Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 10311–10315, November 1992
Cell Biology

Basonuclin: A keratinocyte protein with multiple paired zinc fingers

HUNG TSENG AND HOWARD GREEN

Department of Cellular and Molecular Physiology, Harvard Medical School, Boston, MA 02115

Contributed by Howard Green, July 21, 1992

ABSTRACT **A cDNA clone has been prepared from mRNA of cultured human keratinocytes. The sequence of the cDNA reveals that in the C-terminal two-thirds of the corresponding protein (basonuclin), there are three separated pairs of adjacent zinc fingers. The amino acid sequence of each pair is homologous to that of the single pair of zinc fingers of the *Drosophila* transcription factor encoded by *disco*. Near the C-terminal end of basonuclin and on the surface of a putative α-helix, there is a stripe of serine residues similar to that of the transcription factor PRDII-BF1. Basonuclin possesses a sequence of six amino acids quite similar to one present in the myogenic family of proteins, including Myf5; this sequence is located in the Ω loop of the myogenic proteins but within a zinc finger of the keratinocyte protein. As basonuclin is present mainly in the nuclei of the basal cell layer, its regulatory function is likely to be exerted prior to the process of terminal differentiation.**

In eukaryotes, two structures that confer on proteins the ability to bind DNA in a sequence-specific manner are the zinc finger motif and the basic helix–loop–helix (HLH) motif (1). The protein TFIIIA, first identified as a transcription factor for RNA polymerase III in the laboratories of D. D. Brown and R. G. Roeder, was found to contain tandem 30-amino acid repeats that bind zinc (2, 3). Characteristically, a zinc finger contains pairs of cysteines and histidines arranged according to the formula: C-X_{2-4}-C-X_{12}-H-X_{3-4}-H (4, 5). First sequence analysis and then NMR and x-ray crystallographic studies have revealed that the zinc finger is a compact structure containing a DNA-contacting α-helix folded against a β-hairpin (5–8). The zinc finger motif possesses the same compact structure whether or not its recognition sequence is present (7).

The HLH proteins possess a basic DNA-binding region located N terminal to the two helices and their intervening loop. Binding of these proteins to DNA requires dimerization (9–12). Among proteins containing the basic-HLH motif are the members of the MyoD family of regulatory proteins specific to skeletal muscle. These proteins are able to induce myogenesis in various nonmyoblast cells (13–15). Both basic and HLH domains of MyoD are necessary for its initiation of myogenesis in such cells and for activation of transcription of muscle-specific genes. Thinking that the control of differentiation in the keratinocyte might also depend on proteins of basic-HLH structure, we examined keratinocyte cDNA for the presence of a sequence homologous to Myf5, a member of the MyoD family. Since among the known basic-HLH proteins the amino acid sequence is particularly conserved in the helical regions (16), we used a probe containing the HLH region of Myf5 to screen a cDNA library prepared from keratinocyte mRNA. One of the cDNA clones isolated in this way has now been sequenced.* The region of homology between this cDNA and Myf5 is very short, and the cDNA was found to encode, not the expected HLH protein, but a zinc finger protein of an unusual kind.

MATERIALS AND METHODS

cDNA Library Construction. Keratinocyte strain FRTS was obtained from the foreskin of a newborn. It was cultivated using supporting 3T3 cells (17) under conditions described earlier (18, 19).

Total RNA was isolated from the cultured keratinocytes by lysing the cells grown in 100-mm dishes with a solution containing 6 M guanidinium thiocyanate, 5 mM sodium citrate (pH 7.0), 0.1 M 2-mercaptoethanol, and 0.5% sarkosyl (20). The cell lysate was then overlaid on 2 ml of a solution containing 5.7 M CsCl and 0.1 M EDTA at pH 7.0, and centrifuged in a Beckman SW50.1 rotor for 30 hr. The RNA pellet was resuspended in a solution containing 1% sarkosyl and 5 mM EDTA, extracted twice with phenol/chloroform, and precipitated with 2.5 vol of ethanol. mRNA was purified using Stratagene's Poly(A)Quik kit. A cDNA library was constructed using BRL's Superscript cDNA synthesis and cloning kit. Approximately 1 μg of mRNA was used in the cDNA synthesis and the double-stranded cDNA was directionally cloned into the *Sal* I–*Not* I sites of the pSPORT-1 (BRL) vector. The final library consisted of ≈1.6 × 10^6 independent colonies with a recombinant frequency of about 95%.

