Protein Science (1995), *4*:2279–2288. Cambridge University Press. Printed in the USA.

Drosophila engrailed-1,10-phenanthroline chimeras as probes of homeodomain–DNA complexes

CLARK Q. PAN,[1] RALF LANDGRAF,[2] AND DAVID S. SIGMAN[1,2,3]

[1] Molecular Biology Institute, University of California at Los Angeles, Los Angeles, California 90095-1570
[2] Department of Biological Chemistry, University of California at Los Angeles, Los Angeles, California 90095-1570
[3] Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, California 90095-1570

(RECEIVED June 8, 1995; ACCEPTED August 18, 1995)

Abstract

We have converted the *Drosophila* engrailed homeodomain into a sequence-specific nuclease by linking the protein to the chemical nuclease 1,10-phenanthroline-copper (OP-Cu). Unique cysteines were introduced at six positions into the homeodomain by site-directed mutagenesis for the covalent attachment of OP-Cu. The varied DNA-binding affinity and specificity of these mutants and the DNA cleavage pattern of their OP-Cu derivatives allowed us to assess the crystal structure of the engrailed homeodomain–DNA complex. We have also achieved site-specific double-stranded DNA scission with one of the homeodomain mutants, E28C, which has the potential of being used to identify engrailed binding sites in the genome. Because the homeodomain is so well conserved among members of the homeodomain-containing protein family, other homeodomain proteins can be converted into nucleases by attaching OP-Cu at position 28 of their homeodomains.

Keywords: chemical nuclease; homeodomain; 1,10-phenanthroline-copper; protein–DNA interaction

The homeodomain is a 60-amino acid DNA-binding motif that plays a crucial role in eukaryotic gene regulation during development (Gehring et al., 1994a, 1994b; Kenyon, 1994; Krumlauf, 1994; Lawrence & Morata, 1994; Weigel & Meyerowitz, 1994). It was initially discovered in a family of *Drosophila* proteins that control segmentation and assign anatomical features to the appropriate segment. This family of homeodomain-containing proteins now numbers more than 300, including members in the fungal, animal, and plant kingdom (Duboule, 1994). Their homeodomains are highly conserved in primary amino acid sequence and perhaps more so in three-dimensional structure. Several structural models of homeodomain–DNA complexes have been determined. X-ray crystal structures of *Drosophila* engrailed (Kissinger et al., 1990), yeast MAT α2 (Wolberger et al., 1991), and mammalian Oct-1 (Klemm et al., 1994) homeodomains bound to DNA and an NMR structure of the *Drosophila* Antennapedia homeodomain–DNA complex (Otting et al., 1990) were very similar. These structures revealed an N-terminal arm that makes contacts with the first two bases of the consensus 5′-TAATNN-3′ sequence in the minor groove and a core region of three α-helices, in which one of the helices interacts with the last four bases of the consensus sequence in the major groove. Because homeodomains show little sequence selectivity and tend to bind nonconsensus (A+T)-rich sequences in vitro, residues outside of the homeodomain and additional factors may be involved in sequence-specific recognition in vivo (Kornberg, 1993). Identifying the DNA sequences targeted by homeodomain proteins in vivo, however, has so far proven to be difficult because many homeodomain proteins were isolated based on sequence homology with existing members of the family.

A well-characterized example of a homeodomain protein is engrailed. It has been found in various species of arthropods, annelids, and vertebrates (Duboule, 1994). In *Drosophila*, the segment-polarity gene *engrailed* and its partner *wingless* are required for defining the segmental borders of the embryo (Siegfried et al., 1992). These two genes are expressed in adjacent stripes of cells that eventually become the anterior and posterior limits of each parasegment. The initiation of their expression is regulated by the pair rule genes, but their continued expression during development is supported by intercellular communication. *Engrailed* signaling is required to maintain *wingless* expression, which in return is essential to the continuous expression of *engrailed*.

DNA-binding proteins can be converted into site-specific nucleases by linking them to the chemical nuclease 1,10-phenanthroline-copper (OP-Cu) (Pan et al., 1994b). The *trp* repressor was the first protein converted into a chimeric nuclease by using iminothiolane to bridge 1,10-phenanthroline (OP) to surface lysines on the protein (Chen & Sigman, 1987). More recently, a two-step protocol has been developed. It involved mutagen-

Reprint requests to: David S. Sigman, Molecular Biology Institute, University of California at Los Angeles, Los Angeles, California 90095-1570; e-mail: sigman@ewald.mbi.ucla.edu.

izing selected amino acid residues to cysteines and then alkylating the sulfhydryl groups with an iodoacetamide derivative of OP. This protocol was initially applied to the λ phage cro protein (Bruice et al., 1991) and subsequently to the *Escherichia coli trp* repressor (Sutton et al., 1993) and the factor for inversion stimulation (Fis) protein (Pan et al., 1994a) in our laboratory and to the catabolite activator protein (CAP) by Ebright and colleagues (Pendergrast et al., 1994). These chimeric scission reagents can be used to test predictions based on high-resolution NMR and X-ray crystal structures and to locate binding sites of proteins within genomes. The important regulatory functions of homeodomain proteins make this DNA-binding motif an attractive candidate for the conversion to a scission reagent, especially because strong structural and sequence similarity indicates that the identification of a suitable position for the attachment of OP to any member of this family can serve as a guideline for most other homeodomain proteins.

While the studies described here were in progress, Shang et al. (1994) converted the homeodomain protein Msx-1 into a scission reagent and were able to identify its binding site in the transcriptional control region of the developmental regulatory gene Wnt-1. The choice of position for OP attachment was based on the structure of engrailed and was the equivalent of the engrailed homeodomain amino acid 10 (Fig. 6A). This approach utilizes a residue at the periphery of the protein that does not participate in any crucial protein-DNA contact and gives the OP considerable freedom of movement. If the cysteine residue is modified with OP through an acetylglycylamino spacer, OP-Cu can be close enough to the C-1H target to carry out scission. The Msx-1-OP chimera was used to identify a site within a 3.6-kb fragment of DNA even though the reported scission yield was only 5–15%. These experiments have demonstrated that amino acid 10 of the Msx-1 homeodomain is close to the 5′ end of the consensus 5′-(C/G)TAATTG-3′ sequence in the Msx-1–DNA complex. This suggested that the orientation of the Msx-1 homeodomain relative to DNA is analogous to that of Antp, engrailed, MAT a2, and Oct-1 homeodomains.

Our own work has focused exclusively on the engrailed homeodomain itself because of the availability of a high-resolution X-ray structure for this particular homeodomain–DNA complex. In order to optimize the scission efficiency, six positions on the protein have been mutagenized and derivatized with OP with different spacers. This has allowed us to evaluate the model of the engrailed homeodomain–DNA complex derived from the X-ray crystal structure and to define further the properties of these chimeras that lead to efficient scission agents.

