Volume 321, number 2,3, 267–273 FEBS 12405 April 1993

The structure of neutrophil defensin genes

Rose Linzmeier, David Michaelson, Lide Liu and Tomas Ganz

Will Rogers Institute Pulmonary Research Laboratory at UCLA, Department of Medicine, UCLA School of Medicine, Los Angeles, CA, USA

Received 17 March 1993

Defensins are a family of microbicidal peptides abundant in the granules of mammalian neutrophils, in rabbit alveolar macrophages, and in human and murine intestinal Paneth cells. We cloned and sequenced the genes of three neutrophil-specific defensins. Human HNP-1 and HNP-3 are nearly identical and rabbit NP-3a is closely related. The four known neutrophil-specific defensin genes are strikingly similar in the structure and organization of their three exons and two introns, but the three defensin genes expressed in macrophages (MCP-1 and -2) or Paneth cells (HD-5) are organized differently: HD-5 has only two exons, and MCP-1 and -2 have a comparatively short first intron. The diverse genomic organization of defensin genes may contribute to their cell-specific expression.

Neutrophil; Gene structure; Defensin

1. INTRODUCTION

Polymorphonucleated leukocytes (neutrophils), the most numerous phagocytes circulating in blood and the main cellular constituents of pus, kill bacteria and fungi by exposing them to reactive oxygen intermediates and granule-associated microbicidal proteins [1–3]. Members of the defensin family are among the most abundant of the microbicidal proteins. The human neutrophil peptides HNP-1, -2 and -3 constitute 5–7% of the protein of human neutrophils [4]. The fourth human neutrophil defensin, HNP-4, is approximately one hundred times less abundant [4,5]. The six rabbit defensins, neutrophil peptides NP-1, -2, -3a, -3b, -4 and -5, comprise 15–20% of rabbit neutrophil protein [6]. Two defensins have also been found in guinea pig neutrophils [7,8]. Defensin expression is not limited to neutrophils. Two defensins, MCP-1 and -2, expressed in alveolar macrophages [6,9] are identical to two of the rabbit neutrophil peptides, respectively, NP-1 and -2. The Paneth cells of the mouse small intestine have been shown to contain a large amount of mRNAs coding for typical defensin peptides, termed cryptdins [10]. Recently, the mRNAs coding for a fifth and sixth human defensin (HD-5 and -6) were found to be expressed only in the Paneth cells of the small intestine [11,12].

Defensins are small cationic peptides 29–34 amino acids long. They are arginine-rich and contain six conserved cysteine residues that form three disulfide bonds [13]. In vitro, defensins have microbicidal and cytotoxic activities [4,14] that may be related to their ability to permeabilize membranes [15,16], but also manifest chemotactic and endocrine regulatory effects. Individual members of the defensin family differ remarkably in the spectrum and potency of their biological activities [17].

In order to facilitate studies of tissue-specific expression of the various members of the defensin family, we report the cloning and sequencing of three neutrophil peptide genes, the human HNP-1 and -3 and the rabbit NP-3a, and analyze the differences among these genes and the genes coding for defensins expressed in intestinal Paneth cells and alveolar macrophages.

Correspondence address: T. Ganz, Department of Medicine, UCLA Center for the Health Sciences, Rm. 37-131, Los Angeles, CA 90024-1736, USA. Fax: (1) (310) 206-8766.

Abbreviations: cDNA, complementary DNA; GNCP, guinea pig neutrophil cationic peptide; HD, human defensin; HNP, human neutrophil peptide; MAC, macrophage; MCP, macrophage cationic peptide; mRNA, messenger RNA; NP, neutrophil peptide; PCR, polymerase chain reaction; PMN, polymorphonuclear leukocyte.

The nucleotide sequences of genes HNP-1, HNP-3 and NP-3a have been deposited with Genbank, accession numbers L12690, L12691, and M64599, respectively.