Preparation of the Clone Used To Probe the Library. The partial cDNA clone Myf5/18-07, obtained from the ATCC, lacks the first 178 nucleotides of the full-length cDNA (21) but retains its basic-HLH motif. The entire insert of the clone was excised by *Eco*RI digestion and purified by agarose gel electrophoresis. The purified insert was labeled by the method of random primer extension (22).

Screening. The amplified cDNA library was plated at ≈2000–3000 colonies per 15-cm Petri dish. Colonies were lifted onto a nitrocellulose filter (Schleicher & Schuell), lysed, and denatured according to ref. 23. Hybridization was performed at 59–60°C for 16–18 hr in a buffer described previously (24). The filters were then washed at 60°C in a solution containing 1 M NaCl and 1% SDS. The colonies were stained with 1% methyl green (Fluka) (24) before exposure to x-ray film (X-Omat XAR 5; Kodak). The stain made it easier to align the colonies with the film image. Two hundred colonies that showed hybridization were picked and rapid DNA preparations were made from them. The inserts of these clones were excised by *Eco*RI and *Xba* I digestion, separated on 1% agarose gel, and subjected to Southern analysis with the Myf5 probe. Those colonies whose DNA hybridized to the Myf5 probe were purified by plating at low density. To map the sequence of Myf5 cDNA responsible for hybridization, purified inserts of clone Myf5/18-07 were digested with a panel of restriction enzymes (*Hin*cII, *Hin*fI,

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.

Abbreviation: HLH, helix–loop–helix, zinc-finger protein.

*The sequence reported in this paper has been deposited in the GenBank data base (accession no. L03427).

Nco I, *Pst* I, *Pvu* II, and *Sph* I), separated on a 1% agarose gel, and transferred to nitrocellulose filters. The filter was then probed with purified inserts of each of the keratinocyte cDNA clones. Clones that showed hybridization to the HLH region were further analyzed. In the reciprocal mapping, the region of the cDNA clones that hybridized to the Myf5 HLH motif was located by digestion of the inserts with various restriction enzymes and Southern analysis. The complementary fragments were then sequenced.

Sequencing. To sequence the full-length cDNA, unidirectional deletions were created from either end of the inserts of the two longest cDNA clones (M752A-63 and M752A-173) with Promega's Erase-a-Base kit. Clones of nested deletions were isolated and double-stranded DNA sequencing was done by the method of dideoxynucleotide termination using the Sequenase Version 2.0 DNA sequencing kit (United States Biochemical). Each nucleotide position in the reported sequence has been sequenced at least once in each direction.