Results

Design of the engrailed-OP derivatives

Virtually quantitative DNA scission has been observed with two homodimeric binding proteins with relatively high affinities for their respective binding sites. With the *E. coli trp* repressor, the OP was positioned rigidly near the dyad axis of the protein. In the case of the CAP, efficient scission was achieved when OP was placed on the periphery because the DNA is sharply bent toward the protein's flanks (Pendergrast et al., 1994). Scission efficiency appears to be optimal if the OP-linked copper ion is held rigidly in the minor groove, within 3–4 Å of the C-1H.

In contrast to other proteins that have been converted into nucleases, the homeodomain binds to the DNA presumably as a monomer with no protein-induced bends in the DNA. This not only eliminates the use of the dyad axis as a stable environment to mount OP-Cu for attack of the minor groove, it also leads to a relatively low stability of the homeodomain–DNA complex. Our work with the engrailed homeodomain therefore focuses on alternative methods of targeting to allow efficient scission in a class of proteins that may act as monomers with a rather low affinity for their respective binding sites and do not bend the DNA.

Positions selected for mutation

We have made six cysteine mutants (K2C, R3C, T6C, L13C, E28C, and A54C) of the engrailed homeodomain to allow the covalent linkage of each mutant with OP (Fig. 6A). The different mutants represent three distinct approaches to the conversion of the homeodomain into a scission reagent. A54C was made to test the option of major groove targeting, which, based on the data accumulated thus far, was unlikely to succeed, but had never been tested experimentally. L13C and E28C are remote from the recognition element; therefore, the tethered OP should not interfere with protein–DNA interactions. Finally, K2C, R3C, and T6C were designed to utilize the minor groove binding arm as the structural element that should keep OP close to its minor groove target.

OP derivatives

The first phenanthroline derivative synthesized suitable for targeted scission was 5-iodoacetamido-1,10-phenanthroline (IAOP) (Chen & Sigman, 1987). Ease of synthesis was the primary reason for choosing this structure. Straightforward syntheses of 5-iodoacetylglycylamino-1,10-phenanthroline (IAGOP) and 5-iodoacetyl-β-alanylamino-1,10-phenanthroline (IAβAOP) via phthaloyl glycyl and β-alanyl derivatives are now available. The β-alanyl derivative is superior to the glycyl derivative because the latter readily undergoes an as yet incompletely characterized rearrangement.

DNA affinity and specificity

We have performed gel-retardation experiments (Fig. 1) to analyze the relative specific and nonspecific DNA-binding affinity of engrailed homeodomain mutants and their OP-Cu derivatives (estimated apparent K_d's shown in Table 1). The specific DNA-binding site was the same one used for co-crystallization of the engrailed homeodomain–DNA complex. This sequence, which in fact might not be a relevant in vivo site, includes the consensus 5′-TAAT-3′ sequence cloned into the pUC19 vector. A 55-bp piece that contains the cloned site was used as the probe for "specific binding" and a 180-bp DNA fragment of identical specific radioactivity and concentration, derived from the same restriction digest, was used to provide an internal standard for "nonspecific" binding.

Major groove targeting

Chemical analysis has identified the deoxyribose C-1H as the prime target for oxidative scission by the free $(OP)_2Cu^+$ com-

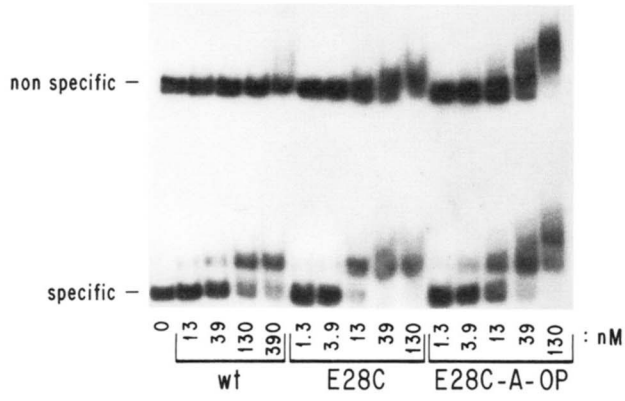

Fig. 1. Gel-retardation assay of wild-type (wt) engrailed homeodomain, mutant E28C, and derivatized E28C-A-OP binding to two DNA fragments. The 55-bp "specific" piece of DNA contains the 21-bp sequence used by Kissinger et al. (1990) for co-crystallization with the engrailed homeodomain. The 180-bp "nonspecific" DNA fragment is derived from the same restriction digest that released the specific fragment.

plex (Sigman, 1990). This position is within the minor groove and protein–OP constructs have been designed assuming a similarity in chemical mechanism. However, the oxidative scission catalyzed by the 1:1 OP–Cu complex attached to the protein could conceivably proceed through an alternative mechanism. If this were the case, the placement of OP on the recognition helix might force such an alternative mechanism of scission in the major groove to become dominant and could provide a high yield of scission due to the rigid placement of the OP provided by the recognition helix. Only Ala-54 appears to be oriented properly with the major groove for the OP-Cu without direct involvement in DNA recognition. However, the A54C mutation abolished specific DNA binding. Even if major groove reactivity is possible, identifying a residue to modify without interfering with specific binding is unlikely.

Placement at periphery

The mutations of either residue 13 or 28 at the outer flanks of the protein had no effect on the specificity of DNA binding, as is expected from their location. The placement of OP at either one of these positions is analogous to the derivatization of position 10 of Msx-1. Glutamate 28 is on the opposite side of the N-terminal arm, far away from most of the specific contacts between the homeodomain and DNA. In the crystal structure, the shortest distance from the side chain of residue 28 to the DNA is 5.9 Å: between one of the carbonyl oxygens of glutamate and one of the phosphate oxygens between guanine 18 and adenosine 19 on the bottom strand. Removing this negative charge close to the DNA in the E28C mutant and E28C-A-OP derivative increased the overall DNA affinity by an order of magnitude (Table 1). This suggests that glutamate 28 is close enough to the phosphate backbone to have a destabilizing impact on the protein–DNA complex. Because the electrostatic repulsion is between the glutamate and the DNA backbone but not the bases, no increase in the binding specificity has been observed when it is replaced. In fact, E28C-A-OP had a lower binding specificity, perhaps due to the nonspecific interaction contributed by OP-Cu.

In the gel matrix, E28C-A-OP (short three-atom acetamido linker arm) yields 10–20% scission centered at adenine 21 on the top strand (Fig. 2A, lane 2), and approximately 5% scission cen-

Table 1. *Estimated K_d's (nM) of the engrailed homeodomain–DNA complexes*

	Specific	Nonspecific	Specific/nonspecific
Wild type	90	1,000	11
K2C	80	170	2
K2C-A-OP	30	130	4
T6C	16	90	6
T6C-A-OP	29	200	7
L13C	260	3,000	12
L13C-A-OP	460	720	2
L13C-AG-OP	110	540	5
E28C	6	65	11
E28C-A-OP	8	46	6

