GENOME-WIDE IDENTIFICATION OF MAMMALIAN DEFENSINS AND FUNCTIONAL ANALYSIS OF A NOVEL DEFENSIN-RELATED PEPTIDE

By

AMAR AJIT PATIL

Bachelor of Veterinary Science & Animal Husbandry
Dr. P.D.K.V. Agricultural University
Akola, Maharashtra, India
2000

Master of Veterinary Science in Animal Biotechnology
Indian Veterinary Research Institute
Izatnagar, Uttar Pradesh, India
2002

Submitted to the Faculty of the Graduate College of the Oklahoma State University in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY December, 2006
GENOME-WIDE IDENTIFICATION OF MAMMALIAN
DEFENSINS AND FUNCTIONAL ANALYSIS OF
A NOVEL DEFENSIN-RELATED PEPTIDE

Dissertation Approved:

Guolong Zhang
Dissertation Adviser

Clinton Krehbiel
Chair

Haobo Jiang

Lin Liu

A. Gordon Emslie
Dean of the Graduate College
ACKNOWLEDGEMENT

I would like to take this opportunity to thank everybody who helped me to obtain my doctoral degree. I will always cherish my memory of all those who helped to enrich experiences of my life.

I express profound gratitude towards my advisor, Dr. Guolong Zhang, for his consistent support, constructive counsel, painstaking efforts and inspiring guidance. His role throughout my graduate studies has been much more than formal obligation as an advisor. I shall always appreciate his faith in my capabilities. I would like to thank my supervisory committee members, Drs. Clinton Krehbiel, Rodney Geisert, Haobo Jiang and Lin Liu for taking time to serve on committee and their help and support. I would also like to thank my lab members, Jessie Xiao, Yibin Cai, Yugendar Bommineni, Anthony Luyai, and Willie Li for their pleasant company, help and advice.

Words are not enough to describe my parent’s sacrifices and patience for my education. Their love, affection and prayer always inspire me during ups and downs of my life. I would like to thank the most beautiful and wonderful thing that has ever happened to me, my wife Yogita. She has been very supportive of my research endeavor and has sacrificed a lot of weekend time so that I can finish my Ph.D. I would also like to thank my in-laws for their encouragement and patience as they are waiting for both of us to visit India. Lastly, I would like to thank my brother Atul, brother-in-law Yogesh and the rest of my family and friends for their support.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. REVIEW OF LITERATURE</td>
<td>3</td>
</tr>
<tr>
<td>Classification of host defense peptides</td>
<td>4</td>
</tr>
<tr>
<td>Pleotropic effects of host defense peptides</td>
<td>7</td>
</tr>
<tr>
<td>Structure-activity relationship</td>
<td>14</td>
</tr>
<tr>
<td>Strategies for peptide design and synthesis</td>
<td>18</td>
</tr>
<tr>
<td>Current therapeutic development and clinical application</td>
<td>22</td>
</tr>
<tr>
<td>Conclusion and future prospect</td>
<td>24</td>
</tr>
<tr>
<td>References</td>
<td>25</td>
</tr>
<tr>
<td>III. RAPID EVOLUTION AND DIVERSIFICATION OF MAMMALIAN α-DEFENSINS AS REVEALED BY COMPARATIVE ANALYSIS OF RODENT AND PRIMATE GENES</td>
<td>53</td>
</tr>
<tr>
<td>Abstract</td>
<td>54</td>
</tr>
<tr>
<td>Introduction</td>
<td>55</td>
</tr>
<tr>
<td>Materials and methods</td>
<td>58</td>
</tr>
<tr>
<td>Results</td>
<td>62</td>
</tr>
<tr>
<td>Discussion</td>
<td>74</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>79</td>
</tr>
<tr>
<td>References</td>
<td>80</td>
</tr>
<tr>
<td>Figure legends</td>
<td>86</td>
</tr>
<tr>
<td>Supplementary material</td>
<td>96</td>
</tr>
<tr>
<td>IV. CROSS-SPECIES ANALYSIS OF MAMMALIAN β-DEFENSIN GENE FAMILY: PRESENCE OF SYNTENIC GENE CLUSTERS AND PREFERENTIAL EXPRESSION IN THE MALE REPRODUCTIVE TRACT</td>
<td>99</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CHAPTER III</strong></td>
<td></td>
</tr>
<tr>
<td>1. Sequences of primers used for RT-PCR analysis of rat α-defensins</td>
<td>89</td>
</tr>
<tr>
<td>2. Demonstration of positive selection in 11 phylogenetically</td>
<td>90</td>
</tr>
<tr>
<td>independent pairs of mammalian α-defensin genes</td>
<td></td>
</tr>
<tr>
<td><strong>CHAPTER IV</strong></td>
<td></td>
</tr>
<tr>
<td>Supplementary table 1. GenBank accession numbers of mammalian</td>
<td>152</td>
</tr>
<tr>
<td>β-defensin cDNA sequences</td>
<td></td>
</tr>
<tr>
<td>Supplementary table 2. Sequences of primer used for RT-PCR analysis of</td>
<td>156</td>
</tr>
<tr>
<td>rat β-defensins</td>
<td></td>
</tr>
<tr>
<td><strong>CHAPTER V</strong></td>
<td></td>
</tr>
<tr>
<td>1. Antibacterial spectrum of rattusin</td>
<td>186</td>
</tr>
<tr>
<td>2. Antibacterial activity of pro-rattusin</td>
<td>187</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure Page

CHAPTER III

1. Amino acid sequence alignment of novel α-defensins and α-defensin-related sequences in the primates, rat and mouse……………………………... 91
2. Tissue expression patterns of novel rat β-defensins ……………………………92
3. Phylogenetic tree of mammalian α-defensins………………………………... 93
4. Alignment of the second exon sequences of rodent α-defensin-related sequences with three rat enteric α-defensins……………………………………...94
5. Genomic organizations of α-defensin clusters in the primates, rat, and mouse……………………………………………………………………….95
Supplementary Figure 1. Sequence alignment of the exon encoding the signal peptide and prosegments of three orthologous genes in the human (DEFA7P), chimpanzee (PTAD7), and olive baboon (PAAD3)………………….97
Supplementary Figure 2. Amino acid sequence alignment of a total of 82 known and newly identified α-defensins and α-defensin-related sequences in the primates and glires……………………………………………….98

CHAPTER IV

1. Amino acid sequence alignment of rat β-defensins…………………………139
2. Identification of novel β-defensins in human and mouse……………………140
3. Genomic organization of syntenic mammalian β-defensin genes………………141
4. Structural organizations of mammalian β-defensin genes……………………142
5. Phylogenetic tree of mammalian β-defensins………………………………...143
6. Tissue expression patterns of rat β-defensins ........................................144
7. Developmental regulation pattern of rat β-defensins in epididymis and testis ……………………………………………………………………..145
Supplementary Figure 1. Amino acid sequence alignment of the canine β-defensin gene family…………………………………………………………..148
Supplementary Figure 2. Amino acid sequence alignment of the human β-defensin gene family………………………………………………………….149
Supplementary Figure 3. Amino acid sequence alignment of the mouse β-defensin gene family………………………………………………………….150
Supplementary Figure 4. Genomic organization of five syntenic β-defensin gene clusters in the human and chimpanzee.................................151

CHAPTER V

1. Nucleotide sequence of the rattusin gene.........................................................188
2. Sequence alignment of rattusin with representative Paneth cell α-defensins, namely mouse cryptdin-4 and human defensin-5..............189
3. Expression pattern of the rattusin gene across gastrointestinal axis...........190
4. RP-HPLC profile of reduced and oxidized rattusin.........................................191
5. Kinetics of killing of E. coli O157H7.................................................................192
6. Kinetics of killing of S. aureus........................................................................193
7. Effect of salinity on the antibacterial activity of rattusin against S. aureus and E. coli O157H7 in presence of increasing concentrations of NaCl........194
8. Effect of Mg++ on the antibacterial activity of rattusin and cryptdin-4 against S. aureus and E. coli O157H7.....................................................195
9. Absence of cytotoxicity of rattusin and cryptdin-4 to CACO-2 cells..........196
CHAPTER I
INTRODUCTION

Antimicrobial peptides (AMPs) are found in virtually every living being as a part of the innate immune system. In vertebrates, there are two major families of AMPs, namely defensins and cathelicidins. They kill a broad range of pathogens ingested by phagocytes and also help fight off pathogens trying to enter through mucosal surfaces and skin. The antibacterial action of AMPs is mediated primarily through binding and disruption of microbial membranes.

In addition to the antimicrobial activity, they exert various modulatory effects on the innate and adaptive immune system. Many AMPs are directly chemotactic for neutrophils, monocytes, and lymphocytes. They are capable of inducing the production of an array of proinflammatory cytokines and chemokines from immune and epithelial cells. They also play a role in differentiation and maturation of dendritic cells, which are professional antigen-presenting cells of the immune system. Paradoxically, some AMPs also have anti-inflammatory effects by preventing the release of tumor necrosis factor-α from the immune cells induced by lipopolysaccharide (LPS), a major cell wall component of Gram-negative bacteria. Due to their pleotropic effects on the immune system, AMPs are also being called host defense peptides (HDPs).
Recently, there is a sharp rise in the incidence of antibiotic-resistant infections with some bacteria even showing resistance to multiple antibiotics. Due to their potent antimicrobial and immunomodulatory effects, AMPs are being actively explored as potential antibacterial drugs especially against antibiotic-resistant bacteria. It is extremely difficult for bacteria to gain resistance to these peptides as opposed to traditional antibiotics. Several of such peptides are in clinical trials to treat both topical and systemic infections.

In this report, we present the latest advances on defensins and cathelicidins pertaining to their pleotropic activities and potential as antibacterial drugs (Chapter II). We have discovered the complete repertoire of the α- and β-defensin gene families in various mammalian species including the human, rat, mouse, chimpanzee and dog through a comprehensive bioinformatics approach (Chapter III & IV). Furthermore, we characterized *in vitro* antibacterial properties of rattusin, a novel α-defensin-related peptide in rat (Chapter V). Overall findings are summarized and future prospect are also discussed (Chapter VI).
CHAPTER II
REVIEW OF LITERATURE

Discovery of penicillin and a host of other antibiotics in the middle of last century led to a decrease in infectious disease mortality by 20-fold from year 1900 to 1980 (170). Since then, widespread and sometimes misuse of antibiotics in clinical practice and in the food animals has caused bacteria to quickly acquire resistance to antibiotics (31, 41, 79, 167). These days, >70% hospital-acquired bacterial infections show resistance to at least one antibiotic (www.niaid.nih.gov/factsheets/antimicro.htm). For example, resistant strains have been found with *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Enterococci*, *Neisseria gonorrhoeae*, *Salmonella enterica* serovar Typhi (http://www.cdc.gov/drugresistance/diseases.htm), and certain bacteria have become resistant to multiple antibiotics (125).

Furthermore, antibiotic resistance increases the cost of treatment through longer duration of hospitalization and use of expensive drugs to treat infection. The estimated cost of treating antibiotic-resistant infections in the US is $30 billion (101). This cost keeps on rising as the number of bacterial strains resistant to antibiotics are increasing and the pharmaceutical industry is unable to bring novel antibiotics to treat resistant infections (174). Therefore, discovery of novel antimicrobial drugs with a very low risk of resistance development are urgently required. Cationic host defense peptides (HDPs)
constitute a potential novel class of antimicrobials by virtue of their bactericidal as well as immunomodulatory properties (32, 87, 88).

These HDPs are small (12 to 50 amino acids), cationic (+2 to +10 charge due to the presence of multiple arginine and lysine residues), and amphipathic. They are capable of killing Gram-positive and Gram-negative bacteria and fungi and inhibiting the replication of enveloped viruses like influenza A virus, human papilloma virus, vesicular stomatitis virus and human immunodeficiency virus (58, 63). Antibacterial activity of HDPs is exerted by physical disruption of bacterial membranes mainly (15). Due to this physical mechanism of action, cationic HDPs kill a broad range of bacteria including antibiotic-resistant strains with a low risk of developing resistance. This property combined with their modulatory effects on innate and adaptive immune system, have made them very attractive candidates for antibacterial drug development, particularly against antibiotic-resistant strains (43, 44, 87, 197). This review will focus on the antibacterial and immunomodulatory properties of vertebrate HDPs.

CLASSIFICATION OF HOST DEFENSE PEPTIDES

Vertebrates mainly produce two major families of HDPs called defensins and cathelicidins (36, 72, 140, 162, 194, 195). Based on the structure difference, HDPs can be classified into distinct classes such as α-helical, β-sheet, cyclic and extended peptides (87). Defensins are amphipathic molecules characterized by the presence of six conserved cysteine residues, and a number of hydrophobic and cationic residues. Defensins are
further divided into three different subclasses, namely α-, β- and θ-defensins (36, 72, 140). This classification is based on the spacing of cysteine residues and the pattern of disulfide bonds. α-defensins consist of three disulfide bonds between Cys1-Cys6, Cys2-Cys4 and Cys3-Cys5, whereas in β-defensins, cysteines are connected between Cys1-Cys5, Cys2-Cys4 and Cys3-Cys6. α-defensins are mainly produced by leukocytes and Paneth cells of small intestine, whereas β-defensins are expressed mainly by mucosal epithelial cells and skin (36, 72, 140). In humans, there are four neutrophil α-defensins [human neutrophil peptide (HNP)-1 to 4] and two enteric α-defensins [human defensin (HD)-5 and 6] (113, 124, 127). In rhesus monkey, myeloid as well as intestinal α-defensins have been identified (154, 155). While multiple enteric α-defensins are present in mice, neutrophil α-defensins are absent (30, 114). Rat neutrophils express three α-defensins, and mRNA for six additional α-defensins have been identified in small intestine (120). Over 35 β-defensins have been identified in humans and rodents (121, 123).

Tertiary structures of several α- and β-defensins are available (4, 48, 51, 52, 59, 69, 86, 119, 135, 137, 145, 204). All defensins possess triple stranded antiparallel β-sheet with a hairpin conformation stabilized by three disulfide bonds. In the case of human β-defensin (HBD)-1 to 3, mouse β-defensin (MBD)-7, and 8 and penguin sphenecin, there is an additional N-terminal α-helical segment involving the first cysteine (4, 51, 52, 69). However, such a segment is absent in bovine neutrophil β-defensin (BNBD)-12 (204). α-defensins also lack the N-terminal helical segment. Cryptdin-4, which is a potent mouse Paneth cell α-defensin, has a different orientation of β-hairpin and lower β-sheet content
than other defensins. These unique structural features with its high cationicity have been suggested to be responsible for high antibacterial potency of cryptdin-4 (59). All α- and β-defensins possess a conserved GXC motif involving cysteine 4, which forms a bulge essential for correct folding of defensins (182, 191).

β-defensins are cyclic peptides expressed by leukocytes of old world primates (156) and possess activity against a wide spectrum of bacteria, fungi, and viruses (156, 172, 190). They are produced by ligation of two different α-defensin-like peptides with premature termination of the open reading frame between the third and fourth conserved cysteine (156). Each peptide thus contributes three cysteines and cyclizes in a head-to-tail manner with formation of three disulfide bonds (156). NMR structure of rhesus β-defensin (RTD)-1 shows the presence of two flexible β-strands connected by a hairpin at both ends (165). Unlike other classes of host defense peptides, RTD-1 is not amphipathic (165). In humans and chimpanzee, β-defensin genes are pseudogenes with a stop codon in the signal sequence leading to early termination of translation (100, 121).

Cathelicidins constitute another major family of HDPs in vertebrates and are characterized by the presence of an N-terminal conserved pro-sequence called cathelin (162, 194, 196). This cathelin pro-sequence is connected to a 20-80-residue, diversified cationic segment at the C-terminus (162, 194, 196). Both cathelin and C-terminal mature peptide play a role in host defense (195). Cathelicidins are stored in the granules of neutrophils. In addition, they are also expressed by skin and various mucosal epithelial surfaces. To date, cathelicidins have been found in a number of vertebrate species
including primates, glire, canine, porcine, bovine, ovine, caprine, equine, chicken and fish (21, 22, 162, 180). In cattle, sheep, and pigs, multiple cathelicidins are present, whereas in primates and glire only one member is present in each species (162, 200).

Even though all cathelicidins are amphipathic and cationic, they adopt either α-helical, β-sheet, or extended structures. Examples of α-helical peptides include human LL-37, mouse CRAMP, sheep SMAP-29, and chicken fowlicidins (110, 149, 151, 180, 192). Guinea pig CAP-11 possesses a single cysteine towards the C-terminus and exists in the dimeric form due to the formation of an intermolecular disulfide bond (98). Some cathelicidins are rich in certain amino acids, which render them extended structures. Extended peptides include bovine indolicidin (Trp and Pro rich), porcine PR-39 (Pro and Arg rich), and porcine prophenin (Pro and Phe rich) (18, 45, 68). Porcine protegrins and bovine and ovine dodecapeptides are looped β-sheet peptides due to the presence of intramolecular disulfide bonds (66, 177). Protegrins consist of four cysteines forming two intramolecular disulfide bonds, whereas bovine and ovine dodecapeptides have two cysteines each forming one disulfide bond.

**PLEOTROPIC EFFECTS OF HOST DEFENSE PEPTIDES**

Defensins and cathelicidins are either produced constitutively or induced in response to infection or injury (194, 195). All these released HDPs exert a host defense role by directly killing a broad spectrum of bacteria. Recent evidence also shows a novel way of antibacterial action, in which macrophages acquire granules of apoptotic
neutrophils, thereby acquiring HDPs to kill intracellular *Mycobacterium tuberculosis* (152).

Studies involving knock-out and transgenic mice have provided convincing evidence showing an indispensable role of HDPs in protection against infections. Mice deficient in *matrix metalloproteinase 7* (*MMP-7*), which is responsible for conversion of inactive precursors of enteric α-defensins into active forms, become more susceptible to a lethal *S. typhimurium* infection (175). Similarly, mice deficient in β-defensin-1 (*Defb1*) show decreased clearance of *Staphylococci* in the bladder (91) and exhibit enhanced colonization of *H. influenzae* in the lung (92). Role of cathelicidin (mouse CRAMP) in protecting skin infection from Group A *Streptococci* is evident by the presence of large and persistent skin lesions in mice lacking *mCRAMP* (105). Conversely, transgenic mice possessing porcine cathelicidin *PR-39* display increased resistance to skin infections by Group A *Streptococci* (71). Transgenic mice expressing human α-defensin HD-5 are resistant to a lethal challenge of *S. typhimurium* infection (132).

Different models have been proposed to explain the mechanism of antibacterial action of HDPs (15). Initial interaction of cationic peptide occurs with negatively charged molecules on bacteria such as lipopolysaccharide (LPS) of Gram-negative bacteria and lipoteichoic acid of Gram-positive bacteria, followed by formation of either transient or stable pores across membranes. Rabbit neutrophil defensins form short-lived pores, whereas human neutrophil defensins form stable channels (60, 77, 176). The formation of stable channels correlates with oligomerization of human neutrophil defensins in contrast
to the monomeric form adopted by rabbit neutrophil defensins. Mouse cryptdin-4 interacts with phospholipid head groups through charge-charge interactions between the cationic peptide and negatively charged membrane lipids and disrupts bilayer head groups (133, 134). Rhesus θ-defensin-1 (RTD-1), which lacks an amphipathic structure, has been shown to induce membrane thinning and formation of pores across membranes (17, 173). Additionally, defensins can also interfere with intracellular processes. Human neutrophil peptides 1-3 (HNP 1-3) completely inhibit the synthesis of DNA, RNA and protein at sub-lethal concentrations (73). Killing of bacteria by defensins can also occur through activation of cell-wall lytic enzymes (131).

In the case of cathelicidins, the mechanism of action on model membranes is studied in more detail than that for defensins. Human LL-37 disrupts model membranes through formation of transmembrane torroidal pores lined by polar lipid head groups and peptides (47). Porcine protegrin-1, a β-sheet cathelicidin, also permeabilizes model membranes through formation of torroidal pores (82, 83, 183). Ovispirin, a peptide derived from ovine cathelicidin SMAP-29, permeabilize bacterial membranes through the “carpet” mechanism by binding to lipid bilayer in parallel, leading to the formation of micelles when reaching the threshold concentrations (184). Cathelicidins also kill bacteria by interfering with intracellular processes. Porcine PR-39 kills bacteria by inhibition of bacterial DNA and protein synthesis (8). PR-26, an N-terminal version of PR-39, causes filamentation of S. typhimurium, suggesting the inhibition of cell division (142). Bovine indolicidin also inhibits DNA synthesis resulting in filamentation of
Escherichia coli (150). Indolicidin has been shown to directly bind HIV-1 DNA through covalent cross-linking, resulting in inhibition of integrase activity (84).

In addition to direct bacterial killing, HDPs exert important modulatory activities on the innate and adaptive immune system. Defensins and cathelicidins possess chemotactic activities on various immune cells. HNPs are chemotactic for monocytes, T cells, and immature dendritic cells (24, 158, 185). The identities of receptors used by HNPs are not clear. HBD-2 is chemotactic for immature dendritic cells, memory T cells, TNF-α-treated neutrophils, and mast cells through interactions with CCR-6 chemokine receptor (102, 186). HBD-3 and HBD-4 are chemotactic for monocytes (37, 38). Similarly, murine β-defensins are also capable of attracting immature dendritic cells and help induce antitumor immunity (6). LL-37 is chemotactic for neutrophils, monocytes, T cells and mast cells (29, 103). PR-39 is a chemoattractant for neutrophils (55). Bovine-bactenecin derived peptide, Bac2A, is a chemoattractant for monocytes (12). Mouse CRAMP is a chemoattractant for neutrophils and monocytes (67). Cathelicidins induce chemotaxis by binding to formyl peptide receptor. Recruitment of immune cells to the site of infection leads to elicitation of cell-mediated and humoral immune responses to fight off infection.

HDPs are also able to induce the release of cytokines and chemokines from immune cells and epithelial cells. Human neutrophil defensins up-regulate chemokine IL-8 production by human airway epithelial cells (168). Cryptdin-3, but not cryptdin-4, induces secretion of IL-8 from human intestinal cells (76). In peripheral blood
mononuclear cells (PBMCs), HBD-2 induces IL-8, IL-6, IL-10, MCP-1, MIP-1β, RANTES, IL-1β, ENA-78, GRO, IGFBP-3 and EGF, whereas HBD-1 induces IL-8, IL-10, IL-6 and MIP1β, and HBD-3 stimulates IL-8, MCP-1 and IGFBP-3 production (9). HBD-2, 3 and 4, but not HBD-1, induces expression of IL-6, IP-10, IL-10, RANTES, MCP-1, MIP-3α at both mRNA and protein levels in skin keratinocytes (104). LL-37 induces secretion of IL-8, IL-6, IL-1β, and TNF-α in human corneal endothelial cells (56) and IL-1α, IL-6, IL-8, TNF-α, and GM-CSF in keratinocytes (13). LL-37 also enhances IL-8 release from human airway smooth muscle cells (205). Thus, chemokines and proinflammatory cytokines released by different cell types under influence of HDPs might help recruit immune cells and modulate the microenvironment, ultimately leading to robust immune responses.

Given the immunostimulatory activities exhibited by HDPs, one would expect that they have adjuvant-like effects on the immune system. Indeed, both defensins and cathelicidins potentiate immune responses to particularly non-immunogenic antigen. HNPs administered together with an antigen induce antigen-specific IgG response in mice (75). HNPs help induce tumor-specific IgG and T cell responses, thereby increasing resistance of mice to a tumor challenge (157). HBD-1 also induces antigen-specific IgG response in mice (16). A DNA vaccine encoding a fusion protein of non-immunogenic lymphoma antigen with mouse β-defensin-2 and 3 generates tumor-specific protective humoral and cell-mediated immunity (6). Similarly, LL-37, when fused with a tumor antigen and administered as a DNA vaccine, generates humoral and cellular immune responses and increases survival against a tumor challenge (3). However, the mechanism
by which these HDPs exert adjuvant-like effects remains elusive (10). It is likely that defensin and cathelicidins act as a carrier and target the fusion partner to antigen-presenting cells by binding to chemokine receptors. It is also possible that these HDPs alter the micro-environment by inducing cytokines and chemokines, thereby influencing differentiation and maturation of dendritic cells indirectly. The third scenario is that HDPs directly interact with dendritic cells and modulate their differentiation. Nevertheless, further research is needed to study the immunomodulatory effect of HDPs to facilitate the development of HDPs as adjuvants for vaccine use (5).

