

# **Analysis of Two Human Gene Clusters Involved in Innate Immunity**

Dissertation

zur Erlangung des Doktorgrades  
der Mathematisch-Naturwissenschaftlichen Fakultät  
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Zhihong Wu

Kiel

Mai 2005

Referent: Prof. Dr. rer. nat., Dr. h. c. Thomas C. G. Bosch

Korreferent: Prof. Dr. rer. nat. Jens-Michael Schröder

Tag der mündlichen Prüfung: 2005-12-08

Zum Druck genehmigt: Kiel, den 2005-12-21

-----  
Der Dekan

# Contents

---

<b>List of Abbreviations</b> .....	i
<b>1 Introduction</b> .....	1
1.1 Chemical antibiotics: an unsolved and growing problem .....	1
1.2 The importance of the innate immune response in living organisms .....	2
1.3 Antimicrobial peptides (AMPs) .....	3
1.3.1 Discovery of AMPs .....	4
1.3.2 Classification of AMPs .....	4
1.3.3 The importance of AMPs in human skin .....	6
1.4 The barrier function of human skin .....	7
1.4.1 The cornified cell envelope (CCE) .....	8
1.4.2 The epidermal differentiation complex (EDC) .....	9
1.4.3 The functions of the SFTP family members .....	10
1.5 Two emerging novel families of AMPs .....	12
1.6 Objective of this study .....	14
<b>2 Materials and Methods</b> .....	16
2.1 Patients .....	16
2.2 Materials .....	17
2.3 Tissues and RNA preparation .....	18
2.4 Cell Culture .....	18
2.5 Conventional RT-PCR .....	19
2.6 Real-time RT-PCR .....	20
2.7 <i>In-silico</i> cloning of the novel human SFTP and SPINK genes .....	21
2.8 5' and 3' Rapid Amplification of cDNA Ends (RACE) .....	23
2.9 Long-distance PCR and the short-range PCR-walking strategy .....	25

2.10	Subcloning and expression. . . . .	25
2.10.1	Preparation of expression constructs. . . . .	25
2.10.2	Induction and expression of the polyhistidine-tagged fusion protein . . . . .	26
2.10.3	Digestion and purification of the recombinant protein . . . . .	28
2.10.4	Protein analysis . . . . .	28
2.11	Generation of goat antibodies . . . . .	28
2.12	Immunoblot analysis. . . . .	29
2.13	Mass spectrometry . . . . .	29
2.14	Enzymatic Assays . . . . .	30
2.15	Antimicrobial activity . . . . .	30
<b>3</b>	<b>Results</b> . . . . .	<b>31</b>
3.1	The human clustered S100 fused-type protein (SFTP) gene family . . . . .	31
3.1.1	Characterization of the full-length human hornerin cDNA . . . . .	31
3.1.2	Characterization of three novel SFTP members cDNAs. . . . .	32
3.1.3	Genomic structure of human SFTP genes. . . . .	34
3.1.4	Tissue distribution of human SFTP expression. . . . .	36
3.1.5	Expression of human SFTP genes in skin disease tissues. . . . .	38
3.1.6	Comparative analysis of human SFTP mRNA-expression in proliferating and differentiating human keratinocytes . . . . .	41
3.2	Identification of four novel human kazal-type serine protease inhibitor-like cDNAs (SPINKs) . . . . .	44
3.2.1	Characterization of the full-length human LEKTI-2 cDNA. . . . .	44
3.2.2	Characterization of SPINK9 homologous cDNAs. . . . .	46
3.2.3	Genomic structure of human SPINK9 and its homologs genes. . . . .	50
3.2.4	Expression analysis of the human SPINKs . . . . .	55
3.3	Functional analysis of the human hornerin and LEKTI-2 . . . . .	57
3.3.1	Purification and Characterization of recombinant hornerin fragments. . . . .	57
3.3.2	Purification and Characterization of recombinant LEKTI-2. . . . .	63
3.3.3	Analyses of protease activity of HRNR3. . . . .	64

3.3.4 Analyses of antimicrobial activity of human hornerin fragments and LEKTI-2. ....	64
3.3.5 Generation of polyclonal goat antisera against hornerin fragments and LEKTI-2. ....	67
<b>4 Discussion</b>	70
4.1 The human clustered S100 fused-type protein (SFTP) gene family .....	71
4.2 The human clustered kazal-type serine protease inhibitor (SPINK) gene family	74
4.3 Structural and functional implications for the novel SFTP/SPINK gene products based on comparative sequence analysis .....	80
4.4 The SFTP genes expression upregulated by cross-talking of the mitogenactivated protein kinases pathways .....	85
4.5 The SPINK7-v2: a chimeric protein encoded by two chromosomes .....	86
4.6 The SPINK9 product: LEKTI-2 .....	88
4.7 The hornerin .....	90
4.8 The polyclonal goat antisera against hornerin fragments and LEKTI-2 .....	91
<b>Abstract</b>	92
<b>Zusammenfassung</b>	95
<b>Reference</b>	98
<b>Appendix</b>	110
5.1 Nucleotide and amino acid sequences of human THHL1 cDNA .....	110
5.2 Nucleotide and amino acid sequences of human repetin cDNA. ....	111
5.3 Nucleotide and amino acid sequences of human hornerin cDNA. ....	112
5.4 Nucleotide and amino acid sequences of human ifapsoriasin cDNA. ....	113
<b>Erklärung</b>	114
<b>Acknowledgments</b>	115
<b>Curriculum Vitae</b>	116

## List of Figures

---

1.1	Preparative C8 RP-HPLC analysis of heparin-bound proteins from pooled stratum corneum extracts of healthy person's heel callus. . . . .	15
3.1	Schematic structure of human hornerin protein. . . . .	33
3.2	Schematic structure of human repetin protein. . . . .	35
3.3	Schematic structure of human ifapsoriasin protein. . . . .	35
3.4	Comparison of the genomic structure of all the human SFTP genes. . . . .	36
3.5	RT-PCR analysis of human SFTP expression. . . . .	37
3.6	Expression of human SFTPs in skin disease tissues. . . . .	40
3.7	Human SFTP-expression in human foreskin-derived keratinocytes in response to calcium or RA. . . . .	42
3.8	The ability to induce SFTPs gene expression is dependent on crosstalking of MAPKs. . . . .	43
3.9	The ability of calcium to induce SFTP promoter activity is dependent on crosstalking of MAPKs. . . . .	45
3.10	Nucleotide and amino acid sequences of human SPINK9. . . . .	47
3.11	Nucleotide and amino acid sequences of human SPINK6. . . . .	48
3.12	Nucleotide and amino acid sequences of human SPINK7. . . . .	49
3.13	Nucleotide and amino acid sequences of human SPINK10. . . . .	51
3.14	Genomic structure of the human SPINK9. . . . .	52
3.15	Genomic structure of the human SPINK10. . . . .	52
3.16	Genomic structure of the human SPINK6. . . . .	53

3.17	Genomic structure of the human SPINK8. . . . .	54
3.18	Genomic structure of the human SPINK7. . . . .	55
3.19	Genomic PCR analysis of the human <i>SPINK7</i> gene. . . . .	56
3.20	RT-PCR analysis of human SPINKs mRNA-expression. . . . .	57
3.21	RT-PCR analysis of human SPINKs expression in tumor skin. . . . .	58
3.22	Purification and characterization of the His-tagged fusion protein His-HRNR1. . . . .	59
3.23	Purification and characterization of the hornerin fragment HRNR2. . . . .	60
3.24	Purification and characterization of the hornerin fragment HRNR3. . . . .	61
3.25	Purification and characterization of the hornerin fragment HRNR4. . . . .	62
3.26	SDS-PAGE analyses of fusion proteins and cleaved mature polypeptides. . . . .	63
3.27	Purification and characterization of recombinant LEKTI-2. . . . .	65
3.28	ESI-MS spectra of HRNR3 in 10 mM NH <sub>4</sub> HCO <sub>3</sub> buffer. . . . .	66
3.29	Titrations of goat antisera in dot-blot assays. . . . .	68
3.30	Titrations of goat anti-HRNR4 antisera in a dot-blot assay. . . . .	69
4.1	Multiple alignment of the conserved domains of the SFTP proteins. . . . .	72
4.2	Comparative analysis of a cluster of human SFTPs at 1q21.3 with the corresponding region of mouse genome. . . . .	73
4.3	Unrooted cladogram of the human SFTP and S100A family and their ortholog in chimpanzee, mouse and rat. . . . .	75
4.4	Multiple alignment of the kazal domains of the human SPINKs protein family. . . . .	76
4.5	Schematic representation of subcellular destination of the human SPINK9. . . . .	77
4.6	Comparative analysis of a cluster of human SPINKs at 5q32 with the corresponding region of mouse and rat genomes. . . . .	78
4.7	Multiple alignment of the kazal domains of the SPINKs protein family. . . . .	80
4.8	Unrooted cladogram of the human SPINK family and their ortholog in mouse and rat. . . . .	81

4.9 Comparative analysis of the human epidermal differentiation complex (EDC) at 1q21 with the corresponding region of chimpanzee, mouse, rat and dog genome. . . . .	84
4.10 Comparison of the promotor structure of the different human SFTP genes. .	95



## List of Tables

---

1.1	AMPs—structural classes . . . . .	5
2.1	Characteristics of patients . . . . .	16
2.2	Primer sequences used for RT-PCR and quantitative real-time RT-PCR . . . . .	20
2.3	PCR conditions for each primer pair shown in Table 2.2 . . . . .	20
2.4	Primer sequences used for 5'RACE . . . . .	23
2.5	Primer sequences used for 3'RACE . . . . .	24
2.6	Primer sequences used for protein expression . . . . .	27
3.1	Realtime RT-PCR analysis of human SFTP transcripts. . . . .	38
3.2	Expression analysis of human SFTPs in tumor skin. . . . .	39
3.3	MBCs and LD90s of hornerin peptides and LEKTI-2. . . . .	66

## List of Abbreviations

$\alpha$	alpha
ACN	Acetonitril
AMP	antimicrobial peptide
AP-1	Transcription activator protein 1
ATCC	American Type Culture Collection
BCC	basal cell carcinoma
BEC	bronchial epithelial cells
bp	base pair
BLAST	Basic Local Alignment Search Tool
cDNA	complementary DNA
CCE	cornified cell envelope
DEPC	Diethylpyrocarbonate
DMEM	Dulbeccos Modified Eagle Medium
DMSO	Dimethylsulfoxid
DNA	Deoxyribonucleic Acid
dNTP	Desoxy-Nucleotide-5'-triphosphate
DTT	Dithiothreitol
EBI	European Bioinformatics Institute
<i>E. coli</i>	<i>Escherichia coli</i>
EDC	epidermal differentiation complex
EDTA	Ethylen-diaminetetraacetate
EMBL	European Molecular Biology Laboratory
ERK	Extracellular Signal-Regulated Protein Kinase
<i>et al.</i>	<i>et alii</i>
FCS	Fetal calf serum
FLG	filaggrin
GAPDH	Glycerinaldehyd-3-Phosphate Dehydrogenase
HCl	Acide chlorhydrique
HPLC	High performance liquid chromatography
hr	hour
HRNR	hornerin
IF	intermediate filament
IFPS	ifapsoriasis
IPTG	isopropyl thio- $\beta$ -D-galactoside
JNK	c-Jun-N-Terminal Kinase
kb/kbp	kilobases/kilobase pairs
kD or kDa	kilo-Dalton (protein molecule mass)
KLH	keyhole limpet hemocyanin

---

LEKTI	lympho-epithelial kazal-type-related inhibitor
LEP	late envelope proteins
M	molar (mol/l)
MAPK	Mitogen-Activated Protein Kinase
MEK	MAPK kinase
mRNA	messenger RNA
MgCl <sub>2</sub>	Magnesiumchlorid
min	minute
ml/mg	milliliter/milligram
ng	nanogram
OD	Optical Density
ORF	Open Reading Frame
PAM	pathogen associated molecule
PBS	phosphate buffered saline
PDB	Protein Data Bank
PFLG	profilaggrin
PL	precancerous lesions
RA	retinoic acid
RACE	Rapid amplification of cDNA ends
RPTN	repetin
RT	Reverse Transcription
SCC	squamous cell carcinoma
SDS	Sodium dodecyl sulfate
SFTP(s)	S100 fused-type protein(s)
SMART	Simple Modular Architecture Research Tool
SPINK(s)	kazal-type serine protease inhibitor(s)
SPRR	small proline rich proteins
TAE	Tris-Acetate-EDTA
TEC	trachea epithelial cells
Temp.	Temperature
TFA	trifluoroacetic acid
THH	trichohyalin
THHL1	trichohyalin-like 1
TSB	Trypticase-Soy Broth
TSS	transcriptional start site
μl/μg	microliter/microgram

# Chapter 1

## Introduction

---

The continuous use of antibiotics has resulted in multi-resistant bacterial strains all over the world. The emergence of resistance to commonly used antibiotics has stimulated the search for new naturally occurring bactericidal and fungicidal agents that may have clinical utility. Antimicrobial peptides (AMPs) are widely distributed effector molecules of innate immunity in animal kingdom that provide a first line of defense against an array of microorganisms (Liu et al., 2000; Vizioli and Salzet, 2002). AMPs of animal origin may be an effective alternative or additive of conventional antibiotics for therapeutic use (Hancock and Chapple, 1999). Phase I and II trials have shown a limited resistance of AMPs for the bacterial strains tested (Zasloff, 2002). These features make the antibiotic peptides a powerful arsenal of molecules that could be the antimicrobial drugs of the new century.

### **1.1 Chemical antibiotics: an unsolved and growing problem**

Antibiotics are chemical substances produced by microorganisms that either destroy or inhibit the growth of other microorganisms. Antibiotics can have a broad spectrum of activity, which means that they are active against a wide range of microorganisms including both pathogenic and nonpathogenic strains.

Moreover, many microorganisms are able to produce enzymes that neutralize antimicrobial compounds. Others can acquire resistance to weak, short-term or repeated exposure to antibiotics. Antibiotic resistance will, in the very near future, leave healthcare

professionals without effective therapies for bacterial infections. As an example, it is now estimated that about half of all *Staphylococcus aureus* strains found in many medical institutions are resistant to antibiotics such as methicillin (Roder et al., 1989).

Consequently, the priority for the next decades should be focused on the development of alternative drugs and/or the recovery of natural molecules that would allow the consistent and proper control of pathogen-caused diseases. Ideally, these molecules should be as natural as possible, with a wide range of action over several pathogens, easy to produce, and not prone to induce resistance.

## **1.2 The importance of the innate immune response in living organisms**

Living organisms are exposed daily to microbes including pathogens, and in order to defend themselves against the abrasive environment, they have developed potent defense mechanisms that are part of innate and adaptive immunity. In invertebrates, only innate immunity, recognized as an ancient mechanism, is present. In vertebrates, defense reactions rely on both innate and adaptive immunity (Hoffmann et al., 1999). The innate immune system provides a first line of host defense against many common microorganisms and is essential for the control of common bacterial infections (Medzhitov and Janeway, 2002). All metazoans have inborn defense mechanisms that constitute innate immunity (Ezekowitz and Hoffmann, 1996).

There are a number of components of host innate immunity. These include physical elements such as the barrier formed by the epithelial cells, and local production of proteins, like cytokines, as well as neutrophil recruitment (Nemes and Steinert, 1999; Neutra et al., 2001). Other components of the innate immune response include the complement system, scavenger cells (macrophages and neutrophils), and the production of antimicrobial substances, which are mainly peptides or polypeptides (Krisanaprakornkit et al., 2000).

Our understanding of innate immune mechanisms has recently undergone a major advance with the discovery of toll-like receptors (TLRs). These are a series of germ-

line encoded pattern-recognition receptors (PRRs) which recognize conserved pathogen-associated molecular patterns (PAMPs) (Akira and Sato, 2003). PRRs can act to enhance the binding of PAMPs to cell surface receptors, to activate inflammatory responses or to induce phagocytosis of the microbe, or to effect a combination of these actions (Barton and Medzhitov, 2002; Takeda and Akira, 2003; Underhill and Ozinsky, 2002). Intracellular PRRs called NODs have been described; these are involved in generating an inflammatory response to specific components of the bacterial cell wall (Inohara and Nunez, 2003). These PRRs provide the body with the ability to detect and respond to a wide variety of bacteria found in various compartments of the body.

Certainly, all the components of innate immunity mentioned above play a role in host defense. The overall response is the result of the complex interplay between a number of different bacterial molecules reacting with their corresponding receptors on the cell surface. However, an essential element to the concept of innate immunity is immediate responsiveness to danger. Recruitment of neutrophils and macrophages takes time, usually hours, whereas pathogenic bacteria might replicate rapidly *in vivo*. In a complex system such as in humans, an invading microorganism can simply be eliminated by this primary reaction - the innate response - without requiring an activation of the adaptive immunity, the next step in this complex cascade (Bals, 2000). Antimicrobial peptides (AMPs) have emerged as key elements for inhibition of microbial proliferation prior to recruitment of cellular immune defense elements.

### **1.3 Antimicrobial peptides (AMPs)**

Antimicrobial peptides (AMPs) are ubiquitous molecules as they are found in microorganisms as well as in plants (Cammue et al., 1992), invertebrates and also vertebrates (Lehrer and Ganz, 1996; Schröder, 1999; Zasloff, 2002). They represent the first line of defense against pathogens acting as effector molecules of innate immunity. The emergence of resistance to commonly used antibiotics has stimulated the search for new naturally occurring bactericidal and fungicidal agents that may have clinical utility.

### 1.3.1 Discovery of AMPs

In the late 1870s scientists were searching for an agent to kill microbes without causing unacceptable damage to the hosts. In the search for this agent, Ehrlich started to work on mammalian granulocytes and noted the different staining characteristics of these cell (Spitznagel, 1997). In 1890s, Petterson found that aqueous extracts of pus from human emphysema had antimicrobial activity (Spitznagel, 1997). In following decades, especially after the discovery of penicillin, lots of antimicrobial peptides were isolated, such as tyrocidine and gramicidin antimicrobial peptides from *Bacillus brevis*, mellitin from bee venom, and cationic antimicrobial proteins from rabbit (Zeya and Spitznagel, 1969). However, an important facet of this defense mechanism was revealed two decades ago by the discovery of cecropins, inducible antimicrobial peptides in the giant silk moth *Hyalophora cecropia*. In 1981, Boman and his associates reported that these widespread biochemical defense peptides act as antibiotics (Steiner et al., 1981). At first it was thought that these antimicrobial peptides were unique to insects, but later they were isolated from other animals including mammals revealing that these peptides were widely distributed in the animal kingdom and provide enormous survival benefits to the host (Zasloff, 1992). Because these peptides are very potent against bacteria, but have no toxic or hemolytic effect on host cells, and have a wide taxonomic distribution, their discovery led to the start of a new era in studies of animal antimicrobial peptides (Boman, 1995).

### 1.3.2 Classification of AMPs

A large number of AMPs has been found in insects, amphibians and mammals. AMPs are widely distributed in nature and represent an ancient mechanism of host defence. More than 1700 such AMPs have been discovered, and an updated list can be found at the website: <http://research.i2r.a-star.edu.sg/Templar/DB/ANTIMIC/>. Most of the antimicrobial peptides are cationic although some anionic peptides have been reported (Fales-Williams et al., 2002).

On the basis of their structural features, cationic peptides can be divided into four

classes in Table 1.1 (Tossi and Sandri, 2002; Zasloff, 2002):

Table 1.1: AMPs—structural classes

Cl.	Structure	Example	Origin
1	Linear, $\alpha$ -helix	Magainin Clavanin	Frog Tunicate haemocytes
	Linear, 2 $\alpha$ -helix	Cecropin A	Insect haemolymph
2	Extended, high proportion of a single amino acid	Bactenecin <sup>7</sup> (Proline-rich)	Bovine neutrophils
		Holotricin <sup>3</sup> (Glycin-rich)	Insect haemolymph
		Histatin (Histidine-rich)	Primate saliva
		Indolicidin (Tryptophan-rich)	Bovine neutrophils
3	1-disulfate, loop	Bactenecin	Bovine neutrophils
		Ranalexin	Frog skin
4	2-disulfide, $\beta$ -sheet	Protegrins	Pig leukocytes
	3-disulfide, $\beta$ -sheet	Tachyplesins	Horseshoe crab haemocytes
		HNPs ( $\alpha$ -defensin)	Human neutrophils
	3-disulfide, $\beta$ -sheet + $\alpha$ -helix	HBDs ( $\beta$ -defensin)	Human skin
	3-disulfide, $\beta$ -sheet + 2 $\alpha$ -helix	Phormecin (Insect defensin)	Insect haemolymph
	4-disulfide, $\beta$ -sheet + $\alpha$ -helix	Thionins ( $\gamma$ -thionin)	Plant tissue
		AFP-1	Plant tissue
		MGD-1	Mollusc haemolymph

(1) linear peptides forming-helical structures. This family includes the cecropins from the haemolymph of insects (Hultmark et al., 1980; Andreu and Rivas, 1998; Boman, 1998; Zheng and Zheng, 2002), the magainins isolated from *Xenopus* skin (Zasloff, 1987; Bechinger et al., 1993; Simmaco et al., 1998), and the histone derived compounds isolated from fish epithelia (parasin) (Park et al., 1998), the cytoplasm of murine macrophages (Hiemstra et al., 1993) and human wound fluids (Frohm et al., 1996);

(2) cysteine-rich open-ended peptides containing single or several disulfide bridges (Lehrer and Ganz, 2002). This family mainly contains the thionins isolated from plants (Bohlmann et al., 1998) and the defensins, a highly complex group of 4-kDa open-ended cysteine-rich peptides arranged with different structural motifs. Defensins have been isolated from mammals, insects and plants and they serve as effector molecules of innate immunity, providing an efficient initial defence against infectious pathogens (Ganz, 1999; Selsted et al., 1984). Many defensin-like antimicrobial peptides have been identified from scorpions (Cociancich et al., 1993; Ehret-Sabatier et al., 1996) and mussels (Charlet et al., 1996). In mammals,  $\alpha$ - and  $\beta$ -defensins are characterized by an antiparallel- $\beta$  sheet



structure, stabilized by three disulfide bridges (Zasloff, 2002). Some of them naturally exist as cyclic molecules such as the  $\theta$ -defensins (Tang et al., 1999; Lehrer and Ganz, 2002).

(3) molecules rich in specific amino acids such as proline, glycine, histidine and tryptophan. Those proline- and glycine-rich are mostly from insects and active against Gram-negative bacteria (Bulet et al., 1999; Otvos, 2000); while those enriched in histidine are particularly basic, mostly from mammals (Pollock et al., 1984) and primarily directed against fungal pathogens (Tsai and Bobek, 1998). Only two peptides enriched in tryptophan residues have been described, both derived from porcine cathelicidin precursors (Schibli et al., 2002).

### 1.3.3 The importance of AMPs in human skin

As described above, more than 1700 of naturally occurring peptide antibiotics have been discovered based on their ability to inhibit the growth of microbial pathogens. These AMPs participate in the innate immune response by providing a rapid first-line defence against infection. Current evidence shows that cathelicidins and defensins act as both natural antibiotics and as signalling molecules that activate host cell processes involved in immune defence and tissue repair. Alterations in the expression pattern of AMPs have been associated with a variety of pathological processes. Ongoing and future studies are likely to implicate AMPs in several unexplained human inflammatory disorders and to provide novel therapeutic approaches for the treatment of these diseases.

AMPs play also a magic role in the skin defense system. The skin serves as a protective barrier between internal organs and the outside environment. Skin epithelium functions naturally as a physical barrier. It has an active immunological role in antigen processing and presentation as well as production of cytokines and defense molecules such as AMPs. The number of reports demonstrating the presence and upregulation of antimicrobial peptides in human skin is increasing and reflects the significance of these peptides in cutaneous innate immunity (Bardan et al., 2004).

Various AMPs, like the cathelicidin LL-37/hCAP18,  $\beta$ -defensins hBD1–4, RNase-7, psoriasin and dermcidin have been found in human skin (Frohm et al., 1997; Harder et al.,

1997; Harder et al., 2001; Harder and Schröder, 2002; Gläser et al., 2005; Schitteck et al., 2001), protecting skin against invading microorganisms (Peschel, 2002; Bardan et al., 2004; Harder and Schröder, 2005). These peptides are produced by epithelial cells and have been proposed to provide a first line of defence against infection by acting as 'natural antibiotics'. The human LL-37/hCAP18, HBD2, HBD3, and RNase-7 are expressed by human skin keratinocytes at sites of inflammation in disorders such as psoriasis, nickel contact dermatitis, systemic lupus erythem and Netherton syndrome and function primarily in the response to injury (Frohm et al., 1997; Harder et al., 1997; Harder et al., 2001; Harder and Schröder, 2002). The dermcidin is specifically and constitutively expressed in the sweat glands, secreted into the sweat and transported to the epidermal surface (Schitteck et al., 2001). Recently, the S100A7/psoriasin was characterized as a new generation of AMPs in human healthy skin (Gläser et al., 2005).

Interestingly, the activity of AMPs is not limited to the killing of microorganisms. In addition, several other actions have been identified. In fact, their multiple roles as mediators of inflammation with effects on epithelial and inflammatory cells has been demonstrated. They are involved in such diverse processes as proliferation, immune induction, wound healing, cytokine release, chemotaxis, protease/antiprotease balance, and redox homeostasis (Ganz, 2002; Cole et al., 2003; Com et al., 2003; Liu et al., 2003).

## **1.4 The barrier function of human skin**

The outermost skin layer, the epidermis, provides the first line of defense against pathogens. Historically, keratinocytes, which form 95% of all epidermal cells, were believed to function purely in maintaining the structure of the epidermis via their production of cytokeratins and in the maintenance of a physical barrier to a variety of exogenous microorganisms. The stratified epithelia is the most superficial layer of the skin and is subdivided into five layers, the stratum germinativum, the stratum spinosum, the stratum granulosum, the stratum lucidum and the stratum corneum. The stratum corneum, the final product of epidermal cell differentiation, forms a protective barrier from physical,

chemical and mechanical insults including microorganisms whilst protecting body homeostasis such as temperature regulation and prevention of fluid loss (Nemes and Steinert, 1999).

A major portion of the barrier function of skin is due to the presence of the cornified cell envelope (CCE), which is a specialized structure formed immediately beneath the plasma membrane of keratinocytes during the last stage of their differentiation (Steinert and Marekov, 1999). The mechanical behavior of the CCE is characterized by its rigidity and insolubility (Regnier et al., 1993; Nemes and Steinert, 1999).

#### **1.4.1 The cornified cell envelope (CCE)**

The stratum corneum composed of multiple layers of flattened dead cells, forms a tough, impenetrable barrier that shields the underlying living cells from hazards of the surrounding environment (Fuchs et al., 1994; Polakowska and Haake, 1994; Ishida-Yamamoto and Lizuka, 1998; Steinert and Marekov, 1999). Corneocytes are the end-products of the terminal differentiation pathway of keratinocytes in these tissues. Their organelles, having broken down, consist simply of a keratin filament matrix encased within the CCE. The CCE is assembled late in the pathway when it replaces the cytoplasmic membrane, and consists of a layer of cross-linked protein coated with covalently bound lipids (Swartzendruber et al., 1987). Among the many morphologic changes that occur in normal epidermis during the transition from granular to cornified cells are the dissolution of the nucleus and other organelles, the aggregation of the keratin intermediate filament (IF) network into macrofibrils, and the formation of a cornified cell envelope consisting of loricrin, the small proline-rich proteins, involucrin, and other proteins that are crosslinked by transglutaminases in the stratum corneum (Nemes and Steinert, 1999; Presland and Dale, 2000; Kalinin et al., 2001). However, the maintenance of barrier functions depends on epithelial interactions with the dermis via the continuous process involving keratinocyte proliferation and terminal differentiation.

### 1.4.2 The epidermal differentiation complex (EDC)

In humans, at least 47 genes specially involved in the differentiation process have been discovered and are clustered within the epidermal differentiation complex (EDC) in chromosome 1q21 (Volz et al., 1993; Mischke et al., 1996). They encode structural as well as regulatory proteins which cooperate during epidermal differentiation. They also show evolutionarily conserved structural relationship together with functional interdependence of the encoded proteins and can be grouped into three structurally related gene families (Marenholz et al., 2001).

The members of the first family encode the CCE precursor proteins. They include involucrin, loricrin, small proline rich proteins (SPRR) (reviewed by Tesfaigzi and Carlson, 1999) and 18 late envelope proteins (LEP) (Marshall et al., 2001) that are crosslinked by the action of keratinocyte transglutaminase at the cell periphery (review by Grenard et al., 2001). Loricrin, involucrin, SPRR, LEP and the distantly related NICE-1 (Marenholz et al., 2001) are encoded by genes with similar structures. The proteins encoded have homologies in the terminal protein domains and contain a variable number of internal tandem repeats specific for each protein, and are major precursors for the building of the CCE.

The second family encode S100 calcium-binding proteins (Schafer and Heizmann, 1996; Heizmann et al., 2002). The S100 proteins comprise a multigene family of low molecular weight proteins that engage in multiple functions in a wide variety of cell types and tissues (Heizmann et al., 2002; Donato, 2003). S100 proteins are characterized by common structural motifs including two EF hands (helix-loop-helix calcium-binding domains) that are separated by a hinge region and flanked by amino- and carboxy-terminal domains. The canonical C-terminal EF-hand binds calcium with a 100-fold higher affinity than the N-terminal, non-conventional, EF-hand (Gribenko and Makhatadze, 1998; Zimmer et al., 2003). They are regulatory proteins primarily involved in different steps of the calcium signal transduction pathway regulating cell shape, cell cycle progression, and differentiation (Eckert et al., 2004). Among the 21 S100 proteins that have been cloned to date, eleven, including S100A2, S100A3, S100A4, S100A6, S100A7, S100A8, S100A9,

S100A10, S100A11, S100A12, and S100A15, are expressed in the human epidermis or in cultured keratinocytes (Boni et al., 1997; Xia et al., 1997; Broome et al., 2003). They may play a role in the pathogenesis of epidermal diseases such as psoriasis, skin cancer, and skin inflammation (Eckert et al., 2004). Originally characterized as a marker of psoriasis, S100A7 (also called psoriasin) overexpression is seen in many inflammatory skin diseases, including atopic dermatitis, mycosis fungoides, Darrier's disease, and inflammatory lichen sclerosus and atrophicus (Madsen et al., 1991; Algermissen et al., 1996; Broome et al., 2003). More interestingly, psoriasin was found to confer potent bactericidal activity against *E.coli* in skin (Gläser et al., 2005).

The third family of genes encode intermediate filament-associated S100 fused-type proteins (SFTP family). The proteins encoded combines EF hands and internal tandem repeats. They comprises profilaggrin (Fleckman et al., 1985) , trichohyalin (Lee et al., 1993), and c1orf10 (Xu et al., 2000). This family also includes repetin (Krieg et al., 1997) and hornerin (Makino et al., 2001) in mice. The SFTPs are synthesized in the granular layer of the epidermis and conjoin with the keratin filaments of keratinocytes during cornification (O'Guin et al., 1992; Resing et al., 1993). Functionally, they are associated to keratin intermediate filaments and partially crosslinked to the CCE, as is summarized in the next section.

### 1.4.3 The functions of the SFTP family members

Profilaggrin (PFLG) is a major protein component of the keratohyalin granules of mammalian epidermis. It is initially expressed as a large polyprotein precursor which is subsequently processed to functional filaggrin (FLG) units in the terminal phase of epidermal differentiation (Resing et al., 1995). The FLGs show wide species variations and their aberrant expression has been implicated in a number of keratinizing disorders (Baden et al., 1974; Holbrook et al., 1982; Sybert et al., 1985). PFLG is a large, highly phosphorylated, insoluble protein consisting of multiple (10-12 in human, 20 or more in rodents) FLG units joined by linker peptides. PFLG is expressed in the granular layer where it is localized within electron-dense nonmembrane bound keratohyalin granules (Presland and Dale,

2000). This sequestration into keratohyalin is believed to protect the epithelial cell from the deleterious effects of premature keratin aggregation by FLG (Dale et al., 1997; Kuechle et al., 1999; Presland et al., 2001). Recent studies have shown that the profilaggrin N-terminal region (which contains the calcium-binding elements) is specifically removed from the protein by proteolysis and localizes to the nuclei of keratinocytes destined to undergo terminal differentiation. FLG binds to cytoplasmic keratin IFs to form the macrofibrils that are retained in cornified cells, whereas the N-terminal peptide localizes to nuclei of epidermal granular and transition cells (Ishida-Yamamoto et al., 1998; Presland and Dale, 2000). Therefore, PFLG processing provides a potential link between two major events of terminal differentiation: the aggregation of keratin filaments and nuclear dissolution (Pearton et al., 2002).

Trichohyalin (THH) is expressed in specialized epithelia that are unusually mechanically strong, such as the inner root sheath cells of the hair follicle (Rothnagel and Rogers, 1986). In pig, tongue THH is sequentially subjected to post-synthetic modifications by peptidylarginine deaminases, which convert many of its arginines to citrullines, and by transglutaminases, which introduce intra- and interprotein chain cross-links, to form highly cross-linked rigid structures in the cells (Tarcza et al., 1997). In the mouse hair follicle, trichohyalin is a multi-functional cross-bridging protein that functions in the inner root sheath and perhaps in other specialized epithelial tissues by conferring to and coordinating mechanical strength between their peripheral cell envelope barrier structures and their cytoplasmic keratin filament networks (Steinert et al., 2003).

C1orf10 is highly expressed only in the esophagus of normal adult tissues and was undetectable in a total of 15 other tissues examined (Xu et al., 2000). The expression of C1orf10 might be unique to esophageal cells and that loss of its expression may play a role in the development of esophageal cancer (Yagui-Beltran et al., 2001).

Other SFTP members have also been identified in mice. The hornerin (HRNR) protein was detected in keratohyalin granules of epidermal granular cells together with profilaggrin (Makino et al., 2001). The repetitive domain of HRNR was found to be rich in glycine, serine, and glutamine. HRNR was expressed highly in the tongue, esophagus, forestom-

ach, and skin among the adult mouse tissues examined, all of them cornifying stratified epithelium. In skin, HRNR was expressed in the upper epidermis of newborn mouse skin, as was profilaggrin, but the amount of hornerin mRNA was more than fourfold greater than the amount of profilaggrin mRNA (Makino et al., 2003). Repetin (RPTN) is expressed in the stratified internal epithelia of forestomach and tongue and to a lesser degree in normal skin epidermis, where it is restricted to the differentiated suprabasal cell layers (Krieg et al., 1997). Increased expression of RPTN has been found in loricrin-deficient mice to maintaining skin barrier function (Djian et al., 2000; Koch et al., 2000), which implies that compensatory mechanisms exist in epithelia that can overcome the absence of involucrin.

Altogether, all SFTP genes share a common gene structure and all SFTP proteins share similar structural characteristics. Human SFTP genes (profilaggrin, trichohyalin and C1orf10), located in the EDC region, share the same genomic structure, which has three exons and two introns with the typical GT-AG structure. The first exon is located in the 3-terminal untranslated region while the second exon encodes the S100 domain. The third exon is very big and encodes the most part of its protein including the EF-hand and multiple tandem repeats. Therefore, these three proteins together with mouse repetin and hornerin, have one S100 domain at the N terminus followed by the second Ca<sup>2+</sup> binding EF-hand domain and they are therefore thought to have evolved through fusion between genes of the S100 Ca<sup>2+</sup>-binding proteins and genes of cornified envelope precursor proteins.

## 1.5 Two emerging novel families of AMPs

Two emerging novel families of AMPs have been neglected, compared with the well studied cathelicidin and defensins families.

The S100 proteins comprise a family of 21 low molecular weight (9-13 kDa) proteins that are characterized by the presence of two calcium-binding EF-hand motifs (reviewed by Eckert et al., 2004). Fourteen S100 protein genes are located within the epidermal differentiation complex on human chromosome 1q21 and are expressed in normal and/or diseased epidermis. They exist in cells as anti-parallel hetero- and homodimers and

upon calcium binding interact with target proteins to regulate cell function. They have also been associated with a large number of human diseases including inflammation, Down's syndrome, neurological disorders, cystic fibrosis, myocardial infarction, psoriasis, and cancer (Heizmann et al., 2002). The structural organization of S100 genes is highly conserved (Zimmer et al., 1996; Nakamura et al., 1998). The typical gene, with some notable exceptions such as S100A4, S100A5, and S100A14, consists of three exons, where the first exon has the 5' untranslated sequences, the second exon contains ATG and codes for the N-terminal EF-hand, and the third exon encodes the carboxy-terminal canonical EF-hand (Heizmann et al., 2002; Pietas et al., 2002). The amino-terminal loop in exon 2 is S100 specific, with a lower  $\text{Ca}^{2+}$ -binding affinity than the canonical  $\text{Ca}^{2+}$ -binding loop within the carboxy-terminal EF-hand of exon 3. The S100 genes are found in all vertebrates examined so far such as human, bovine, porcine, rat, mouse, rabbit, avian, and frog (Kube et al., 1991; Zimmer et al., 1996; Nakamura et al., 1998).

Two S100 proteins, calprotectin and S100A7/psoriasin have been associated with antimicrobial activity (Steinbakk et al., 1990; Miyasaki et al., 1993; Nisapakultorn et al., 2001; Gläser et al., 2005). Calprotectin is a noncovalently complexed heterodimer of two small anionic proteins S100A8-S100A9 (Johne et al., 1997) and is constitutively expressed in squamous mucosal epithelial cells (Wilkinson et al., 1988). Psoriasin is key to the resistance of skin against *E. coli* (Gläser et al., 2005). Both calprotectin and psoriasin bind zinc, thereby depriving microorganisms of this essential nutrient as an antimicrobial mechanism.

