Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 3414–3418, April 1992
Biochemistry

Dorsal–ventral patterning in *Drosophila*: DNA binding of snail protein to the single-minded gene

(transcriptional control/zinc fingers/central nervous system development)

YUMI KASAI, JOHN R. NAMBU, PAUL M. LIEBERMAN, AND STEPHEN T. CREWS*

Molecular Biology Institute, Department of Biology, University of California, Los Angeles, CA 90024

Communicated by H. Ronald Kaback, January 9, 1992

ABSTRACT **The *Drosophila* snail gene is required for proper mesodermal development. Genetic studies suggest that it functions by repressing adjacent ectodermal gene expression including that of the single-minded (*sim*) gene. The snail gene encodes a protein with a zinc-finger motif, and here we report that the snail gene product is a sequence-specific DNA binding protein. The snail protein recognizes a 14-base-pair consensus sequence that is found nine times in a 2.8-kilobase *sim* regulatory region. These results provide evidence for the direct control of *sim* transcription by snail.**

During development of the blastoderm embryo, activity of a group of maternal and zygotic genes that control dorsal–ventral patterning results in the formation of distinct cellular domains, including presumptive mesoderm, mesectoderm, neuroectoderm, dorsal–lateral ectoderm, and extraembryonic membranes (ref. 1; Fig. 1*A*). The mesectodermal cells derive from two single cell-wide stripes that lie between the presumptive mesoderm and neuroectoderm. The mesectodermal cells give rise to the neurons, glia, and other specialized cells that lie along the central nervous system (CNS) midline and are anatomically, developmentally, and functionally distinct from the lateral CNS (2, 3). The single-minded (*sim*) gene encodes a basic helix–loop–helix transcription factor that is required for proper development of the CNS midline lineage (2, 4, 5). The *sim* gene is initially transcribed near the end of the cellular blastoderm stage in the two stripes of mesectodermal precursor cells (ref. 2; Fig. 1*B*) and continues to be expressed in the CNS midline cells throughout embryogenesis (6). Because of its initial transcription in the blastoderm, expression of *sim* provides a useful indicator for mesectodermal specification.

Recent work has indicated that specification of the mesectodermal cells results from the action of genes that specify dorsal–ventral polarity in the embryo, including the dorsal, twist, and snail transcription factors (1, 4, 7–10). Both dorsal mutants and twist/snail double mutants result in the loss of *sim* expression, indicating their requirement for *sim* transcription (1). The roles of dorsal and twist are presently unclear, but one or both are likely to activate *sim* in the ventral region of the blastoderm (1). Both proteins are expressed as gradients along the ventral side of the embryo and are expressed in the mesectodermal nuclei (11–14). The effects of snail mutations on *sim* expression are particularly striking and result in ectopic ventral *sim* expression in cells that normally comprise the mesoderm (refs. 1 and 4; Fig. 1*D*). snail is expressed in the mesodermal anlage and is absent in detectable amounts from the adjacent mesectodermal cells (refs. 8, 10, and 15; Fig. 1*C*). These results are consistent with the idea that snail is a ventral repressor of *sim* and plays a role in establishing the ventral boundary of *sim* expression.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

The ventral expression of snail protein shortly precedes the appearance of *sim* transcripts. snail encodes a nuclear protein that contains five transcription factor IIIA-like zinc fingers and is likely to bind DNA (15, 16). The timing of snail's embryonic expression pattern and its zinc-finger motifs suggest that snail may directly control *sim* transcription. In this paper, we show that snail is a sequence-specific DNA binding protein and that it binds to multiple sites within *sim* regulatory DNA.

MATERIALS AND METHODS

Whole-Mount *in Situ* Hybridizations. Whole-mount *in situ* hybridizations were carried out with digoxigenin-labeled DNA probes according to the method of Tautz and Pfeifle (17). Probes used included full-length *sim* and snail cDNA clones and the *Escherichia coli lacZ* gene.