Conversely, some HDPs are also able to suppress the production of proinflammatory cytokines such as TNF-α and IL-1β in response to LPS stimulation. Cathelicidins, including human LL-37, bovine indolicidin, rabbit CAP18, and chicken fowlicidins, are thought to neutralize the effects of LPS through direct binding to and competing with LPS for its receptor, CD14 (128, 180). However, recent evidence indicates that LL-37 and bovine BMAP-27 suppress LPS-induced responses by inhibiting the transcription of proinflammatory genes (89, 90). Both peptides reduce LPS-induced translocation of NF-κB subunits, p65 and p50, into nucleus. Thus, overall neutralization of LPS effects seems to occur through multiple mechanisms (89). Protection against LPS-induced sepsis by cathelicidins has been demonstrated in animal models (26, 33, 39, 40). In contrast, only one human β-defensin, DEFB-123, has been found to prevent LPS-induced TNF-α secretion in vitro and protect mice against LPS-induced septic shock (93).
In addition to endotoxins, several defensins were recently shown to be able to neutralize several types of exotoxins. HNPs neutralize anthrax lethal toxin and protect mice against lethal effects of the toxin (61). HNP-1 inhibits lethal toxin by inhibiting the cleavage of mitogen-activated protein kinase kinase by toxin. HNPs neutralize diphtheria toxin and \textit{Pseudomonas} exotoxin A by competitively inhibiting toxin-mediated ADP-ribosylation of eukaryotic elongation factor 2 (62). \(\theta\)-defensins also neutralize anthrax lethal toxin, in addition to killing bacilli and germinating spores of \textit{Bacillus anthracis} (171).

HDPs also play a role in the resolution of inflammation and wound healing. HNPs stimulate proliferation of epithelial and fibroblast cells (1, 96) and enhance wound closure of the airway (2). HBD-2, -3 and -4 promote wound healing by promoting keratinocyte migration and proliferation through binding to epidermal growth factor receptor (EGFR) (104). Cathelicidin LL-37 promotes angiogenesis and neovascularization of endothelial cells (65) and is involved in re-epithelization of skin wounds (46). LL-37 induces keratinocyte migration through transactivation of EGFR (161). LL-37 also stimulates airway epithelial cell proliferation and wound closure through EGFR (141). Rat CRAMP promotes healing of gastric ulcers by inducing epithelial cell proliferation through TGF-\(\alpha\)-dependant transactivation of EGFR (189). Porcine PR-39 facilitates the formation of new functional blood vessels in a mouse model through inhibition of proteasome-mediated degradation of hypoxia induced factor 1-\(\alpha\) protein (74). PR-39 also helps in wound healing through induction of syndecan-1 and 4, which
are important in regulating cell proliferation and migration in response to heparin binding growth factors (35).

**STRUCTURE-ACTIVITY RELATIONSHIPS**

Structural and physicochemical properties such as disulfide bond, charge, length, helicity, hydrophobicity and amphipathicity play a critical role in antibacterial, immunomodulatory, and cytotoxic activities of HDPs. It is possible to design peptides with improved antibacterial and immunomodulatory activities and reduced cytotoxicity for therapeutic application by varying one or more of the above properties.

In the case of defensins, conserved disulfide bonds are dispensable for the antibacterial activity. Defensin variants lacking disulfide bonds or having a varying number of disulfide bonds display similar antibacterial activities to the parent peptide (53, 64, 80, 81, 178). Absence of disulfide bonds also does not affect the cytotoxicity of defensins (64). However, the disulfide motif is required for *in vivo* stability of defensins. Absence of disulfide bonds renders cryptdin-4 susceptible to MMP-7, a protease which converts cryptdin-4 precursors into active forms (80, 143). Another function of disulfide bonds in defensins is to maintain the chemotactic activity as observed for HBD-3 (178). Furthermore, even with intact disulfide bonds, mutations in N-terminal helix containing primarily hydrophobic residues of HBD-1 result in a significant reduction of its binding to the CCR-6 receptor, underscoring the importance of these residues in chemotaxis (122). A conserved salt bridge in HNP-2, which is dispensable for the antibacterial
activity, is necessary to maintain its stability against protease degradation (179). The same conserved salt bridge in cryptdin-4 is not required for the antibacterial activity or proteolytic stability (129).

Antibacterial activity of defensins is mainly dictated by net positive charges. In the case of cryptdin-4, charge reversal from cationic arginines to anionic aspartic acids abolishes the antibacterial activity, irrespective of the location of arginine (153). There is a direct positive correlation between the number of charged residues and antibacterial activity as seen with HBD-1 (122). A combination of strong cationicity and moderate hydrophobicity is necessary for the antibacterial activity of HBD-3, whereas high hydrophobicity results in an increase in cytotoxicity to mammalian cells (64). Arginine residues in defensins also have a role in modulating in vivo functions. ADP ribosylation of arginines of HNP-1 in the respiratory tract results in a decreased cytotoxicity and antibacterial activity without altering the chemotactic and IL-8-inducing ability (116).

In the case of α-helical peptides, an appropriate balance of the net positive charge, α-helical content and amphipathicity dictate the antibacterial activity (28, 163). Recently, this was demonstrated with a group of de novo synthesized peptides based on the template obtained by comparing naturally occurring α-helical peptides (198, 199, 201). These synthetic peptides show reduced potency when the net positive charge is decreased to +3 or below. Increasing the charge above +6 fails to enhance the antibacterial potency. Additionally, peptides that are both α-helical and amphipathic have broad-spectrum antibacterial activity. However, peptides with a rigid α-helix possess a strong hemolytic
activity and a poor antibacterial activity. An increase in the flexibility of α-helix by introducing a glycine or proline is generally associated with a decrease in cytotoxicity. Peptides also display a decreased cytotoxicity with moderate hydrophobicity, and a further decrease in hydrophobicity reduces the antibacterial activity.

Most of the above described principles are also true for naturally occurring α-helical cathelicidins. In the case of ovispirin, which is the N-terminal α-helical segment of SMAP-29 with strong hemolytic and cytotoxic activities (136), a replacement of either isoleucine 10 with glycine or isoleucine 7 with threonine enhances the flexibility and substantially reduces hemolytic and cytotoxic activities (136). In contrast, despite being a rigid helical peptide, CAP-18 shows a low hemolytic activity. Antibacterial activity of CAP-18-derived peptides with moderate to high helicity correlates well with the net positive charge than helicity (23, 166). CAP-18-derived peptides with very low helicity exhibit reduced antibacterial activity, despite introductions of additional positively charged residues (164). Some peptides contain a hinge around the center of α-helix consisting of PXXP or PXXXP motif. This hinge divides the peptide into the N- and C-terminal helices. When the N-terminal helix binds to membrane, the hinge bends and allows the insertion of the C-terminal helix into lipid bilayer (187). In the case of highly hydrophobic BMAP-27 and 28, maintenance of the N-terminal amphipathic α-helix while reducing hydrophobicity diminishes the cytotoxicity without changing the antibacterial activity (146). Removal of the four N-terminal hydrophobic residues of LL-37 gives rise to a less cytotoxic derivative without affecting membrane permeabilization or the antibacterial activity (110).
Successful attempts have been made to identify short fragments of naturally occurring α-helical peptides with improved antibacterial and immunomodulatory activities and less cytotoxicity. LL-37-derived peptides with N-terminal hydrophobic amino acids removed, display similar antibacterial, LPS-neutralizing, and chemotactic activities as LL-37, but with significantly reduced cytotoxicity (25, 99, 144). Recently, a large panel of LL-37-derived short fragments were compared for their antibacterial and immunostimulatory activities (13). Two LL-37 fragments, namely RK-31 and KS-30, possess a dramatically improved antibacterial activity with little capacity to stimulate the release of IL-8 from keratinocytes. These two fragments also display slightly reduced toxicity. Interestingly, both fragments have been reported to be present in human sweat, suggesting this gain of function demonstrated in vitro has in vivo implications (95).

For guinea pig CAP-11 peptide, the antibacterial activity of the N-terminal fragment of 18 residues can be further improved by increasing amphipathicity (109). Interestingly, the C-terminal α-helix of chicken fowlicidin-1 is the major region for antibacterial, cytotoxic and LPS-neutralizing activities with three hydrophobic residues just before N-terminal α-helix also sharing cytotoxic and LPS-binding activities. Removal of these three hydrophobic residues results in an almost 10-fold reduction in cytotoxicity and 60-fold reduction in hemolysis without significantly affecting LPS-neutralizing and antibacterial activities (181). Angiogenic and NF-κB-inhibiting activities of PR-39 have been attributed to the N-terminal 11 residues with three arginines being essential for both activities (34). Furthermore, the N-terminal arginines of PR-39 are critical for the antibacterial activity, induction of syndecans and interaction with the SH3
domain (20), whereas the C-terminal part of PR-39 is important only for antibacterial activity but not for interactions with the SH3 domain in mammalian cells (20).

**STRATEGIES FOR PEPTIDE DESIGN AND SYNTHESIS**

Although HDPs have broad-spectrum antibacterial activities, the issues related to antibacterial potency, stability, toxicity, cost and salt sensitivity need to be addressed in order to develop them into therapeutic drugs. A number of strategies for peptide design and synthesis are evolving. Besides structure-based rational design as depicted in the previous section, other strategies include amidation, acylation, cyclization, dimerization, D-enatiomerization, development of peptide mimetics, and high throughput synthesis and screening of synthetic peptides.

Various types of modifications of peptide termini, backbone or amino acid residues are being tried to improve the antibacterial activity and stability of peptides. Amidation of the C-terminus has been shown to increase the antibacterial activity of porcine cathelicidin, tritrpticin (138, 188). N-terminal acylation or C-terminal amidation also protects peptide from being cleaved by exopeptidases. Such *de novo* designed peptides with both termini modified retain the antibacterial activity (99). Peptide backbone can be modified by introducing carbamate bonds or reduced amide bonds (70, 107). Reduced amide bond increases the net positive charge and antibacterial activity. Introduction of carbamate bonds reduces helicity and the hemolytic activity. Another modification of backbone is the inclusion of peptoid residues, which are helix-disrupting...
amino acids. Such modifications can decrease the cytotoxicity of HDPs with increased activity towards bacteria (147, 203). Backbone modification also increases half lives of peptides in serum as they become less susceptible to proteolysis (107). Incorporation of unnatural amino acids into synthetic antibacterial peptides has been tried successfully to increase the resistance of peptides to serum proteases without affecting the antibacterial activity (108).

Incorporation of D-amino acids into HDPs is another promising approach to achieve desirable biological properties. Diastereomeric synthetic peptides containing D-amino acids tend to be less hemolytic and cytotoxic while retaining the antibacterial activity (115, 117, 148, 201). Consistently, diastereomeric peptides are incapable of inserting into the lipid bilayer mimicking mammalian cell membranes, whereas peptides containing only L-amino acids do so readily (118). D-amino acids modulate peptide cytotoxic activity by changing helicity and hydrophobicity. Furthermore, diastereomers maintain the antibacterial activity in serum and are resistant to degradation by proteases (117). Such diastereomers, but not L-amino acid peptides, protect mice against antibiotic-sensitive as well as -resistant infections, suggesting improved stability of diastereomers in vivo (14).

Cyclization of certain peptides has been shown to increase the stability of peptides in vivo (42). Cyclic peptides with alternating D- and L-amino acids are resistant to proteases and highly stable when administered in vivo (27). Similarly, cyclo CP-11, a peptide derived from bovine indolicidin, has a prolonged half-life in the presence of
trypsin as compared to linear CP-11 (130). Cyclization also enhances the antibacterial activity and reduces the hemolytic activity in several cases. A cyclic peptide with D- and L-amino acids shows reduced hemolytic activity and broad spectrum antibacterial activity relative to the linear analog (111). It correlates with an increased ability of the cyclic peptide to interact with and permeabilize negatively charged membranes and reduced affinity for zwitterionic membranes. Cyclization also renders peptides resistant to salt. The most convincing examples of salt-insensitive cyclic peptides are θ-defensins. These peptides, as discussed in the previous section, are naturally occurring cyclic defensins of old world primates. Antibacterial activity of cyclic RTD-1 is insensitive to salt. However, such an activity is completely lost in the presence of physiological concentrations of salt when peptide is linearized (156). De novo designed cyclic peptides, resembling θ-defensins and porcine protegrins, have also been shown to act in a salt-insensitive manner (94). Even naturally occurring HDPs can be cyclized to acquire salt-insensitive antibacterial activities. For example, end-to-end cyclized rabbit neutrophil peptide-1 (NP-1) is able to maintain its activity in the presence of salt, whereas wild-type NP-1 is inactivated (193).

Dimerization of peptides through covalent linkage by disulfide bonds may also increase antibacterial potency and render peptides salt-insensitive. Evidence for this argument comes from naturally occurring peptide dimers. Dimerization of cryptdin-related CRS peptides in mice increases antibacterial potency (54). Dimeric Defr1, a mouse β-defensin-related peptide, retains the antibacterial activity in the presence of physiological salt concentrations (19). Cathelicidins, PMAP-36 and CAP-11, also exist
in dimeric forms in neutrophils; however, dimerization does not increase antibacterial potency (109, 139), although it modulates the kinetics of killing and membrane permeabilization in the case of PMAP-36 (139). Increasing evidence suggests a positive association between dimerization and the protease stability of peptides (126). Cross linking of two tryptophans in indolicidin confers protease resistance without altering the antibacterial activity (112).

Synthetic combinatorial libraries (SCL) have also been employed to identify small peptides (hexapeptides) with the antibacterial activity (7). However, constructing such libraries is cost-prohibitive especially for longer peptides. Another approach to produce a large number of peptides is high throughput, in situ peptide synthesis on a cellulose support by pipetting robots (50). Peptides are subsequently cleaved from the support and the antibacterial activity is studied in a high throughput manner against *Pseudomonas aeruginosa* expressing a luciferase reporter (50). It is estimated that 100,000 individual peptides can be produced by this approach using two robots and screened for antibacterial activities in a year. Using this approach, bovine bactenecin derivatives with increased antibacterial potency have been obtained (49). A recent report described the rational design of antimicrobial peptides based on a linguistic model using naturally occurring HDPs (78). This language is based on grammars that use common sequence patterns found in natural peptides. Using this strategy, a large number of de novo peptides can be synthesized and screened for the antibacterial activity.
Abiotic oligomeric and polymeric compounds are also being synthesized to mimic the structural properties of HDPs. Such compounds are resistant to proteases. For example, oligomeric compounds like acrylamide and oligourea and synthetic polymers like phenylene ethynylene are able to fold into amphipathic structures. These compounds possess antibacterial activities against a broad range of bacteria including antibiotic-resistant bacteria (106, 159, 160, 169). Theses abiotic compounds are selectively active against bacterial membranes. They insert themselves into the bacterial model membranes and are less selective towards mammalian model membranes (57). Additional investigations are warranted to study the \textit{in vivo} toxicity and antibacterial efficacy of these compounds.

\textbf{CURRENT THERAPEUTIC APPLICATION AND CLINICAL DEVELOPMENT}

Due to a range of beneficial effects conferred by HDPs including broad-spectrum antibacterial activity and immunomodulatory effects, efforts are ongoing to develop them as novel antibiotics to treat infections. The companies involved in their commercial development are mostly small biotechnology companies. Their efforts have been met with mixed success. A number of peptides are being developed in various stages of clinical trials and are yet to be approved by FDA. The diseases against which peptides are developed include ventilator-associated pneumonia, in-dwelling catheter-associated infections, oral candidiasis, sepsis, cystic fibrosis, acne, and bacterial meningitis (85, 202).
The most successful peptide to date is bovine indolicidin-based omiganan pentahydrochloride 1% gel, also known as MX-226, which has been shown to reduce catheter-associated infections by 49% in a phase III clinical trial (58, 85). Another indolicidin-based peptide, MX-594AN, significantly reduced severity of acne lesions in a phase II trial (58, 85) (www.migenix.com). Pexiganan, a cationic peptide based on frog magainin, was the first peptide to enter clinical trials for treatment of diabetic foot ulcers. However, FDA rejected its approval despite its effectiveness, since it did not confer any additional benefits than treatments currently available to treat diabetic foot ulcers (85). Iseganan, a peptide based on porcine protegrin, did not provide obvious benefits in ventilator-associated pneumonia and was withdrawn from clinical trials (85).

Compared to other HDPs, not even a single defensin has reached clinical trials. The disadvantages associated with defensins are low potency, salt sensitivity, and significant costs in producing disulfide-containing peptides on a large scale. However, plectasin, a novel defensin from fungus *Pseudoplectania nigrella*, can be produced and purified on an industrial scale using a fungal expression system (97). Plectasin has a half life of 50 min in serum, presumably due to increased stability towards proteases provided by disulfide bonds. Furthermore, it protects mice from experimental peritonitis and pneumonia (97).

Mixed successes are not deterring companies from developing HDPs as antibacterial therapeutics. However, the trend is to develop smaller versions of peptides, which are cheaper to synthesize. For example, Helix BioMedix is developing short
hexapeptides to treat bacterial and fungal infections (202). Also, antibacterial activities of HDPs can be separated from immunomodulatory activities as discussed in the previous section. It will not be surprising if companies start utilizing immunomodulatory fragments, rather than antibacterial fragments for clinical development. Inimex pharmaceutical has identified a lead compound based on LL-37. This lead peptide lacks antibacterial activity, but is able to protect against infection in animal models due to its immunomodulatory activities (11). Helix BioMedix is developing HB107, a peptide based on cecropin, for wound healing rather than anti-infective purposes (202). Some companies are even developing peptide mimetics and oligomeric compounds to treat infections, since such drugs show more stability in serum than HDPs due to their resistance to proteases (202).

**CONCLUSION AND FUTURE PROSPECT**

HDPs possess antimicrobial as well as immunomodulatory activities. They have tremendous potential as novel anti-infective drugs. However, there is a need to address issues such as antibacterial potency, toxicity, stability, salt sensitivity, and cost of production. A variety of strategies are emerging to tackle these issues including structure-based rational design, amidation, cyclization, dimerization, amino acid modification, and peptide mimetics. Another important issue is the length of peptide as it is directly related to the production cost. Efforts should be directed to make HDPs as small as possible without compromising their biological activities. In addition to recombinant production, making flocks of transgenic animals producing a large amount of peptides is a practical
approach to produce HDPs. Taken together, identification of novel HDPs, study of their structure activity relationship, and optimal peptide design will help drive anti-infective peptides into clinics.

REFERENCES


22. **Chang CI, Zhang YA, Zou J, Nie P, and Secombes CJ.** Two cathelicidin genes are present in both rainbow trout (Oncorhynchus mykiss) and atlantic salmon (Salmo salar). *Antimicrob Agents Chemother* 50: 185-195, 2006.


35. **Gallo RL, Ono M, Povsic T, Page C, Eriksson E, Klagsbrun M, and Bernfield M.** Syndecans, cell surface heparan sulfate proteoglycans, are induced by a proline-rich antimicrobial peptide from wounds. *Proc Natl Acad Sci USA* 91: 11035-11039, 1994.


56. **Huang LC, Petkova TD, Reins RY, Proske RJ, and McDermott AM.**


101. **Niederman MS.** Impact of antibiotic resistance on clinical outcomes and the cost of care. 

102. **Niyonsaba F, Iwabuchi K, Matsuda H, Ogawa H, and Nagaoka I.** Epithelial cell-
derived human beta-defensin-2 acts as a chemotaxin for mast cells through a pertussis

    A cathelicidin family of human antibacterial peptide LL-37 induces mast cell

104. **Niyonsaba F, Ushio H, Nakano N, Ng W, Sayama K, Hashimoto K, Nagaoka I, Okumura K, and Ogawa H.**
    Antimicrobial Peptides Human beta-Defensins Stimulate
Epidermal Keratinocyte Migration, Proliferation and Production of Proinflammatory

    Innate antimicrobial peptide protects the skin

106. **Nusslein K, Arnt L, Rennie J, Owens C, and Tew GN.** Broad-spectrum antibacterial

107. **Oh JE, Hong SY, and Lee KH.** Design, synthesis and characterization of antimicrobial
pseudopeptides corresponding to membrane-active peptide. *J Pept Res* 54: 129-136,
1999.

108. **Oh JE and Lee KH.** Synthesis of novel unnatural amino acid as a building block and its


CHAPTER III
RAPID EVOLUTION AND DIVERSIFICATION OF MAMMALIAN
\( \alpha \)-DEFENSINS AS REVEALED BY COMPARATIVE ANALYSIS OF RODENT
AND PRIMATE GENES

Amar Patil\(^1\), Austin L. Hughes\(^2\), and Guolong Zhang\(^1\)

\(^1\)Department of Animal Science, Oklahoma State University, Stillwater, OK 74078;
\(^2\)Department of Biological Sciences, University of South Carolina, Columbia, SC 29208.

The following chapter has been published in the *Physiological Genomics* 2004, 20:1-11
and used with permission of American Physiological Society.
ABSTRACT

Mammalian α-defensins constitute a family of cysteine-rich, cationic antimicrobial peptides produced by phagocytes and intestinal Paneth cells, playing an important role in innate host defense. Following comprehensive computational searches, here we report the discovery of complete repertoires of the α-defensin gene family in the human, chimpanzee, rat, and mouse with new genes identified in each species. The human genome was found to encode a cluster of 10 distinct α-defensin genes and pseudogenes expanding 132 kb continuously on chromosome 8p23. Such α-defensin loci are also conserved in the syntenic chromosomal regions of chimpanzee, rat, and mouse. Phylogenetic analyses showed formation of two distinct clusters with primate α-defensins forming one cluster and rodent enteric α-defensins forming the other cluster. Species-specific clustering of genes is evident in non-primate species, but not in the primates. Phylogenetically distinct subsets of α-defensins also exist in each species with most subsets containing multiple members. In addition, natural selection appears to have acted to diversify the functionally active mature defensin region but not signal or prosegment sequences. We concluded that mammalian α-defensin genes may have evolved from two separate ancestors originated from β-defensins. The current repertoire of the α-defensin gene family in each species are primarily a result of repeated gene duplication and positive diversifying selection after divergence of mammalian species from each other, except for the primate genes, which were evolved prior to the separation of the primate species. We argue that the presence of multiple, divergent subsets of α-defensins in each
species may help animals to better cope with different microbial challenges in the ecological niches which they inhabit.

**INTRODUCTION**

The defensins of vertebrates constitute a large family of cationic antimicrobial peptides that are characterized by the presence of six cysteine residues at defined positions. According to the spacing pattern of cysteine residues, these defensins are further classified into three subfamilies, namely α-, β- and θ-defensins with each consisting of a unique six-cysteine motif and a disulfide bonding pattern (6, 16, 21, 29). For example, the consensus α-defensin motif is C-X1-C-X3-4-C-X9-C-X6-10-C-C, where C1-C6, C2-C4 and C3-C5 form three intramolecular disulfide bridges. While β-defensins have been found in most vertebrate species with a much wider tissue expression pattern, α-defensins are specific to mammals and are mainly produced by leukocytes of myeloid origin and Paneth cells of small intestine (6, 16, 21, 29). On the other hand, θ-defensins are believed to have evolved from α-defensins and have only been discovered in leukocytes of primates (6, 16, 20).

Defensins possess potent antimicrobial activity against a broad range of bacteria, fungi, and enveloped viruses (6, 16, 21, 29). In addition, certain defensins are also capable of inducing maturation of dendritic cells and spermatocytes, inducing secretion of chloride and proinflammatory cytokines from epithelial cells, and chemoattracting immune and inflammatory cells (6, 16, 40). The mechanism by which defensins kill
microbes involves initial electrostatic interactions of cationic peptides with negatively charged phospholipids on the microbial membranes, followed by permeabilization of bacterial membrane and cell death (6). In contrast to most conventional antibiotics that kill microbes by certain biochemical mechanisms, such a physical mechanism of action confers on defensins a broad-spectrum antimicrobial activity with equal efficacy against antibiotic-resistant strains and little risk of developing resistance (8, 43). Consequently, defensins are being explored as a new class of antimicrobials mainly to control antibiotic-resistant microorganisms.

To date, a number of α-defensins have been discovered in human, rhesus macaque, rabbit, guinea pig, mouse, and rat (6, 16, 21). For example, three myeloid α-defensin genes (DEFA1/2, DEFA3 and DEFA4) and two enteric α-defensin genes (DEFA5/HD5 and DEFA6/HD6) have been found in humans, and a group of 19 highly homologous enteric α-defensins (cryptdins) have also been reported in mice. All α-defensins are synthesized from 80-105 amino acid precursors with each composed of a short N-terminal signal sequence (~20 amino acids), an anionic prosegment (40-50 amino acids), and a C-terminal cationic mature peptide (30-35 amino acids) (6, 16, 21). The α-defensins are stored as either inactive proforms or mature active forms in the cytoplasmic granules of phagocytes and Paneth cells and released (and processed if necessary) in response to microbial infection or cholinergic stimulation (6, 16, 21). The enzyme responsible for processing and activation of intestinal α-defensins is metalloproteinase matrilysin (MMP7) in mice (38) and trypsin in humans (7), respectively. The contribution of α-defensins to intestinal mucosal immunity has been best demonstrated by the
evidence that transgenic mice expressing human \textit{DEFA5} or \textit{HD5} became fully resistant to oral lethal infection with \textit{Salmonella typhimurium} (27) and that MMP7-deficiency rendered mice more susceptible to \textit{S. typhimurium} infection (38).

