MOLECULAR AND CELLULAR BIOLOGY, Jan. 1995, p. 129–140 Vol. 15, No. 1
0270-7306/95/$04.00+0

DNA-Binding Specificity of the Cut Repeats from the Human Cut-Like Protein

RYOKO HARADA,[1,2] GINETTE BÉRUBÉ,[1,2] OWEN J. TAMPLIN,[1,3] CLAUDE DENIS-LAROSE,[1,2] AND ALAIN NEPVEU[1,2,3,4]*

Molecular Oncology Group[1] and Departments of Medicine,[2] Oncology,[4] and Biochemistry,[3] McGill University, Royal Victoria Hospital, Montreal, Quebec, Canada H3A 1A1

Received 27 April 1994/Returned for modification 16 June 1994/Accepted 23 September 1994

The *Drosophila* Cut and mammalian Cut-like proteins contain, in addition to the homeodomain, three other DNA-binding regions called Cut repeats. Cut-like proteins, therefore, belong to a distinct class of homeodomain proteins with multiple DNA-binding domains. In this study, we assessed the DNA-binding specificity of the human Cut repeats by performing PCR-mediated random oligonucleotide selection with glutathione S-transferase fusion proteins. Cut repeat 1, Cut repeat 3, and Cut repeat 3 plus the homeodomain selected related yet distinct sequences. Therefore, sequences selected by one of the fusion proteins were often, but not always, recognized by the other proteins. Consensus binding sites were derived for each fusion protein. In each case, however, some selected sequences diverged from the consensus but were confirmed to be high-affinity recognition sites by electrophoretic mobility shift assay. We conclude that Cut DNA-binding domains have broad, overlapping DNA-binding specificities. Determination of dissociation constants indicated that in addition to the core consensus, flanking sequences have a moderate but significant effect on sequence recognition. Evidence from electrophoretic mobility shift assay, DNase footprinting, and dissociation constant analyses strongly suggested that glutathione *S*-transferase/Cut fusion proteins bind to DNA as dimers. The implications of these findings are discussed in relation to the DNA-binding capabilities of Cut repeats. In contrast to other studies, we found that the human Cut-like protein does not preferably bind to a site that includes an ATTA homeodomain-binding motif. Here we demonstrate that the native human Cut-like protein recognizes more efficiently a site containing an ATCGAT core consensus flanked with G/C-rich sequences.

Regulation of gene expression by sequence-specific transcription factors requires that these factors first bind to specific sites in DNA. Therefore, much of the research on specific gene regulation has been focused on the prerequisites for specific DNA binding. The identification of specific DNA recognition sites has been facilitated by the recent development of the method of PCR-mediated random oligonucleotide selection (9). Using this procedure, high-affinity binding sites have been identified for transcription factors with no known target. For example, the c-Myc basic region–helix-loop-helix motif was found to specifically bind to a CA(CG)TG consensus site (8). The same procedure, when used with other transcription factors, has enabled the definition of a whole spectrum of specific DNA sequences that can be recognized. GATA transcription factors, in particular, were found to bind with high affinity to DNA sites that diverged from the established consensus binding site (29, 39).

Protein domains necessary and sufficient for specific DNA binding have been defined through mutational analysis as well as gene-swapping experiments. This led to the identification of different types of DNA-binding domains that are shared among several transcription factors and are conserved throughout evolution. The homeodomain was originally identified by sequence comparison of *Drosophila* gene products that control embryonic development (reviewed in reference 37). The homeodomain is a 61-amino-acid DNA-binding domain encoded by the homeobox (reviewed in references 44 and 52). Homeodomains were eventually found in proteins throughout eukaryotes including mammals, nematodes, yeasts, and plants. Many of these proteins were shown to play a role in development or cellular differentiation, indicating that basic developmental mechanisms have been conserved through evolution. For example, mammalian Hox genes showing sequence conservation with *Drosophila* homeotic selector genes have been shown, like their fly counterparts, to play a role in the formation of the anterior-posterior axis (reviewed in reference 37).

Other conserved DNA-binding domains have been found to be associated with a homeodomain. The first of these domains to be identified were the POU-specific and Paired domains, found in POU and Paired classes of transcription factors. The POU-specific domain is a 75- to 82-amino-acid domain which is separated from the homeodomain by a short variable linker region (47; reviewed in reference 43). These three protein segments, taken together, form what is called the POU domain, so named because it was originally found in the pituitary-specific Pit-1/GHF1, the Oct-1 and Oct-2 mammalian transcription factors, and the *Caenorhabditis elegans* cell lineage control gene *unc-86* (17, 25, 47). High-affinity DNA binding by the POU domain requires the participation of both the POU-specific domain and the homeodomain (5, 55). The Paired domain is a 128-amino-acid domain encoded by the paired box originally identified in the *Drosophila* segmentation genes *paired* and *gooseberry* (15, 51). The paired box was subsequently detected in mammalian Pax genes as well as in genes from other vertebrates (reviewed in reference 21). Molecular analysis of specific DNA binding by Paired proteins indicated that the Paired domain can function either autonomously or with the homeodomain (51). Sequence analysis of homeodomain proteins led to the suggestion that other protein motifs, in addition to the POU and Paired domains, may play a role in

* Corresponding author. Mailing address: Department of Oncology, McGill University, Royal Victoria Hospital, 687 Pine Ave. West, Montreal, Quebec, Canada H3A 1A1. Phone: (514) 842-1231, ext. 5832. Fax: (514) 843-1478.

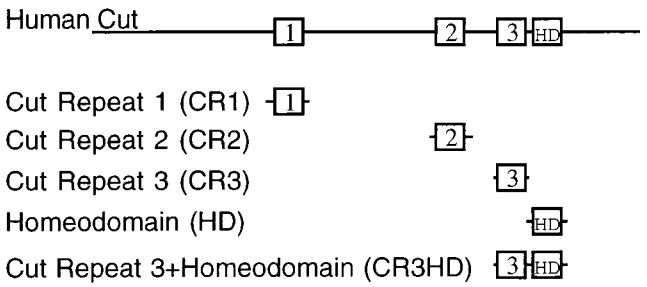

FIG. 1. Schematic representation of the fusion proteins used in this study. A representation of the human Cut protein is displayed at the top, with the Cut repeats and the homeodomain depicted as boxes. Below are shown the protein segments present in the GST and MBP fusion proteins.

DNA binding. These include POU-related regions in the liver-specific transcription factor HNF-1/LBF-1 and in the fungal mating-type protein β1-1 and the Cys-His-rich domain called LIM (7, 18, 19, 27, 38, 41, 53).

The *Drosophila* Cut protein and its mammalian counterparts, the human CCAAT displacement protein (CDP), the canine Clox (Cut-like Hox) protein, and the murine Cux protein, belong to a unique class of homeodomain proteins (3, 11, 40, 54). First, the Cut homeodomain is distinct in that it harbors a histidine at the ninth amino acid of the third helix. This amino acid has been shown in some proteins to determine the specificity of binding to the two bases following the TAAT core. Also, distinct classes of homeodomain proteins contain different amino acids at this position (52). Second, Cut-like proteins contain three conserved 73 amino acid motifs, called the Cut repeats, which share more than 50% amino acid identity with each other (3, 11, 40). Cut repeats have recently been shown to bind specifically to DNA (2, 23). Cut-like proteins therefore represent a novel class of homeodomain proteins with multiple DNA-binding domains.

The biochemical activities as well as the biological functions of Cut proteins in *Drosophila melanogaster* and in mammals remain to be defined. The pattern of expression and the phenotype of mutants in *D. melanogaster* suggest that this protein is involved in cell specification in several tissues, including the external sense organs, the Malpighian tubules, muscles, and the tracheal system (10, 13, 14, 26, 33). Thus, in various *Cut* mutants, these structures did not develop and, when tested, ectopic expression of Cut did not induce cell differentiation but changed the cell type specificity of differentiating cells (10, 12, 14, 26, 32). Evidence accumulated thus far indicates that the mammalian Cut-like proteins act as negative regulators of gene expression (3, 16, 40, 54). In transient transfection experiments, Clox, Cux, and human Cut/CDP have been shown to repress transcriptional activity of the beta major histocompatibility complex enhancer, the *Ncam* promoter, and the c-*myc* promoter, respectively (3, 16, 54). In addition, higher expression of several genes, including the cytochrome β heavy-chain (gp91-phox) gene, was shown to correlate with down-regulation of CDP binding activity upon cellular differentiation. This finding suggests that CDP represses the expression of these genes in undifferentiated cells (6, 45, 48, 49).

The recent identification of mammalian Cut-like proteins as potential transcriptional regulators of several genes revealed that sequences with little apparent homology can serve as recognition sites for these proteins (3, 16, 40, 54). These findings raised the question as to whether binding to dissimilar sequences was rendered possible by the presence of multiple DNA-binding domains in the Cut-like proteins. In this study, we have examined the sequence specificity as well as binding affinity of the human Cut DNA-binding domains. Results from site selection, gel retardation, and DNase footprinting analyses indicated that different Cut DNA-binding domains exhibit broad and overlapping DNA-binding preferences. Consensus binding sites were derived; however, some divergent sequences were bound with similar affinity. Also, sites selected for one Cut DNA-binding domain were frequently, but not always, recognized by other domains. The consensus binding site for the Cut repeat 3-plus-homeodomain fusion protein did not include an ATTA homeodomain-binding motif, and binding assays with Cut-like proteins from COS and HeLa cells confirmed that the presence of such a motif did not make a higher-affinity recognition site.