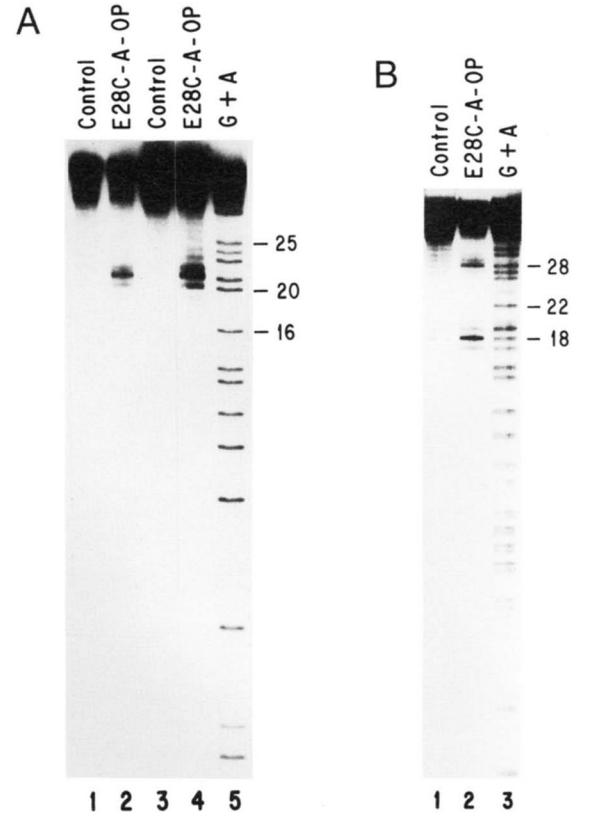

Fig. 2. DNA scission patterns by the engrailed homeodomain E28C-A-OP on a 50-bp substrate. The DNA sequence is numbered in the same way as in Figure 5. **A:** Top strand DNA scission. The free DNA band (lane 1) and E28C-A-OP:DNA complex band (lane 2) from the gel retardation assay were isolated and subjected to DNA scission. DNA was also treated in solution with MPA/Cu (lane 3) and MPA/Cu + E28C-A-OP (lane 4). **B:** Bottom strand gel-matrix DNA scission. Lane 1 depicts the free DNA band and lane 2 shows the E28C-A-OP:DNA complex band from the gel retardation assay.

tered at guanine 18 on the bottom strand (Fig. 2B, lane 2) of a 50-bp substrate. The solution cleavage efficiency is slightly weaker (Fig. 2A, lane 4). The 3′-staggered pattern of cleavage, apparent when cleavage is examined on both strands, implies a minor-groove approach of OP-Cu to DNA. The scission efficiency on the top strand is higher than the bottom strand because the linker has to loop around the phosphate backbone of the bottom strand to reach the cleaved guanine 18.

We also performed gel-matrix DNA scission by E28C-A-OP on a 110-bp DNA substrate (Fig. 3). This DNA fragment was released from pUCenB by a double restriction digest with *Eco*R I and *Hin*d III. For the top strand, the strong scission centered at adenine 21 is again seen (Fig. 3A, lane 2). Double-stranded DNA scission was also detected (Fig. 3B, lane 2; Fig. 3C).

The derivatization of position 13 with IAOP resulted in a surprising reduction in the ability of the protein to bind DNA specifically (Table 1), a feature that was not observed for E28C and not reported for Msx-1-OP derivatized at position 10. This effect on the DNA-binding affinity is reduced when position 13 is derivatized with another OP derivative possessing a longer six-atom acetylglycylamino linker (L13C-AGOP). Specific scission can be observed at guanine 10 of the top strand and thymidine 8 of the bottom strand (Fig. 5), but the yield is below that found for residue 28. The analysis of the location of scission sites obtained with the longer linker arm indicates that the OP-Cu must reach around the minor groove binding element of engrailed to reach its target. The effect of the shorter linker arm on the binding specificity is therefore probably due to a direct competition between OP-Cu and the minor groove binding element. The C_α–C_β trajectory of residue 13 is such that the OP would be directed toward the face of the DNA helix, which is covered by the minor groove binding arm, whereas residue 10 is more likely to direct OP to the opposite face thereby avoiding this conflict (Fig. 6A). In summary, both L13C-AG-OP and E28C-A-OP were successful in creating a DNA scission reagent. Although these chimera confirmed the orientation of the engrailed homeodomain on the recognition sequence, only modest scission yields were obtained due to the flexibility of the OP-Cu and the low affinity of the protein for the recognition sequence (on average approximately 10^{-8} M).

The minor groove binding element

Though successful in generating a scission reagent, the modification of residues 10, 13, and 28 located at the outer flank of Msx-1 and engrailed did not provide the scission yield that would