2. MATERIALS AND METHODS

A rabbit sperm genomic library in EMBL4 λ phage was the generous gift of Dr. Katherine Knight, University of Chicago, Chicago, IL. The library plated on *Escherichia coli* strain LE392 was screened with the *Eco*RI fragment of NP-3a cDNA [18]. A human lung fibroblast genomic library in the Lambda Fix II vector (Stratagene) plated on *E. coli* strain PLK17 was probed with a PCR-amplified fragment of HNP-1 cDNA, containing bases 65–329 [19]. Hybridizing plaques taken through two or three successive screenings were used for DNA preparations. DNA fragments identified by Southern analysis were subcloned into either the Bluescript SK(−) plasmid or M13mp18 and grown in *E. coli* XL1-Blue. Single-stranded DNA was sequenced using the Sequenase Version 2.0 kit (United States Biochemical). Primers

used for sequencing were the M13 universal primer (United States Biochemical) or specific sequencing primers produced by Dr. Thomas Sutherland and Dr. Dohn Glitz (UCLA). All DNA sequences reported here were confirmed by sequencing both DNA strands. DNA sequence analysis was performed with the help of PC-GENE (Intelligenetics, Palo Alto, CA) programs and the Signal Scan 3 program [20].

3. RESULTS AND DISCUSSION

3.1. *Cloning and sequencing of the human neutrophil genes HNP-1 and -3*

Using the PCR-amplified HNP-1 cDNA as a probe eight clones were recovered from the human lung fibroblast Lambda Fix II library. Five of these coded for HNP-1, one coded for HNP-3 and the remaining two were not identified.

Comparison of the gene and cDNA sequences showed each gene to consist of three exons and two introns (Fig. 1). Each splice site junction follows the GT–AG rule with variable homology to the canonical exon–intron junction consensus sequences [21,22]. The first exon of each gene codes for the 5′ untranslated region. The second exon contains the translation initiation consensus sequence [23], the signal peptide and most of the spacer segment (propiece) removed during post-translational processing to mature peptide. The third exon codes for the mature protein.

The polyadenylation site is located within the third exon, twelve nucleotides from its end [24]. A putative TATA box is located 68 bases upstream from the cDNA start. This sequence, TTAAATA, is the same as that found in the NP-3a, GNCP-2, MCP-1 and MCP-2 defensin genes. There are three (HNP-3) or four (HNP-1) CAAT boxes upstream of the TATA box. The sequence of the HNP-1 and -3 genes is in complete agreement with the sequence of their cDNAs [19]. Each gene encodes a 94 amino acid preproprotein which is known to be processed through intermediates to the 30 amino acid mature defensins [25].

The HNP-1 and -3 genes are nearly identical. Within the 3.7 kb sequenced only seventeen base-pair differences were found. Fourteen of these are located more than 350 nucleotides upstream of the presumed mRNA start, one is in the second intron, and one in the 3′ untranslated region of the mRNA. The remaining difference is in the first codon of the mature protein coding region, leading to the known amino acid difference in the mature peptides: HNP-1 starts with alanine while HNP-3 begins with aspartic acid. Since all donors express HNP-1 peptide and about 90% express HNP-3 (unpublished), we surmise that the genes are non-allelic. The high degree of homology suggests these genes have arisen from a very recent duplication.

3.2. *Cloning and sequencing of the rabbit neutrophil gene NP-3a*

Of the five clones strongly hybridizing to the NP-3a probe, three were sequenced wholly or in part and found to belong to the same gene, NP-3a. The gene and cDNA [18] sequence were in agreement, with the exception of a one base-pair difference at position 2,151 at the end of the second exon (C in the cDNA and a T in the gene). Both versions code for a valine. The genomic sequence of this region was obtained from three different clones and all three were in agreement. This discrepancy can be attributed to allelic variation.