Preparation of snail Protein. A full-length snail cDNA clone was isolated from a *Drosophila* embryonic 3- to 12-hr λgt10 library constructed by L. Kauver and T. Kornberg (University of California, San Francisco). The snail protein was prepared by using the bacterial T7 expression system (18). An *Nco* I site was created at the snail translational start site by *in vitro* mutagenesis (19). A 1.2-kilobase (kb) *Nco* I/*Hin*dIII fragment, encoding the full-length snail protein, was cloned into pET-8C resulting in pET-snail. *E. coli* BL21 (DE3) containing either pET-8C or pET-snail was induced by the addition of 0.4 mM isopropyl β-D-thiogalactopyranoside. The cells were grown at 30°C for 2 hr and centrifuged; the pellet was resuspended in 1/50th vol of 40 mM Tris·HCl, pH 7.0/0.5 mM EDTA/1 mM dithiothreitol/10 μM $ZnCl_2$/1 mM phenylmethylsulfonyl fluoride. The suspensions were frozen in dry ice. The same volume of the above solution was added and suspensions were thawed. They were then treated with lysozyme (0.5 mg/ml) on ice for 15 min. Then 0.05% Nonidet P-40 was added, and incubation continued for 15 min. The solution was then brought up to 1 M NaCl and left on ice for 10 min, and the slurry was centrifuged at 35,000 rpm in a Beckman SW50.1 rotor for 1 hr at 4°C. Most of the snail protein was in the resulting pellet. The snail protein was renatured following a variation of the protocol described by Hager and Burgess (20). The protein pellet was resuspended in 5 M urea, allowed to sit at 20°C for several hours, and centrifuged at 35,000 rpm for 1 hr at 20°C. The urea in the supernatant was slowly removed by dialysis using decreasing concentrations of urea. The renatured proteins were finally dialyzed and stored in 20 mM Hepes, pH 7.9/100 mM KCl/10 μM $ZnCl_2$/20% (vol/vol) glycerol/1 mM dithiothreitol/1 mM phenylmethylsulfonyl fluoride. The size of the overexpressed snail protein was the predicted 43 kDa as determined by SDS/PAGE. The final concentrations of the extracts were determined by the Bradford assay (Bio-Rad).

DNA Binding Experiments. All DNA–protein binding reactions were performed in the presence of 12 mM Hepes, pH 7.9/100 mM KCl/10 μM $ZnCl_2$/12% glycerol/1 mM dithio-

Abbreviation: CNS, central nervous system.

*To whom reprint requests should be addressed.