The family of kazal-type protease inhibitors, belongs to MEROPS inhibitor family I1, clan IA (Rawlings et al., 2004). They inhibit serine peptidases of the S1 family (InterPro accession No. IPR001254), such as trypsin and elastase (Rawlings et al., 2004), and are named **serine protease inhibitor, kazal-type (SPINK)**. The Kazal-type motif is characterized by six cysteine residues distributed over a region of about 40–60 amino acids with other residues conserved especially adjacent to Cys3 and a tyrosine located between Cys3 and Cys4, which makes it different from the Kunitz-type serine protease inhibitor sequence motif (Rawlings et al., 2004; Williamson et al., 1984). These proteins contain between 1 and 7 Kazal-type inhibitor repeats (Williamson et al., 1984; Laskowski et al., 1987). There



are a number of inherited diseases that are attributable to abnormalities in peptidase inhibitors. These include forms of emphysema, epilepsy, hereditary angioneurotic oedema and Netherton syndrome (Lomas et al., 2002; Ritchie, 2003; Lehesjoki, 2003; Bitoun et al., 2002). An imbalance between the active enzymes and their natural inhibitors leads to the accelerated destruction of connective tissue, which is associated with the pathology of diseases such as arthritis, cancer, multiple sclerosis and cardiovascular diseases (Peress et al., 1995; Wernicke et al., 1996; Steinman, 1996). The potential for using specific enzyme inhibitors as therapeutic agents to redress this balance has led to intensive research focused on the design, synthesis and molecular deciphering of low-molecular-mass inhibitors of this family of proteins (Johnson et al., 1987; Bottomley et al., 1998; Ahrens et al., 1996). However, PSKP-1, one SPINK isolated from the skin of *Phyllomedusa sauvagii*, shows inhibitory activity against a serum prolyl endopeptidase, but is unable to inhibit trypsin, chymotrypsin, V8 protease, or proteinase K (Gebhard et al., 2004). More interestingly, PSKP-1 can be rendered active against trypsin by active-site site-specific mutagenesis, has bactericidal activity, and induces agglutination of red cells at micromolar concentrations (Gebhard et al., 2004).

## 1.6 Objective of this study

Our research group recently has isolated two antibacterial peptides from skin-derived stratum corneum extracts, one with a N-terminal protein sequence similar to SFTPs (tentatively termed SFT-AMP) and another with homology to kazal-type protease inhibitors (tentatively termed KT-AMP) (Fig. 1.1). The main objective of the study is to characterize and clone these novel AMPs. In detail, our aims are

- 1) to identify and characterize the human genes encoding for KT-AMP and SFT-AMP;
- 2) to resolve the related potential gene clusters expressed in the epidermis;
- 3) to reveal potential control elements and signaling pathways governing expression of the gene cluster;

4) to uncover genes and processes implicated in skin diseases or tumors associated with the two gene clusters.

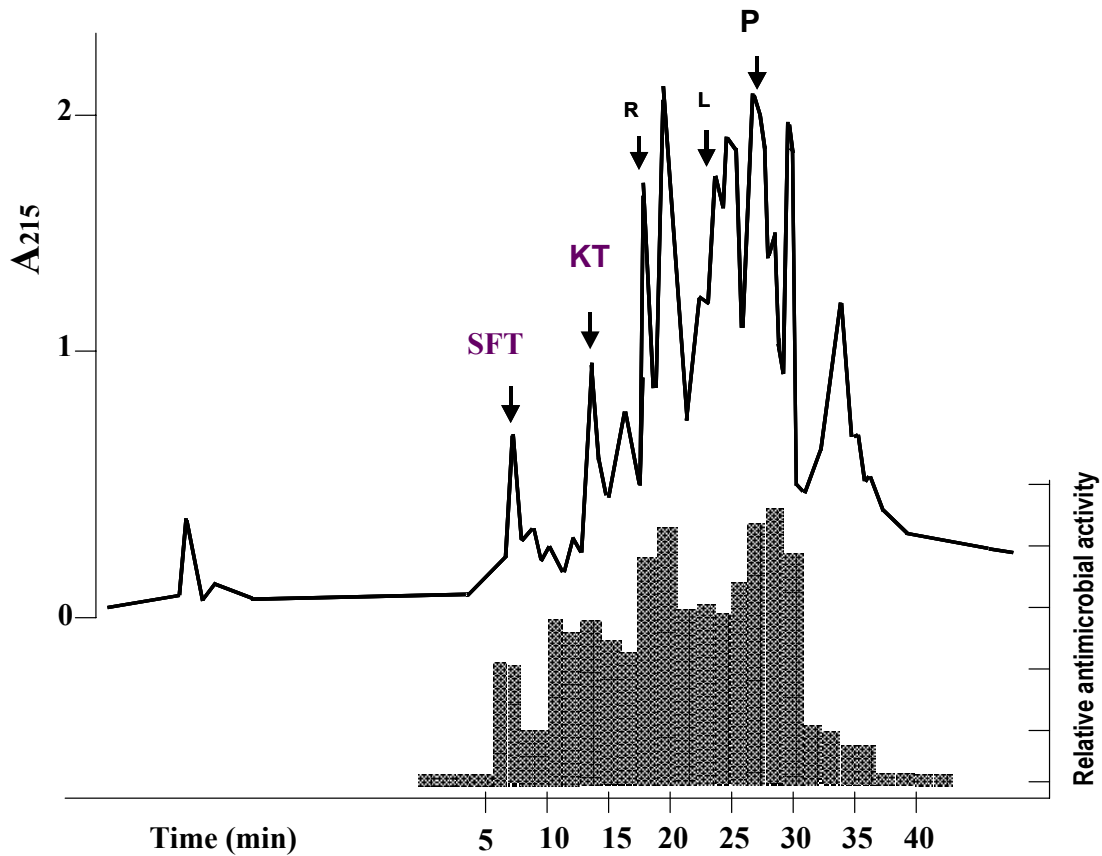


Figure 1.1: Preparative C8 RP-HPLC analysis of heparin-bound proteins from pooled stratum corneum extracts of healthy person's heel callus. Bars indicate the *E. coli*-killing activity of HPLC fractions. Two fractions marked with KT and SFT contain the novel antimicrobial peptides KT-AMP and SFT-AMP, which comprise the main emphasis in this study. Other HPLC fractions contain already known antimicrobial proteins such as psoriasin (P), lysozyme (L) and RNase-7 (R).

# Chapter 2

## Materials and Methods

---

### 2.1 Patients

The cDNA generated from skin biopsies of different patients was kindly provided by Dr. Regine Gläser (Hautklinik Kiel). A total of 17 unrelated individuals were investigated with basal cell carcinoma (BCC), squamous cell carcinoma (SCC), precancerous lesions or psoriasis vulgaris from the north of Germany (Table 2.1).

Table 2.1: Characteristics of patients

Patient	Sex	Age	Localization	Diagnose
1b	W	90	lower leg	Basal Cell Carcinoma
14b	M	51	temporal	Basal Cell Carcinoma
43b	M	83	temporal	Basal Cell Carcinoma
56b	M	74	preauricular	Basal Cell Carcinoma
95b	M	69	cheek	Basal Cell Carcinoma
63b	M	85	cheek	Squamous cell carcinoma
73b	M	81	capillitium	Squamous cell carcinoma
86b	M	68	helix	Squamous cell carcinoma
109b	M	86	shoulder	Squamous cell carcinoma
130b	M	88	helix	Squamous cell carcinoma
28b	M	94	mast	Precancerous Lesions
91b	M	78	shoulder	Precancerous Lesions
111b	M	81	retroauricular	Precancerous Lesions

Each tumor specimen or psoriasis lesions was divided into matched fractions for RNA isolation. All subjects were informed about the contents and aims of the study and gave

their written consent. All human materials were collected in compliance with the laws and guidelines of the ethics committee of the medical faculty of Christian-Albrechts University Kiel, Germany.

## 2.2 Materials

The cell line HaCat was obtained from the Deutsches Krebsforschungszentrum (Heidelberg, Germany). 6-well and 12-well culture Plates were obtained from Becton Dickinson (Heidelberg, Germany). Cell culture flasks (200 ml; 75 cm<sup>3</sup>) were obtained from Sarstedt (Nümbrecht). DMEM (Dulbecco's modified Eagle Medium) and FCS (fetal calf serum) were obtained from Cell Concepts (Umkirch, Germany). Dulbecco's PBS was obtained from PAA Laboratories GmbH (Cölbe, Germany). Blue-Dextran and dNTPs were obtained from Pharmacia (Freiburg, Germany). Agarose was obtained from Biozym Hess (Oldendorf, Germany). Acetonitril, H<sub>2</sub>O (for HPLC) and methanol was obtained from Pro-mochem (Wesel, Germany). LightCycler-FastStart DNA Master SYBR Green I and Stericup Filtrations system was obtained from Roche (Mannheim, Germany). Glycogen proteinase K was obtained from Boehringer (Mannheim, Germany). Millipore YM 3 and YM 30 (3 kDa and 30 kDa) were obtained from Millipore GmbH (Eschborn, Germany). Glycerol was obtained from USB (Cleveland). Sojamehl-Pepton was obtained from Unipath LTD. (Hampshire). Superscript Reverse Transcriptase was obtained from Invitrogen. Nitrocellulose and blotting paper were purchased from Schleicher and Schuell (Dassel, Germany). Restriction endonucleases were from New England Biolabs. The following materials were obtained from Gibco (Eggenstein, Germany): Agarose (ultrapure), Bacto-Trypton, Bacto-Yeast-extract, EDTA, L-Glutamin, Penicillin-Streptomycin-solution, Trypsin and Trypsin-EDTA. The following materials were obtained from Sigma (Deisenhofen, Germany): Agar, Bovines Serum Albumin (BSA), Bromphenolblue, Chloroform, Formamid, DEPC, DMSO, DTT, EpiLife, Ethidiumbromid, Gelatine, Glutaraldehyde, Isopropanol, L-Tryptophan, Magnesiumchlorid (MgCl<sub>2</sub>),  $\beta$ -Mercaptoethanol, Tween, Xylencyanol. The carrier protein keyhole limpet hemocyanin (KLH) was purchased from Pierce Chemical Company. All

other chemicals unmentioned here were of analytical grade and purchased from Merck (Darmstadt, Germany).

### 2.3 Tissues and RNA preparation

Total RNAs were isolated from normal human skin, larynx, pharynx, adenoid, polyp, tonsil, tongue, psoriasis lesions and cultured keratinocytes by using TRIzol reagent (Invitrogen). All other RNAs were purchased from Clontech. The samples were stored at  $-80^{\circ}\text{C}$  until use.

### 2.4 Cell Culture

Primary human foreskin keratinocytes were prepared from a pool of neonatal foreskins obtained from routine surgery using the Rheinwald-Green method (Rheinwald and Green, 1975). Keratinocytes were isolated by using dispase incubation of foreskin tissue to allow for dermalepidermal separation. Epidermal specimens were trypsinized and plated on standard tissue culture dishes. Cells were maintained in EpiLife media (Sigma) supplemented with 10% fetal calf serum containing 10 U/ $\mu\text{l}$  penicillin and 0.1  $\mu\text{g}/\mu\text{l}$  streptomycin (Life Technologies). The second passage cells were maintained and utilized when 75% confluent. Cells were separated to 6-well plates and then stimulated with 1.7 mM  $\text{Ca}^{2+}$  or  $10^{-7}$  M retinoic acid (RA) (Sigma) for the indicated time periods. In some cases cells were treated with 0.5  $\mu\text{g}/\mu\text{l}$  of PD98059, a potent inhibitor of the MAPK kinase (MEK1/2) (Favata et al., 1998), 0.5  $\mu\text{g}/\mu\text{l}$  of SB203580 or SB202190, an inhibitor of the p38 kinase (Cuenda et al., 1995), SB202474, a p38 inactive analog or/and 2.5  $\mu\text{g}/\mu\text{l}$  of SP600125, an inhibitor of the JNK kinase obtained from Calbiochem in 50% DMSO. Each experiment was performed at least in twice.

## 2.5 Conventional RT-PCR

The cDNA was synthesized from 2  $\mu$ g of total RNA with the SuperScript II system (Invitrogen). A total of 2  $\mu$ g of total RNA was treated with RNase-free Dnase (Roche Diagnostics, Mannheim, Germany) and reverse transcribed with an oligo(dT)<sub>18</sub> primer and Superscript II RNaseH<sup>-</sup> reverse transcriptase (Invitrogen) in a 20  $\mu$ l reaction volume under the following conditions: 60 min at 42°C in 50 mM Tris/HCl, pH 8.3, containing 75 mM KCl, 3 mM MgCl<sub>2</sub>, 10 mM dithiothreitol and 0.5 mM dNTPs, followed by enzyme inactivation for 15 min at 70°C. The reaction mixture was diluted 10 times with H<sub>2</sub>O for PCR experiments. All gene-specific PCR primers (Table 2.2) were designed based on PRIMER3.0 (<http://frodo.wi.mit.edu>) to amplify products in the range 150 to 370 bp and to span the first exon-intron boundary. The conditions for PCR were optimized in the Robocycler gradient 96 (Stratagene) and subsequently used for the following conventional PCR and Real-time quantitative RT-PCR. The GAPDH sequence and a cloned corresponding cDNA sequence were used for normalization and to calibrate standard curves, respectively. Conventional PCR was carried out with Advantage 2 Polymerase Mix (Clontech) using a 'touch-down' PCR strategy.

For THHL1, THH, PFLG, IFPS and C1orf10, the thermal cycling conditions comprised: firstly 1 min at 95°C followed by 5 cycles at 95°C for 20 s, 68°C for 45 s; 5 cycles of 95°C for 20 s, 66°C for 20 s, 68°C for 45 s; and finally 32 cycles of 95°C for 20 s, 64°C for 30 s, 68°C for 45 s.

For RPTN, the thermal cycling conditions comprised: firstly 1 min at 95°C followed by 5 cycles at 95°C for 20 s, 72°C for 45 s; 5 cycles of 95°C for 20 s, 70°C for 45 s; and finally 32 cycles of 95°C for 20 s, 68°C for 45 s.

For HRNR, SPINK6, SPINK7 and SPINK8, the thermal cycling conditions comprised: firstly 1 min at 95°C followed by 5 cycles at 95°C for 20 s, 68°C for 45 s; 5 cycles of 95°C for 20 s, 66°C for 20 s, 68°C for 45 s; and finally 32 cycles of 95°C for 20 s, 64°C for 45 s, 68°C for 45 s.

For SPINK9, the thermal cycling conditions comprised: firstly 1 min at 95°C followed

by 5 cycles at 95°C for 20 s, 72°C for 45 s; 35 cycles of 95°C for 20 s, 70°C for 45 s; and finally 72°C for 10 min.

Specificity of the resulting PCR products was verified by purificating the amplified fragments with the QiaQuik spin PCR purification kit (Qiagen), cloning into pGEM-T vector (Promega) and subsequent sequencing.

Table 2.2: Primer sequences used for RT-PCR and quantitative real-time RT-PCR

Gene	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
THHL1	GCTGCCCCACGTTCTACCT	CATGAAGGACACAGGGCTGA
THH	TGGTGAAGTGGGTTTACTTGAA	GAAGCACAGCTCCAAATTCC
RPTN	CAAGCAGGTTACCCGTACTTG	CATTTGGTCTCTGGAGGATGTCTC
HRNR	CTCTGGTGAGCTAGGTTACTCA	TGCAAGATGATATCCACAGTATCTG
PFLG	GGCTCCTTCAGGCTACATTCTA	ATCTGGATTCTTCAGGATTTCG
IFPS	ACCAGGGTTCACCTAAACTTGCA	ATGACATCCACTGTGTCTGGATC
C1orf10	GTTGACTTCAAAGATGCCTCAGT	GTGGGGTTTCACAATCACATC
SPINK6	ACCTCAGCTGGACAAAGCAG	TGGCAAGTCACCAAGAAACA
SPINK7	TGCTTGGCTAGCTCTGCATT	AAAGCGCAAGAGGGAGATTG
SPINK8	CATATTCAGCCACCTCATCCAG	AAGCAAGTGTAGTTACTCTGGTACCC
SPINK9	ACTTGCAACCATGTTCAAGTATAGA	AACTTTAGAACAGAAGAAGCAATCA
GAPDH	CCAGCCGAGCCACATCGCTC	ATGAGCCCCAGCCTTCTCCAT

Table 2.3: PCR conditions for each primer pair shown in Table 2.2

Gene	cDNA amplicon(bp)	Annealing Temp(°C)	Detection Temp(°C)
THHL1	212	63	76
THH	178	63	78
RPTN	175	67	80
HRNR	215	61	79
PFLG	211	63	76
IFPS	211	64	80
C1orf10	160	63	84
SPINK6	322	63	
SPINK7	341	63	
SPINK8	367	63	
SPINK9	175	62	78
GAPDH	360	66	

## 2.6 Real-time RT-PCR

Quantitative real-time PCR was performed by using LightCycler (Roche) with the SYBR Green I detection system (Roche) under the indicated optimized conditions in Table 2.3.

For each gene, a primer pair flanking the first exon-intron boundary within each gene was constructed and synthesized (Table 2.2). The cDNA equivalent to 10 ng of RNA served as template for PCR in a volume of 10  $\mu$ l. Each PCR mixture was prepared as follows: 6.4  $\mu$ l of water, 1.2  $\mu$ l of MgCl<sub>2</sub> (4 mM), 0.2  $\mu$ l of each primer (4  $\mu$ M), 1.0  $\mu$ l of FastStart Master SYBR Green I mixture. Samples were amplified in capillaries after initial denaturation at 95°C for 10 min for 45 cycles: 10 s of denaturation at 95°C, 5 s of annealing at initial 68°C ('touchdown' of -1°C/cycle to 62°C) and 10 s of extension at 72°C. SYBR Green I fluorescence was measured at 72°C at the end of each cycle.

Melting curves were generated after each run to confirm amplification of specific transcripts (cooling of the samples at 65°C for 15 s, heating at 0.2°C/s to 95°C with continuous measurement of the fluorescence). The specificity of each gene products was verified by sequencing. The housekeeping gene GAPDH (glyceraldehyde phosphodehydrogenase) was amplified with each cDNA in a separate PCR to allow calculation of the relative transcripts. The plasmid DNA with the corresponding insert was linearized with restriction digestion and used for construction of standard curves. Exact quantification of the DNA content was done on a GeneQuant II RNA/DNA calculator (Pharmacia Biotech). For standard curve acquisition, eight serial dilutions of linearized plasmid DNA were prepared, representing  $30 - 3.0 \times 10^8$  double-stranded DNA molecules.

## 2.7 *In-Silico* cloning of the novel human SFTP and SPINK genes

All the known SFTP protein sequences in human, mouse and rat were retrieved from the NCBI database ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), which included the five known SFTP protein sequences, namely human profilaggrin (GenBank accession No. A45135), trichohyalin (GenBank accession No. CAH70024), tumor-related protein (GenBank accession No. BC030807.1), mouse repetin (GenBank accession No. NP\_033126) and hornerin (GenBank accession No. NP\_598459). All of the selected sequences were aligned by the ClustalW algorithm on the EMBL server ([www.ebi.ac.uk/emb/](http://www.ebi.ac.uk/emb/)). The consensus protein sequences were extracted and used as *in silico* probes to scan the human



genome on the Ensemble Genome Browser ([www.ensembl.org/Homo\\_sapiens/](http://www.ensembl.org/Homo_sapiens/)). The exon-intron boundaries of SFTP-like candidate genes were identified using FuzzyFinder (<http://www.cse.ucsc.edu/~kent/fuzzyFind/fuzzyFind.html>). All the homolog SFTP sequences were used to recover the original ORFs for partial sequences of possible SFTP genes. After duplicates were removed, all the different sequences were mapped to genomic locations using the BLAT algorithm (<http://genome.ucsc.edu/cgi-bin/hgBlat>). Likewise, the kazal-domain amino acid sequence was used as a query to screen the SPINK family members.

The mouse and rat ortholog of human SFTPs and SPINKs were identified by tBLASTN searches via the BLAT algorithm with the human SFTP/SPINK protein sequences against the mouse and rat genome, respectively. The data obtained were contrasted with gene predictions from the mouse and rat Genome Initiative. Additional mouse and rat SFTP/SPINK homologies were identified by using the predicted mouse/rat SFTP/SPINK cDNA as *in silico* probe to scan the mouse/rat chromosome on the Ensemble Genome Browser. The data obtained in the mouse and rat chromosome 18 were contrasted with gene positional identity and exon/intron conservation to human SPINKs.

The signal anchor predictions were done using SignalP version 2.0 (<http://www.cbs.dtu.dk>); potential protein coding regions were identified with SMART (<http://www.smart.ox.ac.uk>); potential post-translational modifications were determined using NetPhos2.0 (<http://www.cbs.dtu.dk>) and NetNGlyc1.0 (<http://www.cbs.dtu.dk>); determination of the modular structure of the proteins was performed using the PROFILES-CAN Server from the Swiss Institute for Bioinformatics, the SMART (Simple Modular Architecture Research Tool) site and the Conserved Domain Database at NCBI. Alignment of multiple predicted protein sequences was done by using ClustalW. The resulting profile alignments were used to construct a phylogenetic tree with ClustalW (<http://bioweb.pasteur.fr/seqanal/interfaces/clustalw.html#trees>) which was visualized with the cladogram tree printer (<http://bioweb.pasteur.fr/seqanal/interfaces/drawgram-simple.html>) and edited as a postscript file.

## 2.8 5' and 3' Rapid Amplification of cDNA Ends (RACE)

Total RNA was extracted from human foreskin-derived keratinocytes using TRIzol reagent. After DNase I (Roche) treatment to exclude contamination with genomic DNA, 3  $\mu$ g of DNA-free total RNA was then used for first-strand cDNA synthesis for 5'RACE and 3'RACE respectively with SMART RACE cDNA Amplification Kit (Clontech) according to the manufacturer's protocol. The template RNA was digested with RNase H (Roche) and the reaction mixture was diluted 10 times.

Table 2.4: Primer sequences used for 5'RACE

Primer	Primer sequence
5'-RACE CDS Primer:	5'-(T) <sub>25</sub> N <sub>-1</sub> N-3'
THHL1-5'Race	5'-CTGTCCACTGACCATCTCCGGTGGTTG-3'
THHL1-5'RaceNest	5'-GTGGTTGCCTGAACATCCACATCATCTAGC-3'
THH-5'RACE	5'-AATGCTGCTCCATCACAATCATGAGAGAC-3'
THH-5'RaceNest	5'-CACAGCTCCAAATTCCTTTCAAGGAGG-3'
RPTN-5'RACE	5'-GCAGGCTTGGACCAACTGGAACACCAAC-3'
RPTN-5'RaceNe	5'-AGCCGTGTTGGCCATAGTTGGGAGACTG-3'
HNR-5'RaceNest	5'-GACTGATGGGAGTCGGAGTTTTGCTCAC-3'
IFPS-5'RACE	5'-TGGTGTCGGTGACCACGCCTATGCTTC-3'
IFPS-5'RaceNest	5'-GCTGAGGACCTTGTTGCAGGCCATAGTCA-3'
SPINK6-5'Race	5'-AGGCACATTATTGCCATATGTCTGGCCATC-3'
SPINK6-5'RaceNest	5'-GCCACAGTGTGGGTTAGATTCCCGAGTG-3'
SPINK6-3'RACE	5'-GCCAAACAGACGAAACAGATGGTTGACT-3'
SPINK6-3'RaceNest	5'-CACAGCTCCAAATTCCTTTCAAGGAGG-3'
SPINK6-5'Race	5'-TGCCATCAGATCCACAAATTGGATCATA-3'
SPINK6-5'RaceNest	5'-GCAGTCAACCATCTGTTTCGTCTGTTTGG-3'
SPINK7-5'RACE	5'-CTCAACACACAGAATGCAGGGATTATCATAGG-3'
SPINK7-5'RaceNest	5'-TAGGTTATAAAATTGGTGCCGACATGGGTTG-3'
SPINK8-5'Race	5'-ACACTCATTCTGGAAAGTGTGGCCATTTGAG-3'
SPINK8-5'RaceNest	5'-GGTGTCTGCACATTGGGGCAGTCTG-3'
SPINK9-5'Race	5'-TGCCATCAGATCCACAAATTGGATCATA-3'
SPINK9-5'RaceNest	5'-GCAGTCAACCATCTGTTTCGTCTGTTTGG-3'
SPINK10-5'RACE	5'-CATTCCATCAGATCCACAGTGAAGTGTGTAG-3'
SPINK10-5'RaceNest	5'-ATGGTACATGCCTCAGAATACCTGGTACTG-3'

5' RACE were performed essentially according to the Clontech SMART RACE cDNA Amplification protocol. Reactions were set up as follows: 0.5 ng of adapted cDNA, 0.2  $\mu$ M 10  $\times$  UPM primer, 0.2  $\mu$ M gene-specific primer, 5  $\mu$ l of 10  $\times$  reaction buffer (supplied with Advantage polymerase), 0.2 mM dNTP, 1  $\mu$ l of BD Advantage 2 Polymerase mix, in a final volume of 50  $\mu$ l. RACE PCR was performed on MJ Research Peltier Thermal cycler PTC-200 and reaction cycles were as follows: 95°C for 1 min; 5 cycles of 95°C for 30 s, 72°C for 3 min; 5 cycles of 95°C for 30 s, 70°C for 3 min; 25 cycles of 95°C for 30 s, 68°C for 3 min;

and a final extension of 10 min at 68 °C. Then, 0.5 µl of PCR products was employed as template for a second round PCR (Nested PCR) under the following conditions: 1 min at 95 °C, 30 cycles of 20 s at 95 °C and 3 min at 70 °C, and a final extension of 10 min at 70 °C. The gene-specific primers were generated with the predicted exons nucleotide sequences above and listed in Table 2.4.

Table 2.5: Primer sequences used for 3'RACE

Primer	Primer sequence
3' RACE CDS Primer	5'-AAGCAGTGGTATCAACGCAGAGTAC
THHL1-3'Race	5'-GATGCCTCAGCTCCTGAGAAATGTCCTCTG-3'
THHL1-3'RaceNest1:	5'-GCATCAGGCCAGAGCTTTGCACTG-3'
THHL1-3'RaceNest2:	5'-CACAAACATATAACAGGGGACTTCCACTTG-3'
RPTN-3'Race	5'-CCTCCTCAAGCAGGTTACCCGTA-3'
RPTN-3'RaceNest1	5'-GGACACAGGACAGGCGAACCCATAA-3'
RPTN-3'RaceNest2	5'-AAGGGCTTTTGGTTCTTTGAGCAGCA-3'
RPTN-3'RaceNest3	5'-CCACTCCTCAAGGAACCTAGATTTGGTG-3'
HRNR-3'Race	5'-AGTCTGGATCGAGACCATAACAAGAAAGTGGGA-3'
HRNR-3'RaceNest1	5'-GATGACTCACCAGCACCAAGAGGAAC-3'
HRNR-3'RaceNest2	5'-AGCATGAGTCTGCCTCCCGTCACTCTT-3'
HRNR-3'RaceNest3	5'-ATGGAAGTGGCTCAGGGCAAGATGGGTA-3'
HRNR-3'RaceNest4	5'-AACCATGATGGGGGAAGTCTGGCTCA-3'
IFPS-3'Race	5'-ACCAGGGTCACTTAAACTTGCA-3'
IFPS-3'RaceNest1	5'-GATGTCATCATGCATATGCTGGATCGAG-3'
IFPS-3'RaceNest2	5'-AGGCTTGAGTCAGTCCCTGCGGTTCCG-3'
IFPS-3'RaceNest3	5'-TGGGCCAGGGTGAATCTCAACAAGTAGAG-3'
IFPS-3'RaceNest4	5'-GTTTCGAGACCAGCCTGGCCAACATA-3'
IFPS-3'RaceNest5	5'-TCAGGAGGCTGAGGCAGGAGAATTG-3'
SPINK6-3'RACE	5'-GTGAGTTCCAGGACCCCAAGGTCTACTG-3'
SPINK6-3'RaceNest	5'-CCACACTGTGGCTCTGATGGCCAGA-3'
SPINK7-3'RACE	5'-GGTCAACCCCTGCCCTGGCTTATATCAAC-3'
SPINK7-3'RaceNest	5'-CAACCCATCTGCGGCACCAATTTTATAACCTA-3'
SPINK8-3'RACE	5'-CGATGTAAAATGTATATCCCACTGGACCCTGA-3'
SPINK8-3'RaceNest	5'-ATGCAGACTGCCCAATGTGACAGCA-3'
SPINK9-3'RACE	5'-GCCAAACAGACGAAACAGATGGTTGACT-3'
SPINK9-3'RaceNest	5'-ACCACCAGGACAACAGAGATTTTGTATC-3'
SPINK10-3'RACE	5'-ATTCAGTGTACCAGGTATTCTGAGGCATGTAC-3'
SPINK10-3'RaceNest	5'-TGGACTACACTTCACTGTGGATCTGATGGA-3'

The first strand cDNA (3'-ready cDNA) synthesis for 3' RACE was performed according to the manual of the SMART RACE cDNA Amplification Kit (Clontech) using the 3'-RACE CDS Primer provided by the kit. The template RNA was digested with RNase H (Roche) and the reaction mixture was diluted 10 times. Based on the sequence of the 5' RACE product, the sense gene specific primer was designed and synthesized (Table 2.5). The 3' RACE PCR was carried out using primer and UPM under the following condition: 0.5 ng of adapted cDNA, 0.2 µM 10 × UPM primer, 0.2 µM gene-specific primer, 5 µl of 10 × reaction buffer (supplied with Advantage polymerase), 0.2 mM dNTP, 1 µl of BD

Advantage 2 Polymerase mix, in a final volume of 50  $\mu$ l. Reaction cycles were as follows: 95°C for 1 min; 5 cycles of 95°C for 30 s, 72°C for 3 min; 5 cycles of 95°C for 30 s, 70°C for 3 min; 25 cycles of 95°C for 30 s, 68°C for 3 min; and a final extension of 10 min at 68°C. Then, 0.5  $\mu$ l of PCR products was employed as template for a second round PCR (Nested PCR) under the following conditions: 1 min 95 °C, 30 cycles of 20 s 95°C and 3 min 70 °C, and a final extension of 10 min at 70 °C. The PCR product was purified and cloned into the pGEM-T cloning vector (Promega) followed by sequencing.

## 2.9 Long-distance PCR and the short-range PCR-walking strategy

BD Advantage 2 Polymerase mix was used for long-distance PCR of THHL1, SPINK6, SPINK7-v2, SPINK8 and SPINK9. To determine the full-length sequences of human repetin, ifapsoriasin and hornerin cDNA, an alternative approach was used, namely the short-range PCR-walking strategy. Briefly, based on the newly defined 5'- and 3'-terminal sequence, the middle part of each mRNA sequence was then predicted by *in silico* analysis and used for primer design. Each primer pair (see appendix) was expected to produce about 1 kb length of overlapping cDNA.

## 2.10 Subcloning and expression

### 2.10.1 Preparation of expression constructs

Six recombinant proteins were expressed in this study, which included human LEKTI-2, LEKTI-3 and four hornerin fragments (HRNR1, residues 49–98; HRNR2, residues 1075–1172; HRNR3, residues 2591–2662; and HRNR4, residues 2727–2850). The specific primer pairs were listed in Table 2.6.

The regions of the corresponding cDNAs longer than that of the inserts coding for the mature LEKTI-2, LEKTI3 and four hornerin fragments were synthesized by RT-PCR from foreskin-derived keratinocytes RNA, as described. PCR was performed with *Pfu* DNA polymerase (Promega). Except where indicated, PCR amplifications were performed in

50  $\mu$ l reaction volumes in the presence of the appropriate Tris-based buffer, using 2.5 U polymerase, 100 mM each dNTP, 100 ng each primer and 50 ng cDNA template. The PCR mixtures were denatured by heating at 98°C for 30 s. Thirty cycles of amplification were performed using the following conditions: 30 s at 98°C; 45 s at Primer  $T_m - 5^\circ\text{C}$ ; 2 min at 72°C.

For generating the inserts, a second nested PCR was performed with the specific primers pairs containing the *Bgl* II and *Not* I digestion sites under the following conditions: 45 s at 98°C; 5 cycles (45 s at 98°C; 45 s at 55°C; 1 min at 72°C). 25 cycles (45 s at 98°C; 1 min at 72°C). The PCR products were gel purified, ligated into pGEM-T vector and verified by sequencing. The inserts were cut with *Bgl* II and *Not* I, gel purified and cloned into the expression vector pET-32a (Novagen) that had been double-digested with *Bgl* II and *Not* I. The *E. coli* host strain BL21

### 2.10.2 Induction and expression of the polyhistidine-tagged fusion protein

The expression construct was transformed in *E. coli* BL21\beta-D-galactoside (IPTG) for 3 h at 37°C. After incubation, cultures were harvested by centrifugation at 5000 $\times$ g for 10 min at 4°C and resuspended in 8 ml of 1 $\times$ LEW buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, pH 8.0). Resuspended cells were subjected to one cycle of freeze-thawing and disrupted by sonication in an ice bath until the required visual viscosity was obtained. The separation of the cytoplasmic soluble

Table 2.6: Primer sequences used for protein expression

PCR product	Primer	Primer sequence (5' to 3')
LEKTI-2	sense	ACGGACACCAGGTCACCTCTTTCCCTACATC
	antisense	TCTACATATGGTGATGAGTAGGCAATGTGGCA
lekti3	sense	ACCTCAGCTGGACAAAGCAG
	antisense	TGGCAAGTCACCAAGAAACA
hrnr-1	sense	CCTACAAGGCGTCACACTGTTCATCG
	antisense	TCAGTCCCAGGTTTAAGACTTCCTCTGACG
hrnr-2	sense	CTCTACCCATGGGCAACACGGTCT
	antisense	ATGCTGAGTGCAACCAGAGGACTGC
hrnr-3	sense	CGGCCAGCGTGGGTCTGGGTCA
	antisense	TGTAGAACCATGCTGTCCTTGGCTACAGAAG
hrnr-4	sense	AACCTTTGACCAGGAGGGATCTAGCA
	antisense	CAGTCTTTGGAAGGAACCCATAACA
pET-LEKTI-2	sense	ACTGAGATCTGGGTACCGACGACGACGACAAGACGAAACAGATGGTTGACTGC
	antisense	ATTTGCGGCCGCTCAACTTTTCCAAAATGTACAAA
pET-lekti3	sense	ACTGAGATCTGGGTACCGACGACGACGACAAGCAGGTTGACTGTGGTGAGTTC
	antisense	ATTTGCGGCCGCTCAGCATTTTCCAGGATGCTT
pET-hrnr-1	sense	ACTGAGATCTGGGTACCGACGACGACGACAAGGATGACACTCACCAGCACCA
	antisense	ATTTGCGGCCGCTCAAGACTGCCCTGACCCAGAC
pET-hrnr-2	sense	ACTGAGATCTGGGTACCGACGACGACGACAAGACCCATGGGCAACACGGTTC
	antisense	ATTTGCGGCCGCTCAAGACTCGTGGTGACCAAAGC
pET-hrnr-3	sense	ACTGAGATCTGGGTACCGACGACGACGACAAGAGTCAGCATGGGTCTGGCT
	antisense	ATTTGCGGCCGCTCACAACCGGACCCATGTCGG
pET-hrnr-4	sense	ACTGAGATCTGGGTACCGACGACGACGACAAGTCAGGACAGTCATCAACCTTTG
	antisense	ATTTGCGGCCGCTCACTGATAAAAAGTAGCACCTCTGC

fraction (CSF) from the cytoplasmic insoluble fraction and cell debris was performed by centrifugation at 15000×g for 30 min at 4°C. The CSF was filtered (0.45  $\mu$ m-pore-size filter; Schleicher & Schuell GmbH) and the clarified supernatant was used for purification of the polyhistidine-tagged protein with Protino<sup>®</sup>Ni prepared columns (Macherey-Nagel). The fusion proteins were eluted with 1×Elution buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 250 mM imidazole, pH 8.0). The eluate pH was adjusted to 3.0 to 4.0 with 2% (v/v) trifluoroacetic acid (TFA) and the resulted mixture was purified by centrifugation at 8000×g for 3 min. The clarified supernatant was subjected to preparative wide-pore C8 RP-HPLC (RP8) with a column (SP250/10 Nucleosil 300-7 C8; Macherey-Nagel) that was previously equilibrated with 0.1% (v/v) TFA in HPLC-grade water containing 10% acetonitrile. Fusion proteins were eluted with a gradient of increasing concentrations of acetonitrile containing 0.1% (v/v) TFA (flow rate, 3 ml/min). Fractions containing uv(215 nm)-absorbing material were collected, lyophilized and analyzed by ESI-QTOF-mass spectrometry.

### 2.10.3 Digestion and purification of the recombinant protein

The His-tagged fusion proteins purified from HPLC-RP8 were cleaved from their His-tag with enterokinase EKMax according to the manufacturer's suggestion (Invitrogen). The digestion mixture contained 1 unit of EKMax per 250  $\mu\text{g}$  of the fusion protein in a volume of 500  $\mu\text{l}$  and was dialyzed for 3 h (for hornerin peptides) or 16 h (for LEKTI-2 and lekti3) at 37°C. The dialyzed sample was centrifuged at 8000 $\times$ g for 3 min at 4°C. The separation of the target peptide and His-tag from the fusion protein was achieved by a reverse-phase high-performance liquid chromatography (RP-HPLC). Briefly, the supernatant was collected and injected onto a Jupiter-5 $\mu$ -C4-300A HPLC column (Phenomenex) equilibrated with 0.1% TFA in water. Peptides were eluted with a gradient of increasing concentrations of acetonitrile containing 0.1% (v/v) TFA (flow rate, 0.5 ml/min) (see figure legends for details of the gradients used). Fractions of each peak were collected. The concentrations of proteins present in HPLC fractions were estimated by ultraviolet absorbance integration at 215 nm using ubiquitin (Sigma) for calibration.