Comparison of the gene and cDNA sequence showed NP-3a has the same organization as HNP-1 and -3. The first exon codes for the 5′ untranslated region; the second exon for the translation initiation consensus sequence [23], the signal sequence and most of the propiece; and the third exon for the mature protein portion. The splice sites for the two introns follow the GT–AG rule and there is some homology to the more extended splice site consensus sequences [21,22]. The consensus polyadenylation site [24] and the presumed TATA box are noted in Fig. 2. As in HNP-1, four CAAT boxes were found upstream of the TATA box (Fig. 2).

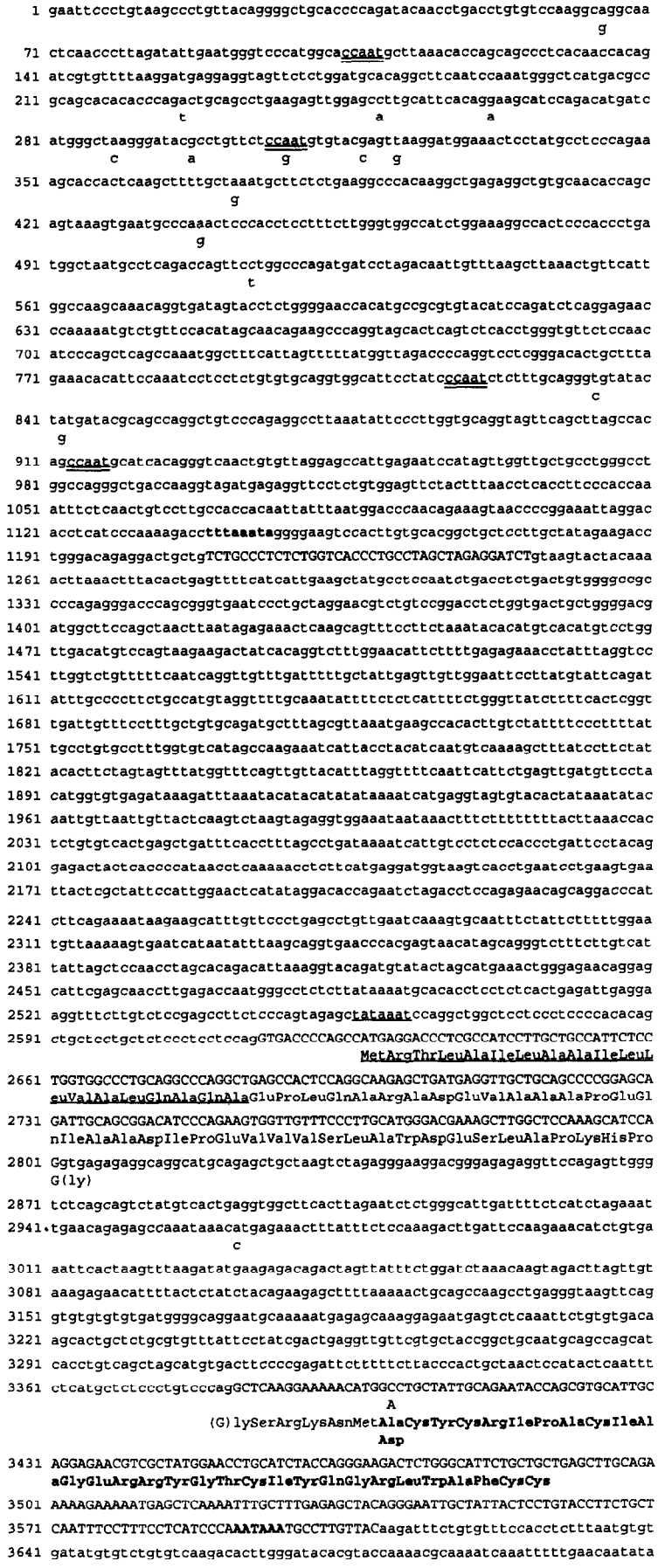
3.3. *Organization of defensin genes*

The structures of seven genes encoding defensins from three mammalian species expressed in three different tissues are compared in Fig. 3. The genes fall into three structural groups, each with a unique pattern of cell-specific expression.