The genomes of several evolutionarily divergent vertebrate species have been sequenced, including zebrafish, Japanese pufferfish (\textit{Fugu rubripes}), chicken, dog, rat, mouse, chimpanzee and human. Availability of such a large amount of sequence information provides a timely opportunity to search for the possible existence of \(\alpha\)-defensins in these species. We hypothesized that identification of the entire repertoire of \(\alpha\)-defensin genes across a range of animal species will reveal the origin and evolutionary relationships of this important gene family. This also serves as a first step to study the role of \(\alpha\)-defensins in host immunity and explore their potential as novel antimicrobials. Following comprehensive genome-wide screening, here we report identification of the complete repertoires of \(\alpha\)-defensin genes in the human, chimpanzee, mouse, and rat with novel sequences being discovered in each species. However, none of \(\alpha\)-defensins have been found in any species other than glires (mouse, rat, guinea pig, and rabbit) and primates (human, chimpanzee, olive baboon, and rhesus macaque). In addition, we provide strong evidence showing that a rapid duplication and positive diversifying selection of \(\alpha\)-defensin genes have occurred following divergence of mammalian species.
MATERIALS AND METHODS

Computational search for novel α-defensins

To identify potential novel sequences in all vertebrate species with whole genome sequences available, all known α-defensin peptide sequences were individually queried against expressed sequence tags (EST), nonredundant sequences (NR), unfinished high throughput genomic sequences (HTGS), and whole genome shotgun sequences (WGS) in the GenBank by using the TBLASTN program (1) with the default settings on the NCBI web site (http://www.ncbi.nlm.nih.gov/BLAST). The vertebrate species examined included zebrafish, Japanese pufferfish, chicken, dog, rat, mouse, chimpanzee, and human. All potential hits were then examined for presence of the characteristic α-defensin motif or highly conserved signal/prosegment sequence. For every novel α-defensin sequence identified, additional iterative BLAST searches were performed as described above until no more novel sequences can be revealed. Because mammalian defensins tend to form clusters (17, 30, 31), all genomic sequences containing α-defensins were also retrieved from the GenBank in order to discover potential novel sequences with distant homology. The nucleotide sequences between two neighboring defensin genes were translated into six open reading frames and individually compared with the two defensin peptide sequences for the presence of α-defensin motif or signal/prosegment sequence by using the BLASTP (1) on the NCBI web site (http://www.ncbi.nlm.nih.gov/blast/bl2seq/) and/or ClustalW program (version 1.82) (35) on the European Bioinformatics Institute web site (http://www.ebi.ac.uk/clustalw).
Prediction of full-length coding sequences and genomic structures of α-defensins

All known α-defensin precursors are encoded in two separate exons separated by a short intron of less than 2 kb with one exon encoding signal/prosegment sequence and the other exon encoding the mature peptide containing the six-cysteine α-defensin motif (22). If either the signal/prosegment sequence or α-defensin motif of a novel gene was missing in a genomic sequence within a 2-kb distance, an approximately 5-kb sequence flanking the α-defensin motif or signal/prosegment sequence was retrieved to identify the full-length coding sequence and to derive the structural organization of that novel α-defensin gene by using a combination of GenomeScan (41), GENSCAN (4) and/or GeneWise (3). In the case of the rat, all computational predictions of coding sequences, except for the pseudogenes, have also been confirmed by cloning and sequencing of their respective RT-PCR products amplified from appropriate tissues.

Identification and characterization of α-defensin gene clusters

To determine the relative position and orientation of each defensin in the genome, individual defensins were searched against the assembled genomes of human (NCBI Build 35), chimpanzee (NCBI Build 1 Version 1), mouse (NCBI Build 33), and rat (BCM Version 3.1) released in May 2004, November 2003, May 2004, and June 2003, respectively. The BLAT program (14) was used for gene mapping through the UCSC Genome Browser (http://genome.ucsc.edu). The chromosomal locations of the α-defensin gene clusters of human, mouse, and rat were revealed by using the Map Viewer Program (http://www.ncbi.nlm.nih.gov/mapview).
Sequence alignment and molecular evolutionary analysis

Multiple sequence alignments were carried out by using the ClustalW program (version 1.82) (35) (http://www.ebi.ac.uk/clustalw). The neighbor-joining method (26) was used to construct the phylogenetic tree by calculating the proportion of nucleotide or amino acid differences (p-distance), and the reliability of each branch was tested by 1000 bootstrap replications. Pairwise comparisons of nucleotide sequences at the codon level was carried out by using the method of Nei and Gojobori (19) to estimate the number of nonsynonymous substitutions per nonsynonymous site (dN) and the number of synonymous substitutions per synonymous site (dS) with the Jukes-Cantor correction for multiple substitutions. Construction of the phylogenetic tree and pairwise comparison of nucleotide or amino acid sequences were carried out by using the MEGA software version 2.1 (15).

RT-PCR analysis of tissue expression patterns of α-defensins

A total of 26 different tissue samples were harvested from healthy, 2-month old Sprague-Dawley rats. Bone marrow progenitor cells were collected from femur, followed by hypotonic lysis of erythrocytes as described (44, 45). Only lineages of white blood cells were used for RNA isolation. Total RNA was extracted using Trizol (Invitrogen, Carlsbad, CA). A panel of 12 human tissue RNA, including bone marrow, small intestine, kidney, trachea, lung, liver, heart, spleen, skeletal muscle, testis, prostate, and uterus, were also purchased from BD Biosciences Clontech (Palo Alto, CA) and used for evaluating the expression pattern of DEFA7. For each human and rat RNA, 4 µg were reverse transcribed with random hexamers and Superscript II reverse transcriptase by
using a first-strand cDNA synthesis kit (Invitrogen) according to the instructions. The subsequent PCR was carried essentially as described (44, 45). Briefly, 1/40 of the first-strand cDNA from each tissue were used to amplify α-defensins and glyceraldehydes 3-phosphate dehydrogenase (GAPDH) with gene specific primers. Every pair of primers were designed from less conservative regions located on different exons to aid in specific amplification of target genes and in distinguishing PCR products amplified from cDNA versus genomic DNA (Table 1). The PCR program used was: 94°C denaturation for 2 min, followed by different cycles of 94°C denaturation for 20 sec, 55°C annealing for 20 sec, and 72°C extension for 40 sec, followed by a final extension at 72°C for 5 min. The sequence of primers and number of PCR cycles used are shown in Table 1. The PCR products were analyzed by electrophoresis on 1.2% agarose gels containing 0.5 µg/ml ethidium bromide. The specificity of each PCR product was confirmed by cloning of the PCR product into T/A cloning vector, followed by sequencing of the recombinant plasmid.

GenBank accession numbers

All cDNA sequences of new α- and θ-defensins described in this manuscript have been submitted to the GenBank. Accession numbers are as follows: rat, Defa6-Defa11 (AY623750-AY623755), Defa12-ps-Defa14-ps (AY746422-AY746424), Defa-rs1 (AY623756); mouse, Defcr20-Defcr25 (AY746425-AY746430), Defcr26 (AY766470), Defa-ps1 (AY746431), CRS1C-2 (AY761183), CRS1C-3 (AY761184), CRS4C-6 (AY761185); human, DEFA7P-DEFA11P (AY746432-AY746436);
chimpanzee, \(PTAD1-PTAD11-ps\) (AY746437-AY746445), \(PTTD1-ps-PTTD2-ps\) (AY746446-AY746447); and olive baboon, \(PAAD1-PAAD3\) (AY746448-AY746450).

**RESULTS**

*Genome-wide identification of \(\alpha\)-defensins in rodents and primates*

To identify possible existence of novel \(\alpha\)-defensin sequences in the vertebrate species whose genomes have been sequenced, all known \(\alpha\)-defensin peptide sequences were individually queried against EST, NR, HTGS, and WGS sequences of zebrafish, Japanese pufferfish, chicken, dog, rat, mouse, chimpanzee, and human in the GenBank by using the TBLASTN program (1). For every novel \(\alpha\)-defensin sequence identified, additional iterative BLAST searches were performed until no more novel sequences could be revealed. All genomic sequences containing \(\alpha\)-defensins were also retrieved from the GenBank and translated into six open reading frames and curated for the presence of the \(\alpha\)-defensin motif or signal/prosegment sequence to reveal potential sequences with distant homology. Whenever necessary, a combination of gene prediction programs, including GenomeScan (41), GENSCAN (4) and/or GeneWise (3), were used to predict full-length coding sequences of novel \(\alpha\)-defensins from genomic sequences.

As a result, a number of sequences have been discovered in the human, chimpanzee, rat and mouse. In the case of human, five novel \(\alpha\)-defensin genes were found in addition to five previously characterized ones, \(DEFA1-6\) (6, 21). Among new \(\alpha\)-defensin genes identified is \(DEFA7\), which consists of a canonical six-cysteine \(\alpha\)-defensin
motif and an N-terminal signal/prosegment sequence highly homologous to known human α-defensins (Fig. 1A).

Because DEFA7 lacks the start codon in the consensus position that initiates the translation of all other α-defensins, we sought to confirm if it is a functional gene. However, RT-PCR analysis revealed no expression of DEFA7 in any of 12 human tissues examined, including small intestine, bone marrow, and kidney, three of which are the most dominant sites to produce α-defensins in other mammalian species. Although it is still possible to be transcribed in other places, DEFA7 is most likely to be a pseudogene and therefore, is designated as DEFA7P. Such results are also consistent with the fact that there is no single EST sequence for DEFA7P in the GenBank, despite of nearly 5.8 million ESTs being deposited as of September 1, 2004.

In addition to DEFA7P, four α-defensin pseudogenes, namely DEFA8P-11P, have also been discovered in the human (Fig. 1A). Without a premature stop codon, each pseudogene would otherwise encode a peptide that is highly similar to a typical α-defensin (Fig. 1A). Identification of additional five α-defensin genes in the extensively studied human genome highlighted the power of our computational search strategy.

A total of 11 novel α-defensins, including four pseudogenes, have also been identified in the chimpanzee genome. These genes, which were termed Pan troglodytes α-defensins (PTAD) 1-11, are all human α-defensin orthologs (Fig. 1A), and therefore are expected to show a similar tissue expression pattern to human genes. PTAD1 and PTAD4
are predicted to be neutrophil-specific, whereas \textit{PTAD5} and \textit{PTAD6} are expected to be expressed mainly in Paneth cells of small intestine. Interestingly, \textit{PTAD7}, the chimpanzee ortholog of \textit{DEFA7P}, appears to be a functional gene with the start codon at the consensus position (Fig. 1A), but its expression site(s) remain(s) to be studied. Four orthologs of human $\alpha$-defensin pseudogenes are also present in the chimpanzee, namely \textit{PTAD8-ps} to \textit{PTAD11-ps}.

In addition to $\alpha$-defensins, two pseudogenes for $\theta$-defensin, \textit{PTTD1-ps} and \textit{PTTD2-ps}, have also been found in the chimpanzee (Fig. 1A). Both \textit{PTTD1-ps} and \textit{PTTD2-ps} are highly homologous to each other and to known human $\theta$-defensin pseudogene, \textit{DEFT1P} (6, 16) (Fig. 1A), but they differ with \textit{DEFT1P} in the locations of premature stop codons. \textit{DEFT1P} contains two stop codons at the amino acid positions 17 and 77; however, only the former stop codon is preserved in \textit{PTTD1-ps} and \textit{PTTD2-ps}.

Three new $\alpha$-defensins, namely \textit{PAAD1-3}, were also identified in two bacterial artificial chromosome (BAC) clones (accession no. AC116558 and AC116559) of olive baboon (\textit{Papio anubis}) (Fig. 1A). \textit{PAAD1-2} are orthologous to \textit{DEFA5}, whereas \textit{PAAD3} is a \textit{DEF7P} homolog. However, unlike \textit{DEF7P} that contains a single nucleotide mutation at the start codon, \textit{PAAD3} is composed of a stretch of completely different nucleotides around the start codon (See supplementary Fig. 1), and therefore it is also predicted to be a pseudogene. This prediction is further supported by the fact that, despite of extensive experimental screening efforts (33, 34), no \textit{PAAD3}-like sequences have been found to be produced in either jejunum or leukocytes of rhesus macaque, which is a member of
cercopithecids (Old World Monkeys) as olive baboon. Presumably, such PAAD3-like gene(s) are also inactive in rhesus macaque; however, it remains to be experimentally verified.

In the rat, a total of 14 α-defensin genes have been discovered, including four known ones (RatNP1/2, RatNP3, RatNP4, and RD5) (5, 42). Among ten novel rat α-defensins (Defa), six genes (Defa6-11) encode peptide sequences varying 87-104 amino acid residues in length with a canonical six-cysteine α-defensin motif (Fig. 1B). Also identified were three α-defensin pseudogenes, namely Defa12-ps, Defa13-ps, and Defa14-ps, with each having a premature stop codon at amino acid positions 34, 29, and 88, respectively (Fig. 1B). In addition, one α-defensin related sequence (Defa-rs1) was also found in the rat. Despite of the conservation in the N-terminal signal/prosegment sequence, Defa-rs1 has only five cysteine residues at the C-terminus with no characteristic α-defensin motif (Fig. 1B). Similarly, two groups of cryptdin-related sequences (CRS1C and CRS4C) with the signal/prosegment sequences highly homologous to α-defensins, but with a different number and spacing pattern of cysteines were reported earlier in mice (13). Such CRS peptides were shown recently to form covalent homo- or heterodimers and are capable of killing bacteria (9). Therefore, it is tempting to speculate that Defa-rs1, the only α-defensin -related sequence in the rat, associates with itself by forming homodimers and participates in innate defense in the gut as other enteric α-defensins.
Among 19 α-defensins (cryptdins) reported in mice, 17 belong to the highly homogeneous cryptdin 1 (Defcr1) subgroup with less than 5 amino acids in difference in the mature sequence and many are believed to be derived from allelic variants of the same gene (22). Defcr4 and Defcr5 represent distant cryptdin family members and their paralogous genes have not been reported in mice. Among eight novel α-defensins that we have identified in the mouse, five sequences (Defcr23-Defcr26 and Defcr-ps1) can be grouped to the Defcr1 subgroup, whereas the other three (Defcr20-Defcr22) are similar to Defcr4 (Fig. 1C). In addition, three novel cryptdin-related sequences were also found in mice (Fig. 1D). Among them, CRS1C-2 and CRS1C-3 are highly homologous to CRS1C-1, which contains a total of 11 cysteines with eight forming C-X-Y triplet repeats. On the other hand, CRS4C-6 belongs to a distant member of the CRS4C subgroup. Instead of consisting seven C-X-Y repeats, CRS4C-6 has an extra cysteine located close to the N-terminal end of the mature sequence. Unlike other CRS4C peptides that form intermolecular covalent dimmers (9), CRS4C-6 is likely to function as a monomer or non-covalently associated dimmers/oligomers. Interestingly, none of α-defensin sequences have been found in the species other than glires and primates. This is perhaps not surprising, given the fact that no α-defensins have been reported in any other vertebrate species despite of the extensive searching efforts in the past (6, 21).

Tissue expression patterns of rat α-defensins

A semi-quantitative RT-PCR was used to analyze tissue expression patterns of rat α-defensins. As shown in Fig. 2, all rat α-defensins exhibited two distinct patterns of expression. One group, which include RatNP1/2, RatNP3, and Defa12-ps, are
preferentially expressed in bone marrow with or without expression in small intestine (Fig. 2). Defa7, Defa10, and Defa11 also belong to this group with little but specific expression in bone marrow (data not shown). In contrast, the other group consisting of RD5, Defa6, Defa8, Defa9, and Defa-rs1 are specifically expressed in small intestine (Fig. 2). These genes are highly transcribed in jejunum and ileum with a low level of expression in duodenum. Interestingly, such a grouping is also consistent with the degree of conservation in the signal peptide sequences, which are more homologous within each group (top panel vs. bottom panel, Fig. 1B). It is surprising to see the expression of RatNP3 and Defa12-ps in both bone marrow and small intestine, given the fact that all known α-defensins are produced either by myeloid or Paneth cells, but not by both cell lineages (21). The cells that express rat α-defensins in the intestinal tract are presumably Paneth cells, which produce enteric α-defensins in humans and mice (21). It is worthy noting that Defa-rs1 is expressed in the jejunum and ileum at the comparable levels to most other enteric defensins (Fig. 2), reinforcing the notion that it most likely will function in vivo.

Molecular evolutionary analyses of mammalian α-defensins

To analyze the phylogenetic relationships among all known α-defensins in glires and primates including the newly identified ones, the proportion of amino acid differences (p-distance) of full-length peptides was calculated by the neighbor-joining method (26) (See supplementary Fig. 2 for alignment of peptide sequences of 82 α-defensins used). As shown in Fig. 3A, two major clusters are evident with α-defensins in the primates with rabbit and guinea pig forming one cluster supported by a bootstrap
value of 65%, and mouse and rat enteric α-defensins forming the other cluster with a bootstrap value of 98%. Rat α-defensins appear to be unique in that the genes of myeloid origin formed a separate cluster from the genes of enteric origin. Rat myeloid α-defensins tend to group with primate, rabbit, and guinea pig α-defensins.

Perhaps the most striking feature of the phylogenetic tress is that α-defensins formed species-specific clusters in the non-primate species (rat, mouse, guinea pig, and rabbit), implying that these genes may have undergone repeated duplication after these species diverged from a common ancestor. In contrast, no species-specific clustering was demonstrated among four primate species examined (human, chimpanzee, rhesus macaque, and olive baboon) (Fig. 3A). Instead, several distinct subclusters of α-defensins exist across the primate species. For example, chimpanzee and human are in a complete analogy with α-defensins present in all subclusters. The α-defensins of Old World Monkeys are located in three subclusters with human and chimpanzee DEFA1-3, DEFA5, and DEFA7P, respectively. These results suggested that many of the primate genes were likely evolved before the divergence of these species. The primate-specific θ-defensins, which exist in both cercopithecids and hominids, are clearly clustered together with the DEFA10P gene lineage, reinforcing the notion that θ-defensins were evolved from α-defensins after divergence of the primates from other mammalian species (6, 16, 20, 21).

To minimize possible biases of selection pressure exerted on exon (and therefore peptide) sequences, the evolutionary relationships of the α-defensin genes were further evaluated by analysis of the intron sequences within the open reading frame. The p-
distance of the introns of 59 α-defensins with known gene sequences was calculated by the neighbor-joining method (the intron sequences used will be available upon request) (26). Consistent with the results obtained from the full-length peptide sequences, a similar two major clusters of genes were obtained, in which mouse and rat enteric α-defensins were separated from others, supported by a bootstrap value of 100%, and hence were most likely evolved from a different ancestor (Fig. 3B). On the other hand, rat myeloid gene introns are clearly clustered together with primate α-defensins and therefore were originated from the same ancestral gene. As for rat and mouse enteric genes, there are two distinct subclusters with each being supported by a bootstrap value of 100%, implying that mouse cryptdins (Defcr) share the same ancestor with rat RD5, Def13-ps, and Defa14-ps, whereas mouse cryptdin-related sequences (CRS1C and CRS4C) were derived from the same primordial gene as rat Defa6, Defa8, and Defa9, and Defa-rs1 (Fig. 3B). It is obvious that each subcluster has undergone a repeated duplication and diversification to give rise to multiple members in each species after these mammalian species diverged from a common ancestor.

Despite of significant sequence conservation in the first exon (Fig. 1C and 1D) and intron (data not shown), substantial differences exist in the cysteine motif encoded by the second exon between canonical α-defensins and three subgroups of related sequences in rats and mice, namely Defa-rs1, CRS1C, and CRS4C (Fig. 1C and 1D). To further address the origin and evolution of the cysteine motif of α-defensin-related sequences, nucleotide sequences of the second exon were compared with rat enteric α-defensins (Defa6, Defa8, and Defa9) in the same subcluster (Fig. 3B). It is evident that Defa-rs1
shares a high homology with three rat α-defensins throughout the entire coding region in
the second exon (Fig. 4), suggesting that, in addition to the first exon and intron, the
second exon of Defa-rs1 was also originated from the same ancestral gene as rat enteric
α-defensins. In fact, the change in the number and pattern of cysteines in Defa-rs1 was
mainly because of nucleotide deletions (Fig. 4).

As for mouse CRS1C and CRS4C, both subgroups are highly homologous in the
second exon (Fig. 4), and therefore share the same ancestor. The difference in the
cysteine motif of CRS1C and CRS4C peptides were primarily because of deletions and
insertions rather than nonsynonymous substitutions (Fig. 4). Although CRS peptides
share a less homology with rat enteric α-defensins and Defa-rs1, patches of identical
nucleotides are evident among them throughout the entire second exon (Fig. 4), implying
that, similar to Defa-rs1, the changes in the second exon of CRS genes were likely to be a
result of sequence diversification, instead of exon shuffling, which commonly occurred
during the evolution of many gene families (23). Coupled with the fact that both intron
and the first exon are highly conserved among classic α-defensins and related sequences,
the full-length CRS genes most likely were derived from the same ancestral genes for rat
enteric α-defensins and Defa-rs1. Apparently, CRS genes and Defa-rs1 were
independently evolved from classic α-defensins in mice and rats, respectively, following
divergence of these two rodent species from each other.

Collectively, our results suggested that all mammalian α-defensins may have
evolved from two ancestral genes with each giving rise to one major cluster of daughter
genes. The absence of α-defensins in non-mammalian species clearly indicated that α-defensins have evolved after the divergence of mammals from other vertebrate species. Formation of species-specific clusters, particularly in non-primate species, strongly suggested that many α-defensins have evolved independently after these mammalian species diverged from a common ancestor. It is worthy noting that, because of the lack of complete repertoires of α-defensin genes in the rabbit and guinea pig, the origin of α-defensins in these two species remains inconclusive. Analysis of coding sequences placed them in the same cluster as primate genes (Fig. 3A). However, comparison of intron sequences indicated that guinea pig myeloid defensins (GNCP1A, GNCP1B and GNCP2) and rabbit kidney defensins (RK1 and RK2) may share the same ancestor with rat and mouse enteric genes, while rabbit myeloid defensins (NP1 and NP2) were likely evolved from the ancestral genes for primates (Fig. 3B).

To further understand the driving force for sequence divergence of α-defensins during evolution, we tested whether positive Darwinian selection has occurred by estimating the number of synonymous substitutions per synonymous site (dS) and the number of nonsynonymous substitutions per nonsynonymous site (dN) substitution for different regions of mammalian α-defensins using the method of Nei and Gojobori (19). We computed these values for 11 pairs of closely related genes, which were phylogenetically and thus statistically independent of each other (24). Consistent with earlier results (10, 12), mean dN was significantly greater than mean dS in the mature peptide, but there was no significant difference between mean dS and mean dN in the signal peptide or prosegment (Table 2). These results support the hypothesis that natural
selection has acted to diversify the functionally active mature defensin region but not other portions of the molecules (10, 12).

**Genomic organization of mammalian α-defensin gene clusters**

Searching through HTGS and WGS sequences in the GenBank led to identification of a number of genomic sequences containing α-defensin genes in the human, chimpanzee, rat and mouse genome (data not shown). To determine the relative position and orientation of each defensin on the chromosome, individual defensins were searched against the assembled human, chimpanzee, rat and mouse genomes by using the BLAT program (14) through the UCSC Genome Browser. In the case of human, all 11 known and newly identified α- and θ-defensins genes were found to form a continuous cluster expanding 132 kb on the chromosome 8p23 (Fig. 5A). All human defensins are transcribed from the same direction from centromere to telomere with the θ-defensin pseudogene (DEFT1P) residing in the center. The chimpanzee genome encodes an α-defensin cluster of 10 distinct genes, which expanded 117 kb in the proximal region of chromosome 7 (Fig. 5A). The α-defensin clusters in the chimpanzee and human are nearly in perfect synteny. The only difference is the presence of two copies of DEFA1 and DEFT1P in the human rather than a single copy in the chimpanzee. Such syntenic α-defensin loci are also present in rats and mice. In the case of rat, all 14 α-defensin genes are clustered within a 311-kb distance on the chromosome 16q12.4-q12.5 (Fig. 5B), whereas six mapped mouse α-defensin genes are located continuously on chromosome 8A1.3 (Fig. 5C).
It is noted that all α-defensin genes in both rodents and primates reside within a β-defensin cluster (data not shown) and are adjacent to β-defensin 1 (Defb1), which is evolutionarily conserved across the mammalian species (Fig. 5). Therefore, the syntenic location and physical proximity of α-defensins in these species provided additional evidence to support the conclusion that α-defensins have arisen from common ancestor(s) by gene duplication followed by diversification.