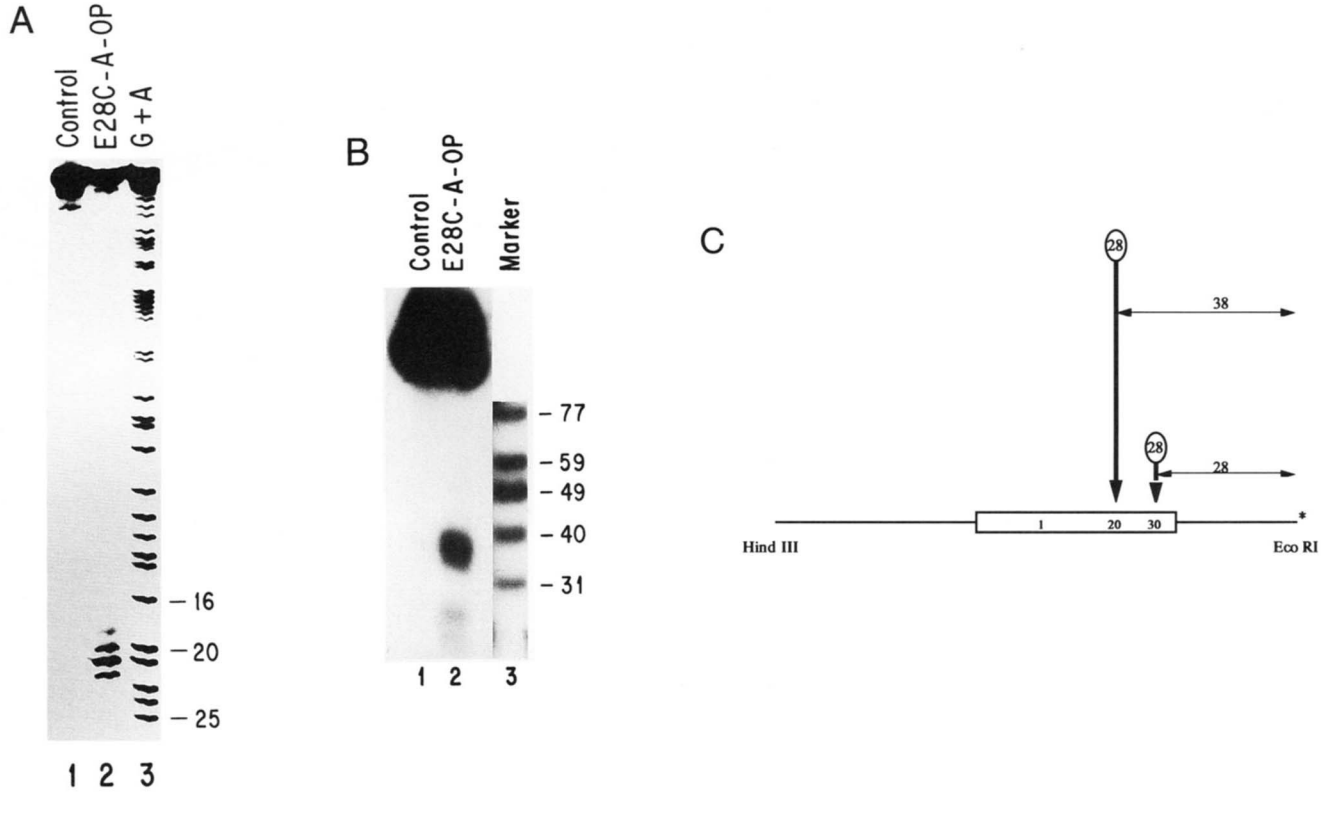

Fig. 3. Gel-matrix DNA cleavage by the engrailed homeodomain E28C-A-OP on a 110-bp substrate. The DNA sequence is numbered in the same way as in Figure 5. **A:** Top strand DNA scission. The free DNA band (lane 1) and E28C-A-OP:DNA complex band (lane 2) from the gel retardation assay were isolated and subjected to DNA scission. **B:** Double-strand DNA scission. Lane 1 depicts the free DNA band and lane 2 shows the E28C-A-OP:DNA complex band from the gel retardation assay. **C:** An illustration of the 110-bp substrate. The 50-bp substrate used in Figure 2 is boxed with the DNA sequence numbered in the same way as in Figure 5. The 110-bp substrate is 3′-radiolabeled at the *Eco*R I site. The size of the two fragments generated from the E28C-A-OP cleavage is shown as the distance to the *Eco*R I site. The relative efficiency of DNA cleavage is approximated by the length of the arrows.

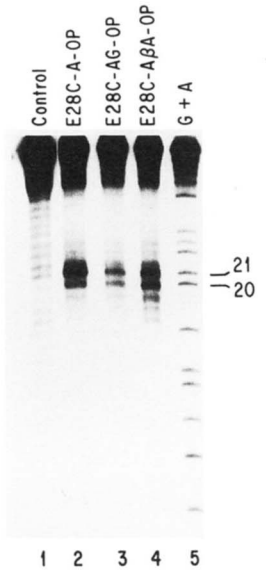

Fig. 4. Gel-matrix DNA cleavage on a 50-bp substrate by OP-Cu covalently linked to the engrailed homeodomain E28C through different spacers. The DNA sequence is numbered in the same way as in Figure 5. Top strand DNA scission is shown here. Free DNA band (lane 1) and homeodomain–DNA complex band of E28C-A-OP (lane 2), E28C-AG-OP (lane 3), and E28C-AβA-OP (lane 4) from the gel retardation assay were isolated and subjected to DNA scission.

be required to screen large genomic DNA for binding sites. The N-terminal minor groove binding arm of the homeodomain provides an interesting target for derivatization with OP in that it might allow a design in which OP is kept rigidly fixed in the minor groove. In earlier work, when OP was linked to the minor groove binding ligand Hoechst 33258, cleavage was observed at 1/1,000 the concentrations as that observed with free OP (Chen et al., 1993). Figure 6B highlights the minor-groove recognition element. Residues 0–2 and the side chain of Arg-3 are not resolved in the crystal structure. We mutagenized residues 2, 3, and 6 into cysteines for subsequent attachment of OP. These three residues have very different characteristics. Lys-2 has the highest degree of flexibility and a mutation in this position is very unlikely to affect the binding behavior. Arg-3 is implicated in the crystal structure as possibly providing a contact to thymidine 12 on the bottom strand, although the side chain is poorly resolved and is not as conserved as Arg-5. The possibility existed therefore, that the contact made by Arg-3 is of minor importance for binding and that a conjugated OP, which possesses some DNA affinity by itself, might compensate for the loss of binding affinity. The introduction of a positive charge after the chelation of copper might mimic the replaced arginine even closer. Finally, Thr-6, like Lys-2, is not involved in minor groove binding but is firmly constrained by minor groove contacts of adjacent Arg-5.

The RC3 mutation caused a complete loss of specific affinity. All residual DNA-binding affinity was nonspecific. Alkylation of R3C with iodoacetamido-OP did not restore binding. The K2C mutation on the other hand reduced the affinity for the specific binding site approximately fivefold, indicating that it is more important for protein–DNA interaction than the crystal structure has indicated (Table 1). However, these interactions are probably not specific because attaching OP to K2C caused a partial recovery of binding affinity and specificity. This could be due to hydrophobic interaction of the OP replacing charge interactions of the lysine side chain. The derivatized protein does cut the top strand of the recognition sequence at thymine 9, but scission is too weak for an accurate assessment of the yield

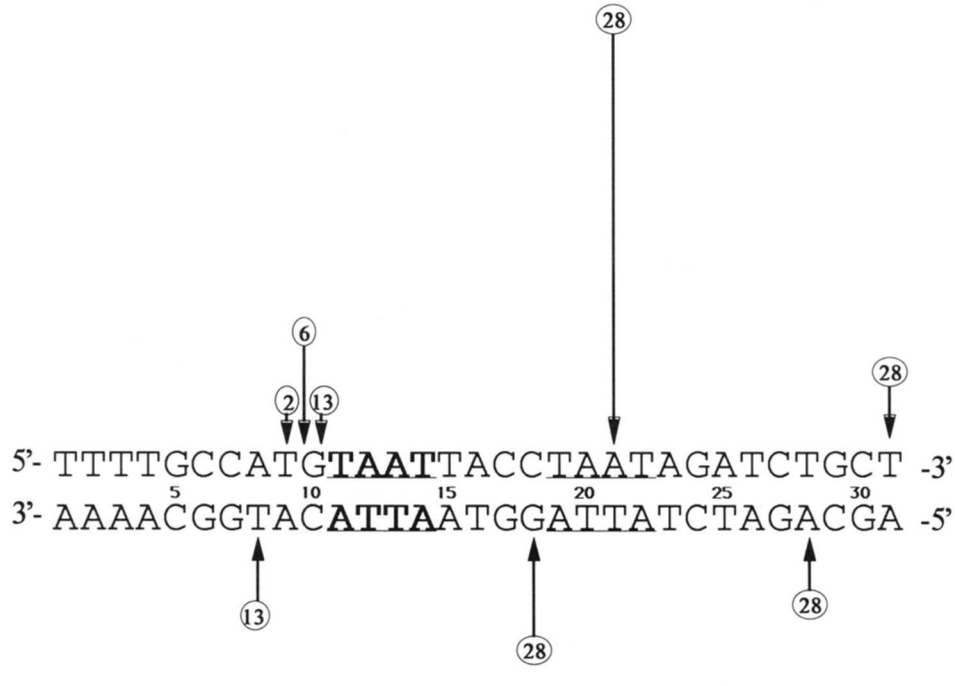

Fig. 5. Summary of the DNA scission patterns of the 50-bp substrate by the engrailed homeodomain chimeras K2C-A-OP, T6C-A-OP, L13C-AG-OP, E28C-A-OP. The 5′-TAAT-3′ sequences of the two engrailed homeodomain binding sites are underlined with that of the 5′-TAATTA-3′ bolded. The relative DNA cleavage efficiencies are roughly proportional to the length of each arrow.