MATERIALS AND METHODS

Plasmid construction. Plasmids for expression of the glutathione *S*-transferase (GST)/Cut fusion proteins were prepared by inserting various fragments derived from cDNAs for the human Cut protein into the bacterial expression vector pGEX-3X (Pharmacia). The sequence of the human Cut protein has been published as the sequence for human CDP, and the cDNA sequence (HSCDP) can be obtained from GenBank (accession number M74099 [40]). The nucleotide and amino acid numbers used hereafter are taken from this cDNA sequence and its deduced amino acid sequence. For Cut repeat 1, an *Eco*RI (nucleotides [nt] 1605)-*Bam*HI (nt 2019) fragment was treated with Klenow enzyme and inserted into the *Sma*I site of pGEX-3X (the *Eco*RI site was added at nt 1605 during cDNA cloning). For Cut repeat 2, an *Rsa*I (nt 2861)-*Pst*I (nt 3153) fragment was inserted into the *Eco*RI site of pGEX-3X after treatment of the fragment and the vector with T4 DNA polymerase and Klenow enzyme, respectively. For Cut repeat 3, a *Cau*II (nt 3413)-*Rsa*I (nt 3737) fragment was inserted by blunt ligation into the *Eco*RI site of pGEX-3X after treatment of both the vector and the fragment with Klenow enzyme. For Cut repeat 3 plus homeodomain, a *Sau*96I (nt 3379)-*Sau*96I (nt 3982) fragment was treated with Klenow enzyme and inserted into the *Sma*I site of pGEX-3X. For Cut homeodomain, a *Bst*XI (nt 3625)-*Apo*I (nt 3963) fragment was treated with T4 DNA polymerase and inserted into the *Sma*I site of pGEX-3X. Plasmid for expression of the maltose-binding protein (MBP)/Cut repeat 3-plus-homeodomain (MBP/CR3HD) fusion protein was prepared by inserting the *Sau*96I (nt 3379)-*Sau*96I (nt 3982) fragment from the human *Cut* cDNA into the *Eco*RI site of the bacterial expression vector pMal-C2 (New England Biolabs); both vector and insert were treated with Klenow enzyme before ligation. For transient transfection in COS cells, a Cut-expressing vector, pSG5-Cut, was prepared by inserting a fragment from the human *Cut* cDNA into the *Eco*RI site of the pSG5 vector (Stratagene). The *Cut* cDNA fragment was from nt 1605 to 5376 in the HSCDP sequence (40). This includes the coding sequence for the three Cut repeats and the Cut homeodomain. *Eco*RI sites were added during the cDNA cloning procedure.

Expression and purification of the fusion proteins. Plasmid vectors expressing GST/Cut fusion proteins were introduced in *Escherichia coli* DH5. Induction of expression and purification of GST and MBP fusion proteins were done as previously described (22, 34, 46). Glutathione-Sepharose was purchased from Pharmacia (catalog no. 17-1756-01); amylose resin was purchased from New England Biolabs (catalog no. 800-21S).

Denaturation and renaturation of GST fusion proteins. Our protocol was derived from previous studies by Aceto et al. (1) and Sacchetta et al. (43a). Bacterial extracts containing GST/Cut fusion proteins of different molecular weights were either mixed together or treated separately. Bacterial extracts were mixed with 2 volumes of a solution of 8 M urea (final concentration, 5.3 M) and incubated at 4°C for 30 min with gentle agitation. Samples were then dialyzed against a solution of 0.1 M KH_2PO_4–1 mM dithiothreitol–1 mM urea at 4°C for 2 h before purification by affinity chromatography over a glutathione-Sepharose column. Integrity of GST fusion proteins at all steps was verified by polyacrylamide gel electrophoresis and Coomassie blue staining.

Electrophoretic mobility shift assays (EMSA). Reactions were performed with 50 ng of purified GST-Cut fusion proteins or 15 μg of total protein from mininuclear extract in a final volume of 20 μl in 10 mM Tris (pH 7.5)–25 mM NaCl–1 mM $MgCl_2$–5 mM EDTA (pH 8.0)–1 mM dithiothreitol–5% glycerol. Five micrograms of bovine serum albumin was included in the reaction mixture. Proteins from mininuclear extracts were preincubated with 2 μg of poly(dI-dC) for 5 min at room temperature. When specified, bacterially expressed proteins were preincubated with 50 ng of poly(dI-dC) for 5 min at room temperature. End-labeled double-stranded oligonucleotides (20,000 cpm, ~10 pg) were added, and samples were further incubated for 15 min. Samples were loaded on a 5% polyacrylamide gel (30:1 acrylamide/bisacrylamide), and complexes were separated at 8 V/cm for 2 h in 0.5× Tris-borate-EDTA. Gels were dried and visualized by autoradiography.

PCR-mediated random site selection. Binding site selections were performed essentially as described previously (9). Fifty nanograms of purified fusion pro-

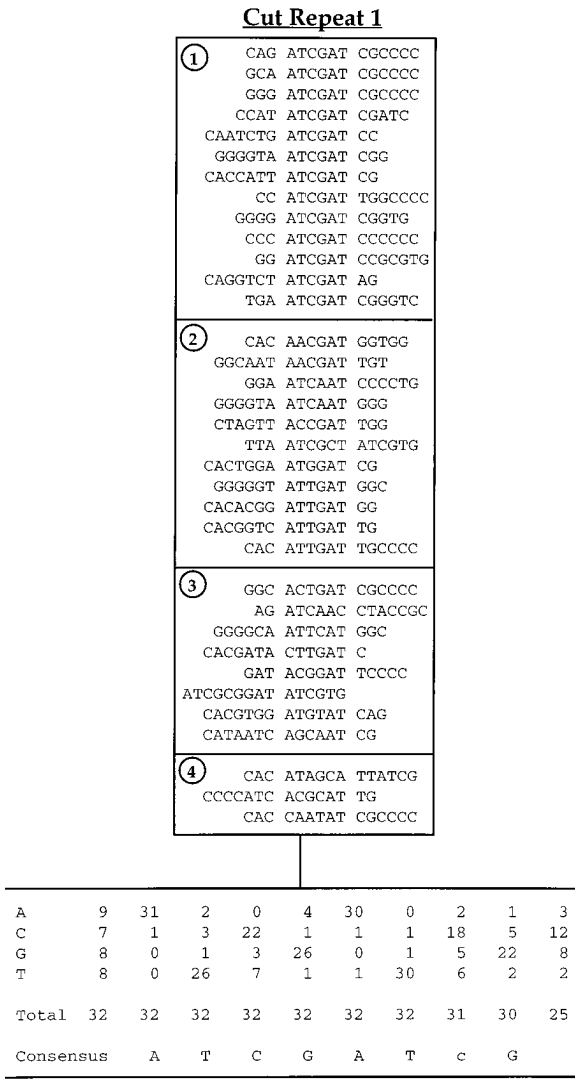

FIG. 2. Compilation of sequences selected by Cut repeat 1. Sequences were selected by using purified GST/Cut repeat 1 proteins and PCR-generated random oligonucleotides. After five rounds of selection, selected oligonucleotides were cloned and analyzed by DNA sequencing. The DNA sequences were organized to fit the best possible alignment. On the basis of the frequency of each nucleotide at each position, a core consensus sequence was deduced. Sequences were classified in different groups according to their similarity to the derived consensus binding site. Group 1 includes sequences which share the consensus binding site, whereas sequences in groups 2 and 3 deviate from the consensus site by one and two nucleotides, respectively, and group 4 comprises more divergent sequences. In the box at the bottom, uppercase letters indicate bases in highly constrained sequences, whereas lowercase letters indicate bases that are moderately constrained. Sequences in group 4 were not included in the numerical description of the consensus.

teins was incubated for 15 min at room temperature with ~10 ng of labeled, double-stranded PCR-generated oligonucleotides in binding buffer containing 50 ng of poly(dI-dC) (100 ng for the homeodomain; see Results). Oligonucleotides were random at 15 positions and flanked on either side by 15 nt which included a *Pst*I and an *Xba*I restriction site, respectively. The sequence of the oligonucleotide used was 5′-AGACCTGCAGTCTGCN$_{15}$CTGTCGTCTAGAGGA-3′. Protein-DNA complexes were separated from the free oligonucleotides by electrophoresis on a 5% polyacrylamide gel (30:1 acrylamide/bisacrylamide). Following exposure of the wet gels, protein-DNA complexes were visualized, excised, and eluted overnight at 37°C in 300 μl of 10 mM Tris HCl (pH 8)–1 mM EDTA. After ethanol precipitation, one half of the DNA was amplified by PCR for 25 cycles. PCR cycles were as follows: 1 min at 94°C, 1 min at 50°C, and 1 min at 72°C. Part of the PCR product (200 ng) was radiolabeled with T4 polynucleotide kinase and purified on 8% nondenaturing polyacrylamide gel. Five to 10 ng of this purified labeled DNA was used for the next round of selection. After the fifth cycle, the PCR products were digested with *Pst*I and *Xba*I and cloned into the plasmid Bluescript KS (Stratagene). Sequencing of the inserts were performed with the T7 polymerase sequencing kit (U.S. Biochemical).

Determination of dissociation constants. The dissociation constants (K_D) value was determined by EMSA. DNA concentrations of the oligonucleotides were determined fluorimetrically. A range of end-labeled double-stranded oligonucleotides were incubated for 20 min with a fixed amount of each protein in the absence of a nonspecific competitor. After separation of the protein-DNA complexes from free DNA by electrophoresis on a gel, the radioactivity in the bound and free DNA was determined in a liquid scintillation counter (LKB). [Bound DNA]/[free DNA] was plotted against [bound DNA].

DNase footprinting assay. The Bluescript SK vector (Stratagene), which contains a *Cla*I site (ATCGAT) as well as a CCAAT box, was chosen for this analysis. The plasmid was ^{32}P end labeled at the *Not*I site with T4 polynucleotide kinase and cleaved with *Pvu*II and *Sst*I. After electrophoresis through a 5% polyacrylamide gel, the 209-bp *Not*I-*Pvu*II labeled fragment was purified by passive elution in 10 mM Tris HCl (pH 7.5)–1 mM EDTA. The ATCGAT and CCAAT sequences start at positions +56 and +91, respectively, relative to the labeled nucleotide. DNase footprinting was carried out essentially as described elsewhere (16a, 20). End-labeled DNA (100,000 cpm per reaction) was incubated with variable quantities of proteins, in the presence of 50 ng of poly(dI-dC), for 15 min at room temperature in a final volume of 25 μl in 10 mM Tris (pH 7.5)–25 mM NaCl–1 mM $MgCl_2$–1 mM dithiothreitol–5% glycerol–4% (wt/vol) polyvinyl alcohol. For GST/Cut repeat 3 (GST/CR3) and GST/homeodomain (GST/HD), NaCl was replaced with potassium glutamate. Then 50 μl of 10 mM $MgCl_2$–5 mM $CaCl_2$ was added, tubes were incubated for 90 s, 2 μl of DNase at 2.5 μg/ml was added, tubes were incubated for 90 s, 90 μl of DNase stop solution (20 mM EDTA, 1% sodium dodecyl sulfate, 0.2 M NaCl) was added, and tubes were mixed by vortexing. Following phenol-chloroform extraction and ethanol precipitation, samples were electrophoresed through an 8% denaturing polyacrylamide gel (30:1 acrylamide/bisacrylamide) in 1× Tris-borate-EDTA. Gels were visualized by autoradiography without prior drying.

Transient transfection and mininuclear extract. COS cells were grown in Dulbecco modified Eagle medium supplemented with 10% fetal bovine serum. The expression vector pSG5Cut was introduced by the DEAE-dextran technique (36). Ten micrograms of plasmid DNA was added to 2×10^6 cells in 150-mm-diameter plates. Cells were treated with chloroquine on the next day and harvested 48 h later. Mininuclear extracts were prepared from transfected and untransfected cells by a published procedure (31).