### 2.10.4 Protein analysis

Proteins were separated by 10% NuPAGE<sup>®</sup>Novex<sup>®</sup> Bis-Tris gel electrophoresis with MES SDS buffer (Invitrogen). Proteins in Bis-Tris gels were stained with AgNO<sub>3</sub> (Sigma) (Blum et al., 1987) or Coomassie blue R-250 (Sigma). The SeeBlue<sup>®</sup>Plus2 Pre-stained Standard marker (Invitrogen) was used as a molecular mass marker (49, 38, 28, 17, 14, 6 and 3 kDa). The polypeptide purity and molecular masses were assessed with ESI tandem MS on a Q-TOF2 (Micromass, Manchester, U.K.) mass spectrometer.

## 2.11 Generation of goat antibodies

A total of 1.0 mg of protein mixture including His-tagged fusion protein and purified peptide was conjugated by the glutaraldehyde method (Avrameas and Ternynck, 1969; Briand et al., 1985) to maleimide-activated keyhole limpet hemocyanin (KLH) (protein-KLH 1 : 1, wt/wt) and subsequently mixed with 1.0 mg of His-tagged protein for use as immunogens.

Immunization of a goat was carried out three times at 2-week intervals. Thirteen days after the final immunization, blood from each goat was collected and incubated at 37°C for 1 h to enhance clot formation and then incubated at 4°C overnight to encourage clot retraction. The serum was separated from the clot and stored at -70°C until required.

## 2.12 Immunoblot analysis

The proteins separated by RP-HPLC were blotted to a nitrocellulose membrane (0.2- pore size; Bio-Rad) with amounts of antigen and antiserum dilutions as indicated in Fig. 3.29 and Fig. 3.30. After blotting, the membrane was saturated at room temperature for 1 h with PBS/M [1×PBS containing 5% (wt/vol) freeze-dried low-fat milk] and then washed three times with PBS/T [1×PBS containing 0.05% (vol/vol) Tween]. The membrane was incubated at room temperature for 1 h with a polyclonal antiserum diluted 1 : 1000 in PBS/T/M (1×PBS containing 0.05% Tween and 5% freeze-dried low-fat milk). After three washes with PBS/T, the membrane was incubated at room temperature for 1 h with goat anti-rabbit immunoglobulin G (heavy plus light chains)-horseradish peroxidase conjugate (Bio-Rad) diluted 1 : 30000 in PBS/T/M. The membrane was washed three times with PBS/T and twice with PBS. The substrate (SuperSignal West Dura extended duration substrate; Pierce) was deposited for 5 min onto the membrane. The chemiluminescence was detected by a CD-camera (Diana, Raytest).

## 2.13 Mass spectrometry

Electrospray-ionization mass spectrometry analyses were done in the positive ionization mode with a quadrupole orthogonal accelerating time-of-flight mass spectrometer (QTOF-II hybrid mass spectrometer; Micromass).



## 2.14 Enzymatic Assays

For analysis of self-digestion, aliquots containing 20  $\mu\text{g}$  of HRNR3 from RP4-HPLC were lyophilized and diluted with 100 mM ammonium bicarbonate, pH 8.9. The solutions were incubated at 37°C for 0 h, 6 h and 16 h, respectively and subject to mass spectrometry.

## 2.15 Antimicrobial activity

The antimicrobial activities of of hrnr-2, hrnr-3, hrnr-4 and LEKTI-2 were assayed against the Gram-positive bacteria *Staphylococcus aureus* ATCC 6538 and *Enterococcus faecalis* ATCC 29212, the Gram-negative *Escherichia coli* ATCC 11303 and *Pseudomonas aeruginosa* ATCC 10145, and the yeast *Candida albicans* ATCC 24433.

To measure the microbicidal activity, the purified peptide was estimated with a microdilution assay system (Steinberg and Lehrer, 1997). Test organisms were incubated for 3 h at 37°C with purified LEKTI-2 or hornerin fragments (freshly lyophilized from HPLC purifications and stored as a pool in 10 mM sodium phosphate buffer at  $-80^{\circ}\text{C}$ ) in 10 mM sodium phosphate buffer, pH 7.4, containing 1% (v/v) trypticase soy broth (TSB). The antibiotic activity was analyzed by plating of serial dilutions of the incubation mixtures and determination of the number of CFU the next day. The limit of detection (1 colony per plate) was equal to  $1 \times 10^2$  CFU/ml.

# Chapter 3

## Results

---

In our previous work, two putative new antimicrobial peptides were isolated and purified from pooled stratum corneum extracts of healthy person's heel callus, S100 fused-type antimicrobial peptide (SFT-AMP) and kazal-type antimicrobial peptide (KT-AMP). Their N-terminal amino acid sequences were determined by the Edman degradation method, which are shown as following: SFT-AMP (SQHGSGSGHSSGYGQHGSR) and KT-AMP (TKQMVDCQHYYKKLPPGTTRFCHHMYYPPIIG).

### **3.1 The human clustered S100 fused-type protein (SFTP) gene family**

#### **3.1.1 Characterization of the full-length human hornerin cDNA**

To identify the cDNA sequence and the corresponding gene of the new antimicrobial peptide SFT-AMP, the N-terminal 19-amino-acid sequence of the natural peptide SFT-AMP was used as a query to blast against the human EST database and genome data by the tBLASTn algorithm on the NCBI server. This search led to a 100%-matched genomic locus located on the genomic contig NT\_029289 on chromosome 1q21.3. To generate a partial cDNA clone for this fragment, PCR amplification was performed using reverse-transcribed RNA from human foreskin-derived keratinocytes and specific oligonucleotides derived from the DNA sequence. Based on the resulted cDNA sequence from cloning

and sequencing of the PCR-amplified product, we designed specific oligonucleotides and performed 3'/5' RACE. However we failed to get any band from RACE.

To determine the full-length cDNA sequence for human SFT-AMP, comprehensive bioinformatic analysis was then used to help predict the gene structure for SFT-AMP. The SFT-AMP gene is located between trichohyalin (THH) and profilaggrin (PFLG), which is a locus for the epidermal differentiation complex (EDC) region, a highly conserved synteny to mouse chromosome 3F2 ([http://www.ensembl.org/Homo\\_sapiens/synteniview](http://www.ensembl.org/Homo_sapiens/synteniview)) (Fig. 4.2). According to the gene order on mouse EDC, the human SFT-AMP seemed to be a putative synteny to the mouse hornerin (GenBank Accession No. NM. 133698). The amino-acid sequence of the mouse hornerin was then used as a query to blast against human EST database and genome data by the tBLASTn algorithm to look for the candidate DNA sequence on the Ensemble Genome Browser. This method allowed us to find two exons encoding the potential human hornerin, one of which contained the locus encoding SFT-AMP peptide and many GC-rich repeats. Based on the predicted two-exon DNA sequence, a series of specific oligonucleotides were designed to be used for 3'/5' RACE and PCR walking. The cDNA sequence was completed by combining the overlapping sequences from each PCR product and submitted to GenBank with an accession No. AY396741 (Fig. 5.3). The full-length cDNA (9632 bp) contained an ORF consisting of 8553 bp, from nucleotides 80 to 8632, and a polyadenylation signal (AATAAA) situated 14 nucleotides 5' of the poly(A) tail. The deduced protein chain consisted of 2850 amino acids, containing a S100 domain (residues 4–46), a EF-hand domain (residues 51–80), a spacer sequence (residues 82–246), six tandem repeats A crosslinking with five tandem repeats B (Fig. 3.1). The mature peptide SFT-AMP is represented by the repeat unit A6 (residues 2591–2684, Fig. 3.1A).

### 3.1.2 Characterization of three novel SFTP members cDNAs

To identify candidate novel S100 fused-type protein (SFTP) members genes within the human genome, we used the tBLASTn algorithm and the 80-amino-acid sequences of the two conserved EF-hand domains of all the known SFTP members, such as human THH,

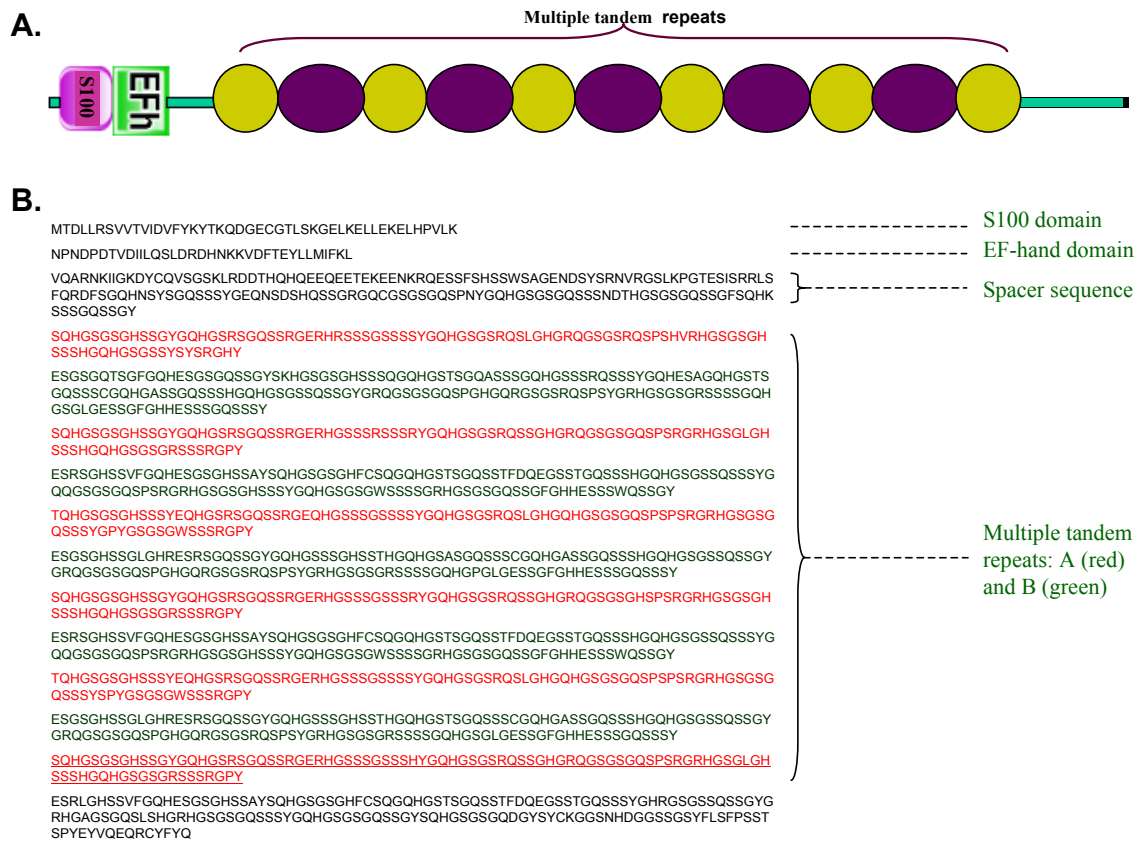


Figure 3.1: Schematic structure of human hornerin protein. A, the schematic structure of hornerin protein. The hornerin proprotein contains a S100 domain, an EF-hand domain, a spacer sequence and multiple tandem repeats; B, the full-length amino-acid sequence of hornerin and the localization of conserved motifs and repeats (repeat A marked red while repeat B green). The mature peptide SFT-AMP isolated was underlined red.

PFLG, hornerin (HRNR), C1orf10 and mouse repetin (RPTN), as a query against human genomic data (Build 33, available on April 30, 2003) to search for DNA loci with potential EF-hand domains. This search led to the identification of a contig NT\_032962 in the human chromosome 1q21.3 (Fig. 4.2). A total of seven putative homologous loci were clustered at this contig. Among them, three correspond respectively to known genes, *THH*, *PFLG* and *C1orf10*, one to *HRNR* while the other three are putative novel gene loci which were designated later as trichohyalin like 1 (*THHL1*), *RPTN*, and *ifapsoriasin* (IFPS). The exon 3 of the novel gene was predicted on the DNA sequence of the contig NT\_032962. By using the PCR-walking strategy and the SMART RACE procedure as described above, corresponding *THHL1*, *RPTN*, and *IFPS* overlapping cDNA clones were obtained and

sequenced.

**THHL1.** In both the 5' and 3' experiment, a single extended product was detected by using the specific primers based on the related S100/EF-hand DNA sequence. The full-length cDNA (3603 bp) of THHL1 was completed with the two overlapping cDNA sequences from sequencing results of RACE products, which contained an ORF consisting of 2715 bp and a poly(A) signal (AATAAA) situated 12 nucleotides 5' of the poly(A) tail (GenBank accession No. AY456639). The deduced protein chain consisted of 904 amino acids containing a S100 domain (residues 4–46), a EF-hand domain (residues 52–80) and an AAA-ATPase domain (residues 170–526).

**RPTN and IFPS.** In the 5' experiment, a single extended product was detected by using the specific primers of each gene. The PCR-walking strategy was used to confirm the middle cDNA sequence while the 3' SMART RACE was used to determine the 3'-end cDNA sequences. All the partial cDNA clones were obtained and sequenced. The full-length cDNA was completed with the overlapping cDNA sequences from sequencing results of both PCR-walking and RACE products. The full-length RPTN cDNA (3569 bp) contained an ORF consisting of 2355 bp and a poly(A) signal (AATAAA) situated 15 nucleotides 5' of the poly(A) tail (GenBank accession No. AY396742). The deduced RPTN protein chain consisted of 784 amino acids containing a S100 domain (residues 4–46), a EF-hand domain (residues 51–80), a spacer sequence (residues 81–268), thirteen tandem repeats with 24 amino-acid residues each and the C-terminal sequence (Fig. 3.2A and B). The full-length IFPS cDNA (9117 bp) contained an ORF consisting of 7176 bp and a poly(A) signal (AATAAA) situated 9 nucleotides 5' of the poly(A) tail (GenBank accession No. AY827490). The deduced IFPS protein chain consisted of 2391 amino acids containing a S100 domain (residues 4–46), a EF-hand domain (residues 51–80), two spacer sequences (residues 81–268) and two kinds of tandem repeats (Fig. 3.3).

### 3.1.3 Genomic structure of human SFTP genes

The genomic structure of human SFTP genes was analyzed by using the BLAT algorithm to compare each mRNA sequence with the human genome sequence assembled in May 2004



in the UCSC server. All introns were flanked by the consensus donor and acceptor splice sites. The putative TATA box was found on about 25 bp before the upstream genomic sequence of the transcription start site (TSS) site of each gene. All the SFTP genes consisted of three exons, exon 1 being untranslated and exon 2 containing the ATG start codon and encoding the S100 domain; exon 3 was the largest and contained the high G+C content of the underlying DNA and the TAA/TGA/TAG stop codon (Fig. 3.4).

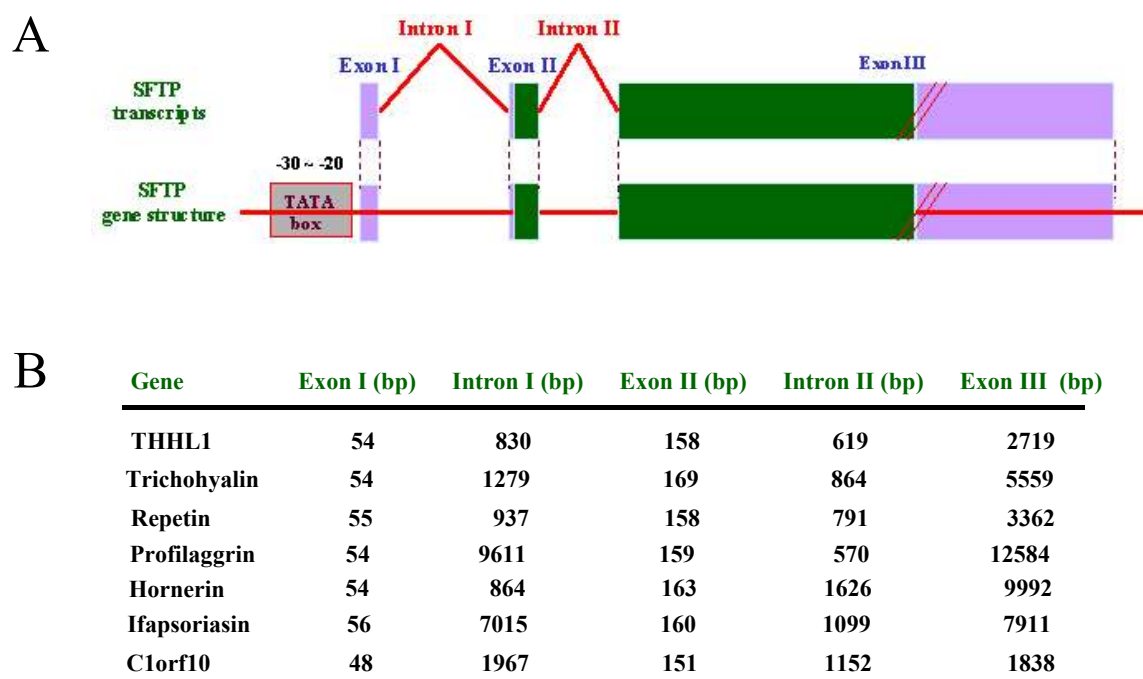


Figure 3.4: Comparison of genomic structure of the human SFTP genes. A, Schematic representation of the genomic and exon/intron consensus organization for the human SFTP genes. Boxes represent exons, green-colored regions indicate the coding sequence, and intervening lines denote introns. B, Exon/intron sizes and boundaries.

### 3.1.4 Tissue distribution of human SFTP expression

To determine tissue distribution of each human SFTP gene expression, we performed RT-PCR and real-time PCR analyses (Fig. 3.5 and Table. 3.1). To prevent the possibility of contamination from genomic DNA, RT samples were treated with DNase I (to destroy any remaining genomic DNA in our sample) before PCR amplification. To further exclude the possibility that our positive signal may result from genomic DNA amplification,

intron-spanning primer pairs specific to each gene were designed according to the cDNA sequences of exons 1 and 2. From all the 23 tissues analyzed, SFTP members were not only strongly expressed in keratinocytes but also in other tissues. *IFPS* mRNA was restricted to skin and small intestine and *THH* mRNA was restricted only to skin. *THHL1* also showed a narrow expression profile and its mRNA was detected in nasal epithelia, tongue and neutrophils and weakly expressed in skin. Other SFTP members (*HRNR*, *PFLG*, *RPTN* and *C1orf10*) showed a broad expression profile. More interestingly, *HRNR* mRNA was also detected in bone marrow and neutrophils.

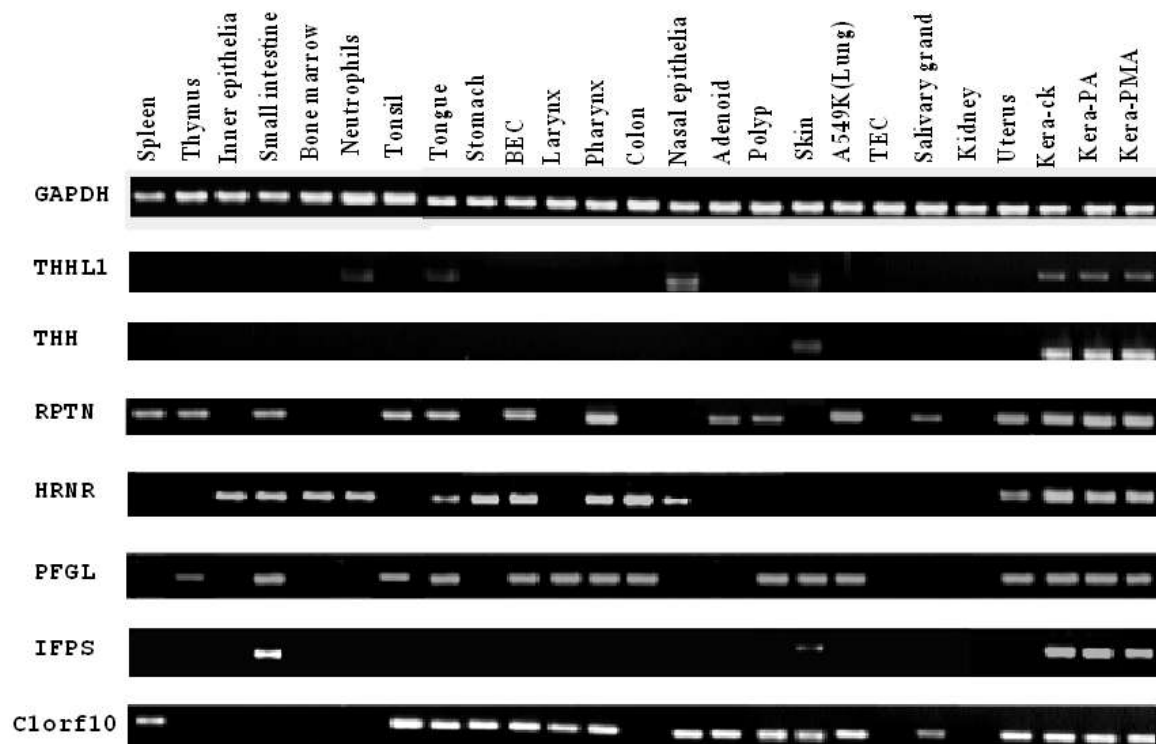


Figure 3.5: RT-PCR analysis of human SFTP expression. RNA was isolated from the indicated human tissues and cells. The RNA was treated with reverse transcriptase followed by PCR amplification with a SFTP-specific primer pair or a GAPDH-specific primer pair and the analysis of amplification products by agarose gel electrophoresis. The housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was employed as a loading control. BEC, bronchial epithelial cells; TEC, trachea epithelial cells; Kera-ck, untreated keratinocytes; Kera-PA, keratinocytes treated with *Pseudomonas aeruginosa*; Kera-PMA, keratinocytes treated with phorbol-myristate-acetate (PMA).



Table 3.1: Realtime RT-PCR analysis of human SFTP transcripts. A transcript copy number determined by real-time PCR using cDNA as a template. Results represent the average of two independent cDNA samples. In all cases, copy numbers were normalized against GAPDH transcripts and shown by 1000 times. BEC, bronchial epithelial cells; TEC, trachea epithelial cells.

Tissue	THHL1	THH	RPTN	HRNR	PFLG	IFPS	C1orf10
Spleen	0	0	0	0	0	0	0.278
Thymus	0	0	1.32	0	5.54	0	0
Inner epithelia	0	0	0	0	0	0	0
Small intestine	0	0	0.0325	0	0.110	0.0581	0
Bone marrow	0	0	0	0.0803	0	0	0
Neutrophils	20.3	0	0	0.0887	0	0	0
Tonsil	0	0	176	0	1.26	0	47.9
Tongue	990	0	139	0.0575	0.633	0	114
Stomach	0	0	0	0.067	0	0	1.93
BEC	6.66	0	157	0.0627	3.27	0	29.8
Larynx	0	0	0	0	0	0	0.0792
Colon	0	0	0	0.546	0	0	0
Nasal epithelia	72200	0	2.16	0.0791	0	0	3.44
Adenoid	0	0	14.0	0	0	0	0.478
Polyp	0	0	0	0	0	0	44.4
A549K	0	0	0	0	0.0091	0	0
TEC	0	0	14.3	0	0	0	0
Salivary gland	0	0	0	0	0	0	4.45
Kidney	0	0	0	0	0	0	0
Uterus	0	0	53.1	16.2	1.44	0	457
skin	3.09	37.2	59.3	0	32070	1672	5.12
Keratinocytes	0.329	2.27	7.81	43.1	441	6.27	10.7

### 3.1.5 Expression of human SFTP genes in skin disease tissues

To analyze expression of SFTPs in skin cancer tissues, realtime RT-PCR was performed. The expression levels of all SFTP family members were examined with paired tissue specimens (Table 3.2). No significant difference in SFTPs levels between skin precancerous lesions and paired normal tissues were detected, although the relative amounts of SFTPs levels differed among cases.

The expression levels of PFLG and IFPS were found to be decreased in SCC cases but were increased in BCC cases. The most significant difference was found in three cases. In case BCC-63, the C1orf10 expression level was increased 987 times in BCC lesion compared

to the normal tissue. In case SCC-86, the *hornerin* expression level was increased 190 times in involved lesion than the paired normal tissue. In case SCC-63, the *repetin* expression level was increased 78 times in involved lesion than the paired normal tissue.

Table 3.2: Expression analysis of human SFTPs in tumor skin samples. The transcript copy number was determined by real-time PCR using cDNA as a template. Always paired tissue specimens were analyzed (a, uninvolved tissue; b, involved tissue). In all cases, mRNA copy numbers were indicated as the absolute value per 1 ng total RNA. NS, normal skin; BCC, Basal Cell Carcinoma; SCC, Squamous cell carcinoma; PL, Precancerous Lesions.

Tissue	THHL1	THH	RPTN	HRNR	PFLG	IFPS	C1orf10
Normal skin	0.04	33.3	1.38	0.26	81.3	6.8	1.45
PL-28a	11.2	170	26.6	0.11	5428	528	105
PL-28b	4.5	34.1	19.2	0.44	1882	171	38.0
PL-91a	1.49	79.8	19.7	0.35	4867	219	53.6
PL-91b	2.45	59.5	26.7	0.53	9718	323	271
PL-111a	3.5	47.4	10.5	0.21	4248	168	63.1
PL-111b	3.12	44.4	4.89	0.19	845	71.5	10.5
SCC-63a	1.43	14.3	29.3	0.3	9212	559	278
SCC-63b	12.86	105	383	3.31	34600	1415	4225
SCC-73a	4.67	77.5	22.4	1.97	4840	442	202
SCC-73b	0.75	39.1	25.7	0.73	19570	711	237
SCC-86a	10.6	194	5.13	0.23	17610	541	244
SCC-86b	1.43	34.1	73.4	43.4	48670	875	191
SCC-109a	0.38	24.5	10.7	0.42	4155	550	180
SCC-109b	0.76	1.21	9.31	0.49	214	19.0	5.97
SCC-130a	2.16	10.4	11.2	0.28	8709	263	93
SCC-130b	1.52	0.82	3.69	0.3	540	63.1	17.9
BCC-1a	0.11	13.4	12.0	0.91	21300	486	87.7
BCC-1b	0.14	3.44	110	11.9	6634	167	532
BCC-14a	21.83	410	5.13	0.22	7480	187	111
BCC-14b	1.37	56.7	13.4	0.97	2506	119	109900
BCC-43a	4.72	158	6.37	0.26	6844	455	77.5
BCC-43b	0.22	0	7.82	1.42	848	76.1	9072
BCC-56a	0.52	29	3.6	0.17	9829	169	82.7
BCC-56b	0.12	0	11.8	0.19	1885	99.4	2679
BCC-95a	4.91	93.4	9.11	0.26	4115	133	88
BCC-95b	3.30	56.3	25.4	1.21	14860	647	424

To obtain more clues, expression of SFTPs in other skin diseases tissues were also examined (Fig. 3.6). The relative amounts of SFTPs (including *PFLG*, *IFPS*, *C1orf10*, *RPTN* and *THH*) levels were decreased in the involved skin tissue with inflammation, psoriasis

or DTC (acute allergic contact dermatitis) when compared with the paired normal tissues. The highest significant difference was found in *IFPS* levels in the case PS-3 (Fig. 3.6).

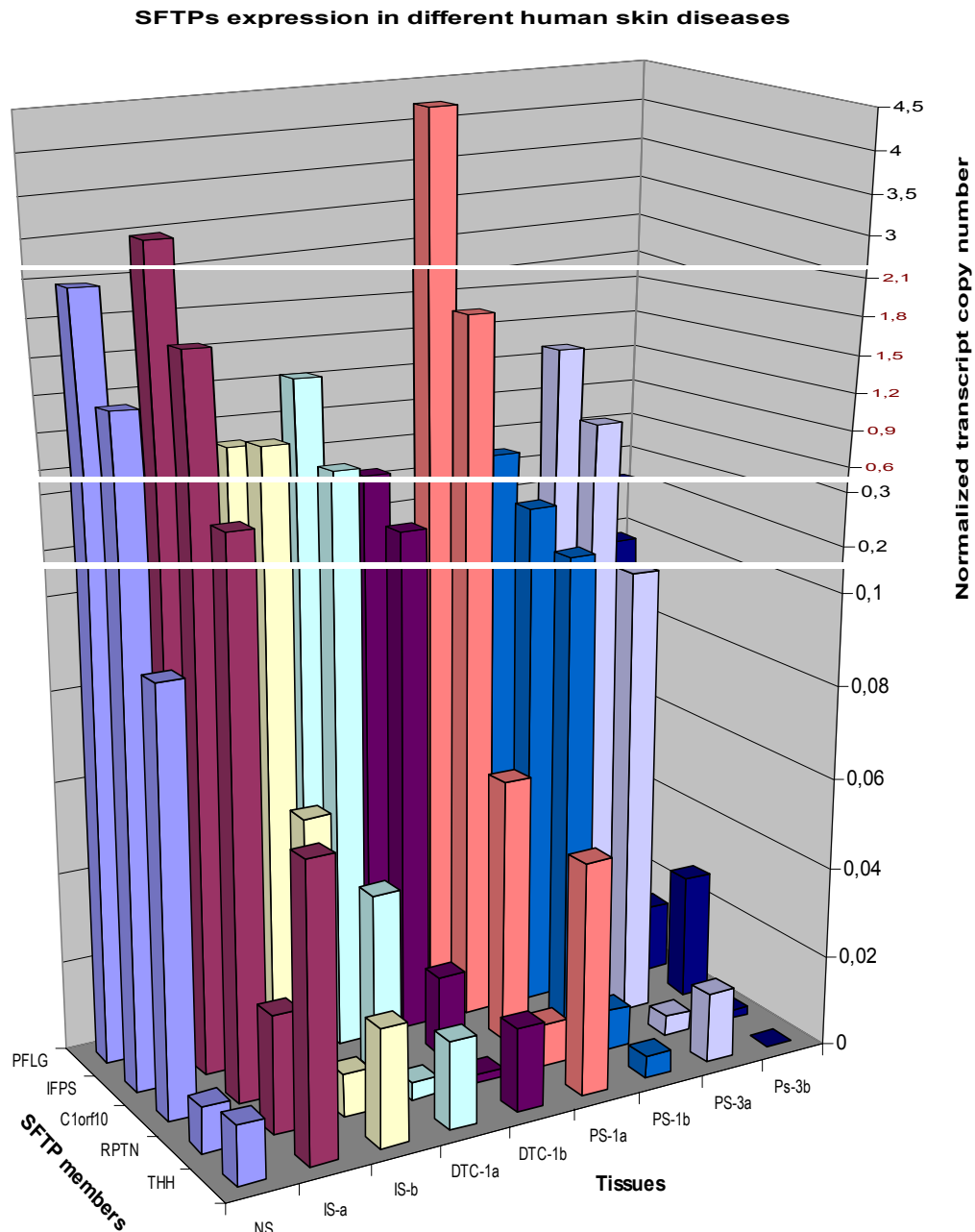


Figure 3.6: Expression of human SFTPs in skin disease tissues. *SFTP* mRNA was assayed in the lesional (a) and nonlesional (b) skin of four patients with inflammation, psoriasis vulgaris or acute allergic contact dermatitis and the normal skin of four healthy controls by real-time quantitative RT-PCR. A transcript copy number was determined by real-time PCR using cDNA as a template. In all cases, copy numbers were normalized against the copy number of the GAPDH gene. NS, normal skin; IS, inflamed skin; DTC, acute allergic contact dermatitis; PS, psoriasis lesions.

### 3.1.6 Comparative analysis of human SFTP mRNA-expression in proliferating and differentiating human keratinocytes

We compared the expression of SFTP family members in human foreskin-derived keratinocytes and examined whether or not this expression is altered by addition of retinoic acid (RA) or extracellular calcium. As shown in Fig. 3.7, 1.7 mM calcium increased the mRNA levels of all SFTP family members in keratinocytes after 2-day incubation, while  $10^{-7}$  M RA depressed completely their mRNA levels during eight days incubation. Calcium-induced expression of SFTP genes was associated with profound morphological changes, consistent with the onset of the differentiation program. Calcium significantly increased mRNA levels of *PFLG*, *RPTN*, *IFPS*, *C1orf10* and *THHL1* after 2 days and these increases were maintained over the course of six days exposure. The expression of *THH* and *HRNR* was also induced slightly after 2 days, although their relative mRNA levels were found to be very low. These results suggest that all SFTP genes are late differentiation markers of human keratinocytes.

To determine whether the SFTP promoter activity is related to upstream MAP kinase pathways, cells were treated with selective pharmacological inhibitors of ERK, JNK and p38. As shown in Fig. 3.8, either p38 inhibitors SB202190 and SB203580 or the MEK inhibitor PD98059 had no effect on SFTPs expression whereas the JNK inhibitor SP600125 upregulated SFTPs expression. Compared with calcium-dependent expression, the JNK inhibitor produced higher induction of *RPTN*, *PD98059* or *C1orf10* promoter activity than that of *PFLG* or *HRNR*.

Given that SFTP promoter activity is strongly regulated by MAPK pathways, we speculated whether the calcium-mediated induction was dependent on this signalling. To address this issue, cells were treated with different combinations of selective pharmacological inhibitors of ERK, JNK and p38. As shown in Fig. 3.9, these drugs each affected the ability of calcium to induce *RPTN/IFPS* promoter activity. The combination of PD98059 and SP202190 markedly inhibited calcium-mediated induction, whereas the combination of SP600125 and SP202190 or the three combination of SP600125, PD98059 and SP202190

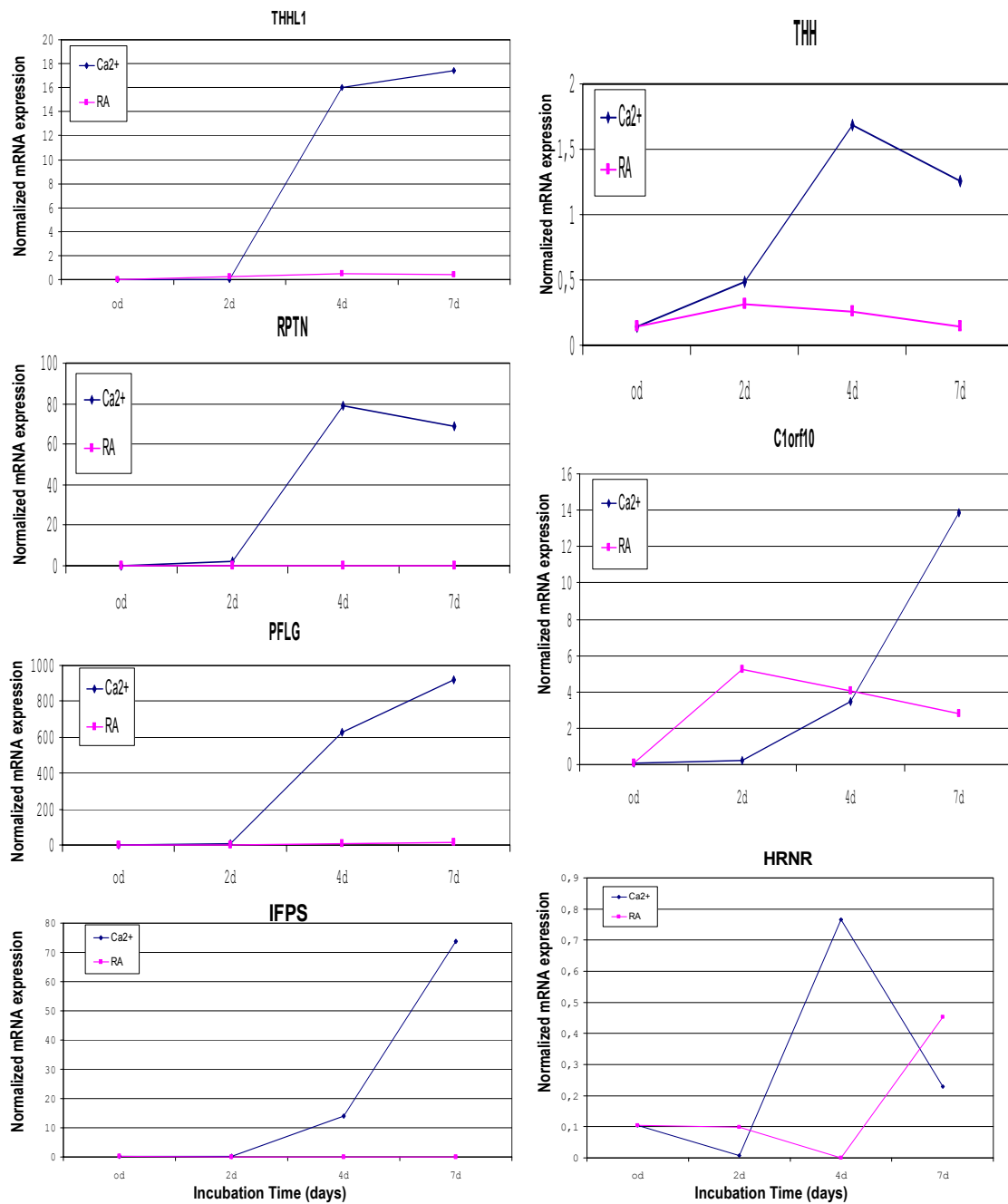


Figure 3.7: Human SFTP-expression in human foreskin-derived keratinocytes in response to calcium or retinoic acid. Second passage human keratinocytes were treated with 1.7 mM calcium or  $10^{-7}$  M RA. The cells were harvested at day 2, day 4 and day 7 respectively. The mRNA for SFTP genes were analyzed by real-time quantitative RT-PCR. A transcript copy number was determined by real-time PCR using cDNA as a template. In all cases, copy numbers were normalized against the copy number of the GAPDH gene and expressed as copies per 1000 GAPDH copies.