Comparison of the open reading frames with genomic sequences of all known and newly identified α-defensins (including α-defensin-related sequences) revealed a highly conserved gene structure, which is composed of two exons separated by a 500-700 bp intron, except for DEFA5/PTAD5, DEFA6/PTAD6, RK1, RK2, PAAD1 and PAAD2, whose introns vary from 848-1646 bp (data available upon request). Consistent with earlier findings, signal sequence and most of the prosegment of α-defensins are encoded in the first exon, whereas the second exon primarily encodes the mature α-defensin sequences. However, based on α-defensin genes studied thus far, it appears that genes of myeloid origin consist of an additional exon encoding 5’-untranslated region, whereas enteric genes are composed of only two exons (21). The lack of full-length cDNA sequences for these newly identified novel sequences prevented us from predicting the genomic locations of 5’- and 3’-untranslated regions and thus possible existence of additional introns.
DISCUSSION

Sequence diversity of mammalian α-defensins

Following systemic, genome-wide screenings, the complete repertoires of the α-defensin gene family, including a number of novel genes, have been identified in the human, chimpanzee, rat, and mouse (Fig. 1). Although it is highly unlikely, we could not rule out the possibility that additional α-defensin genes with distant homology might be uncovered in these species by different computational methods such as the use of Hidden Markov models (18, 30, 31). Alignment of all known α-defensin peptide sequences revealed the conservation of signal and negatively charged prosegment sequences as well as the characteristic six-cysteine-containing α-defensin motif, with exception of pseudogenes and α-defensin-related sequences (Fig. 1). Given the significance of six cysteines in maintaining the spatial structure and biological functions, most pseudogenes have at least one mutation at the conserved cysteine position (Fig. 1), consistent with their dysfunctional nature.

Among all newly identified genes in the primates, chimpanzee PTAD7 is unique in that it represents a distant, putative functional gene. Unlike a typical α-defensin that is positively charged, the putative mature sequence of PTAD7 contains two arginines (R) and two glutamic acids (E), and therefore is predicted not to have a net positive charge under physiological conditions. However, its human ortholog, DEFA7P, appears to be a pseudogene because of an absence of the start codon and hence lack of transcription in a wide range of tissues. Interestingly, PAAD3, a DEFA7P ortholog in olive baboon, also
appears to be a pseudogene with no start codon at the canonical position (see supplementary Fig. 1). The unique sequence and evolutionary dynamics of the PTAD7 gene lineage in different primary species implied that PTAD7 might function differently in the chimpanzee from other α-defensins. Consistent with its novelty, PTAD7 is the only known putative functional α-defensin with a nonsynonymous mutation to serine at the fourth canonical cysteine position (Fig 1A).

As opposed to many β-defensins, most α-defensins contain no or only one amino acid following the last two cysteines (Fig. 1). However, Defa8, Defa9 and Defa11 in the rat share a unique feature of unusually long C-terminal tail (9-14 amino acids) rich in charged and polar uncharged amino acids. Such C-terminal tails, which are predicted to be exposed to the surface to influence the overall charge and amphipathicity, might confer a different antimicrobial spectrum and/or efficacy on those peptides, as the net charge and amphipathicity are strongly correlated with antimicrobial activity of defensins (25, 28, 32). Further studies are warranted to assess the importance of such long C-terminal tails in antimicrobial activity as well as in other biological functions.

Evolution of mammalian α-defensins

All α-defensins have been found only in certain mammals, but not in any other vertebrate species that we examined, suggesting that α-defensin genes have appeared following mammalian divergence. The presence of α-defensin loci in syntenic chromosomal regions of different mammalian species (Fig. 5) is indicative of a common ancestor. Phylogenetic analyses of both full-length peptide and intron sequences of α-
defensins revealed two distinct clustering (Fig. 3), implying that they may have independently evolved from two separate ancestral genes. One ancestral gene has undergone significant duplication and diversification giving rise to enteric-specific α-defensins in the rat and mouse, whereas the other ancestral gene has evolved to the genes in the primates as well as in rat myeloid cells.

Phylogenetic analysis of primate α-defensins revealed an interesting evolutionary pattern and pointed out possible presence of subgroups of α-defensins that are specific in the hominid lineage. Despite extensive searches for the repertoire of α-defensins in both leukocytes (34) and small intestines (33), neither DEFA4- nor DEFA6-like genes have been found in rhesus macaques (Fig. 1A and Fig. 3). Such genes are also absent in nearly 1000 olive baboon genomic sequences deposited in the GenBank as of September 1, 2004. Therefore, it is likely that DEFA4/PTAD4 and DEFA6/PTAD6 lineages are hominid-specific genes. On the other hand, albeit of a single copy of DEFA5/PTAD5 gene encoded in human and chimpanzee genomes, multiple members of the DEFA5-like genes are present in both rhesus macaque (RED1-6, RMAD4/5, and RMAD6/7) and olive boboon (PAAD1 and PAAD2) (Fig. 3A). Apparently, the ancestral gene for DEFA5 has undergone a different evolutionary pattern: It duplicated and expanded in cercopithecids, but remained a single copy in hominids. In addition, the DEFA5 ancestral gene has diverged in rhesus macaques to acquire a new ability for certain offspring genes to express in myeloid cells (e.g., RMAD4/5, and RMAD6/7), in addition to intestinal Paneth cells (e.g., RED1-RED6) as in humans.
With regard to the evolution of mammalian α-defensins, Bevins et al. proposed a model, which was based on dot matrix sequence comparisons of introns of a few mouse and human genes, that α-defensins were derived from two ancestral genes for human enteric DEFA5 and DEFA6, and that a subsequent homologous unequal meiotic crossover of DEFA5 and DEFA6 generated a hybrid gene that was further evolved to present day myeloid α-defensins (2). Although such a model might be true in primate species (2), our results clearly argued against its generalization to non-mammalian species, because the mouse genome apparently encodes neither DEFA5/DEFA6-like genes nor myeloid α-defensins. Therefore, mouse α-defensin genes were apparently not derived from homologous crossover of the primordial genes for DEFA5 or DEFA6. Furthermore, because of presence of multiple DEFA5 genes expressed in both small intestine and leukocytes in cercopithecids, but an absence of DEFA6 lineages, the validity of Bevins’ evolutionary model in primate species remains to be examined with availability of additional primate genomes.

As for the origin of mammalian α-defensins, they were most likely evolved from β-defensins after mammals diverged from other vertebrates, primarily because of the physical proximity on the chromosome as well as a similarity in spatial structure and biological activities (6, 39). As suggested by Nguyen et al. (20), appearance and further diversification of α- and θ-defensins from already a large β-defensin gene family is probably because of the need for antiviral defense in certain mammalian and/or primate species, as these newly evolved α- and θ-defensins have acquired novel lectin-like activity and are capable of inhibiting the entry of viruses (such as HIV) to host cells (36,
It is possible that divergence of α-defensin-related sequences from canonical α-defensins in rodents could lead to some novel activities beyond antimicrobial. Further functional characterization of these α-defensins will shed light on the significance of their diversification during evolution and facilitate their development as a new class of antimicrobial agents.

Our earlier comparative analysis of β-defensins in the chicken, rodents, and human revealed that most gene lineages are conserved across mammalian species and thus were evolved before the divergence of mammals from each other (39). In contrast, α-defensins tend to form species-specific clusters particularly in non-primate species (Fig. 3). Distinct subsets of α-defensin genes also exist with most subsets containing multiple members in each species. It is even true with the rat and mouse (Fig. 3), which only diverged 17-25 million years ago. It is rather unique that most α-defensin genes duplicated very recently in the evolutionary time and have undergone significant, but independent expansion in different species. Clearly, the β-defensin gene family is evolutionarily older than α-defensins. As compared to the former, α-defensin genes duplicated and expanded much more rapidly. Calculation of the rates of synonymous vs. nonsynonymous substitutions of 11 pairs of representative α-defensins revealed that positive Darwinian selection appears to have acted to diversify α-defensins, particularly in the mature peptide region (Table 2).

Gene duplication followed by positive selection indeed has been observed in several gene families involved in immune responses (11). Divergence of these immune
genes often leads to either an additional layer of functional redundancy or acquisition of functional novelties, both of which conceivably help the hosts cope more effectively with a broad range of pathogens. In fact, α-defensins have been shown to exhibit selectivity against varied microorganisms, and a modest difference in the primary sequence could have a significant impact on the antibacterial spectrum and/or potency (22). Therefore, as different mammals live in quite different ecological niches, the production of species-specific α-defensins would presumably allow them to better respond to the specific microbial challenges that they face. The presence of α-defensin-related sequences in rodents and functional divergence of θ-defensins from α-defensins in certain primate species provide additional evidence supporting this notion.

ACKNOWLEDGEMENTS

We thank Lin Liu, Narendranath Reddy, Manoj Bhaskaran and Marcia Howard for providing access to rat tissues and Donald Skinner-Noble for critical reading of the manuscript. We are also grateful for the constructive comments from three anonymous reviewers.

GRANTS

This work was supported by Oklahoma Center for Advancement of Science and Technology Grant HR03-146. We also acknowledge the support of Oklahoma Agricultural Experiment Station.
REFERENCES


FIGURE LEGENDS

Fig. 1. Amino acid sequence alignment of novel α-defensins and α-defensin-related sequences in the primates (A), rat (B) and mouse (C and D). Panel C shows novel cryptdins identified in mice, whereas Panel D demonstrates new cryptdin-related sequences. The dashes are created to optimize the alignment, and canonical cysteines in the mature defensin regions are shaded. The letter ‘X’ in bold denotes the presence of premature stop codon in each pseudogene. Note that computational predictions of all novel rat α-defensin peptide sequences, except for the pseudogenes, have been confirmed by cloning and sequencing of their respective RT-PCR products amplified from appropriate tissues. Abbreviations: DEFA/Defa, α-defensin; DEFT1P, human θ-defensin 1 pseudogene; PTAD, chimpanzee (Pan troglodytes) α-defensin; PTTD, Pan troglodytes θ-defensin; RMAD, rhesus macaque myeloid α-defensin; RED, rhesus macaque enteric α-defensin; RTD, rhesus macaque θ-defensin; PAAD, Papio anubis α-defensin; RatNP, Rat neutrophil α-defensin; Defa-rs1, rat α-defensin-related sequence; RD5, rat α-defensin 5; Defcr, mouse α-defensin-related cryptdin; CRS1C, mouse cryptdin-related sequence 1C; CRS4C, mouse cryptdin-related sequence 4C; P/ps, pseudogene.

Fig. 2. Tissue expression patterns of novel rat β-defensins by RT-PCR. See Materials and Methods for details. The number of PCR cycles was optimized for each gene, and the specificity of each PCR product was confirmed by sequencing. The house-keeping gene, GAPDH, was used for normalization of the template input.
Fig. 3. Phylogenetic tree of mammalian α-defensins. The tree was constructed by calculating the proportion difference (p-distance) of aligned amino acid sites of full-length peptide sequences (A) or p-distance of aligned nucleotide sites of the entire intron sequences (B) according to the neighbor-joining method. The reliability of each branch was tested using 1000 bootstrap replications. Numbers on the branches indicate the percentage of 1000 bootstrap samples supporting the branch. Only branches supported by a bootstrap value of at least 40% are shown. Abbreviations: NP/MCP, rabbit neutrophil α-defensin/macrophage cationic peptide; RK, rabbit kidney α-defensin; GNCP, guinea pig neutrophil cationic peptide. See Fig. 1 legend for additional abbreviations.

Fig. 4. Alignment of the second exon sequences of rodent α-defensin-related sequences with three rat enteric α-defensins. For clarity, the last ~75 nucleotides of introns were also aligned and the exon/intron conjunction sites are indicated with a vertical line. Conserved regions are shaded and the stop codon of each gene boxed. See Fig. 1 legend for abbreviations.

Fig. 5. Genomic organizations of α-defensin clusters in the primates (A), rat (B), and mouse (C). The position of each gene is represented by a solid vertical bar and the width of each bar is proportional to the size of each gene. The direction of transcription is indicated by the triangle. The genes with solid triangles are transcribed in the direction opposite to the ones with open triangles. The gene clusters were derived from the assembled genomes of human (NCBI Build 35), chimpanzee (NCBI Build 1 Version 1), rat (BCM Version 3.1) and mouse (NCBI Build 33). Note that the α-defensin cluster is
located in a syntenic region adjacent to β-defensin 1 (*DEFB1*) in each species. In mice, a number of additional α-defensin genes are encoded in two other large but unmapped clusters, and are not indicated here for the sake of clarity. Abbreviations: Tel, telomere; Cen, centromere. See Fig. 1 legend for additional abbreviations.
# Table 1. Primer sequences used for RT-PCR analysis of rat α-defensins

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward</th>
<th>Reverse</th>
<th>Product Size (bp)</th>
<th>No. of Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cDNA</td>
<td>Genomic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RatNP1/2</td>
<td>CCATGAGGACGCTCACTCTGTGCT</td>
<td>CAAATGGATTCTTCTTGGTCGG</td>
<td>316</td>
<td>833</td>
</tr>
<tr>
<td>RatNP3</td>
<td>CAAATGGATTCTTCTTGGTCGG</td>
<td>CTTCTTCTTCTTCTTTGTGCTCA</td>
<td>286</td>
<td>793</td>
</tr>
<tr>
<td>RD5</td>
<td>GAAGACACTTGTCTCCTCTTTCTG</td>
<td>TGTTGCAGATCCCCCATAATGCCCT</td>
<td>245</td>
<td>861</td>
</tr>
<tr>
<td>Defa6</td>
<td>GAAGACACTTGTCTCCTCTTTCTG</td>
<td>TATTTCTTTAGTTCCGAGCTTTTCG</td>
<td>228</td>
<td>815</td>
</tr>
<tr>
<td>Defa8</td>
<td>GAAGACACTTGTCTCCTCTTTCTG</td>
<td>AATAGACCTTGGAGCACTAGACC</td>
<td>258</td>
<td>840</td>
</tr>
<tr>
<td>Defa9</td>
<td>GAAGACACTTGTCTCCTCTTTCTG</td>
<td>TATTTCTTTAGTTCCGAGCTTTTCG</td>
<td>305</td>
<td>890</td>
</tr>
<tr>
<td>Defa10</td>
<td>CAAATGGATTCTTCTTTGGTCGG</td>
<td>ATCCGAGCATTGAGCACTAGACC</td>
<td>263</td>
<td>765</td>
</tr>
<tr>
<td>Defa11</td>
<td>CAAATGGATTCTTCTTTGGTCGG</td>
<td>TTGTTCATAGTGAGCACTAGACC</td>
<td>293</td>
<td>810</td>
</tr>
<tr>
<td>Defa12-ps</td>
<td>CAAATGGATTCTTCTTTGGTCGG</td>
<td>AATAGAGTACCTTTGAGCACTAGACC</td>
<td>269</td>
<td>769</td>
</tr>
<tr>
<td>Defa-rs1</td>
<td>GAAGACACTTGTCTCCTCTTTCTG</td>
<td>ATGCTGTCACTTTGAGCACTAGACC</td>
<td>261</td>
<td>834</td>
</tr>
<tr>
<td>GAPDH</td>
<td>GTGAAGGGTGGAGTGAGAC</td>
<td>GAGATGAGTACCTTTGAGCACTAGACC</td>
<td>356</td>
<td>&gt;390</td>
</tr>
</tbody>
</table>
Table 2. Demonstration of positive selection in 11 phylogenetically independent pairs of mammalian α-defensin genes.

<table>
<thead>
<tr>
<th>Region</th>
<th>$d_S$</th>
<th>$d_N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal peptide</td>
<td>0.031 ± 0.014</td>
<td>0.023 ± 0.008</td>
</tr>
<tr>
<td>Prosegment</td>
<td>0.066 ± 0.026</td>
<td>0.066 ± 0.022</td>
</tr>
<tr>
<td>Mature defensin</td>
<td>0.122 ± 0.041</td>
<td>0.182 ± 0.047$^b$</td>
</tr>
</tbody>
</table>

$^a$ Mean (± S.E.) numbers of synonymous substitutions per synonymous site ($d_S$) and of nonsynonymous substitutions per nonsynonymous site ($d_N$). The following sister pairs of closely related genes were compared: from Primates, DEFA5 vs. RMAD4/5, DEFA1 vs. DEFA3, RMAD1/2 vs. RMAD8; from rabbit NP1/MCP1 vs. NP2/MCP2, NP4 vs. NP5; from rodents, GNCP1A vs. GNCP2, defcr5 vs. defcr6, defcr20 vs. defcr21, RatNP3 vs. RatNP4, rat defa6 vs. rat defa8, rat defa7 vs. rat defa11.

$^b$ Significant difference between $d_S$ and $d_N$ at P<0.01 level (paired t-test, two-tailed).
Figure 1
Figure 2
Figure 3
Figure 4
Figure 5
Supplementary Fig. 1. Sequence alignment of the exon encoding the signal peptide and prosegments of three orthologous genes in the human (DEFA7P), chimpanzee (PTAD7), and olive baboon (PAAD3). The encoded peptide sequence for each exon is also indicated. Note that the start codon in PTAD7 is mutated in both DEFA7P and PAAD3 as highlighted in orange color, and that a string of sequences around the start codon (underlined) in PAAD3 are completely different from those in DEFA7P and PTAD7. 

Abbreviations: DEFA7P, human α-defensin 7 pseudogene; PTAD7, Pan troglodytes α-defensin 7; PAAD3, Papio anubis α-defensin 3.

Supplementary Fig. 2. Amino acid sequence alignment of a total of 82 known and newly identified α-defensins and α-defensin-related sequences in the primates and glires. The dashes are created to optimize the alignment, and canonical cysteines in the mature defensin regions are shaded. The premature stop codon in each pseudogene is omitted. Note that computational predictions of all novel rat α-defensin peptide sequences, except for the pseudogenes, have been confirmed by cloning and sequencing of their respective RT-PCR products amplified from appropriate tissues. Abbreviations: DEFA/Defa, α-defensin; DEFT1P, human θ-defensin 1 pseudogene; PTAD, chimpanzee (Pan troglodytes) α-defensin; PTTD, Pan troglodytes θ-defensin; RMAD, rhesus macaque myeloid α-defensin; RED, rhesus macaque enteric α-defensin; RTD, rhesus macaque θ-defensin; PAAD, Papio anubis α-defensin; RatNP, Rat neutrophil α-defensin; Defa-rs1, rat α-defensin-related sequence; RD5, rat α-defensin 5; Defcr, mouse α-defensin-related
cryptdin; CRS1C, mouse cryptdin-related sequence 1C; CRS4C, mouse cryptdin-related sequence 4C; NP/MCP, rabbit neutrophil α-defensin/macrophage cationic peptide; RK, rabbit kidney α-defensin; GNCP, guinea pig neutrophil cationic peptide; P/ps, pseudogene.

Supplementary Fig. 1
CHAPTER IV

CROSS-SPECIES ANALYSIS OF THE MAMMALIAN B-DEFENSIN GENE FAMILY:
PRESENCE OF SYNTENIC GENE CLUSTERS AND PREFERENTIAL EXPRESSION
IN THE MALE REPRODUCTIVE TRACT

Amar A. Patil¹, Yibin Cai¹, Yongming Sang², Frank Blecha² and Guolong Zhang¹

¹Department of Animal Science, Oklahoma State University, Stillwater, OK 74078;
²Department of Anatomy and Physiology, Kansas State University, Manhattan, KS 66506.

The following chapter has been published in the Physiological Genomics, 2005, 23:5-17 and
used with permission of American Physiological Society.
ABSTRACT

Mammalian β-defensins are an important family of innate host defense peptides with pleiotropic activities. As a first step to study the evolutionary relationship and biological role of the β-defensin family, we identified their complete repertoires in the human, chimpanzee, mouse, rat, and dog following systemic, genome-wide computational searches. Although most β-defensins genes are composed of two exons separated by an intron of variable length, some contain additional one or two exons encoding an internal pro-sequence, a segment of carboxyl terminal mature sequences or untranslated regions. Alternatively spliced isoforms have also been found with several β-defensins. Furthermore, all β-defensin genes are densely clustered in 4-5 syntenic chromosomal regions with each cluster spanning <1.2 Mb across the five species. Phylogenetic analysis indicated that, although the majority of β-defensins are evolutionarily conserved across species, subgroups of gene lineages exist that are specific in certain species, implying that some β-defensins originated after divergence of these mammals from each other, while most others arose before the last common ancestor of mammals. Surprisingly, reverse transcriptase-PCR revealed that all but one rat β-defensin transcripts are preferentially expressed in the male reproductive tract particularly in epididymis and testis, except that Defb4, a human β-defensin-2 ortholog, is more restricted to the respiratory and upper gastrointestinal tracts. Moreover, most β-defensins expressed in the reproductive tract are developmentally regulated with enhanced expression during sexual maturation. Existence of such a vast array of β-defensins in the male reproductive tract suggests that these genes may play a dual role in both fertility and host defense.
INTRODUCTION

Defensins and defensin-like molecules comprise a diverse group of cationic antimicrobial peptides characterized by the presence of multiple cysteine residues and a highly similar tertiary structure known as the defensin motif (14, 28, 35, 46, 60). Defensin-like genes have been discovered in many species including plants, fungi, arthropods, mollusks, reptiles, birds, and mammals (13, 14, 28, 36, 47, 52, 55, 57). Defensins in vertebrate species are further classified into three families, namely α-, β-, and θ-defensins, based on the spacing pattern of six cysteine residues (10, 14, 28). Although the evolutionary relationship between vertebrate defensins and defensins in other species remains unclear, phylogenetic analysis revealed that a primordial β-defensin gene is the common ancestor for all vertebrate defensins (57). β-defensins are believed to evolve before the divergence of mammals from birds (57) and gave rise to α-defensins in glires and primates after they diverged from other mammalian species (36). θ-defensins further originated from α-defensins following separation of primates from other mammals (33, 36).

In contrast with α- and θ-defensins that are produced primarily in the granules of either leukocytes or intestinal Paneth cells, β-defensins are unique in that they are primarily produced by nongranular mucosal epithelial cells lining the respiratory, gastrointestinal, and genitourinary tracts (14, 28, 35, 46, 60). In addition to their broad-spectrum antimicrobial activity against bacteria, fungi, and certain enveloped viruses, β-defensins have been recognized recently to resemble chemokines both structurally and functionally (14, 28, 46, 60). Recent studies also revealed that certain β-defensins are actively involved in sperm maturation and capacitation (41, 65, 68, 70), physiological properties that are not related to host defense.
All β-defensin genes encode a precursor peptide that consists of a hydrophobic, leucine-rich signal sequence, a pro-sequence, and a mature six-cysteine defensin motif at the carboxyl terminus (14, 28). In most cases, β-defensin precursors are encoded in two separate exons separated by an intron of variable length with one exon encoding the signal and pro-sequence and the other encoding primarily the mature peptide (14, 28). It is generally believed that mature β-defensins need to be cleaved from the pro-sequences by proteolytic enzymes to become biologically active, although such proteases are yet to be identified.

A number of putative β-defensins were reported recently in humans and mice (47); however, later identification of a novel human sequence implied possible existence of additional β-defensins in those two species (42). In addition, the evolutionary relationships among β-defensin genes across mammalian species are poorly understood, although limited data are available on β-defensin evolution within species or among the primates (9, 22, 32, 48).

Recent availability of complete genome sequences of a number of phylogenetically distinct mammalian species provides an excellent opportunity to identify the entire β-defensin gene family members in these species. Here, we report the discovery of complete repertoires of β-defensins in the human, chimpanzee, mice, rat, and dog by using a systemic computational search strategy that we developed recently (36, 57). We showed that β-defensin genes form 4-5 syntenic clusters in these mammals. Although the majority of β-defensin genes are evolutionarily conserved across rodents, canines and primates, a few gene lineages exist only in certain species. We also examined the tissue expression patterns of the entire β-defensin gene family in the rat by reverse transcriptase-PCR. To our surprise, all but one of the β-defensins were found to be
preferentially expressed in the male reproductive system, particularly in testis and different segments of epididymis. The only exception, Defb4, is more restricted to the respiratory and upper gastrointestinal tracts with virtually no expression in the male reproductive system. Furthermore, we provide evidence that many β-defensins are developmentally regulated and the expression levels are elevated with age. These findings argue that the major biological functions of β-defensins may be related to reproduction and fertility, in addition to innate host defense.