RESULTS

Sequences selected for GST/Cut repeat fusion proteins include TATNG, CCAAT, and ATNNAT motifs. Cut repeats were originally identified in the *Drosophila* Cut homeodomain protein as three repeats of 73 amino acids and later found in three mammalian homologs, human CDP, canine Clox, and murine Cux (3, 40, 54). From sequence comparison, Cut repeats are well conserved between *D. melanogaster* and mammals, suggesting that they carry an important biological function. Indeed, we and others have previously demonstrated that Cut repeats can function as DNA-binding domains (2, 23). In the present study, we wished to determine the DNA-binding specificity of the Cut repeats from the human CDP/Cut-like protein. For this purpose, we used the procedure of PCR-mediated random oligonucleotide site selection. Various protein segments were expressed in bacteria as GST fusion proteins and purified by affinity chromatography over a glutathione-Sepharose column. The fusion proteins used in this study are diagrammatically represented in Fig. 1. Radiolabeled, double-stranded oligonucleotides were used in EMSA with each of the fusion proteins. Oligonucleotides were random at 15 positions and flanked on either side by 15 nt which allowed both annealing of primers and cleavage by restriction enzymes for cloning. Cut repeat-binding sites were selected by isolating the lower-mobility protein-DNA complex separated by polyacrylamide gel electrophoresis, followed by PCR amplification of the isolated DNA for subsequent EMSA. The binding-site selection was carried out at least five times (eight times for Cut repeat 3); the selected oligonucleotides were then cloned and sequenced. A compilation of sequences selected by Cut repeat 1 is presented in Fig. 2. A consensus binding site, 5′-ATCGAT-3′, was derived, and a significant number of CCAAT (ATTGG,

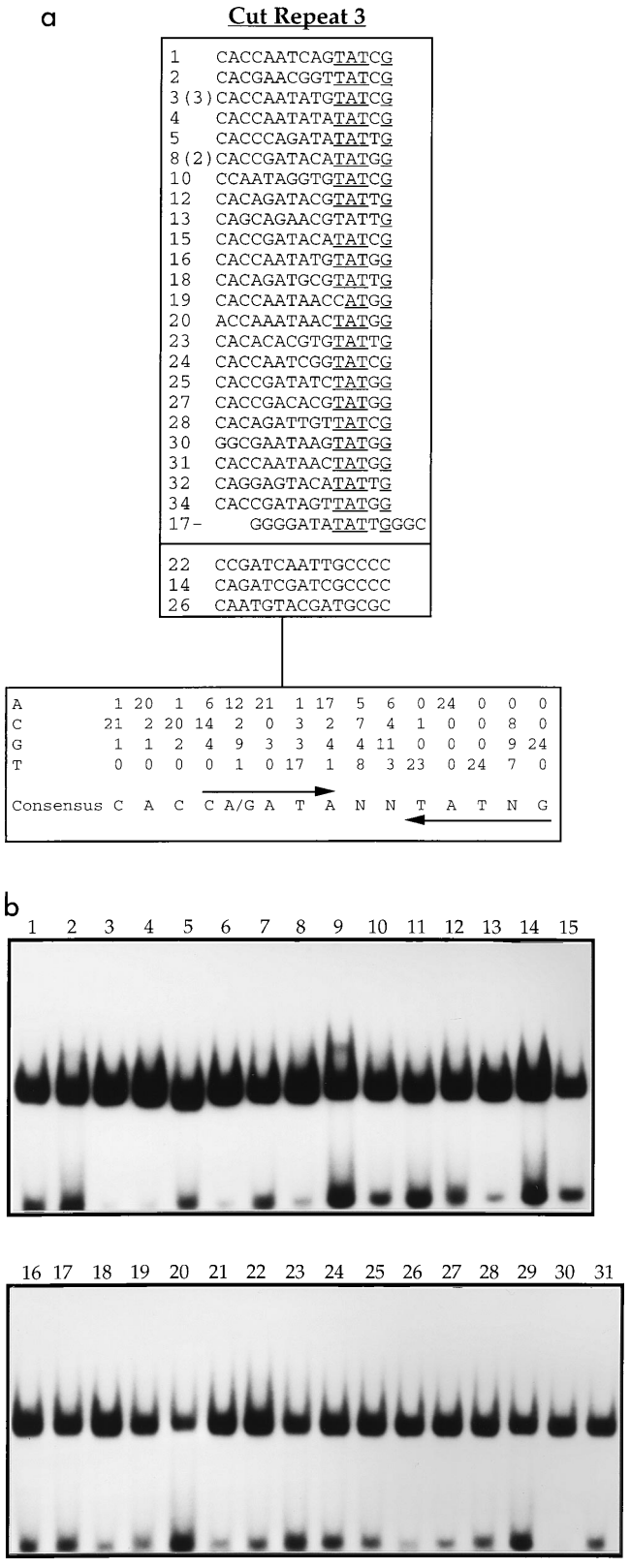

FIG. 3. Compilation of sequences selected by Cut repeat 3. (a) Sequences were selected by using purified GST/CR3 fusion proteins and PCR-generated random oligonucleotides. After eight rounds of selection, selected oligonucleotides were cloned and analyzed by DNA sequencing. The numbers in parentheses indicate the number of times a particular sequence was found. On the basis of the frequency of each nucleotide at each position, a core consensus sequence was deduced. Sequences 22, 14, and 26 could not be aligned with the others and were not included in the calculation of the consensus. The position of a palindromic sequence is shown by arrows. (b) End-labeled double-stranded oligonucleotides generated by PCR were incubated with 50 ng of purified GST/CR3 fusion proteins in the presence of 50 ng of poly(dI-dC). The specific sequence of each oligonucleotide is shown in the list of selected sites.

in the reverse orientation) and closely related sequences were also noted. A significant number of clones diverged from the consensus but were still found to be good binding sites (see Fig. 5; also data not shown). Sequences were therefore classified in different groups according to their similarity with the derived consensus binding site. Group 1 includes sequences which share the consensus binding site, whereas sequences in groups 2 and 3 deviate from the consensus site by one and two nucleotide(s), respectively, and group 4 comprises more divergent sequences.

Oligonucleotides selected by Cut repeat 3 after five cycles included a significant number of ATNNAT, TATNG, and CC AAT or CCGAT sequences. It was not possible, however, to align a majority of sites according to either one of these motifs. Selection was therefore continued for three more cycles. Selected sites after eight cycles are presented in Fig. 3. All but three sequences could be aligned according to TATNG. The derived consensus binding site, CACCNATANNTATNG, comprised a palindrome, and curiously, almost all selected sequences, when aligned, were in the same orientation. This finding suggested that the flanking, invariant sequences were possibly part of the binding site. Interestingly, when the flanking sequences were juxtaposed to the central core, the palindrome that is part of the consensus binding site could be extended to the right: CACCNATANNTATNGCTG. To ensure that selection had been brought to completion and that divergent sequences were true binding sites, we verified that all 31 cloned sites were bound with similar affinities (Fig. 3b).

In the case of Cut repeat 2, no sequences were selected or amplified, even though the procedure was repeated several times with various fusion proteins containing Cut repeat 2 either alone or with variable lengths of adjacent protein segments. These results are in contrast with those of two other groups which showed strong DNA binding by Cut repeat 2 (2, 4). Curiously, one of these groups did not observe DNA binding by Cut repeats 1 and 3 (2). The reason for these differences is currently not known, but it does not appear to involve sequence divergence between Cut repeats isolated from different mammalian species (3, 16, 40).

In contrast with the Cut repeats, when we performed the same procedure with the Cut homeodomain, a large fraction of the PCR-mediated random sequences were already present in the retarded protein-DNA complex after the first selection cycle, indicating that the Cut homeodomain can bind with a relatively high affinity to nonspecific sequences. Sequences obtained after seven selection cycles could not be aligned to a single consensus. Similar results were obtained when the procedure was repeated in the presence of 100 ng of various nonspecific competitor DNA [poly(dIdC), lambda DNA, or salmon sperm DNA]. We conclude that the Cut homeodomain, when acting alone, does not exhibit a strong preference for specific DNA sequences.

The flexible consensus obtained with Cut repeat 3 suggested that this domain had a broad DNA-binding specificity. By analogy with the POU domain, we considered that Cut repeat 3 may bind DNA with higher specificity and better affinity when acting in conjunction with the Cut homeodomain which is located next to it in the protein (2, 55). We therefore re-

Cut Repeat 3+HD

1				AC	ATCGAT	CCACCTG				
			CAGG	ATCGAT	CCCCC					
			CACG	ATCGAT	CCCTC					
			CAC	ATCGAT	CCCCCC					
			G	ATCGAT	CTGGGGGC					
			TG	ATCGAT	CGCCGCC					
		GGGCTA	ATCGAT	GTG						
		CCC	ATCGAT	TCGGTG						
		GGGCA	ATCGAT	GGG						
	CAGGGGT	ATCGAT	GC							
		CAC	ATTGAT	CCCCCC						
		CACA	ATTGAT	CCCCC						
	CACGCCA	ATTGAT	CA							
		GCC	ATTGAT	CCCCCC						
		CAC	ATTGAT	GGCCCC						
		GGAGGC	ATTGAT	GTG						
		CACGGG	ATTGAT	GGT						
		CACACA	ATTGAT	TGG						
		CACAT	ATTGAT	CTGG						
2			CAC	AATGAT	CCCCCC					
			GGG	ATGGAT	GTGGGG					
			CACA	ATGGAT	CCCCC					
			CAC	ATAGAT	CCGGTA					
		ACAGG	GTCGAT	GGTG						
		GGG	GTCGAT	CGGGTG						
	GGGGGCGG	GTCGAT	C							
		CAC	AATGAT	TGCCCC						
		GCC	AATGAT	TGCCCC						
		GGGGCG	TTCGAT	GTG						
		TGCG	ATAGAT	TGGTG						
3			GCCA	ATGCAT	CCCCC					
		CACCGA	ATACAT	GGG						
		CAG	GTCCAT	CGCCCC						
4			CACC	ATGCAC	CCTCC					

A	8	28	3	3	0	33	0	0	1	1
C	11	0	0	15	3	0	0	18	13	14
G	12	4	0	3	30	0	0	10	12	13
T	2	1	30	12	0	0	33	5	6	2
Total	33	33	33	33	33	33	33	33	32	30
Consensus	G/C	A	T	C/T	G	A	T	C/g	C/G	C/G

FIG. 4. Compilation of sequences selected by the Cut repeat 3-plus-homeodomain protein. DNA was selected by using purified GST/CR3HD fusion proteins and PCR-generated random oligonucleotides. After five rounds of selection, selected oligonucleotides were cloned and analyzed by DNA sequencing. The DNA sequences were organized to fit the best possible alignment. On the basis of the frequency of each nucleotide at each position, a core consensus sequence was deduced. Sequences were classified in different groups according to their similarity to the derived consensus binding site. Group 1 includes sequences which share the consensus binding site, whereas sequences in groups 2 and 3 deviate from the consensus site by one and two nucleotides, respectively, and group 4 comprises more divergent sequences. In the box at the bottom, uppercase letters indicate bases in the highly constrained sequence, whereas lowercase letters indicate bases that are moderately constrained. Sequences in group 4 were not included in the numerical description of the consensus.

peated the site selection with a fusion protein containing Cut repeat 3 and the Cut homeodomain. The recombinant protein is described in Fig. 1, and the compilation of selected sequences is presented in Fig. 4. The consensus for Cut repeat 3 plus homeodomain was 5′-AT(C/T)GAT-3′, and we also noted some preference for C or G at flanking positions. Interestingly, this consensus is closely related to that of Cut repeat 1. These results suggest that the presence of the Cut homeodomain may modify the binding specificity of Cut repeat 3 or, alternatively, that the homeodomain may be involved in cooperative DNA binding with Cut repeat 3. In conclusion, consensus binding sites were derived for Cut repeat 1, Cut repeat 3, and Cut repeat 3 plus homeodomain. However, in each case, a significant number of selected sequences diverged from the consensus.

