Formation of a Monomeric DNA Binding Domain by Skn-1 bZIP and Homeodomain Elements

T. Keith Blackwell,* Bruce Bowerman,† James R. Priess, Harold Weintraub

Maternally expressed Skn-1 protein is required for the correct specification of certain blastomere fates in early *Caenorhabditis elegans* embryos. Skn-1 contains a basic region similar to those of basic leucine zipper (bZIP) proteins but, paradoxically, it lacks a leucine zipper dimerization segment. Random sequence selection methods were used to show that Skn-1 binds to specific DNA sequences as a monomer. The Skn-1 basic region lies at the carboxyl terminus of an 85–amino acid domain that binds preferentially to a bZIP half-site and also recognizes adjacent 5′ AT-rich sequences in the minor groove, apparently with an amino (NH_2)-terminal "arm" related to those of homeodomain proteins. The intervening residues appear to stabilize interactions of these two subdomains with DNA. The Skn-1 DNA binding domain thus represents an alternative strategy for promoting binding of a basic region segment recognition helix to its cognate half-site. The results point to an underlying modularity in subdomains within established DNA binding domains.

In *Caenorhabditis elegans*, the maternal *skn-1* gene product is required during early embryonic development for the correct specification of pharyngeal, intestinal, and muscle cell fates (*1*, *2*). In embryos produced by homozygous *skn-1* mutants, precursors of these cells instead differentiate along alternative fate pathways, which include formation of excess hypodermal (or "skin") cells (*1*). Skn-1 is a nuclear protein that is detectable between the two- and eight-cell stages of development and is asymmetrically localized to posterior blastomeres (*3*). Maternally expressed genes that are required for regulating the localization (*par* and *mex* genes), and possibly the activity (*pie-1* gene), of the Skn-1 protein have been identified (*2*, *3*). Mutations in these genes result in *skn-1*–dependent, ectopic formation of pharyngeal, intestinal, and muscle cells (*2*), suggesting that Skn-1 participates in committing blastomeres to these various differentiation pathways during normal development. However, it is still not known how Skn-1 performs these functions and no Skn-1–dependent downstream target genes or gene products have been identified.

Near the COOH-terminus of the predicted Skn-1 protein is an amino acid sequence (Fig. 1) (*1*) that is remarkably similar to the basic regions (BR) of the BR leucine zipper (bZIP) family of DNA-binding transcriptional regulators (*4*). A similar COOH-terminal BR is also encoded by a related *C. elegans* gene (*skn-1*–related gene, *srg-1*, Fig. 1) (*1*). The bZIP DNA binding domains are composed of a BR segment that directly contacts the DNA (*5*) and an immediately COOH-terminal "leucine zipper" α-helical dimerization segment (*6*). The conserved amino acid residues with which a bZIP BR makes contact with DNA are present in the Skn-1 BR (Fig. 1); this similarity suggests that Skn-1 might also be a DNA binding protein. However, Skn-1 differs from bZIP proteins in a critical and surprising aspect: it lacks a leucine zipper. In bZIP proteins the leucine zipper segment is essential for DNA binding because it mediates dimerization, which these proteins require for DNA binding, and also because it orients and stabilizes the adjacent BR segment, which is apparently unstructured when it is not bound to DNA (*7*–*9*). Because Skn-1 does not contain a leucine zipper and if it indeed binds to DNA, it must do so in a way that is distinct from that of bZIP proteins.

Using a random-sequence selection strategy, we have shown that Skn-1 binds to specific DNA sequences in vitro. The consensus-preferred Skn-1 binding site is composed of two contiguous sequence motifs. One of these sequence motifs is identical to one-half of the palindromic site recognized by certain bZIP proteins, and the other is an adjacent 5′ AT-rich region. Skn-1 binds to these sites at high affinity as a monomer. This DNA binding activity resides within 85 amino acid residues which lie at the Skn-1 COOH-terminus, and which we have designated the Skn domain. At the COOH-terminus of the Skn domain is its BR, and at its NH_2-terminus is a sequence that is identical to the peptide "arm" with which Antennapedia, a *Drosophila* homeodomain protein, recognizes bases in the minor groove (*10*). A series of biochemical and mutagenesis experiments suggest that a Skn-1 monomer binds to its cognate site in a bipartite manner, placing the BR into the major groove of the bZIP half-site, and the arm into the minor groove of the AT-rich region. The intervening Skn domain residues appear to stabilize the complex and properly orient the COOH-terminal BR and NH_2-terminal arm for binding to DNA. Our results show that Skn-1 (and the product of *srg-1*) define an unusual class of DNA binding proteins, and support the idea that maternally expressed Skn-1 functions by regulating transcription of zygotically expressed target genes. These results also demonstrate an underlying modularity in subdomains within these various DNA binding domains, and reveal that either of two distinct mechanisms can lead to and promote binding of a BR segment to its cognate half-site.

Binding of Skn-1 to DNA as a monomer, with both bZIP and homeodomain elements. We used the selection and amplification of binding sites (SAAB) method (*11*) of reiterative nucleic acid selection (*11*, *12*) to identify sequences bound by a

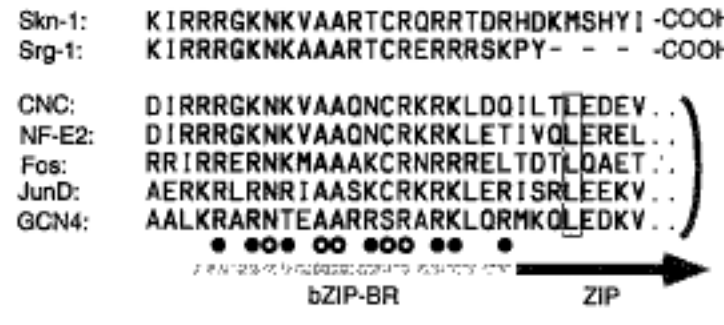


Fig. 1. The Skn-1 basic region. Comparison of the COOH-terminus of Skn-1 with that of its homolog Srg-1, and with the indicated bZIP protein residues. Amino acids that are identical to, or conservative changes from, the corresponding Skn-1 or Srg-1 residues are shaded. Open and closed circles indicate GCN4 residues that were shown by x-ray crystallography to directly contact bases in the major groove, and backbone phosphates, respectively (*5*). The NH_2-terminal end of the leucine zipper (ZIP) is indicated as defined in (*5*), and its first conserved leucine is boxed. Sequences are derived from the following sources: Skn-1 and Srg-1 (*1*); Cap 'n Collar (CNC) (*45*); NF-E2 (*46*); Fos (*47*); JunD (*48*); and GCN4 (*5*).

T. K. Blackwell, B. Bowerman, and J. R. Priess are in the Department of Basic Sciences, and H. Weintraub is in the Howard Hughes Medical Institute, at the Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, WA 98104, USA.

*Present address: Center for Blood Research, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA.

†Present address: Institute of Molecular Biology and Department of Biology, University of Oregon, Eugene, OR 97403, USA.

glutathione-S-transferase (GST)–Skn-1 fusion protein in vitro. Within 38 DNA molecules derived from five rounds of selection for Skn-1 binding, 19 GTCAT and 17 ATCAT sequences were present (*13*) (Fig. 2), frequencies that are far greater than would be predicted from chance [1 in 512 base pairs (bp) for each]. This finding is striking because GTCAT is the optimal half-site for bZIP proteins related to the transcription factor GCN4, which binds as a dimer preferentially to the palindromic elements 5′-ATGACTCAT-3′ or 5′-ATGACGTCAT-3′; these sites differ only by the overlap at the center base (*14*). Only one of the 38 selected sequences contained two contiguous GTCAT or ATCAT half-sites (*15*), suggesting that Skn-1 recognizes a single half-site motif. Skn-1 also selected an AT-rich region immediately adjacent to these bZIP half-sites (Fig. 2), suggesting that these bases also form part of the binding site.

