[Frontiers in Bioscience 3, d985-988, September 1, 1998]

BASONUCLIN, A ZINC FINGER PROTEIN ASSOCIATED WITH EPITHELIAL EXPANSION AND PROLIFERATION

Hung Tseng

Department of Dermatology and Cancer Center University of Pennsylvania School of Medicine, Philadelphia, PA 19104.

Received 6/16/98 Accepted 7/10/98

TABLE OF CONTENTS

1. ABSTRACT

Basonuclin is a zinc finger protein first described as a keratinoycte specific protein present in various stratified squamous epithelia found in epidermis, esophagis, cornea and virgina. Recent studies showed that its distribution also includes the germ cells of testis and ovary. The presence of basonuclin appeared always related to the cellular proliferative ability not just cell division, for it was found both in dividing and quiescent cells. Basonuclin disappeared when a cell became post-mitotic. This review examines the latest findings about the distribution, molecular and cellular biology of basonuclin and discusses its possible role in cell proliferation.

2. INTRODUCTION

Basonuclin was discovered serendipitously during a search of keratinocyte transcription regulators with a helix-loop-helix (HLH) domain similar to that of the myogenic family of proteins such as MyoD, Myf5 etc. The myogenic factors work cooperatively to initiate the myogenesis in myoblasts by inducing transcription of muscle-specific genes (1). Assuming that the control of keratinocyte differentiation might rely on similar proteins, we examined, by colony hybridization using a Myf5 cDNA as probe at low stringency, a human keratinocyte cDNA library for the presence of such sequences. Hybridizing clones were selected by the following criteria for further analysis: 1) the nucleic acid sequence homology detected by hybridization must reflect amino acid sequence homology to the HLH domain; 2) keratinocyte specific expression and 3) mRNA level must change during the transition in which keratinocytes switch from proliferation to differentiation. A cDNA clone was thus selected and its sequence revealed that it was a novel protein containing six typical C2H2 zinc finger motifs grouped into three well separated pairs (2). Its limited homology to the HLH motive resided within the second zinc finger in between the two histidine residues. An antiserum raised against the N-terminus portion of the protein detected a nuclear protein in the basal layer of the epidermis (figure 1), hence the name basonuclin (3).

3. STRUCTURAL FEATURES

The human basonuclin cDNA sequence revealed a number of structure features that are shared either among a number of transcription regulators or unique to basonuclin. All structural elements discussed below are highly conserved between human and mouse (4), suggesting their relevance to the function of the protein.

3.1. The zinc fingers

Basonuclin possesses three pairs of C2H2 type zinc fingers, which are well spread within the C-terminal two thirds of the primary sequence (figure 2). The six fingers are of two types, zinc fingers 1, 3, and 5 are related closely to zinc finger 1 of *Drosophila* protein *disconnected (disco)*; and zinc fingers 2, 4, and 6, which have a larger spacer region (6-8 amino acid residues) between the two conserved histidines, are homologous to zinc finger 2 of *disco*. The evolutionary conservation of these zinc fingers is remarkable: the first pair of zinc finger of human basonuclin shares 33 of 46 residues with the only pair of fingers of *Drosophila disco*, a 75% identity. The two proteins share no sequence otherwise. Similarly, the first and the third pairs of human and mouse basonuclin zinc fingers diverges by only one amino acid residue for each pair (1.5%), despite numerous silent nucleotide substitutions (8.2%). The sequence of the second pair of zinc finger appears less conserved, there have been 15 amino acid replacements between human of mouse (22.7%). Such a high degree of sequence conservation in the first pair of zinc fingers spanning the evolutionary distance of human, mouse and fly suggests that the sequence has been restrained from diversification and therefore required for the function of the protein.

3.2. The nuclear localization signal

(NLS) of basonuclin, PKKKSRK, matches well with the consensus sequence of the class A NLS, in which a helix breaker, a proline or a glycine, is followed by at least three basic amino acids. Basonuclin's NLS, which remains unchanged between human and mouse (4), is located near

