binding sites with native SO peptide corresponded to the cleavage sites, indicating that the attachment of OP did not alter the DNA binding specificity of the peptide. The two binding sites identified had the core sequence 5′-(C/T)GATAC-3′. Although the homeodomain consensus sequence 5′-TAAT-3′ is present in multiple copies within the DNA analyzed, these were not recognized. Interestingly, the core sequence 5′-(C/T)GATAC-3′ is related to the Pbx1- binding site of 5-′TTGAT-3′ with the identical bases underlined (Lu & Kamps, 1996). The numerous scission events observed in the experiments (both the double-strand cleavage screen and the nucleotide resolution mapping experiment) indicate that there are many secondary sites that can be detected. We have sought to identify high-affinity sites by *in vitro* selection (Blackwell & Weintraub, 1990; Tuerk & Gold, 1990). Although nanomolar concentrations of SO peptides were sufficient to cause gel shifts of the initial random pool of double-stranded oligonucleotides, affinities failed to increase over multiple rounds of selection (T. R. Hazbun and M. C. Mossing, unpublished results). These results are also in accord with reports of promiscuous *in vitro* binding by other homeodomains (Kornberg, 1993). The SO domains studied here may be less selective than other isolated homeodomains that have been studied *in vitro*.

Use of the OP−Cu Moiety To Find and Investigate DNA-Binding Sites. The DNA cleaving agent OP-Cu has been used to test predictions of high-resolution structures of protein−DNA complexes and to confirm DNA-binding sites. The OP moiety was used to confirm structural predictions of the Engrailed homeodomain−DNA complex (Pan et al., 1995) and identify binding sites for the Msx-1 homeodomain (Shang et al., 1994a), CAP protein (Pendergrast et al., 1994), and Fis protein (Pan et al., 1996). In the case of the conjugated Fis protein, a search for binding sites in a bacterial DNA substrate was performed without previous knowledge of the presence of binding sites. Our work extends earlier work by finding DNA-binding sites for SO-OP without any previous knowledge of the SO HtH domain DNA-binding specificity.

The SO-OP peptides generated scission predominantly at the 5′ end of the footprint similar to other N-terminally derivatized homeodomains which cleave at the 5′ end of their respective DNA-binding sites (Pan et al., 1995; Shang et al., 1995). A series of Engrailed−OP conjugates derivatized at various positions helped assess the Engrailed cocrystal structure and specifically the N-terminal arm position relative to the DNA (Pan et al., 1995). The fact that cleavage occurred indicates that Cys 224 is positioned near the DNA at some stage during DNA binding by the SO peptides. Further investigation and finer scale mapping of SO-OP cleavage of specific nucleotides within the DNA-binding site may be useful as protein structural information becomes available.

Basic N-Terminal Residues Homologous to the Homeodomain N-Terminal Arm Do Not Contribute to in Vitro DNA-Binding Specificity. Complementary to the identification of DNA sequences recognized by a protein is the identification of the protein sequences responsible for DNA binding. The addition of N-terminal residues homologous to Exd could potentially alter DNA-binding specificity. The similarity in cleavage patterns of pUC119 between the three SO-OP conjugated peptides indicated that these residues do not alter DNA-binding specificity or affinity. We do not believe that the conjugation of OP to Cys 224 has prevented contributions from the N-terminal extensions to DNA-binding specificity, since DNase I footprinting of each of the peptides gives qualitatively similar protection patterns

whether or not Cys 224 is conjugated to OP-Cu (data not shown). If these residues are involved in DNA binding, then they require cooperative interactions with other proteins or need other interactions within the SO protein to assist in specific DNA binding.

Comparison with Other DNA-Binding Systems. Recently DNA binding has been demonstrated *in vitro* for AREC3 (Kawakami et al., 1996), another member of the SO family. AREC3 was initially identified as an activity in nuclear extracts from mammalian tissue culture cells which bound specifically to the control region (ARE) of the Na,K-ATPase α1 subunit gene. Expression of AREC3 fusion proteins has allowed the identification of fragments of AREC3 which are active in DNA binding. Full specific binding activity has been reported for proteins with N- and C-terminal deletion boundaries which correspond to residues 150 and 270 of SO, respectively. Interestingly, proteins in which the α3 homology has been deleted but N-terminal sequences retained, as well as a minimal homeodomain fragment corresponding to residues 200−288 of SO, have been reported to show "nonspecific" binding. The ARE sequence used for these studies, GGNGNCNGGTTGC, shows no recognizable homology to the sequences identified here. It may be that additional sequence specificity will be revealed in SO fragments which contain more N-terminal sequences.

Homeodomains were initially thought of as self-sufficient, independent DNA-binding domains. Recent evidence indicates that other parts of the protein and other proteins can modulate specificity. For example, association of PBC proteins with Hox proteins can alter the specificity and affinity of DNA binding (Mann & Chan, 1996). The SO HtH domain has homology to the interaction loop between α1 and α2 of the Pbx homeodomain. OP-conjugated SO peptides may therefore be a useful tool to study cooperative DNA binding between SO and possible interacting partners.

Regardless of whether additional SO sequences or the cooperation of other proteins is required for full specificity, it is important to first investigate the structural and functional characteristics of the minimal domain. The studies reported here are the first steps in this process.