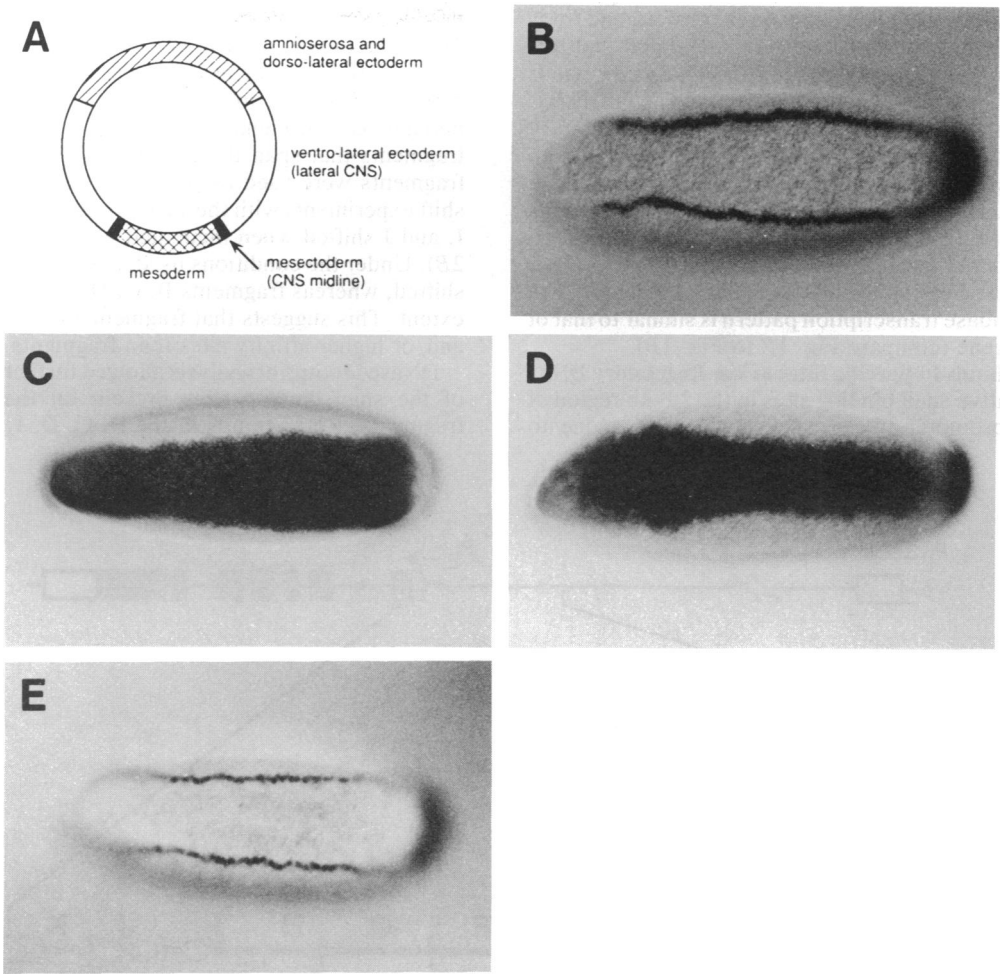

FIG. 1. (*A*) Schematic cross-section of a *Drosophila* embryo at the blastoderm stage of development showing four principle dorsal–ventral subdivisions. The CNS is derived from two distinct groups of cells that lie on both sides of the embryo: the mesectoderm and ventral–lateral ectoderm. The mesectodermal cells give rise to the neurons and glia that lie along the midline of the CNS. The ventral–lateral ectoderm gives rise to the lateral cells of the CNS and the ventral–lateral epidermis. (*B*) Whole-mount *in situ* hybridization of a *sim* cDNA probe to a wild-type cellular blastoderm embryo. Note the two single cell-wide stripes of hybridization representing the mesectodermal cells. (*C*) Whole-mount *in situ* hybridization of a snail cDNA probe to a wild-type cellular blastoderm embryo. Hybridization is restricted to the mesodermal precursor cells that lie between the mesectodermal cells. (*D*) Whole-mount *in situ* hybridization of a *sim* cDNA probe to a snail$^{4.26}$ homozygous mutant embryo. Note the expanded staining in most, but not all, cells that would normally give rise to the mesodermal precursor cells. (*E*) Whole-mount *in situ* hybridization of a β-galactosidase gene probe to a cellular blastoderm embryo containing P[*2.8sim/lacZ*]. Two stripes of hybridization appear in the location of the mesectodermal precursors at the same time that the native *sim* gene is initially transcribed. This indicates that the 2.8-kb *sim* genomic restriction fragment contains sequences sufficient for normal *sim* gene early promoter transcription. All embryos show ventral views with the anterior side on the left.

threitol/0.05% Nonidet P-40/bovine serum albumin (20 μg/ml)/poly(dI-dC)·poly(dI-dC) (100 μg/ml). Renatured snail protein or control extracts (0.2 μg) were preincubated at room temperature for 10 min. Then 2×10^3 dpm of ^{32}P-end-labeled restriction fragment was added, bringing the final reaction volume to 25 μl and the incubation continued for 30 min. DNA–protein complexes were subjected to gel electrophoresis and mobility-shift analysis (21). The reaction mixtures were immediately loaded onto a 4% polyacrylamide gel in 0.5× TBE (1× TBE = 90 mM Tris/64.6 mM boric acid/2.5 mM EDTA, pH 8.3) and electrophoresed at 25°C. Control experiments indicate that the mobility shifts were dependent on the presence of snail protein in the reaction mixture. Addition of the pET-8C extract or the Epstein–Barr virus Zta DNA-binding transcription factor did not shift the fragments. Double-stranded competitor oligonucleotides were used in the reactions to indicate specificity of snail binding and have the following sequences: the *sim* oligonucleotide is TCAGTTGCAAACAGGTGATTGCAGG, and the nonspecific oligonucleotide derived from the adenovirus E1B promoter is CTTAAAGGGTATATAATGCGCCGTG.

DNase I Footprinting Experiments (22). Binding of DNA to snail protein was carried out under the same conditions used for mobility-shift assays. In a 50-μl reaction volume, 1×10^4 dpm of ^{32}P-labeled restriction fragment was incubated with protein for 30 min. Five microliters of 10× DNase I buffer (100 mM $MgCl_2$/35 mM $CaCl_2$) containing 0.04 units of DNase I per μl (Promega) was added, and incubation continued for 1 min at 25°C. Reactions were terminated by the addition of 50 μl of stop buffer (40 mM EDTA/2% SDS/20 mM NaCl/0.2 mg of tRNA per ml) followed by extraction with phenol/chloroform, ethanol precipitation, and electrophoresis on a 7 M urea/6% polyacrylamide sequencing gel.