Four defensins expressed exclusively in neutrophils, human HNP-1 and -3, rabbit NP-3a and guinea pig GNCP-2 [26], have three exons, each coding for the same approximate region of the gene product. The length of the first intron is 1,369 in HNP-1 and -3, 1,232 in NP-3a and 1,217 in GNCP-2. The second intron, in each case less than half the length of the first, is 588 nucleotides in HNP-1 and -3, 577 in NP-3a and 561 in GNCP-2. The rabbit MCP-1 and -2 genes are expressed in alveolar macrophages as well as neutrophils. The number and content of the exons of these genes is the same as the neutrophil-specific genes [27], however, there are about 800 fewer nucleotides in the first intron. The HD-5 gene, expressed in Paneth cells, is organized differently than the other genes. It contains only two

→

Fig. 1. The nucleotide sequence of the HNP-1 and HNP-3 genes. The HNP-1 gene sequence is shown with differences in HNP-3 indicated below. The first nucleotide of an *Eco*RI site is numbered as base 1. Introns are in lower-case letters, exons are in upper-case letters. The end of exon 3 is deduced from the cDNA polyadenylation site. The preproprotein sequences are shown in three letter code with the signal sequence underlined. The TATA-like box, polyadenylation signal and mature protein are shown in bold. CAAT boxes upstream of the TATA-like box are double underlined. The TATA box upstream of the second exon is homologous to that of the HD-5 promoter and is underlined.