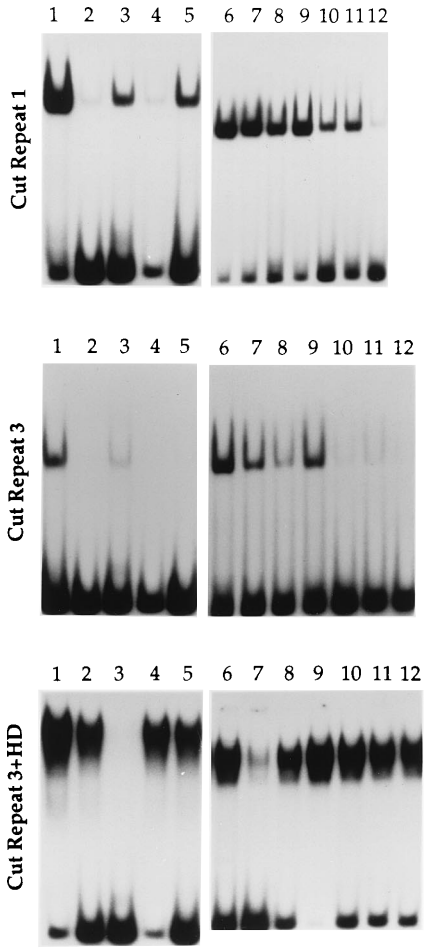

FIG. 5. EMSA using GST/Cut fusion proteins and selected binding sites. End-labeled double-stranded oligonucleotides generated by PCR were incubated with 50 ng of purified GST/Cut fusion proteins in the presence of 50 ng of poly(dI-dC). The specific sequence of each oligonucleotide, as well as the fusion protein used for its selection, is given in Table 1.

Different Cut repeats show related but distinct sequence specificity. To confirm that the selected sequences are true protein-binding sites, we performed EMSA with probes selected for the various Cut fusion proteins. This assay was also expected to establish whether different fusion proteins would exhibit distinct DNA-binding specificities. In total, 12 probes were tested. The binding of Cut fusion proteins to these various sites is presented in Fig. 5. The sequences of the probes are listed in Table 1, together with the DNA-binding domain used for their selection and the relative binding affinities of each protein. All probes were well recognized by the fusion protein for which they had been selected. This was true not only for probes from group 1, which contain a perfect consensus binding site, but also for probes from groups 2, 3, and 4, whose sequences diverged from the consensus (Fig. 5, probes 3, 4, 5, 7, 8, 10, 11, and 12). We conclude from this observation that selected binding sites which diverged from the consensus binding sites are recognized with similar affinities. This finding suggests that Cut repeats can tolerate a certain variability in their recognition sites. It is also possible, as suggested by subsequent experiments (see below), that sequences flanking the core consensus play a role in determining binding affinity.

TABLE 1. DNA-binding affinity of GST/Cut fusion proteins for selected binding sites

	Selection[a]	Sequence of binding site[b]	Binding affinity		
			CR1	CR3	CR3+HD
1	CR3+HD (1)	CACACA ATTGAT TGG	+++	+	+++
2	CR3+HD (1)	CAC ATTGAT CCCCCC	+/−	−	+++
3	CR1 (3)	GGGGCA ATTCAT GGC	+	+/−	−
4	CR3+HD (2)	CACAT ATTGAC CTGG	+/−	−	+++
5	CR3+HD (2)	CAC AATGAT TGCCCC	+	−	+++
6	CR1 (1)	TGA ATCGAT CGGGTC	+++	++	++
7	CR1 (2)	TTA ATCGCT ATCGTG	+++	+	+
8	CR1 (2)	CAC AACGAT GGTGG	++	+/−	++
9	CR1 (1)	CAG ATCGAT CGCCCC	+++	+	+++
10	CR3+HD (2)	GGG ATGGAT GTGGGG	+	−	+++
11	CR3+HD (3)	GCCA ATGCAT CCCCC	+	−	+++
12	CR3+HD (4)	CACC ATGCAC CCTCC	+/−	−	+++

[a] The fusion protein used in the selection is indicated together with the group (in parentheses) to which the sequence belongs on the basis of its resemblance to the deduced consensus binding site. Proteins are designated as in Fig. 1.
[b] Some sequences are in the reverse orientation to facilitate the comparison.

A second observation is that sequences selected by one fusion protein were most often recognized by one or more other fusion proteins. In particular, three of the probes were bound very well by all fusion proteins (Fig. 5, probes 1, 6, and 9). This is in accordance with the fact that selected sites for the various fusion proteins contained similar sequence motifs. Also, the fusion protein containing the Cut homeodomain in addition to Cut repeat 3 exhibited binding preferences that are clearly distinct from that of the Cut repeat fusion proteins. For example, probes 2, 4, and 12 were recognized by Cut repeat 3 plus homeodomain but not by Cut repeat proteins, while the reverse was true for probes 3 and 7. Altogether, these results demonstrate that different Cut repeats or the combination of Cut repeat 3 and the Cut homeodomain recognize closely related sequences but still have distinct DNA-binding specificities.

Sequences flanking the core consensus binding site influence binding affinity. A close examination of the results in Fig. 5 also revealed that sequences flanking the core consensus site can have a significant effect on binding by Cut repeats 1 and 3. This is exemplified by the results obtained with probes 1 and 2 (Fig. 5). To further study the sequence requirements for DNA binding, we analyzed the relative binding affinity of Cut repeats to synthesized double-stranded oligonucleotides. The sequences of these binding sites are displayed in Table 2. Probes A and B contain the ATCGAT core which is commonly found within selected sites. These two probes differ on the 3′ side of the core: probe A contains the sequence CGCCCC, which was frequently selected, whereas probe B contains an unrelated sequence. The relative binding affinities of each fusion protein was first assessed by EMSA (Fig. 6). Probe A represented a better binding site than probe B, suggesting that sequences flanking the ATCGAT consensus are also part of the recognition site. To assess with more accuracy the contribution of flanking sequences, we determined the apparent dissociation constant of each fusion protein to probes A and B. In this experiment, DNA at various concentrations was incubated with a fixed amount of purified proteins. Following a 20-min incubation, the free and bound DNA were separated by polyacrylamide gel electrophoresis. After quantification, the concentrations of bound over free DNAs were plotted against that of free DNA. The Scatchard plots are shown in Fig. 7, and the derived dissociation constants are presented in Table 2. It is important to emphasize that calculations of dissociation constants did not take into account the monomeric versus dimeric nature of the fusion proteins (see below); therefore, the values presented here are valid only for the purpose of comparing the affinities of different GST fusion proteins. For this reason, these values are referred to as apparent K_Ds. The effect of 3′ flanking sequences on the binding affinity appears in each case to be in the range of about 1.5- to 2-fold. The comparison of apparent K_Ds obtained with different GST/Cut fusion proteins indicates that Cut repeat 1 exhibited a more than 10-fold-higher affinity than Cut repeat 3 for probes A and B. However, when the Cut homeodomain was present with Cut repeat 3, DNA-binding affinity was increased by more than 2 orders of magnitude. These results suggest that Cut repeat 3 and the Cut homeodomain cooperate to bind to DNA with higher affinity.

DNase footprinting analysis of DNA binding by MPB/Cut and GST/Cut fusion proteins. To analyze further the DNA-binding specificity of Cut DNA-binding domains, we performed a DNase footprint analysis using a DNA fragment containing an ATCGAT consensus as well as a CCAAT motif. Several observations can be made from this analysis. Using 50 ng of proteins, we observed DNA protection with GST/CR3HD, GST/CR1, and MBP/CR3HD, but not with GST/CR3 or GST/HD (Fig. 8A). A larger amount of the last two proteins was needed to obtain a footprint (Fig. 8B, lanes 6, 7,

TABLE 2. Dissociation constants of Cut fusion proteins for consensus binding sites

Probe[a]	Sequence of consensus binding site	K_D (M)			
		GST/CR1	GST/CR3	GST/CR3HD	MBP/CR3HD
A	AG ATCGAT CGCCCC	3.4×10^{-10}	7.9×10^{-9}	3.1×10^{-11}	4.6×10^{-10}
B	AG ATCGAT CAGACT	5.9×10^{-10}	1.2×10^{-8}	7.5×10^{-11}	ND

[a] Probes A and B contain the ATCGAT core, which represents a common consensus binding site for all Cut fusion proteins, but differ at positions 3′ to the core consensus.

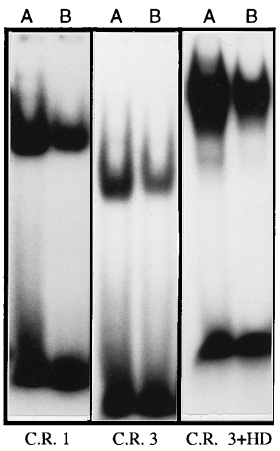

FIG. 6. EMSA using GST/Cut fusion proteins and consensus binding sites with different flanking sequences. End-labeled double-stranded oligonucleotides were incubated with 50 ng of purified GST/Cut fusion proteins in the presence of 50 ng of poly(dI-dC). Probes A and B contain the ATCGAT core, which represents a common consensus binding site for all Cut fusion proteins but differ at positions 3′ to the core consensus. The sequences of oligonucleotides A and B are given in Table 2.