RESULTS

Germ-Line Transformation Experiments Define *sim* Regulatory DNA. The *sim* gene consists of eight exons scattered over 20 kilobases (kb) of DNA (ref. 4; S.T.C., J. Thomas, Y.K., M. G. Muralidhar, and J.R.N., unpublished data). It possesses two promoters: P_E drives transcription in the

cellular blastoderm through germ-band extension, and P_L is activated during germ-band extension and later controls midline glial *sim* expression (refs. 4 and 6; Fig. 2*A*). Previous germ-line transformation experiments, in which *sim* DNA is fused to *lacZ*, have shown that P_E and all of the regulatory elements necessary for early *sim* transcription are in a 7.8-kb *Bam*HI fragment (4). We now show that a 2.8-kb *sim* genomic fragment contains P_E and all of the necessary elements required for *sim* midline expression (Fig. 2*A*). This is demonstrated by *in situ* hybridization of a β-galactosidase DNA probe to embryos bearing the P[*2.8sim/lacZ*] element. The early β-galactosidase transcription pattern is similar to that of the native *sim* gene (compare Fig. 1*E* to Fig. 1*B*).

snail Protein Binds to Specific Sites in *sim* Regulatory DNA. To identify putative snail binding sites in the 2.8-kb region of the *sim* early promoter, we used a snail cDNA clone to produce snail protein by bacterial expression methods (18). This protein was used in electrophoretic mobility-shift and DNase I footprinting assays. The 2.8-kb *sim* genomic fragment, shown by germ-line transformation to contain the *sim* midline pattern elements, was digested into 11 restriction fragments that span the entire region (Fig. 2*A*). The DNA fragments were used as probes in electrophoretic mobility-shift experiments with the snail protein. Only probes B, C, D, I, and J shifted when incubated with the snail protein (Fig. 2*B*). Under the conditions used, almost all of fragment J was shifted, whereas fragments B, C, D, and I shifted to a lesser extent. This suggests that fragment J has more binding sites and/or higher-affinity sites than fragments B, C, D, and I.

DNase I footprint analysis allowed further characterization of the snail binding sites present on the five restriction fragments. Both strands of the B, C, D, I, and J fragments