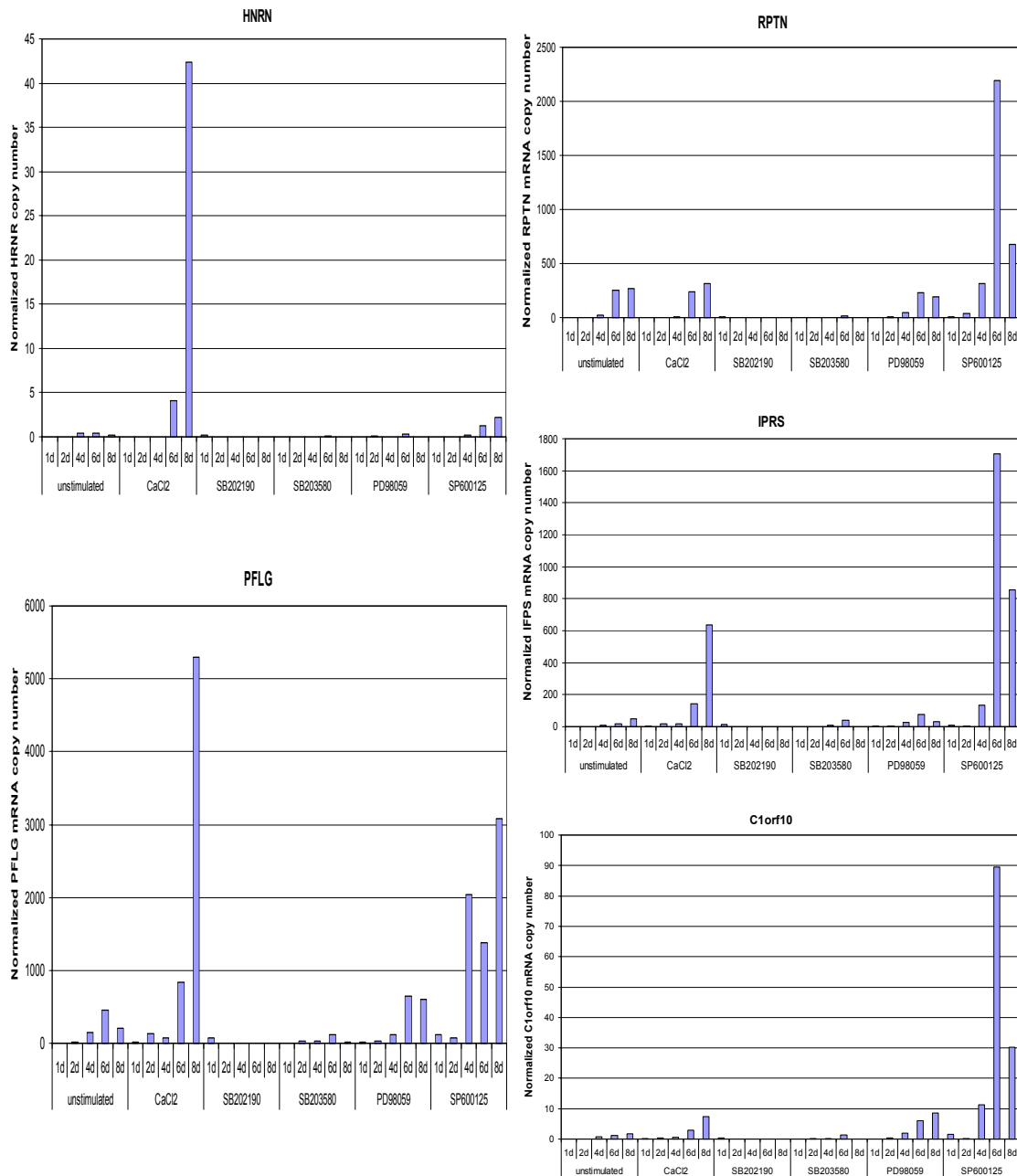


Figure 3.8: The ability to induce SFTP promoter activity is dependent on crosstalking of MAPKs. Second passage primary human keratinocytes were treated with 1.7 mM calcium, 10 mM MEK inhibitor PD98059, 10 mM JNK inhibitor SP600125, 5 mM p38 inhibitor SB202190 or SB203580. The cells were harvested at day 2, day 4, day 6 and day 8 respectively. The mRNA for SFTP genes were analyzed by real-time quantitative RT-PCR. A transcript copy number was determined by real-time PCR using cDNA as a template. In all cases, copy numbers were normalized against the copy number of the GAPDH gene and expressed as copies per 1000 GAPDH copies.

produced low induction of *RPTN*, or *IFPS* promoter activity with a peak at day 8 after treatment. Conversely, the three combination of SP600125, PD98059 and SP202474 (a p38 inactive analog) resulted in the lowest effect on the calcium-mediated induction. The *RPTN/IFPS* expression reached a peak at day 2 after treatment, then gradually decreased, but remained at high levels even at day 8. These results indicate that calcium-mediated changes in *RPTN* and *IFPS* expression are dependent on crosstalking of ERK, p38/MAP kinase and JNK. Both ERK and p38 positively regulate SFTPs expression in response to calcium whereas JNK inhibits SFTPs expression.

## **3.2 Identification of four novel human kazal-type serine protease inhibitor-like cDNAs (SPINKs)**

### **3.2.1 Characterization of the full-length human LEKTI-2 cDNA**

To identify the human LEKTI-2 cDNA sequence, we first used the N-terminal 30-amino-acid sequence of the natural peptide KT-AMP as a query and the tBLASTn algorithm against human genome data to look for the candidate DNA sequence on the NCBI server. This search led to a best-matched fragment within one putative exon, located on the genomic contig NT\_029289, which corresponds to amino acid residues 5-30 of KT-AMP (Fig. 3.10A). To generate a partial cDNA clone for this fragment, we performed PCR amplification using reverse-transcribed RNA from human foreskin-derived keratinocytes and specific oligonucleotides derived from the predicted cDNA sequence. Based on the resulted cDNA sequence from cloning and sequencing the PCR-amplified product, we designed specific oligonucleotides and performed 3'/5' RACE. This method allowed us to amplify two fragments containing the 5' and 3' end of the full-length cDNA, respectively. The cDNA sequence was completed by combining the overlapping sequences from each PCR product (Fig. 3.10B). The full-length cDNA (453 bp) contained an ORF consisting of 258 bp, from nucleotides 56 to 313, and a polyadenylation signal (AATAAA) situated 14 nucleotides 5' of the poly(A) tail. The deduced protein chain consisted of 86 amino acids.

## IFPS and RPTN expression in keratinocytes

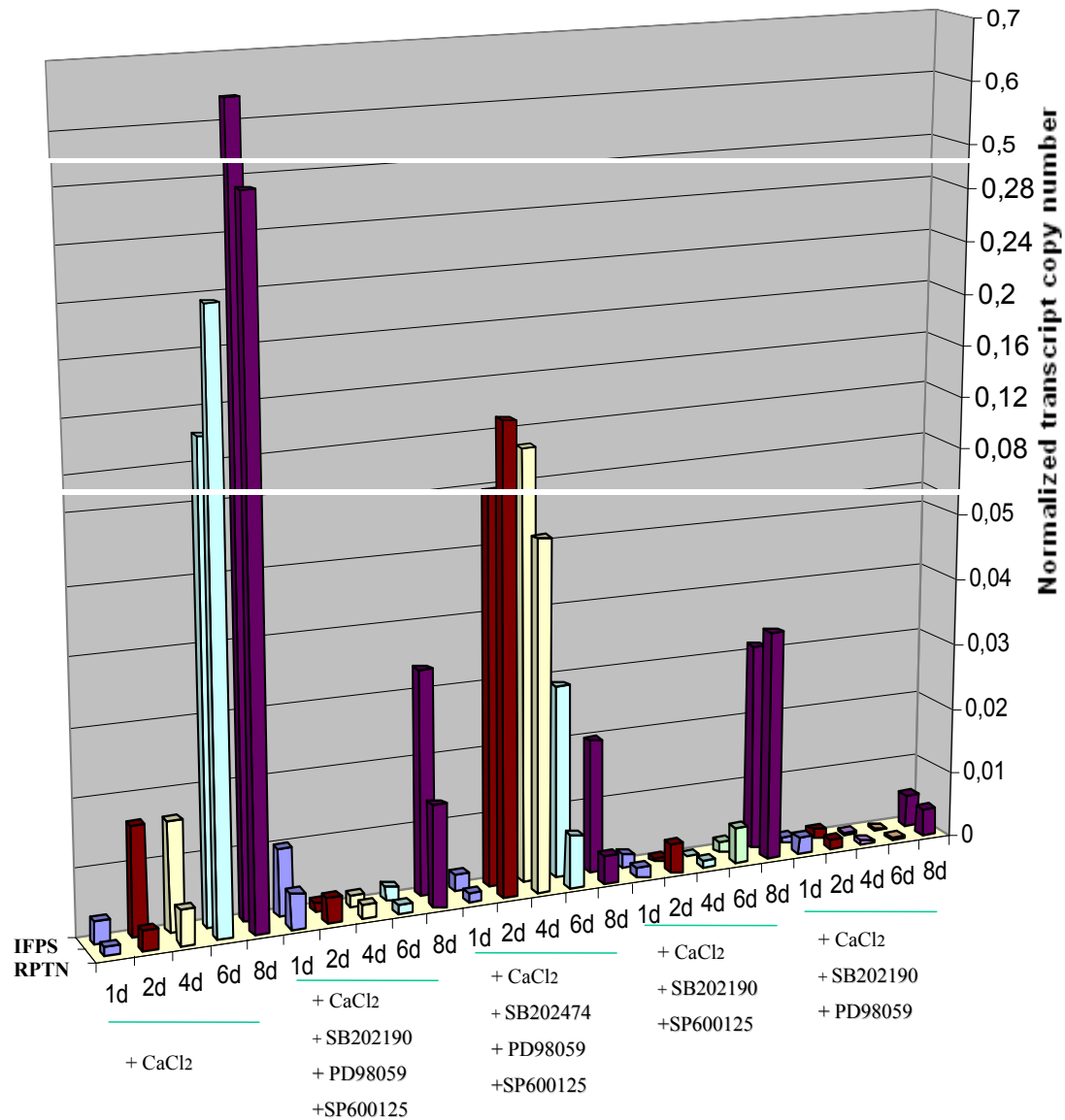


Figure 3.9: The ability of calcium to induce SFTP promoter activity is dependent on crosstalking of MAPKs. Second passage human keratinocytes were treated with different combinations of 10 mM MEK inhibitor PD98059, 10 mM JNK inhibitor SP600125, 5 mM p38 inhibitor SB202190 and 5 mM p38 inactive analog SB202474 for 2h and maintained in EpiLife medium containing 1.7 mM calcium. The cells were harvested 2 d, 4 d, 6 d and 8 d respectively. The mRNA for *RPTN* and *IFPS* were analyzed by real-time quantitative RT-PCR. A transcript copy number was determined by real-time PCR using cDNA as a template. In all cases, copy numbers were normalized against the copy number of the GAPDH gene and expressed as copies per 1000 GAPDH copies.



The 16 residues from the first Met represents a leader sequence containing a signal peptide. A short hydrophobic region located between amino acid 3 and 25 that was predicted by the SignalP application encodes the transmembrane signal anchor. This putative signal anchor is followed by a typical so-called kazal domain that spanned the residues 26-61, which corresponds to the native KT-AMP peptide, as predicted by the SMART program (Fig. 3.10 C). With high homology to domains 2 and 15 of a multi 'Lympho-Epithelial-Kazal-type-Inhibitor' (LEKTI), the matured protein was designated as LEKTI-2 and the gene was designated as SPINK9 (serine protease inhibitor, kazal type 9) (GenBank accession No. AY396740).

### 3.2.2 Characterization of SPINK9 homologous cDNAs

To identify candidate SPINK9 homologs genes within human genome, chromosome 5q32, we used the tBLASTn algorithm and the 61-amino-acid sequence of LEKTI-2 kazal domain as a query against human genomic data (Build 33, available on April 30, 2003) to search for DNA loci with potential kazal domain. This search led to the identification of a contig NT\_029289 in the human chromosome 5q32. A total of 8 putative homologous loci were clustered at 5q32. Among them, three correspond respectively to known genes, *SPINK1* (GenBank accession No. NM\_003122), *SPINK5* (GenBank accession No. NM\_205841) and *ECG2* (GenBank accession No. NM\_032566), one to SPINK9 while the other four are putative novel gene loci which were designated here as *SPINK6*, *SPINK7*, *SPINK8* and *SPINK10*. By using RT-PCR and the SMART RACE procedure as described above, corresponding *SPINK6*, *SPINK7*, *SPINK8* and *SPINK10* full-length cDNA clones were obtained from keratinocytes and sequenced. During the course of this work, nucleotide and protein sequence databases (mainly GenBank and SwissProt) were periodically searched. The mRNA or predicted sequences for *SPINK6*, *SPINK7* and *SPINK8* have recently appeared in GenBank database though detailed cloning and characterization of these mRNA and protein sequences has not been reported.

**SPINK6.** In the 5' experiment, two extended products were detected with different transcription start site (TSS) but sharing the same first exon. In the 3' experiment, a single

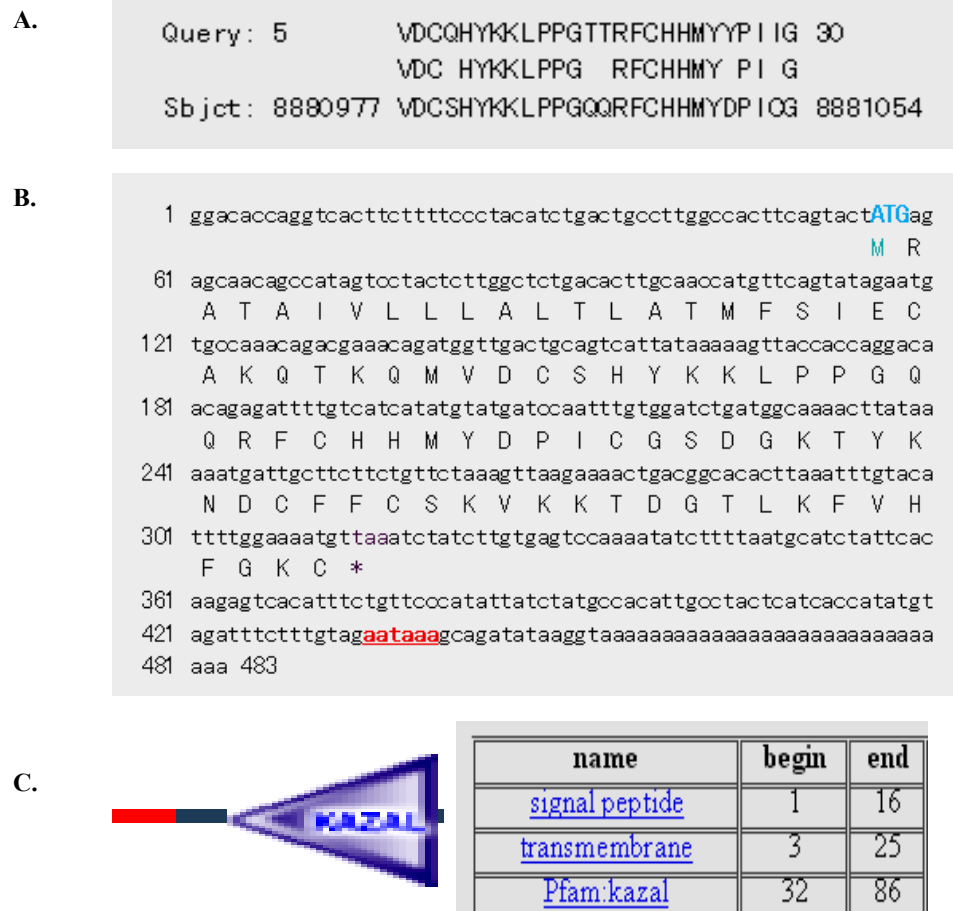


Figure 3.10: Nucleotide and amino acid sequences of human SPINK9. A, the best-matched fragment located on the genomic contig NT\_029289 was found by the tBLASTn program. B, the full-length nucleotide sequence of human SPINK9 cDNA (453 bp) was verified by direct cDNA cloning and sequencing and aligned with the predicted open reading frame amino acid sequence. The underlined red nucleotides indicated the canonical poly(A) addition signal. C, schematic representation of the predicted domain architecture of human LEKTI-2. The locations of the signal peptide (marked red) and kazal domain (arrow) are indicated.

extension product was detected which contained a polyadenylation signal (AATAAA) situated 214 nucleotides 5' of the poly(A) tail. The full-length cDNA (946 bp for the longer transcript; 883 bp for the shorter one) contained an ORF consisting of 243 bp (Fig. 3.11A). Compared with the newly identified SPINK6/MGC21394 (GenBank accession No. NM\_205841) (Strausberg et al., 2002), all the three *SPINK6* transcripts contain the same ORF though showing both 5' and 3'-end extensional transcription in keratinocytes. The deduced protein chain consisted of 80 amino acids with a leader sequence containing a signal peptide (residues 1–24) while the residues 7–26 represent a transmembrane domain.

The last 54 residues include the kazal domain (Fig. 3.11).

```

▼TSS1
1  gtgaatcatccctataatgtagtgaaactttagaaagaagagccgggaggatgtattggt
▼TSS2
61  tgttaggaaaaatgtaggctaccagtagaaaaatgacattctctattaataagatctgaggt
121 gcgacacacataattgtcccaatttttaagattgatggggagcatgaagcatttttttaa
181 tgtggtggcaggccccattaaatgcataaactgcataggactcatgtggtctgaatgtat
241 tttagggctttctgggaattgtcttgacagagaacctcagctggacaaagcagccttgat
301 ctgagtgagctaactgacacaatgaaactgtcaggcatgtttctgctcctctctctggct
      M K L S G M F L L L S L A
361 cttttctgctttttaacaggtgtcttcagtcagggaggacaggttgactgtgggtgagttc
      L F C F L T G V F S Q G G Q V D C G E F
421 caggacaccaaggtctactgcaactcgggaatctaaccacactgtggctctgatggccag
      Q D T K V Y C T R E S N P H C G S D G Q
481 acatatggcaataaatgtgccttctgtaaggccatagtgaaaagtgggtgaaagattagc
      T Y G N K C A F C K A I V K S G G K I S
541 ctaaagcatcctggaaaaatgctgaggttaaagccaatgtttcttggtgacttgccagctt
      L K H P G K C *
601 tgcagccttcttttctcacttctgcttataacttttgctgggtgattcctttaattcataa
661 agacatacctactctgcctgggtcttgaggagtccaatgtatgtctattttctcttgattc
721 acttgtcaataaaagtacattctgcaaaagcattgactgtgttcttactttgagatcaaga
781 aaatatatacatacaacagagaataattgaaactgggataaaacttggatcaagaaacctat
841 tttcctgatctaagttaaccatccagtcattgtgagatcatagggaaaatcccttcctttc
901 tcttattcttagaatctccatggtttataatgaaggattttgacaaca 946

```

Figure 3.11: Nucleotide and amino acid sequences of human SPINK6. The full-length nucleotide sequence of human SPINK6 cDNA was aligned with the predicted open reading frame amino acid sequence. The underlined nucleotides indicated the canonical poly(A) addition signal. The two transcript variants of SPINK6 are indicated. TSS1, transcription start site 1; TSS2, transcription start site 2.

**SPINK8.** In the 5' experiment, three extended products were detected. In the 3' experiment, a single extension product was detected which contained a polyadenylation signal (AATAAA) situated 20 nucleotides 5' of the poly(A) tail. Three PCR fragments were obtained after a long-distance PCR with a pair of specific primers derived from 5' and 3' end. The full-length cDNA of SPINK8-v1(1022 bp) contained an ORF consisting of 285 bp, from nucleotides 532 to 816, which is exactly the same as that of the newly predicted SPINK5L2 mRNA sequence (GenBank accession No. XM\_376433). The deduced protein chain consisted of 94 amino acids with a leader sequence containing a signal peptide (residues 1–24) while the residues 4–26 are a transmembrane domain. The residues 47–93 are a kazal domain. The other two variants (1171 bp for SPINK8-v2; 1088 bp for SPINK8-v3) didn't encode any kazal domain.

**SPINK7.** In both the 5' and 3' experiment, a single extended product was detected. The full-length cDNA (522 bp), designated here SPINK7-v2 (SPINK7 splicing variant

2), contained an ORF consisting of 214 bp and a poly(A) signal (AATAAA) situated 18 nucleotides 5' of the poly(A) tail (Fig. 3.12A). The deduced protein chain consisted of 71 amino acids. Compared with the newly identified human SPINK5L2 partial mRNA sequence (GenBank accession No. NM\_001001325; designated here SPINK7-v1) (Puente and Lopez-Otin 2004), SPINK7-v2 contains the same kazal domain (residues 13-71) as v1 (Fig. 3.12B–D). In contrast, SPINK7-v1 possessed a leader sequence containing a signal peptide (residues 1–22) and a transmembrane domain (residues 7–26), as is typical for a SPINK protein, while SPINK7-v2 contains no signal peptide or transmembrane domain.

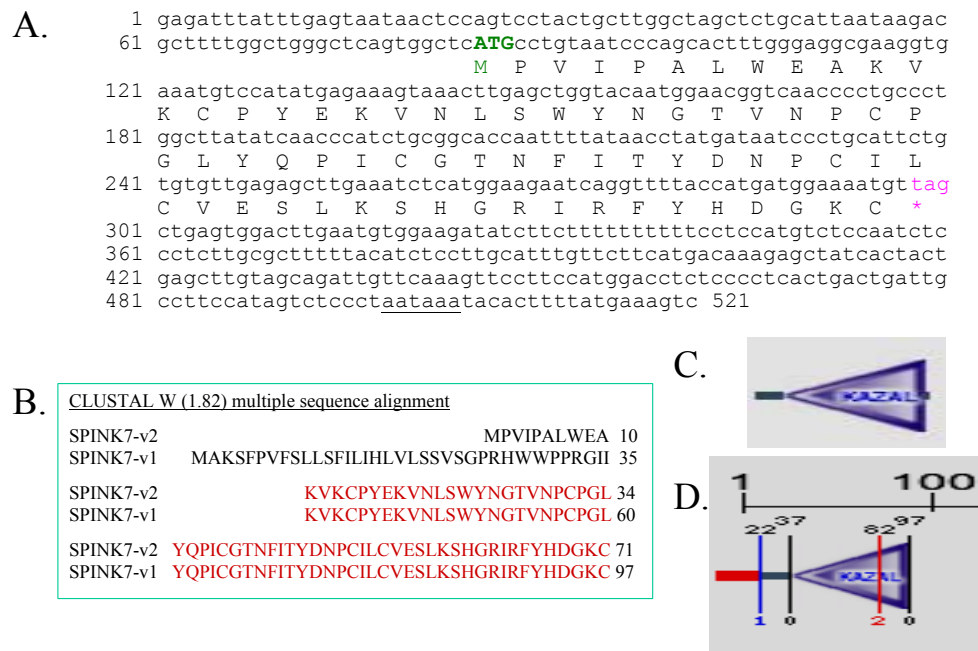


Figure 3.12: Nucleotide and amino acid sequences of human SPINK7. A, The full-length nucleotide sequence of human SPINK7-v1 cDNA was aligned with the predicted open reading frame amino acid sequence. The underlined nucleotides indicated the canonical poly(A) addition signal; B, Alignment of the full-length protein sequences of the both SPINK7 variants; C, SPINK7-v1 domain architectures. The location of kazal domain (arrow) is indicated; D, SPINK7-v2 domain architectures. The locations of the signal peptide (marked red) and kazal domain (arrow) are indicated.

**SPINK10.** In the 3' experiment, a single extension product was successfully detected but in the 5' experiment no product was obtained. The deduced amino acid sequence of the 3' RACE PCR product then was used as a query to BLASTp and tBLASTn against public protein and EST databases on the NCBI server. This search led to a best-matched ho-

mologous protein, the double-headed protease inhibitor (DHPI) from dog submandibular gland (Swissprot accession No. P01002), and two EST sequences from bovine gut (EMBL accession Nos. CK946321 and CK946653) which can be assembled as a putative full-length cDNA sequence encoding the possible bovine DHPI. The predicted bovine DHPI amino acid sequence was used as a query to tBLASTn against human genome data to look for the 5' end cDNA sequence of human DHPI. After being defined by computer analysis of exon-intron boundary, the full-length cDNA (1172 bp), designated here SPINK10, was generated which contained an ORF consisting of 405 bp and a poly(A) signal (ATTAAA)(Fig. 3.13A). The deduced protein chain consisted of 134 amino acids with a leader sequence containing a signal peptide (residues 1–21) and two kazal domains (residues 29–83 and 86–134) (Fig. 3.13B).

### 3.2.3 Genomic structure of human SPINK9 and its homologs genes

The genomic structure of human novel SPINKs genes was analysed by using the BLAT algorithm to compare each mRNA sequence with the human genome sequence assembled in May 2004 in the UCSC server (<http://genome.ucsc.edu/cgi-bin/hgBlat>). The mRNA sequence of SPINK9 was aligned perfectly to a unique genome region on the SPINK9 gene. These alignment data clearly indicate that each unique mRNA segment represents an individual exon. Thus, according to the specific location of each mRNA segment and the distance between each segment on the gene, the exon/intron structure of this gene can be predicted. All introns were flanked by the consensus donor and acceptor splice sites.

The SPINK9 gene consists of four exons, exon 4 being the largest and exon 2 being the smallest; exon 1 contains the ATG start codon, and exon 4 contains the TAA stop codon (Fig. 3.14).

The SPINK10 gene consists of six exons, exon 6 being the largest and exon 2 being the smallest; exon 1 contains the ATG start codon, and exon 6 contains the TGA stop codon (Fig. 3.15).

The SPINK6 gene consists of four exons, exon 4 being the largest and exon 2 being the smallest; exon 1 contains the ATG start codon, and exon 4 contains the TGA stop codon.

**A**

```

1 tcttctctctctggactccttggctcagagcagcatatctccatcattggcattccaccat
61 cactATGaagggttcactgcttttgccatccttggctagcagtcactgcttgggtac
   M K G F T A F A I L G L A V T A W A T
121 ctcgccatatgcaaaagggacaaagtgaaactgttccagatacattaaaggcttaaagac
   S P Y A K G T K V N C S R Y I K G L K T
181 tgcattgccaagagactggagaccaatatgtggcacagatcagaaaacttacagtaatga
   A C P R D W R P I C G T D Q K T Y S N E
241 atgcatgttttgcatgcaaaaccaggaagtagatcaagaatttcaactcagaaaacttca
   C M F C M Q N Q E V D Q E F Q L R K L H
301 tgaaggtaaatgtcagattcagtgaccaggtattctgaggcatgtaccatggactacac
   E G K C Q I Q C T R Y S E A C T M D Y T
361 acttcactgtggatctgatggaaatgtatattccaacagatgtaccttttgaataccgt
   L H C G S D G N V Y S N R C T F C N T V
421 tgtgaagagccaaggtgcaatttggttgaaaattatggattgtgctgaatctccctgat
   V K S Q G A I W L K N Y G L C *
481 gcttaaggatatcaagtctccttggcagattccacttagataaccagtttgtcttggac
541 ggctgaggaggcaagaataatctggtacaaacatggatttgagcttccaacaaagtatt
601 ctcagcaaggaactgactatcagtcctatctgtggatgtgattcttctcttaaggagaat
661 atctatggtcattttcctaagctaaaaatcatatagaaagattttctctatttctatata
721 aattgggaactagatattatacacacaatggtgagaagagaatatgcttatacatgttt
781 aaaagactcaattgttttaagtaaacatgagtcagaggatggatctgtccattggctca
841 tattatctttggcacactaatgaagtgaaaaatttagatttggagaagtgagaatccaa
901 aatactatacactaatcagaacagaccagtataacgtatattttagggtgcatcattta
961 aagaggactcagtcacgtatgggacatgcaaaaaaggcattatgttgaagggagagaga
1021 cattaaaaacaagatctgattgattaatgttgaagtcactaaaaggcctcaggaaaata
1081 tgatggagggtgggaatgatgagggtatattcaaaaccatctgagagagaaactccagaa
1141 ttttctcatctagcatgccatagaagaagcacc 1172

```

**B**

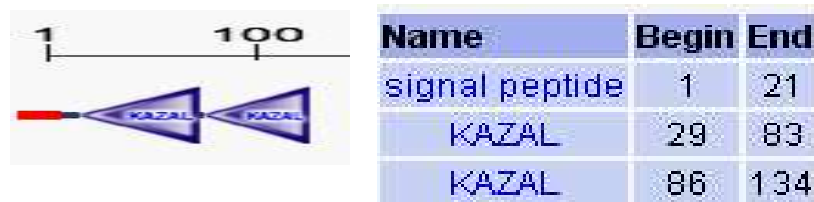


Figure 3.13: Nucleotide and amino acid sequences of human SPINK10. A, The predicted full-length nucleotide sequence of human SPINK10 cDNA was aligned with the predicted open reading frame amino acid sequence. The 3'-end sequence from nucleotides 318 to 1172 was verified by direct cDNA cloning and sequencing while the 5' end sequence was predicted by bioinformatic analysis. The underlined nucleotides indicated the canonical poly(A) addition signal. B, SPINK10 domain architectures. The locations of the signal peptide (marked red) and kazal domains (arrow) are indicated.

This structure also demonstrates the different transcription start sites of exon 1 to produce the long and short form mRNAs. All the mRNAs (SPINK6-v1, -v2 and -v3) are produced by utilizing all exons (thus the same protein) except that the SPINK6-v3 mRNA is obtained by utilizing shortened 5' and 3' UTR (Fig. 3.16).

The SPINK8 gene consists of eight exons including five constant exons (c1-5) and three variable exons (v1-3), c1 being the largest and v3 being the smallest; c3 contains the ATG start codon, and c5 contains the TGA stop codon (Fig. 3.17). This structure also demonstrates the alternative splicing of exon v1-3 to produce the three form mRNAs (SPINK8-v1, -v2 and -v3). The SPINK8-v1 mRNA is produced by utilizing exons c1-5 and

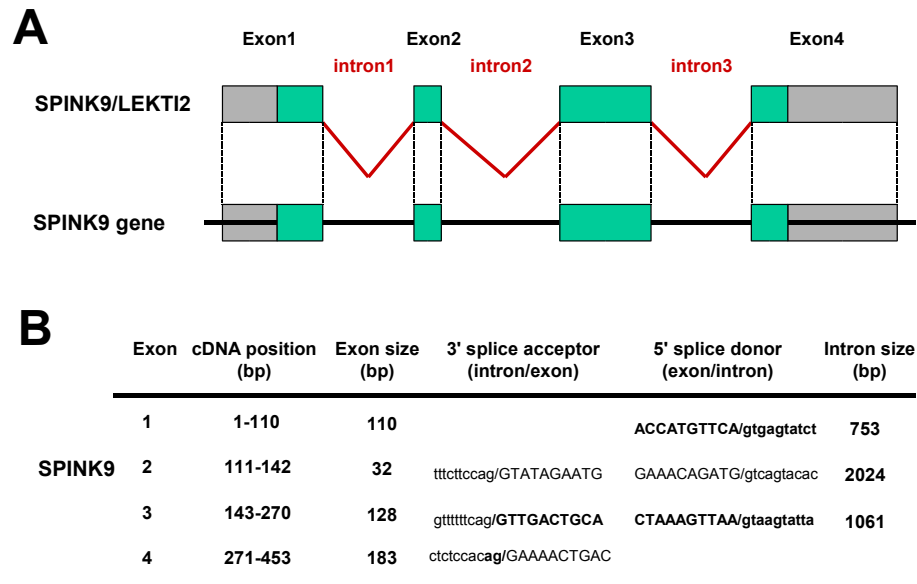


Figure 3.14: Genomic structure of the human SPINK9. A, Schematic representation of the genomic and exon/intron organization for the human SPINK9. Boxes represent exons, green-colored regions indicate the coding sequence, and intervening lines denote introns. B, Exon/intron sizes and boundaries. The splice acceptor and donor sequences are shown in upper and lower case letters, respectively.

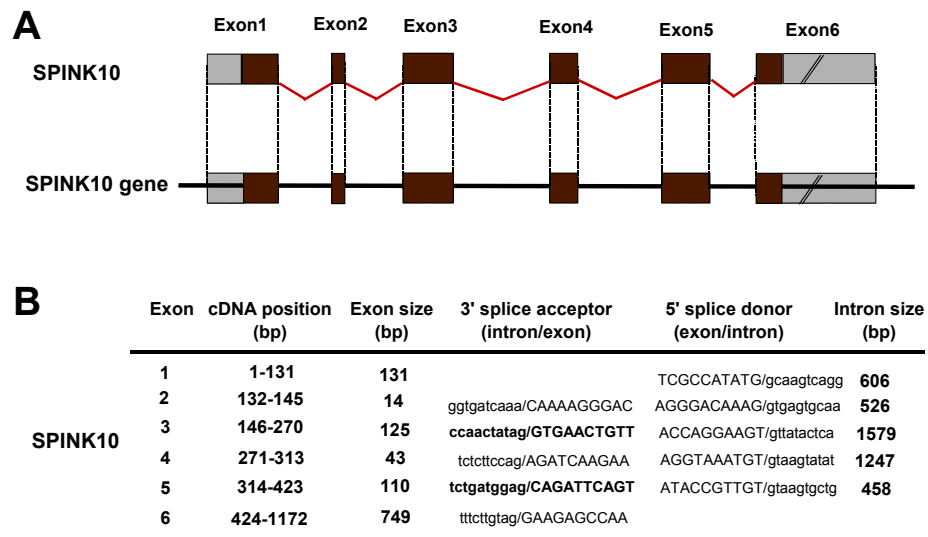


Figure 3.15: Genomic structure of the human SPINK10. A, Schematic representation of the genomic and exon/intron organization for the human SPINK10. Boxes represent exons, brown-colored regions indicate the coding sequence, and intervening lines denote introns. B, Exon/intron sizes and boundaries. The splice acceptor and donor sequences are shown in upper and lower case letters, respectively.

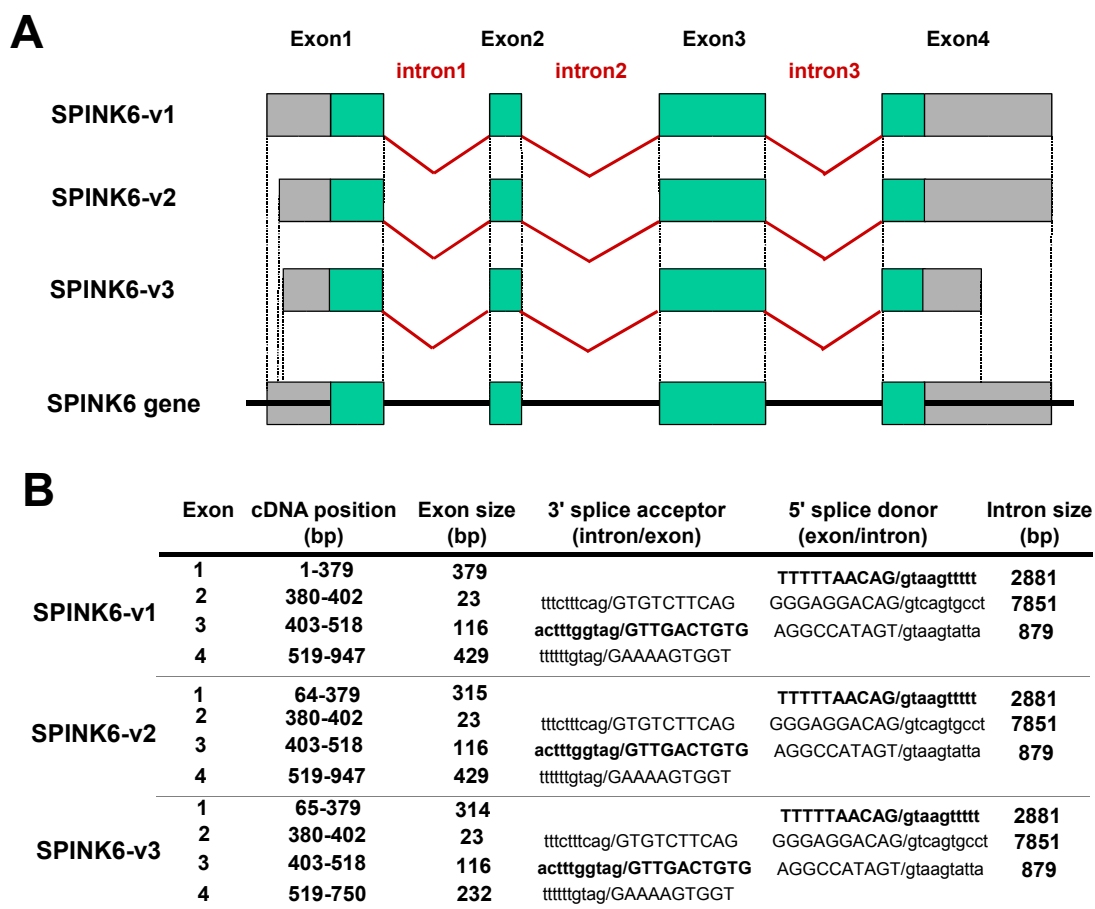


Figure 3.16: Genomic structure of the human SPINK6. A, Schematic representation of the genomic and exon/intron organization for the three splice variants. Boxes represent exons, green-colored regions indicate the coding sequence, and intervening lines denote introns. B, Exon/intron sizes and boundaries for the mRNA isoforms (SPINK6-v1, -v2 and -v3). The splice acceptor and donor sequences are shown in upper and lower case letters, respectively.

v3, whereas the SPINK8-v2 mRNA is obtained by utilizing all but exon v2 and that the SPINK8-v3 mRNA is obtained by utilizing all but exon v1. More interestingly, the exon v3 for SPINK8-v3 adopts a different splice acceptor.