9, and 10), and surprisingly, a similar and relatively large region of DNA was protected, suggesting that Cut repeat 3 and the Cut homeodomain may have overlapping DNA-binding preferences. When the two domains were present on the same protein, as in GST/CR3HD and MBP/CR3HD, not only was a lesser amount of protein needed, but the footprints were smaller (compare lane 4 with lanes 6, 7, 9, and 10 in Fig. 8B). This result emphasizes the gain both in DNA-binding affinity and specificity when Cut repeat 3 and the Cut homeodomain are present on the same protein.

The region containing the ATCGAT consensus was protected by all fusion proteins. This confirms that this sequence is recognized by every Cut DNA-binding domain. On the other hand, although CCAAT motifs and closely related sequences were found in a significant fraction of the sites selected by each fusion protein, the region containing a CCAAT motif was protected only by GST/CR1. This may indicate that the affinity of Cut repeat 1 for CCAAT sites is higher than that of other Cut DNA-binding domains. We cannot, however, exclude the possibility that the surrounding sequence environment is important in determining whether a CCAAT site will be recognized by a particular Cut repeat. We note, for example, that selected sites for Cut repeat 3 contain, in addition to the CCAAT sequence, a TAT motif that is not present here.

Finally, the sizes of the footprints are quite revealing. Interestingly, the footprints generated by GST/CR3HD and GST/CR1 around the ATCGAT and CCAAT motifs were of 32 bp, whereas the MBP/CR3HD protein produced a 16-bp footprint which corresponded to the top half of the GST/CR3HD footprint (Fig. 8A, lanes 2, 4, and 12; Fig. 8B, lanes 2 and 4). This result could be interpreted to mean that GST/Cut fusion proteins bind to DNA as dimers, a notion that is in accordance with the dimeric nature of GSTs. This hypothesis was further tested in the experiments described below.

GST/Cut repeat fusion proteins bind to DNA as dimers. One way to assess whether DNA binding involves a dimeric protein is to perform EMSA with a mixture of proteins of different sizes. Heterodimeric protein-DNA complexes should migrate with an intermediate mobility in comparison with the two homodimeric protein-DNA complexes. The formation of heterodimers involves prior dissociation of the dimerization partners followed by reassociation of monomeric subunits to generate homo- and heterodimers. Previous studies have demonstrated that GST enzymes exist as stable dimers that can be reversibly dissociated by incubation in the presence of a low concentration of a chemical denaturant (1, 43a). A similar approach was taken here. Bacterial extracts containing GST/CR3 or GST/CR3HD fusion proteins of different sizes were mixed, incubated in the presence of 5.3 M urea for 30 min, and then dialyzed extensively before purification by affinity chromatography on glutathione-Sepharose beads. As controls, GST/Cut fusion proteins were prepared in the same manner but without prior mixing of the bacterial extracts. Protein-DNA complexes of intermediate mobility were reproducibly observed with GST/Cut fusion proteins purified from mixed extracts (Fig. 9A, lanes 2 and 5; Fig. 9B, lanes 5 and 6), whereas each of the controls generated a unique protein-DNA complexes (Fig. 9A, lanes 1, 3, 4, and 6; Fig. 9B, lanes 1 to 4). This result strongly suggests that GST/Cut fusion proteins exist as dimers. Furthermore, consistent with the stable nature of GST dimers, when EMSA were performed with GST/Cut fusion proteins of different size but without prior treatment with a denaturant, protein-DNA complexes of intermediate mobility were not observed (data not shown).

MPB/CR3HD and GST/CR3HD fusion proteins exhibit different dissociation constants. Determination of the apparent K_Ds should be affected by the monomeric or dimeric nature of the binding protein. Dimerization would have the effect of lowering the apparent K_D. We therefore calculated the apparent K_D of MBP/CR3HD and compared it with that of the corresponding GST/CR3HD fusion protein (Fig. 7). The value obtained for MBP/CR3HD, 4.6×10^{-10} M, indicated a lower binding affinity than that of the corresponding GST/CR3HD fusion protein, 3.1×10^{-11} M. The difference in apparent K_Ds is in agreement with the notion that MBP/CR3HD bound to DNA as a monomer and GST/CR3HD bound as dimers.

Cut-like proteins expressed in mammalian cells bind with high affinity to the ATCGAT consensus binding site. The results presented above showed that the sequence ATCGAT represents a good binding site for several of the Cut DNA-binding domains. To confirm that a human Cut protein which contains multiple DNA-binding domains can bind to this sequence, COS cells were transfected with the vector pSG5Cut expressing a fragment of the human *Cut* cDNA encoding the three Cut repeats and the Cut homeodomain. Nuclear extracts prepared from untransfected and transfected COS cells were used in EMSA with double-stranded oligonucleotide A as a probe (Fig. 10, lanes 6 to 10). A protein-DNA complex of slower mobility and of increasing intensity is visible in lanes 8 to 10, which contain, respectively, 1, 2, and 4 μl of nuclear extract from transfected COS cells. We conclude that the proteins expressed from the pSG5Cut vector can bind with high affinity to the Cut consensus binding site. A common protein-DNA complex was also present in all samples, from either transfected or untransfected cells. This finding suggested that COS cells contain an endogenous binding activity for oligonucleotide A. To verify whether this binding activity involved an endogenous Cut protein, EMSAs were performed in the absence or presence of monoclonal antibodies against the human Cut protein (Fig. 10, lanes 11 to 16). For this experiment, nuclear extracts were prepared from COS and HeLa cells. In each case, the unique protein-DNA complex disappeared when binding was performed in the presence of either of the two anti-Cut monoclonal antibodies. These results demonstrate that mammalian Cut proteins can bind with high affinity to the ATCGAT consensus binding site.

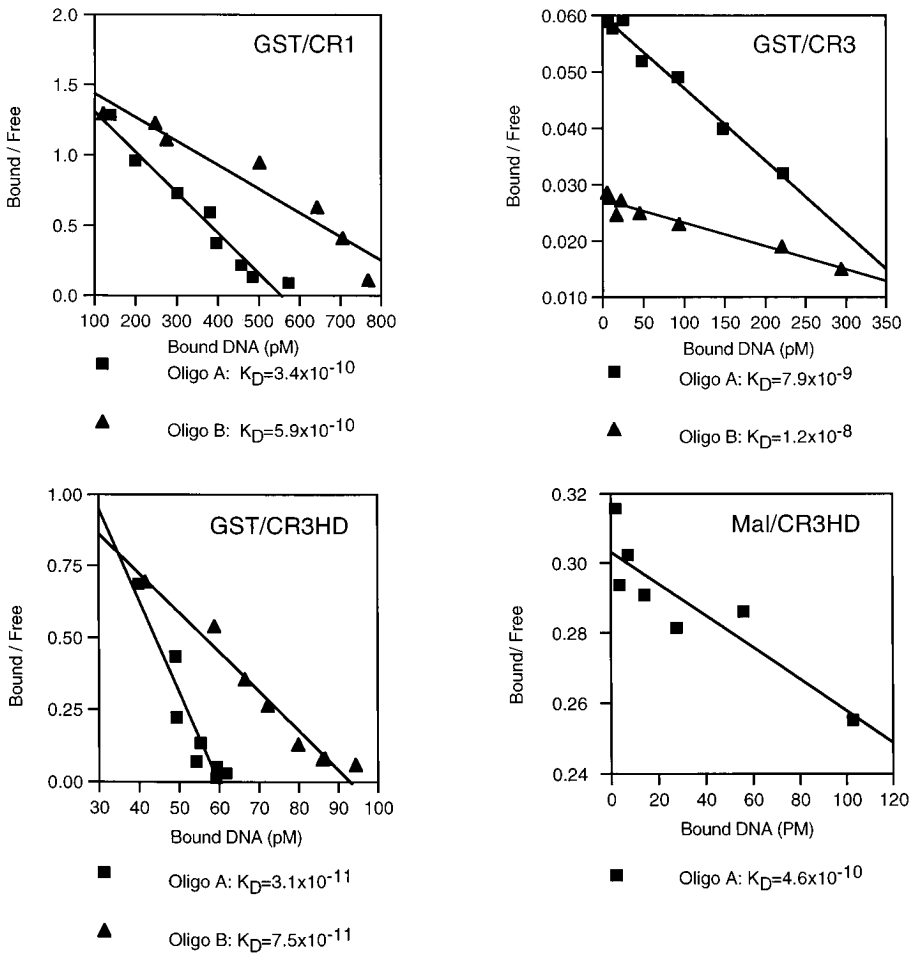

FIG. 7. Apparent K_Ds of Cut fusion proteins for consensus binding sites. Using end-labeled double-stranded oligonucleotides and purified GST/Cut and MBP/Cut fusion proteins, a range of DNA concentration was incubated with a fixed amount (50 ng) of each protein in the absence of nonspecific competitor DNA. After separation of the protein-DNA complexes from free DNA on a gel, the radioactivity in the fractions of bound and free DNA was calculated in a liquid scintillation counter (LKB).

The mammalian Cut-like proteins do not bind with higher affinity to a sequence that contains the ATTA homeodomain motif. As this work was being completed, a study from another group reported a Cut consensus binding with a similar ATC GAT core but with different flanking sequences which included a bona fide ATTA homeodomain motif (2). We decided to compare the affinities of mammalian Cut proteins for the two different Cut consensus binding sites. Oligonucleotides encoding the two sequences were labeled to the same specific activity, and an equal amount of each probe was used in EMSA with nuclear extracts from untransfected and transfected COS cells (Fig. 10, lanes 1 to 10). The results clearly indicate that mammalian Cut proteins do not have a higher affinity for a sequence that contains an ATTA motif.