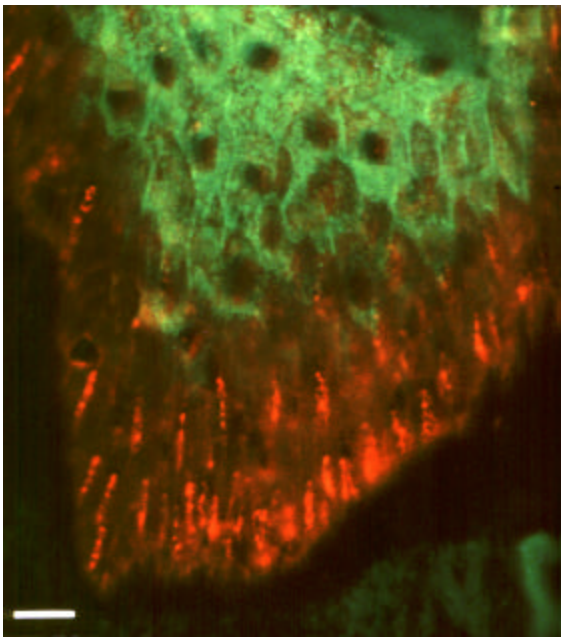

Figure 1. Basonuclin in epidermis. A cryosection of plantar skin was stained with a rabbit, anti-human-basonuclin antiserum (red) and a mouse, monoclonal antibody agaist involucrin (green) (Hudson *et al.* 1992), a keratinocyte differentiation marker. Basonuclin appeared clustered within the nucleus, suggesting its association with sub-nuclear structures. Bar, 20 μm.

the middle of the primary sequence of the protein between the first and the second pair of fingers (figure 2). Recently, it was found that phosphorylation of the Serine 541 (underlined) at the C-terminus side of the NLS PKKKSRKSS had a profound inhibitory effect on the translocation of the protein into the nucleus (5).

3.3. The serine stripe

The serine stripe is a novel structure first found in basonuclin (2), of which 12% are serines. Within the amino acid residues 858-875 in between the second and the third pairs of zinc fingers, eight serines are arranged to form a stripe on the surface of a putative alpha-helix. A similar arrangement was then found in another zinc finger protein, PRDII-BF1 (2), a well characterized transcription regulator for the interferon promoter (6). Interestingly, between the two serine stripes, the only similarity other than the serine arrangement is the absence of proline or glycine, whose presence would interfere with the formation of an alpha-helix. Again, the serine stripes of human and mouse basonuclin are identical at protein level, despite a 6.9% nucleotide substitution rate (4). The serine stripe may serve as an interphase for protein-protein interactions.

3.4. Myf-5 homology

The sequence similarity between basonuclin and Myf5 is limited to only 26 nucleotides or 6 amino acid residues (NPNPRL). It resides within the histidine spacer of zinc finger 2 of basonuclin as well as disco and the loop region of the HLH motif (2). This is a very intriguing finding and its significance has not been understood. The loop region of Myf5 was thought to contain a highly compact structure termed omega loop (7, 8), with a sequence NPNQRL that was conserved in every myogenic protein from *Caenorhabditis elegance* to man. The loop regions of other HLH proteins, however, contain different sequences. But at the N-terminus of type A influenza virus neuraminidases, a related sequence NPNQKI was found. This sequence was the most conserved also in all viral strains from various geographical regions (9, 10, 11).

4. TISSUE AND CELLULAR DISTRIBUTION

By immunocytochemistry, basonuclin was first detected in the nucleus of basal keratinocytes of plantar epidermis (3). At the tip of the rete ridges, where the epidermis is the thickest, some suprabasal cells next to the basal layer also contained basonuclin. Basonuclin stain appeared as dots within the nucleus, suggesting their association with certain sub-nuclear structure (figure 1). In the epidermis, basonuclin disappeared when keratinocytes entered terminal differentiation, as indicated by the non-overlapping staining patterns of basonuclin and involucrin, the latter being a marker of keratinocyte differentiation (3). Basonuclin was found also in skin appendages such as hair follicles. In anagen phase of the hair cycle while hair is growing, basonuclin containing cells were located within the outer root sheath throughout the length of the follicle and in the bulb where the follicular epithelial cells were in contact with the mesenchymal cells in the dermal papilla. In the out root sheath, the subcellular distribution of basonuclin appeared strictly nuclear, but became cytoplasmic within the bulb, where terminal differentiation occurred. Results of basonuclin immunocytochemical staining of other stratified squamous epithelia, such as cornea, esophagus and vagina, suggested that basonuclin was a common component in this type of tissue. In hyperplasia such as basal cell and squamous cell carcinoma, the level of basonuclin was elevated. In psoriasis, its distribution extended to the suprabasal layers (Tseng, Kupper and Green, unpublished) .