The SPINK7 gene consists of at least four exons according to the newly isolated SPINK5L2 mRNA partial sequence (SPINK7-v1) from human skin (5' and 3' UTR sequences remain unknown); exon 1 contains the ATG start codon, and exon 4 contains the TAG stop codon (Fig. 3.18). However, to our surprise, the SPINK7-v2 mRNA is produced by utilizing exons 3 and 4 on the gene locus 5q32 while its 5'-end mRNA leader sequence



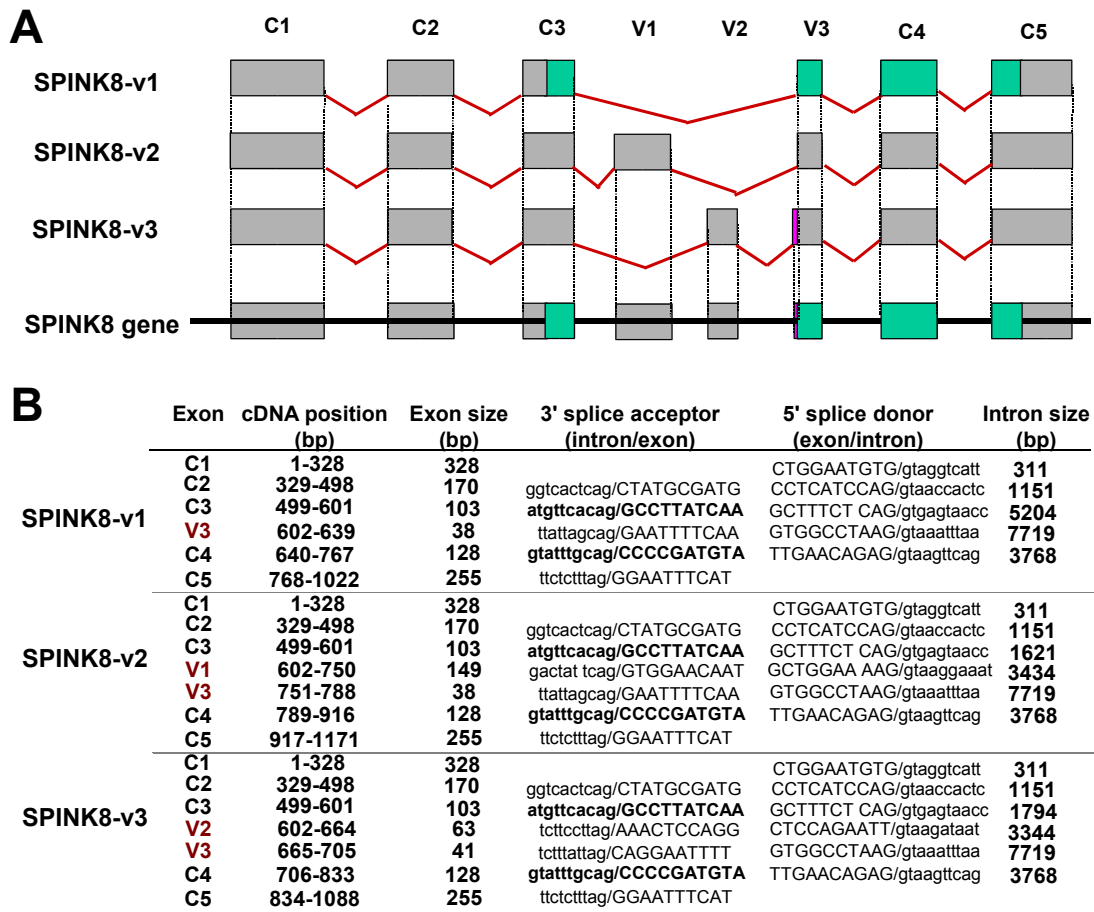


Figure 3.17: Genomic structure of the human SPINK8. A, Schematic representation of the genomic and exon/intron organization for the three splice variants. Boxes represent exons, green-colored regions indicate the coding sequence, and intervening lines denote introns. B, Exon/intron sizes and boundaries for the mRNA isoforms (SPINK8-v1, -v2 and -v3). The splice acceptor and donor sequences are shown in upper and lower case letters, respectively.

(117 nucleotides) is obtained by utilizing a foreign exon from the reversed complement strand at human chromosome 7q11.23. The foreign exon contains a retroviral sequence (nucleotides 1–66) and an Alu sequence (nucleotides 67–117).

To understand the chimeric structure of SPINK7-v2 mRNA, PCR analysis was performed by using genomic DNAs as templates (Fig. 3.19). Four pairs of specific primers were designed with the retrieved genomic DNA sequences at 5q32 on the NCBI server. All the PCR products were sequenced. The results showed that both the commercial genomic DNA (CG-DNA) and the fresh-prepared genomic DNA (KG-DNA) isolated from

foreskin-derived keratinocytes possess a common 3973 bp region upstream of the gene exon 3. By using the fourth pair of primers to generate an expected product "D", a related PCR product was found in CG-DNA but not in KG-DNA. This suggests that there could be a genomic breakage point, a big-fragment integration or interruption within the related region in KG-DNA.

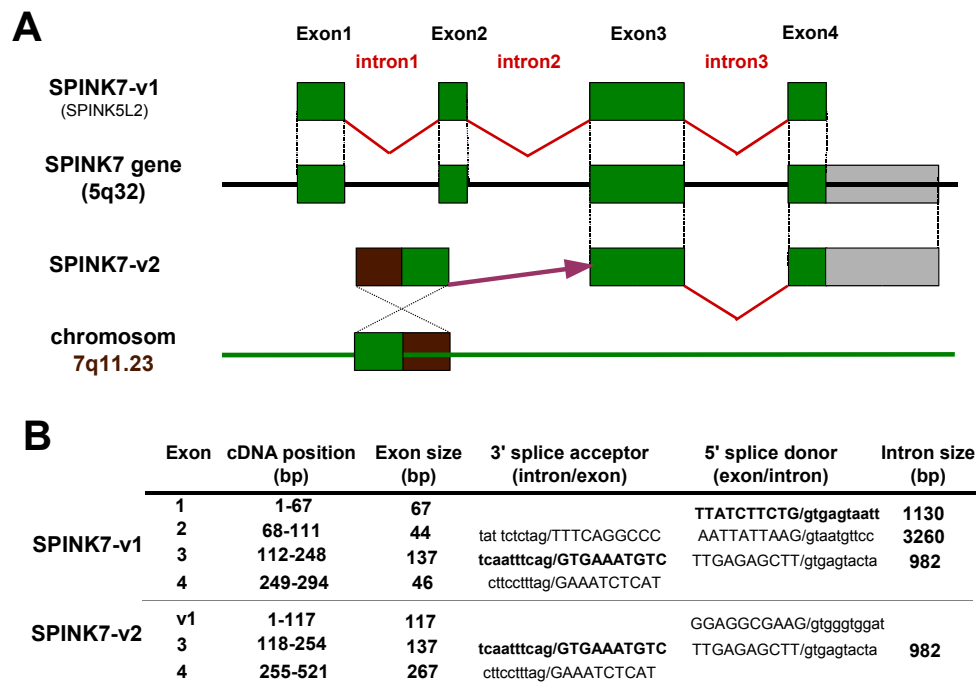


Figure 3.18: Genomic structure of the human SPINK7. A, Schematic representation of the genomic and exon/intron organization for the two mRNA-splice variants. Boxes represent exons, green-colored regions indicate the coding sequence, brown-colored regions indicate the untranslated sequence and intervening lines denote introns; B, Exon/intron sizes and boundaries for the mRNA isoforms (SPINK7-v1 and -v2). The splice acceptor and donor sequences are shown in upper and lower case letters, respectively.

### 3.2.4 Expression analysis of the human SPINKs

To determine the expression pattern of the newly identified human SPINKs genes, RT-PCR analyses were performed on twenty-three adult tissue samples (Fig. 3.20). From all the 23 tissues analyzed, only foreskin-derived keratinocytes presented a clearly detectable signal for all SPINK mRNAs tested, while no detectable levels for SPINK6, SPINK7-v2 mRNA or SPINK9 were seen in normal skin. SPINK7-v2 mRNA was restricted to bone marrow.

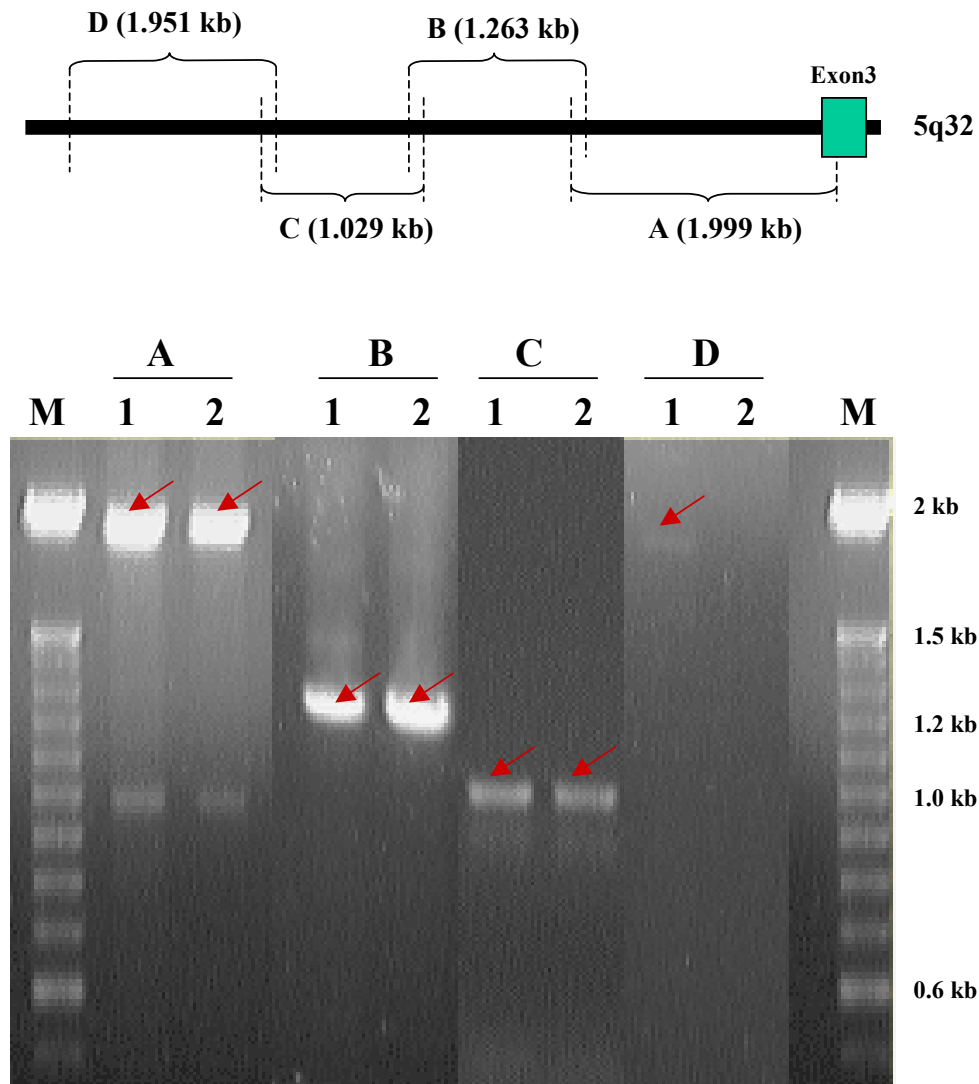


Figure 3.19: Genomic PCR analysis of the human *SPINK7* gene. Two kinds of genomic DNA were used: 1, a commercial human genomic DNA from Promega; 2, the fresh-prepared genomic DNA isolated from foreskin-derived keratinocytes. The PCR amplification was performed with a specific primer pair (upper panel) and the analysis of amplification products was done by agarose gel electrophoresis (lower panel). Lanes A–D each corresponds to the expected PCR products on the upper panel. The arrow-marked bands were confirmed by sequencing.

Moderate expression of *SPINK8-v1* was detected in lung and low expression in skin and colon. *SPINK9* mRNA was highly expressed in thymus but also low in tonsil, adenoid and brachial epithelial cells (BEC). *SPINK6* mRNA was expressed in thymus, small intestine, neutrophils, BEC, larynx, pharynx, colon and polyp.

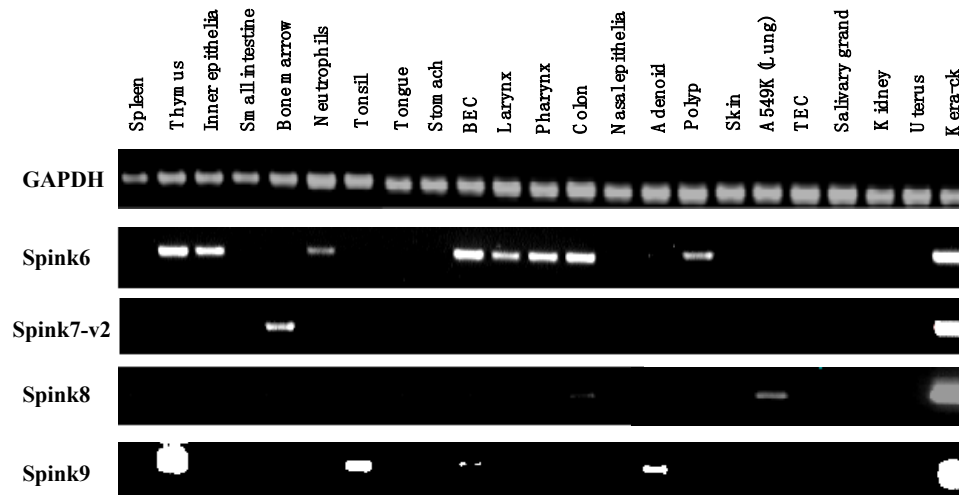


Figure 3.20: RT-PCR analysis of human SPINKs mRNA-expression. RNA was isolated from human tissues and cells as indicated. The RNA was treated with reverse transcriptase followed by PCR amplification with a SPINK-specific primer pair (lower and middle panel) or a GAPDH-specific primer pair (upper panel) and the analysis of amplification products was performed by agarose gel electrophoresis. GAPDH was employed as a loading control. BEC, bronchial epithelial cells; TEC, tracheal epithelial cells; Kera-ck, untreated primary keratinocytes.

To analyze the expression of *SPINKs* in skin cancer tissues, we examined the expression of the four *SPINK* genes by RT-PCR with paired tissue specimens (Fig. 3.21). *SPINK6* exhibited a complex expression profile among different cases. *SPINK7-v2* mRNA was detected only in the involved tissue of one patient with precancerous lesions. High expression of *SPINK9* was found in two cases, namely PL-91 and SCC-63. In contrast to only *SPINK8-v1* mRNA transcript detected in the colon, skin and keratinocytes (Fig. 3.21), a complex alternative splicing pattern was found in tumor skin. A total of three different *SPINK8* mRNA transcripts (*SPINK8-v1*, *-v2* and *-v3*) were isolated and confirmed by sequencing (Fig. 3.17).

### 3.3 Functional analysis of the human hornerin and LEKTI-2

#### 3.3.1 Purification and Characterization of recombinant hornerin fragments

The PCR fragment encoding protein fragments consisting of 50, 98, 72 or 124 amino acid residues of hornerin was subcloned into an expression vector (pET32a) to yield pET32a-

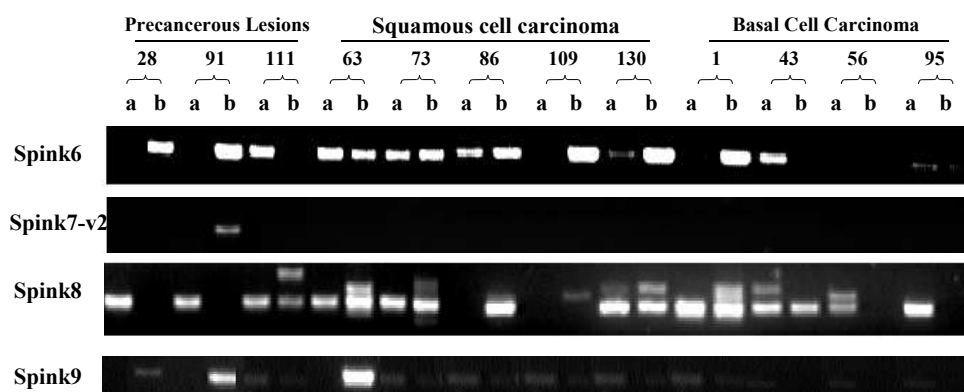


Figure 3.21: RT-PCR analysis of human SPINKs expression in tumor skin. Paired tissue specimens were analyzed (a, uninvolved tissue; b, involved tissue). NS, normal skin; BCC, basal cell carcinoma; SCC, squamous cell carcinoma; PL, precancerous lesions.

HRNR-1, pET32a-HRNR-2, pET32a-HRNR-3 and pET32a-HRNR-4, respectively. The expression of recombinant fusion proteins in *E. coli* BL21(DE3)pLysS was induced by the addition of isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) as described in Materials and Methods. The soluble and insoluble cytoplasmic fractions, separated from induced *E. coli* cells and analyzed by Tricine-SDS-PAGE, revealed that the fusion protein accumulated essentially as soluble material in the cytoplasmic fraction of *E. coli* (data not shown). Subsequently, the fusion protein was purified by Protino<sup>®</sup>Ni prepared columns kits and then by RP8-HPLC (Fig. 3.22 A, Fig. 3.23 A, Fig. 3.24 A and Fig. 3.25 A). Mass analysis indicated post-translational removal of the initial methionine (Fig. 3.22 C, Fig. 3.23 C, Fig. 3.24 C and Fig. 3.25 C).

The (Asp)<sub>4</sub>-Lys sequence located upstream of the insert protein sequence is the target of enterokinase (EK). This enzymatic cleavage was used to recover the related protein. The (His)<sub>6</sub>-HRNR-(1-4) fusion proteins were subjected to 3-hour cleavage process. The EK-treated fractions were purified further by RP4-HPLC (Fig. 3.22 B, Fig. 3.23 B, Fig. 3.24 B and Fig. 3.25 B) and were subjected to mass analysis (Fig. 3.23 D, Fig. 3.24 D and Fig. 3.25 D). Mass spectrometry confirmed the EK proteolytic products of three fusion proteins corresponding to the polypeptides HRNR2, HRNR3 and HRNR4. However, within the resulting RP4-HPLC peaks of EK-treated His-HRNR1, there was no exact mass that corresponded to the polypeptide HRNR1.

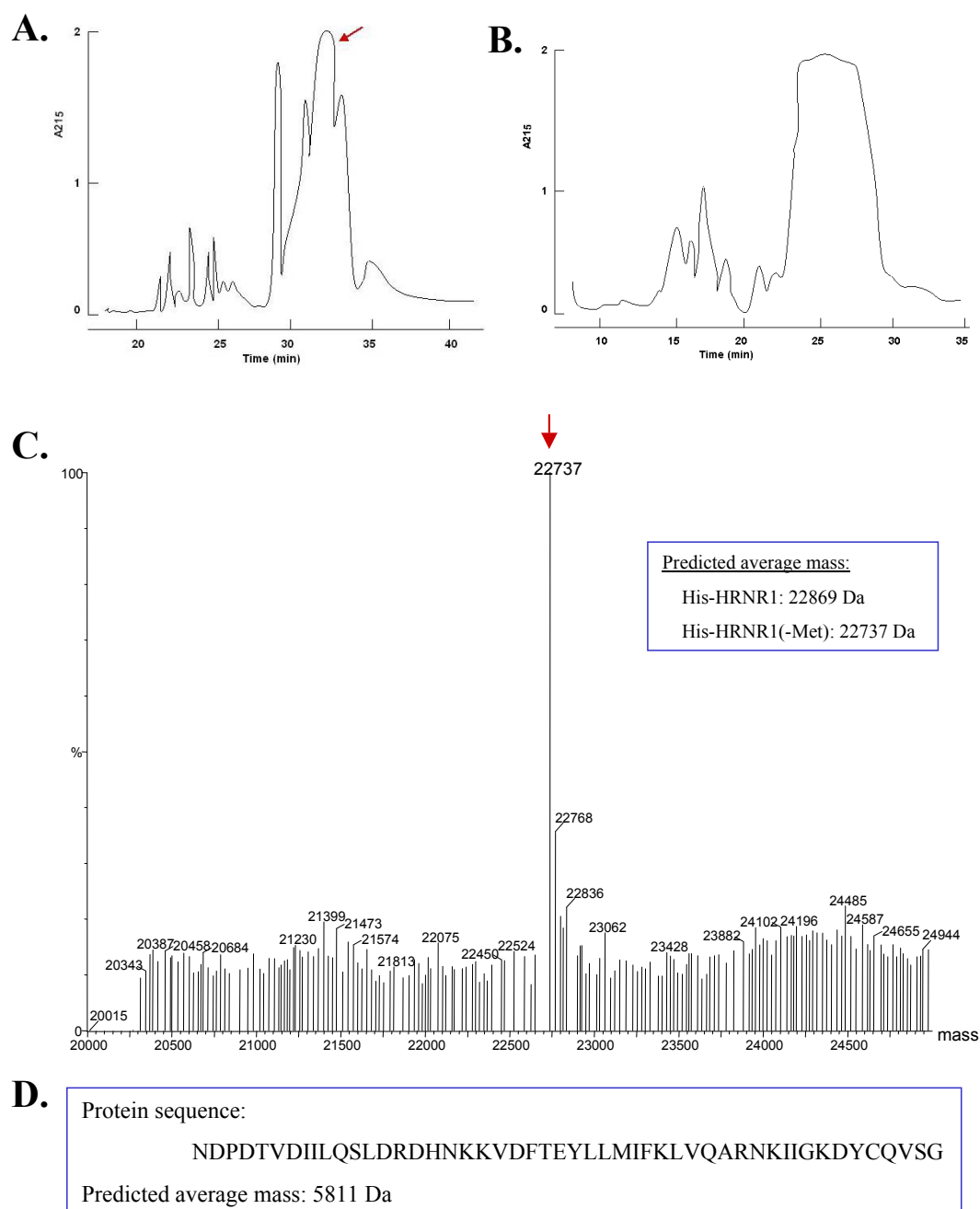


Figure 3.22: Purification and characterization of the His-tagged fusion protein His-HRNR1. A, the Ni-column-bound His-tagged fusion proteins of *E. coli* extracts were separated by RP8-HPLC, and the fraction containing the highest peak was collected. The red arrow indicates the highest peak; B, the enterokinase-treated fraction was purified by RP4-HPLC, and the resulting peaks were collected and subjected to mass analysis; C, ESI-MS spectrum of the purified His-HRNR1 revealed an exact mass of 22737 Da. The red arrow indicates the mass corresponding to the predicted mass; D, the amino acid sequence and predicted mass of HRNR1.

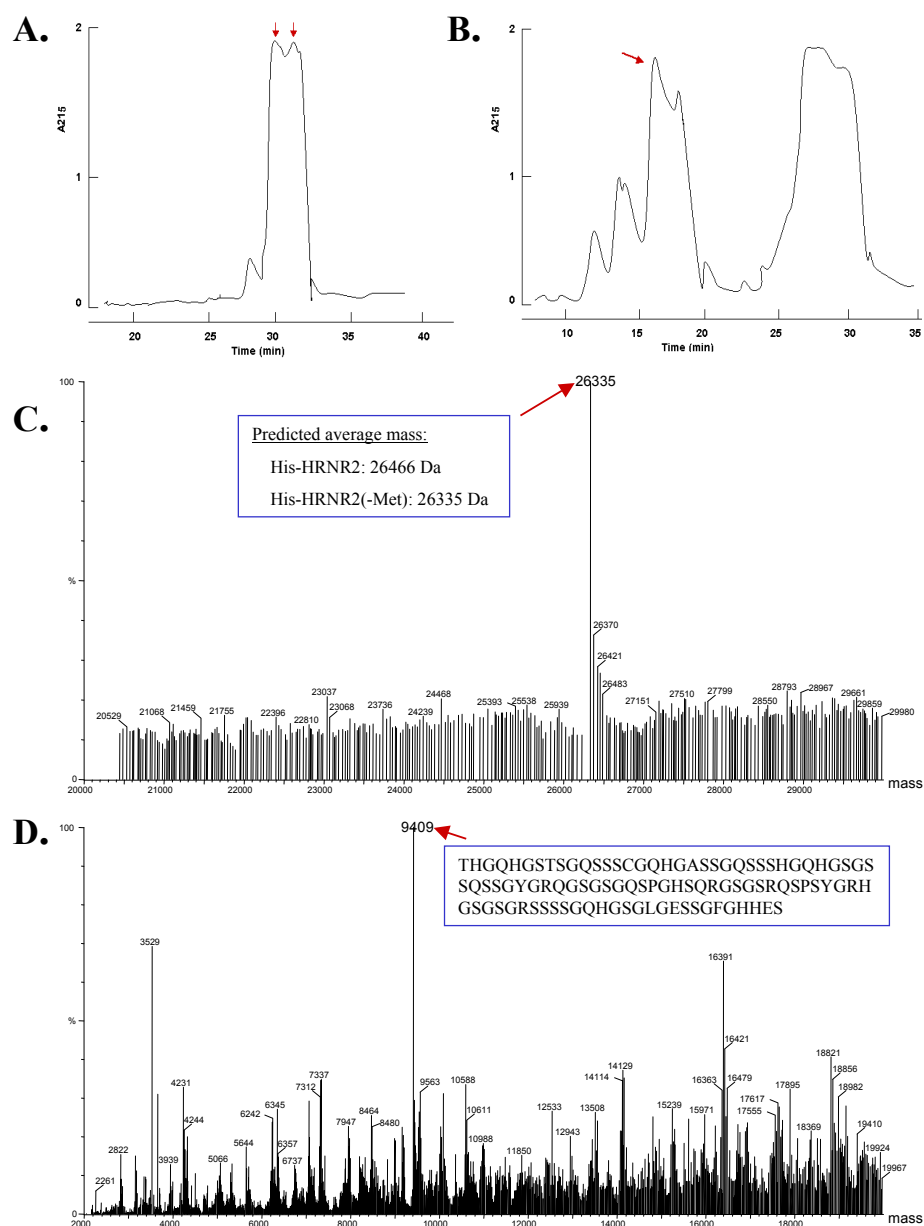


Figure 3.23: Purification and characterization of the hornerin fragment HRNR2. A, the Ni-column-bound His-tagged fusion proteins of *E. coli* extracts were separated by RP8-HPLC, and the fraction containing the highest peak was collected. The red arrow indicated the highest peak; B, the enterokinase-treated fraction was purified by RP4-HPLC, and the resulting peaks were collected and subjected to mass analysis. The red arrow indicates the peak with an ESI-MS spectrum shown in Fig. D; C, ESI-MS spectrum of the purified His-HRNR2 revealed an exact mass of 26335 Da. The red arrow indicates the mass corresponding to the predicted mass of His-HRNR2; D, ESI-MS spectrum of the purified recombinant HRNR2 revealed an exact mass of 9409 Da. The red arrow indicates the mass corresponding to the predicted mass of HRNR2. The amino acid sequence of HRNR2 is shown in the box.

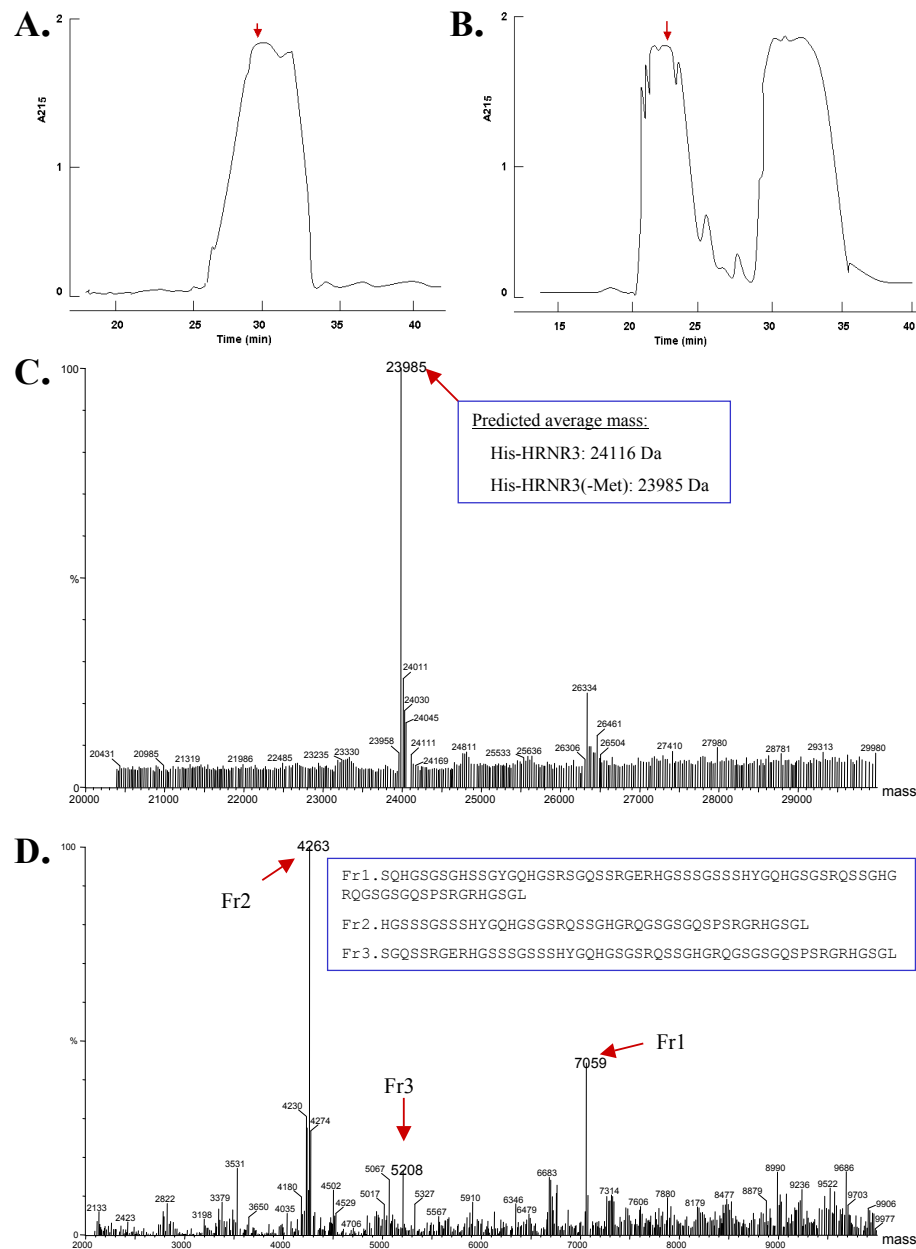


Figure 3.24: Purification and characterization of the hornerin fragment HRNR3. A, the Ni-column-bound His-tagged fusion proteins of *E. coli* extracts were separated by RP8-HPLC, and the fraction containing the highest peak was collected. The red arrow indicated the highest peak; B, the enterokinase-treated fraction was purified by RP4-HPLC, and the resulting peaks were collected and subjected to mass analysis. The red arrow indicates the peak with an ESI-MS spectrum shown in Fig. D; C, ESI-MS spectrum of the purified His-HRNR2 revealed an exact mass of 23985 Da. The red arrow indicates the mass corresponding to the predicted mass. D, ESI-MS spectrum of the purified recombinant HRNR3 revealed an exact mass of 7059 Da. The red arrow indicates the mass corresponding to the predicted mass. The amino acid sequences of HRNR3 fragments are shown in the box.



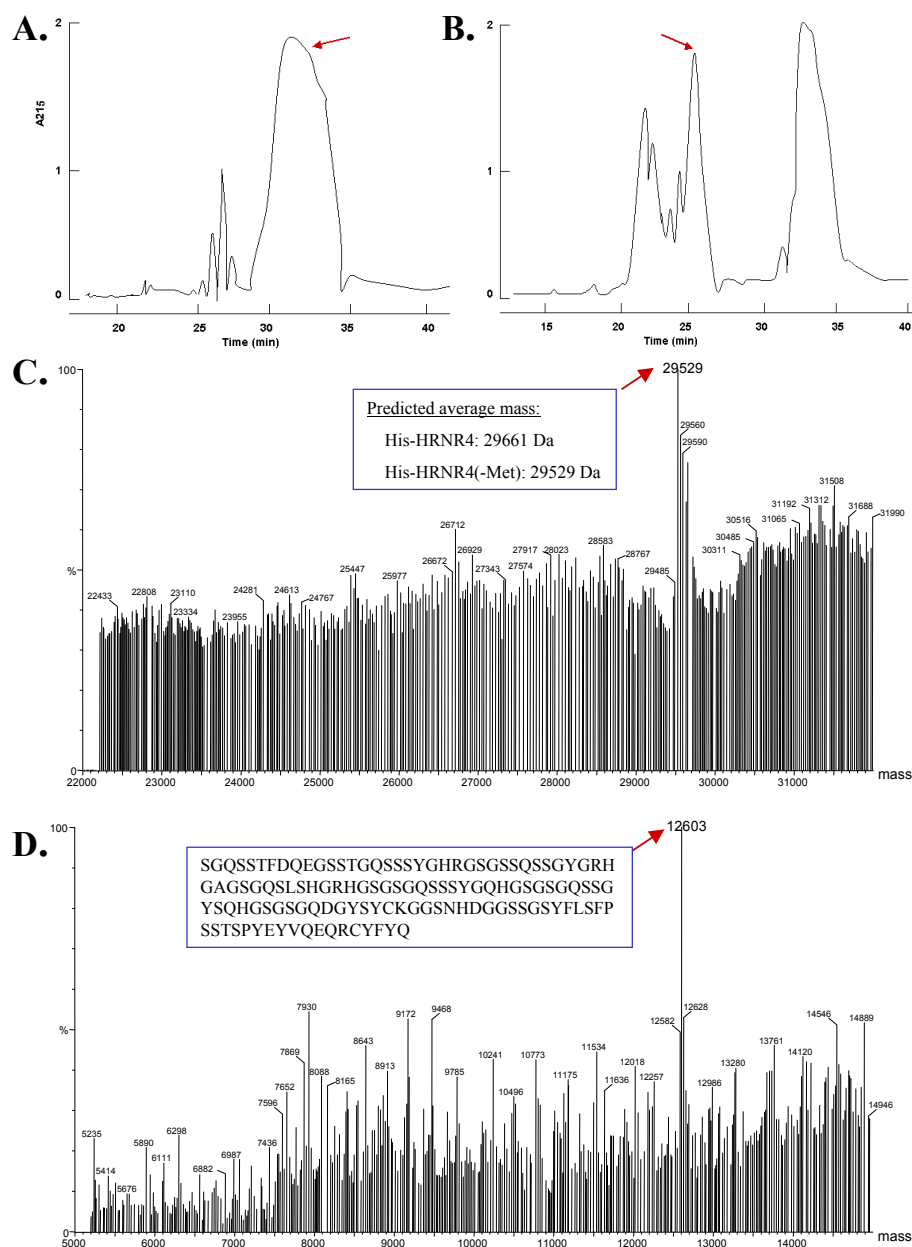


Figure 3.25: Purification and characterization of the hornerin fragment HRNR4. A, the Ni-column-bound His-tagged fusion proteins of *E. coli* extracts were separated by RP8-HPLC, and the fraction containing the highest peak was collected. The red arrow indicates the highest peak; B, the enterokinase-treated fraction was purified by RP4-HPLC, and the resulting peaks were collected and subjected to mass analysis. The red arrow indicates the peak with an ESI-MS spectrum shown in Fig. D; C, ESI-MS spectrum of the purified His-HRNR2 revealed an exact mass of 29529 Da. The red arrow indicates the mass corresponding to the predicted mass. D, ESI-MS spectrum of the purified recombinant HRNR4 revealed an exact mass of 12603 Da. The red arrow indicates the mass corresponding to the predicted mass. The amino acid sequence of HRNR4 is shown in the box.

The RP4-HPLC pattern confirmed the purity of the polypeptides. However, mass spectrometry returned multiple polypeptide mass peaks for HRNR2, HRNR3 and HRNR4. These results confirmed that HRNR2, HRNR3 and HRNR4 were purified in the flowthrough fraction. Moreover, the apparent fractions of HRNR2, HRNR3 and HRNR4 were not recognized in SDS-Page analysis by Coomassie staining (Fig. 3.26) or silver staining (data not shown).

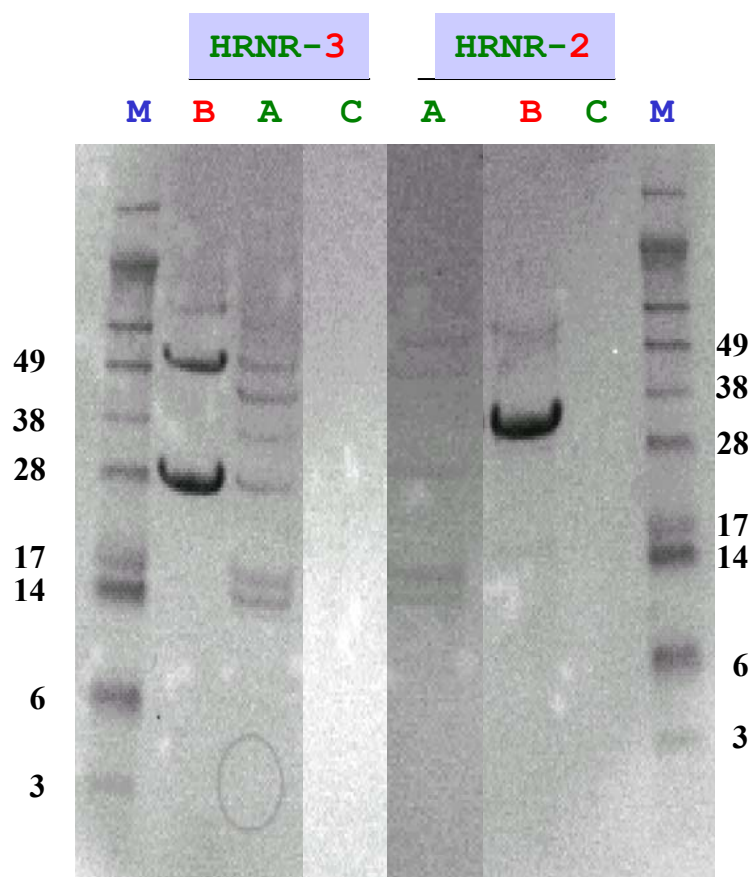


Figure 3.26: SDS-PAGE analyses of fusion proteins and cleaved mature polypeptides. Two hornerin-fragments-fusion proteins (HRNR2 and HRNR3) were analyzed. A, total proteins before induction; B, total proteins after 3 h induction of 1 mM IPTG; C, the RP4-HPLC purified polypeptides HRNR2 and HRNR3.