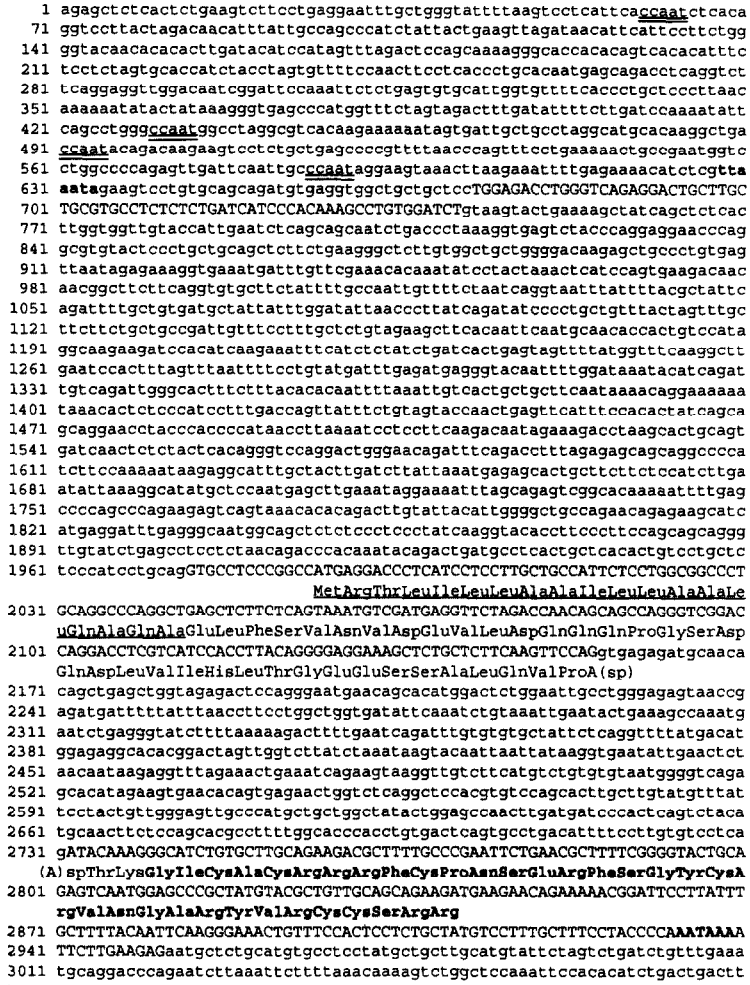


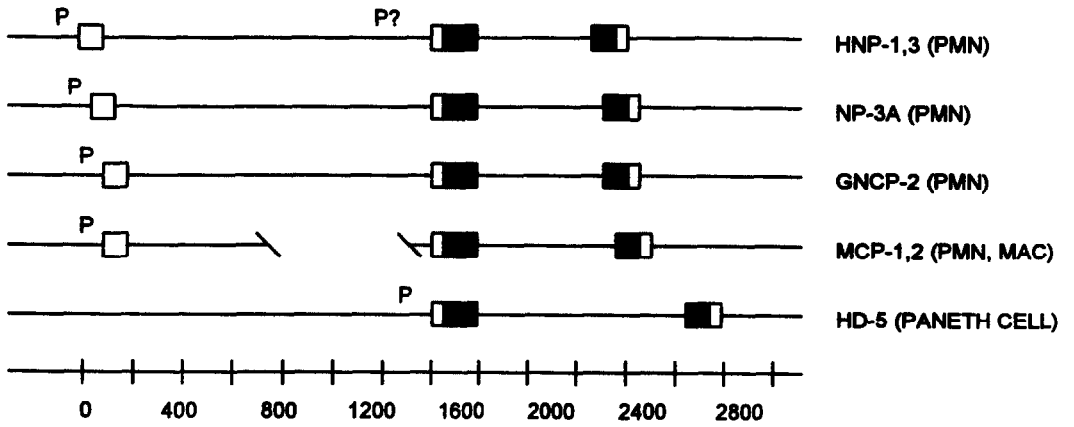
Fig. 2. The nucleotide sequence of the NP-3a gene. Nucleotides comprising introns are shown in lower-case letters. The 3′ end of the third exon was deduced from the cDNA polyadenylation site. The deduced amino acid sequence of the signal peptide (underlined) and the propeptide are shown. The mature protein is shown in boldface. The TATA-like box and polyadenylation signal are also in boldface. Upstream CAAT boxes are double underlined.

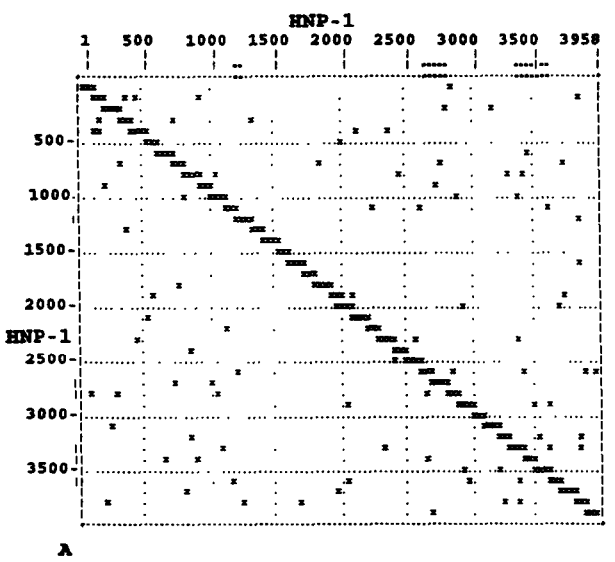
exons, the first coding for the 5′ untranslated region and prepro portion and the second for the mature protein portion [11]. The single intron is 980 nucleotides, intermediate in length between the first and second introns of the neutrophil genes. Interestingly, the human neutrophil genes contain a TATA box in a position analogous to that of the HD-5 gene (underlined in Fig. 1).