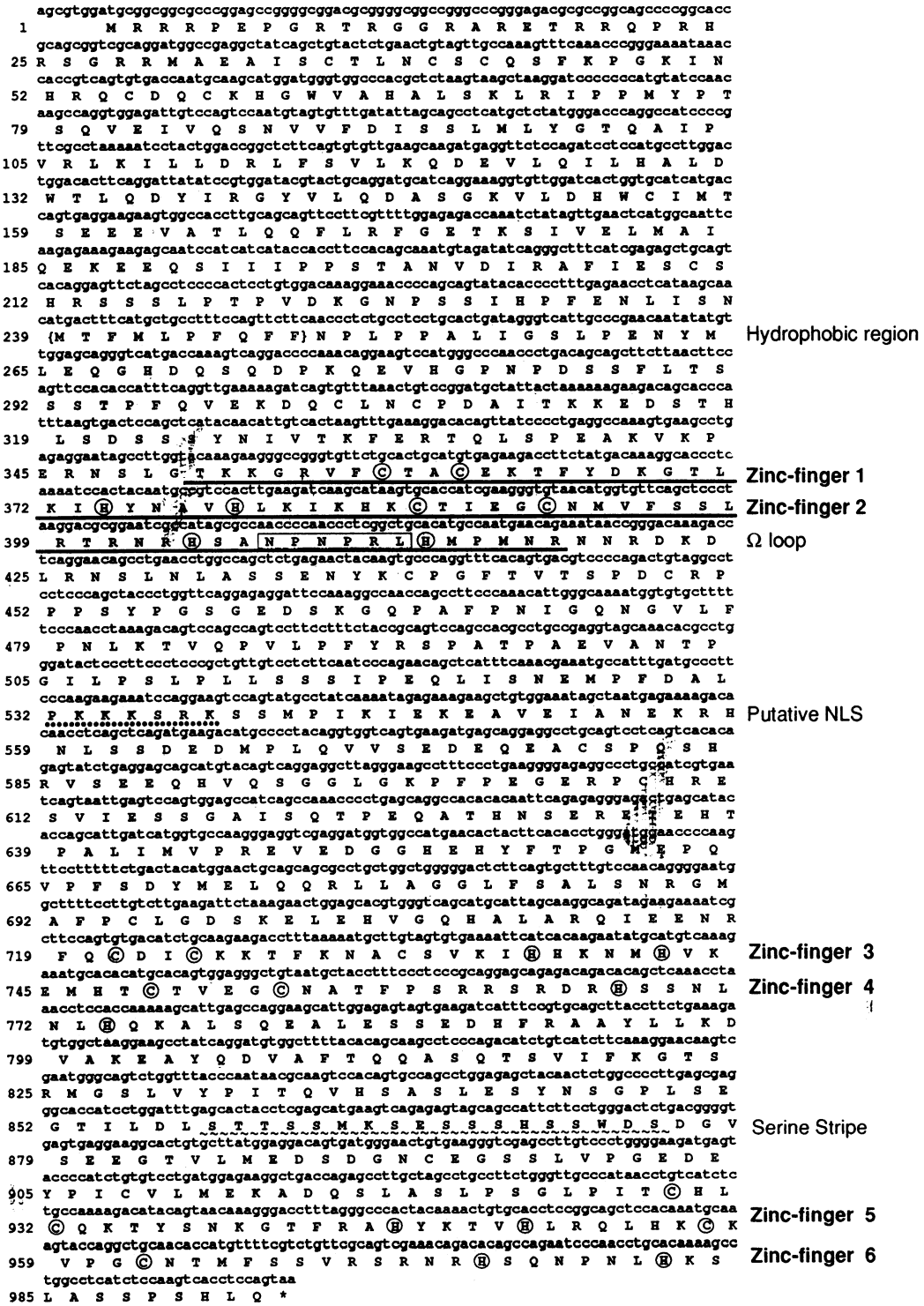

FIG. 1. Coding region of basonuclin. The principal features of the sequence are the cysteines and histidines of the six zinc fingers (circled), the region homologous to *disco* (underlined), a hydrophobic region (parentheses), the Ω loop (framed), the nuclear localization sequence (NLS, underdotted), and the serine stripe (underscalloped).

The region homologous to the Myf5 HLH motif was sequenced by reactions primed with oligonucleotides flanking that region (New England Biolabs).

RESULTS

Search for a Keratinocyte cDNA Homologous to Myf5. A cDNA library was constructed with mRNA isolated from cultured human keratinocytes. Two hundred thousand colonies from this library were screened with the Myf5 cDNA clone and after clonal purification, >20 clones were confirmed to be true positives.

The regions in Myf5 cDNA responsible for hybridization were then mapped for each positive clone. One clone, M752A, containing an insert of 3.8 kilobases (kb), was found to hybridize to a 266-base-pair (bp) *Pst* I restriction fragment of Myf5 containing its HLH region. As will be shown below, the nucleotide sequence of M752A complementary to Myf5 is confined to the loop of its HLH region and is only 26 nucleotides in length. We have given the name basonuclin to the keratinocyte protein encoded by M752A.

Nucleotide Sequence of Basonuclin. In Northern analysis of keratinocyte mRNA, the basonuclin sequence of 3.8 kb detected a single mRNA of 4.6 kb. Several cDNA clones with inserts of the same size as the mRNA were isolated from a more limited library prepared from size-fractionated inserts of the original library. When four such independent clones were sequenced, three had identical 5′ termini, ending with a guanine, suggesting that they might be the full-length cDNAs. The sequence shown in Fig. 1 was that obtained by sequencing of two of the full-length cDNAs and the original partial cDNA.

In the 4611 bp of the cDNA sequence, there is a single large open reading frame. If the first ATG is the translation initiator, the encoded polypeptide should contain 993 amino acids. However, the nucleotides neighboring that ATG should make it a weak initiator, whereas the ATG at codon 30 should be a strong initiator (25). We can therefore not exclude that some or all of the polypeptide molecules would lack the first 29 amino acids.