These predictions were confirmed by analysis of individual selected molecules for binding by Skn-1, and by examination of how base methylation with dimethyl sulfate interfered with Skn-1 binding (*16*). Dimethyl sulfate methylates the N-7 position of guanine, located in the major groove of DNA, and the N-3 position of adenine, which is in the minor groove (*17*). In bZIP proteins, when bound to DNA, the BR forms an α helix that extends from the NH_2-terminal of the leucine zipper through the major groove, where it contacts bases and backbone phosphates (5). Consistent with recognition in the major groove, methylation of guanines within the bZIP half-site interfered with Skn-1 binding to a relatively high-affinity site (Fig. 3, A and B). Interference also occurred at adenines along the bottom strand (Fig. 3, A and B), suggesting the possibility of binding in the minor groove, which is not characteristic of a bZIP BR (5). Other selected sites were characterized by similar methylation interference patterns (*16*), an indication that the bZIP half-site sequence and surrounding bases constitute a complete Skn-1 binding site. The preference of Skn-1 for A and T bases that are 5′ of the selected GTCAT and ATCAT sequences (Fig. 2) is consistent with these methylation interference data, which suggest base contact at these positions. Full-length Skn-1 produced by in vitro translation binds to a labeled oligonu-

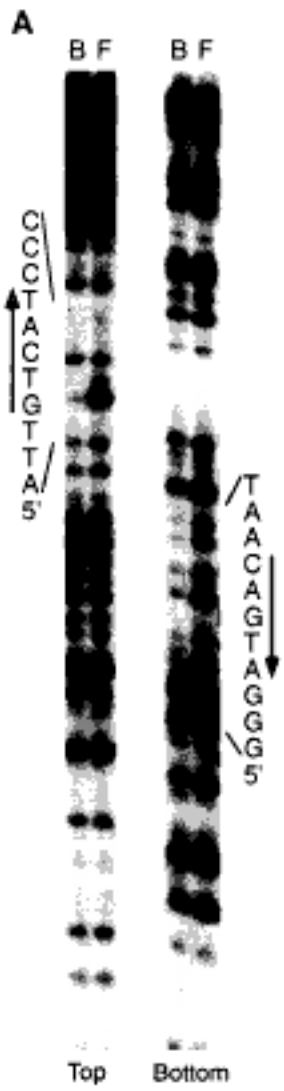
GTCAT SITES:

G:	1	1	1	1	GTCAT	5	1	0
A:	3	10	9	3	GTCAT	1	7	7
T:	9	4	8	14	GTCAT	4	3	7
C:	6	4	1	1	GTCAT	9	8	5
	$\overline{G}$	A	A/T	T	GTCAT	$\overline{A}$	$\overline{G}$	$\overline{G}$
	-4	-3	-2	-1	01234	5	6	7

ATCAT SITES:

G:	2	1	2	0	ATCAT	4	1	2
A:	3	7	5	4	ATCAT	0	6	5
T:	2	6	6	7	ATCAT	4	3	5
C:	6	0	1	2	ATCAT	6	4	2
	N	A/T	A/T	A/T	ATCAT	$\overline{A}$	$\overline{G}$	N
	-4	-3	-2	-1	01234	5	6	7

Fig. 2. Selected Skn-1 binding sites. Sequences that flank GTCAT and ATCAT sequences selected in the SAAB protocol (*13*) are tabulated. These sites are numbered according to the convention for bZIP protein binding sites, so that 0 indicates the base at the center of the inverted repeat, with apparent sequence preferences summarized below each set. Three selected molecules each contain an ATCAT motif that overlaps the 3′ AT sequence of a GTCAT or ATCAT motif, and thus may be artifactual. Only the remaining selected ATCAT sites, which are designated as nonoverlapping, are tabulated. Because one site had an indeterminant base at the −1 position (*15*), in the ATCAT compilation the total listed at −1 is 13. Bases that are selected against are indicated by a line above the letter, and A/T indicates a preference for either base.



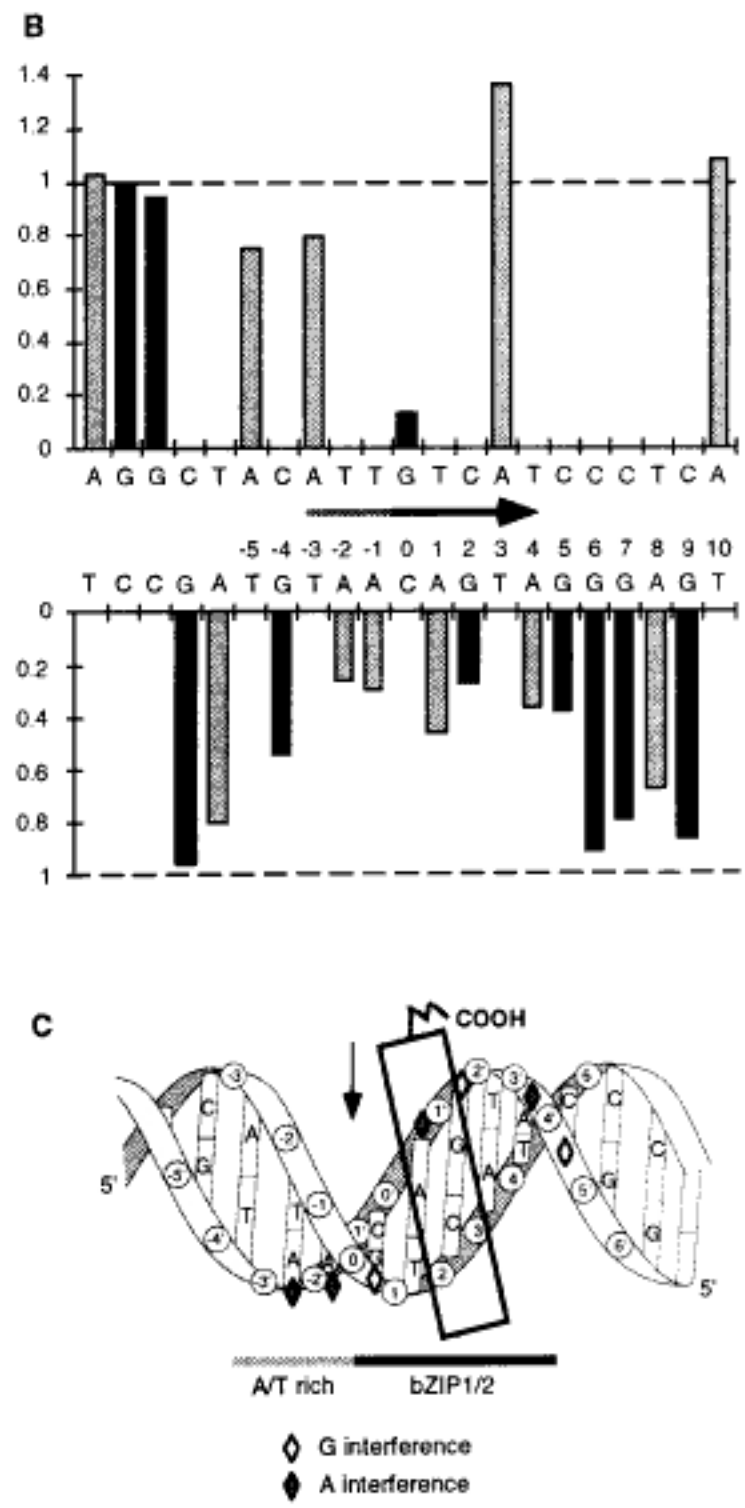


Fig. 3. Methylation interference with Skn-1–DNA binding. (**A**) Autoradiogram of a methylation interference analysis (*49*). End-labeled and partially methylated DNA was incubated with GST–Skn-1 protein, and bound, and free molecules were separated by EMSA. After cleavage at methylated G and A residues, these samples were subjected to electrophoresis on a sequencing gel. Individual labeled DNA strands are indicated below the gels as top and bottom. Arrows indicate the bZIP half-site, and B and F the bound and free EMSA fractions, respectively. (**B**) Bar graph of the relative methylation interference at the indicated site positions, which are numbered as in Fig. 2. The gel autoradiographed in (A) was analyzed on a PhosphorImager, the data were plotted as the ratios of bound to free at each position, and the ratios were normalized to approximately 1 where no interference was observed. Only the relevant positions are shown. The bZIP half-site is indicated as in (A), and the preferred AT-rich region by a shaded line. Black bars denote G residues, and shaded bars A residues. (**C**) A cartoon illustrating the positions at which G and A methylation interference occurred (bound/free <50%). A black box indicates the presumed approximate location of the α-helical BR in the major groove.

cleotide containing its preferred binding site (SK1) (Fig. 4B, lane 2) (*18*), and binds at a similar level to an ATCAT site which is otherwise identical, but does not bind to a sequence in which the bZIP half-site has been mutated (*15*). The preferred Skn-1 binding site thus consists of a single consensus bZIP half-site which can be varied to ATCAT, and which lies immediately 3′ of an AT-rich sequence.