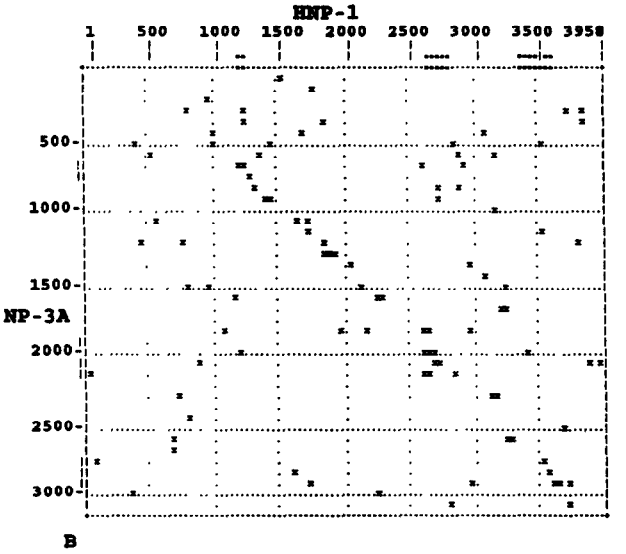
Using the PC-GENE Pustell program an analysis of the similarity of these defensin genes was made. When HNP-1 is compared to itself (Fig. 3A), most of the homology falls along the diagonal, showing that the gene contains few repeated sequences. HNP-1 and NP-3a are closely related, even outside the coding regions, as reflected in the clear and uninterrupted diagonal shown in Fig. 3B. NP-3a and MCP-1 are also quite homologous. The break in their homology due to the

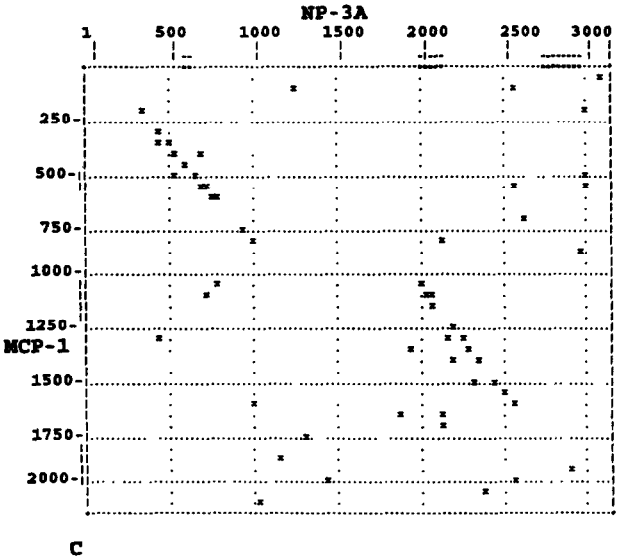
→

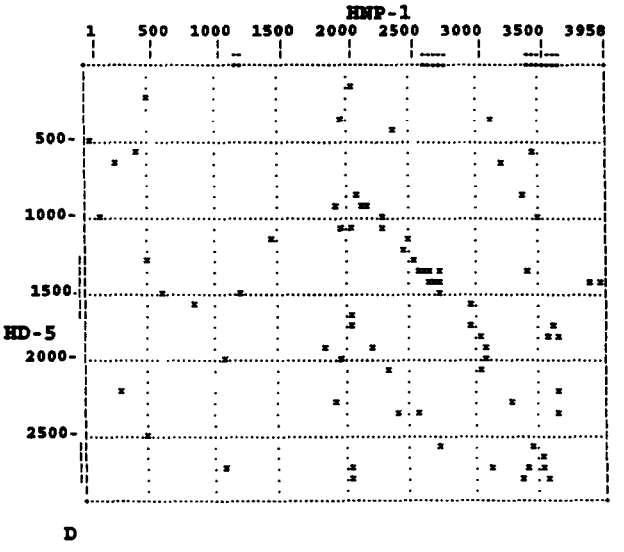
Fig. 3. Comparison of the organization and homology of defensin genes. The top panel shows schematic representations of the organization of seven defensin genes. Open boxes indicate non-coding exon regions. Shaded boxes show translated coding regions. The 3′ and 5′ flanking regions and introns are represented by straight lines. The break in the first intron in the MCP-1 and -2 indicates the location of a possible 800 base-pair deletion. P indicates the location of promoters deduced from the location of the TATA boxes. The P? stands for a TATA box in the first intron of HNP-1 and -3 the location of which is analogous to that found in HD-5. The cell type in which each of the genes is expressed is shown in parentheses. The lower panels (A–D) show four graphs generated by the PC-GENE Pustell program, with parameters range = 4, scale = 1. The numbers on the axes refer to the nucleotide position of the gene indicated. Double lines along the axes indicate the positions of the exons. An area of identity is indicated by an X.