The polypeptide contains six zinc finger motifs in widely spaced pairs, the first pair beginning at residue 358 and the last at residue 957, close to the C terminus. A hydrophobic region extends between residues 239 and 248. At amino acid residue 532, between the second and third zinc fingers, there is a potential nuclear localization signal. This sequence, PKKKSRK, is in good agreement with the class A nuclear localization signal consensus sequence (26), in which there is a helix breaker, proline or glycine, followed by at least three basic amino acids in a hexa- or octapeptide. PKKKSRK is also consistent with the consensus X X K-R/K-X-R/KX, third X not asparagine (27).

Relatedness of the Zinc Fingers of Basonuclin to those of *disco*. When a 60-amino acid sequence of basonuclin, including zinc fingers one and two and encompassing the hexapeptide homologous to Myf5 (see below), was used to search the GenBank protein sequence data base (28), the most highly related sequence found was that of the *Drosophila* gene *disco* (disconnected) (29, 30). The *disco* gene is expressed in many different tissues, but its function is thought to be in the establishment of stable connections between the larval optic nerves and the brain; there is no obvious relation in function between *disco* and basonuclin.

Most of the nucleotide sequence of basonuclin has no homology to that of *disco*, but there is very strong similarity between the three pairs of zinc fingers of basonuclin and the single pair of zinc fingers of *disco*. Zinc fingers 1, 3, and 5 of basonuclin are homologous to the first zinc finger of *disco* and zinc fingers 2, 4, and 6 of basonuclin are homologous to the second zinc finger of *disco* (Fig. 2). The first group contains the characteristic sequence C-X_2-C-X_{12}-H-X_4-H (5) and the second group contains the sequence C-X_4-C-X_{12}-H-X_{6-8}-H.

In the first zinc fingers of basonuclin and *disco*, 13 of the 20 residues located between the first cysteine and the last histidine are identical (65%). Zinc fingers 3 and 5 of basonuclin share eight and seven residues, respectively, in common with the first zinc finger of *disco*.

In the corresponding span between the first cysteine and last histidine of the second zinc fingers of basonuclin and *disco*, 22 of 26 residues are identical (85%). In this region, 13 residues are identical in all three even-numbered fingers of basonuclin and in zinc finger 2 of *disco* as well. The X_{12} region between the second cysteine and the first histidine is known to contain amino acid residues that contact bases in DNA (5, 7, 8); in this region of the even-numbered zinc fingers of basonuclin and *disco*, arginine, a common base-contacting amino acid in other zinc finger proteins, is uniformly present