The lack of any sequence repeat or symmetry in the preferred Skn-1 binding site suggests that Skn-1 might bind to DNA as a monomer, in contrast to the dimeric mode of DNA binding characteristic of bZIP proteins (*7*). When two DNA binding protein species that form dimers have different electrophoretic mobilities, exchange among

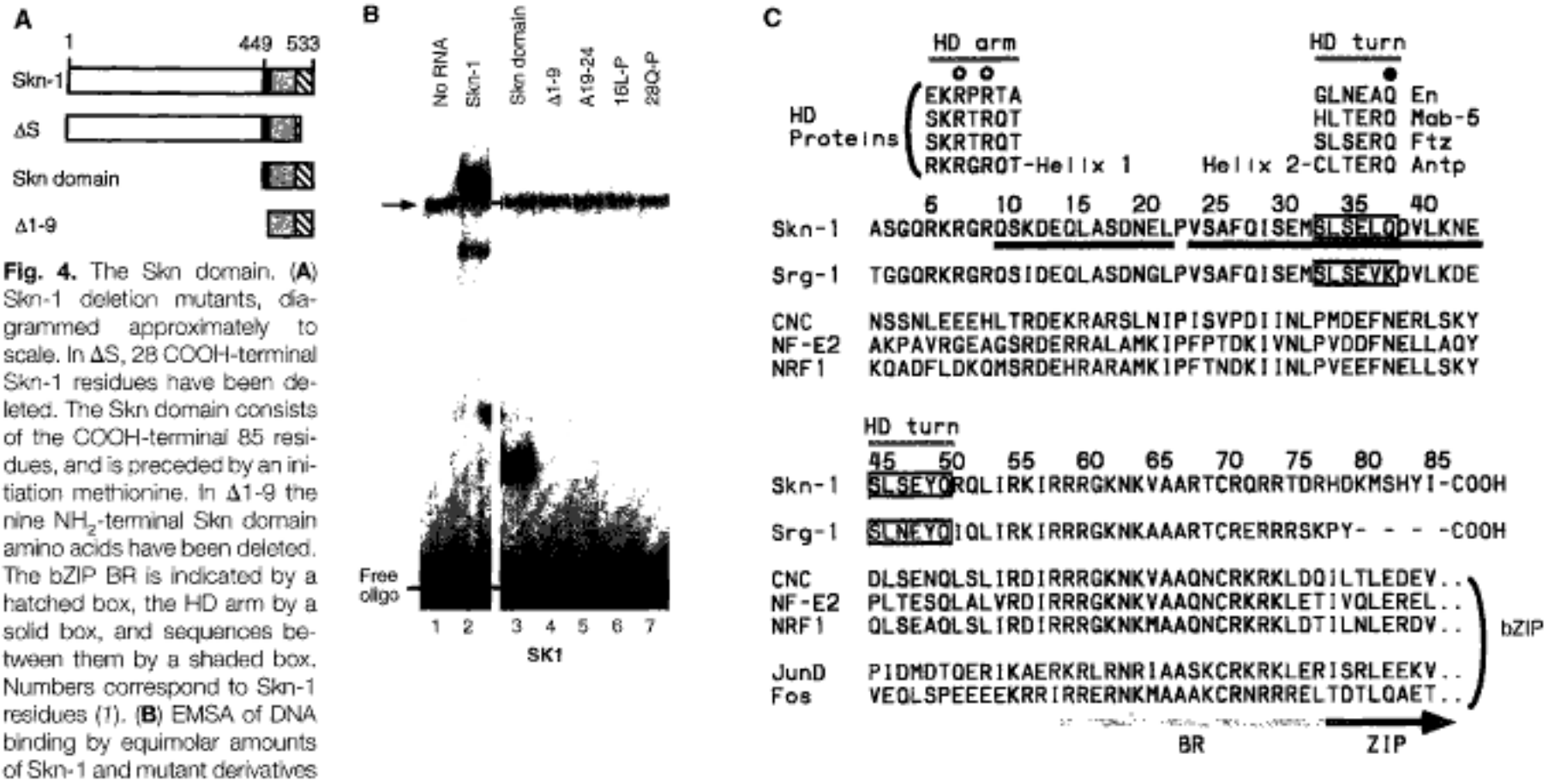


Fig. 4. The Skn domain. (**A**) Skn-1 deletion mutants, diagrammed approximately to scale. In ΔS, 28 COOH-terminal Skn-1 residues have been deleted. The Skn domain consists of the COOH-terminal 85 residues, and is preceded by an initiation methionine. In Δ1-9 the nine NH_2-terminal Skn domain amino acids have been deleted. The bZIP BR is indicated by a hatched box, the HD arm by a solid box, and sequences between them by a shaded box. Numbers correspond to Skn-1 residues (*1*). (**B**) EMSA of DNA binding by equimolar amounts of Skn-1 and mutant derivatives (*50*) to the SK1 probe, which was synthesized based on the preferred Skn-1 binding consensus (Fig. 2). The mutants analyzed in lanes 4 to 7 were derived from the Skn domain, and numbered as in (C). For example, A19–24 denotes substitution of alanine for residues 19–24, and 16L-P indicates substitution of a proline for leucine at that position. A sample that contained unprogrammed lysate is labeled No RNA, and an arrow indicates binding by lysate components. (**C**) Skn domain residues are compared with bZIP and partial homeodomain (HD) sequences as in Fig. 1. Homeodomain residues that abut the COOH-terminus of helix 2, and constitute a turn and the beginning of helix 3 (*22*, *28*), are designated HD turn and indicated by a shaded line like that indicating the HD arm. The locations of homeodomain helices 1 and 2 are indicated, but are not numbered proportionally to Skn domain residues. Homeodomain residues that contact bases in the minor groove are indicated by open circles, and that contact backbone phosphates by a closed circle (*22*). Skn-1 residues that are similar to the HD turn are boxed, and those that Chou-Fasman or Robson-Garnier analyses predict to be α-helical are designated with a black bar underneath. Leucine zipper and BR segments are designated as in Fig. 1. The protein sequences are derived from the following sources: Skn-1 (*1*); Srg-1 (*1*, *51*); Cap'n Collar (CNC) (*45*); NF-E2 (*46*); NRF1 (*52*); Antennapedia (Antp), Fushi-tarazu (Ftz), Mab-5, and Engrailed (en) (*27*); Fos (*47*); and JunD (*48*).