Surveys of RNA from human tissues by Northern blot or RT-PCR revealed the presence of basonuclin mRNA in spleen, thymus, mammary gland, placenta, brain and heart. The highest concentration, however, was found in testis, which contains no keratinized epithelium. Testicular basonuclin appeared restricted to the germ cell lineage in the stratified cuboidal epithelium of the seminiferous tubules, although its potential presence in the Sertoli cells had not been definitively ruled out (12). In mouse testis, in contrast to its strict association with keratinocytes in the proliferative compartment of the epidermis, basonuclin was expressed in both mitotic spermatogonia as well as meiotic spermatocytes and spermatids. Remarkably, it was even found in the highly differentiated spermatozoa, where it was no longer nuclear, but associated with the midpiece, the centrioles and possibly the acrosomes (13, 12). It has been controversial, however, as to the subcellular location of basonuclin in the developing mouse testicular germ cells. By *in situ* hybridization, Yang *et al.* detected basonuclin message in the earliest newborn mouse seminiferous tubules, but their anti-human-basonuclin antibodies could stain basonuclin only in extranuclear locations in secondary spermatids (13). On the other hand, Mahoney et al (12), using an anti-mouse-basonuclin antiserum, detected basonuclin in the nuclei of germ cells from spermatogonia to spermatids, before the protein moved to extra-nuclear locations during spermiogenesis. Mahoney *et al.* (12) also described

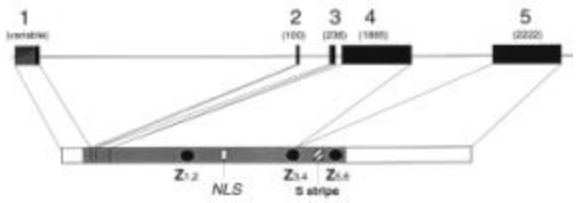

Figure 2. The genomic and mRNA structure of the basonuclin gene. The upper part of the figure depicts the gemonic structure. The filled boxes are exons and the numbers in the parentheses indicate their length in base pairs. The gray box with a dash-line border indicates that the 5' of the first exon should be extended according to a recent finding (Wei and Tseng unpublished). The lower part of the figure depicts the mRNA structure and the locations of some sequence features of the protein. The coding region is shown as a shaded box, with 5' and 3' untranslated regions shown as open boxes on the left and right, respectively. Z, zinc finger, NLS, nuclear localization signal, S stripe, serine stripe. The multiple 5' ends of the mRNA are indicated as an open box with a dash-line border.

basonuclin's presence in the nucleus of developing oocytes, the only cells expressing high level of basonuclin in the ovary. Basonuclin's wider distribution in both dividing and differentiated testicular germ cells may explain why RT-PCR analysis of testis tissue RNA gave a stronger signal than that of the skin, where basonuclin was restricted to the basal cells of the epidermis. By the same token, the lower level basonuclin message in thymus, spleen, mammary gland and placenta may indicate that like the ovary, these tissues could contain a small number of cells that express high level of basonuclin.

5. THE GENE STRUCTURE AND THE mRNA

Human basonuclin gene contains five exons spanning nearly 29 kb of genomic DNA (figure 2). The first three exons are relatively small, all zinc fingers are encoded in the fourth and the fifth exons. A notable feature of the gene is its sizable 17 kb second intron. The gene has been mapped to Chromosome 15 by screening a panel of human-rodent somatic cell hybrids (14).

Initial Northern analysis showed basonuclin mRNA as a single band of 4.6 kb. But recent RNase protection analysis of keratinocyte cellular RNA suggested that basonuclin mRNA contained multiple 5' ends mapped over a region of 400 bases of the genomic sequence. The 5' end of the longest mRNA (basonuclin mRNA I) was approximately 350 bases from that of the cDNA clone. This discrepancy might be explained by the fact that the genomic sequence near the 5' of the cDNA clone (-250 to +187, the beginning of cDNA being +1) has a GC content of over 80%, in contrast to the average 40-50% of the flanking sequences (14). RNA sequences with such a high GC content usually contain stable secondary structures that may hinder the reverse transcription process. Indeed, evidence from RNase resistance assay and computer analysis showed that a large and very stable stem-loop structure (deltaG < 210 kcal/mol) resembling a cruciform is likely formed by sequence between -212 to -59 and interferes with reverse transcription (Wei Tang and Hung Tseng, unpublished). It does not seem likely that more transcription initiation sites or exons lay further upstream. There is, however, an alternative exon (BSN1b) located downstream from the first reported exon (BSN1a) (13). The nucleotide sequences of human and mouse BSN1b are virtually identical, a remarkable degree of conservation unusual between nucleotide sequences of the two species (15).