DISCUSSION

The *Drosophila* Cut protein and its mammalian counterparts, the human CDP/Cut, canine Clox, and murine Cux proteins, belong to a unique class of homeodomain proteins (3, 11, 40, 54). These proteins have similar structural organizations, with three Cut repeats followed by a Cut-type homeodomain. Cut repeats are three conserved 73-amino-acid motifs originally shown to share from 52 to 63% amino acid identity with each other in *Drosophila* Cut and later found to be highly conserved in mammals (3, 11, 40, 54). The high degree of conservation of Cut repeats suggests that they may have an important biological function. Indeed, we and others have previously shown that Cut repeats can function as DNA-binding domains (2, 23). In this study, we have addressed the DNA-binding specificity of the human Cut DNA-binding domains, using the procedure of PCR-mediated random site selection with a bacterially expressed GST/CR1, GST/CR3, or GST/CR3HD fusion protein. Our results show that different Cut DNA-binding domains exhibit overlapping but distinct sequence specificities. Consensus binding sites were derived for Cut repeat 1, Cut repeat 3, and Cut repeat 3 plus homeodomain; however, in each case, some selected sequences diverged from the consensus by one, two, or even more nucleotides. This was not due to an incomplete selection procedure, since these sequences were shown in EMSA to be recognized with high affinity. Therefore, these results indicate that Cut repeats can tolerate a certain degree of flexibility in their DNA targets. A similar relaxed sequence specificity was found for the GATA factors (29, 39). One possible explanation for the apparent relaxed binding specificity of Cut repeats comes from the observation that sequences flanking the core consensus site are also part of the recognition site. This was first seen in EMSA

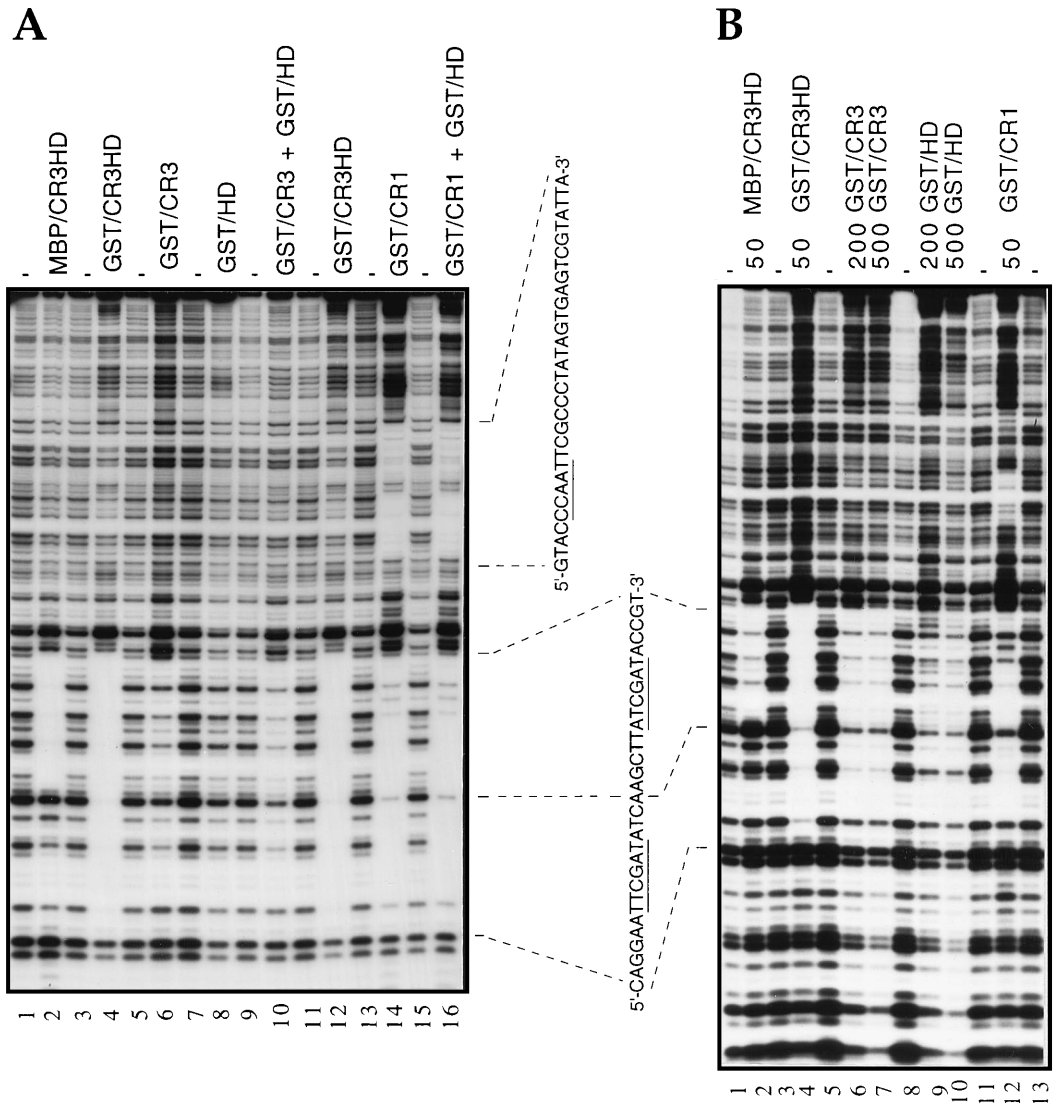

FIG. 8. DNase footprinting analysis of Cut fusion proteins. (A) Fifty nanograms of each fusion protein was incubated with labeled DNA. Odd-numbered lanes, control DNase digestion in the absence of protein. Proteins used for other lanes are indicated at the top. (B) Variable amounts (indicated in nanograms at the top) of proteins were used. Odd-numbered lanes, control DNase digestion in the absence of protein.

using a panel of selected binding sites. Furthermore, when dissociation constants for different sites were compared, specific 3′ flanking sequences were shown to have a twofold effect on the binding affinity. Therefore, as was suggested for the GATA factors, the observed divergences between some selected sites could be explained on the basis that binding affinities depend not exclusively on the conformity of each sequence with the core consensus binding site but possibly also on the conjoined frequencies of specific nucleotides at determined positions in the core and in flanking sequences (39).

Four cDNAs encoding mammalian homologs of the *Drosophila* Cut protein have been isolated so far, using four apparently divergent DNA-binding sites in the protein purification procedure or in the cDNA plaque screening: the FP and the ME1a1 sequences for the human CDP/Cut, the βe2 subelement for the canine Clox protein, and the *Ncam* promoter a sequence for the murine Cux protein (3, 16, 40, 54). In addition, we have previously demonstrated that the human Cut protein can bind to the C3S sequence (23). The C3S sequence includes the ATCGAT motif which corresponds to the consensus binding site for Cut repeat 1 and Cut repeat 3 plus homeodomain. We also note that multimerization of the βe2 subelement by blunt-end ligation would reconstitute the sequence ATCGAT. This helps explain why a dimer of the βe2 subelement was a good binding site for Clox whereas the monomer was not (3). The FP sequence contains several AT doublets as well as a CCAAT motif. Such motifs as well as closely related sequences were found in a significant fraction of the sites selected for the Clox and the human Cut DNA-binding domains (reference 2 and 4 and this study). The ME1a1 sequence contains a GATC motif which is also found in 16 of 34 sequences selected by the GST/CR3HD fusion protein. In summary, each of the previously identified binding sites contains a sequence motif that is found among the sequences obtained by in vitro site selection. This confirms that these apparently divergent sequences are genuine binding sites for Cut-like proteins. On the other hand, taken together, these findings also suggest that the mammalian Cut-like proteins

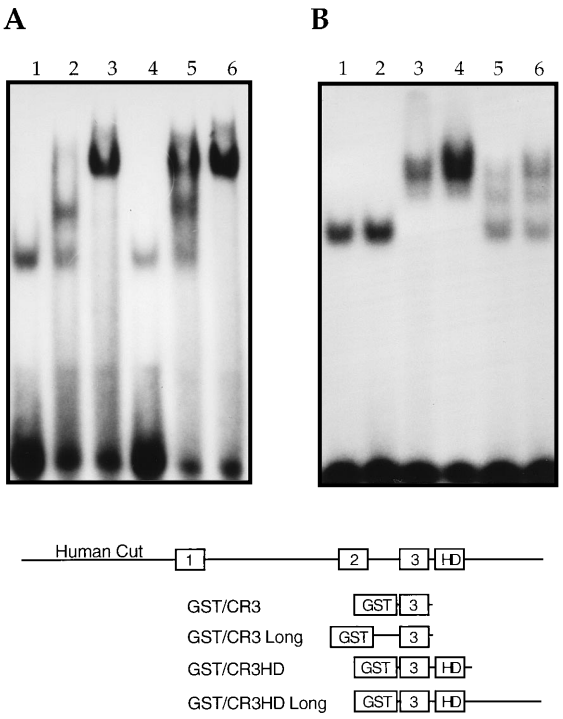

FIG. 9. EMSA using GST/Cut fusion proteins that had been incubated in 5.3 M urea. Bacterial extracts, either mixed or treated separately, were incubated in the presence of 5.3 M urea for 30 min and then dialyzed extensively before purification of GST fusion proteins. (A) GST/CR3 fusion proteins of different molecular weights were incubated with end-labeled double-stranded oligonucleotides C3S (5′-AAAAGAAGCTTATCGATACCGT-3′) in the presence of 50 ng of poly(dI-dC). Lanes: 1, GST/CR3 (10 ng); 2, GST/CR3 plus GST/CR3 long (50 ng); 3, GST/CR3 long (100 ng); 4, GST/CR3 (10 ng); 5, GST/CR3 plus GST/CR3 long (100 ng); 6, GST/CR3 long (100 ng). (B) GST/CR3HD fusion proteins of different molecular weights were incubated with end-labeled double-stranded oligonucleotides A (see Table 2) in the presence of 50 ng of poly(dI-dC). Lanes: 1 and 2, 5 and 10 ng of GST/CR3HD; 3 and 4, 5 and 10 ng of GST/CR3HD long; 5 and 6, 5 and 10 ng of GST/CR3HD plus GST/CR3HD long.

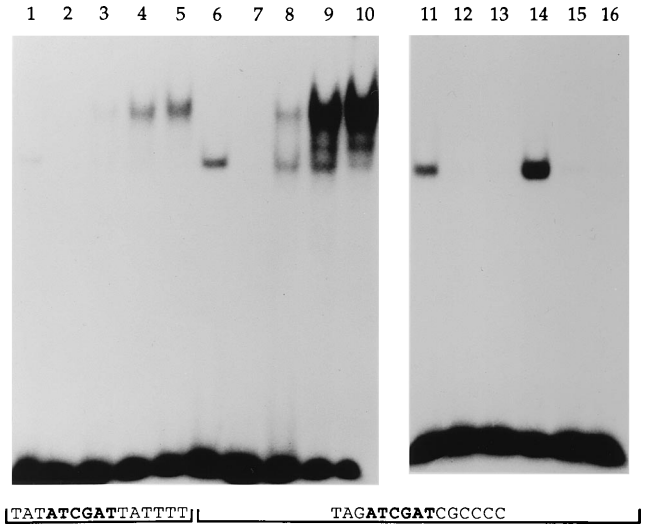

FIG. 10. EMSA using nuclear extracts from COS and HeLa cells. Nuclear extracts prepared from untransfected (COS and HeLa) and transfected (COS) cells were used in EMSA with double-stranded oligonucleotides containing an ATCGAT core with (lanes 1 to 5) or without (lanes 6 to 16) an ATTA homeodomain-binding motif. Transfected COS cells received the vector pSG5Cut, which encodes a human Cut-like protein with the three Cut repeats and the Cut homeodomain. Lanes: 1 and 6, 2 μl of nuclear extract from untransfected COS cells; 2 and 7, probe alone; 3 and 8, 4 and 9, and 5 and 10, 1, 2 and 4 μl, respectively, of nuclear extract from transfected COS cells; 11 to 13, 2 μl of nuclear extract from untransfected COS cells incubated with monoclonal antibodies against hemagglutinin (lane 11) or human Cut-like proteins (antibodies A and W3) (lanes 12 and 13); 14 to 16, 2 μl of nuclear extract from untransfected HeLa cells incubated with monoclonal antibodies against hemagglutinin (lane 14) or human Cut-like proteins (A and W3) (lanes 15 and 16).

have a relaxed DNA-binding specificity and that they are able to recognize a wide range of sequences. Indeed, a significant fraction of selected binding sites did not perfectly conform to the derived consensus site yet represented high-affinity binding sites.