**MATERIALS AND METHODS**

*Computational search for novel β-defensins*

To identify potential novel β-defensins in the human, mouse, rat, dog, and chimpanzee, systemic computational searches were performed essentially as we described (36, 57). In brief, all known β-defensin peptide sequences were individually queried against expressed sequence tags (EST), nonredundant sequences (NR), unfinished high throughput genomic sequences (HTGS), and whole genome shotgun sequences (WGS) in GenBank by using the TBLASTN program (1) with the default settings on the NCBI web site ([http://www.ncbi.nlm.nih.gov/BLAST](http://www.ncbi.nlm.nih.gov/BLAST)). We then examined all potential hits for the presence of the characteristic β-defensin motif or conserved signal peptide and prosequence. Additional iterative BLAST searches were performed for every novel β-defensin sequence identified as described above until no more novel sequences were revealed. Because mammalian defensins tend to form clusters (47, 48), all genomic sequences containing β-defensins were also retrieved from GenBank to discover potential novel sequences with distant homology. The nucleotide sequences between two neighboring defensin genes were translated into six open reading frames
and individually compared with the two defensin peptide sequences for the presence of β-
defensin motif or signal peptide/pro-sequence by using the BLASTP program (1) on the NCBI web site (http://www.ncbi.nlm.nih.gov/blast/bl2seq/) and/or ClustalW program (version 1.82) (53) on the European Bioinformatics Institute web site (http://www.ebi.ac.uk/clustalw).

**Prediction of full-length coding sequences and genomic structures of β-defensins**

In most cases, signal and pro-sequence of β-defensins are encoded in a separated exon from the mature peptide-encoding exon. If either the signal/pro-sequence or β-defensin motif of a novel gene was missing in a genomic sequence, a 5-15 kb sequence flanking sequence was retrieved to identify the full-length coding sequence and to derive the structural organization of that novel β-defensin gene by using a combination of GenomeScan (61), GENSCAN (5) and/or GeneWise2 (4). All cDNA sequences of novel β-defensins described in this study have been submitted to GenBank and the accession numbers are listed in Supplementary Table 1.

**Chromosomal mapping of β-defensin gene clusters**

The BLAT program (25) was used to determine the relative position and orientation of each defensin in the genome through the UCSC Genome Browser (http://genome.ucsc.edu). Individual defensins were searched against latest versions of the assembled genomes of human (NCBI Build 35 Version 1), chimpanzee (NCBI Build 1 Version 1), mouse (NCBI Build 32), rat (BCM Version 3.1), and dog (NCBI Build 1 Version 1) released in June 2004, November 2003, September 2004, June 2003, and August 2004, respectively.
**Sequence alignment and phylogenetic analysis of β-defensins**

Multiple sequence alignments were carried out by using the ClustalW program (version 1.82) (53). The neighbor-joining method (43) was used to construct the phylogenetic tree by calculating the proportion of amino acid differences (p-distance), and the reliability of each branch was tested by 1000 bootstrap replications.

**Semi-quantitative RT-PCR analysis of tissue and developmental expression patterns of β-defensins**

A total of 29 different tissue samples (see Fig. 6B) were harvested from healthy, 2-month old Sprague-Dawley rats. Total RNA was extracted using Trizol (Invitrogen). For each RNA sample, 4 μg were reverse transcribed with random hexamers and Superscript II reverse transcriptase by using a first-strand cDNA synthesis kit (Invitrogen) according to the instructions. Subsequent PCR was performed with DNA Engine (Model PTC-200, MJ Research) essentially as described (36, 57, 69). Briefly, 1/40 of the first-strand cDNA from each tissue were used to amplify β-defensins and glyceraldehydes 3-phosphate dehydrogenase (GAPDH) with gene-specific, exon-spanning primers (Supplementary Table 2). The PCR program used was: 94°C denaturation for 2 min, followed by different cycles of 94°C denaturation for 20 sec, 55°C annealing for 20 sec, and 72°C extension for 40 sec, followed by a final extension at 72°C for 5 min. The number of PCR cycles was optimized for each gene to ensure linear amplification (Supplementary Table 2). Specificity of each PCR product was confirmed by cloning the PCR product into the pGEM-T Easy vector (Promega), followed by sequencing of the recombinant plasmid.
To examine developmental expression of the reproductive tract-specific β-defensins, testes were harvested from Sprague-Dawley rats on 1, 15, 30, 60 and 90 days after birth; and epididymis were collected on 7, 15, 30, 60 and 90 days after birth with 2-4 animals per time point. RNA isolation and semi-quantitative RT-PCR were performed as described above. Signal intensity of each gene-specific PCR product was quantified by using Scion Imaging software (http://www.scioncorp.com), followed by normalization against the GAPDH signal amplified from the same cDNA sample, and the results were expressed as relative level of expression.

RESULTS

Discovery of mammalian β-defensin gene repertoires

Based on the conservation of the characteristic six-cysteine-containing β-defensin motif, an iterative, genome-wide homology search strategy was developed to identify complete repertoires of the β-defensin gene family in the human, chimpanzee, mouse, rat, and dog as we previously described (36, 57). Because such computational searches are biased toward the defensin motif, the nucleotide sequences identified were primarily composed of the exons encoding the cysteine-spanning region. To identify the first exon sequences, 5-15 kb of finished or unfinished genomic sequence containing each putative defensin gene were retrieved from GenBank and used for computational prediction by using GenomeScan (61), GENSCAN (5), and/or GeneWise2 (4). As a result, a number of novel, full-length defensin precursor sequences with conserved six cysteines and signal peptides have been identified in each species and submitted to the GenBank. All new genes were named by following the recommendations of the
HUGO and Mouse Gene Nomenclature Committees and kept in consistence with the published sequences.

In the case of rat, a total of 42 β-defensin (Defb) genes and a pseudogene (Defb16-ps) were discovered, including four known ones (24, 29, 41, 68) (Fig. 1). Among all rat genes, 12 β-defensins, namely Defb1, 4, 14, 22, 24, 27, 29, 33, 36, 39, 51, and 52, were found in the EST database (data not shown), from which full-length peptide precursor sequences can be deduced (Fig. 1). All the remaining genes were identified from WGS and HTGS sequences with first exons being predicted, except for Defb19. Consistent with earlier observations, the predicted signal sequences are highly homologous with known β-defensin genes in other mammalian species, rich in leucine and hydrophobic residues (Fig. 1). The accuracy of the exon-intron boundary predicted for each gene has also been confirmed by direct sequencing of RT-PCR product with exon-spanning primers (Supplementary Table 2). Inability to identify the signal sequence for Defb19 may be due to mis-annotation of genome sequences or presence of premature stop codons in its first exon.

It is noted that all rat β-defensin genes were named according to their mouse orthologs as previously described and approved by the Mouse Nomenclature Committee (47). Consequently, rat β-defensin-1 (24), β-defensin-2 (24), E-3/2D6 (41, 68), and Bin-1b/EP2e (29) have been renamed as Defb1, Defb4, Defb22, and Spag11e, respectively. An extra internal exon encoding an additional segment of pro-sequence has been predicted for Defb12, Defb52, and Spag11c, which have also been experimentally confirmed by sequencing of RT-PCR products in both
cases. Presence of an extra exon in Spag11c has also been shown previously in its orthologs (HE2c/EP2c/SPAG11c) in the human and rhesus monkey (11, 12).

A total of 43 novel β-defensins genes and pseudogenes were discovered in the canine genome (Supplementary Fig. 1). All sequences were named so as to maintain consistency with the nomenclature of human β-defensins. The canine β-defensin-1 (CBD1) gene that we identified earlier was re-named CBD122 (44). Five other CBD genes with no human orthologs were named CBD138-CBD142. Among all canine sequences identified, eight (CBD1, CBD102, CBD103, CBD108, CBD122, CBD124, CBD139, and CBD40,) were also supported by at least one EST sequence (data not shown). First exons of all but two canine β-defensins were also predicted. While most CBD genes consist of two exons separated by an intron, CBD105, CBD118, CBD122, and CSPAG11c (canine ortholog of human SPAG11c/HE2c) were predicted to have an extra exon (data not shown).

Although a number of β-defensins have been described in humans and mice with the use of hidden Markov models (HMM) (47), later discovery of another novel sequence in the human (42) prompted us to re-screen the human and mouse genome sequences for possible presence of additional new sequences by using the strategy that we described above. Albeit of low throughput, our method yielded a total of 39 human and 52 mouse β-defensin genes and pseudogenes, including four novel human (DEFB133, 134, 135, and 136) and six new mouse β-defensins (Defb49, 50, 51, 52, 53, and 54-ps). All new β-defensin sequences consist of the characteristic signal sequence and a conserved defensin motif (Fig. 2A). Moreover, most of their orthologs are present in dogs and/or rodents except for mouse Defb53 and Defb54-ps, which
appear to be mouse-specific (Fig. 2B). To further support their authenticity, all new genes were found to locate inside the same genomic clusters as others (Fig. 3).

We also predicted first exons for most human (Supplementary Fig. 2) and mouse genes (Supplementary Fig. 3), which were lacking in the earlier publication (47). The predicted sequences and their intron-exon boundaries are consistent with the rat orthologous genes, which have been confirmed by direct sequencing. Such sequences are also in agreement with a few full-length human and mouse β-defensins as described in the follow-up studies (3, 32, 39, 42, 48, 58, 66).

We also searched the chimpanzee genome and found a repertoire of 37 β-defensin genes that share 91-100% identity to their human orthologs in each case (data not shown), except that DEFB110 and DEFB128 orthologs were not found the chimpanzee genomic sequences currently available in GenBank. Because of their conservation in dogs and rodents, there is little reason to suspect that DEFB110 and DEFB128 are missing in the chimpanzee, but instead it is most likely due to low coverage (4-fold) of the current genome sequences.

Alignment of all β-defensins revealed a high degree of diversity in the primary amino acid sequence in the mature region despite of the conservation of six cysteine residues, indicative of their possible nonredundant biological functions. Noticeably, the DEFB126/CBD126/Defb22 lineage (Fig. 1 and Supplementary Figs. 1, 2, and 3) and dog-specific CBD141 (Supplementary Fig. 1) contain additional four amino acids between the third and fourth cysteine as compared to most other β-defensins. Moreover, three canine- and rodent-specific β-defensin lineages, namely
CBD138/Defb33, CBD139/Defb52, and CBD140/Defb51, consist of 3-4 extra amino acids between the second and third cysteine but 2-3 fewer amino acids between the third and fourth cysteine (Fig. 1 and Supplementary Figs. 1 and 3). Conceivably, such variations are expected to alter their overall tertiary structures and possibly functional properties from conventional β-defensins.

Strikingly, in contrast to most characterized β-defensins that contain less than 5 amino acids following the last cysteine, the majority of the family members in fact have much longer and mostly linear C-terminal tails. The role of such long tails remains largely a mystery. However, several β-defensins were recently found to be heavily glycosylated in vivo in the tail region (30, 41, 65, 68). Furthermore, rat Defb22/E-3/2D6 and rhesus DEFB118/ESC42 contain lectin and trefoil-like motifs, respectively, which potentially facilitate oligomerization and interactions of proteins with target cell membrane (30, 41).

Chromosomal clustering of β-defensins

To identify genomic organization of β-defensins, individual β-defensin genes were searched against the most current versions of the assembled human, mouse, rat, chimpanzee, and dog genomes by using the BLAT program (25) through the UCSC Genome Browser. It is apparent that β-defensin genes in each species are clustered densely in 4-5 different chromosomal regions with each cluster spanning between 55-1118 kb (Fig. 3). All clusters are syntenic across species because orthologous genes are extensively conserved with the same order and orientation for most genes in each cluster (Fig. 3).
In the rat genome (BCM Version 3.1), all 43 β-defensin genes have been mapped and form four distinct clusters expanding 812 kb, 78 kb, 55 kb, and 193 kb continuously on chromosomes 16q12.5, 15p12, 9q12-q13, and 3q41, respectively (Fig. 3). Such syntenic clusters are also conserved in mice (NCBI Build 32) with four clusters spanning 1118 kb, 71 kb, 56 kb, and 164 kb on chromosomes 8qA1.3-A2, 14qC3, 1qA3, and 2qH1, respectively (Fig. 3). However, the genomic contig containing mouse Defb33, Defb51, and Defb52 have not been assigned to a specific region on mouse chromosome 8. The location of mouse Defb48-ps currently remains unmapped. Mouse Defb32 reported earlier (47) was found on chromosome 6. However, we suspect that Defb32 is unlikely to be an authentic β-defensin gene, because of the facts that 1) only five instead of six cysteines are present in the putative mature region, 2) such a cysteine-containing region is negatively instead of positively charged, 3) no characteristic first exon sequence could be identified within a 15-kb upstream sequence, and 4) no orthologs can be found in other mammals. It is also noteworthy that we could not find mouse Defb31 and Defb47-ps (47) in either EST, WGS, HTGS, or assembled genome sequences that are publicly available in GenBank.

The canine genome (NCBI Build 1 Version 1) also encodes four β-defensin clusters expanding 356 kb, 76 kb, 99 kb, and 373 kb on the syntenic regions on chromosomes 16, 25, 12, and 24, respectively (Fig. 3). However, human β-defensin genes in the most current genome assembly (NCBI Build 35 Version 1) were annotated in five separate clusters on three different chromosomes with two on chromosome 8p23.1 (Fig. 3A and 3B), one on 6p12.3 (Fig. 3C), and two on 20p13 and 20q11.21 (Fig. 3D). Apparently, two human β-defensin clusters on the p and q arms of chromosome 20 share a common ancestor with a continuous synthetic cluster in
rodents and dogs (Fig. 3D). In contrast, another two human gene clusters on 8p23.1 are only separated by less than 4 Mb, but their counterparts in rodents and dogs are located on two different chromosomes (Fig. 3A and 3B). The current chimpanzee genome (NCBI Build 1 Version 1) also encodes five major β-defensin clusters in almost perfect synteny with human genes (Supplementary Fig. 4). It is noteworthy that the α-defensin cluster is located within the largest β-defensin cluster in rodents and humans, but is apparently missing in the canine genome (Fig. 3A), consistent with the fact that α-defensins have only been discovered in glires and primates (36).

Earlier studies indicated that a varied copy number of α- and β-defensin genes are present on chromosome 8p23.1 of human individuals (19, 51). Because of such complexity, current annotation of several human β-defensins in this region, such as DEFBI07, 108, 109, 130, and 131, shows an obvious discrepancy in the location and order of their orthologous genes with other species (Fig. 3A and 3B). Thus, the accuracy of the current human genome assembly in this region remains to be refined and corrected in the future (51).

Nevertheless, our current version of human and mouse β-defensin genomic clusters manifests a significant improvement over an earlier attempt (47). For example, several genes in human 8p23.1 and their orthologous loci in mice have been re-ordered, which agrees with their syntenies in other species (Fig. 3A) and are also consistent with more recent studies (48, 51). Moreover, an ambiguous cluster containing DEFBI30 and DEFBI31 (47) has been mapped to human chromosome 8p23.1 with three additional novel genes (DEFBI34, 135, and 136) being
added (Fig. 3B). Another novel human β-defensin gene (DEFB133) has also been mapped to 6p12.3 and its orthologs have been found and localized in rodent and canine genomes (Fig. 3C).

**Structural organizations and transcriptional flexibility of β-defensin genes**

A comparison of all available EST with genomic sequences of β-defensins across mammalian species revealed that, similar to enteric α-defensins (35), most β-defensin genes are composed of two separate exons with the first exon encoding the signal and pro-sequence and the second exon encoding primarily the mature sequence containing the defensin motif. The intron sizes of β-defensins vary greatly, spanning 1-10 kb in most cases, in contrast with α- and θ-defensins consisting of introns of usually <1 kb (36). However, we also found additional five distinct patterns of the β-defensin gene structure (Fig. 4). For example, mouse Defb21 (EST accession number AK076875), mouse Defb44 (AK079042), rat Defb29 (CK839317), rat Defb51 (BM390654), and canine CBD102 (DN346173), CBD139 (CO687407), and CBD140 (CO692331) have EST sequences showing the presence of additional one or two exons encoding the 5'-untranslated region (UTR), reminiscent of myeloid α-defensin genes (35). On the other hand, mouse Defb30 (AK078987) contains an additional exon encoding the 3'-UTR. Mouse Defb28 (AV044615) and canine CBD122 (BM537999) are unique in that they are composed of three exons with the last exon encoding a segment of the mature sequence downstream of the six cysteines, while the first two exons encoding the signal, pro- and, majority of mature sequences. A few genes, such as SPAG11c /CSPAG11c/Spag11c (11, 12) and DEFB105/CBD105/Defb12 (data not shown), exhibit a three-exon structure, containing an extra intron presumably encoding an additional segment of pro-sequence. Interestingly, rat Defb52 (CR466423) is composed of four exons with three exons encoding the entire open reading frame and the fourth exon
encoding a part of 5'-UTR. Additional gene structures are expected to be revealed with availability of more mRNA and EST sequences for β-defensins.

It is well known that the SPAG11 gene is capable of differentially utilizing different exons to produce multiple isoforms of β-defensin-like sequences (11, 12). Such alternative splicing now does not appear to be unique to the SPAG11 gene. We have found two EST sequences (AK020304 and BB787829) that consist of the first exon of mouse Defb17 and the second exon of Defb16, suggesting that these two genes are actually alternatively spliced isoforms. Such splicing is most likely evolutionarily conserved, because of the absence of signal sequence for its orthologous gene, DEFB110/CBD110/Defb16 in the human, chimpanzee, dog, and rat. Therefore, primate DEFB110 and DEFB111, canine CBD110 and CBD111, and rat Defb16-ps and Defb17 are likely to be two isoforms of the same gene. Similarly, human DEB119 and DEFB120 have been shown recently to share the same first exon encoding the signal sequence (39). CBD122, on the other hand, has three isoforms sharing identical signal, pro- and the majority of mature sequences with the only difference in the carboxyl terminal tail after the last cysteine (44). Collectively, acquirement of additional exons and frequent alternative splicing add additional layers of complexity in β-defensin gene transcription and regulation, which presumably allow the host to better cope with invading infections.

*Phylogenetic analysis of β-defensins*

To reveal the evolutionary relationships of all known β-defensins including newly identified ones in rodents, dogs, and primates, we sought to construct a phylogenetic tree by using the neighbor-joining method (43). Interestingly, genes within a genomic cluster tend to
form a separate clade from the genes in other clusters (data not shown), implying that many of them most likely evolved by gene duplication events. Indeed, a number of highly similar paralogous genes are present in each species and physically located adjacent to each other. Consistent with the genetic mapping data (Fig. 3), many β-defensins have obvious orthologs with minimum sequence variation across four mammalian species and are clustered together (Fig. 5), meaning that these lineages arose before the last common ancestor of rodents, canines, and primates. Some even form well-supported clusters with different chicken β-defensins, suggesting their existence prior to the divergence of mammals from birds (57). The conservation of these genes during evolution may be indicative of their functional significance.

Conversely, subgroups of β-defensins also exist that are specific in mice, rats, and dogs, implying that these genes originated after these species separated and have undergone different evolutionary patterns. For example, two well-supported clades of rodent-specific genes (Clade I: Defb2, 9, 10, and 11; and Clade II: Defb37, 38, 39, 40, and 50) exist (Fig. 5). High sequence similarity within each clade supports the notion that these rodent-specific β-defensin genes arose from gene duplication and positive diversifying selection as previously described (32, 48). In contrast to the Clade II genes with the orthologs present in both rodent species, many mouse β-defensins in Clade I have no obvious orthologs in rats, implying that many Clade I genes were likely to have emerged only after the mouse-rat split about 41 million years ago (27), whereas all Clade II genes apparently appeared prior to the split. Another large cluster of rodent genes, namely Defb3, 4, 5, 6, 7, and 8, form a common clade with human DEFB4/hBD-2 gene (Fig. 5). Therefore, it is likely that such a gene lineage duplicated and expanded significantly in rodents, but remained unchanged in primates and was lost in the canine lineage. Alternatively, multiple
such ancestral gene lineages were lost in primates and canines, but retained in rodents. However, significant homology of murine genes in this cluster as well as their physical vicinity (Fig. 3A) supports the former conclusion.

Multiple gene lineages specific to dogs and primates are also present, including \textit{DEFB104/CBD104}, \textit{DEFB108/CBD108}, \textit{DEFB114/CBD114}, \textit{DEFB120/CBD120}, \textit{DEFB121/CBD121}, \textit{DEFB133/CBD133}, and \textit{DEFB134/CBD134} (Figs. 3 and 5). Conversely, \textit{CBD138/Defb33}, \textit{CBD139/Defb52}, and \textit{CBD140/Defb51} (Fig. 3A) only exist in rodents and dogs, but are absent in primates (Fig. 5). Dog-specific genes also exist, such as \textit{CBD141} and \textit{CBD142} (Fig. 3D). However, no primate-specific \(\beta\)-defensins were found, suggesting that gene duplication did not occur in the primates after their divergence. It is noteworthy that \textit{CBD102} is not the canine ortholog of human \textit{DEFB4}, murine \textit{Defb2} or \textit{Defb4}, but instead it is specific to the dog, paralogous to \textit{CBD103} (Fig. 5 and Supplementary Fig. 1). Therefore, the \textit{CBD102} gene is most likely to evolve from duplication of the \textit{CBD103} gene after separation of canines from other mammals.

\textit{Tissue expression pattern of rat \(\beta\)-defensins}

We next examined the tissue expression patterns of the entire \(\beta\)-defensin gene family in the rat as a first step toward understanding their \textit{in vivo} biological functions. Semi-quantitative RT-PCR was performed with all identified rat \(\beta\)-defensin genes by using a panel of 29 different tissues from healthy, 2 month-old Sprague-Dawley rats. Primers were designed from two different exons to differentiate between the amplicons from cDNA and those from genomic DNA and verify predicted exon-intron junctions (Supplementary Table 2). Different numbers of
PCR cycles were performed for each gene to ensure the linearity of amplification. Specificity of each amplicon was confirmed by direct sequencing of the RT-PCR product or recombinant plasmid containing the RT-PCR product.

To our surprise, nearly all β-defensins were found to be expressed preferentially in the male reproductive system particularly in epididymis and testis (Fig. 6A) with no or minimum expression in most other tissues (data not shown). More strikingly, the reproductive tract-specific genes are expressed differentially in three different regions of epididymis, namely caput (head), corpus (body), and cauda (tail) (Fig. 6A). For example, Defb12/35, 15/34, 17, 18, 21, 25, 41, Spag11c, and Spag11e are more abundantly expressed in the caput of epididymis, whereas Defb9, 10, 11, 13, and 40 are restricted to the cauda (Fig. 6A). On the other hand, Defb24 and Defb33 are highly specific to testis (Fig. 6A). Our results are consistent with earlier studies on the expressions of rat Defb1/RBD-1 (8, 24) and Spag11e/Bin-1b (29) and also largely agree with the expressions of their human and mouse orthologs (39, 42, 58, 59, 66).

In addition to a canonical Spag11c/EP2c transcript (3, 12, 56), an alternatively spliced longer transcript was detected in the rat (Fig. 6A) and has been deposited in GenBank under accession number DQ012093. Such a transcript is unique in that it contains an extra 60 bp at the 3’-end of the first exon, resulting in a longer pro-sequence, which is not found in either human or rhesus monkey (11, 12). To further illustrate the difference in the SPAG11/HE2/EP2 locus between primates and rodents, the d isoform, which is composed of the first two exons of the c form and the last exon of the e form, was not detected in the rat (data not shown), but is readily detectable in the primates (11, 12).
Although the majority of genes showed the most abundant expression in testis or epididymis of healthy adult rats, *Defb1* and *Defb42* are also expressed at considerable levels in kidney (Fig. 6A) and *Defb24* in spleen and ovary (data not shown). *Defb36* was found at a lesser extent on mucosal surfaces lining the respiratory, gastrointestinal, and reproductive tracts as well as on skin (Fig. 6B). The only obvious exception is *Defb4/RBD-2*, which shows virtually no expression in either testis or epididymis, but instead is more restricted to the respiratory tract, particularly in lung (Fig. 6B). *Defb4* transcript was also found at low levels in the upper gastrointestinal tract as well as in vagina (Fig. 6B), which is consistent with the previously published report (24). The biological significance of such concentrated, but apparently differential production of so many β-defensins in testis and different regions of epididymis remains largely unknown.