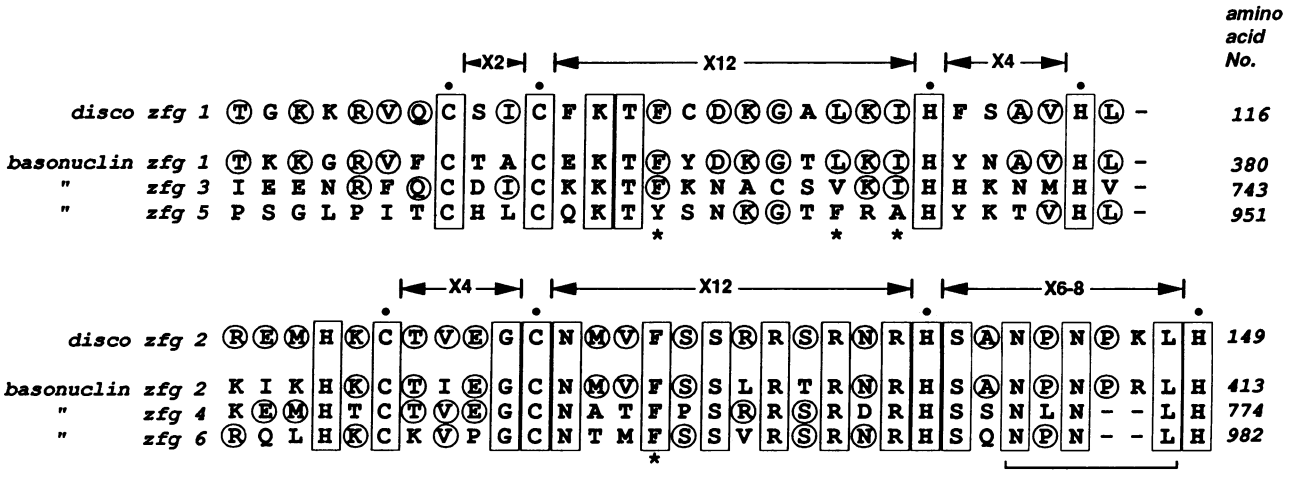

FIG. 2. Six zinc fingers of basonuclin. The zinc fingers are of two kinds and occur in pairs: the odd-numbered fingers are similar to zinc finger 1 of *disco*, and the even-numbered fingers are similar to zinc finger 2 of *disco*. All the fingers belong to the C_2-H_2 family (overdotted residues), but whereas the odd-numbered fingers have the typical four-amino acid residues between the Hs, the even-numbered fingers have six or eight. Residues identical in three fingers of basonuclin and one of *disco* are indicated by vertical boxes. Other amino acids shared by *disco* and one or more, but not all, corresponding fingers of basonuclin are circled. One or three positions in the X_{12} region are occupied by hydrophobic residues (indicated by *). In the even-numbered fingers there are three invariant arginine residues, possibly base-binding, three invariant asparagines, and two invariant serines. The bracketed six-amino acid sequences of zinc fingers 2 of basonuclin and *disco* are homologous to a six-amino acid sequence in the loop region of the myogenic proteins, which do not contain zinc fingers.

at three positions (Fig. 2). In two other positions, a serine and an asparagine are also invariant, and in the X_{6-8} region, two asparagines, a serine, and a leucine are invariant. All of the zinc fingers have an aromatic residue at the fourth position of the X_{12} region, but only the odd-numbered fingers have hydrophobic residues at positions 10 and 12.

We have isolated and partially sequenced the basonuclin gene of the mouse. In a region of 162 nucleotides within zinc fingers 1 and 2, there are 12 nucleotide substitutions, all of them silent. It is therefore clear that these zinc fingers are under strong selective pressure.

From this comparison of basonuclin and *disco*, it seems that, in spite of the differences in overall amino acid sequence and ultimate function, the similarities between their zinc fingers are very strong and indicate similarities in their DNA-binding sites. The separate conservation of amino acid identity by the odd- and the even-numbered zinc fingers of basonuclin implies that these two classes recognize different DNA sequences. This is consistent with the recent demonstration that a pair of zinc fingers is the smallest recognition unit conferring specific DNA binding on the transcription repressor PRDI-BF1 (31). The presence of multiple similar pairs of zinc fingers in basonuclin might reasonably increase the DNA-binding activity of the protein.

Relation of Basonuclin to Other Zinc Finger Proteins. The first, third, and fifth zinc fingers of basonuclin have the usual four-amino acid spacer between the two histidines (Fig. 2). In the second, fourth, and sixth zinc fingers, the second histidine is not located at its usual position but is displaced downstream, so that the spacer between the two histidine residues is extended to six or eight amino acids. Basonuclin appears to belong to an emerging group of serine-rich proteins with scattered zinc fingers, of which some possess extended histidine spacers (Table 1). Basonuclin resembles PRDII-BF1, a transcription factor for the interferon promoter studied in the laboratory of T. Maniatis; these two proteins are quite different in amino acid sequence, but both contain zinc fingers organized in pairs, and in both proteins the second member of each pair has an extended histidine spacer (32). The two pairs of fingers in PRDII-BF1 have numerous similarities in sequence and both recognize the same DNA sequence. Although the solitary third zinc finger in PRDII-BF1 (amino acids 955–980) was reported as having a structure of C-X_2-C-X_{12}-H-X_4-C, it can also be seen as C-X_2-C-X_{12}-H-X_8-H, for there is a histidine located three amino acid residues C terminal to the cysteine. Alternation of zinc fingers with short and long spacers is not seen in the other *Drosophila* proteins with pairs of separated zinc fingers listed in Table 1, with the exception of the hunchback protein. Repeated zinc finger pairs in basonuclin may well be the result of duplications, but there is no trace of duplicated sequence outside of the zinc finger regions.