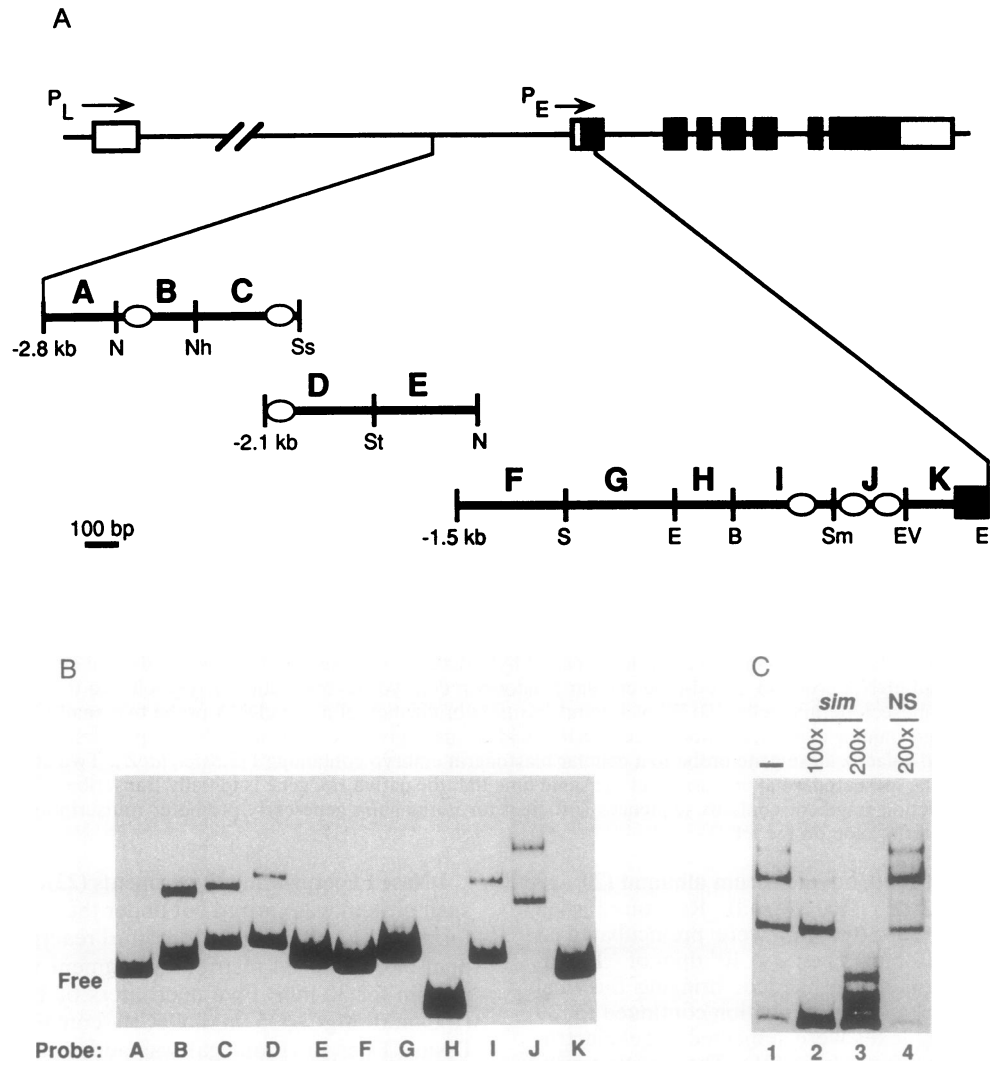

FIG. 2. The snail protein binds at multiple sites within the *sim* early promoter region. (*A*) Schematic representation of the *sim* gene structure and probes used for assays. Top line depicts genomic structure of the *sim* gene, showing the eight exons and two promoters (P_E and P_L). Solid boxes indicate coding regions and open boxes represent untranslated regions. Initial embryonic transcription of the *sim* gene occurs from P_E. The 2.8-kb *sim* genomic DNA shown by germ-line transformation to be sufficient to drive correct early *sim* transcription is shown below. Letters A–K above the partial restriction map indicate DNA fragments used in the mobility-shift and DNase I footprinting assays. The 5′ ends of A, D, and F are artificially introduced *Kpn* I sites not present in the *sim* gene. Ovals indicate the location of snail binding sites as determined by DNase I footprinting experiments. B, *Bam*HI; E, *Eco*RI; EV, *Eco*RV; N, *Nde* I; Nh, *Nhe* I; S, *Sal* I; Sm, *Sma* I; Ss, *Ssp* I; St, *Sty* I. (*B*) Mobility-shift analysis of snail protein binding to *sim* DNA. Full-length snail protein was incubated individually with each of the ^{32}P-end-labeled *sim* restriction fragments (lanes A–K). The protein/DNA mixtures were electrophoresed on a polyacrylamide gel and subjected to autoradiography. The specific probe present in the DNA binding reaction is indicated below each lane. (*C*) snail binding to DNA is sequence specific. ^{32}P-labeled restriction fragment J was incubated with snail protein and in the presence or absence of various competitor DNAs and was analyzed by mobility-shift assays. Lanes: 1, snail protein added without competitor DNA; 2 and 3, snail protein added with 100× or 200× molar excess of the *sim* oligonucleotide containing sites Sna-5a and Sna-5b; 4, snail protein added with 200× molar excess of an oligonucleotide containing a nonspecific (NS) DNA sequence.

were footprinted in the presence or absence of snail protein and the results are summarized in Fig. 3. In the presence of snail protein, two sites on fragment J (Sna-4 and Sna-5) are well protected from DNase I cleavage (Fig. 3*A*). Fragments B and I both contain a single site of protection (Sna-1 and Sna-3, respectively; data not shown) that is only weakly protected compared to those on fragment J. Fragments C and D share one weakly protected site (Sna-2) in the region where the two probes overlap (data not shown). The sequences that are protected on all five restriction fragments are shown in Fig. 3*B*. The Sna-5 site spans 47 base pairs (bp); the Sna-4 site spans 34 bp; and the Sna-3, Sna-2, and Sna-1 sites span approximately 33, 20, and 21 bp, respectively. Alignment of protected regions reveals a 14-bp consensus sequence (ANCACCTGTTNNCA) that is present once in Sna-1, Sna-2, and Sna-3; twice in Sna-4; and four times in Sna-5. The four regions in Sna-5 can be arranged as two inverted repeats. In each of the Sna-5 inverted repeats, the two conserved elements overlap one another by 10 bp, resulting in an 18-bp sequence with dyad symmetry. In Sna-4, the two conserved sites are in inverted orientation with respect to one another and are separated by 9 bp.