### 3.3.2 Purification and Characterization of recombinant LEKTI-2

The PCR fragment encoding the protein consisting of 61 amino acid residues of LEKTI-2 was subcloned into an expression vector (pET32a) to yield pET32a-LEKTI-2. The ex-

pression purification of recombinant fusion protein in *E. coli* BL216-LEKTI-2 fusion protein was subjected to an overnight cleavage process in the presence of enterokinase. The enterokinase-treated fractions were purified further by RP4-HPLC and were subjected to mass analysis (Fig. 3.27 B and D). Mass spectrometry revealed a single polypeptide mass of 7058 Da for LEKTI-2, in full agreement with the expected theoretical molecular mass (7064 Da) minus six hydrogens due to forming of three disulfide bonds (6 Da).

### 3.3.3 Analyses of protease activity of HRNR3

The complex ESI-MS spectra of hornerin fragments suggested an autoproteolytic process. To prove this hypothesis, the purified recombinant hornerin fragment HRNR3 was tested in 100 mM NH<sub>4</sub>HCO<sub>3</sub> solution at 37°C for 16 h. ESI-MS analyses revealed multiple polypeptide mass peaks, three of which were clearly mapped to two c-terminal sequences (4262 Da and 5207 Da) and one N-terminal sequence (6394 Da) of HRNR3, which were cleaved after the arginine residue (Fig. 3.28).

### 3.3.4 Analyses of antimicrobial activity of human hornerin fragments and LEKTI-2

Antimicrobial activities of HRNR2, HRNR3, HRNR4 and LEKTI-2 were assayed against the Gram-positive bacteria *Staphylococcus aureus* ATCC 6538 and *Enterococcus faecalis* ATCC 29212, the Gram-negative *Escherichia coli* ATCC 11303 and *Pseudomonas aeruginosa* ATCC 10145, and the yeast *Candida albicans* ATCC 24433 (Table 3.3), using the microdilution assay system, as described in the Materials and Methods section. In the antimicrobial assay system used, recombinant LEKTI-2 exerted antifungal activity against *C. albicans* ATCC 24433 with a LD 90 at 50 µg/ml, while recombinant hornerin fragments displayed activity, with degree of variation between the target microorganisms. HRNR2 and HRNR4 exhibited weak activity against the Gram-negative bacteria tested, but showed higher

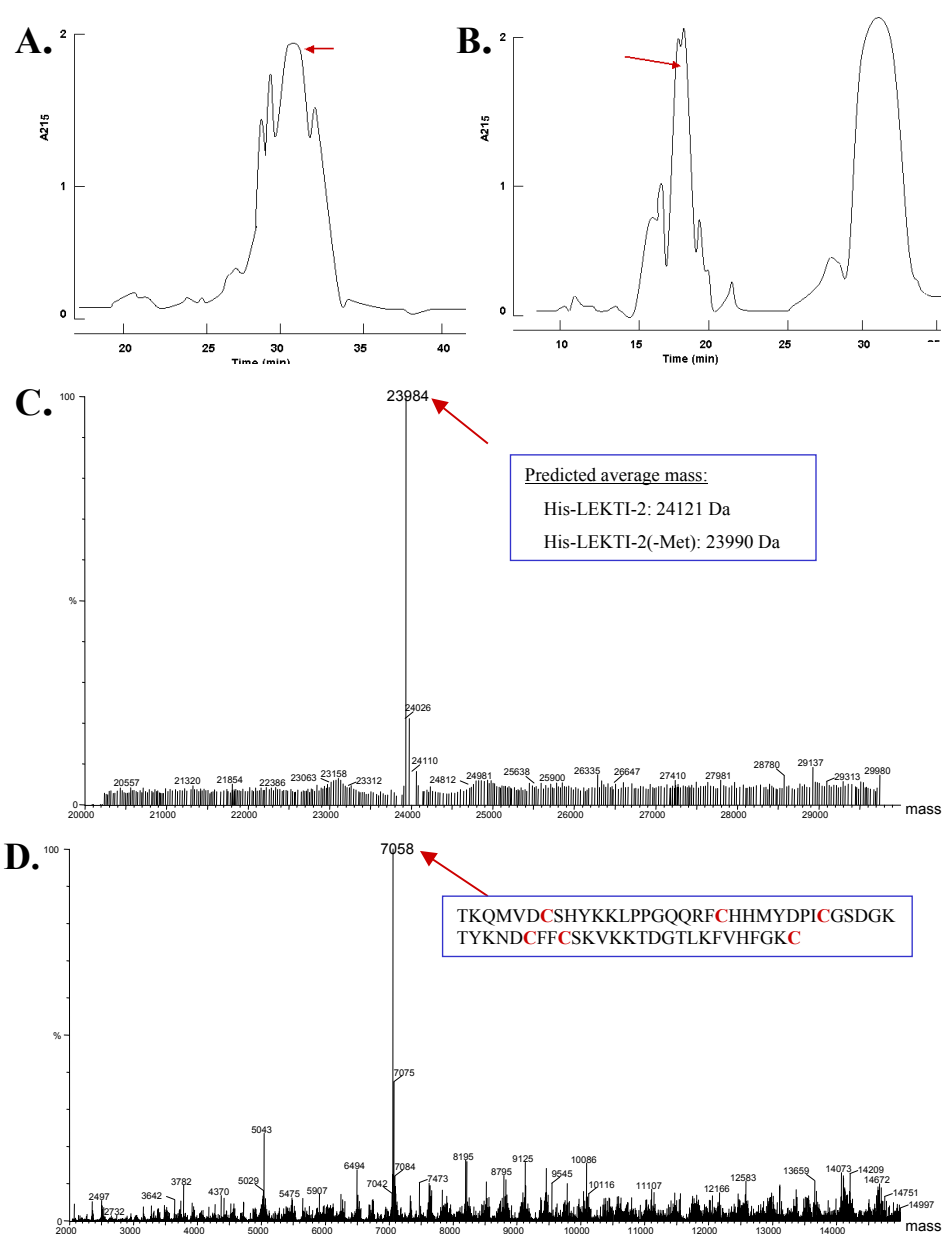


Figure 3.27: Purification and characterization of recombinant LEKTI-2. A, the Ni-column-bound His-tagged fusion proteins of *E. coli* extracts were separated by RP8-HPLC, and the fraction containing the highest peak was indicated by a red arrow and collected; B, the enterokinase-treated fraction was purified by RP4-HPLC, and the resulting peaks were collected and subjected to mass analysis. The red arrow indicates the peak with an ESI-MS spectrum shown in Fig. D; C, ESI-MS spectrum of the purified His-LEKTI-2 revealed an exact mass of 23984 Da. The red arrow indicates the mass corresponding to the predicted mass. D, ESI-MS spectrum of the purified recombinant LEKTI-2 revealed an exact mass of 7058 Da. The red arrow indicates the mass corresponding to the predicted mass. The amino acid sequence of LEKTI-2 was inserted.

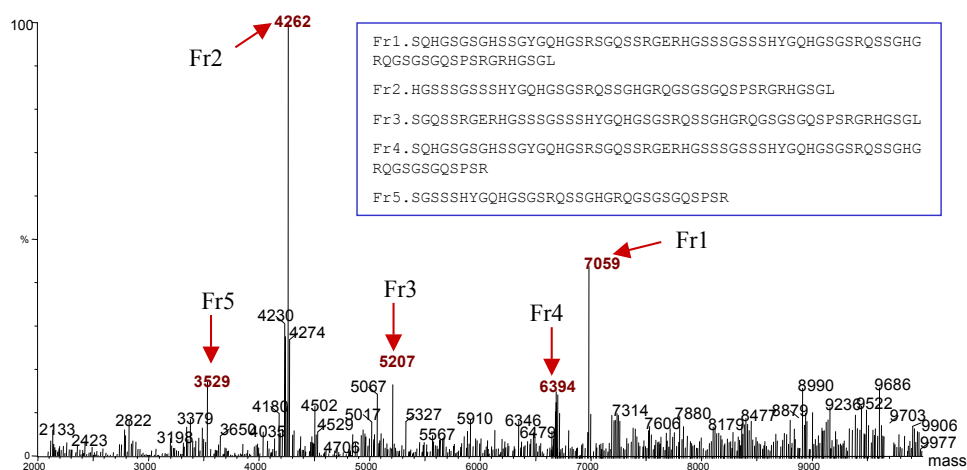


Figure 3.28: ESI-MS spectra of HRNR3 in 10 mM  $\text{NH}_4\text{HCO}_3$  buffer. The red arrow indicates the mass corresponding to the predicted mass. The amino acid sequences of HRNR3 fragments are shown in the box

activity against the yeast *C.albicans*. HRNR3 was almost 10 times more active against *E.coli* and *P.aeruginosa* than HRNR2 and HRNR4. The Gram-positive bacteria *S. aureus* ATCC 6538 and *E. faecalis* ATCC 29212 were resistant to the peptides.

Table 3.3: MBCs and LD90s of hornerin peptides and LEKTI-2. MBC, minimal bactericidal concentration (indicates 99.9% killing).

Organisms	value	HRNR2	HRNR3	HRNR4	LEKTI-2
<i>E. coli</i> ATCC 11303	MBC ( $\mu\text{g/ml}$ )	>100	>100	>100	>100
	LD90 ( $\mu\text{g/ml}$ )	>100	12.5	>100	>100
<i>P. aeruginosa</i> ATCC 10145	MBC ( $\mu\text{g/ml}$ )	>100	50	>100	>100
	LD90 ( $\mu\text{g/ml}$ )	100	6.25	>100	>100
<i>S. aureus</i> ATCC 6538	MBC ( $\mu\text{g/ml}$ )	>100	>100	>100	>100
	LD90 ( $\mu\text{g/ml}$ )	>100	>100	>100	>100
<i>E. faecalis</i> ATCC 29212	MBC ( $\mu\text{g/ml}$ )	>100	>100	>100	>100
	LD90 ( $\mu\text{g/ml}$ )	>100	>100	>100	>100
<i>C. albicans</i> ATCC 24433	MBC ( $\mu\text{g/ml}$ )	>100	25	>100	>100
	LD90 ( $\mu\text{g/ml}$ )	12.5	3.125	100	50

### 3.3.5 Generation of polyclonal goat antisera against hornerin fragments and LEKTI-2

Recombinant proteins (hornerin fragments HRNR2, HRNR3 and HRNR4 and LEKTI-2) were mixed with their corresponding His-tagged fusion protein and conjugated to maleimide-activated keyhole limpet hemocyanin (KLH). The conjugate was used in the immunization of goats. Thirteen days after the final immunization, blood from each goat was collected and the serum was separated from the clot.

The activity of the polyclonal goat antiserum was first tested by dot-blot assays using various purified antigens purified from RP-HPLC. 1 : 1000 and 1 : 3000 dilutions of each antiserum were tested with antigen amounts ranging from 0.1 ng to 1000 ng. As shown in Fig. 3.29, the anti-LEKTI-2 antiserum reacted with the LEKTI-2 fraction tested at both dilutions; however, the 1 : 3000 dilution was two dilution steps less sensitive. Though the anti-LEKTI-2 antiserum showed cross-reactivity with the HRNR2 fraction, a faint reaction was seen only with high amounts ( $\geq 30$  ng) of antigen and was clearly distinguishable from reactions that were still observed with 0.1 ng of LEKTI-2 antigen. Likewise, anti-HRNR2 antiserum gave a similar reaction pattern with the HRNR2 fraction and the LEKTI-2 fraction, but less than 1 ng of antigen was not detected at both dilutions.

The activity of the antisera against HRNR4 (Fig. 3.30) was tested using a dot blot assay with a variety of RP-HPLC fractions containing natural hornerin fragments, recombinant HRNR4 or recombinant LEKTI-2 at a serum dilution of 1 : 1000 or 1 : 3000. The sera gave a positive reaction with natural hornerin fragments from fraction 8 to 15 and recombinant HRNR4 at both dilutions; however, the 1 : 3000 dilution was only one dilution step less sensitive. The sera did not react with recombinant LEKTI-2.

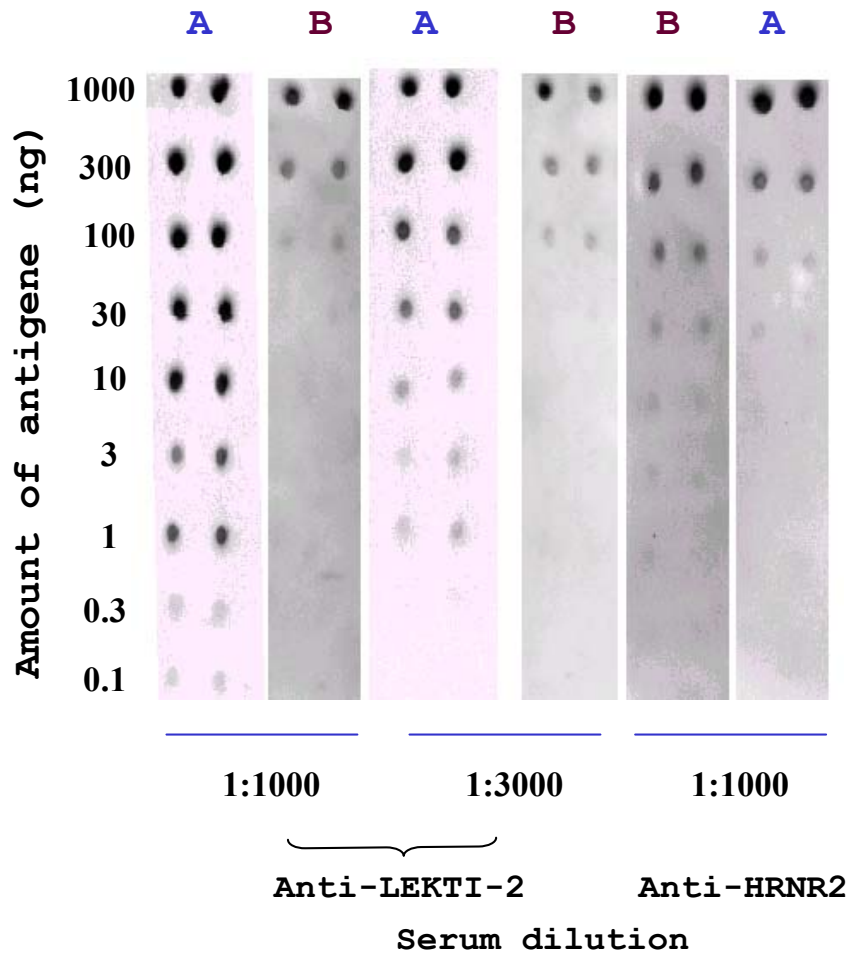


Figure 3.29: Titrations of goat antisera in dot-blot assays. Amounts of 0.1 to 1000 ng of LEKTI-2 (A) or HRNR-2 (B) were spotted onto nitrocellulose membranes and reacted with polyclonal goat antiserum against LEKTI-2 or HRNR2 at the indicated dilutions.

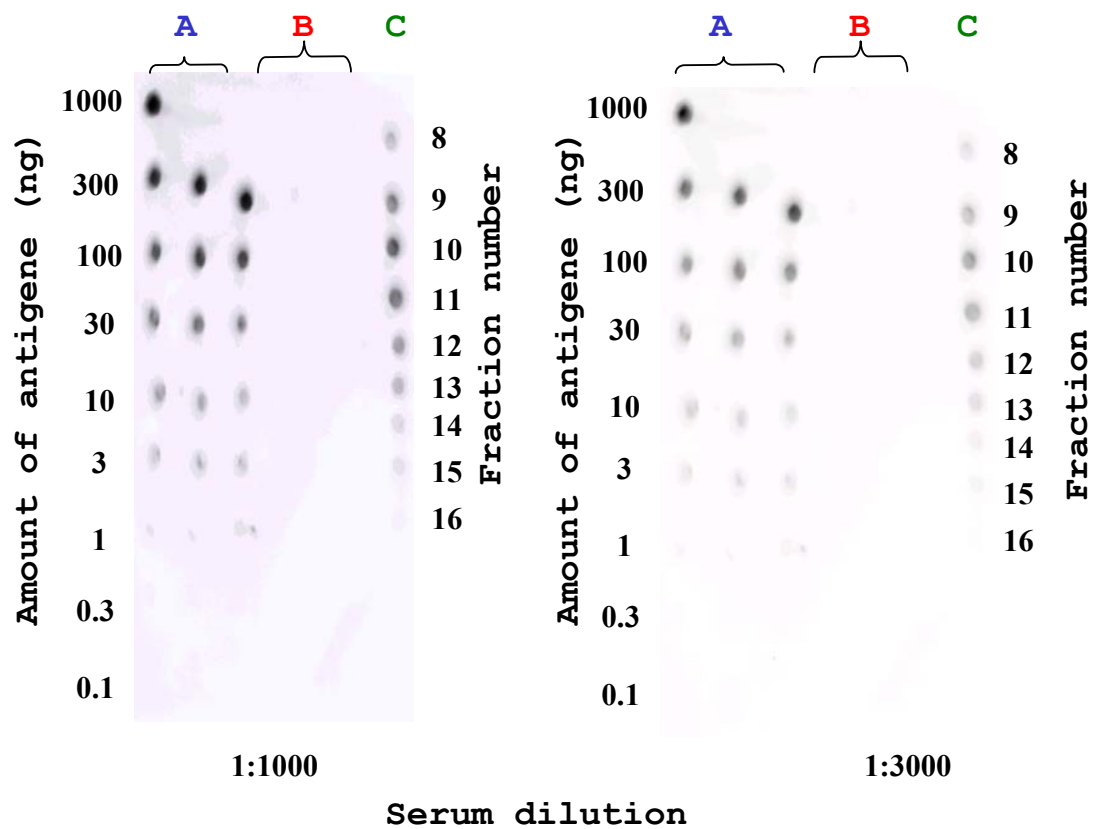


Figure 3.30: Titrations of goat anti-HRNR4 antisera in a dot-blot assay. Amounts of 0.1 to 1000 ng of recombinant HRNR4 (A) or LEKTI-2 (B) and 1  $\mu$ l of RP-HPLC fractions containing natural hornerin fragments (C) were spotted onto nitrocellulose membranes and reacted with polyclonal goat antiserum against HRNR4 at the indicated dilutions.



# Chapter 4

## Discussion

---

An initial clue about the function of the hornerin and LEKTI-2 genes was obtained by the previous work in our group, which revealed that the stratum corneum of healthy skin contained marked antimicrobial activity that was related to the presence of several antibacterial molecules. The data presented in this thesis elucidated their gene structures, determined their full-length cDNA sequences and characterized part of their protein functions. Remarkably, these genes appear to be constitutively present and belong to two multigene families, the SFTP gene family clustered in chromosome 1q21.3 and the SPINK gene family clustered in chromosome 5q32. Therefore, also included here is a complete dissection of both gene clusters and comparative analysis of their gene expression and their putative gene evolution.

It is widely accepted that the prevalence of gene families in vertebrates has important functional consequences, such as the frequent occurrence of unexpectedly weak mutant phenotypes due to redundancies between paralogues. An exhaustive characterization of the composition of the studied multigene families is a prerequisite for effective clarification of the gene function. This is currently made possible by the availability of an increasing number of sequenced genomes in vertebrates.

## 4.1 The human clustered S100 fused-type protein (SFTP) gene family

Using the public available *in-silico* tools, we found that the human SFTP gene family consists of seven members, including the previously characterized *profilaggrin* (*PFLG*) (Fleckman et al., 1985), *trichohyalin* (*THH*) (Lee et al., 1993), and *c1orf10* (Xu et al., 2000) and four so far uncharacterized genes, *THHL1*, *repetin* (*RPTN*), *hornerin* (*HRNR*) and *ifap-soriasin* (*IFPS*). Based on the retrieved DNA sequences, the 5' and 3' cDNA sequences of the four novel members were determined by RACE RT-PCR. Unfortunately, except for *THHL1*, the long-distance PCR failed to isolate the full-length cDNA sequences of the related genes. Sequence analysis indicated that each of the novel three genes (*RPTN*, *HRNR* and *IFPS*) contains many GC-rich repeats within the exon 3, which was a technique problem for all PCR-based methods and cDNA isolation techniques because high GC-containing templates generate frequently local secondary structure, leading to inefficient PCR (Filichkin and Gelvin, 1992). Then an alternative short-distance RT-PCR walking strategy was adopted to determine the full sequences. This technical approach taken would seem to be the only reliable way to identify the exact sequence of SFTP mRNAs, or similarly repetitive protein structures.

To clarify the grouping of family members, BLAT was used to identify SFTPs genomic sequences (Kent, 2002). Aligning mRNA and genomic DNA sequence revealed intron/exon structures. In addition, chromosomal locations of SFTP family members were also identified. Interestingly, the intron/exon structure of the SFTP genes is strictly conserved in the human chromosome 1q21.3: a small untranslated exon 1 followed by a small exon 2 containing the translation start site ATG and coding for the S100 domain and a large exon 3 coding the EF-hand domain and the rest part of the big proprotein (Fig. 3.4). They were closely clustered in a head-to-tail mode without any gene gap unrelated. Intron/exon structures can thus be used as a criterion for the identification of SFTP homologs in distant species.

To identify their primary protein structures, the amino acid sequences of the N-terminal

two EF-hand domains of the human SFTP proteins were compared with that of all the synteny in mouse and rat (Fig. 4.1). All SFTPs matched perfectly the consensus pattern for two domains: S-100/ICaBP type calcium binding protein signature (Kligman and Hilt, 1988) and EF-hand calcium-binding domain (PROSITE PDOC00018). In contrast, most parts of these proteins showed low-homology relationship (data not shown).

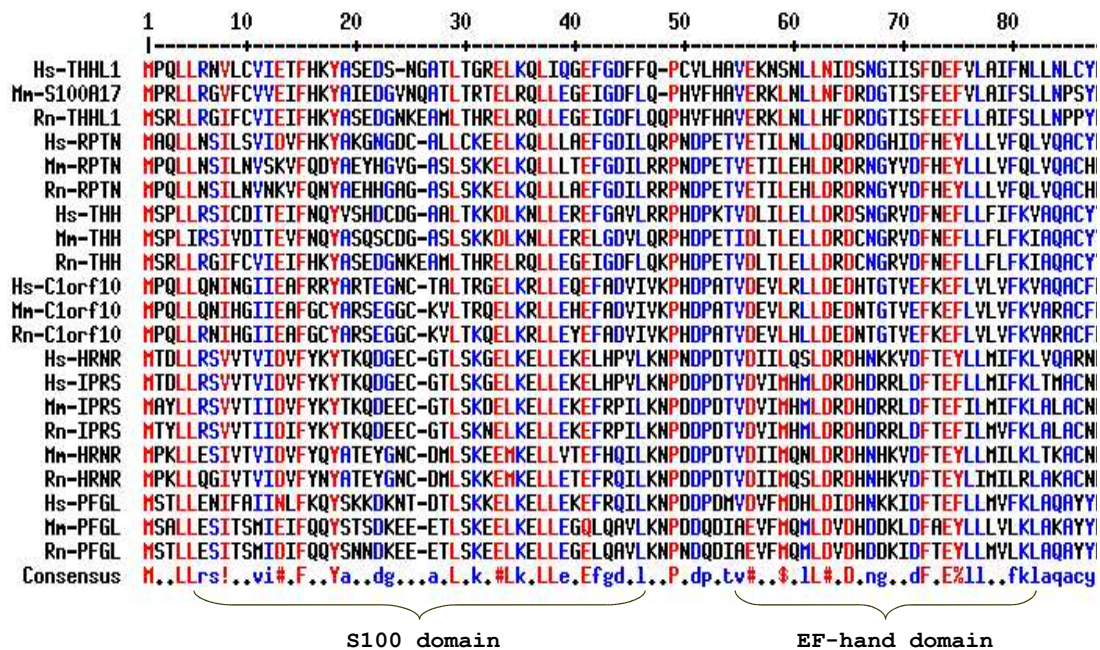


Figure 4.1: Multiple alignment of the conserved domains of the SFTP proteins. The amino acid sequences of the partner S100 and EF-hand domains have been truncated to improve the alignment. The conserved amino acids are shaded in red. Roman numerals at the top correspond to residue numbers. The domain pattern is indicated at the bottom. Hs, *homo sapiens*; Mm, *Mus musculus*; Rn, *Rattus norvegicus*.

To identify the *Mus musculus* SFTPs ortholog to human SFTPs, tBLASTN searches via the BLAT algorithm were carried out with the full-length sequences of the human SFTP proteins against the mouse and rat genome, respectively. The data obtained were contrasted with gene predictions from the mouse Genome Initiative. For all the genes (*THHL1*, *THH*, *RPTN*, *HRNR*, *PFLG*, *IFPS* and *C1orf10*) a genomic structure of coding exons was found to be totally conserved, including the gene order and the positions of the exon-intron boundaries. The mouse SFTP orthologs mapped to chromosome region 3F2 (Fig. 4.2).

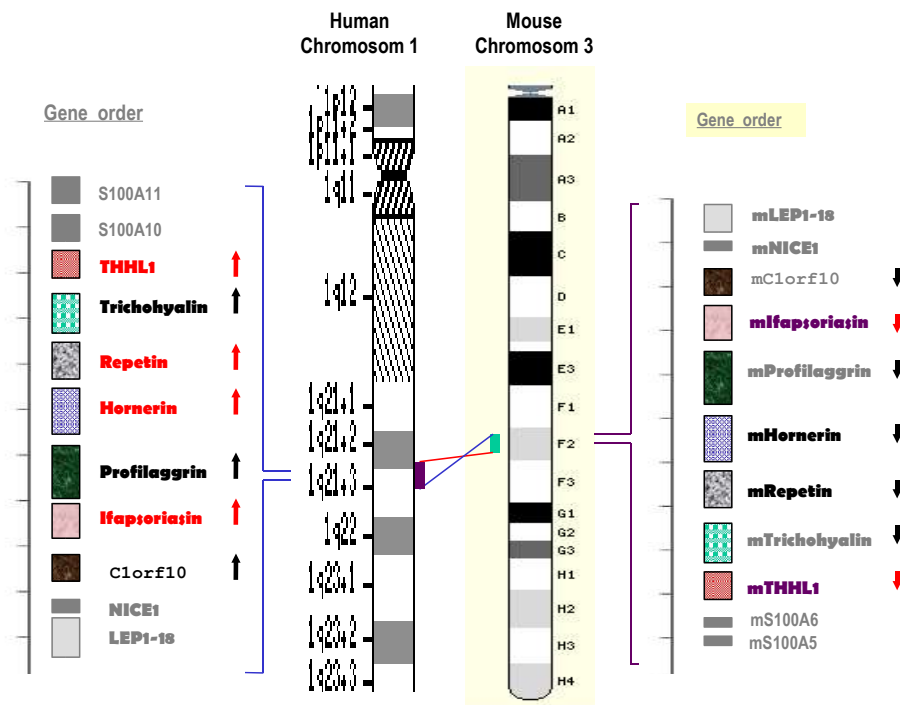


Figure 4.2: Comparative analysis of a cluster of human SFTPs at 1q21.3 with the corresponding region of mouse genome. The order and orientation of genes are indicated by black/red arrowheads. Orthology is indicated by connecting lines.

For a closer inspection of the relationship of all the protein sequences, their N-terminal 100-amino-acid sequences were aligned using DiAlign on the Genomatix server (<http://www.genomatix.de/cgi-bin/dialign/dialign.pl>) to the currently known S100A-type proteins selected from the NCBI protein database. The phylogenetic tree was generated from the sequence alignment and shown as an unrooted cladogram (Fig. 4.3). According to this tree, each human SFTP genes has an extremely conserved counterpart in chimpanzee (*Pan troglodytes*), mouse and rat. The tree can be primarily subdivided into several clades, distinguished by the sequence variability. The SFTPs protein sequences have clustered together into a single clade comprising all the 28 SFTPs isoforms distributed within the down half of the phylogenetic tree separately from all the human S100 sequences. This Clade contains two sub-branches, one including profilaggrin, hornerin and Ifaporsiasin,

another including repetin, C1orf10, trichohyalin and THHL1. In addition, hornerin and Ifapsoriasin proteins have grouped together separately from the profilaggrin, whereas THHL1 and trichohyalin have grouped together separately from repetin and C1orf10.

## 4.2 The human clustered kazal-type serine protease inhibitor (SPINK) gene family

A total of three members of the SPINK gene family have been previously characterized, which include *SPINK1* (Horii et al., 1987), *SPINK5* and *ECG2* (Cui et al., 2003). In this study five so far uncharacterized SPINK genes, *SPINK6*, *SPINK7*, *SPINK8*, *SPINK9* and *SPINK10*, were identified and found to be clustered within a very small region in chromosome 5q32. The genomic structure of the SPINK gene cluster is less conserved compared to that of the SFTP gene cluster. However, analysis from amino acid sequences revealed a high overall sequence homology and indicated that all eight SPINK members have an identical modular structure that was specific for kazal-type serine protease inhibitors. This included a short, but variable, N-terminal cytoplasmic tail, the putative signal anchor, and one-to-several C-terminal kazal domains. Except for the *SPINK1*, all the other SPINKs members showed the same transcriptional direction (Fig. 4.6).

To identify their primary protein structures, the amino acid sequences of the kazal domains of the human putative SPINKs identified at 5q32 were compared with the typical kazal type family members and *SPINK5* domain 2 and 15 (Fig. 4.4). All SPINKs matched perfectly the consensus pattern for kazal domains: C-x(7)-C-x(6)-Y-x(3)-C-x(2,3)-C (PROSITE PDOC00254), and therefore are homologous to PSTI (PDB code: 1tgs), acrosin inhibitor (PDB code: 2bus), ovomucoid (PDB code: 1ovo and 2ovo), and human *SPINK1* (PDB code: 1cgi). Besides the consensus kazal residues, SPINKs exhibit asparagine at position 36, a structurally important residue strongly conserved among serine protease inhibitors (Schechter and Berger, 1968). Interestingly, the *SPINK9* mature peptide is extremely basic, having a net charge of +9.13, while a seminal plasma inhibitor IIa from Cow (*Bos taurus*)(PDB code:2bus) has the next larger charge at +9.02 (Strop and Wuthrich,

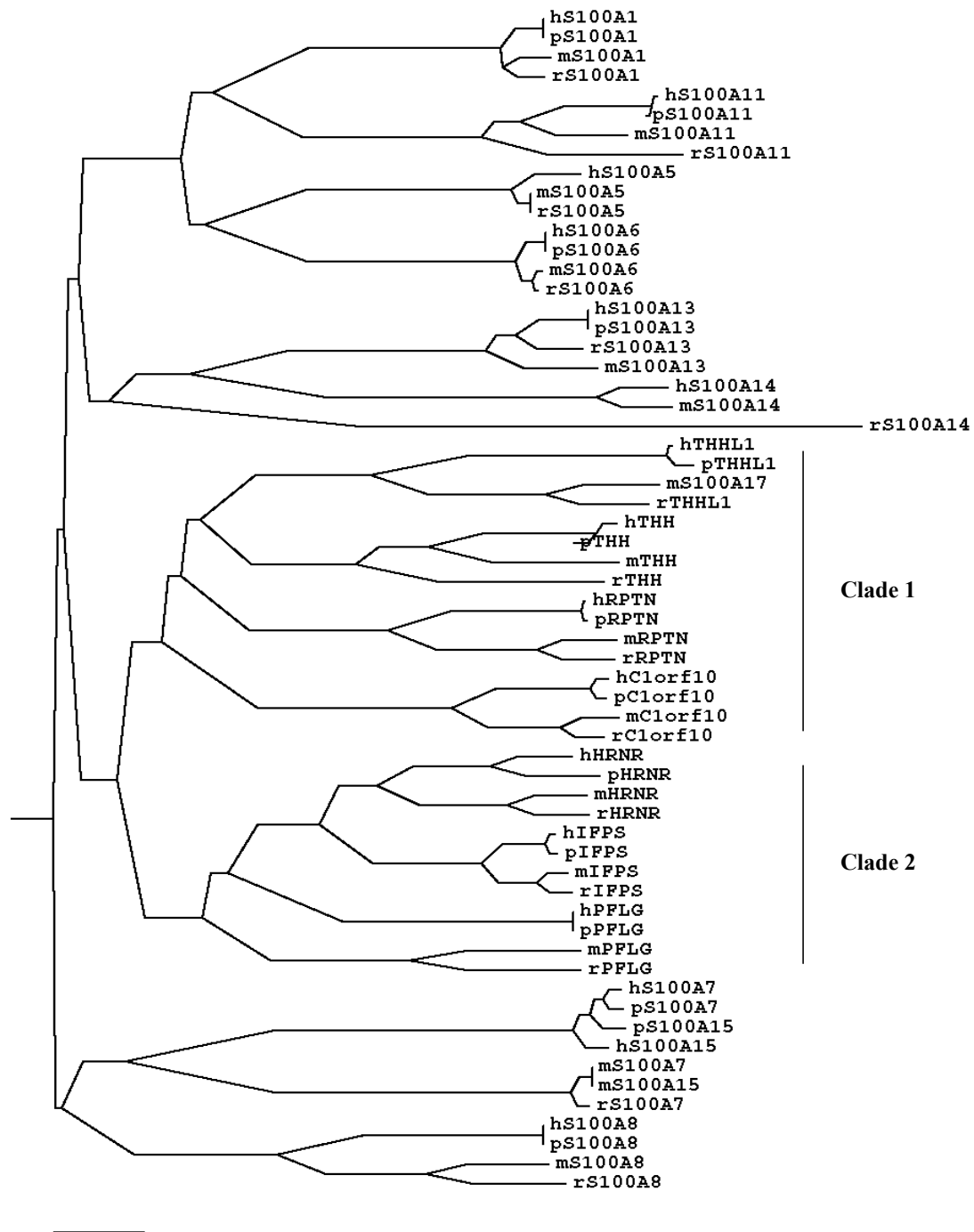


Figure 4.3: Unrooted cladogram of the human SFTP and S100A family and their orthologs in chimpanzee, mouse and rat. The tree, based on N-terminal 100 amino acid residues of each protein, shows the relationships among known S100 domains from SFTP and S100 families.

1983).

The neural network-based tool SignalP allows the discrimination of subcellular des-

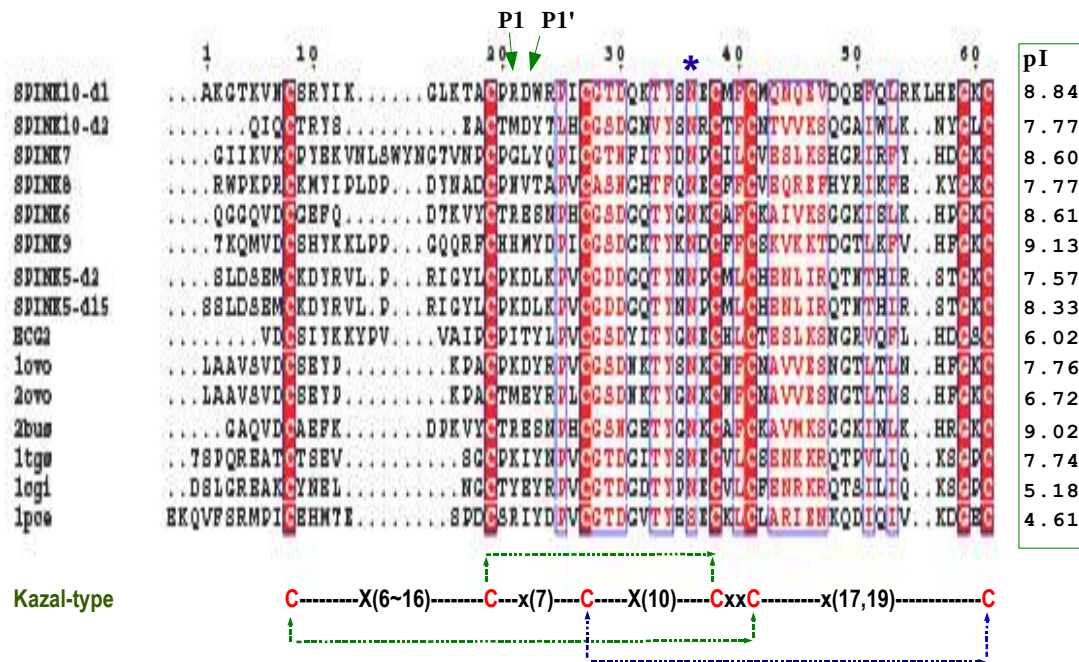


Figure 4.4: Multiple alignment of the kazal domains of the human SPINKs protein family. The partner kazal domain sequences have been truncated to improve the alignment. The conserved amino acids are shaded in red. The asterisk points to the conserved Asn residue. The two arrows above the alignment mark the active residue position P1 and P1'. Roman numerals at the top correspond to residue numbers. The disulfide pattern found in the kazal type is indicated at the bottom. The theoretical pIs (isoelectric points) are listed on the right-hand side.

tion of newly identified proteins based on the N-terminal sequence information only (Dyrlov Bendtsen et al., 2004). The five human SPINKs (SPINK6-v1, SPINK7-v1, SPINK8-v1, SPINK9 and SPINK10) contained a N-terminal signal targeting sequence but not any other signals for specific retrieval, retention, or commitment, so that they are predicted to be transported to the cell surface. Besides the signal peptide, there was also a predicted transmembrane (TM) helix located at the N-terminus by using the SOSUI algorithm (Mitaku et al., 2002). Therefore, the luminal protein of human SPINKs might be secreted constitutively to the extracellular space and their TM domains will reside at the plasma membrane. In contrast, SPINK7-v2 was distinguished by the absence of signal peptide or TM, which suggests SPINK7-v2 function as a soluble protein inside the cell. As shown as an example, Figure 4.5 showed the human SPINK9 protein was predicted to be secreted to the extracellular space and anchored at the plasma membrane with the TM trace sequence

(IECAKQ). The 6-residue TM anchor was cleaved further before SPINK9 became a mature protein.