It is noted that we failed to detect the expression of *Defb3*, *Defb5*, and *Defb16-ps* following multiple attempts with different combinations of primer pairs. Although we cannot rule out possibility of wrong computational prediction of first exons, it is most likely that *Defb3* and *Defb5* are expressed at extremely low levels in healthy adult rats, but may be inducible in response to infection and inflammation. *Defb16-ps* could be a non-expressing pseudogene. Alternatively, these three genes could be expressed in the tissues other than the ones that we examined. Because of the inability to identify the first exon of *Defb19*, we could not detect its gene expression by RT-PCR either.
Developmental regulation of rat β-defensins

Semi-quantitative RT-PCR was performed to study developmental expression patterns of a few selected rat genes that are specific to epididymis and/or testis. Rat testes were collected on 1, 15, 30, 60, and 90 days after birth; and epididymis were collected on 7, 15, 30, 60, and 90 days after birth with 2-4 animals used per time point. As shown in Fig. 7A, expressions of all selected β-defensins (Defb1, 15, 29, 30, 42, and 49) in epididymis are significantly elevated during development with virtually no expression on day 7 but peaked quickly at one month. Defb42 is slightly different in that it shows little expression up to one month, but the expression level is dramatically enhanced at two months of age and increased up to three months after birth. Such a pattern is reminiscent of rat Spag11e/Bin-1b, whose transcript does not peak in epididymis until four months and remains elevated in rats of 2 years of age (29).

However, β-defensin genes in testis showed a different pattern. Although most epididymal β-defensin genes were not expressed on day 7, many genes in testis display considerable expression soon after birth, with some (Defb24 and Defb29) expressed constitutively in rats of all ages studied and some (Defb27 and Defb36) were upregulated and peaked at one month of age (Fig. 7B). Defb33, on the other hand, behaves similar to epididymal Defb42 (Fig. 7A) in that it is not detected in testis until one month, but continues to increase for at least three months after birth (Fig. 7B). It is interesting to note that Defb29 is developmentally regulated in the epididymis (Fig. 7A), but exhibits a constitutive pattern in the testis (Fig. 7B), implying the presence of certain epididymis-specific enhancer(s) or testis-specific inhibitory factor(s).
The constitutive expression of several β-defensins in testis, but not epididymis, of newborn animals is presumably required to protect testis and particularly germ cells from invading infections before the time of sexual maturation. Simultaneous production of such a large array of defensins in various segments of the male reproductive tract with different developmental expression patterns implies that these molecules may play a nonredundant role in the reproductive process.

DISCUSSION

Sequence diversity of mammalian β-defensins

Through systemic genome-wide screening, we have discovered a total of 39, 37, 43, 52, and 43 β-defensin genes and pseudogenes in the human, chimpanzee, dog, mouse, and rat, respectively. Alignment of all β-defensins revealed a high degree of conservation in the spacing pattern of six cysteines with a consensus pattern of C-X$_{5.8}$-C-X$_{3.7}$-C-X$_{5.13}$-C-X$_{4.7}$-CC. The invariantly spaced cysteines form a rigid, triple antiparallel β-sheet structure and are believed to be critically important in maintaining the functional activities (14, 28). However, we observed the absence of a canonical cysteine or presence of an extra cysteine in a few putative functional defensins. For example, human DEFB107 and mouse Defb8/Defr1 have a nonconservative mutation in the first cysteine, whereas mouse and rat Defb50 lack the second conserved cysteine (Supplementary Figs. 2 and 3). DEFB133/Defb49 lineage, on the other hand, contains a mutation in either the fifth or sixth cysteine in the mouse and human (Supplementary Figs. 2 and 3), but is fully conserved with six canonical cysteines in the rat (Fig. 1).
Extra cysteines have also been found within the six-cysteine motif of three peptides, namely human DEFB119 and DEFB132 and mouse Defb5 (Supplementary Figs. 2 and 3). In contrast, the instance of cysteines occurring outside the defensin motif appears to be much higher with multiple members being found in each species. Among 39 human β-defensins, seven such peptides are found (DEFB105, 106, 112, 117, 126, 127, and 133) (Supplementary Fig. 2). Interestingly, five are conserved across species, including DEFB105/CBD105/Defb12/Defb35, DEFB106/CBD106/Defb15/Defb34, DEFB112/CBD112, DEFB117/CBD117/Defb19, and DEFB126/CBD126/Defb22.

It is largely unknown how a missing or an extra cysteine would affect the tertiary structures and functional properties of β-defensins. Obviously, the existence of an odd number of cysteines will potentially oligomerize through intermolecular disulfide bridging. Indeed, such β-defensins as Defb8/Defr1 (6) and DEFB126/Defb22/ESP13.2/E-3/2D6 (41, 65, 68) have been shown to form covalently bounded homodimers in vitro and/or in vivo. Such dimerization in fact leads to enhanced antibacterial activities in the case of Defb8/Defr1 (6). Not surprisingly, a few β-defensins with six cysteines also have a tendency to form dimers or oligomers, which are believed to facilitate membrane interaction and permeabilization (20, 45).

It is noteworthy that our strategy yielded a total of 10 additional novel β-defensins in humans and mice as compared with the strategy based on the Hidden Markov Model (HMM) (47). Although we believe that the current list most likely represents the complete repertoires of β-defensins in these species, we could not rule out the possibility that additional β-defensin-related genes with remote similarity might be uncovered in these species. In fact, a group of
human SPAG11/HE2/EP2 isoforms were recently found to share identical signal and pro-sequences with similar antimicrobial properties with classic β-defensins, but differ significantly in the number and cysteine spacing pattern in the mature sequence (3, 56, 62, 63). Such α-defensin-related sequences have also been discovered in mice (21, 23) and rats (36). Nevertheless, the current study represents the most comprehensive attempt to identify mammalian canonical β-defensin gene family members with the characteristic six-cysteine motif.

_Evolutionary relationships of mammalian β-defensins_

β-defensins are the most ancient family of vertebrate defensins as compared to α- and θ-defensins (57). β-defensin-like sequences have been found in the rattlesnake (34), orange-spotted grouper (GenBank accession number AY129305), and zebrafish (Zhang, G., unpublished results), suggesting that the primordial gene for β-defensins appeared prior to the fish-primate split about 450 million years ago (27). In contrast, α-defensins were derived from β-defensins after divergence of glires and rodents from other mammals about 91 million years ago, whereas θ-defensins did not appear until primates separated from other mammals around 23 million years ago (27, 33, 36).

Our earlier study discovered a single cluster consisting of 13 β-defensin genes encoded in the entire chicken genome (57). Comparative analysis of the chicken and mammalian β-defensin gene clusters revealed that two clusters on human 8p23.1 and their orthologous loci in other mammalian species (Fig. 3A and 3B) are syntenic with the chicken gene cluster, implying that the ancestral genes in these two mammalian clusters evolved prior to the bird-mammal
separation (57). Many genes in these two ancient clusters were demonstrated to have undergone significant repeated duplication and positive diversification in humans and mice after their divergence (32, 48), which further gave rise to other β-defensin clusters in mammals during evolution, presumably as a result of chromosomal translocation and expansion of certain gene lineages.

Consistent with the evolutionarily active nature of the two ancient clusters, subgroups of genes exist in these two regions that are specific in rodents, dogs and/or primates (Fig. 5). This notion is further reinforced by the presence of multiple highly homologous α- and β-defensins and emergence of α- and θ-defensins in rodents and primates in these regions (Fig. 3A). In contrast, the remaining genomic clusters are rather static throughout evolution with the presence of orthologs across mammalian species, except that CBD141 and CBD142 on chromosome 24 are specific to dogs and was likely duplicated and diverged from adjacent genes (Fig. 3D) after the dog diverged from other mammals.

Collectively, these results suggest that the evolution of mammalian β-defensins is an extremely dynamic and active process. Individual gene lineages were derived at different evolutionary times. Although some β-defensins evolved prior to divergence of mammals from fish, many appear to be mammal-specific with a few emerging as recently as the rat-mouse split, which occurred only 41 million years ago (27). However, no gene duplication occurred prior to the separation of chimpanzee from humans at about 5.5 million years ago (27), because of the complete orthology of β-defensin genes in these two species (Supplementary Fig. 4).
What is the driving force for the presence of such a large number of divergent β-defensins in mammals? One plausible explanation is that sequence diversification may confer functional novelties on different β-defensins for each species to better cope with a variety of microbial threats from the environment. Consistent with this hypothesis, different β-defensins (15, 16, 18) and even orthologous genes in different primates (2) have been found to differ quite dramatically in their antimicrobial potency and spectrum. Furthermore, expression of a large array of β-defensins in the male reproductive tract is presumably to safeguard the reproductive process, enhance fertility, and sustain species survival, as detailed in the next section.

**Dual role of region-specific expression of β-defensins in the male reproductive tract**

Obviously, optimal reproductive function ensures survival of the species, which needs full protection of germ cell precursors in testis, sperm in different stages of maturation in epididymis and after deposition in the female reproductive tract. Infection of testis and epididymis may result in temporary or permanent loss of fertility by disrupting specialized environment of these organs conducive for sperm storage and maturation (37). As adaptive immunity is largely absent in the male reproductive system, the host must evolve alternative and effective mechanisms of protection of sperm from infectious agents.

Region-specific expression of all but one rat β-defensins in the male reproductive system (Fig. 6) suggests that these molecules may constitute an essential component in maintaining the normal reproductive process. In line with this, many testis- and epididymis-specific β-defensins were shown to be antimicrobial and capable of protecting sperm from infections (3, 8, 29, 56, 62, 64). In addition to being microbicides, β-defensins expressed in different regions of epididymis
appear to be actively involved in the reproduction. These defensins are mainly produced by epithelial cells lining the epididymal duct, regulated by androgens, secreted in luminal plasma, and bind preferentially to the surface of maturing but not immature sperm (29, 30, 41, 65, 68, 70). Rat Spag11e/Bin-1b has been shown to enhance sperm maturation by inducing Ca\(^{2+}\) uptake and subsequent motility and progressive movement of immature spermatozoa (70). Moreover, another epididymis-specific β-defensin, macaque DEFB126/ESP13.2, coats the entire ejaculated sperm and masks zona pellucida ligands on the sperm surface, but becomes dissociated when sperm are fully capacitated, suggesting that DEFB126 may be an important decapacitation factor on the sperm surface that needs to be removed prior to sperm-zona interaction and fertilization (54, 65).

Besides β-defensins, α-defensins have also been reported in the male reproductive tract (8, 17). Human cathelicidin LL-37/hCAP-18, which belongs to another important family of vertebrate antimicrobial peptides (40, 67), is expressed by epididymal epithelia and is abundantly present in the seminal fluid with ability to bind to sperm (31). Furthermore, the female reproductive tract also produces α- and β-defensins albeit at low levels (26, 38, 50). Coupled with the fact that many β-defensins in epididymis and testis are developmentally regulated with elevated expression at the time of puberty and sexual maturation (Fig. 7), these results strongly favor the argument that β-defensins, perhaps together with many other antimicrobial peptides, have a dual function in both innate host defense and fertility.

Defb4/RBD-2 represents the only β-defensin that is expressed most abundantly in lung with virtually no expression in the reproductive tract (Fig. 6B), and thus it may play an important
role in airway defense. Its human ortholog, DEFB4/hBD-2 was shown to be the predominant β-defensin in human neonatal lung and developmentally regulated (49) and its concentration is inversely correlated with the severity of lung disease in cystic fibrosis patients (7). Further studies are needed to identify the regulatory mechanisms that specify the expression of defensins in the male reproductive tract vs. respiratory tract. Functional divergence of β-defensins produced in the same and different regions of the reproductive tract also remains to be investigated and these molecules may have potential as fertility and contraceptive drugs.

ACKNOWLEDGEMENTS

We would like to thank Lin Liu and Kathy Swenson for assistance with rat tissue harvesting and Yanjing Xiao for her initial help with the developmental regulation of rat β-defensins.

GRANTS

This work was supported by the Oklahoma Center for the Advancement of Science and Technology Grant HR03-146 and Oklahoma Agricultural Experiment Station project H-2507.

REFERENCES


**FIGURE LEGENDS**

Fig. 1. Amino acid sequence alignment of rat β-defensins. Dashes are inserted to optimize the alignment and conserved residues are shaded. Internal exon sequences of *Defb12, Defb52, and Spag11c* have been deleted to maximize the alignment. Long carboxyl terminal tails of several β-defensins as well as an additional segment of amino terminal sequence preceding the canonical start codon of *Defb17* are not shown for the sake of clarity. The letter ‘X’ denotes the presence of premature stop codon in the pseudogene. All computationally predicted novel rat β-defensin sequences except for *Defb3, Defb5*, and *Defb16p* have been confirmed by sequencing of their respective RT-PCR products amplified from appropriate tissues. The failure to amplify *Defb3, Defb5*, and *Defb16p* is likely due to no or extremely low expression in healthy animals. Abbreviations: *Defb*, β-defensin; *SPAG11*, sperm-associated antigen 11; *p*, pseudogene.
Fig. 2. Identification of novel β-defensins in humans and mice. A, Amino acid sequence alignment of novel sequences with DEFB1/Defb1. Dashes are inserted to optimize the alignment and conserved residues are shaded. The letter ‘X’ denotes the presence of premature stop codon in the pseudogene. B, Phylogenetic tree of novel β-defensins in the human and mice demonstrating their conservation in other mammalian species, except for mouse Defb53 and Defb54p. The tree was constructed by the neighbor-joining method based on the proportion difference (p-distance) of aligned amino acid sites of full-length peptide sequences. 1000 bootstrap replicates were used to test the reliability of each branch. Numbers on the branches indicate the percentage of 1000 bootstrap samples supporting the branch. Only branches supported by a bootstrap value of at least 50% are shown. Abbreviations: DEFB/Defb, β-defensin; CBD, canine β-defensin; mDefb, mouse β-defensin; rDefb, rat β-defensin; P/p, pseudogene.

Fig. 3. Genomic organization of syntenic mammalian β-defensin gene clusters. Four different β-defensin clusters (A, B, C, and D) are present in the rat, mouse, and dog, whereas five clusters exist in the human and chimpanzee due to the split of the Cluster D in the rodents and dog into two separate loci in the primates. The position of each gene is represented by a vertical bar. The direction of transcription is indicated by the triangle. The orthologous genes are linked across species and paralogous genes are also connected with dashed lines. However, in panel A, murine Defb13, Defb15, and Defb35 that are orthologous to DEFB107/CBD107, DEFB106/CBD106/Defb34, and DEFB105/CBD105/Defb12 gene lineages, respectively, are not linked for simplicity. Mouse Defb2, 9, 10, and 11 are not orthologs of the rat genes with corresponding numbers, and therefore they are not linked. The gene mapping was derived from
the assembled genomes of human (NCBI Build 35 version 1), rat (BCM version 3.1), mouse (NCBI Build 32), and dog (NCBI Build 1 Version 1), respectively. The genomic cluster containing mouse Defb33, 51, and 52 is not mapped on chromosome 8, and therefore, is placed within slanted lines based on the locations of their orthologs on rats and dogs. Note that an α-defensin gene cluster is located within a β-defensin locus in the human, mouse, and rat, but is absent in the dog genome (A). Abbreviations: Tel, telomere; Cen, centromere; P/p, pseudogene; 11e/11c, SPAG11e/SPAG11c.

Fig. 4. Structural organizations of mammalian β-defensin genes. Although most β-defensin genes are composed of two exons (boxes) separated by an intron (lines) of variable length, certain genes in mice (Defb21 and Defb44), rats (Defb29 and Defb51), and dogs (CBD102, CBD139, and CBD140) contain additional one or two exons encoding 5’-untranslated regions. On the other hand, mouse Defb30 consists of an extra 3’-untranslated region. An extra exon encoding the carboxyl terminal mature sequence exists in mouse Defb28 and canine CBD122, whereas an internal extra pro-sequence is present in SPAG11c/SPAG11c/Spag11c and DEFB105/CBD105/Defb12 gene lineages. In addition to an extra internal exon, rat Defb52 comprises a fourth exon encoding the 5’-untranslated region. Solid boxes refer to the coding exons, whereas the open boxes refer to the noncoding exons.

Fig. 5. Phylogenetic tree of mammalian β-defensins. Only β-defensins in the rat chromosome 16q12.5 and its syntetic loci in the human, dog, and mouse are shown. The tree was constructed by the neighbor-joining method based on the proportion difference (p-distance) of aligned amino acid sites of full-length peptide sequences. 1000 bootstrap replicates were used to test the
reliability of each branch. Numbers on the branches indicate the percentage of 1000 bootstrap samples supporting the branch. Only branches supported by a bootstrap value of at least 30% are shown. Abbreviations: DEFB, human β-defensin; CBD, canine β-defensin; mDefb, mouse β-defensin; rDefb, rat β-defensin; SPAG11/spag11, sperm-associated antigen 11; CSPAG11, canine sperm-associated antigen 11; P/p, pseudogene;

Fig. 6. Tissue expression patterns of rat β-defensins by RT-PCR. Lanes: 1, kidney; 2, testis; 3, caput of epididymis; 4, corpus of epididymis; 5, cauda of epididymis; and 6, vas deferens. Defb12/35 and Defb15/34 have identical expression patterns and are therefore combined in a single panel. Notice that most β-defensins are expressed preferentially in the male reproductive system with no or minimum expression in other tissues (A), except that Defb4 and Defb36, which have more universal distribution patterns (B). See Materials and Methods for details. The number of PCR cycles was optimized for each gene, and the specificity of each PCR product was confirmed by sequencing. The house-keeping gene, GAPDH, was used for normalization of the template input.

Fig. 7. Developmental regulation pattern of rat β-defensins in epididymis (A) and testis (B) by RT-PCR. The signal intensity of each gene-specific PCR product was digitized and quantitated, followed by normalization against the GAPDH signal amplified from the same cDNA sample. The results are expressed as means ± SEM of 2-4 animals per data point.
Fig. 1
Fig. 2
Fig. 3
Fig. 4
Fig. 5
<table>
<thead>
<tr>
<th>Defb1</th>
<th>Defb2</th>
<th>Defb9</th>
<th>Defb10</th>
<th>Defb11</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defb12/35/15/34</td>
<td>Defb13</td>
<td>Defb14</td>
<td>Defb17</td>
<td>Defb18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defb20</td>
<td>Defb21</td>
<td>Defb22</td>
<td>Defb23</td>
<td>Defb24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defb25</td>
<td>Defb26</td>
<td>Defb27</td>
<td>Defb28</td>
<td>Defb29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defb30</td>
<td>Defb33</td>
<td>Defb37</td>
<td>Defb38</td>
<td>Defb39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defb40</td>
<td>Defb41</td>
<td>Defb42</td>
<td>Defb43</td>
<td>Defb44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spag11c</td>
<td>Spag11e</td>
<td>Defb49</td>
<td>Defb51</td>
<td>Defb52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>Defb4</th>
<th>Defb36</th>
<th>GAPDH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6
Fig. 7
SUPPLEMENTARY MATERIAL

Supplementary Fig. 1. Amino acid sequence alignment of the canine β-defensin gene family. Dashes are inserted to optimize the alignment and conserved residues are shaded. Internal exon sequences of CBD105 and CSPAG11c as well as the last several amino acids of the first exon sequence of CBD112 have been deleted to maximize the alignment. Long carboxyl terminal tails of several β-defensins are not shown for the sake of clarity. The letter ‘X’ denotes the presence of premature stop codon in the pseudogene. First exons for some β-defensins were not identified. Abbreviations: CBD, canine β-defensin; CSPAG11, canine sperm-associated antigen 11; P, pseudogene.

Supplementary Fig. 2. Amino acid sequence alignment of the human β-defensin gene family. Dashes are inserted to optimize the alignment and conserved residues are shaded. Internal exon sequences of DEFB105 and SPAG11c as well as the last several amino acids of the first exon sequence of DEFB112 have been deleted to maximize the alignment. Long carboxyl terminal tails of several β-defensins are not shown for the sake of clarity. The letter ‘X’ denotes the presence of premature stop codon in the pseudogene. First exons for some β-defensins were not identified. Abbreviations: DEFB, β-defensin; SPAG11, sperm-associated antigen 11; P, pseudogene.

Supplementary Fig. 3. Amino acid sequence alignment of the mouse β-defensin gene family. Dashes are inserted to optimize the alignment and conserved residues are shaded. Internal exon sequences of Defb12, Defb52, and Spag11c have been deleted to maximize the alignment. Long
carboxyl terminal tails of several β-defensins are not shown for the sake of clarity. The letter ‘X’ in bold denotes the presence of premature stop codon in the pseudogene. First exons for some β-defensins were not identified. Abbreviations: Defb, β-defensin; p, pseudogene.

Supplementary Fig. 4. Genomic organization of five syntenic β-defensin gene clusters in the human and chimpanzee. The position of each gene is represented by a vertical bar. The direction of transcription is indicated by the triangle. The orthologous genes are linked across species. The gene mapping was derived from the assembled genomes of human (NCBI Build 35 Version 1) and chimpanzee (NCBI Build 1 Version 1), respectively. It is noted that chimpanzee DEFB130 and DEFB131 are not conclusively mapped, whereas DEFB110 and DEFB128 are missing in the current genome assembly. Abbreviations: Tel, telomere; Cen, centromere; P, pseudogene; 11e/11c, SPAG11e/SPAG11c.
Supplementary Fig. 1
Supplementary Fig. 2
## Supplementary Table 1. GenBank accession numbers of mammalian β-defensin cDNA sequences