The specificity of snail protein for DNA binding sites is also shown by competition experiments. Binding of snail protein to fragment J was inhibited by the presence of a 200-fold molar excess of an oligonucleotide containing Sna-5a and Sna-5b binding sites (Fig. 2*C*, compare lanes 1 and 3). However, an equivalent amount of an oligonucleotide containing an unrelated DNA sequence failed to inhibit snail binding (Fig. 2*C*, lane 4).

DISCUSSION

The *Drosophila* snail protein has five zinc fingers of the transcription factor IIIA Cys_2-His_2 class (16); a homolog of

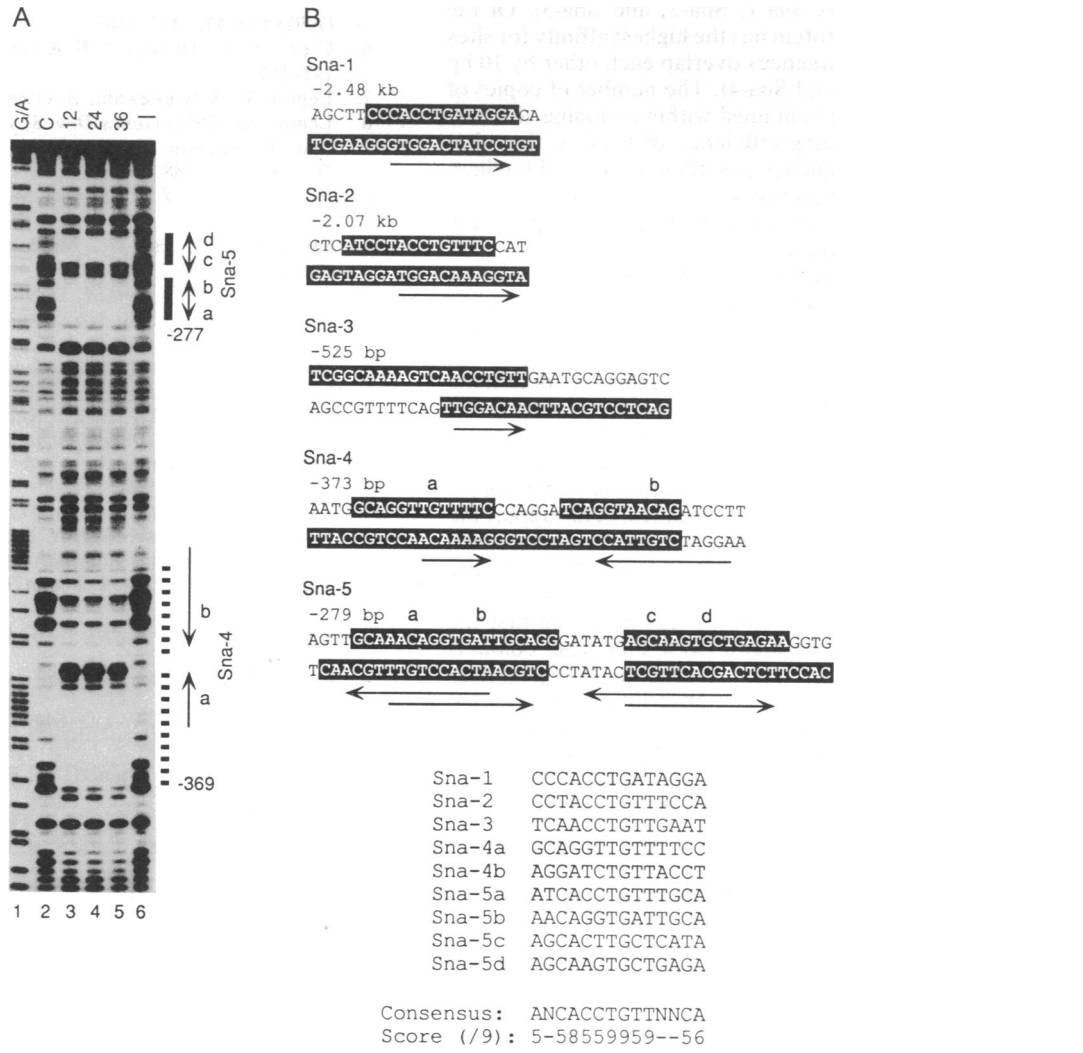

FIG. 3. DNase I footprint analysis indicates that snail recognizes a 14-bp consensus sequence. (*A*) DNase I footprints of binding sites for snail in the *sim* early promoter region. Fragment J was end-labeled on the noncoding strand with ^{32}P and incubated with increasing amounts of snail protein extract. Lane 3, 12 μg of extract; lane 4, 24 μg of extract; lane 5, 36 μg of extract. In lane 2, 12 μg of extract prepared from bacteria containing the control plasmid pET-8C was used. Lane 1 is a Maxam and Gilbert sequencing ladder. In lane 6, no bacterial extract was present. Regions of protection are indicated by the solid (strong protection) and dashed (weaker protection) lines shown on the right of the autoradiogram. Strongest protection was seen for the Sna-5 binding site. Vertical arrows indicate the orientations of the snail binding consensus sequences located at the protected sites (see *B*). Note the appearance of several DNase I hypersensitive sites located near the protected regions. (*B*) Consensus snail binding sequence in the *sim* early promoter region. Sequence analysis of the five snail binding regions indicates that there is a 14-bp consensus sequence found nine times. Sequences from the five protected regions (Sna-1, -2, -3, -4, and -5) are shown together with the consensus snail binding sequence below. The sequences protected from DNase I digestion are highlighted, and the extent and orientation of the consensus sites are indicated by arrows. The positions are relative to the start of the long *sim* open reading frame. Sna-1 and Sna-3 binding sites are located in restriction fragments B and I, respectively. The Sna-2 binding site is contained in both fragments C and D, and the remaining six binding sites are included within fragment J (see also Fig. 2*A*).