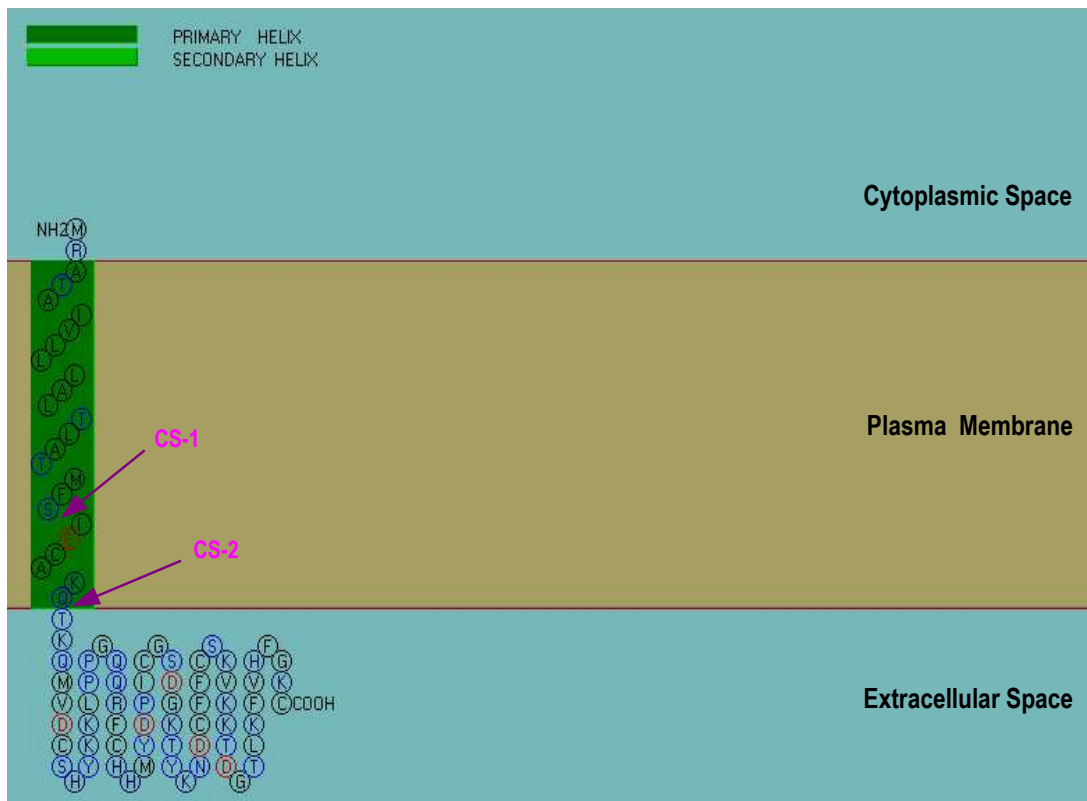


Figure 4.5: Schematic representation of subcellular destination of the human SPINK9. The schematic map was generated by the SOSUI algorithm. The subcellular destination was predicted by the TargetP program while the cleavage site of signal peptide (CS-1) was done by the SignalP program. The second arrow with CS-2 points to the second cleavage site between the raw and mature protein.

The SPINK inhibitors belong to a family of proteins which inhibit a number of serine proteases (such as trypsin and elastase). To date, 7 different mammalian SPINKs have been identified, of which 5 are found in humans. As shown above we identified 5 novel SPINKs at 5q32.

For the *Mus musculus* and *Rattus norvegicus* SPINKs orthologs, tBLASTN searches via the BLAT algorithm were carried out with the human SPINK kazal domains against the mouse and rat genome respectively. The data obtained were contrasted with gene predictions from the mouse and rat Genome Initiative. In most genes (SPINK1, SPINK3, SPINK5, SPINK6, SPINK7, SPINK8 and ECG2) a genomic structure of coding exons is



totally conserved, including the positions of the exon-intron boundaries. The mouse SPINK orthologs mapped to two separate chromosome regions from SPINK3 (SPINK3 to 18C and ECG2 to 18E1). Likewise, the rat SPINK orthologs mapped to two separate chromosome regions from SPINK1 (SPINK1 to 18p11, and ECG2 to 18q12.1) (Fig. 4.6). In contrast, the mouse and rat genome lacked two ortholog genes for human SPINK9 and SPINK10.

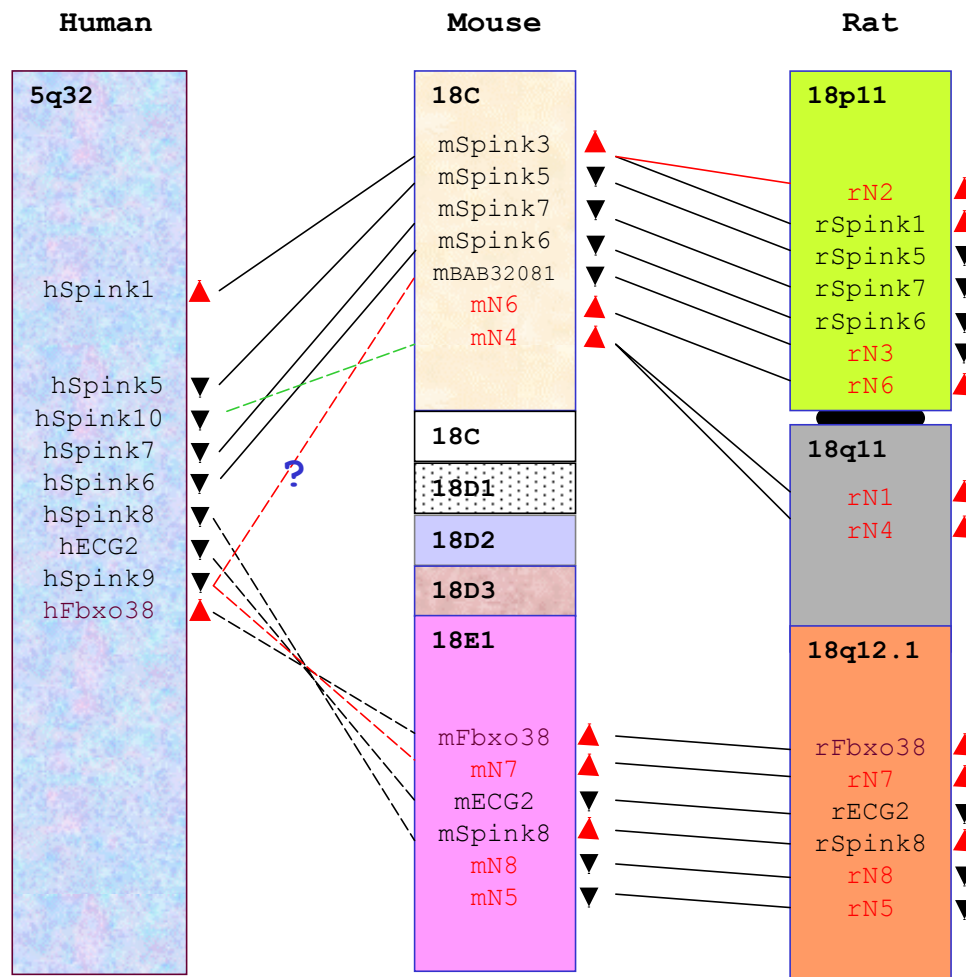


Figure 4.6: Comparative analysis of a cluster of human SPINKs at 5q32 with the corresponding region of mouse and rat genomes. The order and orientation of the genes are indicated by black/red arrowheads. Orthology is indicated by connecting lines.

For the *M. musculus* and *R. norvegicus* additional SPINKs homolog, tBLASTN searches were carried out with the human SPINK kazal domains against the mouse and rat genome on the Ensemble Genome Browser. The data obtained in the mouse and rat chromosome

18 were contrasted with gene positional identity and exon/intron conservation to human SPINKs. We predicted seven novel SPINKs at mouse 18C–18E1 and eight at rat 18p11–18q12.1. However, no recognizable mouse/rat ortholog to human SPINK9 or SPINK10 has been found in the corresponding syntenic region of mouse/rat chromosome. Mouse N7 (or rat N7) is localized in the putative ortholog site to human SPINK9 but they share only a low homology. As the ortholog of human SPINK10 is also present in pig, cow and dog, the gene represents a mammal-specific gene compared with rodents and a very recent duplication event in mammals leading to the generation of a gene (double-headed protease inhibitor) distantly related to SPINKs (Gururajan et al., 1998; Velasco et al., 1998).

For a closer inspection of the relationship of all the SPINKs protein sequences, they were aligned using DiAlign on the Genomatix server to the current 5 kazal-type motifs selected from the PDB database (Fig. 4.7). A single conserved domain became evident which is a kazal domain at the C-terminus and is encoded by one exon, which suggested that the SPINKs genes likely have been resulted from a gene duplication event.

To further explore the structural and evolutionary relationships between the newly identified SPINKs and other members of the kazal-type family, we next performed a computational phylogenetic tree analysis. The phylogenetic tree was generated from the sequence alignment seen above and shown as an unrooted cladogram (Fig. 4.8). According to this tree, most human SPINKs genes (SPINK1, SPINK2, SPINK4, SPINK5, SPINK6, SPINK7, SPINK8, SPINK9 and ECG2) have an extremely conserved counterpart in mouse and rat. The tree can be primarily subdivided into four clades, distinguished by the sequence variability within the kazal-domain. Clade 1 comprises 3 SPINKs isoforms (SPINK5, SPINK9 and ECG2) distributed within the upper half of the phylogenetic tree. Clade 2 comprises 2 SPINKs isoforms (SPINK7 and SPINK8) and expanded rodent N7. Clade 3 comprises 2 SPINKs isoforms (SPINK4 and SPINK6) and 4 expanded rodent SPINKs (N4, N5, N6 and N8) distributed within the down half of the phylogenetic tree. Clade 4 comprises 3 SPINKs isoforms (SPINK1, SPINK2 and SPINK10) and expanded rat N2.

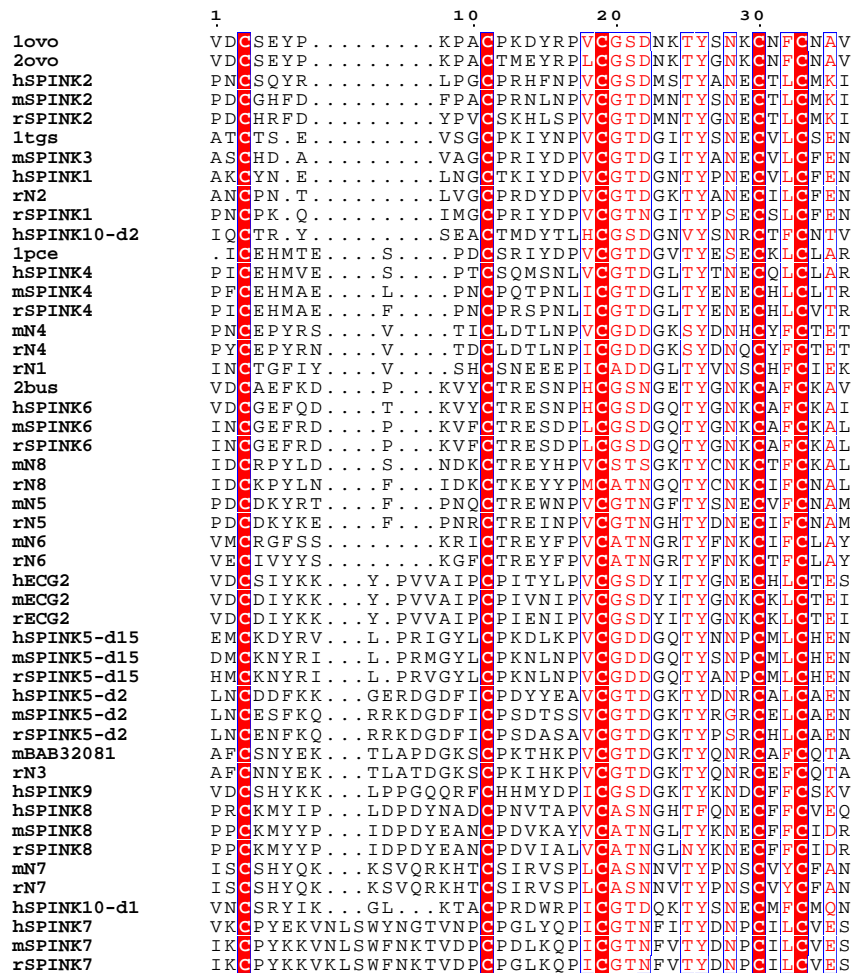


Figure 4.7: Multiple alignment of the kazal domains of the SPINKs protein family. The partner kazal domain sequences have been truncated to improve the alignment. Numbers in parentheses refer to the kazal-like region used for alignment if the sequence contains more than one repeat. The reserved amino acids are shaded in red.

### 4.3 Structural and functional implications for the novel SFTP/SPINK gene products based on comparative sequence analysis

The clustered organization of SFTP and SPINK genes provides a powerful opportunity to examine gene gain and loss in evolution because physical linkage is a key diagnostic feature which allows homology to be established unambiguously. The availability of several completed eukaryotic genome sequences allowed us to carry out large-scale searches for novel

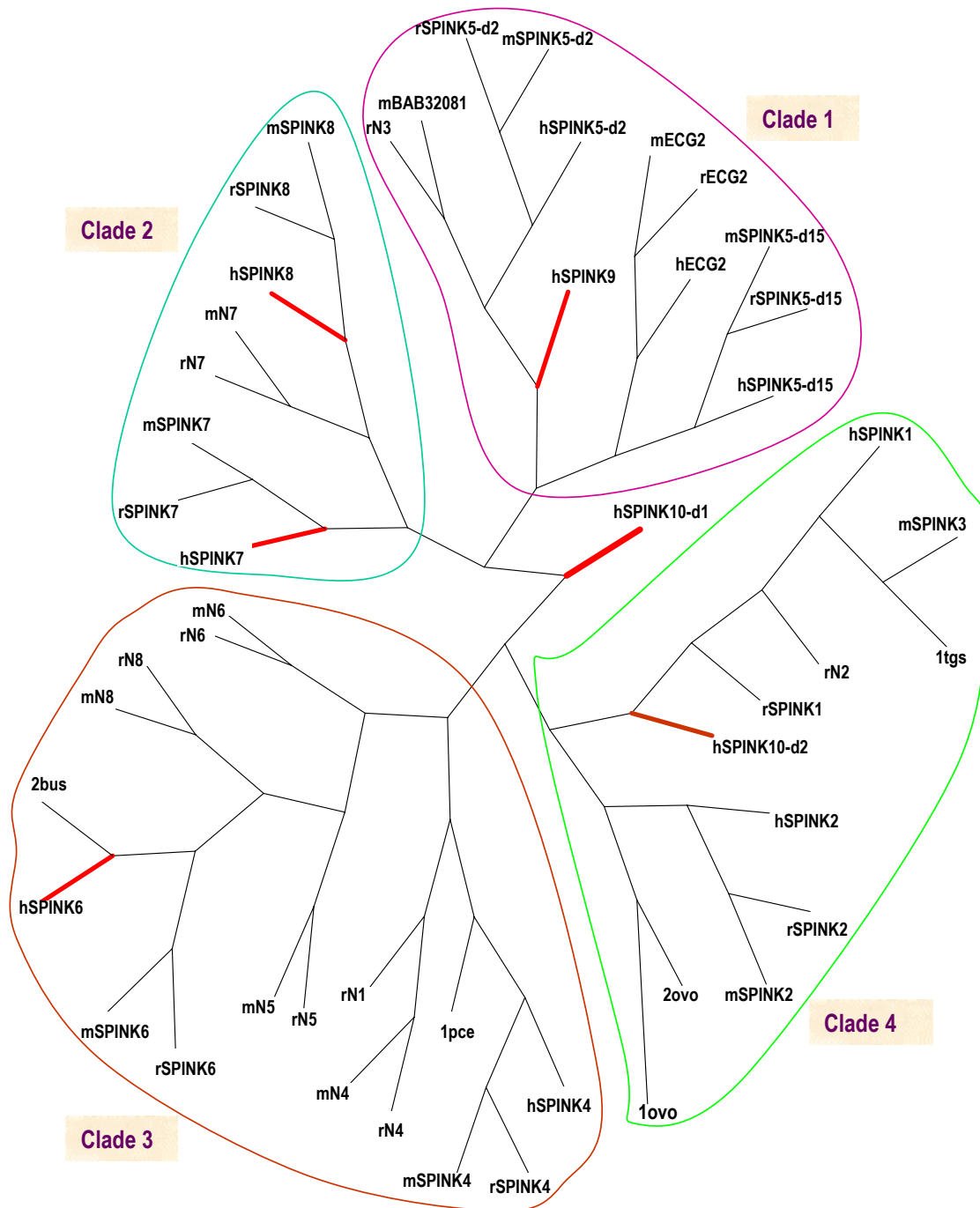


Figure 4.8: Unrooted cladogram of the human SPINK family and their ortholog in mouse and rat. The tree was subdivided into four clades.

members of previously identified protein families and to perform "phylogenomic" analyses, i.e. functional predictions via phylogenetic analysis. The premise of phylogenomics (Eisen, 1998) is that only the analyses carried out for complete sets of proteins obtained

from whole genomes allow to make accurate inferences of the duplication/speciation patterns, identification of orthologs and paralogs, and mapping on known functions onto the evolutionary tree (Zmasek and Eddy, 2001; 2002). Thereby, functions of uncharacterized genes can be predicted by their phylogenetic position relative to the characterized ones, and by analyzing the rates and patterns of gene evolution (Eisen, 1998; Eisen and Wu, 2002; Eisen and Fraser, 2003).

Based on the phylogenetic considerations supported by structure prediction and identification of characteristic conserved protein domain, all so far uncharacterized ortholog members of the SFTP/SPINK family in chimpanzee, mouse, rat and dog were predicted. In addition to such an ancient evolutionary origin, a multicistronic gene cluster sometimes results from horizontal transfer between species (Lawrence and Roth, 1996; Xu et al., 1998).

Fig. 4.2 summarized comparative chromosome localization of human and mouse SFTP genes. The human, mouse and rat (not shown) SFTP genes cluster at syntenic loci (5q32, 3F2 and 2q34, respectively). The relative positions and opposing orientations of human *hTHHL1*, *hTHH*, *hRPTN*, *hHRNR*, *hPFLG*, *hIFPS* and *hC1orf10* are matched in the mouse locus by *mTHHL1*, *mTHH*, *mRPTN*, *mHRNR*, *mPFLG*, *mIFPS* and *mC1orf10*, which confirms that human SFTPs and mouse SFTPs are orthologues. Thus the phylogenetic relationship of the SFTP genes, reflected in Fig. 4.3, is supported by the analysis of gene structure. Comparative genomic analysis, on the other hand, localized ortholog matches of the human EDC region (between *S100A11* and *S100A1*) in chimpanzee, mouse, rat and dog (Fig. 4.9). The EDC contains two highly conserved segments: segment A, between *S100A11* and *SPRR2C*, which contains the SFTP family; segment B, between *S100A9* and *S100A1*. The segment A is localized just before the segment B in human and chimpanzee, whereas the segment A is inverted and localized after the segment B in mouse and rat. In dog, both were apparently redistributed to two different chromosomes, the segment A being localized in chromosome 17 and the segment B being inverted and localized in chromosome 7. These data suggest the SFTP gene cluster be an extremely conserved gene family within mammals and that genomic identity and locations of the SFTP gene cluster date back as far as the dawn of mammals. This would suggest an early process of cluster nucleation (Glusman et

al., 2000). This was consistent with the notion that the cluster has undergone few species-specific rearrangements or segmental duplications in the past 15 million years (Bailey et al., 2004; Zhou and Mishra, 2005). There may be a basic function shared by all family members that has been co-opted in mammals to become part of the CCE structure.

Fig. 4.6 summarized comparative chromosome localization of human, mouse and rat SPINKs genes. In most genes (*SPINK1*, *SPINK3*, *SPINK5*, *SPINK6*, *SPINK7*, *SPINK8* and *ECG2*) a genomic structure of coding exons is totally conserved, including the positions of the exon-intron boundaries. The ortholog relationship was confirmed by phylogenetic analysis (Fig. 4.8). In contrast to the closely clustered genome structure of the human SPINKs at chromosome region 5q32, the mouse SPINKs orthologs mapped to two separate segment regions, *SPINK3* cluster to 18C and *ECG2* to 18E1. Likewise, the rat SPINKs orthologs mapped to two separate chromosome regions, *SPINK1* to chromosome region 18p11 and *ECG2* to chromosome region 18q12.1. However, no recognizable mouse/rat ortholog was detected for human *SPINK9* and *SPINK10* in the corresponding syntenic region, and moreover the mouse and rat SPINKs have showed expanded orthologs. One of these (mN7 or rN7) is localized at the putative genome site ortholog to human *SPINK9* but with a low homology and an inverted transcription direction. On the other hand, mouse BAB32081 seems ortholog to h*SPINK9* according to the tree. Nevertheless, the human *SPINK9* accumulated more mutations than the other family members in genome evolution so that it possibly gains new function. This would suggest that the human SPINK cluster displays a more complex evolutionary history.

Therefore, identification of the above two gene clusters confirms that selection can act on entire groups of genes, leading to joint transfers of genes between genomes (Koonin et al., 2001; Lawrence, 1997), concerted gene loss (Aravind et al., 2000), gene fusion events (Snel et al., 2000), co-regulation of genes through common regulatory elements (Pilpel et al., 2001), and the creation and maintenance of operons containing nonhomologous but co-transcribed genes (Lawrence, 2002; Lathe et al., 2000). In addition, it has also been confirmed conserved gene clusters accurately convey functional coupling between the genes present in them (Overbeek et al., 1999). Moreover, evolutionarily selected (and

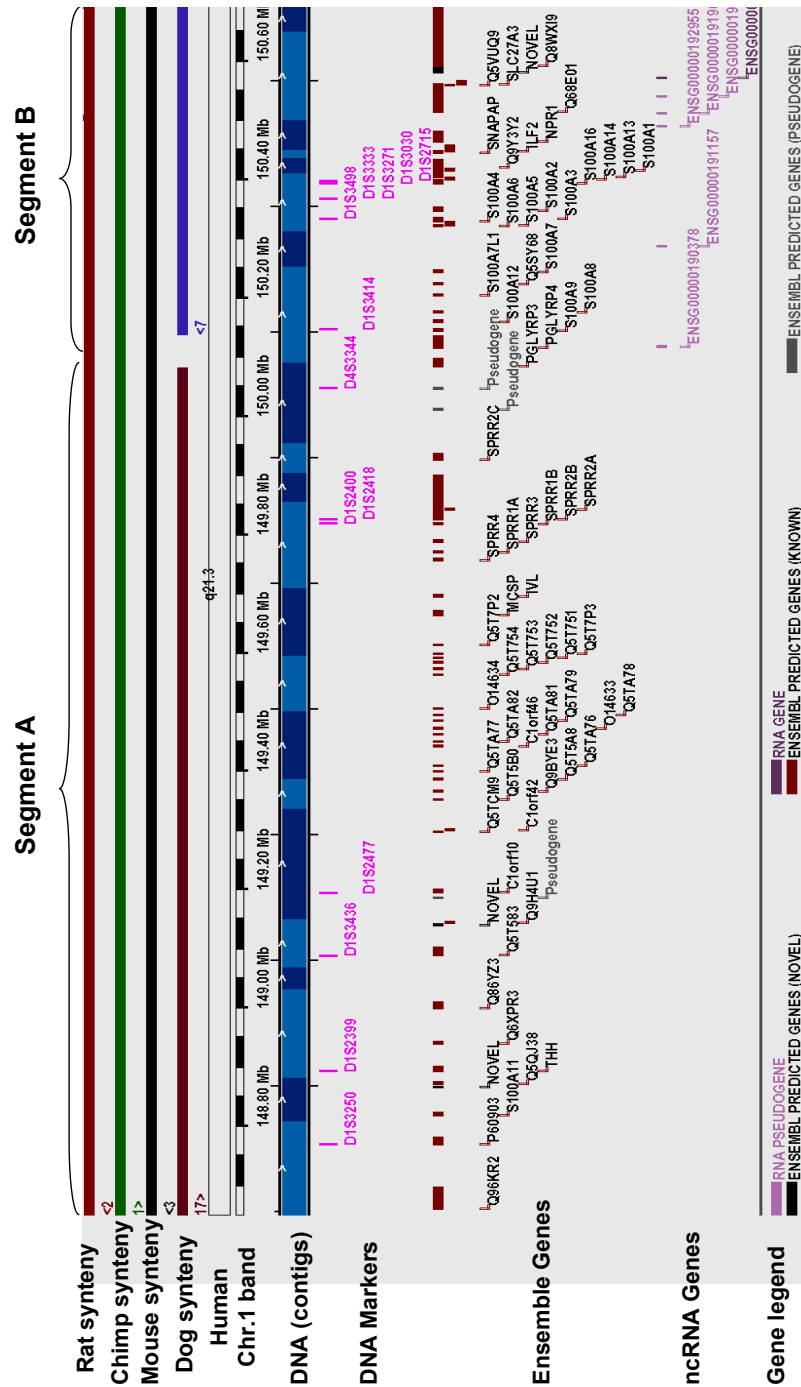


Figure 4.9: Comparative analysis of the human epidermal differentiation complex (EDC) at 1q21 with the corresponding region of chimpanzee, mouse, rat and dog genome. The human EDC region between *S100A11* and *S100A1* was used for ortholog analysis on Ensemble Genome Server. Orthology segment is indicated by parallel lines.

thereby, functional) gene clusters make a pave to be physically linked together (and thereby, co-regulated).

It is of great interest to compare the equivalent chromosomal regions in the genomes of different species and in particular of the mouse, for which many genetic aberrations have been mapped and it's much easier to perform a gene knock-out analysis. This may provide insights into the evolution of a gene family as well as help in the identification of candidate genes for human pathologies. Moreover, there are accumulated data suggesting that human 1q21 be associated with some skin tumors (Kimchi et al., 2005) and psoriasis (Giardina et al., 2004) while 5q32 be associated with the 5q-syndrome and atopy (Boulton et al., 2002; Marsh et al., 1994). The peak of linkage of AD and psoriasis on chromosome 1q21 overlies the human epidermal differentiation complex (EDC) (Mischke et al., 1996).

#### **4.4 The SFTP genes expression upregulated by cross-talking of the mitogen-activated protein kinases pathways**

The conserved clusters are likely to represent functionally coupled genes, such as those forming operon structures for co-expression and/or those encoding physically interacting protein subunits (Tamames et al., 1997; Dandekar et al., 1998; Overbeek et al., 1999). The proximal promoter of each of the SFTP genes has been analyzed in human and encompasses closely situated recognition motifs for transcription factors such as activator protein-1 (AP1) (Jang et al., 1996), Ets (Andreoli et al., 1997), POU domain (Jang et al., 2000) and TATA box binding protein (TBP) (Timmers et al., 1992), which are known to coordinately regulate the activity of the promoter by binding of particular members of each family to their respective sites (Jang et al., 2000) (Fig. 4.10).

In the present study, it was shown that SFTP gene members have a distinct tissue-specific expression, though all are expressed in keratinocytes. In cultured primary keratinocytes, the expression of SFTP genes was low during the proliferative phase and highly increased upon differentiation of the cells (Fig. 3.7). This is in agreement with all SFTPs



genes being late differentiation markers of human keratinocytes, whose mRNA levels respond to extracellular signals such as elevated calcium concentrations (Eckert and Welter, 1996). Notably, calcium-induced expression of the SFTP genes can be inhibited by either p38 inhibitors SB202190/SB203580 or the MEK inhibitor PD98059 (Fig. 3.9). Interestingly, the JNK inhibitor SP600125 can stimulate increased SFTPs expression at a high level though lower than these under calcium induction (Fig. 3.8). We concluded that the silence of SFTP promoter activity was not dependent on ERK or p38/MAP kinase but dependent on JNK. This may indicate that the SFTP genes are under a common regulator(s) during the keratinocyte differentiation. One could speculate that there are common constitutive factors that keep the SFTP gene family locus in a transcriptionally open position but only during a differentiating phase, and in addition to this there are tissue specific factors that apply a specific restriction on the SFTP gene.

#### **4.5 The SPINK7-v2: a chimeric protein encoded by two chromosomes**

The present study characterizes a novel human gene, SPINK7-v2, that is expressed in the bone marrow, keratinocytes and one case of precancerous lesions. Its full-length mRNA sequence was experimentally cloned from keratinocytes. A transcription variant of SPINK7-v2, SPINK5L2 (renamed to SPINK7-v1 here), has been recently identified in human and mouse (Puente and Lopez-Otin, 2004). The genomic structure of both variants explicitly demonstrates how the mRNA sequences are produced and further corroborates the accuracy of the two mRNA sequences. As shown in Fig. 3.18, the SPINK7-v1 mRNA sequence can be perfectly aligned with specific regions on the gene sequence, while the SPINK7-v2 mRNA sequence can be perfectly aligned with 2 complete exons on the gene sequence at 5q32 and a specific inverted region at chromosome 7q11.23. It is of particular interest that, although the sequence of the SPINK7-v1 mRNA was found in the GenBank database without a report of its detailed cloning and characterization, no sequences resembling the SPINK7-v2 have been reported, nor is it predicted by computational analysis.

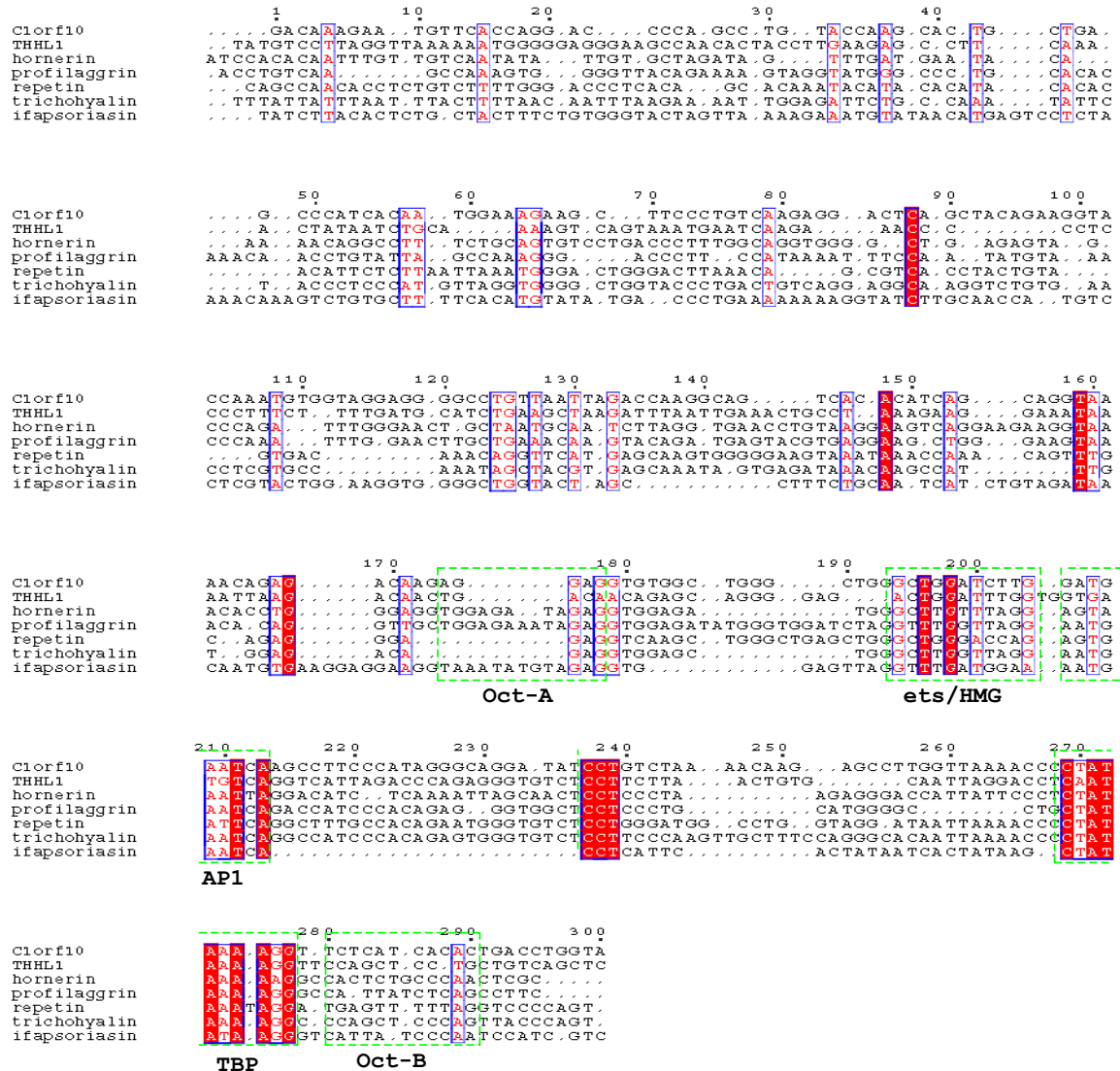


Figure 4.10: Comparison of the promoter structure of the different human SFTP genes. An alignment of the promoter region 300-bp upstream of the transcription start site of the eight characterized human SFTP gene products were performed by the T-coffee program. The transcription element effector sites important for epithelia were predicted by the TESS program and indicated by framed lines. The dots indicating gaps were inserted for optimal sequence alignment.

Thus this is the first identification of the accurate genomic structure of this chimeric gene.

A comparison of the protein-domain structure of all members of the human SPINK family, was performed to predict possible functions of the new members. All members display a highly conserved kazal domain, the essential components for serine protease inhibitors. The difference between the two isoforms of SPINK7 is the substitution of a signal-peptide

(SP) in the amino terminus of SPINK7-v2 which may result in the secretory disturbance. Thus it can be predicted that these two isoforms may have differential substrate specificities or other protein-protein interaction differences. It may also be that they are localized in different cell compartments. These predictions still need to be experimentally verified.

Transposable elements (TEs)-transposons and retroposons, are a major source of genetic change, including the creation of novel genes, the alteration of gene expression in development, and the genesis of major genomic rearrangements (Lozovskaya et al., 1995). Transposition of a TE into or near a particular host gene might impose novel developmental control on this host gene. Recent observations have shown that a number of sequences of known mobile elements were frequently inserted into the DNA of a gene regions and can influence the regulation of a gene's expression (Britten, 1997). A mobile element has also been found to be inserted in many unrelated genes in *Rickettsia conorii* suggesting the potential role of selfish DNA in the creation of new protein sequences (Ogata et al., 2000). In this study a human endogenous retrovirus sequence and an Alu sequence were found in the SPINK7-v2 transcript. Both of the retrovirus sequence and Alu sequence have replaced exons 1 and 2 of the gene *SPINK7* and generated a SPINK7-v2 mRNA with a 5' sequence derived from the TE. Vansant and Reynolds have observed that Alu sequences include functional binding sites for retinoic acid (RA) receptors (Vansant and Reynolds, 1995), suggesting that the SPINK7-v2 gene containing Alu sequence in its promoter might be involved in diverse and yet interconnected biological processes, such as embryogenesis, growth, and differentiation (De Luca, 1991; Ross et al., 2000).

#### **4.6 The *SPINK9* product: LEKTI-2**

The family of kazal-type protease inhibitors (SPINKs) inhibit serine peptidases of the S1 family, such as trypsin and elastase (Rawlings et al., 2004). During the last decade there has been intensive investigation into the biological effects of SPINKs. To date, three members of this family: SPINK1, ECG2 and SPINK5/LEKTI have been identified at chromosome 5q32. SPINK1 acts as the first line of defence against prematurely activated trypsinogen

by inhibiting approximately 20% of total trypsin activity within the pancreas (Horii et al., 1987). SPINK1 has been found to be associated with chronic pancreatitis (Chen et al., 2000; Witt et al., 2000). ECG2 may play an important role in the carcinogenesis of esophageal cancer (Cui et al., 2003). LEKTI is a protease inhibitor protein and is encoded by the gene *SPINK5* (Mägert et al., 1999). LEKTI regulates proteolysis in epithelia formation and keratinocyte terminal differentiation, and its defects lead to overdesquamation of corneocytes and skin barrier dysfunction (Komatsu et al., 2002). Several mutations in the *SPINK5* gene can cause Netherton syndrome, a rare autosomal recessive disorder characterized by defective cornification of the skin and TH2-skewed immunologic alterations resembling atopic dermatitis (AD) (Chavanas et al., 2000). Recently it has been revealed that LEKTI deficiency causes abnormal desmosome cleavage in the upper granular layer through degradation of desmoglein 1 due to stratum corneum tryptic enzyme and stratum corneum chymotryptic enzyme-like hyperactivity (Descargues et al., 2005). This leads to defective stratum corneum adhesion and resultant loss of skin barrier function. This work implies that SPINKs might be key regulators of epidermal protease activity.

The present study dissects the SPINK cluster at chromosome 5q32. This result is the first step toward comprehensive characterization of the putative genes for skin diseases. As a good example, this study characterizes the novel gene *SPINK9* and its mature protein. It is proposed to be a secreted protein. The full-length protein of *SPINK9* is autoprocessed to form the mature protein LEKTI-2 by removing the first 25 amino acids. This autocatalytic cleavage results in a mature LEKTI-2 protein containing a complete kazal domain sequence highly homologous with the domains 2 and 15 of LEKTI. Determination of potential P1 and P1' sites as shown in Fig. 4.4 confirms the supposed function of the protein as a serine proteinase inhibitor. Like LEKTI, LEKTI-2 is also highly expressed in the thymus, which suggests the possibility of a comparable function in the regulation of the maturation of T-lymphocytes.

As yet, the biological function of LEKTI-2 is unclear. One SPINK protein isolated from the skin of *Phyllomedusa sauvagii* (PSKP-1) is described to exhibit antibacterial properties (Gebhard et al., 2004), which suggests that SPINKs might also be involved in innate immunity.

In this study recombinant LEKTI-2 exerts only antifungal activity (Table 3.3), whereas the natural mature LEKTI-2 possesses *E. coli*-killing activity (Fig. 1.1). This difference may result from the wrong folding of recombinant LEKTI-2 in *E. coli* cells. Further investigations are now necessary to determine the target proteinase of LEKTI-2, its function within the thymus and stratum corneum, the inhibitory potency of the Kazal-type domain, and to clarify the pathophysiological role of LEKTI-2.