<table>
<thead>
<tr>
<th>Human</th>
<th>Chimpanzee</th>
<th>Dog</th>
<th>Mouse</th>
<th>Rat</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-defensin</td>
<td>Accession No.</td>
<td>β-defensin</td>
<td>Accession No.</td>
<td>β-defensin</td>
</tr>
<tr>
<td>DEFBI</td>
<td>NM_005218</td>
<td>DEFBI</td>
<td>NM_001009151</td>
<td>CBD1</td>
</tr>
<tr>
<td>DEFBI04</td>
<td>NM_004942</td>
<td>DEFBI02</td>
<td>NM_001009076</td>
<td>CBD102</td>
</tr>
<tr>
<td>DEFBI03</td>
<td>NM_018661</td>
<td>DEFBI03</td>
<td>DQ012056</td>
<td>CBD103</td>
</tr>
<tr>
<td>DEFBI04</td>
<td>NM_080389</td>
<td>DEFBI04</td>
<td>DQ012057</td>
<td>CBD104</td>
</tr>
<tr>
<td>DEFBI05</td>
<td>NM_152250</td>
<td>DEFBI05</td>
<td>DQ012058</td>
<td>CBD105</td>
</tr>
<tr>
<td>DEFBI06</td>
<td>NM_152251</td>
<td>DEFBI06</td>
<td>DQ012059</td>
<td>CBD106</td>
</tr>
<tr>
<td>DEFBI07</td>
<td>DQ119826</td>
<td>DEFBI07</td>
<td>DQ012060</td>
<td>CBD107</td>
</tr>
<tr>
<td>DEFBI08</td>
<td>NM_001002035</td>
<td>DEFBI08P</td>
<td>DQ012061</td>
<td>CBD108</td>
</tr>
<tr>
<td>DEFBI09</td>
<td>DQ012013</td>
<td>DEFBI09</td>
<td>DQ012062</td>
<td>CBD109</td>
</tr>
<tr>
<td>DEFBI10</td>
<td>DQ012014</td>
<td>DEFBI11</td>
<td>DQ012063</td>
<td>CBD110</td>
</tr>
<tr>
<td>DEFBI11</td>
<td>DQ012015</td>
<td>DEFBI12</td>
<td>DQ012064</td>
<td>CBD111</td>
</tr>
<tr>
<td>DEFBI12</td>
<td>DQ012016</td>
<td>DEFBI13</td>
<td>DQ012065</td>
<td>CBD112</td>
</tr>
<tr>
<td>DEFBI13</td>
<td>DQ012017</td>
<td>DEFBI14</td>
<td>DQ012066</td>
<td>CBD113</td>
</tr>
<tr>
<td>DEFBI14</td>
<td>DQ012018</td>
<td>DEFBI15</td>
<td>DQ012067</td>
<td>CBD114</td>
</tr>
<tr>
<td>DEFBI15</td>
<td>DQ012019</td>
<td>DEFBI16</td>
<td>DQ012068</td>
<td>CBD115P</td>
</tr>
<tr>
<td>DEFBI16</td>
<td>DQ012020</td>
<td>DEFBI17</td>
<td>DQ012069</td>
<td>CBD116</td>
</tr>
<tr>
<td>Gene</td>
<td>Accession</td>
<td>Gene</td>
<td>Accession</td>
<td>Gene</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>---------</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td>DEFB117</td>
<td>DQ012021</td>
<td>DEFB118</td>
<td>DQ012070</td>
<td>CBD117</td>
</tr>
<tr>
<td>DEFB118</td>
<td>NM_054112</td>
<td>DEFB119</td>
<td>DQ012071</td>
<td>CBD118</td>
</tr>
<tr>
<td>DEFB119</td>
<td>NM_153289</td>
<td>DEFB120</td>
<td>DQ012072</td>
<td>CBD119</td>
</tr>
<tr>
<td>DEFB120</td>
<td>NM_153323</td>
<td>DEFB121</td>
<td>DQ012073</td>
<td>CBD120</td>
</tr>
<tr>
<td>DEFB121</td>
<td>NM_001011878</td>
<td>DEFB122P</td>
<td>DQ012074</td>
<td>CBD121</td>
</tr>
<tr>
<td>DEFB122P</td>
<td>AY122474</td>
<td>DEFB123</td>
<td>DQ012075</td>
<td>CBD122</td>
</tr>
<tr>
<td>DEFB123</td>
<td>NM_153324</td>
<td>DEFB124</td>
<td>DQ012076</td>
<td>CBD123</td>
</tr>
<tr>
<td>DEFB124</td>
<td>DQ119827</td>
<td>DEFB125</td>
<td>DQ012077</td>
<td>CBD124</td>
</tr>
<tr>
<td>DEFB125</td>
<td>NM_153325</td>
<td>DEFB126</td>
<td>DQ012078</td>
<td>CBD125</td>
</tr>
<tr>
<td>DEFB126</td>
<td>NM_030931</td>
<td>DEFB127</td>
<td>DQ012079</td>
<td>CBD126</td>
</tr>
<tr>
<td>DEFB127</td>
<td>NM_139074</td>
<td>DEFB128</td>
<td>DQ012080</td>
<td>CBD127</td>
</tr>
<tr>
<td>DEFB128</td>
<td>AF525930</td>
<td>DEFB130</td>
<td>DQ012081</td>
<td>CBD128</td>
</tr>
<tr>
<td>DEFB129</td>
<td>NM_080831</td>
<td>DEFB131</td>
<td>DQ012082</td>
<td>CBD129</td>
</tr>
<tr>
<td>DEFB130</td>
<td>DQ012022</td>
<td>DEFB132</td>
<td>DQ012083</td>
<td>CBD130</td>
</tr>
<tr>
<td>DEFB131</td>
<td>AF529414</td>
<td>DEFB133</td>
<td>DQ012084</td>
<td>CBD131</td>
</tr>
<tr>
<td>DEFB132</td>
<td>NM_207469</td>
<td>DEFB134</td>
<td>DQ012085</td>
<td>CBD132</td>
</tr>
<tr>
<td>DEFB133</td>
<td>DQ012023</td>
<td>DEFB135</td>
<td>DQ012086</td>
<td>CBD133P</td>
</tr>
<tr>
<td>DEFB134</td>
<td>DQ012024</td>
<td>DEFB136</td>
<td>DQ012087</td>
<td>CBD134</td>
</tr>
<tr>
<td>DEFB135</td>
<td>DQ012025</td>
<td>DEFB137P</td>
<td>DQ012088</td>
<td>CBD135</td>
</tr>
<tr>
<td>DEFB136</td>
<td>DQ012026</td>
<td>SPAG11c</td>
<td>AF263553</td>
<td>CBD136</td>
</tr>
</tbody>
</table>
The complete repertoires of β-defensins identified in the human, chimpanzee, dog, mouse, and rat are shown together with GenBank accession numbers for their mRNA sequences. Previously identified genes are italicized and listed with a reference accession number (NM) if available. The remaining genes were discovered in this study. *Spag11c-1* is an...
alternatively spliced variant of Spag11c in the rat. Abbreviations: DEFB/Defb, β-defensin; SPAG11, sperm-associated antigen 11; P/ps, pseudogene.
**Supplementary Table 2. Primer sequences used for RT-PCR analysis of rat β-defensins**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer Sequence</th>
<th>Product Size (bp)</th>
<th>No. of Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forward</td>
<td>Reverse</td>
<td>cDNA</td>
</tr>
<tr>
<td>Defb1</td>
<td>TGGACCCTGACTTCACCAGACAT</td>
<td>CTTCAAACCACCTGCAACTCTCTCTG</td>
<td>244</td>
</tr>
<tr>
<td>Defb2</td>
<td>GACTCTCTGCTCTCTGCTGCT</td>
<td>CCCACAAACAAGGTAGAAGCA</td>
<td>172</td>
</tr>
<tr>
<td>Defb3</td>
<td>CTCTTCTCATTTTCCTGGTGCT</td>
<td>GTGCCACAACACTGCCAATCTCTCT</td>
<td>150</td>
</tr>
<tr>
<td>Defb4</td>
<td>CGCATCCATTACTTCTTCTTTCTC</td>
<td>ATCTCACTCTAGCCAGTCCACAA</td>
<td>173</td>
</tr>
<tr>
<td>Defb5</td>
<td>CTCTTCTCATTTTCCTGGTGCT</td>
<td>TCTCAGTCTACCCAGTCCACAA</td>
<td>164</td>
</tr>
<tr>
<td>Defb9</td>
<td>GACTCTCTGCTCTCTGCTGCT</td>
<td>CACATGGGAGGTAGGAAACGAC</td>
<td>163</td>
</tr>
<tr>
<td>Defb10</td>
<td>GACTCTCTGCTCTCTGCTGCT</td>
<td>GTGCAGAGGAAACCACATCGGA</td>
<td>177</td>
</tr>
<tr>
<td>Defb11</td>
<td>GACTCTCTGCTCTCTGCTGCT</td>
<td>GTAACAGCTACCCACGTCCTTCTG</td>
<td>169</td>
</tr>
<tr>
<td>Defb12/35</td>
<td>AGCAGAGGGAGCTTTATATTG</td>
<td>CTCCCGCAGCAGAAAGATTC</td>
<td>219</td>
</tr>
<tr>
<td>Defb13</td>
<td>CGGGCTTGTCTTCTCATCACACAC</td>
<td>GTGTTAAGTGGCTCGGCAGG</td>
<td>143</td>
</tr>
<tr>
<td>Defb14</td>
<td>TCGTATGGTCTTCTGGACCTGCT</td>
<td>CTGCAAGCATTTTTCGACCTCTGT</td>
<td>160</td>
</tr>
<tr>
<td>Defb15/34</td>
<td>GACACGGCTGCTGCACCCACTCA</td>
<td>CATCCCGAGCTCTGGTAGG</td>
<td>272</td>
</tr>
<tr>
<td>Defb16p</td>
<td>TGTCATTGTTTTCGACTAC</td>
<td>GCATACATCCACATCCACAGGAA</td>
<td>216</td>
</tr>
<tr>
<td>Defb17</td>
<td>TGTCATTGTTTTCGACTAC</td>
<td>GCATTCTCTCAGTCCACAGAC</td>
<td>186</td>
</tr>
<tr>
<td>Defb18</td>
<td>TTCTGCTGCTGCTGCTCTTC</td>
<td>TATTCTCGCATACACACAG</td>
<td>173</td>
</tr>
<tr>
<td>Defb20</td>
<td>TCCTTATTTGTGTTTGTGTGT</td>
<td>TTTCTTGTGCCTTTTGTCTCC</td>
<td>181</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------</td>
<td>------------------------</td>
<td>-----</td>
</tr>
<tr>
<td>Defb21</td>
<td>GCTCTTTCTCTGCTTGGCTCTT</td>
<td>GTTTGTAGTGTGGTGTGTCCAG</td>
<td>180</td>
</tr>
<tr>
<td>Defb22</td>
<td>ATGAAGTCCCTATTGTCTGCTC</td>
<td>TTTCCCCCTTCTCCTGTTT</td>
<td>303</td>
</tr>
<tr>
<td>Defb23</td>
<td>GCTTCAGTATCAGGTTGGAATCA</td>
<td>GTATGTGGGTTATTTGCTGTC</td>
<td>197</td>
</tr>
<tr>
<td>Defb24</td>
<td>CCTGCTTTCTGCCATATTTTG</td>
<td>TCTGGAGTCTGTCATGCGGT</td>
<td>268</td>
</tr>
<tr>
<td>Defb25</td>
<td>GGCAAAAGTGGATTCTGCATT</td>
<td>ATAGGACAGGCACACAGAGA</td>
<td>180</td>
</tr>
<tr>
<td>Defb26</td>
<td>CTTTCAATTATCTGTTGTTGCT</td>
<td>CCGAGGTAGTTTTTGGTGTG</td>
<td>214</td>
</tr>
<tr>
<td>Defb27</td>
<td>AACACCCCTGGATTCCCGAAGC</td>
<td>CTTGGTTTCTGTTGCTTGGGA</td>
<td>258</td>
</tr>
<tr>
<td>Defb28</td>
<td>ACATTAAGCTCTGTTCTTGCAC</td>
<td>ACTGTATGTGTCCTCGACTTG</td>
<td>285</td>
</tr>
<tr>
<td>Defb29</td>
<td>TCTGCTGTGTGATTTCTATCC</td>
<td>CTGGCGACATTCAAACCTGAG</td>
<td>359</td>
</tr>
<tr>
<td>Defb30</td>
<td>GTTGATCCTTTGTCTTTTGTC</td>
<td>ACGGGCATTTTCCTTTCTTTC</td>
<td>221</td>
</tr>
<tr>
<td>Defb33</td>
<td>GGCTGCTATTCTCTCTCTCTTC</td>
<td>GGCGATTTTTGTATCGACTTTTGC</td>
<td>172</td>
</tr>
<tr>
<td>Defb36</td>
<td>AGGCCCATCATGAAGCTCTGC</td>
<td>TTCCCCGTTTCCGGGCAGTTAG</td>
<td>227</td>
</tr>
<tr>
<td>Defb37</td>
<td>TGGCTTTCGCTGTTGTCCTCTC</td>
<td>CAGGTGCGGTCACTACAATCAGA</td>
<td>140</td>
</tr>
<tr>
<td>Defb38</td>
<td>TGCTTTTCGCTGCTTTTCTCTC</td>
<td>CGCAGGGCTAACAGAAATAAGAC</td>
<td>145</td>
</tr>
<tr>
<td>Defb39</td>
<td>TGCTTTTCGCTGCTTTTCTCTC</td>
<td>GGCTTGGGAACTCGGATAGACA</td>
<td>205</td>
</tr>
<tr>
<td>Defb40</td>
<td>TGCTTTTCGCTGCTTTTCTCTC</td>
<td>GACTCCAGTTAAACCMAACGACG</td>
<td>201</td>
</tr>
<tr>
<td>Defb41</td>
<td>GATCCCACTCCTTCTGTGCTG</td>
<td>CCCACCATCTCCAACATTTTC</td>
<td>197</td>
</tr>
</tbody>
</table>
RT-PCR reactions were performed with DNA Engine (Model PTC-200, MJ Research). Number of PCR cycles used for evaluating tissue and developmental expression patterns was the same for all relevant genes, except for Defb29, Defb36, and Defb49. For Defb29 amplification, 35, 40, and 30 cycles were used for gene expression in the tissue panel, epididymis, and testis, respectively. For Defb36 amplification, 35 cycles were used for the tissue panel with 32 cycles used for its developmental expression in testis. A total of 35 cycles were performed for tissue expression of Defb49, whereas 30 cycles were used for developmental expression in epididymis.
b Because of 100% identity in the coding sequences between *Defb12* and *Defb35* and between *Defb15* and *Defb34* in the rat, the same primers were used for each pair.
CHAPTER V
RATTUSIN, A NOVEL $\alpha$-DEFENSIN-RELATED PEPTIDE WITH SALT-INDER SENSITIVE ANTIBACTERIAL ACTIVITY
Cationic antimicrobial peptides (CAMPs) are an essential component of the innate immune system. As a major family of CAMPs, defensins are expressed mainly by mucosal epithelial cells and neutrophils, which constantly come into direct contact with microbes. Defensins are capable of killing a broad spectrum of bacteria through physical disruption of membranes. However, all known defensins show a significantly reduced antibacterial activity in the presence of salt and/or divalent cations, thereby limiting their therapeutic potential for the treatment of systemic infections. A genome-wide computational screen of the entire rat genome led to the identification of a novel α-defensin-related gene, designated rattusin. Rattusin shares a highly conserved pro-sequence with mammalian α-defensins. However, unlike defensins with a canonical six-cysteine motif, rattusin consists of only five cysteines with a different spacing pattern. Furthermore, rattusin is preferentially expressed in distal small intestine with potent activities against a broad range of Gram-negative and Gram-positive bacteria, including antibiotic-resistant strains. In contrast to classical α-defensins, rattusin retains its activity in the presence of physiological concentrations of salt and divalent cations. To our knowledge, rattusin is the only defensin-related peptide with salt-insensitive antibacterial activities, making it an attractive therapeutic candidate for diseases like cystic fibrosis and Crohn’s disease.
INTRODUCTION

The emergence of antibiotic-resistant pathogens has become a major health crisis worldwide (10, 15). Novel antimicrobial drugs against resistant microbes are urgently needed. Defensins and cathelicidins, an essential component of the vertebrate innate immune system, are capable of killing a broad spectrum of bacteria with an equal efficiency against both antibiotic-resistant and -susceptible strains and little chance of developing resistance (27). Therefore, these CAMPs are being actively explored as antibacterial drugs. However, the potential of defensins as therapeutics is dampened by a loss of activity in the presence of physiological concentrations of salt (25, 43). As a consequence, none of the defensin-based therapeutics except for a fungal defensin, plectasin, are under clinical development (57).

In vertebrates, there are two major families of antimicrobial peptides, namely defensins and cathelicidins (12, 25, 55, 56). Unlike cathelicidins that are mostly free of cysteines, defensins consist of six conserved cysteine residues in the mature region. Depending upon the disulfide bond connectivity, defensins are further divided into three classes known as α-, β-, and θ-defensins. α-defensins are mainly produced by intestinal Paneth cells and leukocytes, whereas β-defensins are mainly produced by skin keratinocytes as well as mucosal epithelial cells of the gastrointestinal, respiratory and urogenital tracts (4, 12, 25, 32, 34, 42, 43).
In humans, there are only two Paneth cell α-defensins, namely human α-defensin (HD)-5 and 6, whereas multiple Paneth cell α-defensins (known as cryptdins in mice) have been found in mice and rhesus monkey (4, 32-34). Matrilysin or matrix metalloproteinase-7 (MMP-7) is responsible for post-translational processing of inactive cryptdin precursors into antibacterially active peptide (48). Cryptdins are stored as the mature active form in Paneth cell granules and are released in response to bacterial products like lipopolysaccharide (LPS), lipoteichoic acid, lipid A and muramyl dipeptide (3). In contrast, HD-5 is stored in Paneth cell granules as the precursor form and is processed by trypsin during or after secretion into intestinal lumen (13). Recently, a family of cysteine-rich cryptdin-related sequences (CRS) have been found in mouse intestine (19). Evidence for the essential role of α-defensins in protecting enteric infections has been demonstrated convincingly. Transgenic mice expressing HD-5 were resistant to Salmonella typhimurium infection (39). Conversely, mice lacking the Matrilysin gene were unable to produce mature Paneth cell cryptdins and became highly susceptible to S. typhimurium infections (48).

Defensins possess pleiotropic functions. In addition to broad spectrum antibacterial, antiviral and antifungal activities, certain defensins are chemotactic for dendritic cells, mast cells, monocytes, and T cells, and some can also induce maturation of dendritic cells and sperms. Defensins help in wound healing by inducing vascularization, promoting proliferation of epithelial and fibroblast cells, and augmenting wound closure (17, 25, 31, 43). In addition, several human neutrophil α-defensins were shown recently to be capable of neutralizing bacterial toxins (20, 21, 30).
We identified a α-defensin-related gene (Defa-rs1) recently through a computational screen of the rat genome (36), which we re-named rattusin in this report. Unlike classical mammalian defensins with a canonical six-cysteine motif, rattusin consists of five cysteines with a unique cysteine spacing pattern. To test whether rattusin is a bona fide antimicrobial peptide, we synthesized putatively mature rattusin based on its C-terminal sequence and compared its activity with mouse cryptdin-4 and human HD-5, which are the two most potent Paneth cell α-defensins that have been reported.

**MATERIALS AND METHODS**

*RT-PCR analysis of gene expression pattern of rattusin*

Different segments of the gastrointestinal tract were collected from 2-month-old Sprague-Dawley rats. Total RNA was extracted using TRIzol (Invitrogen, Carlsbad, CA). For each RNA sample, 4 µg were reverse transcribed using random hexamers and SuperScript II reverse transcriptase (Invitrogen) according to the manufacturer’s instructions. The subsequent PCR was carried out as described (36). Briefly, 1/40 of the first-strand cDNA was used to amplify rattusin and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) with gene-specific primers (36). The PCR product was analyzed by electrophoresis on 1.2% agarose gel containing 0.5 µg/ml ethidium bromide. The identity of the rattusin PCR product was confirmed by cloning into T/A cloning vector and direct DNA sequencing.
Peptide synthesis and refolding

Since MMP-7 preferentially cleaves the sequence prior to leucine (2, 19, 44, 45), we rationalized that the mature rattusin is likely to start from the first leucine after pro-sequence. Therefore, putatively mature rattusin of 31 amino acids (LRVRRTLQCSRR VCRNTCIRSLRSTYAS) was chemically synthesized in the reduced form using the standard solid-phase synthesis, purified by reverse phase-high pressure liquid chromatography (RP-HPLC) to >95% purity (Bio-synthesis, Lewisville, TX). The mass of peptide was confirmed by matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry (MS) using Voyager DE-PRO (Applied Biosystems, Foster City, CA) housed in the Recombinant DNA/Protein Core Facility at Oklahoma State University.

To oxidize the reduced rattusin, 0.1 mg/ml of the peptide was exposed to O₂ gas for 5 min in 50 mM Tris, pH 8.0, and stirred gently for 48 h at room temperature with the cap open as described (19). Following air oxidation, refolded rattusin was purified by RP-HPLC on a 4.6 × 250 mm Vydac C18 column (Grace Vydac, Hesperia, CA) and BioLogic DuoFlow Liquid Chromatography System (Bio-Rad, Hercules, CA). Buffer A consisted of 5% acetonitrile and 0.18% trifluoroacetic acid (TFA). Buffer B consisted of 90% acetonitrile & 0.15% TFA. The gradient used was 0 to 60% buffer B over 90 minutes at a flow rate of 1 ml/min. Eluted peptide was lyophilized and stored at -80°C until use. The peptide was reconstituted in 0.01% acetic acid and quantified by measuring UV absorbance at 280 nm based on the extinction coefficients of tyrosine and cysteines present in rattusin (41). Recombinant cryptdin-4 was produced in bacteria as described
Plasmid encoding cryptdin-4 was provided by Dr. Andre Ouellette at University of California, Irvine. Synthetic HD-5 was kindly provided by Dr. Wuyuan Lu at University of Maryland, Baltimore (49). Cryptdin-4 and HD-5 were used as the reference peptides.

*Recombinant expression and purification of pro-rattusin*

Recombinant pro-rattusin was produced as a N-terminal 6X histidine-tagged fusion protein using the strategy described previously for cryptdin-4 (41, 44). Restriction sites for *Hind*III and *Bam*HI were included in forward and reverse primers, respectively, as underlined in primer sequences. Additional methionine-encoding *ATG* codon was introduced in the forward primer prior to the start of pro-rattusin nucleotide sequence to enable cyanogens bromide cleavage in order to remove tag sequence. Forward primer was 5’-CGGGATCCATGGATCCATTCCAAGAGGCAGAAG3’, and the reverse primer sequence was 5’-GCCCAGCTTATCTCATCCTTGAGGTCAT-3’. Pro-rattusin cDNA was amplified from rat ileum RNA by using random hexamers, SuperScript II reverse transcriptase, and high-fidelity Platinum *Pfx* DNA polymerase (Invitrogen). The resulting PCR product was gel purified, digested with *Hind*III and *Bam*HI, ligated into pET28a vector (Novagen, Madison, WI), and transformed into *E. coli* DH-5α. The presence of the correct insert was confirmed by DNA sequencing.

Recombinant plasmid was then re-transformed into *E. coli* BL21 (DE3)-Codon Plus-RIL cells (Stratagene, La Jolla, CA). For large-scale production of recombinant pro-rattusin, overnight bacteria were subcultured 1:100 in 2 L of LB broth till OD600 reached
0.4 to 0.6. Recombinant fusion protein was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.4 mM for 6 h. Bacteria were pelleted, resuspended in the buffer containing 20 mM sodium phosphate, 0.5 M NaCl, 20 mM imidazole and 6 M guanidine chloride, pH 7.4, and sonicated on ice. The lysate was further stirred for 1 h at room temperature to ensure complete solubilization of inclusion bodies.

Following centrifugation of the lysate at 30,000 × g for 30 min at 4°C, recombinant His-tagged pro-rattusin was purified from the supernatant using Ni Sepharose 6 fast flow resins (GE Healthcare Biosciences, Piscataway, NJ) according to the manufacturer’s instructions. Fusion protein was then eluted using a buffer contained 20 mM sodium phosphate, 0.5 M NaCl, 0.5 M imidazole and 6 M guanidine chloride, pH 5.0. Following dialysis in 0.01% acetic acid using a Slide-A-Lyzer 3.5 MWCO dialysis cassette (Pierce Biotechnology, Rockford, IL), fusion protein was lyophilized and dissolved in 2.5% TFA at 2.5 mg/ml and incubated with 25 mg/ml CNBr in dark for 24 h to remove the fusion tag (50). The cleaved product was lyophilized, dissolved in Buffer A and subjected to RP-HPLC. Pro-rattusin fraction was then lyophilized, and its molecular weight was confirmed by MALDI-TOF mass spectrometry. The peptide concentration was quantified by measuring UV absorbance at 280 nm.

Bacterial culture and antibacterial assays

All bacterial strains including *Staphylococcus aureus* ATCC 25923, *Listeria monocytogenes* ATCC 19115, *S. aureus* ATCC BAA-39, *S. aureus* ATCC 43300,
Escherichia coli O157:H7 ATCC 700728, Salmonella enterica serovar Typhimurium ATCC 14028, S. enterica serovar Typhimurium DT104 ATCC 700408, and Klebsiella pneumoniae ATCC 13883 were purchased from either ATCC (Manassas, VA) or MicroBiologics (St. Cloud, MN). Bacteria were grown in trypticase soy broth (TSB) overnight and subcultured for 3-4 h at 37°C in a shaking incubator to the mid-log phase. To study the antibacterial spectrum, a broth microdilution assay was used as described (53). Briefly, mid-log phase bacteria were washed with 25 mM sodium phosphate buffer, pH 7.4, and suspended to $5 \times 10^5$ CFU/ml in 25 mM sodium phosphate, pH 7.4, containing 5% TSB with and without 100 mM NaCl. Bacteria (90 µl) were then dispensed in a 96-well tissue culture plate, followed by addition in duplicate of serially diluted rattusin or cryptdin-4 (10 µl). After overnight incubation at 37°C, the minimum inhibitory concentration (MIC) of each peptide was determined as the lowest concentration that gave no visible bacterial growth.

To study the kinetics of bacterial killing, a standard colony counting assay was used as described (53). Rattusin, cryptin-4 and HD-5 were incubated with 90 µl of $5 \times 10^5$ CFU/ml Staphylococcus aureus or E. coli O157:H7 in 25 mM sodium phosphate buffer, pH 7.4, containing 1% TSB with and without 100 mM NaCl. Following incubation at 37°C for 10, 30, 60, 120 and 240 min, bacteria were diluted rapidly with ice-cold PBS and serially plated onto trypticase soy agar (TSA) plates. Viable bacteria were counted after overnight incubation at 37°C. The effect of Mg$^{2+}$ on the antibacterial activity was studied by incubating rattusin and cryptdin-4 for 4 h with 90 µl of S. aureus ($5 \times 10^5$ CFU/ml) at the concentration of 2 µM each or with E. coli O157:H7 at the
concentration of 4 µM each in 1% TSB, 25 mM sodium phosphate buffer, pH 7.4, containing 0, 1, 2 and 5 mM of MgCl₂.

*Cytotoxicity Assay*

The toxicity of rattusin to human Caucasian colon adenocarcinoma (Caco-2) cells (ATCC, Manassas, VA) was measured using alamarBlue dye (Biosource, Camarillo, CA) as described previously (53, 54). Briefly, Caco-2 cells were seeded into a 96 well plate at 5 × 10⁴ cells/ml in Dulbecco’s modified Eagle’s medium (DMEM) with 10% FBS and grown overnight in a humidified 5% CO₂ incubator. Cells were washed once with DMEM, followed by addition of fresh DMEM containing 100 µM of rattusin, cryptdin-4 or chicken fowlicidin-1 and in the presence and absence of 10% FBS. After 18 h incubation, 10 µl of alamarBlue dye was added and cells were further incubated for another 6 h at 37°C. The plate was read with excitation at 545 nm and emission at 590 nm. The percentage of cell death was calculated as [1-(F_peptide-F_background)/(F_acetic acid-F_background)] × 100, where F_peptide is the fluorescence of cells exposed to 100 µM peptide, F_acetic acid is the fluorescence of cells exposed to 0.01% acetic acid only, and F_background is the background fluorescence of 10% alamarBlue dye in cell culture medium without cells.
RESULTS

*Rattusin, an α-defensin-related peptide with preferential expression in the distal small intestine*

Following a comprehensive screening of the entire rat genome, we identified an α-defensin-related gene (36). Similar to α-defensins, this gene is composed of two exons with the first exon encoding the 5’ untranslated region (UTR) and preprosegment and the second encoding a putatively mature peptide and 3’ UTR (Fig. 1). The encoded peptide, named rattusin, shares a high degree of similarity in the signal and pro-sequences with α-defensins (Fig. 2). However, rattusin consists of five cysteines in the C-terminal mature region, instead of canonical six cysteines in α-defensins (36). Furthermore, the cysteine spacing patterns are quite different between rattusin and classical α-defensins (Fig. 2).