EMSA and DNase footprint analysis with GST and MBP fusion proteins indicated that DNA-binding affinity was greatly increased when the Cut homeodomain was present in addition to Cut repeat 3. Moreover, it appears from the site selection analysis and subsequent gel retardation and DNase footprinting studies that DNA-binding specificity was also modified. Since these two domains are situated very close to one another in the native Cut proteins, it is tempting to speculate that they normally function in concert. The cooperation between the Cut homeodomain and Cut repeat 3 is analogous in some respects to that between the POU-specific domain and the POU homeodomain. For both POU and Cut proteins, the highest DNA-binding affinity and specificity is achieved when their respective homeodomains bind to DNA in conjunction with their adjacent DNA-binding domains (5, 55). In both cases, the presence of the second DNA-binding domain slightly modifies the binding specificity (5, 55). In contrast to the POU-specific domain, however, Cut repeats can bind to DNA with high affinity on their own.

Site selection using either GST/Cut homeodomain or GST/CR3H3 fusion proteins did not reveal a significant number of targets containing the homeodomain ATTA-binding motif. This result was not completely unexpected since the Cut homeodomain differs from that of Engrailed and Antennapedia at many positions involved in DNA contacts, and the effects of these differences on DNA-binding specificity have not yet been determined (28, 42). In contrast to our results, a consensus binding site with an ATTA motif was recently reported for Clox, another mammalian Cut-like protein (2). We found the same ATCGAT core consensus but with GC-rich flanking sequences. When both consensus binding sites were tested in parallel by using nuclear extracts from untransfected COS and HeLa cells, the consensus binding site with GC-rich flanking sequences was preferred. It is possible that the Clox protein does not have the same DNA-binding specificity as the monkey or the human Cut-like protein. The canine and human proteins have identical amino acid sequences within the Cut repeats but have several differences outside these domains (3, 40). It is possible that some of these differences, in particular within the Cut homeodomain, affect the DNA-binding specificity.

As this report was under revision, a study of Cut repeat DNA-binding specificity was published by another group (4). Similarly to our study, selected oligonucleotides for Cut repeat 3 contained the sequence TATNG (CNATA, in the reverse orientation). The ATCGAT sequence was part of the consensus binding site for Cut repeat 1 and was also present in 6 of 14 sites selected for Cut repeat 3. Our observation that different Cut repeats display overlapping yet distinct DNA-binding preferences was also made by these authors. On the other hand, the CCAAT or CCGAT motif was found in sequences selected for both Cut repeats 1 and 3 but at frequencies different from those that we found. It is likely that the broad DNA-binding

preference exhibited by Cut repeats is responsible for the fact that similar sequences are found but at variable frequencies in independent binding-site selections.

GSTs function as dimers. Therefore, we tested the hypothesis that GST/Cut fusion proteins bind to DNA as dimers. When GST/Cut fusion proteins of different molecular weights were mixed, denatured, renatured and then used in EMSA, protein-DNA complexes of intermediate mobility were observed. The appearance of such complexes is best explained by assuming that heterodimers of GST/Cut fusion proteins were formed during the denaturation-renaturation process. Additional evidence for GST/Cut fusion protein binding to DNA as dimers was provided from the comparison of DNase footprints produced by MBP/CR3HD and GST/CR3HD fusion proteins. The MBP protein protected a region of 16 bp corresponding to the top half of the 32-bp region protected by the GST protein. Finally, determination of the dissociation constants of these two proteins for a consensus binding site revealed K_Ds of 4.6×10^{-10} M for the MBP/CR3HD protein and 3.1×10^{-11} M for the GST/CR3HD protein. This result is also in agreement with the notion that GST/Cut proteins bind to DNA as dimers.

The three consensus binding sites derived for Cut repeat 1, Cut repeat 3, or Cut repeat 3 plus homeodomain contain a palindromic sequence. While we demonstrated that GST/Cut fusion proteins bind to DNA as dimers, we do not think that palindromic binding sites were selected essentially because dimeric GST/Cut fusion proteins were used in the selection procedure. If this were the case, we would have expected that the monomeric MBP/CR3HD protein would have protected a region more or less corresponding to one side of the palindrome, while the dimeric GST/CR3HD would protect both sides. Instead, the 16-bp region protected by MBP/CR3HD included the ATCGAT palindrome in its center, and the GST/CR3HD protein protected the same 16-bp region plus another region of 16 bp which contained an imperfect palindrome, TTCGAT. These results are best explained by assuming that binding to the ATCGAT sequence is determinant in targeting the GST/CR3HD dimers to this region of DNA and that protection of the adjacent 16 bp results from cooperative binding. In the same line of thinking, it is striking that the footprints produced by GST/CR1 were also of 32 bp and that the ATC GAT or CCAAT sites were positioned not in the middle of these regions but clearly on one side of it. Therefore, we speculate that binding to the CCAAT or ATCGAT sites targeted the GST/CR1 to these regions and that cooperative binding allowed binding to, and protection of, larger regions of DNA.

In addition to the fusion proteins described here, we have recently prepared a set of MBP/Cut fusion proteins: MBP/CR1, MBP/CR3, MBP/HD, and MBP/CR3HD. None of these proteins except MBP/CR3HD were found to bind to DNA (Fig. 7 and 8 and data not shown). Although many factors can explain a negative result, it is striking that the groups that have studied DNA binding of individual Cut repeats invariably have used GST fusion proteins (2, 4). It is tempting to speculate that Cut repeats cannot, as a single unit, form a stable complex with DNA. We hypothesize that DNA binding requires that at least two Cut DNA-binding domains act in concert: either Cut repeat 3 and the Cut homeodomain (as in MBP/CR3HD and GST/CR3HD) or two Cut repeats (as in the GST/CR1, GST/CR2, and GST/CR3 dimers) (references 2 and 4 and this study). In the native Cut proteins, interaction with DNA may involve cooperation of all the DNA-binding domains. Alternatively, we can envisage that the evolutionary conserved region predicted to form a coiled-coil structure at the amino-terminal end of the Cut proteins may allow the formation of Cut homodimers which would then bind to DNA in a similar manner as the GST/Cut homodimers. Future experiments should address this question.

Combined results from several groups indicate that Cut-like proteins contain four DNA-binding domains (references 2, 4, and 23 and this study). This raises interesting questions about the interactions of Cut-like proteins with their target genes in the cell. On one hand, the presence of several DNA-binding domains with related but distinct DNA-binding preferences could obviously increase the repertoire of target genes that can be regulated by Cut-like proteins. Alternatively, we might envisage that targeting to specific genes requires cooperative binding of two, three, or even four DNA-binding domains to distinct sites within the same promoter. We would then predict that the distance between independent binding sites will have an important effect on cooperative binding, as was shown for some bacterial repressors (30). Another aspect worthy of consideration is the effect of cooperative binding on gene expression. As previously demonstrated for multimeric DNA-binding proteins, binding to two sites involves DNA looping, which, in turn, may affect gene expression in a positive or negative manner (24, 35). Another effect of cooperative binding is to permit binding to lower-affinity binding sites. When such a site overlaps with a *trans*-activator binding site, competition for binding may result in repression of gene expression. This mechanism has been proposed to explain repression of the *Ubx* promoter by *eve* protein (50). In this instance, cooperative binding required protein-protein interactions between *eve* monomers. Multimerization would not be a prerequisite for cooperative DNA binding by Cut-like proteins since several DNA-binding domains are present within these proteins. Future experiments will address whether repression by Cut-like proteins involves or even requires cooperative binding to multiple sites.

ACKNOWLEDGMENTS

A.N. and R.H. are the recipients of a scholarship and a fellowship, respectively, from the Fonds de la Recherche en Santé du Québec. This research was supported by grant MT-11590 from the Medical Research Council of Canada and grant 3497 from the National Cancer Institute of Canada to A.N.

REFERENCES

1. **Aceto, A., M. Caccuri, P. Sacchetta, T. Bucciarelli, B. Dragani, N. Rosato, G. Federici, and C. Di Ilio.** 1992. Dissociation and unfolding of Pi-class glutathione transferase. Evidence for a monomeric inactive intermediate. Biochem. J. **285:**241–245.
2. **Andrés, V., M. D. Chiara, and V. Mahdavi.** 1994. A new bipartite DNA-binding domain: cooperative interaction between the Cut repeat and homeo domain of the Cut homeo proteins. Genes Dev. **8:**245–257.
3. **Andres, V., B. Nadal-Ginard, and V. Mahdavi.** 1992. *Clox*, a mammalian homeobox gene related to *Drosophila Cut*, encodes DNA-binding regulatory proteins differentially expressed during development. Development **116:**324–334.
4. **Aufiero, B., E. J. Neufeld, and S. H. Orkin.** 1994. Sequence-specific DNA binding of individual Cut repeats of the human CCAAT displacement/Cut homeodomain protein. Proc. Natl. Acad. Sci. USA **91:**7757–7761.
5. **Aurora, R., and W. Herr.** 1992. Segments of the POU domain influence one another's DNA-binding specificity. Mol. Cell. Biol. **12:**455–467.
6. **Barberis, A., G. Superti-Furga, and M. Busslinger.** 1987. Mutually exclusive interaction of the CCAAT-binding factor and of a displacement protein with overlapping sequences of a histone gene promoter. Cell **50:**347–359.
7. **Baumhueter, S., D. Mendel, P. Conley, C. Kuo, C. Turk, M. K. Graves, C. A. Edwards, G. Courtois, and G. R. Crabtree.** 1990. HNF-1 shares three sequence motifs with the POU domain proteins and is identical to LF-B1 and APF. Genes Dev. **4:**372–379.
8. **Blackwell, T. K., L. Kretzner, E. M. Blackwood, B. Eisensman, and H. Weintraub.** 1990. Sequence-specific DNA binding by the c-Myc protein. Science **250:**1149–1151.
9. **Blackwell, T. K., and H. Weintraub.** 1990. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science **250:**1104–1110.
10. **Blochlinger, K., J. Bodmer, L. Y. Jan, and Y. N. Jan.** 1990. Patterns of