Serine Stripe. The most abundant amino acid in basonuclin is serine (12%). The other proteins listed in Table 1 are also serine-rich. In a small region of basonuclin located between the fourth and fifth zinc fingers (residues 858–875), 10 of 18 amino acids are serines (Fig. 1). This region contains only amino acids that are either favorable (H, M, K, E, and W) or indifferent (S, D, and T) to α-helix formation (36). Helical wheel analysis revealed that 8 of the 10 serines are located on one side of a putative α-helix, forming a serine stripe (Fig. 3). The amino acid sequence of the human zinc finger protein PRDII-BF1 (32) also shows the presence of a serine stripe (between amino acids residues 1328 and 1346) (Fig. 3). Seven of the 10 serines are located at identical intervals in the two proteins. In the PRDII-BF1 sequence, an arginine at amino acid 1339 (position 12 in the serine stripe) intrudes into an otherwise homogeneous serine region. This arginine is specified by codon AGG, which is likely the result of a single nucleotide substitution in serine codon AGT or AGC. Apart

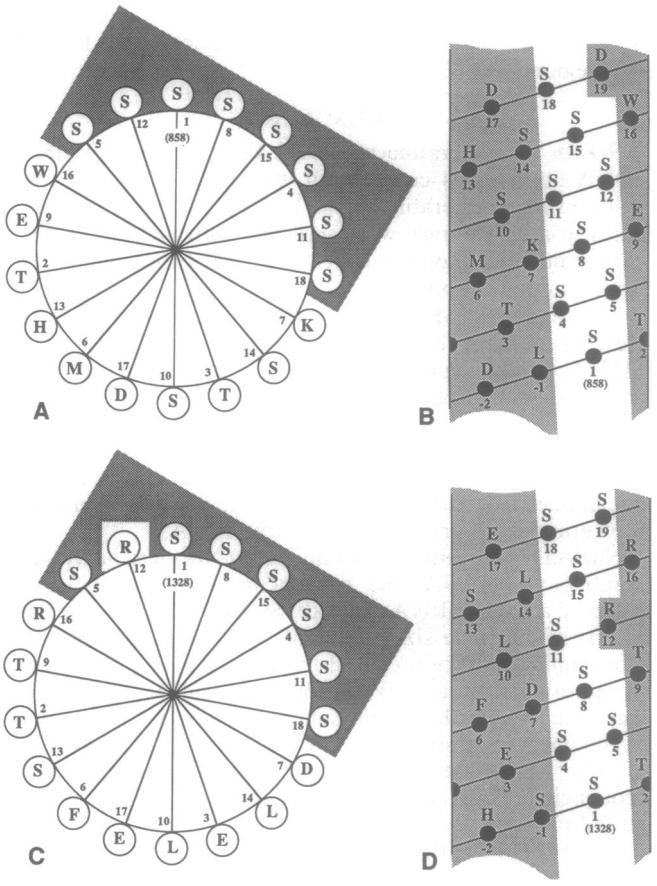

FIG. 3. Serine stripes in basonuclin and PRDII-BF1. Helical wheels (*A* and *C*) and cylindrical plots (*B* and *D*). Serine stripes are indicated by the darkened frame of the wheels and the light background of the cylinders. The stripes are formed by the sequence of residues 858–874 of basonuclin (*A* and *B*) and the sequence of residues 1328–1346 of PRDII-BF1 (*C* and *D*).

Table 1. Serine-rich zinc finger proteins

Gene (species)	Ref.	Serine content, %	Finger organization*	Interval between Hs (amino acid residues)	Presence of serine stripe
Basonuclin (human)		12.0	(2) + (2) + (2)	4, 8; 4, 6; 4, 6	Yes
PRDII-BF1 (human)	32	12.7	(2) + (1)† + (2)	3, 3/5; 8†; 3, 5	Yes
teashirt (*tsh*) (*Drosophila*)	33	12.7	(1) + (1) + (1)	5; 5; 5	No
hunchback (*hb*) (*Drosophila*)	34	11.4	(4) + (2)	3, 3, 3, 5; 3, 5	No
Suvar (3)7 (*Drosophila*)	35	10.0	(1) + (1) + (1) + (1)	5; 5; 5; 5	No
disco (*Drosophila*)	29	12.7	(2)	4, 8	No

*Within parentheses are numbers of directly adjoining zinc fingers; + indicates a separation between fingers of 40–1640 amino acid residues.

†This solitary finger has a structure that has been represented as C_2-H-X_4-C (32). However, since there is another histidine located four residues downstream of the last C, the zinc finger can also be represented as C_2-H-X_8-H.