We revealed earlier that the rattusin mRNA was transcribed abundantly in the small intestine of rat (36). To further analyze the detailed expression pattern of rattusin, RT-PCR was performed using the RNA isolated from different segments of the gastrointestinal tract of 2-month-old healthy rats. As shown in Fig. 3, rattusin was highly expressed in distal jejunum and entire ileum, but not in other parts of the gastrointestinal tract. Cryptdin-4 and HD-5, Paneth cell-specific α-defensins, also show a similar expression pattern in small intestine (9, 35, 47). This suggested that, similar to enteric α-defensins, rattusin is likely to be produced by Paneth cells.
**Antibacterial properties of rattusin**

To test whether putatively mature rattusin is functional or not, we synthesized putatively mature rattusin in the reduced form and further oxidized it. Successful refolding was confirmed by RP-HPLC, showing a decrease in the retention time due to the formation of disulfide bonds and a change in the conformation (Fig. 4). Refolded rattusin was used to study antibacterial activity against several representative Gram-negative and Gram-positive bacteria by a modified broth microdilution assay with and without 100 mM NaCl (53). As shown in Table 1, rattusin exhibited antibacterial activity with the MIC value mostly in the range of 2-4 μM. In most cases, rattusin was as potent as cryptdin-4, the most efficacious mouse Paneth cell α-defensins that has been reported (32). Importantly, rattusin also displayed a similar antibacterial efficiency against two strains of methicillin-resistant *Staphylococcus aureus* (MRSA) and multi-drug resistant *Salmonella typhimurium* DT 104 (Table 1). Importantly, the activity of rattusin remained largely unchanged in the presence of NaCl, in sharp contrast to cryptdin-4, whose activity was abolished by salt.

To study the kinetics of bacterial killing, each representative strain of Gram-negative and Gram-positive bacteria was used in a standard colony counting assay. Rattusin, cryptdin-4 and HD-5 at 4 μM killed *E. coli* by three logs within 2 h (Fig. 5A). In the presence of 100 mM of NaCl, killing of bacteria by rattusin was slightly delayed with a complete killing occurring at 4 h. In contrast, Cryptdin-4 and HD-5 were completely inactivated in 100 mM NaCl (Fig. 5B). A similar trend also occurred with *S. aureus*. All three peptides at 2 μM killed *S. aureus* by two logs in 4 h (Fig. 6A). Killing
by rattusin was largely unaffected by 100 mM NaCl. However, cryptdin-4 and HD-5 were significantly inhibited in the presence of salt (Fig. 6B). Collectively, these results clearly suggested that rattusin with a broad-spectrum, salt-insensitive bactericidal activity is among the most potent defensins. To further study the effect of salinity, the antibacterial activity of rattusin was examined in the presence of increasing concentrations of NaCl. As shown in Fig. 7, rattusin maintained antibacterial activity against both \textit{E. coli} and \textit{S. aureus} in up to 200 mM NaCl.

Divalent cations such as Mg\textsuperscript{++} are usually present at 1 to 2 mM in most biological fluids (6, 7). At such low concentrations, divalent cations are known to inhibit antibacterial activities of cationic peptides (8, 11). To study the effect of Mg\textsuperscript{++} on the antibacterial activity, rattusin and cryptdin-4 were incubated with bacteria in the presence of increasing concentrations of MgCl\textsubscript{2}. There was a dose-dependent loss of the activity of cryptdin-4 against \textit{E. coli} with complete inactivation occurring at 2-5 mM of MgCl\textsubscript{2} (Fig. 8A). In contrast, the activity of rattusin was largely unaffected in the presence of MgCl\textsubscript{2} (Fig. 8A). Mg\textsuperscript{++} did not inhibit the activity of either rattusin or cryptdin-4 against \textit{S. aureus} (Fig 8B). In the case of cryptdin-4, increased killing observed at 1 mM was not observed at higher concentrations of MgCl\textsubscript{2} (Fig 8B).

\textit{Low cytotoxicity of rattusin}

The cytotoxic effect of rattusin to intestinal epithelial cells was studied using human Caco-2 cells. Fowlcidin-1 (100 µM) was used as a positive reference since it showed significant cytotoxicity towards mammalian cells (53). Following treatment of
cells for 24 h, rattusin exhibited no cytotoxicity even at 100 µM, similar to cryptdin-4 (Fig. 9).

Inhibition of the antibacterial activity of rattusin by anionic pro-segment

Recombinant pro-rattusin was produced as an N-terminal 6× histidine-tagged fusion protein followed by removal of the fusion tag by CNBr cleavage. The antibacterial activity of pro-rattusin was studied by a modified broth microdilution assay against two representative Gram-positive and Gram-negative bacteria. Pro-rattusin showed a MIC value of 16 µM against *S. aureus*, but no inhibition of growth was observed for *L. monocytogenes, E. coli* and *S. typhimurium* even at 32 µM, the highest concentration used (Table 2). This is reminiscent of cryptdin-4, whose anionic pro-segment also inhibited the antibacterial activity of C-terminal cationic peptide (44).

DISCUSSION

Rattusin shares a highly conserved signal and pro-sequence with classical mammalian α-defensins, but with a unique C-terminal region (Fig. 2). Our results demonstrate that this C-terminal peptide possesses potent, broad-spectrum antibacterial activities. Recently, a number of cysteine-rich cryptdin-related sequences (CRS) in mice were described, which contain a highly similar prepro-region to mice cryptdins (19). However, the cysteine-spacing patterns of these CRS peptides are different from rattusin.

Similar to cryptdin-4 and HD-5 (9, 35), rattusin is preferentially expressed in distal small intestine. It is tempting to speculate that preferential expression of certain
enteric defensins in the distal small intestine might prevent colonization of microflora from adjacent colon. Consistently, the number of microbes in small intestine is $10^4$-$10^6$ times less than the number of microbes in colon (5). Thus, the number and species of microorganisms in small intestine might be dictated by Paneth cell defensins.

Desirably, rattusin maintains its antibacterial activity in the presence of physiological concentrations of NaCl, which is in sharp contrast to cryptdin-4, HD-5, and other defensins, which show reduced antibacterial activity in the presence of NaCl (12, 25, 38, 43). Maintenance of the antibacterial activity of rattusin in the presence of salt might be due to its distinct spacial structure conferred by the unique cysteine spacing pattern. Cryptdin-4 disrupts bilayer head groups through interfacial binding onto lipid bilayer rather than inserting into the hydrophobic core (40, 41). It interacts favorably with bacterial phospholipid vesicles containing negatively charged phospholipids than neutral phospholipids indicating ionic interaction between positive charged peptide and negative charged lipids (40, 41). Such interaction is more susceptible to inhibition by physiological concentrations of NaCl, which in rodent ileum is maintained at around 162 mM (29). Since rattusin is relatively resistant to NaCl, its mechanism of action is likely to be different from cryptdin-4.

Divalent cations such as Mg$^{++}$ stabilize LPS of Gram-negative bacteria by binding to negatively charged phosphate and pyrophosphate groups of LPS, thereby preventing charge-charge repulsion (16, 28). Antimicrobial peptides must displace these divalent cations in order to interact with the outer membrane (16, 28). A number of defensins and
cathelicidins are inhibited at physiological concentrations of divalent cations (1, 8, 11, 24, 26). Rattusin kills bacteria in the presence of 2-5 mM Mg\textsuperscript{++}, whereas cryptdin-4 is inactivated. This suggested that rattusin is able to displace divalent cations.

Divalent cations are not inhibitors of the antibacterial activity of CAMPs against Gram-positive bacteria since they lack LPS. On the contrary, Mg\textsuperscript{++} potentiated the activity of rattusin against \textit{S. aureus}. However, this finding is not surprising since Mg\textsuperscript{++} causes transcriptional suppression of \textit{dlt} operon in \textit{S. aureus} (22). Genes of \textit{dlt} operon are involved in incorporation of D-alanine into lipoteichoic acid and wall teichoic acid, which introduces positively charged amino group and partially neutralizes negative charges on teichoic acid, thereby providing resistance to CAMPs (23). Therefore, suppression of \textit{dlt} operon by Mg\textsuperscript{++} increases susceptibility of \textit{S. aureus} to rattusin and cryptdin-4.

Similar to pro-cryptdin-4, pro-rattusin generally lacks antibacterial activity (Table 2). Pro-cryptdin-4 does not interact with phospholipid vesicles and lacks a typical β-sheet structure (41). It is likely that such inhibition involves neutralization of positive charges on the C-terminal peptide by the negative charge on the N-terminal pro-segment. This mechanism is further supported by evidence that a replacement of aspartate and glutamate by glycine confers bactericidal and membrane-interacting ability of CAMPs (46).
In conclusion, rattusin is unique α-defensin-related peptide, which is preferentially expressed in distal small intestine. It possesses broad-spectrum antibacterial activity insensitive to salt and divalent cations. The salt-insensitive activity of rattusin may be further exploited for the treatment of cystic fibrosis and Crohn’s disease. Increased salt concentrations are believed to be responsible for inactivation of defensins in the airway of the cystic fibrosis patients (14). In Crohn’s disease, a deficiency of intestinal Paneth cell defensins is associated with an increase in susceptibility to bacterial infections (37). Therefore, exogenous delivery of potent, salt-insensitive rattusin represents a promising therapeutic strategy to treat both topical and systemic antibiotic-resistant infections.

ACKNOWLEDGEMENTS

This work was supported by the Oklahoma Center for the Advancement of Science and technology Grant HR03-146 and Oklahoma Agricultural Experiment Station project H-2507. We would like to thank Yugendar Bomminenni for help with cell culture.

GRANTS

This work was supported by the Oklahoma Center for the Advancement of Science and Technology Grant HR03-146 and Oklahoma Agricultural Experiment Station project H-2507.
REFERENCES


**FIGURE LEGENDS**

Fig.1. Nucleotide sequence of the *rattusin* gene. Numbering is based upon the transcription start site, which is numbered +1 with arrow showing the direction of transcription. Transcribed mRNA sequence is shown in uppercase letters. The open reading frame coding for the rattusin precursor is indicated below each triplet codon. TATA and CAAT boxes in the promoter region are underlined. Stop codon is denoted by an asterisk. The polyadenylation signal is underlined and italicized. The gene structure is
derived from rat EST sequences available in GenBank (Accession Nos. AI639089.1 and CR476138.1) and our previous analysis of *rattusin* gene (36).

Fig. 2. Alignment of rattusin with representative Paneth cell α-defensins, namely mouse cryptdin-4 and human defensin-5 (HD-5). Dashes are created for optimal alignment. Conserved amino acid residues are shaded. The net positive charge on each mature peptide, as underlined, is indicated in parenthesis and length of mature peptide is also indicated. Note that the signal peptide and pro-sequence region of all three peptides are highly homologous, whereas C-terminal mature peptide region are diversified. A conserved six-cysteine motif of cryptdin-4 and HD-5 is absent in rattusin.

Fig. 3. Expression pattern of the rattusin gene across gastrointestinal axis. RT-PCR was used to evaluate the expression of the rattusin transcript. The housekeeping gene, GAPDH, was used to normalize template input.

Fig. 4. RP-HPLC profile of reduced and oxidized rattusin. Reduced synthetic peptide was refolded by air oxidation in 50 mM Tris buffer, pH 8.0, for 48 h. Oxidized rattusin was purified to homogeneity by RP-HPLC. Note that there is a decrease in the retention time of oxidized rattusin due to refolding.

Fig. 5. Kinetics of killing of *E. coli* O157:H7 in the absence (A) and presence (B) of 100 mM NaCl by rattusin, cryptdin-4 and HD-5. *E. coli* O157:H7 ATCC 700728 was incubated with 4 μM rattusin, cryptdin-4, HD-5, or an equal volume of 0.01% acetic acid (no peptide) in duplicate in 25 mM sodium phosphate buffer, pH 7.4, 1% TSB with and without 100 mM NaCl for 10, 30, 60, 120 and 240 min. Surviving bacteria were plated and counted. Data shown are means ± SEM of two independent experiments.

Fig. 6. Kinetics of killing of *S. aureus* in the absence (A) and presence (B) of 100 mM NaCl by rattusin, cryptdin-4 and HD-5. *S. aureus* ATCC 25923 was incubated with 2 μM rattusin, cryptdin-4, HD-5, or an equal volume of 0.01% acetic acid (no peptide) in duplicate in 25 mM sodium phosphate buffer, pH 7.4, 1% TSB with and without 100 mM NaCl.
NaCl for 10, 30, 60, 120 and 240 min. Surviving bacteria were plated and counted. Data shown are means ± SEM of two independent experiments.

Fig. 7. Effect of salinity on the antibacterial activity of rattusin against *S. aureus* (A) and *E. coli* O157:H7 (B). *S. aureus* and *E. coli* were incubated with 2 and 4 μM of rattusin, or an equal volume of 0.01% acetic acid (no peptide) in 25 mM sodium phosphate buffer, pH 7.4, 1% TSB with increasing concentrations of NaCl, for 4 h. Surviving bacteria were plated and counted. Data shown are means ± SEM of two independent experiments.

Fig. 8. Effect of Mg^{++} on the antibacterial activity of rattusin and cryptdin-4 against *S. aureus* (A) and *E. coli* O157:H7 (B). *S. aureus* or *E. coli* were incubated with rattusin, cryptdin-4, or an equal volume of 0.01% acetic acid (no peptide) in 25 mM sodium phosphate buffer, containing 1% TSB with 0, 1, 2 and 5 mM MgCl₂, pH 7.4, for 4 h. Surviving bacteria were plated and counted. Both peptides were used at identical concentrations for each bacterial strain (2 and 4 μM against *S. aureus* and *E. coli*, respectively). Data shown are means ± SEM of 2-3 independent experiments.

Fig. 9. Absence of cytotoxicity of rattusin and cryptdin-4 to Caco-2 cells. Cells were incubated with 100 μM of rattusin, cryptdin-4 or fowlicidin-1 for 24 h in DMEM with and without 10 % fetal bovine serum for 24 hours. Cell viability was measured with an alamarBlue dye-based method. Data are representative of two independent experiments done in duplicate.
Table 1. Antibacterial spectrum of rattusin and cryptdin 4

<table>
<thead>
<tr>
<th>Bacteria</th>
<th>ATCC Number</th>
<th>Rattusin 0 mM NaCl</th>
<th>Rattusin 100 mM NaCl</th>
<th>Cryptdin-4 0 mM NaCl</th>
<th>Cryptdin-4 100 mM NaCl</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gram-negative</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>E. coli</em> O157:H7</td>
<td>700728</td>
<td>2-4</td>
<td>2-4</td>
<td>2</td>
<td>&gt;16</td>
</tr>
<tr>
<td><em>S. typhimurium</em></td>
<td>14028</td>
<td>4</td>
<td>8</td>
<td>8</td>
<td>&gt;16</td>
</tr>
<tr>
<td><em>S. typhimurium</em> DT104</td>
<td>700408</td>
<td>4</td>
<td>4-8</td>
<td>4-8</td>
<td>&gt;16</td>
</tr>
<tr>
<td><em>K. pneumoniae</em></td>
<td>13883</td>
<td>2-4</td>
<td>4-8</td>
<td>8</td>
<td>&gt;16</td>
</tr>
<tr>
<td><strong>Gram-positive</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>S. aureus</em></td>
<td>25923</td>
<td>4</td>
<td>4</td>
<td>1-2</td>
<td>≥16</td>
</tr>
<tr>
<td><em>L. monocytogenes</em></td>
<td>19115</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>&gt;16</td>
</tr>
<tr>
<td><em>S. aureus</em> (MRSA)</td>
<td>43300</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td><em>S. aureus</em> (MRSA)</td>
<td>BAA-39</td>
<td>4-8</td>
<td>4-8</td>
<td>1-2</td>
<td>&gt;16</td>
</tr>
</tbody>
</table>
Table 2. Inhibition of the antibacterial activity of rattusin by the pro-segment

<table>
<thead>
<tr>
<th>Bacteria</th>
<th>ATCC Number</th>
<th>MIC (µM)</th>
<th>Pro-rattusin</th>
<th>Rattusin</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gram-negative</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>E. coli</em> O157:H7</td>
<td>700728</td>
<td>&gt;32</td>
<td>2-4</td>
<td></td>
</tr>
<tr>
<td><em>S. typhimurium</em></td>
<td>14028</td>
<td>&gt;32</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td><strong>Gram-positive</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>S. aureus</em></td>
<td>25923</td>
<td>16</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td><em>L. monocytogenes</em></td>
<td>19115</td>
<td>&gt;32</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 1
<table>
<thead>
<tr>
<th></th>
<th>Signal Peptide</th>
<th>Prosequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rattusin</td>
<td>MKTLILLSALVLLALGVQADPIQEAEEETKTEEQPADEDQDVSVESFGPEASA</td>
<td>(+ 8) 31</td>
</tr>
<tr>
<td>Cryptdin-4</td>
<td>MKTLVLLSALVLLAFGVQADP1ONTDEETFKEEQPGEEDQAVSISFGQGEGSA</td>
<td>(+ 8) 32</td>
</tr>
<tr>
<td>HD-5</td>
<td>MRT1AILAAILLVALQAGAESLQERADEATTQKRSGEDNQDLAISFGNGLSA</td>
<td>(+ 4) 32</td>
</tr>
</tbody>
</table>

Rattusin  
Cryptdin-4  
HD-5  

Fig. 2
<table>
<thead>
<tr>
<th>Tissue</th>
<th>DNA Ladder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stomach</td>
<td>100 bp DNA ladder</td>
</tr>
<tr>
<td>Duodenum</td>
<td></td>
</tr>
<tr>
<td>Proximal jejunum</td>
<td></td>
</tr>
<tr>
<td>Middle jejunum</td>
<td></td>
</tr>
<tr>
<td>Distal jejunum</td>
<td></td>
</tr>
<tr>
<td>Proximal ileum</td>
<td></td>
</tr>
<tr>
<td>Middle ileum</td>
<td></td>
</tr>
<tr>
<td>Distal ileum</td>
<td></td>
</tr>
<tr>
<td>Cecum</td>
<td></td>
</tr>
<tr>
<td>Colon</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 4
Fig. 5
Fig. 6
Fig. 7

A

[Graph A showing survival (CFU/ml) against NaCl (mM) with two lines: one for No peptide and one for Rattusin.]

B

[Graph B showing survival (CFU/ml) against NaCl (mM) with one line for No peptide and one for Rattusin, with survival values below 3 × 10^2.]
Fig. 8
Fig. 9
CHAPTER VI

SUMMARY AND FUTURE PROSPECT

Using a comprehensive computational search strategy, we identified the complete repertoire of the $\alpha$-defensin gene family in the human, chimpanzee, rat, and mouse, including a number of novel genes. $\alpha$-defensin genes were located on syntenic chromosomes in each species. Phylogenetic analysis revealed that $\alpha$-defensin genes in rodents have undergone repeated duplication which gave rise to the current repertoire of genes. We also found that positive selection is acting on the $\alpha$-defensin mature peptide coding region. Further, gene expression studies with rat $\alpha$-defensin genes revealed that they are either expressed in bone marrow or small intestine.

We also identified the entire family members of $\beta$-defensins including a large number of novel genes in the human, chimpanzee, rat, mouse and dog. Each species possessed $>$35 $\beta$-defensin genes encoded in their genome. Such genes were located on 4-5 syntenic chromosomes across species. In contrast to most known genes with two-exon gene structure, several novel $\beta$-defensin genes had three exons. Phylogenetic analysis showed that most of the genes are conserved in all species. However, certain gene lineages were only present in certain species, indicating that these genes appeared by gene duplication after species separation. Gene expression analysis of rat $\beta$-defensins
revealed that all but one are expressed in the male reproductive tract especially in testis and epididymis. Further, this expression was developmentally regulated with the highest level occurring at sexual maturity. This indicates that defensins may have a role in both reproduction and antimicrobial protection.

A novel α-defensin-related gene, which we named rattusin, was identified in the rat and preferentially expressed in small intestine. The encoded peptide displayed an entirely different cysteine-spacing pattern, suggesting that it might have distinct antibacterial properties. Putatively mature rattusin possessed a broad-spectrum antibacterial activity with similar efficiency against both antibiotic-sensitive as well as -resistant bacteria. Further, this activity of rattusin was salt-insensitive although salt slightly delayed kinetics of bacterial killing, in contrast with human defensin-5 and mouse cryptdin-4, both of which are among the most potent defensins identified. Additionally, rattusin was active in the presence of 2-5 mM of Mg++, whereas cryptdin-4 lost its activity completely. Rattusin also displayed little toxicity towards human colon epithelial cells.

To our knowledge, rattusin is the only defensin-related peptide with salt-insensitive antibacterial activities, and therefore, has great potential to become a novel anti-infective drug. Elucidation of its tertiary structure will help explain the structural-basis for such an antibacterial property. In addition, this study has also identified a large number of novel α- and β-defensins. Their specific functions in reproductive health and
fertility largely remain unknown. Further studies are needed to explain the biological significance for such a large number of defensin genes.
VITA

Amar Ajit Patil

Candidate for the Degree of

Doctor of Philosophy

Thesis: GENOME-WIDE IDENTIFICATION OF MAMMALIAN DEFENSINS AND FUNCTIONAL ANALYSIS OF A NOVEL DEFENSIN-RELATED PEPTIDE

Major Field: Animal Science (Molecular Biology)

Biographical:

Personal Data: Born in Belgaum district of India, Son of Mr. Ajit and Mrs. Surekha Patil

Education

2000-2002: Master of Veterinary Science (Animal Biotechnology)
    Indian Veterinary Research Institute, Izatnagar, India
1995-2000: Bachelor of Veterinary Science & Animal Husbandry
    Nagpur Veterinary College, Nagpur
    Dr. P.D.K.V. Agricultural University, Akola, India

Experience

2003-2006: Graduate Research Assistant, Molecular biology lab,
    Dept. Animal Science, Oklahoma State University,
    Stillwater, Oklahoma
2000-2002: Junior Research Fellow, National Biotech. Center
    Indian Veterinary Research Institute, Izatnagar, India
2000: Veterinary Intern
    Various State Government veterinary hospitals, India

Professional Memberships

American Association for Advancement of Science (AAAS)
Scope and Method of Study: The purpose of this study was to identify novel antimicrobial defensins encoded in the genomes of evolutionarily diverse species such as dog, rat, mouse, human and chimpanzee. Our hypothesis was that identification of novel genes will help us to understand the evolution of this important gene family and provide novel antimicrobial candidates. Genome-wide computational searches of various mammals were conducted to identify the defensin genes using known defensins as query. Further, a novel defensin-related sequence named rattusin with a unique cysteine-spacing pattern was synthesized and evaluated for its antibacterial activity to test the hypothesis that it possesses a unique antibacterial activity. Its precursor peptide was also produced recombinantly and tested against pathogenic bacteria. In addition, the gene expression of rattusin was studied across the entire gastrointestinal tract.

Findings and Conclusions: A large number of defensins were found. α-defensin genes were present only in primates and glires, whereas β-defensins genes were identified in all species studied. Computational mapping revealed that all α- and β-defensin genes were clustered on syntenic chromosomes across all species with 4 clusters in each species, except that humans have 5 clusters. Phylogenetic and evolutionary analysis revealed that α-defensin were subjected to rapid evolution and diversification. α-defensin genes were primarily expressed in bone marrow and small intestine, whereas β-defensin genes were mainly expressed in the male reproductive tract. A novel α-defensin-related gene rattusin was identified in rat. This gene displayed a preferential expression in the distal small intestine. Precursor of rattusin did not exhibit significant antibacterial activity, but the C-terminal putatively mature antibacterial peptide showed potent, broad spectrum, salt insensitive antibacterial activity in contrast to cryptdin-4 and human defensin-5. Rattusin also maintained antibacterial activity in the presence of divalent cation, Mg++. Rattusin showed no cytotoxicity to intestinal epithelial cells. Thus, rattusin represents a potential antimicrobial drug to treat infections.