(Fig. 5). The location of the scission site does, however, suggest that the linker arm in fact stretches along the minor groove with a considerable degree of freedom. This high degree of freedom is probably the reason for its marginal yield of scission.

Of the minor groove binding element, mutagenesis of Thr-6 to a cysteine should provide an efficient cleaver. The T6C mutation slightly increased binding (Table 1). Perhaps, the conversion to a slightly smaller amino acid at the densely packed protein-DNA interface creates room for additional contacts with DNA by adjacent residues in the homeodomain. OP-derivatization of T6C increased the affinity even further at the expense of roughly a factor of two in binding specificity. These effects could be due to the possible interactions of OP-Cu with the DNA, as described for K2C-A-OP. Scission by T6C-A-OP occurs exclusively on the top strand of the DNA. The position of cleavage sites and the low yield can be explained by the molecular model for the homeodomain-DNA complex (Fig. 6B). The C_α-C_β trajectory of Thr-6 forces the linker arm to bend back to reach its target, guanine 10, by inserting between the remainder of the N-terminal arm and the DNA top strand. The bottom strand becomes inaccessible to OP-Cu due to a blockage by Arg-5. This constrained placement is likely to result in an equilibrium in which most of the time the attached OP points away from the DNA rather than into the minor groove. Further analysis of the remaining residues close to the minor groove did not identify residues with better orientation characteristics.

Optimizing the scission by chimeras of E28C

A careful study of the crystal structure of the engrailed homeodomain-DNA complex has led us to choose six positions for OP-Cu attachment. Out of these six, residue 28 yields the most efficient DNA scission. OP-Cu linked to position 28 has a clear path to the DNA, whereas those attached to positions 6 and 13 have to compete with the N-terminal arm to reach the DNA. Furthermore, Glu-28 sits on a rigid part of the homeodomain as a part of the second helix so that its OP-Cu should have less flexibility as compared to that of Lys-2. However, modeling of E28C-A-OP complexed to DNA (Fig. 6A) suggested that higher DNA cleavage efficiency may be achieved by increasing the linker length between the engrailed homeodomain and OP. To improve the DNA scission efficiency of E28C-A-OP, the protein was modified with OP derivatives with longer linker arms IAGOP and IAβAOP. Figure 4 shows that E28C-AG-OP cleaves the top strand of the 50-bp substrate with twofold lower efficiency relative to E28C-A-OP (the three-atom long acetamido linker), whereas the slightly longer E28C-AβA-OP yields twofold higher DNA scission efficiency. Thus, the tether length contributes to efficiency, but the precise dependence is not straightforward. Because E28C-AG-OP cleaves the lower strand threefold better than E28C-A-OP (data not shown), more detailed modeling may be essential to understand all the factors that contribute to chimera cleavage efficiency.

A new engrailed homeodomain binding site

E28C-A-OP exhibits highly specific DNA cleavage at nucleotides (nt) 18 and 21, resulting from the homeodomain interacting with the consensus 5′-TAATTA-3′ sequence from nt 11 to 16 (Fig. 5). However, an unexpected, though weaker cleavage is also observed at nt 28 when the 50-bp DNA fragment containing the sequence listed in Figure 5 was treated with E28C-A-OP (Fig. 2B, lane 2). The corresponding 3′-staggered cutting at nt 31 is not resolved from the parent band in the gel shown in Figure 2A but has been seen in other experiments (data not shown). In the experiment with the 110-bp substrate, the top-strand cleavage at nt 31 was run off the gel (Fig. 3A, lane 2), but the double-strand scission at 31 was detected (Fig. 3B, lane 2; Fig. 3C).

A closer examination of the sequence used for DNA scission experiments revealed the presence of the 5′-TAATAG-3′ sequence from nt 19 to 24 located 5′ to the cleaved nt 28 and 31 (Fig. 5). Ades and Sauer have used a gel shift selection and PCR amplification assay to identify the 5′-TAATTA-3′ sequence (the same sequence used in the co-crystallization experiment and in this study) as the consensus engrailed homeodomain binding site (Ades & Sauer, 1994). However, the 5′-TAATAG-3′ sequence has also been identified by their assay as one of the high-affinity binding sites. The difference between this sequence and that used in the co-crystallization experiment at the nucleotides immediately 3′ to the 5′-TAAT-3′ core recognition sequence resulted in different patterns of DNA cleavage by E28C-A-OP. 5′-TAATAG-3′ is cut two nucleotides 3′ relative to 5′-TAATTA-3′ when both sequences are aligned (Fig. 7). This suggests that the contacts made between the engrailed homeodomain and these two sequences have minor differences.

Discussion

We have performed site-directed mutagenesis on the engrailed homeodomain to covalently attach the chemical nuclease OP-Cu at unique positions on the protein. The varied DNA-binding affinity and specificity of these mutants and the DNA cleavage pattern of their OP-Cu derivatives have not only confirmed the crystal structure of the engrailed homeodomain-DNA complex but also provided further insights into how the N-terminal arm is positioned relative to the DNA. Furthermore, we have achieved site-specific double-stranded DNA scission on a 110-bp DNA substrate with one of these mutants, E28C-A-OP, and have improved upon its DNA cleavage efficiency by changing the linker arm length between OP-Cu and the homeodomain. These experiments have provided an opportunity for evaluating systematically the factors that lead to efficient protein-OP scission reagents.

Evaluation of the crystal structure

Our initial goal was to understand how the N-terminal arm (residues 3–9) is oriented in the minor groove because this arm has been proposed to be critical to DNA-binding specificity of the homeodomain (Gehring et al., 1994b). We have introduced three unique cysteine mutations in this region at positions 2, 3, and 6 (Fig. 6B). The backbone and the side chain of Lys-2 and the side chain of Arg-3 are not resolved in the crystal structure. DNA cleavage by K2C-A-OP, on the other hand, allowed us to place the side chain of residue 2 along the minor groove to make critical contacts with the DNA, in agreement with the lowered DNA-binding specificity exhibited by the lysine-to-cysteine mutation. Thr-6 is in the center of the N-terminal arm and is immediately adjacent to Arg-5, the most conserved residue in this region of the homeodomain (Duboule, 1994). The DNA scission pattern of T6C-A-OP permitted us to confirm the predic-