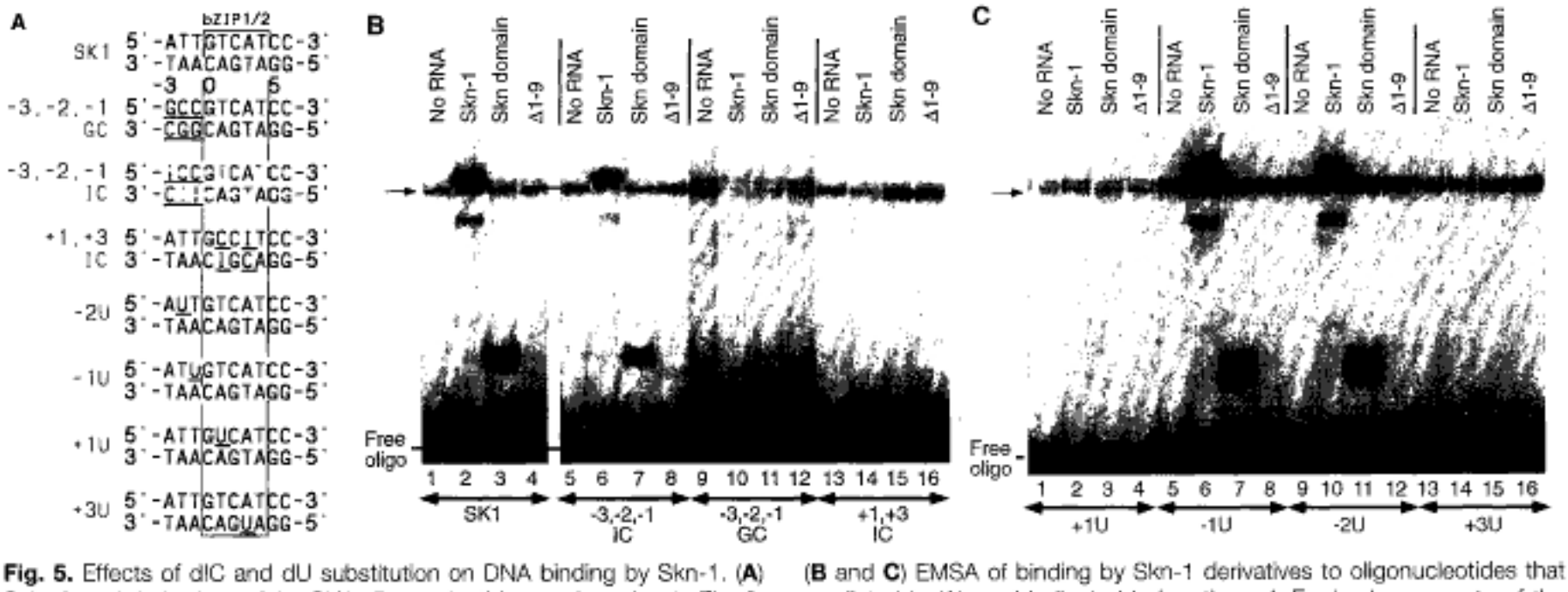
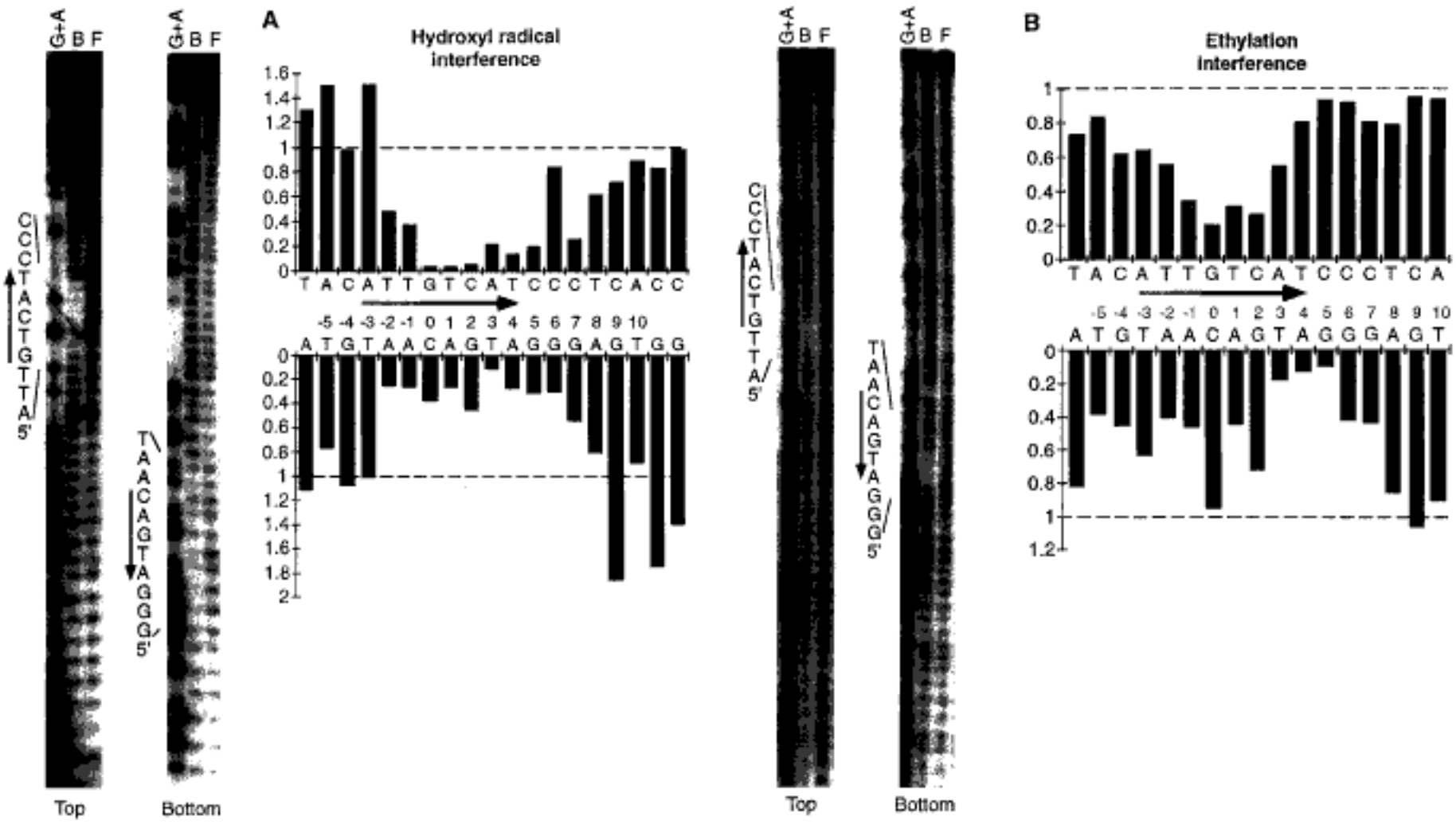


Fig. 5. Effects of dIC and dU substitution on DNA binding by Skn-1. (**A**) Substituted derivatives of the SK1 oligonucleotide, numbered as in Fig. 2. Only 10 bp of each site are shown, with the bZIP half-site indicated in bold type, and inosine by I. Substituted bases are underlined in each sequence. (**B** and **C**) EMSA of binding by Skn-1 derivatives to oligonucleotides that are listed in (A), and indicated below the gel. Equimolar amounts of the indicated proteins were assayed. An arrow indicates binding by lysate components.



dimer subunits produces a heterodimer, which can be detected as a protein-DNA complex of intermediate electrophoretic mobility (*19*). However, when Skn-1 was cotranslated with any of several Skn-1 deletion mutants, no novel intermediate complexes were observed in an electrophoretic mobility shift assay (EMSA). Mixing of in vitro–translated full-length Skn-1 with GST–Skn-1 fusion protein gave a similar result (*15*). The inability to detect dimer formation by this assay is consistent with the possibility that Skn-1 binds to DNA as a monomer. Analysis of Skn-1 deletion mutants (*15*) has localized its DNA binding capability to its 85 COOH-terminal amino acids (the Skn domain) (Fig. 4). An analogous amino acid sequence is also encoded by *srg-1* (Fig. 4C). The Skn domain binds to the SK1 site with an affinity similar to that of the full-length protein (Fig. 4B, lanes 2 and 3). In contrast, DNA binding is dramatically reduced (undetectable) by either deletion of nine amino acids from the NH_2-terminus of the Skn domain (Δ1–9; Fig. 4B, lane 4), or by removal of the BR from Skn-1 (ΔS; Fig. 4A) (*15*). Preliminary measurements by EMSA indicate that a purified Skn domain peptide binds to the SK1 site with an equilibrium dissociation constant (K_d) of approximately 1×10^{-9} M (*20*), an affinity comparable to those of homeodomain and bZIP (*7*, *9*, *21*) proteins for their respective cognate sites.

The essential NH_2-terminal Skn domain residues deleted in the Δ1–9 mutant are remarkably related to the flexible "arm" that lies at the NH_2-terminus of homeodomains (Fig. 4C), which bind to DNA as monomers (*10*, *22*). Structures determined by x-ray crystallography and NMR spectroscopy show that homeodomains otherwise consist of a bundle of three α helices, and that they bind to DNA by inserting a COOH-terminal "recognition helix" (helix 3) into the major groove, and the homeodomain (HD) arm into the minor groove (*10*, *22*). Various proteins bind in the minor groove with similar sequences (*23*–*26*), but we refer to the Skn domain NH_2-terminal arm as the HD arm, because it is nearly identical to the NH_2-terminal arm of the *Drosophila* homeodomain protein Antennapedia (Fig. 4C) (*27*). The amino acid sequence similarity between the Skn domain and the homeodomain is limited to the HD arm, except for a short sequence that is present twice (HD turn; Fig. 4C), and is similar to the bend that immediately abuts the COOH-terminal end of homeodomain helix 2 (*22*, *28*). However, the Skn domain may be analogous to the homeodomain in that its HD arm is also at its NH_2-terminal end, and its presumed major groove recognition helix (the bZIP BR) lies at its COOH-terminus (Fig. 4C).

Binding of the Skn domain in both the major and minor grooves of DNA. The analogy to homeodomain proteins suggests that the Skn domain might bind to DNA in a bipartite manner, with its BR and HD arm segments each functioning like their counterparts in the bZIP and homeodomain protein families, and binding to the major and minor grooves of DNA, respectively. If indeed the Skn-1 BR binds to the bZIP half-site in the major groove and makes the same contacts as in bZIP proteins (*5*), it would be oriented on the binding site as in Fig. 3C, with its COOH-terminus pointing toward the AT-rich portion of the site. When homeodomains bind to DNA, helix 3 is oriented in the major groove so that its COOH-terminal end is pointed toward the bases that are contacted in the minor groove by the HD arm (*10*, *22*). An analogy between the Skn domain and the homeodomain would lead to the prediction that the Skn-1 HD arm would bind to the AT-rich portion of its site in the minor groove. Consistent with this idea, the two residues at which the greatest A methylation interference was observed are in this region (−2 and −1; Fig. 3).