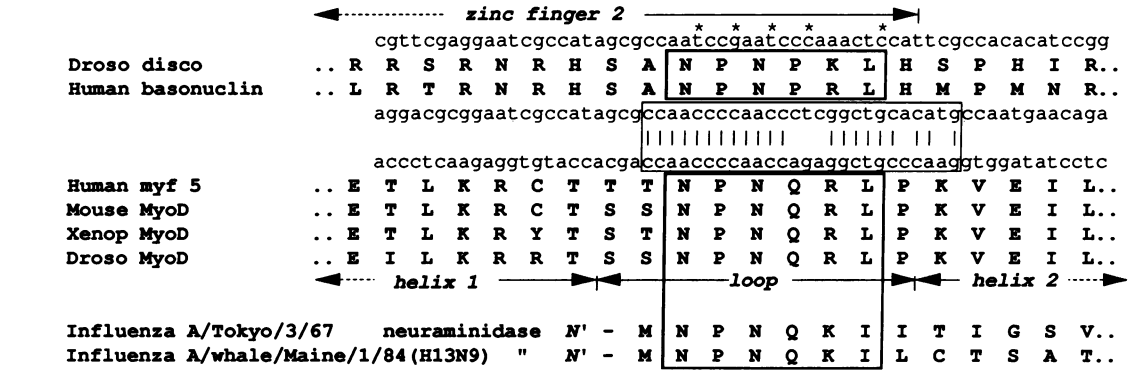

FIG. 4. Comparison of the second zinc fingers of basonuclin and *disco* with the loop region of the MyoD proteins. The amino acid sequence of the Ω loop regions is framed. Apart from the MyoD family, the influenza A proteins are the only ones that share homology with basonuclin and *disco* in this region. The similarity is greatest between basonuclin and Myf5, since 21 nucleotides out of 26 are identical. Within the sequence of 18 nucleotides of the Ω loop, the *disco* gene has 5 nucleotide mismatches with basonuclin (asterisks); all are due to silent substitutions.

from the serines themselves, the serine stripes of the two proteins are not similar in amino acid sequence.

The Similarity Between Basonuclin and Myf5 Is Confined to a Very Short Sequence in the Latter's Loop Region. Comparison of the published sequence of Myf5 (21) with the sequence of basonuclin (Fig. 4) shows that the only region of similarity is between a basonuclin sequence of 26 nucleotides and the sequence of the Myf5 Ω loop, a highly compact surface structure (9, 37). These two sequences are 80% identical. Six consecutive amino acids of the loop region are highly conserved in every other myogenic protein from *Caenorhabditis elegans* to man but are not found in other HLH proteins (16). The only example of a similar sequence in an apparently unrelated protein is NPNQKI, conserved at the extreme N terminus of the type A influenza virus neuraminidases (38–40). In basonuclin, as well as in *disco*, the NPNPK/RL sequence is located within zinc finger two. Each of the codons for the five amino acids identical in *disco* and basonuclin contains a silent mutation, suggesting that this sequence is important enough to be under strong selective pressure. The α-helix-breaking proline residues of the hexapeptide may make possible the coordination of the zinc with the more widely separated histidine residues of zinc finger two. Whether the similarity between this hexapeptide and the loop region of the myogenic proteins, which do not contain zinc fingers, is fortuitous or indicates some common property remains to be determined. Outside this region, neither basonuclin nor *disco* resembles the myogenic proteins in sequence, nor does either possess a basic region corresponding to that of the MyoD family.

Possible Functions of Basonuclin. In experiments to be reported elsewhere, basonuclin mRNA was found to be absent from most tissues whose extracted RNA was examined by Northern blotting. Antisera prepared against basonuclin detected its presence in the nuclei of the deepest layers of the epidermis, particularly in the basal layer, and we have named the protein accordingly. Basonuclin was also found in the nuclei of keratinocytes derived from vaginal epithelium. It seems most likely that basonuclin is a transcription factor specific for squamous epithelium and for the constituent keratinocytes at a stage either prior to or at the very beginning of terminal differentiation. A role related to proliferation or to the potential for proliferation, which is distributed clonally (41), is not excluded.

The excellent technical assistance of Mrs. Liesbeth Brown is gratefully acknowledged. We thank Dr. Raymond S. Brown for advice, encouragement, and many suggestions. This investigation was aided by a grant from the National Cancer Institute.