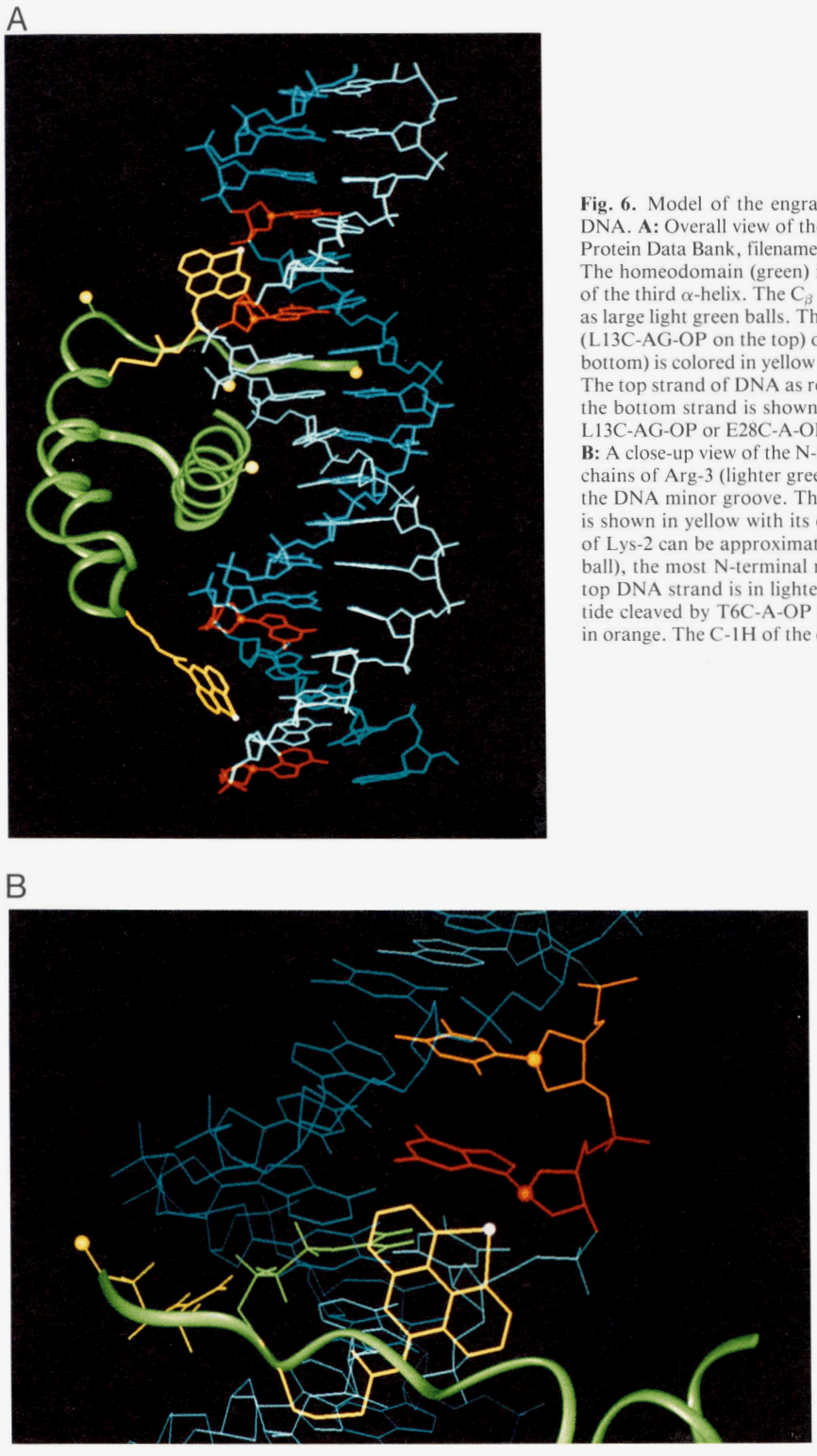

Fig. 6. Model of the engrailed homeodomain OP-Cu derivatives binding to DNA. **A:** Overall view of the homeodomain–DNA structure obtained from the Protein Data Bank, filename "pdb1hdd.ent" (Brookhaven National Laboratory). The homeodomain (green) is oriented such that the N-terminal arm is on top of the third α-helix. The C_β positions of residues 3, 6, 10, and 54 are illustrated as large light green balls. The OP on either the longer acetylglycylamino linker (L13C-AG-OP on the top) or the shorter acetamido linker (E28C-A-OP on the bottom) is colored in yellow with the copper atoms depicted as small pink balls. The top strand of DNA as represented in Figure 5 is colored in lighter blue and the bottom strand is shown in darker blue. The cleaved nucleotides by either L13C-AG-OP or E28C-A-OP are in red with the C-1H highlighted as small balls. **B:** A close-up view of the N-terminal arm of the homeodomain (green). The side chains of Arg-3 (lighter green) and Arg-5 (darker green) insert themselves into the DNA minor groove. The OP on the acetamido spacer linked to position 6 is shown in yellow with its copper atom depicted as a pink ball. The position of Lys-2 can be approximated by the backbone nitrogen of Arg-3 (light green ball), the most N-terminal residue resolved in the crystal structure. Again the top DNA strand is in lighter blue and the bottom in darker blue. The nucleotide cleaved by T6C-A-OP is shown in red and that cleaved by K2C-A-OP is in orange. The C-1H of the cleaved nucleotides are highlighted as colored balls.

tions made by the crystal structure on the orientation of the N-terminal arm relative to the DNA.

Besides focusing on the interaction between the N-terminal arm and the minor groove of DNA, we wanted to determine the overall orientation of the homeodomain relative to its specific DNA-binding site. The three residues that we mutated and modified with OP-Cu are Leu-13 on the left side, Ala-54 in the middle, and Glu-28 on the right side of the homeodomain (Fig. 6B). The DNA scission pattern of the two derivatives that showed significant DNA cutting supports the co-crystal structure of the

sequences	8	23
1	ATGTAATTACCTAATA	
2	ATTTAATTGA	
3	TAATAATAATAATAA	
4	ACCTAATAGATCTGCT	
	16	31

Fig. 7. Four engrailed homeodomain binding sites aligned by the core consensus homeodomain binding sequence 5′-*TAAT*-3′. Sequence 1 has been used by Kissinger et al. (1990) in the co-crystallization experiments with the homeodomain. Sequence 4 has been identified by E28C-A-OP. Both sequences are part of the same DNA substrates we used and are numbered in the same manner as in Figure 5. Sequence 2 represents the consensus of engrailed binding sites in the regulatory region of the *engrailed* gene and sequence 3 is the simple 5′-TAA-3′ repeat that also binds to engrailed.

engrailed homeodomain–DNA complex. For example, L13C-AG-OP cuts on the left side of the DNA site, whereas E28C-A-OP cuts on the right side (Figs. 5, 6B).

Specificity of the engrailed homeodomain–DNA complex

The engrailed homeodomain appears to be quite promiscuous in DNA recognition in vitro. DNAse I footprint experiments have demonstrated that engrailed can bind to clusters of DNA sites in the regulatory region of the *engrailed* gene (Desplan et al., 1988). These sites have the consensus sequence 5′-ATTT AATTGA-3′. However, the protein also binds to a simple sequence of 5′-TAA-3′ repeats. The 5′-TAATTA-3′ sequence used by Kissinger et al. for co-crystallization with the engrailed homeodomain is different from the above two. The only identity between these three sequences is the presence of the 5′-TAAT-3′ (Fig. 7). In our hands, the sequence of the co-crystal structure had only one order of magnitude higher affinity toward the engrailed homeodomain than a "nonspecific" piece of DNA.