We have tested further these predicted major and minor groove interactions by mutagenesis of the Skn-1 binding site. To detect minor groove binding, A·T base pairs in the site were substituted with either inosine (I)·C or G·C pairs (Fig. 5A). Whereas an I·C base pair is indistinguishable from A·T in the minor groove, a G·C base pair differs significantly from A·T in both grooves. Substitution of the AT-rich portion of the SK1 site (Fig. 5A) with G·C

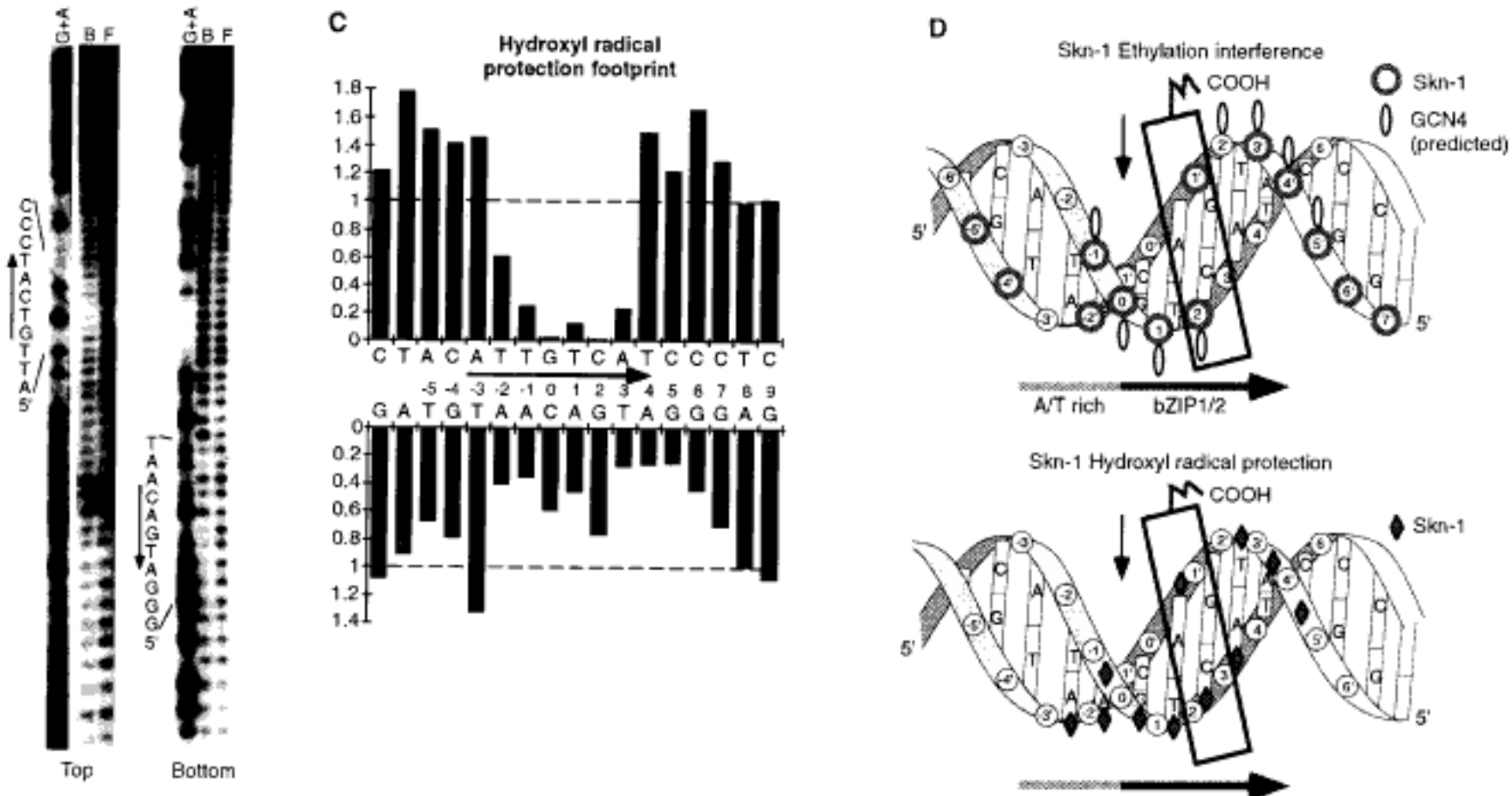


Fig. 6. Interference and protection analyses of Skn-1 DNA binding. (**A**) Hydroxyl radical interference with Skn-1 binding to DNA (*49*). The gel was analyzed by autoradiography and phosphorimaging as in Fig. 3. Site positions were determined by comparison with a G+A track (*49*). (**B**) Ethylation interference with Skn-1 binding to DNA (*49*), analyzed as in (A). (**C**) Skn-1 hydroxyl radical protection footprint (*49*), analyzed as in (A). (**D**) Skn-1 binding interference and protection summary. The top cartoon illustrates Skn-1 ethylation interference (B to F <50%), and the maximum extent to which it would be predicted that a single bZIP BR would make contact with phosphates (based on the GCN4 structure) (*5*). It is not possible to predict these contacts exactly because the GCN4 contacts are different on each half of its binding site (*5*). In the two top diagrams, the approximate predicted position of the α-helical BR is shown by a black box, and the AT-rich and bZIP half-site sequences are indicated as in (A to C). Hydroxyl radical protection footprints for Skn-1 (B to F <50%) and for the α2 homeodomain [taken from (*53*)] are diagrammed in the middle and bottom cartoons, respectively. The extent of protection by α2 is not precisely comparable to the Skn-1 data, because the α2 data were not quantitated. The approximate position (*22*) of homeodomain helix 3 in the major groove is indicated by a black box, of helices 1 and 2 by a shaded box, and of the NH_2-terminus of the HD arm by N.

pairs prevents either full-length Skn-1 or the Skn domain from binding (Fig. 5B, lanes 10 and 11), suggesting that these bases are contacted by the proteins. In contrast, I·C substitution at these positions allows binding of either protein at an affinity comparable to that for SK1 (Fig. 5B, lanes 6 and 7), thus demonstrating that Skn-1 binds to the AT-rich portion of its site in the minor groove. It is interesting that the Δ1–9 Skn domain mutant, which lacks the HD arm and does not appear to bind the SK1 site (Fig. 4B, lane 4; Fig. 5B, lane 4), binds with barely detectable and similar affinities to the −3,−2,−1 GC and −3,−2,−1 IC substituted sites (Fig. 5B, lanes 8 and 12). Removal of the HD arm thus appears to have changed the binding sequence preferences of the Skn domain so that it no longer discriminates against G·C base pairs at these upstream positions, suggesting that these upstream minor groove interactions are mediated by the HD arm (*29*). Consistent with this hypothesis, the Δ1–9 Skn domain mutant appears to bind the −3,−2,−1 GC site with a higher affinity than does the intact Skn domain (Fig. 5B, lanes 10–12), suggesting that presence of the HD arm in intact Skn-1 may clash with the extra groups that G·C base pairs present in the upstream minor groove.

Substitution of I·C for the A·T pairs at +1 and +3 within the bZIP half-site prevents Skn-1 binding (Fig. 5B, lanes 14–16), suggesting a requirement for major groove interactions. We have assayed major groove binding directly by substituting T residues with uracil (U), which results in loss of a methyl group that is often utilized in base-specific interactions in the major groove. For example, conserved alanine residues in the GCN4 BR interact with T methyl groups at the +1 and +3 positions (*5*), at which U substitutions inhibit binding by either Fos/Jun or GCN4 complexes (*30*). In the SK1 site, T-to-U substitution at either of these positions similarly prevents binding by Skn-1 or the Skn domain (Fig. 5C, lanes 2, 3, 14, and 15), but at −1 or −2 this substitution has no effect (Fig. 5C, lanes 6, 7, 10, and 11). These data indicate that major groove interactions are essential for binding of Skn-1 to DNA, and are limited to the bZIP half-site. Full-length Skn-1 and the Skn domain alone have the same pattern of sequence discrimination among all of the I·C, G·C, and U swaps (Figs. 5, B and C), indicating that all of these sequence specificities are accounted for by the 85-residue Skn domain.