1. Harrison, S. C. (1991) *Nature (London)* **353**, 715–719.
2. Brown, R. S., Sander, C. & Argos, P. (1985) *FEBS Lett.* **186**, 271–274.
3. Miller, J., McLachlan, A. D. & Klug, A. (1985) *EMBO J.* **4**, 1609–1614.
4. Klug, A. & Rhodes, D. (1987) *Trends Biochem. Sci.* **12**, 464–469.
5. Gibson, T. J., Postma, J. P. M., Brown, S. & Argos, P. (1988) *Protein Eng.* **2**, 209–218.
6. Berg, J. M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 99–102.
7. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. (1989) *Science* **245**, 635–637.
8. Pavletich, N. P. & Pabo, C. O. (1991) *Science* **252**, 809–817.
9. Murre, C., McCaw, P. S. & Baltimore, D. (1989) *Cell* **56**, 777–783.
10. Davis, R. L., Cheng, P.-F., Lassar, A. B. & Weintraub, H. (1990) *Cell* **60**, 733–746.
11. Lassar, A. B., Buskin, J. N., Lockshon, D., Davis, R. L., Apone, S., Hauschka, S. D. & Weintraub, H. (1989) *Cell* **58**, 823–831.
12. Vinson, C. R. & Garcia, K. C. (1992) *New Biol.* **4**, 396–403.
13. Weintraub, H., Tapscott, S. J., Davis, R. L., Thayer, M. J., Adam, M. A., Lassar, A. B. & Miller, A. D. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5434–5438.
14. Choi, J., Costa, M. L., Mermelstein, C. S., Chagas, C., Holtzer, S. & Holtzer, H. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 7988–7992.
15. Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benezra, R., Blackwell, T. K., Turner, D., Rupp, R., Hollenberg, S., Zhuang, Y. & Lassar, A. (1991) *Science* **251**, 761–766.
16. Benezra, R., Davis, L. L., Lockshon, D., Turner, D. L. & Weintraub, H. (1990) *Cell* **61**, 49–59.
17. Rheinwald, J. G. & Green, H. (1975) *Cell* **6**, 331–334.
18. Allen-Hoffman, B. L. & Rheinwald, J. G. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7802–7806.
19. Simon, M. & Green, H. (1985) *Cell* **40**, 677–683.
20. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
21. Braun, T., Buschhausen-Denker, G., Bober, E., Tannich, E. & Arnold, H. H. (1989) *EMBO J.* **8**, 701–709.
22. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
23. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 1.90–1.104.
24. Tseng, H. & Green, H. (1988) *Cell* **54**, 491–496.
25. Kozak, M. (1989) *J. Cell Biol.* **108**, 229–241.
26. Dang, C. V. & Lee, W. M. F. (1989) *J. Biol. Chem.* **264**, 18019–18023.
27. Chelsky, D., Ralph, R. & Jonak, G. (1989) *Mol. Cell. Biol.* **9**, 2487–2492.
28. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
29. Heilig, J. S., Freeman, M., Laverty, T., Lee, K. J., Campos, A. R., Rubin, G. M. & Steller, H. (1991) *EMBO J.* **10**, 809–815.
30. Lee, K. J., Freeman, M. & Steller, H. (1991) *EMBO J.* **10**, 817–826.
31. Keller, A. D. & Maniatis, T. (1992) *Mol. Cell. Biol.* **12**, 1940–1949.
32. Fan, C.-M. & Maniatis, T. (1990) *Genes Dev.* **4**, 29–42.
33. Fasano, L., Röder, L., Coré, N., Alexandre, E., Vola, C., Jacq, B. & Kerridge, S. (1991) *Cell* **64**, 63–79.
34. Tautz, D., Lehmann, R., Schnürch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K. & Jäckle, H. (1987) *Nature (London)* **327**, 383–389.
35. Reuter, G., Giarre, M., Farah, J., Gausz, J., Spierer, A. & Spierer, P. (1990) *Nature (London)* **344**, 219–223.
36. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 211–222.
37. Leszczynski, J. F. & Rose, G. D. (1986) *Science* **234**, 849–855.
38. Blok, J. & Air, G. M. (1982) *Virology* **121**, 211–229.
39. Colman, P. M. & Ward, C. W. (1985) *Curr. Top. Microbiol. Immunol.* **114**, 178–255.
40. Air, G. M., Webster, R. G., Colman, P. M. & Laver, W. G. (1987) *Virology* **160**, 346–354.
41. Barrandon, Y. & Green, H. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2302–2306.