Because of the heterogeneity at the 5' end of basonuclin mRNA, it may be initiated at multiple non-random sites or post-transcriptionally modified to generate multiple 5' ends. In either case, basonuclin promoter appeared TATA-less (14). Transcription from TATA-less promoter usually requires Sp1 as an activator protein and many Sp1 binding sites were found within the GC rich region, albeit all were downstream from the initiation sites of the longer mRNAs (14, Wei and Tseng unpublished). Despite the heterogeneity at the 5' end of the mRNA, all human basonuclin transcripts shared at the beginning of their reading frame the same two neighboring AUG codons. The first appeared a weaker translation initiator than the second. The mouse basonuclin cDNA lacked the AUG orthologous to the first AUG of human basonuclin. It had been shown that human basonuclin had a slightly slower mobility than that of mouse on SDS-PAGE gels (5). But *in vitro* translation result suggested that both human AUGs could be used (Wei and Tseng, unpublished). In any case, the sequence between the first two human AUGs is unlikely essential for at least the mouse basonuclin, since it was not translated (5).

6. EXPRESSION REGULATION

The expression of basonuclin appeared to be regulated at multiple levels. In cultured keratinocytes, its mRNA was down-regulated during terminal differentiation (3). At present, however, very little is known about the transcriptional regulation of the gene. A recent study of the 5' ends of its mRNA suggested that its translation might also be regulated. A large and very stable secondary structure was likely present at the 5' untranslated region (UTR) of the longest basonuclin transcripts. *In vitro* translation experiments showed that this structure inhibited translation initiated from the first and the second AUGs. It has been postulated that since at least two shorter transcripts did not contain this putative secondary structure, the production of basonuclin protein could be regulated by adding or removing the secondary structure from the 5' UTR of basonuclin mRNA (Wei and Tseng unpublished).

Because of its zinc fingers and a nuclear localization signal, basonuclin is likely a transcription regulator functioning in the nucleus. If this is the case, then a potential way of modifying its activity is to control basonuclin's access to its nuclear targets. It was shown recently that basonuclin's nuclear localization was indeeded regulated. The phosphorylation of a serine residue at the C-terminus of the nuclear localization signal appeared critical for blocking basonuclin's translocation into the nucleus (5).

7. BASONUCLIN AND CELLULAR PROLIFERATION

The tissue and cellular distribution of basonuclin suggested its association with cell proliferation. Basonuclin was found in some of the cells with the highest proliferative

rate in mammals: the germ cells of the seminiferous tubules, keratinocytes in the outer root sheath of the hair follicle and the basal layer of the epidermis. On the other hand, it was also found in cells that are not in mitotic cycle, such as meiotic spermatiocytes and oocytes, or spermatozoa whose haploid genome would not divide again before fertilization. However, despite their different function, destiny and proliferative state, these basonuclin expressing cells shared a common ability to proliferate when in need, which separated them from the post-mitotic somatic cells that had permanently lost their proliferative ability. This made basonuclin different from a widely used cellular proliferative marker, the Ki-67 antigen, which is present strictly in dividing cells (16, 17). In cultured keratinocytes, a group of basonuclin containing cells, often residing in the center of a large colony, lacked Ki-67, suggesting that they were quiescent (3). Under optimal conditions, basonuclin was associated with chromatin in all phases of the cell cycle, including mitotic chromosomes. Concomitant to the rise of involucrin mRNA, an indication that keratinocytes had entered terminal differentiation and become post-mitotic, basonuclin mRNA level dropped to undetectable by Northern analysis (3). These observations suggested that basonuclin was required for the proliferation of certain cell types. Clearly basonuclin is not part of the basic cell cycle machinery since its apparent absence in some highly proliferative cells.

8. PERSPECTIVE

The tissue distribution of basonuclin suggests that its function is cell type specific. Its association with cells with proliferative potential seems to indicate that it is required for proliferation, yet not all highly proliferative cells appear to contain it (e.g. intestinal epithelium). This suggests that basonuclin performs a special function required by certain proliferative cells. What that function is and what cellular need it fulfills should be the focus of future research in this field.