Participation of the entire Skn domain in the protein-DNA complex. To explore how the Skn domain interacts with DNA, we first performed a hydroxyl radical interference assay. Hydroxyl radical attack destroys the DNA backbone sugar residue,

and results in the loss of the corresponding nucleotide (*31*). This assay thus gives a measure of the relative contribution of individual nucleotides to the total binding energy. Hydroxyl radical interference with Skn-1 binding is generally more prominent along the top strand, and occurs wherever sequence preferences were observed except at −3 (Fig. 6A), suggesting that this particular sequence preference does not derive from direct contact. Interference that is apparent between +5 and +7, particularly on the bottom strand, is not correlated with distinct base preferences (Fig. 6A) and thus might be derived from contacts with the phosphate backbone. As would be expected if Skn-1 DNA binding derives entirely from the Skn domain, a Skn-1 deletion mutant consisting of the COOH-terminal 103 amino acids of Skn-1 gives a pattern of interference identical to that of GST–Skn-1 (*15*).

Contacts with DNA backbone phosphates add to the overall binding energy of protein-DNA interactions, and stabilize and orient elements that recognize specific base pairs (*32*). To investigate how Skn-1 might contact DNA phosphates, we assayed the effect of phosphate ethylation on its DNA binding (*17*). Ethylation interference with Skn-1 DNA binding is greatest between +3 and +5 along the bottom strand, and between −1 and +2 on the top strand (Fig. 6B). These particular implied contacts correlate with those made by the GCN4 BR, but ethylation interference is also evident at other positions (Fig. 6, B and D). In general, interference occurs along one face of the DNA helix (Fig. 6D), suggesting that Skn-1 projects its recognition elements into the major and minor grooves from one face of the binding site.

One of the most informative indicators of how Skn-1 forms a complex with DNA is the pattern by which it protects DNA from hydroxyl radical attack. Because hydroxyl radicals are very small, protection from them requires an intimate association with the protein, and can derive either from residues that lie in the minor groove or that otherwise shield the backbone sugars (*31*). A hydroxyl radical protection "footprint" of a bZIP protein is at best very weak and often undetectable (*33*), because the BR binds exclusively in the major groove (*5*). In contrast, Skn-1 gives a marked hydroxyl radical footprint (Fig. 6C), and protects the DNA on both sides of the minor groove in the AT-rich region at −1 and −2, thus providing support for binding in the minor groove at these positions (Fig. 6, C and D). Remarkably, Skn-1 also protects both sides of the major groove where the BR is predicted to bind (Fig. 6, C and D), suggesting that Skn-1 elements other than the BR must lie close to the bZIP half-site portion of its recognition sequence. In similar studies, the homeodomain protein α2 was found to protect both sites of the minor groove through interactions by its HD arm, and to protect both sides of the adjacent major groove by presence of helices 1 and 2 (and the intervening loop) above the backbone where helix 3 is bound (Fig. 6D). This footprint pattern appears to be characteristic of homeodomain proteins although the exact number of positions protected may vary (*34*). The Skn-1 footprint appears to cover a somewhat smaller DNA surface than that of α2, but the similarity in the overall pattern of these footprints is striking (Fig. 6D). This similarity suggests that, although the Skn-1 amino acids in the segment between the HD arm and the BR do not appear to be related to the HD, they may form an analogous structure that lies close to the DNA.

An analogy with the homeodomain would also predict that Skn domain residues between its BR and HD arm, which we refer to as its internal residues, are critical for stability of the overall protein-DNA complex. Structure-prediction algorithms indicate helical character within two regions of the Skn domain internal residues (Fig. 4C). Substitution of a proline residue into either of these putative helices, which would impair α helix formation, eliminated DNA binding (Fig. 4B, lanes 6 and 7), as did spanning them with alanine residues (Fig. 4B, lane 5), which would be predicted to favor formation of one continuous α helix. These findings demonstrate that the Skn domain internal residues are essential for DNA binding, and suggest that they may form a substructure that orients and stabilizes the BR, in much the same way that homeodomain helices 1 and 2 orient and stabilize helix 3, the DNA recognition helix.

Function of Skn-1 as a sequence-specific DNA binding protein. We have established that Skn-1 is a sequence-specific DNA binding protein and that its DNA binding capability is localized to the residues that we have designated the Skn domain. During the earliest stages of *C. elegans* embryogenesis, expression of maternal Skn-1 protein is required for the subsequent correct specification of multiple different cell types (*1*). Our findings suggest that Skn-1 performs these various functions by regulating transcription of downstream genes. Supporting this idea, it appears that the DNA binding capability of Skn-1 is required in vivo, because a previously reported *skn-1* mutation (zu135) (*1*) is a termination codon at residue 28 of the Skn domain (Fig. 4C) (*35*). Furthermore, the Skn-1 residues that lie NH_2-terminal to the Skn domain can act as a powerful transcriptional activator in mammalian cell transfection assays (*36*).

The consensus preferred Skn-1 binding site is a single bZIP half-site and an immediately adjacent 5′ AT-rich sequence (Fig. 2). Determination of this consensus should aid in the identification of genes that are regulated by Skn-1. An interesting potential candidate is the *C. elegans* homolog of MyoD, *hlh-1* (*37*). The *hlh-1* protein is detectable transiently in the daughters of an embryonic cell that contains Skn-1 protein (*3*, *37*, *38*). This early *hlh-1* expression involves two functionally redundant regions 5′ of the *hlh-1* coding sequences; one of these regions contains multiple consensus Skn-1 binding sites (*39*). Variations from the preferred Skn-1

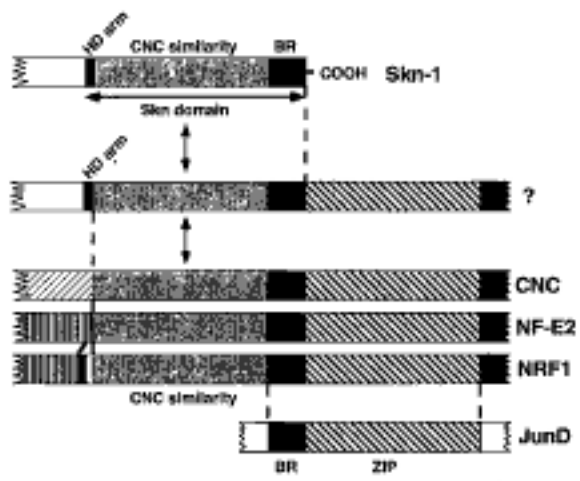


Fig. 7. Comparison of the Skn domain with bZIP protein sequences. This diagram is based on BLAST homology searches and comparison of known sequences (Fig. 4C). Members of the CNC-related bZIP protein subgroup (CNC, NF-E2, and NRF1) share the CNC similarity sequences, and are also related to each other throughout their bZIP motifs and in sequences immediately COOH-terminal to them (Fig. 6) (*45*, *46*, *52*). Their closest other bZIP relatives are the Jun proteins, to which they are related only in the bZIP domain. The Skn domain and the CNC-related bZIP proteins diverge from each other precisely at the junction with the Skn-1 HD arm (Fig. 4C), suggesting conservation of the CNC similarity sequences (which constitute the internal Skn domain sequences) as a functional unit. The upstream regions of Skn-1 do not suggest the origin of its HD arm. Skn-1 and the CNC group could have evolved from a common precursor (indicated as ?) which had both a bZIP domain and an HD arm, and could bind DNA as a dimer. The HD-ZIP proteins identified in *Arabidopsis* are partially analogous to such a precursor; they contain a complete homeodomain which lies immediately COOH-terminal to a leucine zipper, and they bind to DNA as dimers (*54*). Alternatively, a CNC-related bZIP protein might have acquired an HD arm, eventually allowing both loss of the leucine zipper and binding to DNA as a monomer. Identical shadings at corresponding sequences indicate related amino acid sequences. Points of divergence are indicated by dashed lines, except among the CNC-related bZIP proteins, where solid lines indicate divergence.

binding site might also be encountered. For example, the Skn domain can bind as a monomer to consensus palindromic CRE (ATGACGTCAT) and AP1 (ATGACTCAT) bZIP recognition sites with affinities that are two and ten times, respectively, lower than its affinity for the SK1 site (*20*), suggesting the possibility that Skn-1 might compete with bZIP proteins for binding to some target DNA sites in vivo. The Skn-1 preferred binding site is also similar to one (TAAGTCA) recognized by certain *fork head*/HNF-3 proteins (*40*). These proteins have structures similar to homeodomains (*25*). Interestingly, HNF-3 binds to GTCA in the major groove with an α helix distinct in sequence from the Skn-1 BR (*25*). It also recognizes the AT-rich sequence in the minor groove (*25*) and, in principle, it could share binding sites with Skn-1–related proteins.