the snail protein has also been described in vertebrates (23). Our results provide direct evidence that the snail gene encodes a sequence-specific DNA binding protein and that this protein is likely to be a transcriptional regulator. Crystallographic studies on another zinc-finger protein, zif286, have indicated that each finger is likely to interact with three nucleotides within the major groove of the DNA helix (24). It is predicted that a snail protein monomer would bind a 15-bp sequence, a size similar to that obtained for the consensus sequence described here. The consensus sequence, however, does extend beyond the protected regions in several cases (Fig. 3*B*, Sna-1, Sna-2, and Sna-3). Furthermore, snail shows variability in the strength of binding to the different *sim* DNA sites as determined by the electrophoretic mobility-shift and DNase I protection assays. In the presence of snail protein, DNA sequences bearing inverted repeats of the consensus sequence are better protected from DNase I digestion than sites bearing only a single copy of the consensus sequence (Sna-4 and Sna-5 versus Sna-1, Sna-2, and Sna-3). Of the inverted repeats, snail protein has the highest affinity for sites where the consensus sequences overlap each other by 10 bp (Fig. 3, compare Sna-5 and Sna-4). The number of copies of the consensus sequence contained within a binding site thus correlates with the binding efficiency of snail to that site. While the consensus sequence is sufficient for snail binding, the exact sequences within the consensus, adjacent nucleotides, spacing of binding sites, and snail multimerization may play a role in binding affinity.

Genetic and biochemical evidence support the idea that snail is a direct repressor of *sim* transcription. In addition to *sim*, snail appears to repress ventral expression of the neuroectodermal genes achaete-scute T3, enhancer of split m7, and rhomboid (9, 10). It will be interesting to determine whether snail regulates transcription of these and other genes expressed in the neuroectoderm by binding DNA sequences similar to those found within the *sim* gene. The developmental role of zinc-finger proteins as transcriptional repressors is not unique to snail. Genetic and biochemical data support the roles of zinc-finger gap genes such as Krüppel and hunchback in repression of transcription (25–28). In these cases, the gap gene protein binding sites are interspersed among binding sites for transcriptional activators. It is interesting that the consensus snail binding site identified in this paper contains an E-box sequence, CACCTG, similar to that generally bound by basic helix–loop–helix transcription factors (29). The significance of this motif in the *sim* gene is unknown but allows speculation that basic helix–loop–helix transcriptional activators (e.g., twist) may activate *sim* in the mesodermal and mesectodermal anlage in the absence of snail but that snail may block binding and repress expression in the mesoderm. The exact biochemical nature of snail repression of *sim* transcription will be revealed by further biochemical, germ-line transformation, and genetic analyses.