In addition to cleaving at the primary binding site, the E28C-A-OP also cleaved at the sequence 5′-TAATAG-3′. Using a binding site selection assay, the consensus engrailed homeodomain binding sequence has been determined to be 5′-TAATTA-3′ with 93%, 100%, 100%, 98%, 88%, and 65% preference for the indicated nucleotides (Ades & Sauer, 1994). The site that we have identified differs from the consensus sequence at positions five and six, but is one of the 74 binding sites that have been isolated by Ades and Sauer. Although only 3% of the set of 74 binding sites have A at the fifth position, position 6 is G in 24% of this set of binding sites.

Potential use of homeodomain 28-OP to locate homeodomain binding sites

Besides evaluating the model of the engrailed homeodomain–DNA complex based on X-ray crystallography, engrailed-OP chimera may permit the identification of new engrailed binding sites. The most efficient DNA cleavage observed by us is when OP-Cu is attached to residue 28 of the homeodomain. Using a 110-bp substrate, E28C-A-OP not only exhibited double-stranded DNA cleavage of the previously known engrailed binding site (5′-TAATTA-3′) but also revealed a possible new engrailed site (5′-TAATAG-3′). However, the use of this derivative to identify engrailed sites in larger pieces of DNA is hindered by the low intrinsic sequence specificity of the homeodomain. With only one order of magnitude of preference for binding at its specific site when faced with many competing nonspecific sites in a kilobase or megabase piece of DNA, E28C-A-OP will have a difficult time to pick out its specific site.

In vivo studies of homeodomain-containing proteins suggest that their action is much more specific than that observed in vitro by the homeodomain itself (Lawrence & Morata, 1994). This can be explained by regions of the protein outside of the homeodomain or cofactors contributing to the specificity of the protein. The entire engrailed protein may have a higher affinity toward its specific sites than nonspecific ones, and attaching OP-Cu to the whole protein could allow easier identification of its specific DNA sites in the genome. Currently, few cofactors that improve the specificity of homeodomain proteins have been identified, and none has been found for engrailed. However, because the homeodomain-containing proteins are so similar structurally, tethering OP-Cu at residue 28 of those homeodomain proteins with known partners could yield efficient and specific DNA cleavage. Two possible systems that could be explored are the yeast homeodomain protein MAT α2 (Dranginis, 1990; Stark & Johnson, 1994) and POU-type homeodomains (Pomerantz & Sharp, 1994).

Materials and methods

Site-directed mutagenesis and protein purification

The plasmid pETenHB (Poole et al., 1985) containing the engrailed homeodomain in a pET3A vector was kindly supplied by Dr. Thomas Kornberg. We performed two-step PCR mutagenesis to generate the six cysteine mutants. The homeodomain was expressed in BL21 cells. The IPTG-induced cells were centrifuged. The cell pellet was resuspended and sonicated in a minimum volume of 25 mM HEPES pH 7.6, 1 mM PMSF, and 1 mM DTT. Cell debris was removed by centrifugation and the supernatant was loaded onto a Mono S column (Pharmacia). The engrailed homeodomain typically elutes at 400 mM KCl. One exception is the E28C mutant, which elutes from the column at 600 mM KCl.

Recombinant plasmid and DNA substrates

A Pharmacia Gene Assembler was used to make two complementary strands of DNA, each 50 nt in length. This synthetic DNA contains the 21-bp sequence used in the crystallization of homeodomain–DNA complex (Kissinger et al., 1990). The 50-mer was then cloned into the pUC19 vector at *Sma* I. We designate the resulting recombinant plasmid as pUCenB. For gel-retardation assays, pUCenB was triple digested with *Nar* I, *Kpn* I, and *Bam*H I to release the specific (55-bp) and the nonspecific (180-bp) DNA fragments. These two DNA fragments were not separated after the restriction digest to insure equal amounts of each. They were then radiolabeled at the 5′ end, extracted with phenol–chloroform, and desalted together.

Two substrates were used for DNA cleavage reactions. The first one is the synthetic 50-mer; the second is a 110-mer released from pUCenB by a double digest with *Eco*R I and *Hin*d III. This

110-mer substrate contains the 50-mer and allows better resolution of the cleaved products from the parent band in a native polyacrylamide gel.

Gel-retardation assay

The homeodomain was mixed with radiolabeled DNA probes in a solution of 10 mM NaCl, 20 mM Tris, pH 8, 5 mM $MgCl_2$, 25 μg/mL BSA, 1.25 μg/mL poly-dIdC, and 7% glycerol for 10 min. The mix was then loaded on a native 7% 0.5× TB polyacrylamide gel. We omitted EDTA in the gel and running buffer because we have found that EDTA chelates copper and inhibits its DNA cleavage by engrailed-OP derivatives.

OP derivatization

The protein is first fully reduced with a 5–10-fold molar excess of DTT, and the excess DTT is removed with a G25 spin-column. The reduced protein is then derivatized with a 5–10-fold molar excess of IAOP (or other OP derivatives) for more than 5 h, with the excess free OP completely removed by dialysis. All steps were carried out at 4 °C.

DNA cleavage

For gel-matrix scission, a gel-retardation assay is first performed and the bands corresponding to the free DNA and the engrailed-OP:DNA complex were excised and soaked in an equal volume of 100 μM $CuCl_2$, 600 μM MPA, 600 μM H_2O_2, 20 mM NaCl, and 40 mM Tris, pH 8, for 1–2 h. For solution-cutting, the engrailed-OP and its DNA substrate were mixed in a solution of 10 mM NaCl, 20 mM Tris, pH 8, 25 μg/mL BSA, 1.25 μg/mL poly-dIdC, 5 μM $CuCl_2$, and 30 μM MPA for approximately 20 min. We did not include Mg^{2+} in the reaction mixture because it may compete with copper for binding to OP. The scission reactions were quenched with 1.4 mM 2,9-dimethyl-OP and run on denaturing polyacrylamide gels to analyze single-strand DNA cleavage or on native gels to measure double-strand DNA scission.

Synthesis of IAβAOP

N-phthalyl-β-alanine was converted into the acid chloride using PCl_5 in benzene and then added to a solution of 5-amino-1,10-phenanthroline in CH_2Cl_2 containing triethylamine. The reaction mixture was heated at reflux temperature for 1 h. Following concentration on a rotary evaporatory, the reaction was heated at reflux temperature in 95% EtOH containing hydrazine. The desired product, 5-β-alanylamino-1,10-phenanthroline, was purified by silica gel column chromatography after the reaction mixture was acidified to remove phthalic hydrazide, made alkaline to pH 9.0, evaporated to dryness, and dissolved in MeOH. ^{1}H NMR (360 MHz, CH_3OH-d_4) δ 8.98 (dd, J = 4.3, 1.5 Hz, 1 H), 8.90 (dd, J = 4.3, 1.4 Hz, 1 H), 8.44 (dd, J = 8.4, 1.6 Hz, 1 H), 8.18 (dd, J = 8.1, 1.4 Hz, 1 H), 7.98 (s, 1 H), 7.64 (dd, J = 8.4, 4.3 Hz, 1 H), 7.59 (dd, J = 8.1, 4.4 Hz, 1 H), 3.08 (t, J = 6.4 Hz, 2 H), 2.74 (t, J = 6.4 Hz, 2 H). ^{13}C NMR (90 MHz, CH_3OH-d_4) δ 174.27, 150.74, 150.32, 146.76, 144.85, 137.42, 132.77, 132.68, 129.64, 126.13, 124.82, 124.24, 121.81, 39.55, 38.78.