P
P?
HNP-1,3 (PMN)
P
NP-3A (PMN)
P
GNCP-2 (PMN)
P
MCP-1,2 (PMN, MAC)
P
HD-5 (PANETH CELL)
0
400
800
1200
1600
2000
2400
2800


HNP-1
HNP-1
A


HNP-1
NP-3A
B


NP-3A
MCP-1
C


HNP-1
HD-5
D

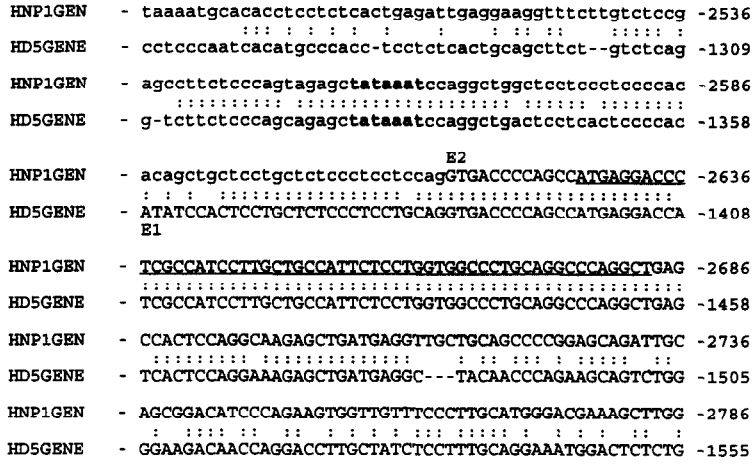


Fig. 4. Region of homology between HNP-1 and the HD-5 promoter. The PC-GENE NALIGN program was used to align the HD-5 and HNP-1 genes. Identical nucleotides are marked with a colon. The HNP-1 intron and HD-5 5′ untranslated region are in lower-case letters. Exons are in upper-case letters. The TATA boxes are in bold type. The HNP-1 signal sequence is underlined. E2 indicates the start of the second exon of HNP-1. E1 indicates the start of the first exon of HD-5.

probable deletion in the first intron of MCP-1 is visualized in Fig. 3C. The human defensin gene, HNP-1, is less closely related to the guinea pig GNCP-2 gene than to the rabbit defensin genes, with areas of homology distributed all along the two genes (data not shown).

As reflected by a truncated diagonal in Fig. 3D, the homology between HNP-1 and HD-5 genes begins just upstream of the HNP-1 second exon and the HD-5 first exon, and the similarity continues through the subsequent intron and the next exon. This suggests that the non-homologous 5′ end of the two genes diverged through deletion or insertion. The striking homology between the HD-5 promoter region and the region upstream of the second exon of HNP-1 and -3 is shown in Fig. 4. The sequences upstream of the TATA boxes is unrelated. Homology is quite strong in the area of the TATA boxes, remains strong through the signal sequence, then diverges somewhat after the start of the proprotein region. The TATA box upstream of the second exon of the HNP genes may be a vestigial remnant from the ancestor gene of the HNP-1, HNP-3 and HD-5 genes. It is also possible that this potential promoter may initiate transcription of HNP-1 and -3 in some cell type other than neutrophils.

The conserved exon–intron structure of neutrophil defensin genes from man, rabbit and guinea pig contrasts with the different layouts of otherwise homologous human and rabbit defensin genes expressed in Paneth cells or macrophages. Gross rearrangements (deletions or insertions), that appear as non-homologous segments in Pustell plots, may have contributed to the differences in tissue specificity of defensin gene expression. These regions could be fertile territory in the search for tissue-specific *cis*-acting sequences.