Conclusions. We have identified a minimal, independently folding DNA-binding domain from the Sine Oculis protein of *Drosophila melanogaster* which contains homology to the helix−turn−helix region of the homeodomain. Copper−phenanthroline conjugates of SO peptides have been used to identify the sequence GATAC as a common feature in DNA sites bound by this domain. Although 63 amino acids are sufficient for DNA recognition, we have presented evidence that additional N-terminal sequences can participate in the folded structure. These studies represent initial steps in the characterization of the structure and function of a new class of helix−turn−helix DNA-binding proteins with important roles in development.

ACKNOWLEDGMENT

We thank David S. Sigman for the gift of IAAOP, Alexander G. Lappin for the use of the CD spectropolarimeter, William C. Boggess for the electrospray mass spectrometry analysis, and Joseph E. O'Tousa and Michelle A. Serikaku for the *so*^{mda} cDNA.

REFERENCES

Ades, S. E., & Sauer, R. T. (1994) *Biochemistry 33*, 9187−9194.
Blackwell, T. K., & Weintraub, H. (1990) *Science 250*, 1104−1110.
Bradford, M. M. (1976) *Anal. Biochem. 72*, 248−254.
Cheyette, B. N., Green, P. J., Martin, K., Garren, H., Hartenstein, V., & Zipursky, S. L. (1994) *Neuron 12*, 977−996.
Clarke, N. D., Kissinger, C. R., Desjarlais, J., Gilliland, G. L., & Pabo, C. O. (1994) *Protein Sci. 3*, 1779−1787.
Damante, G., Tell, G., Leonardi, A., Fogolari, F., Bortolotti, N., Di Lauro, R., & Formisano, S. (1994) *FEBS Lett. 354*, 293−296.
Darby, N. J., Kemmink, J., & Creighton, T. E. (1996) *Biochemistry 35*, 10517−10528.
Gill, S. C., & von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319−326.
Hudson, M. J., & Fried, M. G. (1989) *BioTechniques 7*, 812−815.
Kawakami, K., Ohto, H., Ikeda, K., & Roeder, R. G. (1996) *Nucleic Acids Res. 24*, 303−310.
Kendall, R. L., Yamada, R., & Bradshaw, R. A. (1990) *Methods Enzymol. 185*, 398−407.
Kornberg, T. B. (1993) *J. Biol. Chem. 268*, 26813−26816.
Lu, Q., & Kamps, M. P. (1996) *Mol. Cell. Biol. 16*, 1632−1640.
Mann, R. S., & Chan, S.-K. (1996) *Trends Genet. 12*, 258−262.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499−560.
McGinnis, W., & Krumlauf, R. (1992) *Cell 68*, 283−302.
Mollah, A. K., Aleman, M. A., Albright, R. A., & Mossing, M. C. (1996) *Biochemistry 35*, 743−748.
Oliver, G., Mailhos, A., Wehr, R., Copeland, N. G., Jenkins, N. A., & Gruss, P. (1995a) *Development 121*, 4045−4055.
Oliver, G., Wehr, R., Jenkins, N. A., Copeland, N. G., Cheyette, B. N., Hartenstein, V., Zipursky, S. L., & Gruss, P. (1995b) *Development 121*, 693−705.
Otting, G., Qian, Y. Q., Muller, M., Affolter, M., Gehring, W., & Wuthrich, K. (1988) *EMBO J. 7*, 4305−4309.
Pan, C. Q., Landgraf, R., & Sigman, D. S. (1995) *Protein Sci. 4*, 2279−2288.
Pan, C. Q., Johnson, R. C., & Sigman, D. S. (1996) *Biochemistry 35*, 4326−4333.
Pendergrast, P. S., Ebright, Y. W., & Ebright, R. H. (1994) *Science 265*, 959−962.
Qian, Y. Q., Resendez-Perez, D., Gehring, W. J., & Wuthrich, K. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 4091−4095.
Serikaku, M. A., & O'Tousa, J. E. (1994) *Genetics 138*, 1137−1150.
Shang, Z., Ebright, Y. W., Iler, N., Pendergrast, P. S., Echelard, Y., McMahon, A. P., Ebright, R. H., & Abate, C. (1994a) *Proc. Natl. Acad. Sci. U.S.A 91*, 118−122.
Shang, Z., Isaac, V. E., Li, H., Patel, L., Catron, K. M., Curran, T., Montelione, G. T., & Abate, C. (1994b) *Proc. Natl. Acad. Sci. U.S.A. 91*, 8373−8377.
Sigman, D. S., Kuwabara, M. D., Chen, C. H., & Bruice, T. W. (1991) *Methods Enzymol. 208*, 414−433.
Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60−89.
Taniuchi, H. (1973) *J. Biol. Chem. 248*, 5164−5174.
Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994) *Nucleic Acids Res. 22*, 4673−4680.
Tuerk, C., & Gold, L. (1990) *Science 249*, 505−510.
Vieira, J., & Messing, J. (1987) *Methods Enzymol. 153*, 3−11.
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., & et al. (1994) *Nature 368*, 32−38.
Yang, J. T., Wu, C. S., & Martinez, H. M. (1986) *Methods Enzymol. 130*, 208−269.

BI9625206