expression of *Cut*, a protein required for external sensory organ development in wild-type and *Cut* mutant *Drosophila* embryos. Genes Dev. **4:**1322–1331.
11. **Blochlinger, K., R. Bodmer, J. Jack, L. Y. Jan, and Y. N. Jan.** 1988. Primary structure and expression of a product from *Cut*, a locus involved in specifying sensory organ identity in *Drosophila*. Nature (London) **333:**629–635.
12. **Blochlinger, K., L. Y. Jan, and Y. N. Jan.** 1991. Transformation of sensory organ identity by ectoc expression of Cut in *Drosophila*. Genes Dev. **5:**1124–1135.
13. **Blochlinger, K., L. Y. Jan, and Y. N. Jan.** 1993. Postembryotic patterns of expression of *Cut*, a locus regulating sensory organ identity in *Drosophila*. Development **117:**441–450.
14. **Bodmer, R., S. Barbel, S. Sheperd, J. W. Jack, L. Y. Jan, and Y. N. Jan.** 1987. Transformation of sensory organs by mutations of the *Cut* locus of *D. melanogaster*. Cell **51:**293–307.
15. **Bopp, D., M. Burri, S. Baumgartner, G. Frigerio, and M. Noll.** 1986. Conservation of a large protein domain in the segmentation gene *paired* and in functionally related genes in *Drosophila*. Cell **47:**1033–1040.
16. **Dufort, D., and A. Nepveu.** 1994. The human Cut homeodomain protein represses expression from the c-*myc* promoter. Mol. Cell. Biol. **14:**4251–4257.
16a. **Dynan, W. S., and R. Tjian.** 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell **35:**79–87.
17. **Finney, M., G. Ruvkun, and H. R. Horvitz.** 1988. The C elegans cell lineage and differentiation gene unc-86 encodes a protein containing a homeodomain and extended sequence similarity to mammalian transcription factors. Cell **55:**757–769.
18. **Frain, M., G. Swart, P. Monaci, S. Stampfli, R. Frank, and R. Cortese.** 1989. The liver-specific transcription factor LF-B1 contains a highly diverged homeobox DNA binding domain. Cell **59:**145–157.
19. **Freyd, G., S. Kim, and R. Horvitz.** 1990. Novel cysteine-rich motif and homeodomain in the product of the *Caenorhabditis elegans* cell lineage gene *lin-11*. Nature (London) **344:**876–879.
20. **Galas, D. L., and A. Schmitz.** 1978. DNase footprinting: a simple method for the detection of protein-DNA bindng specificity. Nucleic Acids Res. **5:**3157–3170.
21. **Gruss, P., and C. Walther.** 1992. Pax in development. Cell **69:**719–722.
22. **Guan, C., P. Li, P. D. Riggs, and H. Inouye.** 1987. A vector that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. Gene **67:**21–30.
23. **Harada, R., D. Dufort, C. Denis-Larose, and A. Nepveu.** 1993. Conserved Cut repeats in the human Cut homeodomain protein function as DNA binding domains. J. Biol. Chem. **269:**2062–2067.
24. **Hochschild, A.** 1990. Protein-protein interactions and DNA loop formation, p. 107–137. *In* N. R. Cozzarelli and J. C. Wang (ed.), DNA topology and its biological effects. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
25. **Ingraham, H. A., S. E. Flynn, J. W. Voss, V. R. Albert, M. S. Kapiloff, L. Wilson, and M. G. Rosenfeld.** 1990. The POU-specific domain of Pit-1 is essential for sequence-specific, high affinity DNA-binding and DNA-dependent Pit-1-Pit-1 interactions. Cell **61:**1021–1033.
26. **Jack, J., D. Dorsett, Y. Delotto, and S. Liu.** 1991. Expression of the *Cut* locus in the *Drosophila* wing margin is required for cell type specification and is regulated by a distant enhancer. Development **113:**735–747.
27. **Karlsson, O., S. Thor, T. Norberg, H. Ohlsson, and T. Edlund.** 1990. Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. Nature (London) **344:**879–882.
28. **Kissinger, C. R., B. Liu, E. Martin-Blanco, T. B. Kornberg, and C. Pabo.** 1990. Crystal structure of an Engrailed homeodomain-DNA complex at 2.8 angstrom resolution: a framework for understanding homeodomain-DNA interactions. Cell **63:**579–590.
29. **Ko, L. J., and J. D. Engel.** 1993. DNA-binding specificities of the GATA transcription factor family. Mol. Cell. Biol. **13:**4011–4022.
30. **Krämer, H., M. Niemöller, M. Amouyal, B. Revet, B. von Wilcken-Bergmann, and B. Müller-Hill.** 1987. *lac* repressor forms loops with linear DNA carrying two suitably spaced *lac* operators. EMBO J. **6:**1481–1491.
31. **Lee, K. A. W., A. Bindereif and M. R. Green.** 1988. A small-scale procedure for preparation of nuclear extracts that support efficient transcription and pre-mRNA splicing. Gene Anal. Tech. **5:**22–31.
32. **Liu, S., and J. Jack.** 1992. Regulatory interactions and role in cell type specification of the Malpighian tubulles by the Cut, Kruppel, and caudal genes of Drosophila. Dev. Biol. **150:**133–143.
33. **Liu, S., E. McLeod, and J. Jack.** 1991. Four distinct regulatory regions of the *Cut* locus and their effect on cell type specification in *Drosophila*. Genetics **127:**151–159.
34. **Maina, C. V., P. D. Riggs, A. G. III Grandea, B. E. Slatko, L. S. Moran, J. A. Tagliamonte, L. A. McReynolds, and C. Guan.** 1988. A vector to express and purify foreign proteins in *Escherichia coli* by fusion to, and separation from, maltose binding protein. Gene **74:**365–373.
35. **Mandal, N., W. Su, R. Haber, S. Adhya, and H. Echols.** 1990. DNA looping in cellular repression of transcription of the galactose operon. Gene Dev. **4:**410–418.
36. **McCutchan, J. H., and J. S. Pagano.** 1968. Enhancement of the infectivity of simian virus 40 deoxyribonucleic acid with diethyl aminoethyl-dextran. J. Natl. Cancer Inst. **41:**351–354.
37. **McGinnis, W., and R. Krumlauf.** 1992. Homeobox genes and axial patterning. Cell **68:**283–302.
38. **Mendel, D. B., P. A. Khavari, P. B. Conley, P. B. Graves, L. P. Hansen, A. Admon, and G. R. Crabtree.** 1991. Characterization of a cofactor that regulates dimerization of a mammalian homeodomain protein. Science **254:**1762–1763.
39. **Merika, M., and S. H. Orkin.** 1993. DNA-binding specificity of GATA family transcription factors. Mol. Cell. Biol. **13:**3999–4010.
40. **Neufeld, E. J., D. G. Skalnik, P. M.-J. Lievens, and S. H. Orkin.** 1992. Human CCAAT displacement protein is homologous to the *Drosophila* homeoprotein, *Cut*. Nature Genet. **1:**50–55.
41. **Nicosia, A., P. Monaci, L. Tomei, R. DeFransesco, M. Nuzzo, H. Stunnenberg, and R. Cortese.** 1990. A myosin-like dimerization helix and an extra-large homeodomain are essential elements of the tripartite DNA binding structure of LFB1. Cell **61:**1225–1236.
42. **Otting, G., Y. Q. Qian, M. Billeter, M. Muller, M. Affolter, W. J. Gehring, and K. Wuthrich.** 1990. Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscoppy in solution. EMBO J. **9:**3085–3092.
43. **Ruvkun, G., and M. Finney.** 1991. Regulation of transcription and cell identify by POU domain proteins. Cell **64:**475–478.
43a. **Sacchetta, P., A. Aceto, T. Bucciarelli, B. Dragani, S. Santarone, N. Allocati, and C. Di Ilio.** 1993. Multiphasic denaturation of glutathione transferase B1-1 by guanidium chloride. Eur. J. Biochem. **215:**741–745.
44. **Scott, M. P., J. W. Tamkun, and G. W. Hartzell.** 1989. The structure and function of the homeodomain. Biochem. Biophys. Acta **989:**25–48.
45. **Skalnik, D. G., E. C. Strauss, and S. H. Orkin.** 1991. CCAAT displacement protein as a repressor of the myelomonocytic-specific gp91-phox gene promoter. J. Biol. Chem. **266:**16736–16744.
46. **Smith, D. B., and K. S. Johnson.** 1988. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene **67:**31–40.
47. **Sturm, R. A., and W. Herr.** 1988. The POU domain is a bipartite DNA-binding structure. Nature (London) **336:**601–604.
48. **Superti-Furga, G., A. Barberis, G. Schaffner, and M. Busslinger.** 1988. The -117 mutation in greek HPFH affects the binding of the three nuclear factors to the CCAAT region of the γ-globulin gene. EMBO J. **7:**3099–3107.
49. **Superti-Furga, G., A. Barberis, E. Schreiber, and M. Busslinger.** 1989. The protein CDP, but not CP1, footprints on the CCAAT region of the γ-globulin gene in unfractionated B-cell extracts. Biochim. Biophys. Acta **1007:**237–242.
50. **TenHarmsel, A., R. J. Austin, N. Savenelli, and M. D. Biggin.** 1993. Cooperative binding at a distance by *even-skipped* protein correlates with repression and suggests a mechanism of silencing. Mol. Cell. Biol. **13:**2742–2752.
51. **Treisman, J., E. Harris, and C. Desplan.** 1991. The paired box encodes a second DNA-binding domain in the paired homeo domain protein. Genes Dev. **5:**594–604.
52. **Treisman, J., E. Harris, D. Wilson, and C. Desplan.** 1992. The homeodomain: a new face for the helix-turn-helix? Bioessays **14:**145–150.
53. **Tymon, A. M., U. Kües, W. V. J. Richardson, and L. A. Casselton.** 1992. A fungal mating type protein that regulates sexual and asexual development contains a POU-related domain. EMBO J. **11:**1805–1813.
54. **Valarché, I., J.-P. Tissier-Seta, M. R. Hirsch, S. Martinez, C. Goridis, and J. F. Brunet.** 1993. The mouse homeodomain protein Phox2 regulates Ncam promoter activity in concert with Cux/CDP and is a putative determinant of neurotransmitter phenotype. Development **119:**881–896.
55. **Verrijzer, C. P., M. J. Alkema, W. W. van Weperen, H. C. Van Leeuwen, J. J. Strating, and P. C. van der Vilet.** 1992. The DNA binding specificity of the bipartite POU domain and its subdomains. EMBO J. **11:**4993–5003.