To prepare IAβAOP, iodoacetic anhydride was added to a methanolic solution containing 5-β-alanylamino-1,10-phenanthroline and stirred for 10 min. The product was purified by silica gel chromatography (5% MeOH in $CHCl_3$) to yield a light yellow solid (260 mg, 80%). ^{1}H NMR (360 MHz, CH_3OH-d_4) δ 9.08 (dd, J = 4.3, 1.5 Hz, 1 H), 9.01 (dd, J = 4.4, 1.5 Hz, 1 H), 8.51 (dd, J = 8.4, 1.5 Hz, 1 H), 8.27 (dd, J = 8.1, 1.4 Hz, 1 H), 8.07 (s, 1 H), 7.72 (dd, J = 8.4, 4.4 Hz, 1 H), 7.66 (dd, J = 8.1, 4.4 Hz, 1 H), 3.94 (s, 2 H,-CHCOCH_2I), 3.60 (t, J = 6.5 Hz, 2 H), 2.80 (t, J = 6.5 Hz, 2 H). ^{13}C NMR (90 MHz, DMSO-d_6) δ 170.53, 167.91, 149.86, 149.31, 145.74, 143.74, 135.86, 131.91, 131.73, 128.05, 124.67, 123.62, 122.88, 120.17, 35.67, 35.46, 0.863 (-NHCOCH_2I).

IAGOP was prepared as described previously (Pendergrast et al., 1994).

Acknowledgments

We thank Dr. Young-Moon Cho for the synthesis of the IAGOP and IAβAOP and David Perrin for technical assistance during the synthesis of IAGOP. This work was made possible by National Institutes of Health grant USPHS GM21199 to D.S. C.P. was supported by a training grant from USPHS GM07185 and U.S. Department of Health and Human Services Biotechnology Training Program GNO8375. R.L. was supported by a Merck Academic Development Program Predoctoral Fellowship CO92591.

References

Ades S, Sauer R. 1994. Differential DNA-binding specificity of the engrailed homeodomain: The role of residue 50. *Biochemistry 33*:9187–9194.

Bruice TW, Wise J, Rosser DSE, Sigman DS. 1991. Conversion of lambda phage cro into an operator-specific nuclease. *J Am Chem Soc 113*:5446–5447.

Chen C-hB, Mazumder A, Constant JF, Sigman DS. 1993. Nuclease activity of 1,10-phenanthroline-copper. New conjugates with low molecular weight ligands. *Bioconj Chem 4*:69–77.

Chen C-hB, Sigman DS. 1987. Chemical conversion of a DNA-binding protein into a site-specific nuclease. *Science 237*:1197–1201.

Desplan C, Theis J, O'Farrell P. 1988. The sequence specificity of homeodomain-DNA interaction. *Cell 54*:1081–1090.

Dranginis A. 1990. Binding of yeast a1 and α2 as a heterodimer to the operator DNA of a haploid-specific gene. *Nature 347*:682–685.

Duboule D. 1994. *Guidebook to the homeobox genes*. Oxford, UK: Oxford University Press.

Gehring W, Affolter M, Burglin T. 1994a. Homeodomain proteins. *Annu Rev Biochem 63*:487–526.

Gehring W, Qian Y, Billeter M, Furukubo-Tokunaga K, Schier A, Resendez-Perez D, Affolter M, Otting G, Wuthrich K. 1994b. Homeodomain-DNA recognition. *Cell 78*:211–223.

Kenyon C. 1994. If birds can fly, why can't we? Homeotic genes and evolution. *Cell 78*:175–180.

Kissinger C, Liu B, Martin-Blanco E, Kornberg T, Pabo C. 1990. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: A framework for understanding homeodomain-DNA interactions. *Cell 63*:579–590.

Klemm J, Rould M, Aurora R, Herr W, Pabo C. 1994. Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell 77*:21–32.

Kornberg T. 1993. Understanding the homeodomain. *J Biol Chem 268*: 26813–26816.

Krumlauf R. 1994. *Hox* genes in vertebrate development. *Cell 78*:191–201.

Lawrence P, Morata G. 1994. Homeobox genes: Their function in *Drosophila* segmentation and pattern formation. *Cell 78*:181–189.

Otting G, Qian Y, Billeter M, Muller M, Affolter M, Gehring W, Wuthrich K. 1990. Protein-DNA contacts in the structure of a homeodomain-

DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J 9*:3085–3092.

Pan CQ, Feng J, Finkel SE, Landgraf R, Johnson R, Sigman DS. 1994a. Structure of the *Escherichia coli* Fis protein–DNA complex probed by protein conjugated with 1,10-phenanthroline copper(I) complex. *Proc Natl Acad Sci USA 91*:1721–1725.

Pan CQ, Landgraf R, Sigman DS. 1994b. DNA binding proteins as site-specific nucleases. *Mol Microbiol 12*:335–342.

Pendergrast PS, Ebright YW, Ebright RH. 1994. High-specificity DNA cleavage agent: Design and application to kilobase and megabase DNA substrates. *Science 1994*:959–962.

Pomerantz J, Sharp PA. 1994. Homeodomain determinants of major groove recognition. *Biochemistry 33*:10851–10858.

Poole S, Kauvar L, Drees B, Kornberg T. 1985. The engrailed locus of *Drosophila*: Structural analysis of an embryonic transcript. *Cell 40*:37–43.

Shang Z, Ebright Y, Iler N, Pendergrast S, Echelard Y, McMahon A, Ebright R, Abate C. 1994. DNA affinity cleaving analysis of homeodomain–DNA interaction: Identification of homeodomain consensus sites in genomic DNA. *Proc Natl Acad Sci USA 91*:118–122.

Siegfried E, Chou TB, Perrimon N. 1992. *Wingless* signaling acts through *zeste-white 3*, the *Drosophila* homolog of *glycogen synthase kinase-3*, to regulate *engrailed* and establish cell fate. *Cell 71*:1169–1179.

Sigman DS. 1990. Chemical nucleases. *Biochemistry 29*:9097–9105.

Stark M, Johnson A. 1994. Interaction between two homeodomain proteins is specified by a short C-terminal tail. *Nature 371*:429–432.

Sutton C, Mazumder A, Chen B, Sigman DS. 1993. Transforming the *Escherichia coli* Trp repressor into a site-specific nuclease. *Biochemistry 32*:4225–4230.

Weigel D, Meyerowitz E. 1994. The ABCs of floral homeotic genes. *Cell 78*:203–209.

Wolberger C, Vershon A, Liu B, Johnson A, Pabo C. 1991. Crystal structure of a MAT α2 homeodomain–operator complex suggests a general model for homeodomain–DNA interactions. *Cell 67*:517–528.