The Skn domain and binding of its BR to DNA. The results demonstrate that a Skn domain BR segment, which would not bind well to its cognate half-site as a peptide (*7*), is nevertheless capable of high affinity DNA binding as a part of the Skn domain. The lack of any evidence that the Skn domain can form a dimer, and its recognition of a single bZIP half-site within its preferred binding sequence, indicate that the Skn domain binds to DNA as a monomer. Our thinking about how it might do so has relied heavily on comparisons with the homeodomain. In the homeodomain helix 3, the recognition helix, is precisely oriented and stabilized for binding in the major groove by its interactions with helices 1 and 2, and by binding of the HD arm in the adjacent minor groove (*10*, *22*). The framework of the Skn domain internal sequences, together with its own NH_2-terminal arm, seem to perform a similar function (Figs. 4, 5, 6C, and 6D), stabilizing and orienting the BR as an α helix in the major groove.

Despite these similarities, the Skn domain is clearly different from the standard homeodomain as well as from several related DNA binding domains, including those of the *fork head*/HNF-3 family (*25*), the *hin*-related prokaryotic DNA-invertases (*26*), and histones H1 and H5 (*41*). All of those DNA binding domains are based on globular three-helix bundles, within which helices 2 and 3 correspond to a helix-turn-helix motif, a substructure present within various prokaryotic and eukaryotic DNA binding domains (*42*). In contrast, the internal Skn domain sequence does not resemble any of the homeodomain helices. The internal Skn domain sequences might form a classical homeodomain structure but fail to conform to standard sequence alignments. This is unlikely since there are certain characteristic residues that are nearly invariant in all homeodomain proteins (*27*). Alternatively, the internal Skn domain might be an independent way that has evolved to orient and stabilize its own recognition helix (the BR). We have attempted to test whether the internal Skn domain residues might function as the homeodomain helices 1 and 2, by substituting this region of Antp into the Skn domain. This fusion protein fails to bind well to a Skn-1 site (*15*), but this result is not surprising given some uncertainty in knowing where to join these motifs, and the probable requirement for specific interactions between the bound BR and the internal Skn domain sequences.

The internal Skn domain sequences are, however, related to the corresponding amino acids within the CNC-related bZIP proteins (CNC, NF-E2, and NRF1; Figs. 4C and 7), suggesting that these sequences might all form similar structures. Figure 7 shows a possible evolutionary scheme for divergence of the Skn-1 and CNC-related bZIP protein families from a common precursor. On the basis of the known structure of the homeodomain and the bZIP proteins, it is possible that a protein such as the proposed progenitor in Fig. 7 can accommodate simultaneously both a ZIP segment, which is adjacent to the BR α helix, and a helix 1:helix 2–type "orientation domain" that sits on top of the recognition helix. In recognizing the outside surface of the α-helical BR, this CNC homology region might be analogous to E1A, HMGI/Y, and glucocorticoid receptor proteins, which appear to interact with bZIP domains and may contact particular BR residues (*43*). It would also resemble the so-called "recognition element" that seems to be required for the transcriptional activation function of the MyoD BR (*44*). However, in each of these examples, binding to the BR recognition helix is intermolecular, while for the Skn-1 internal residues and for helices 1 and 2 of the homeodomain, this interaction is intramolecular.

Thus, the Skn domain defines a novel class of DNA binding proteins. It is formed from subdomains that are also present in protein families which are distinct from each other, and utilize different strategies to bind DNA. In the Skn domain the BR, the internal CNC-related sequences, and the HD arm are used in contexts that are not the same as those in which they are commonly found, thus providing the first demonstration of an underlying sub-modularity in these DNA binding domains. Evolution thus appears to have been capable of deriving distinctly different solutions to the problem of how to stabilize and precisely orient a BR segment for making contacts with a specific DNA sequence.

REFERENCES AND NOTES

1. B. Bowerman, B. A. Eaton, J. R. Priess, *Cell* **68**, 1061 (1992).
2. C. C. Mello, B. W. Draper, M. Krause, H. Weintraub, J. R. Priess, *ibid.* **70**, 163 (1992).
3. B. Bowerman, B. W. Draper, C. Mello, J. Priess, *ibid.* **74**, 443 (1993).
4. W. H. Landschulz, P. F. Johnson, S. L. McKnight, *Science* **240**, 1759 (1988).
5. T. E. Ellenberger, C. J. Brandl, K. Struhl, S. C. Harrison, *Cell* **71**, 1223 (1992).
6. E. K. O'Shea, R. Rutkowski, P. S. Kim, *Science* **243**, 538 (1989); E. K. O'Shea, J. D. Klemm, P. S. Kim, T. Alber, *ibid.* **254**, 539 (1991).
7. R. Talanian, C. J. McKnight, P. S. Kim, *ibid.* **249**, 769 (1990).
8. K. T. O'Neil, R. H. Hoess, W. F. DeGrado, *ibid.*, p. 774; L. Patel, C. Abate, T. Curran, *Nature* **347**, 572 (1990); K. T. O'Neil, J. D. Shuman, C. Ampe, W. F. DeGrado, *Biochemistry* **30**, 9030 (1991).
9. M. A. Weiss *et al.*, *Nature* **347**, 575 (1990).
10. G. Otting *et al.*, *EMBO J.* **9**, 3085 (1990).
11. T. K. Blackwell and H. Weintraub, *Science* **250**, 1104 (1990).
12. H. J. Thiesen and C. Bach, *Nucleic Acids Res.* **18**, 3203 (1990); C. Tuerk and L. Gold, *Science* **249**, 505 (1990); A. D. Ellington and J. W. Szostak, *Nature* **346**, 818 (1990).
13. DNA sequences that were bound by a GST–Skn-1 fusion protein (*3*) were isolated from an oligonucleotide library (D9) in which 35 bp of random sequence DNA were flanked by fixed sequences that permitted amplification by the polymerase chain reaction (PCR) (*55*). The SAAB technique was performed essentially as described, except that dithiothreitol (DTT) was present in DNA binding reactions at 5 mM. In the first selection round, the sample was subjected to electrophoresis so that the free D9 template migrated about 1.5 cm into the EMSA gel, then the top 0.25 cm of the gel (including a visible band) were excised for DNA recovery. In subsequent rounds, a single visible band was excised. Throughout this procedure, only a small portion of the input DNA was bound, giving sequential selection for high-affinity Skn-1 binding sequences. The GST–Skn-1 protein was present at 1×10^{-7} M during the first two selection rounds and at 2.5×10^{-8} to 5×10^{-8} M during subsequent rounds, each of which contained DNA at 1×10^{-10} to 5×10^{-10} M. The GST–Skn-1 fusion protein (but not GST alone) bound to a greater fraction of the sequences derived from five rounds of selection and amplification than of the starting D9 library (*15*), indicating enrichment for Skn-1 binding sites (*11*). These selected molecules were then cloned and sequenced as described (*55*). A list of individual selected sequences is available from the author on request.
14. A. R. Oliphant, C. J. Brandl, K. Struhl, *Mol. Cell. Biol.* **9**, 2944 (1989); Y. Nakabeppu, and D. Nathans, *EMBO J.* **8**, 3833 (1989); J. W. Sellers, A. C. Vincent, K. Struhl, *Mol. Cell. Biol.* **10**, 5077 (1990).
15. T. K. Blackwell, B. Bowerman, J. R. Priess, H. Weintraub, unpublished data.
16. Of 16 selected DNA molecules examined by EMSA for binding by GST–Skn-1 protein as described (*55*), the 13 that included a GTCAT or ATCAT sequence (or varied from it at a single base) were bound at a level comparable to the selected population average (*15*). Those that did not were bound at a weak or undetectable level (*15*), and presumably were derived from nonspecific binding. Five selected Skn-1 binding sites were examined in the methylation interference assay (*15*).
17. A. Wissmann and W. Hillen, *Methods Enzymol.* **208**, 365 (1991).
18. Analysis of Skn-1 deletion mutants (*15*) suggests that the faster migrating minor Skn-1 complexes (Fig. 4B, lane 2) derive from internal initiation of translation.
19. I. A. Hope and K. Struhl, *EMBO J.* **6**, 2781 (1987).
20. T. K. Blackwell, unpublished data.
21. M. Affolter, A. Percival-Smith, M. Muller, W. Leupin, W. J. Gehring, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4093 (1990).
22. C. R. Kissinger, B. Liu, E. Martin-Blanco, T. B. Korn-