We appreciate the comments of Al Courey, Judith Lengyel, and members of the laboratory of S.T.C. on the manuscript, and R. L. Cagan for assistance with the figures. We would also like to thank Maria Leptin for communicating her unpublished results. Y.K. is a postdoctoral trainee supported by Department of Health and Human Services National Institutional Research Service Award T 32 CA09056. J.R.N. is supported by a National Institutes of Health postdoctoral fellowship and P.M.L. is supported by an American Cancer Society postdoctoral fellowship. S.T.C. is a Lucille P. Markey Scholar, and this work was supported by a grant from the Lucille P. Markey Charitable Trust and National Institutes of Health Grant R01 HD25251.

1. Rushlow, C. & Arora, K. (1990) *Semin. Cell Biol.* **1**, 137–149.
2. Thomas, J. B., Crews, S. T. & Goodman, C. S. (1988) *Cell* **52**, 133–141.
3. Klämbt, C., Jacobs, J. R. & Goodman, C. S. (1991) *Cell* **64**, 801–815.
4. Nambu, J. R., Franks, R. G., Hu, S. & Crews, S. T. (1990) *Cell* **63**, 63–75.
5. Nambu, J. R., Lewis, J. O., Wharton, K. A. & Crews, S. T. (1991) *Cell* **67**, 1157–1167.
6. Crews, S. T., Thomas, J. B. & Goodman, C. S. (1988) *Cell* **52**, 143–151.
7. Leptin, M. & Grunewald, B. (1990) *Development* **110**, 73–84.
8. Leptin, M. (1991) *Genes Dev.* **5**, 1568–1576.
9. Rao, Y., Vaessin, H., Jan, L. Y. & Jan, Y.-N. (1991) *Genes Dev.* **5**, 1577–1588.
10. Kosman, D., Ip, Y. T., Levine, M. & Arora, K. (1991) *Science* **254**, 118–122.
11. Rushlow, C. A., Han, K., Manley, J. L. & Levine, M. (1989) *Cell* **59**, 1165–1177.
12. Steward, R. (1989) *Cell* **59**, 1179–1188.
13. Roth, S., Stein, D. & Nüsslein-Volhard, C. (1989) *Cell* **59**, 1189–1202.
14. Thisse, B., Stoetzel, C., Gorostiza-Thisse, C. & Perrin-Schmitt, F. (1988) *EMBO J.* **7**, 2175–2183.
15. Alberga, A., Boulay, J.-L., Kempe, E., Dennefeld, C. & Haenlin, M. (1991) *Development* **111**, 983–992.
16. Boulay, J. L., Dennefeld, C. & Alberga, A. (1987) *Nature (London)* **330**, 395–398.
17. Tautz, D. & Pfeifle, C. (1989) *Chromosoma* **98**, 81–85.
18. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) *Methods Enzymol.* **185**, 60–89.
19. Taylor, J. W., Ott, J. & Eckstein, F. (1985) *Nucleic Acids Res.* **13**, 8765–8785.
20. Hager, D. A. & Burgess, R. R. (1980) *Anal. Biochem.* **109**, 76–86.
21. Fried, M. & Crothers, D. M. (1981) *Nucleic Acids Res.* **9**, 6505–6525.
22. Galas, D. & Schmitz, A. (1978) *Nucleic Acids Res.* **5**, 3157–3170.
23. Sargent, M. G. & Bennett, M. F. (1990) *Development* **109**, 967–973.
24. Pavletich, N. P. & Pabo, C. O. (1991) *Science* **252**, 809–817.
25. Stanojevic, D., Hoey, T. & Levine, M. (1989) *Nature (London)* **341**, 331–335.
26. Hülskamp, M. & Tautz, D. (1991) *BioEssays* **13**, 261–268.
27. Pankratz, M. J. & Jäckle, H. (1990) *Trends Genet.* **6**, 287–292.
28. Carroll, S. (1990) *Cell* **60**, 9–16.
29. Murre, C., McCaw, P. & Baltimore, D. (1989) *Cell* **56**, 777–783.