9. ACKNOWLEDGMENT

This work was supported by grants from National Institute in Aging and National Alopecia Areata Foundation. H.T. is a recipient of a career development award from Dermatology Foundation.

10. REFERENCES

1. Weintraub, H., R. Davis, S. Tapscott, M. Thayer, M. Krause, R. Benezra, T. K. Blackwell, D. Turner, R. Rupp, S. Hollenberg, Y. Zhuang & A. Lassar**:** The *myoD* gene family: nodal point during specification of the muscle cell lineage. *Science* 251, 761-766 (1991)

2. Tseng, H. & H. Green**:** Basonuclin: A keratinocyte protein with multiple paired zinc fingers. *Proc Natl Acad Sci USA* 89, 10311-10315 (1992)

3. Tseng, H. & H. Green**:** Association of basonuclin with ability of keratinocytes to multiply and with absence of terminal differentiation. *J Cell Biol* 126(2), 495-506 (1994)

4. Matsuzaki, K., S. Iuchi & H. Green**:** Conservation of human and mouse basonuclins as a guide to important features of the protein. *Gene* 195, 87-92 (1997)

5. Iuchi, S. & H. Green**:** Nuclear localization of basonuclin in human keratinocytes and the role of phosphorylation. *Proc. Nat. Acad. Sci USA* 94, 7948-7953 (1997)

6. Fan, C.-M. & T. Maniatis**:** A DNA-binding protein containing two widely separated zinc finger motifs that recognize the same DNA sequence. *Genes Dev* 4, 29-42 (1990)

7. Murre, C., P. S. McCaw & D. Baltimore**:** A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, *MyoD*, and *myc* proteins. *Cell* 56, 777-783 (1989)

8. Leszczynski, J. F. & G. D. Rose**:** Loops in globular proteins: a novel category of secondary structure. *Science* 234, 849-855 (1986)

9. Blok, J. & G. M. Air**:** Sequence variation at the 3¢ end of the neuraminidase gene from 39 influenza type A viruses. *Virology* 121, 211-229 (1982)

10. Colman, P. M. & C. W. Ward**:** Structure and diversity of influenza virus neuraminidase. *Curr Topics Microbiol Immunol* 114, 178-255 (1985)

11. Air, G. M., R. G. Webster, P. M. Colman & W. G. Laver**:** Distribution of sequence differences in influenza N9 neuraminidase of tern and whale viruses and crystallization of the whale neuraminidase complexed with antibodies. *Virology* 160, 346-354 (1987)

12. Mahoney, M. G., W. Tang, M. M. Xiang, S. B. Moss, G. L. Gerton, J. R. Stanley & H. Tseng**:** Translocation of the zinc finger protein basonuclin from the mouse germ cell nucleus to the midpiece of the spermatozoon during spermiogenesis. *Biol Reprod*, 59, 388-394 (1998)

13. Yang, Z. H., G. I. Gallicano, Q.-C. Yu & E. Fuchs**:** An unexpected localization of basonuclin in the centrosome, mitochondria and acrosome of developing spermatids. *J Cell Biol* 137(May 5), 657-670 (1997)

14. Teumer, J., H. Tseng & H. Green**:** The human basonuclin gene. *Gene* 188, 1-7 (1997)

15. Makalowski, W., J. Zhang & M. S. Boguski**:** Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences. *Genome Res* 6, 846-857 (1996)

16. Gerdes, J., H. Lemke, H. Baisch, H.-H. Wacker, U. Schwab & H. Stein**:** Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67. *J Immunol* 133, 1710-1715 (1984)

17. Gerdes, J., L. Li, C. Schlueter, M. Duchrow, C. Wohlenberg, C. Gerlach, I. Stahmer, S. Kloth, E. Brandt & H.-D. Flad**:** Immunobiochemical and molecular biologic characterization of the cell proliferation-associated nuclear antigen that is defined by monoclonal antibody Ki-67. *Am J Pathol* 138, 867-873 (1991)

Key words: Zinc fingers, cellular proliferation, keratinocytes, germ cells, epidermis.

Send correspondence to: Dr Hung Tseng, Department of Dermatology, 422 Curie Boulevard, Stellar-Chance Laboratories M8c, University of Pennsylvania, Philadelphia, PA 19104, Tel:(215)-898-1981, Fax:(215)-573-9102, E-mail: htsengpe@mail.med.upenn.edu