berg, C. Pabo, *Cell* **63**, 579 (1990); C. Wolberger, A. K. Vershon, B. Liu, A. D. Johnson, C. O. Pabo, *ibid.* **67**, 517 (1991).
23. A. K. Aggarwal, D. W. Rodgers, M. Drottar, M. Ptashne, S. C. Harrison, *Science* **242**, 899 (1988).
24. M. E. A. Churchill and A. A. Travers, *Trends Biochem. Sci.* **16**, 92 (1991).
25. K. L. Clark, E. D. Halay, E. Lai, S. K. Burley, *Nature* **364**, 412 (1993).
26. J.-A. Feng, R. C. Johnson, R. E. Dickerson, *Science* **263**, 348 (1994).
27. M. P. Scott, J. W. Tamkun, G. W. I. Hartzell, *Biochim. Biophys. Acta* **989**, 25 (1989).
28. Y. Q. Qian *et al.*, *Cell* **59**, 573 (1989); C. L. Phillips, A. K. Vershon, A. D. Johnson, F. W. Dahlquist, *Genes Dev.* **5**, 764 (1991).
29. It is possible that the Skn domain also makes other minor groove interactions, as is suggested by the A methylation interference observed at +1 and +4 (Fig. 3). A second minor groove interaction occurs in the *hin* DNA binding domain, which resembles a homeodomain (as described), but in which amino acid residues just COOH-terminal to helix 3 contact bases that are also bound in the major groove by helix 3 residues (*26*). Skn-1 contains eight amino acid residues COOH-terminal to the BR (as defined in bZIP proteins, Fig. 1) which could potentially mediate such interactions, but this particular Skn-1 methylation interference could also be derived from steric effects (*17*).
30. G. Risse, K. Jooss, M. Neuberg, H.-J. Bruller, R. Muller, *EMBO J.* **8**, 3825 (1989); W. T. Pu and K. Struhl, *Nucleic Acids Res.* **20**, 771 (1992); E. E. Blatter, Y. Ebright, R. H. Ebright, *Nature* **359**, 650 (1992).
31. T. D. Tullius and B. A. Dombrowski, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5469 (1986).
32. S. C. Harrison, *Nature* **353**, 715 (1991); C. O. Pabo and R. T. Sauer, *Annu. Rev. Biochem.* **61**, 1053 (1992).
33. D. E. Hill, I. A. Hope, J. P. Macke, K. Struhl, *Science* **234**, 451 (1986); C. R. Vinson, P. B. Sigler, S. L. McKnight, *ibid.* **246**, 911 (1989); M. R. Gartenberg, C. Ampe, T. A. Steitz, D. M. Crothers, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6034 (1990).
34. C. Goutte and A. D. Johnson, *EMBO J.* **13**, 1434 (1994).
35. C. Schubert, B. Bowerman, J. Priess, unpublished data.
36. T. K. Blackwell and H. Weintraub, unpublished data.
37. M. Krause, A. Fire, S. W. Harrison, J. Priess, H. Weintraub, *Cell* **63**, 907 (1990).
38. L. Chen, M. Krause, B. Draper, H. Weintraub, A. Fire, *Science* **256**, 240 (1992).
39. M. Krause, S. Harrison, S. Xu, L. Chen, A. Fire, *Dev. Biol.*, in press.
40. R. H. Costa, D. R. Grayson, J. E. Darnell Jr., *Mol. Cell. Biol.* **9**, 1415 (1989); E. Lai *et al.*, *Genes Dev.* **4**, 1427 (1990); D. G. Overdier, A. Porcella, R. H. Costa, *Mol. Cell. Biol.* **14**, 2755 (1994).
41. M. Suzuki, *EMBO J.* **8**, 797 (1989); V. Ramakrishnan, J. T. Finch, V. Graziano, P. L. Lee, R. M. Sweet, *Nature* **362**, 219 (1993).
42. S. C. Harrison and A. K. Aggarwal, *Annu. Rev. Biochem.* **59**, 933 (1990).
43. J. N. Miner and K. R. Yamamoto, *Genes Dev.* **6**, 2491 (1992); D. Thanos, W. Du, T. Maniatis, *Cold Spring Harbor Symp. Quant. Biol.* **58**, 73 (1993); F. Liu and M. R. Green, *Nature* **368**, 520 (1994).
44. H. Weintraub *et al.*, *Genes Dev.* **5**, 1377 (1991).
45. J. Mohler, K. Vani, S. Leung, A. Epstein, *Mech. Dev.* **34**, 3 (1990).
46. N. C. Andrews, H. Erdjument-Bromage, M. B. Davidson, P. Tempst, S. H. Orkin, *Nature* **362**, 722 (1993).
47. F. Van Streeten, R. Muller, T. Curran, C. Van Beveren, I. M. Verma, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3183 (1983).
48. K. Ryder, A. Lanahan, E. Perez-Albuerne, D. Nathans, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 1500 (1989).
49. In binding interference and protection experiments, a Skn-1 binding site selected in the SAAB assay was amplified in a PCR reaction in which one primer was end-labeled with ^{32}P, as described (*55*). In all these experiments, bound and free DNA fractions were isolated from dried preparative EMSA gels, in which less than 30% of the input DNA was in the bound fraction, and then analyzed on 10% sequencing gels (*55*). The methylation interference assay was performed as in (*55*), except that DNA was cleaved at both G and A residues as in (*56*). In the hydroxyl radical interference assay, end-labeled DNA was cleaved with hydroxyl radicals as described (*31*, *57*) prior to EMSA of Skn-1 binding. In the hydroxyl radical protection footprint, DNA was first incubated with GST–Skn-1 fusion protein for 10 min before the addition of cleavage reagents (*31*, *57*). After cleavage for 10 min, the sample was placed on an EMSA gel. Ethylation interference assays were performed as described (*17*), with PCR primers which expanded the template size to 124 bp (*55*), permitting better resolution of the bottom strand on the sequencing gel. Results obtained with Phosphorimaging (Molecular Dynamics) were analyzed using the program Excel (Microsoft).
50. Skn-1 and its derivatives were expressed by in vitro transcription from Bluescript-based vectors and translation in reticulocyte lysates (Promega), and quantitated (*55*). The ΔS mutant was constructed by restriction fragment deletion, and the Skn domain and Δ9 mutants by the PCR, so that initiation methionines were immediately followed by the Skn-1 residues indicated in Fig. 4C. Nucleotide sequences were confirmed within the amplified regions. Skn domain substitutions were produced by oligonucleotide-directed mutagenesis (Amersham). DNA binding assays contained these various in vitro translated Skn-1 or mutant proteins at concentrations between 0.3×10^{-10} and 2×10^{-10} M, and a DNA concentration of 1×10^{-10} to 5×10^{-10} M, and were performed with ^{32}P-labeled probes essentially as in (*55*). The SK1 oligonucleotide probe consisted of the double-stranded sequence: 5′-CTTATTGTCATCCTTATG, with 5′-GATC overhangs on each end. The various mutant derivatives of this sequence described in Fig. 5A differed from it only at the indicated bases.
51. B. Bowerman and J. Priess, unpublished data.
52. J. Y. Chan, X.-L. Han, Y. W. Kan, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11371 (1993).
53. R. T. Sauer, D. L. Smith, A. D. Johnson, *Genes Dev.* **2**, 807 (1988).
54. I. Ruberti, G. Sessa, S. Lucchetti, G. Morelli, *EMBO J.* **10**, 1787 (1991); G. Sessa, G. Morelli, I. Ruberti, *ibid.* **12**, 3507 (1993).
55. T. K. Blackwell *et al.*, *Mol. Cell. Biol.* **13**, 5216 (1993).
56. D. K. Lee, M. Horikoshi, R. G. Roeder, *Cell* **67**, 1241 (1991).
57. W. Dixon *et al.*, *Methods Enzymol.* **208**, 380 (1991).
58. We thank M. Scott, S. Henikoff, C. Goutte, and T. Ellenberger for helpful discussions; J. Mitchell for technical assistance; and T. Ellenberger for critical reading of the manuscript. Supported in part by a Burroughs Wellcome Fund fellowship of the Life Sciences Research Foundation during the initial stages of this work (T.K.B.); and by the Pathology training grant at the University of Washington and the Poncin foundation (T.K.B.); by a fellowship from the Jane Coffin Childs Medical Research Fund and an American Chemical Society grant (B.B.); by the Howard Hughes Medical Institute (H.W.); and by NIH grants (H.W. and J.R.P.).

23 June 1994; accepted 15 September 1994