Acknowledgements: This study was supported in part by Grants HL35640 and HL46809 from the National Institutes of Health. We acknowledge the essential contributions of Kathleen A. Page in the early stages of the HNP gene studies. Bahram Ghaderi and Joseph Rosales provided excellent technical assistance in the sequencing of HNP-1 and -3.

REFERENCES

[1] Elsbach, P. and Weiss, J. (1992) in: Inflammation: Basic Principles and Clinical Correlates (Gallin, J.I., Goldstein, I.M. and Snyderman, R. eds.) pp. 603–636, Raven Press, New York.
[2] Ganz, T., Selsted, M.E. and Lehrer, R.I. (1986) Sem. Resp. Infect. 1, 107–117.
[3] Spitznagel, J.K. (1984) Contemp. Topics Immunobiol. 14, 283–343.
[4] Ganz, T., Selsted, M.E. and Lehrer, R.I. (1990) Eur. J. Haematol. 44, 1–8.
[5] Wilde, C.G., Griffith, J.E., Marra, M.N., Snable, J.L. and Scott, R.W. (1989) J. Biol. Chem. 264, 11200–11203.
[6] Selsted, M.E., Brown, D.M., De Lange, R.J., Harwig, S.S.L. and Lehrer, R.I. (1985) J. Biol. Chem. 260, 4579–4584.
[7] Selsted, M.E. and Harwig, S.S.L. (1987) Infect. Immun. 55, 2281–2286.
[8] Nagaoka, I., Someya, A., Iwabuchi, K. and Yamashita, T. (1991) FEBS Lett. 280, 287–291.
[9] Selsted, M.E., Brown, D.M., De Lange, R.J. and Lehrer, R.I. (1983) J. Biol. Chem. 258, 14485–14489.
[10] Ouellete, A.J., Greco, R.M., James, M., Frederick, D., Naftilan, J. and Fallon, J.T. (1989) J. Cell. Biol. 108, 1687–1695.
[11] Jones, D.E. and Bevins, C.L. (1992) J. Biol. Chem. 267, 23216–23225.
[12] Jones, D.E. and Bevins, C.L. (1993) FEBS Lett. 315, 187–192.
[13] Selsted, M.E. and Harwig, S.S.L. (1989) J. Biol. Chem. 264, 4003–4007.
[14] Lichtenstein, A., Ganz, T., Selsted, M.E. and Lehrer, R.I. (1986) Blood 68, 1407–1411.
[15] Lehrer, R.I., Barton, A., Daher, K.A., Harwig, S.S.L., Ganz, T. and Selsted, M.E. (1989) J. Clin. Invest. 84, 553–559.
[16] Lichtenstein, A. (1991) J. Clin. Invest. 88, 93–100.

[17] Lehrer, R.I. and Ganz, T. (1990) Blood 76, 2169–2181.
[18] Michaelson, D., Rayner, J., Couto, M. and Ganz, T. (1992) J. Leukoc. Biol. 5, 634–639.
[19] Daher, K.A., Lehrer, R.I., Ganz, T. and Kronenberg, M. (1988) Proc. Natl. Acad. Sci. USA 85, 7327–7331.
[20] Prestridge, D.S. (1991) CABIOS 7, 203–206.
[21] Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349–83.
[22] Mount, S.M. (1982) Nucleic Acid Res. 10, 459–472.
[23] Kozak, M. (1991) J. Cell Biol. 115, 887–903.
[24] Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211–214.
[25] Valore, E.V. and Ganz, T. (1992) Blood 79, 1538–1544.
[26] Nagaoka, I., Someya, A., Iwabuchi, K. and Yamashita, T. (1992) FEBS Lett. 303, 31–35.
[27] Ganz, T., Rayner, J.R., Valore, E.V., Tumolo, A., Talmadge, K. and Fuller, F. (1989) J. Immunol. 143, 1358–1365.