## 4.7 The hornerin

*Hornerin*, as one member of the SFTP gene family, was first found to be expressed in mouse stratified epithelium, including the epidermis, tongue, esophagus, and forestomach (Makino et al., 2001). Mouse hornerin is colocalized with profilaggrin in keratohyalin granules in keratinocytes in the granular layer (Makino et al., 2003). The present study characterizes the *hornerin* ortholog in human by RT-PCR and phylogenetic analysis. Human hornerin was not detected in esophagus (data not shown) but in tongue and forestomach. Unexpectedly, human hornerin is expressed in bone marrow and neutrophils because other SFTP members are not detected there.

The full-length protein of hornerin is autoprocessed to form an array of repeated mature peptides, one of which has been isolated from healthy skin. This autocatalytic cleavage results in numerous peaks in ESI-MS RP-HPLC analysis of natural hornerin-containing stratum corneum extracts from skin (data not shown). This process was also confirmed by recombinant protein analysis shown in Fig. 3.28, suggesting that human hornerin contain trypsin-like protease activity. The recombinant polypeptides display antimicrobial activity to Gram-negative bacteria and fungi, especially the HRNR3 whose amino acid sequence was deduced from the natural mature peptide. HRNR3 showed much higher activity than the other recombinant hornerin fragments in this study (Table 3.3). Detection of microbial activity of hornerin was very surprising, because hornerin is generally viewed as a structural proprotein involved in formation of CCE (Makino et al., 2001). Thus, these data indicate that human hornerin appears to be proteolytically processed *in vivo* to give

rise to antibiotic activity against several microorganisms.

## **4.8 The polyclonal goat antisera against hornerin fragments and LEKTI-2**

Recombinant LEKTI-2 and hornerin-fragments were used to immunize goats. The high activity of the resulting goat antisera against hornerin fragments and LEKTI-2 was confirmed by dot-blot assays using recombinant as well as natural antigens purified by reversed-phase-HPLC. These experiments demonstrated that it is possible to generate polyclonal antibodies against hornerin fragments and LEKTI-2. These new generated antibodies have been found to be an useful tool to detect LEKTI-2 and hornerin fragments in HPLC-fractions of human skin extracts (data not shown). Further investigations are necessary to verify the specificity of the generated antisera (e.g. by using western-blot analysis). Specific antibodies against hornerin fragments and LEKTI-2 would help to investigate protein expression and to determine whether gene expression correlates with protein expression. If the antibodies might work in immunohistochemistry, analysis of the cellular expression pattern of hornerin fragments and LEKTI-2 in various tissues would be possible.

## Abstract

---

Human body surfaces are defended by epithelia, which provide the initial physical barrier against potential harmful microorganisms. Epithelia also provide chemical barriers to microbial colonization including low pH, hydrolytic enzymes, and defense molecules such as antimicrobial peptides (AMPs). The number of reports demonstrating the presence and upregulation of AMPs in human skin is increasing and reflects the significance of these peptides in cutaneous innate immunity. This thesis has performed an exhaustive characterization of two coherent gene clusters, S100 fused-type proteins (SFTP) and kazal-type serine protease inhibitors (SPINK), with a part of the effort to elucidate the molecular mechanism of innate immunity in the skin. Included is expression analyses of various novel genes identified in primary keratinocytes and generation of goat antisera against recombinant hornerin fragments and recombinant LEKTI-2, respectively.

SFTP proteins comprise a conserved family of intermediate filament-associated proteins involved in keratinocyte terminal differentiation. Three human SFTP genes have been identified so far. Here the identification of four additional SFTP members (*THHL1*, *Repetin*, *Hornerin* and *Ifapsoriasin*) is reported. All the seven genes were identified as a cluster localized within the epidermal differentiation complex (EDC) on chromosome 1q21.3, a region of shared synteny with mouse chromosome 3F2 and rat chromosome 2q34. Several points of evidence, including a strong similarity in protein structure, a conserved linear gene order and the positions of the exon-intron boundaries, suggest that the SFTP clusters in human, mouse and rat are orthologous and share a common conserved function. The maintenance of one cognate SFTP gene cluster during vertebrate evolution suggests that their genomic organization should be important for the regulation, expression and function

of SFTP genes.

Serine protease inhibitors kazal type (SPINKs) are a group of small secreted proteins, which play an important role in the pathology of diseases such as arthritis, cancer, multiple sclerosis and cardiovascular diseases. To obtain fundamental information about the human SPINK gene cluster, the genomic structure of 5q32 was analyzed. Five novel members of SPINK genes were identified by *in-silico* analysis and RT-PCR and characterized by sequence homology, conserved kazal domain, a clustered localization and gene expression on keratinocytes. In contrast to the conserved SFTP genes, the SPINK genes showed extensive species differences, including expansions of gene subsets, gene deletions, exon shuffling generating a hybrid gene (*SPINK7*) and accelerated accumulation of mutations in the kazal domain.

To better understand the biology and function of both gene clusters, patterns of SFTP/SPINK gene expression in human foreskin derived primary keratinocytes, human adult tissues, human skin tumors and other diseases were studied. Both gene clusters were expressed in keratinocytes but in a different range of tissue types. Genes of the SFTP cluster were found to have distinct patterns that correlated with variation in diseased lesions. Their mRNA expression was activated upon calcium stimulation resulting in keratinocyte differentiation, and were found to be under upstream positive induction of the ERK/p38 MAPK signal transduction pathway and under negative upstream induction of the JNK pathway, respectively.

Using recombinant expression of hornerin fragments as fusion proteins, surprisingly it was found that the hornerin proprotein was processed by an unknown mechanism (possible self digestion) into multiple novel antimicrobial peptides. This indicated that hornerin not only might function as an intermediate filament-associated structural protein but also might play multiple roles in the regulation of human skin flora as a precursor molecule of antimicrobial peptides. These peptides may help limit infection by potential pathogens in the first few hours following bacterial colonization.

Further generation of recombinant *SPINK9* gene product, LEKTI-2, revealed so far only low antimicrobial activity. It is suggested that the native peptide could be modified by



posttranslational actions, which may enhance antimicrobial activity.

The *SPINK7* gene product appears to be a putative chimeric fusion protein encoded by two chromosomes and is expressed in bone marrow, skin tumor (precancerous lesions) and in cultured keratinocytes. Though the function remains to be characterized, *SPINK7* will be a useful model to investigate the molecular mechanism leading to the expression of one gene derived from different chromosomes.

## Zusammenfassung

---

Die Körperoberflächen des Menschen werden durch die Epithelien geschützt, welche eine erste physikalische Barriere gegen potentiell schädliche Mikroorganismen darstellen. Die Epithelien bieten als weiteren Schutz vor mikrobieller Besiedlung auch eine chemische Barriere, welche durch einen niedrigen pH-Wert, hydrolytische Enzyme und Abwehrmoleküle, wie zum Beispiel antimikrobielle Peptide (AMPs), gekennzeichnet ist. Die zunehmende Anzahl an Veröffentlichungen über das Vorkommen und die Induktion von AMPs in der Haut des Menschen spiegelt die Bedeutung von diesen Peptiden für die angeborene Abwehr („innate immunity“) der Haut wieder. In dieser Arbeit wurde eine eingehende Untersuchung der zwei zusammenhängenden Gen-Cluster von „S100-fused type proteins“ (SFTP) und Serin Protease-Inhibitoren vom Kazal-Typ („kazal-type serine protease inhibitors, SPINK) durchgeführt, um weitere Einblicke in die molekularen Mechanismen der angeborenen Abwehr der Haut zu erhalten. Untersuchungen zur Expression verschiedener in primären Keratinozyten neu identifizierter Gene wurden durchgeführt und Ziegen-Antiseren gegen rekombinante Hornerin-Fragmente und rekombinantes LEKTI-2 hergestellt.

SFTP Proteine gehören zur konservierten Familie der „intermediate filament-associated proteins“, einer Gruppe von Proteinen, die in die terminale Differenzierung der Keratinozyten involviert sind. Bis jetzt wurden drei humane SFTP-Gene identifiziert. In dieser Arbeit wird die Identifikation von vier weiteren SFTP-Mitgliedern (*THHL1*, *Repetin*, *Hornerin* und *Ifapsoriasis*) beschrieben. Alle sieben SFTP-Gene liegen zusammen auf einem Gen-Cluster des epidermalen Differenzierungs-Komplexes („epidermal differentiation complex“, EDC) auf Chromosom 1q21.3, einem chromosomalen Abschnitt mit Ähnlichkeit zu Maus-Chromosom 3F2 und Chromosom

2q34 der Ratte. Eine hohe Ähnlichkeit der Protein-Struktur, eine konservierte lineare Gen-Anordnung und die Exon-Intron-Abfolge legen nahe, dass die SFTP-Cluster im Menschen, Maus und Ratte ortholog sind und eine gemeinsame konservierte Funktion aufweisen. Die Erhaltung eines verwandten SFTP-Gen-Clusters während der Evolution der Vertebraten legt die Vermutung nahe, dass die genomische Organisation der SFTP-Gene im SFTP-Cluster für deren Expression, Regulation und Funktion von wichtiger Bedeutung sein könnte.

Serin Protease-Inhibitoren vom Kazal-Typ (SPINKs) sind eine Gruppe von kleinen sezernierten Proteinen, welche eine wichtige Bedeutung für die Pathologie von Arthritis, Krebs, Multipler Sklerose und kardiovaskulären Erkrankungen besitzen. Um detaillierte Informationen über das SPINK-Gen-Cluster zu erhalten, wurde die genomische Struktur von 5q32 analysiert. Fünf neue Mitglieder der SPINK-Gen Familie konnten durch *in-silico*-Analyse und RT-PCR identifiziert werden. Diese Gene zeichnen sich durch eine hohe Sequenz-Homologie, durch eine konservierte Kazal-Domäne, durch eine geclusterte Lokalisation und durch eine Expression in Keratinozyten aus. Im Gegensatz zu den konservierten SFTP-Genen besitzen die SPINK-Gene große Spezies-abhängige Unterschiede einschließlich einer Expansion bestimmter Genabschnitte, Gen-Deletionen, Exon-Verschiebungen und Entstehung eines Hybrid-Gens (*SPINK7*) und einer erhöhten Häufigkeit von Mutationen in der Kazal-Domäne.

Um die funktionelle und biologische Rolle des SFTP- und SPINK-Gen-Clusters näher zu charakterisieren, wurde die Expression der SFTP- und SPINK-Gene in humanen Vorhaut-Keratinozyten und adulten Geweben einschließlich erkrankten Geweben, wie zum Beispiel Haut-Tumoren, analysiert. Es zeigte sich, dass beide Gen-Cluster in Keratinozyten exprimiert werden, aber in anderen Geweben ein unterschiedliches Expressionsprofil aufweisen. Insbesondere die Expression der Gene des SFTP-Clusters korrelierte mit der Ausprägung von erkrankten Läsionen. Die mRNA-Expression der SFTP-Gene wurde im Zuge der durch erhöhte Kalzium-Konzentrationen ausgelösten Keratinozyten-Differenzierung aktiviert und durch den ERK/p38 MAP-Kinase-Signalweg positiv reguliert und durch den JNK-Signalweg negativ reguliert.

Die Expression von rekombinanten Hornerin-Fragmenten ergab, dass das Hornerin-Protein überraschenderweise durch noch unbekannte Mechanismen (möglicherweise durch autokatalytischen Verdau) in mehrere neue antimikrobielle Peptide prozessiert wird. Dies zeigt, dass Hornerin möglicherweise nicht nur als Struktur-Protein (als ein „intermediate filament-associated Protein“) fungiert, sondern auch eine Bedeutung für die Regulation der menschlichen Hautflora durch die Bereitstellung antimikrobieller Peptide haben könnte. Solche antimikrobiellen Peptide könnten dazu beitragen, schon in den ersten Stunden nach bakterieller Besiedlung die Entstehung von Infektionen zu verhindern.

Die Verwendung von rekombinantem Material des *SPINK9*-Genprodukts LEKTI-2 ergab bisher nur eine niedrige antimikrobielle Aktivität. Es wäre möglich, dass das native Peptid aufgrund von postrationalen Modifikationen eine gesteigerte antimikrobielle Aktivität aufweist.

Das *SPINK7* Genprodukt stellt ein putatives chimäres Fusionsprotein dar, dessen Information auf zwei Chromosomen kodiert ist. *SPINK7* wird im Knochenmark, Haut-Tumoren (präkanzerösen Läsionen) und in kultivierten Keratinozyten exprimiert. Obwohl die Funktion von *SPINK7* noch charakterisiert werden muß, so könnte es doch ein nützliches Modell darstellen, um die molekularen Mechanismen zu analysieren, die zur Expression eines von zwei verschiedenen Chromosomen abstammenden Gens führen.

## References

---

- Ahrens, D., Koch, A.E., Pope, R.M., Stein-Picarella, M. and Niedbala, M.J. (1996) Expression of matrix metalloproteinase 9 (96-kd gelatinase B) in human rheumatoid arthritis. *Arthritis Rheum*, **39**, 1576-1587.
- Akira, S. and Sato, S. (2003) Toll-like receptors and their signaling mechanisms. *Scand J Infect Dis*, **35**, 555-562.
- Algermissen, B., Sitzmann, J., LeMotte, P. and Czarnetzki, B. (1996) Differential expression of CRABP II, psoriasin and cytokeratin 1 mRNA in human skin diseases. *Arch Dermatol Res*, **288**, 426-430.
- Andreoli, J., Jang, S., Chung, E., Coticchia, C., Steinert, P. and Markova, N. (1997) The expression of a novel, epithelium-specific ets transcription factor is restricted to the most differentiated layers in the epidermis. *Nucl. Acids Res.*, **25**, 4287-4295.
- Andreu, D. and Rivas, L. (1998) Animal antimicrobial peptides: an overview. *Biopolymers*, **47**, 415-433.
- Aravind, L., Watanabe, H., Lipman, D.J. and Koonin, E.V. (2000) Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *PNAS*, **97**, 11319-11324.
- Avrameas, S. and Ternynck, T. (1969) The cross-linking of proteins with glutaraldehyde and its use for the preparation of immunoadsorbents. *Immunochemistry*, **6**, 53-66.
- Baden, H.P., Roth, S.I., Goldsmith, L.A. and Lee, L.D. (1974) Keratohyalin protein in disorders of keratinization. *J. Invest. Derm.*, **64**, 411-414.
- Bailey, J.A., Church, D.M., Ventura, M., Rocchi, M. and Eichler, E.E. (2004) Analysis of Segmental Duplications and Genome Assembly in the Mouse. *Genome Res.*, **14**, 789-801.
- Bals, R. (2000) Epithelial antimicrobial peptides in host defense against infection. *Respiratory Research*, **1**, 141 - 150.
- Bardan, A., Nizet, V. and Gallo, R.L. (2004) Antimicrobial peptides and the skin. *Expert Opinion on Biological Therapy*, **4**, 543-549.
- Barton, G. and Medzhitov, R. (2002) Toll-like receptors and their ligands. *Curr Top Microbiol Immunol*, **270**, 81-92.
- Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L.L., Studholme, D.J., Yeats, C. and Eddy, S.R. (2004) The Pfam protein families database. *Nucl. Acids Res.*, **32**, D138-141.
- Bechinger, B., Zasloff, M. and Opella, S.J. (1993) Structure and orientation of the antibiotic peptide magainin in membranes by solid-state nuclear magnetic resonance spectroscopy. *Protein Sci*, **2**, 2077-2084.
- Bitoun, E., Chavanas, S., Irvine, A.D., Lonie, L., Bodemer, C., Paradisi, M., Hamel-Teillac, D., Ansai, S.-i., Mitsuhashi, Y., Taieb, A., de Prost, Y., Zambruno, G., Harper, J.I. and Hovnanian, A. (2002) Netherton Syndrome: Disease Expression and Spectrum of SPINK5 Mutations in 21 Families. *J Invest Derma*, **118**, 352-361.
- Blum, H., Beier, H. and Gross, H.J. (1987) Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis*, **8**, 93-99.
- Bohlmann, B.J. (1999) Treatment of cat scratch disease. *N Engl J Med*, **340**, 1842.
- Bohlmann, H., Vignutelli, A., Hilpert, B., Miersch, O., Wasternack, C. and Apel, K. (1998)

- Wounding and chemicals induce expression of the Arabidopsis thaliana gene *Thi2.1*, encoding a fungal defense thionin, via the octadecanoid pathway. *FEBS Letters*, **437**, 281-286.
- Boman. (1998) Gene-Encoded Peptide Antibiotics and the Concept of Innate Immunity: An Update Review. *Scand J Immuno*, **48**, 15-25.
- Boman, H.G. (1995) Peptide antibiotics and their role in innate immunity. *Annu. Rev. Immunol.*, **13**, 61-92, .
- Boni, R., Burg, G., Doguoglu, A., Ilg, E.C., Schafer, B.W., Muller, B. and Heizmann, C.W. (1997) Immunohistochemical localization of the Ca<sup>2+</sup> binding S100 proteins in normal human skin and melanocytic lesions. *Br J Dermatol*, **137**, 39-43.
- Bottomley, K.M., Johnson, W.H. and Walter, D.S. (1998) Matrix metalloproteinase inhibitors in arthritis. *Enzyme Inhib J*, **13**, 79-101.
- Boulwood, J., Fidler, C., Strickson, A.J., Watkins, F., Gama, S., Kearney, L., Tosi, S., Kasprzyk, A., Cheng, J.-F., Jaju, R.J. and Wainscoat, J.S. (2002) Narrowing and genomic annotation of the commonly deleted region of the 5q- syndrome. *Blood*, **99**, 4638-4641.
- Brade, L., Vielhaber, G., Heinz, E. and Brade, H. (2000) In vitro characterization of anti-glucosylceramide rabbit antisera. *Glycobiology*, **10**, 629-636.
- Briand, J.P., Muller, S. and Van Regenmortel, M.H. (1985) Synthetic peptides as antigens: pitfalls of conjugation methods. *J Immunol Methods*, **78**, 59-69.
- Britten, R.J. (1997) Mobile elements inserted in the distant past have taken on important functions. *Gene*, **205**, 177-182.
- Broome, A.M., Ryan, D. and Eckert, R.L. (2003) S100 protein subcellular localization during epidermal differentiation and psoriasis. *J Histochem Cytochem*, **51**, 675-685.
- Bulet, P., Hetru, C., Dimarcq, J.L. and Hoffmann, D. (1999) Antimicrobial peptides in insects; structure and function. *Dev Comp Immunol*, **23**, 329-344.
- Butler, D., Yasuda, L. and Yao, M. (1995) An intramolecular recombination mechanism for the formation of the rRNA gene palindrome of *Tetrahymena thermophila*. *Mol. Cell. Biol.*, **15**, 7117-7126.
- Cammue, B., De Bolle, M., Terras, F., Proost, P., Van Damme, J., Rees, S., Vanderleyden, J. and Broekaert, W. (1992) Isolation and characterization of a novel class of plant antimicrobial peptides from *Mirabilis jalapa* L. seeds. *J. Biol. Chem.*, **267**, 2228-2233.
- Charlet, M., Chernysh, S., Philippe, H., Hetru, C., Hoffmann, J. and Bulet, P. (1996) Innate Immunity. Isolation of several cysteine-rich antimicrobial peptides from the blood of a mollusc, *Mytilus edulis*. *J. Biol. Chem.*, **271**, 21808-21813.
- Chavanas, S., Bodemer, C., Rochat, A., Hamel-Teillac, D., Ali, M. and Irvine, A.D. (2000) Mutations in *SPINK5*, encoding a serine protease inhibitor, cause Netherton syndrome. *Nat Genet*, **25**, 141-142.
- Chen, J.M., Mercier, B., Audrezet, M.P. and Ferec, C. (2000) Mutational analysis of the human pancreatic secretory trypsin inhibitor (PSTI) gene in hereditary and sporadic chronic pancreatitis. *J Med Genet*, **37**, 67-69.
- Chernysh, S., Kim, S.I., Bekker, G., Pleskach, V.A., Filatova, N.A., Anikin, V.B., Platonov, V.G. and Bulet, P. (2002) Antiviral and antitumor peptides from insects. *PNAS*, **99**, 12628-12632.
- Cociancich, S., Goyffon, M., Bontems, F., Bulet, P., Bouet, F., Menez, A. and Hoffmann, J. (1993) Purification and characterization of a scorpion defensin, a 4kDa antibacterial peptide presenting structural similarities with insect defensins and scorpion toxins.

- Biochem Biophys Res Commun*, **194**, 17-22.
- Cole, A.M., Liao, H.I., Ganz, T. and Yang, O.O. (2003) Antibacterial activity of peptides derived from envelope glycoproteins of HIV-1. *FEBS Letters*, **535**, 195-199.
- Com, E., Bourgeon, F., Evrard, B., Ganz, T., Colleu, D., Jegou, B. and Pineau, C. (2003) Expression of antimicrobial defensins in the male reproductive tract of rats, mice, and humans. *Biology of Reproduction*, **68**, 95-104.
- Cuenda, A., Rouse, J., Doza, Y.N., Meier, R., Cohen, P., Gallagher, T.F., Young, P.R. and Lee, J.C. (1995) SB 203580 is a specific inhibitor of a MAP kinase homologue which is stimulated by cellular stresses and interleukin-1. *FEBS Letters*, **364**, 229-233.
- Cui, Y., Wang, J., Zhang, X., Lang, R., Bi, M., Guo, L. and Lu, S.H. (2003) *ECRG2*, a novel candidate of tumor suppressor gene in the esophageal carcinoma, interacts directly with metallothionein 2A and links to apoptosis. *Biochem. Biophys. Res. Commun.*, **302**, 904-915.
- Dale, B.A., Presland, R.B., S.P., L., Underwood, R.A. and Fleckman, P. (1997) Transient expression of epidermal filaggrin in cultured cells causes collapse of intermediate filament networks with alteration of cell shape and nuclear integrity. *J Invest Dermatol*, **108**, 179-187.
- Dandekar, T., Snel, B., Huynen, M. and Bork, P. (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.*, **23**, 324-328.
- De Luca, L. (1991) Retinoids and their receptors in differentiation, embryogenesis, and neoplasia. *FASEB J.*, **5**, 2924-2933.
- Descargues, P., Deraison, C., Bonnart, C., Kreft, M., Kishibe, M., Ishida-Yamamoto, A., Elias, P., Barrandon, Y., Zambruno, G., Sonnenberg, A. and Hovnanian, A. (2005) Spink5-deficient mice mimic Netherton syndrome through degradation of desmoglein 1 by epidermal protease hyperactivity. *Nat Genet*, **37**, 56-65.
- Djian, P., Easley, K. and Green, H. (2000) Targeted Ablation of the Murine Involucrin Gene. *J. Cell Biol.*, **151**, 381-388.
- Donato, R. (2003) Intracellular and extracellular roles of S100 proteins. *Microsc Res Tech*, **60**, 540-551.
- Dyrlov Bendtsen, J., Nielsen, H., von Heijne, G. and Brunak, S. (2004) Improved Prediction of Signal Peptides: SignalP 3.0. *Journal of Molecular Biology*, **340**, 783-795.
- Eckert, R.L., Broome, A.-M., Ruse, M., Robinson, N., Ryan, D. and Lee, K. (2004) S100 Proteins in the Epidermis. *J Invest Dermatol*, **123**, 23-33.
- Eckert, R.L. and Welter, J.F. (1996) Transcription factor regulation of epidermal keratinocyte gene expression. *Mol Biol Rep*, **23**, 59-70.
- Ehret-Sabatier, L., Loew, D., Goyffon, M., Fehlbaum, P., Hoffmann, J.A., van Dorselaer, A. and Bulet, P. (1996) Characterization of Novel Cysteine-rich Antimicrobial Peptides from Scorpion Blood. *J. Biol. Chem.*, **271**, 29537-29544.
- Eisen, J.A. (1998) Phylogenomics: Improving Functional Predictions for Uncharacterized Genes by Evolutionary Analysis. *Genome Res.*, **8**, 163-167.
- Eisen, J.A. and Fraser, C.M. (2003) Phylogenomics: Intersection of Evolution and Genomics. *Science*, **300**, 1706-1707.
- Eisen, J.A. and Wu, M. (2002) Phylogenetic Analysis and Gene Functional Predictions: Phylogenomics in Action. *Theor Popul Biol*, **61**, 481-487.
- Ezekowitz, R.A.B. and Hoffmann, J. (1998) Innate immunity. *Curr. Opin. Immunol.*, **10**, 9-53.
- Fales-Williams, A.J., Brogden, K.A., Huffman, E., Gallup, J.M. and Ackermann, M.R. (2002)

- Cellular Distribution of Anionic Antimicrobial Peptide in Normal Lung and during Acute Pulmonary Inflammation. *Vet Pathol*, **39**, 706-711.
- Favata, M.F., Horiuchi, K.Y., Manos, E.J., Daulerio, A.J., Stradley, D.A., Feeser, W.S., van Dyk, D.E., Pitts, W.J., Earl, R.A., Hobbs, F., Copeland, R.A., Magolda, R.L., Scherle, P.A. and Trzaskos, J.M. (1998) Identification of a Novel Inhibitor of Mitogen-activated Protein Kinase Kinase. *J. Biol. Chem.*, **273**, 18623-18632.
- Fietz, M.J., Rogers, G.E., Eyre, H.J., Baker, E., Callen, D.F. and Sutherland, G.R. (1992) Mapping of the trichohyalin gene: co-localization with the profilaggrin, involucrin, and loricrin genes. *J Invest Dermatol.*, **99**, 542-544.
- Filichkin, S.A. and Gelvin, S.B. (1992) Effect of dimethyl sulfoxide concentration in specificity of primer matching in PCR. *BioTechniques*, **12**, 828-829.
- Fleckman, P., Dale, B.A. and Holbrook, K.A. (1985) Profilaggrin, a high-molecular-weight precursor of filaggrin in human epidermis and cultured keratinocytes. *J Invest Dermatol.*, **85**, 507-512.
- Frohm, M., Agerberth, B., Ahangari, G., Stahle-Backdahl, M., Liden, S., Wigzell, H. and Gudmundsson, G.H. (1997) The Expression of the Gene Coding for the Antibacterial Peptide LL-37 Is Induced in Human Keratinocytes during Inflammatory Disorders. *J. Biol. Chem.*, **272**, 15258-15263.
- Frohm, M., Gunne, H., Bergman, A.C., Agerberth, B., Bergman, T., Boman, A., Liden, S., Jornvall, H. and Boman, H.G. (1996) Biochemical and antibacterial analysis of human wound and blister fluid. *Eur J Biochem*, **237**, 86-92.
- Fuchs, E., Chan, Y.M., Paller, A.S. and Yu, Q.C. (1994) Cracks in the foundation: keratin filaments and genetic disease. *Trends Cell Biol*, **4**, 321-326.
- Gallo, R., Ono, M., Povsic, T., Page, C., Eriksson, E., Klagsbrun, M. and Bernfield, M. (1994) Syndecans, Cell Surface Heparan Sulfate Proteoglycans, and Induced by a Proline-Rich Antimicrobial Peptide from Wounds. *PNAS*, **91**, 11035-11039.
- Gallo, R.L. and Nizet, V. (2003) Endogenous production of antimicrobial peptides in innate immunity and human disease. *Curr Allergy Asthma Rep*, **3**, 402-409.
- Ganz, T. (1999) IMMUNOLOGY:Enhanced: Defensins and Host Defense. *Science*, **286**, 420-421.
- Ganz, T. (2002) The role of hepcidin in iron sequestration during infections and in the pathogenesis of anemia of chronic disease. *Israeli Medical Association Journal*, **4**, 1043-1045.
- Gebhard, L.G., Carrizo, F.U., Stern, A.L., Burgardt, N.I., Faivovich, J., Lavilla, E. and Ermacora, M.R. (2004) A Kazal prolyl endopeptidase inhibitor isolated from the skin of *Phyllomedusa sauvagii*. *Eur J Biochem*, **271**, 2117-2126.
- Giardina, E., Capon, F., De Rosa, M., Mango, R., Zambruno, G., Orecchia, A., Chimenti, S., Giardina, B. and Novelli, G. (2004) Characterization of the loricrin (LOR) gene as a positional candidate for the PSORS4 psoriasis susceptibility locus. *Annals of Human Genetics*, **68**, 639-645.
- Gläser, R., Harder, J., Lange, H., Bartels, J., Christophers, E. and Schroder, J.-M. (2005) Antimicrobial psoriasin (S100A7) protects human skin from *Escherichia coli* infection. **6**, 57-64.
- Glusman, G., Sosinsky, A., Ben-Asher, E., Avidan, N., Sonkin, D., Bahar, A., Rosenthal, A., Clifton, S., Roe, B. and Ferraz, C. (2000) Sequence, Structure, and Evolution of a Complete Human Olfactory Receptor Gene Cluster. *Genomics*, **63**, 227-245.
- Grenard, P., Bates, M.K. and Aeschlimann, D. (2001) Evolution of Transglutaminase Genes:



- Identification of a Transglutaminase Gene Cluster on Human Chromosome 15q15. Structure of the gene encoding transglutaminase X and a novel gene family member, transglutaminase Z. *J. Biol. Chem.*, **276**, 33066-33078.
- Gribenko, A.V. and Makhatadze, G.I. (1998) Oligomerization and divalent ion binding properties of the S100P protein: A  $\text{Ca}^{2+}/\text{Mg}^{2+}$ -switch model. *J Mol Biol*, **283**, 679-694.
- Gururajan, R., Lahti, J.M., Grenet, J., Easton, J., Gruber, I., Ambros, P.F. and Kidd, V.J. (1998) Duplication of a Genomic Region Containing the Cdc2L1-2 and MMP21-22 Genes on Human Chromosome 1p36.3 and their Linkage to D1Z2. *Genome Res.*, **8**, 929-939.
- Hancock, R.E.W. and Chapple, D.S. (1999) Peptide antibiotics. *Antimicrobial Agents Chemotherapy*, **43**, 1317-1323.
- Harder, J., Bartels, J., Christophers, E. and Schröder, J.M. (2001) Isolation and characterization of human beta-defensin-3, a novel human inducible peptide antibiotic. *J. Biol. Chem.*, **276**, 5707-5713.
- Harder, J. and Schröder, J.-M. (2002) RNase 7, a Novel Innate Immune Defense Antimicrobial Protein of Healthy Human Skin. *J. Biol. Chem.*, **277**, 46779-46784.
- Harder, J. and Schröder, J.-M. (2005) Psoriatic scales: a promising source for the isolation of human skin-derived antimicrobial proteins. *J Leukoc Biol*, **77**, 476-486.
- Harder J, Bartels, J., Christophers E, Schröder JM. (1997) A peptide antibiotic from human skin. *Nature*, **387**, 861.
- Heizmann, C.W., Fritz, G. and Schafer, B.W. (2002) S100 proteins: structure, functions and pathology. *Front Biosci.*, **7**, d1356-1368.
- Herr, W. and Cleary, M.A. (1995) The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. *Genes Dev.*, **9**, 1679-1693.
- Hiemstra, P.S., Eisenhauer, P.B., Harwig, S.S., van den Barselaar, M.T., van Furth, R. and Lehrer, R.I. (1993) Antimicrobial proteins of murine macrophages. *Infect Immun*, **61**, 3038-3046.
- Hoffmann, J.A., Kafatos, F.C., Janeway, C.A., Jr. and Ezekowitz, R.A.B. (1999) Phylogenetic Perspectives in Innate Immunity. *Science*, **284**, 1313-1318.
- Holbrook, K.A., Dale, B.A. and Brown, K.S. (1982) Abnormal epidermal keratinization in the repeated epilation mutant mouse. *J. Cell Biol.*, **92**, 387-397.
- Horii, A., Kobayashi, T., Tomita, N., Yamamoto, T., Fukushige, S., Murotsu, T., Ogawa, M., Mori, T. and Matsubara, K. (1987) Primary structure of human pancreatic secretory trypsin inhibitor (PSTI) gene. *Biochem. Biophys. Res. Commun.*, **149**, 635-641.
- Hultmark, D., Steiner, H., Rasmuson, T. and Boman, H.G. (1980) Insect immunity. Purification and properties of three inducible bactericidal proteins from hemolymph of immunized pupae of *Hyalophora cecropia*. *Eur J Biochem*, **106**, 7-16.
- Inohara, N. and Nunez, G. (2003) NODS: intracellular proteins involved in inflammation and apoptosis. *Nat Rev Immunol*, **3**, 371 -382.
- Ishida-Yamamoto, A. and Iizuka, H. (1998) Structural organization of cornified cell envelopes and alterations in inherited skin disorders. *Exp Dermatol*, **7**, 1-10.
- Jang, S.-I., Karaman-Jurukovska, N., Morasso, M.I., Steinert, P.M. and Markova, N.G. (2000) Complex Interactions between Epidermal POU Domain and Activator Protein 1 Transcription Factors Regulate the Expression of the Profilaggrin Gene in Normal Human Epidermal Keratinocytes. *J. Biol. Chem.*, **275**, 15295-15304.
- Jang, S.-I., Steinert, P.M. and Markova, N.G. (1996) Activator Protein 1 Activity Is Involved in the Regulation of the Cell Type-specific Expression from the Proximal Promoter of

- the Human Profilaggrin Gene. *J. Biol. Chem.*, **271**, 24105-24114.
- Johne, B., Fagerhol, M.K., Lyberg, T., Prydz, H., Brandtzaeg, P., Naess-Andresen, C.F. and Dale, I. (1997) Functional and clinical aspects of the myelomonocyte protein calprotectin. *Mol. Pathol.*, **50**, 113-123.
- Johnson, M.A., Johns, D.G. and Fridland, A. (1987) 2',3'-Dideoxynucleoside phosphorylation by deoxycytidine kinase from normal human thymus extracts: activation of potential drugs for AIDS therapy. *Biochem Biophys Res Commun*, **148**, 1252-1258.
- Kalinin, A., Marekov, L.N. and Steinert, P.M. (2001) Assembly of the epidermal cornified cell envelope. *J Cell Sci*, **114**, 3069-3070.
- Kent, W.J. (2002) BLAT---The BLAST-Like Alignment Tool. *Genome Res.*, **12**, 656-664.
- Kimchi, E.T., Posner, M.C., Park, J.O., Darga, T.E., Kocherginsky, M., Karrison, T., Hart, J., Smith, K.D., Mezhir, J.J., Weichselbaum, R.R. and Khodarev, N.N. (2005) Progression of Barrett's Metaplasia to Adenocarcinoma Is Associated with the Suppression of the Transcriptional Programs of Epidermal Differentiation. *Cancer Res*, **65**, 3146-3154.
- Kligman, D. and Hilt, D.C. (1988) The S100 protein family. *Trends Biochem Sci*, **13**, 437-443.
- Koch, P.J., de Viragh, P.A., Scharer, E., Bundman, D., Longley, M.A., Bickenbach, J., Kawachi, Y., Suga, Y., Zhou, Z., Huber, M., Hohl, D., Kartasova, T., Jarnik, M., Steven, A.C. and Roop, D.R. (2000) Lessons from Loricrin-deficient Mice: Compensatory Mechanisms Maintaining Skin Barrier Function in the Absence of a Major Cornified Envelope Protein. *J. Cell Biol.*, **151**, 389-400.
- Komatsu, N., Takata, M., Otsuki, N., Ohka, R., Amano, O., Takehara, K. and Saijoh, K. (2002) Elevated Stratum Corneum Hydrolytic Activity in Netherton Syndrome Suggests an Inhibitory Regulation of Desquamation by SPINK5-Derived Peptides. *Journal of Investigative Dermatology*, **118**, 436-443.
- Koonin, E.V., Makarova, K.S. and Aravind, L. (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol*, **55**, 709-742.
- Kosak, S.T. and Groudine, M. (2004) Form follows function: the genomic organization of cellular differentiation. *Genes Dev.*, **18**, 1371-1384.
- Kosak, S.T. and Groudine, M. (2004) Gene Order and Dynamic Domains. *Science*, **306**, 644-647.
- Krieg, P., Schuppler, M., Koesters, R., Mincheva, A., Lichter, P. and Marks, F. (1997) Repetin (Rptn), a New Member of the "Fused Gene" Subgroup within the S100 Gene Family Encoding a Murine Epidermal Differentiation Protein. *Genomics*, **43**, 339-348.
- Krisanaprakornkit, S., Kimball, J.R., Weinberg, A., Darveau, R.P., Bainbridge, B.W., and Dale, B.A. (2000) Inducible expression of human beta-defensin 2 by *Fusobacterium nucleatum* in oral epithelial cells: multiple signaling pathways and role of commensal bacteria in innate immunity and the epithelial barrier. *Infect. Immun.*, **68**, 2907-2915.
- Kube, E., Weber, K. and Gerke, V. (1991) Primary structure of human, chicken, and *Xenopus laevis* p11, a cellular ligand of the Src-kinase substrate, annexin II. *Gene*, **102**, 255-259.
- Kuechle, M.K., Thulin, C.D., Presland, R.B. and Dale, B.A. (1999) Profilaggrin Requires both Linker and Filaggrin Peptide Sequences to Form Granules: Implications for Profilaggrin Processing *In Vivo*. *J Invest Dermatol*, **112**, 843-852.
- Laskowski, M. Jr., Kato, I., Ardelt, W., Cook, J., Denton, A., Empie, M.W., Kohr W.J., Park,

- S.J., Parks, K., Schatzley, B.L., Schoenberger, O.L., Tashiro, M., Vichot, G., Wieczorek, A., and Wieczorek, M. (1987) Ovomuroid third domains from 100 avian species: isolation, sequences, and hypervariability of enzyme-inhibitor contact residues. *Biochemistry*, **26**, 202-221.
- Lathé III, W.C., Snel, B. and Bork, P. (2000) Gene context conservation of a higher order than operons. *Trends in Biochemical Sciences*, **25**, 474-479.
- Lawrence, J.G. (1997) Selfish operons and speciation by gene transfer. *Trends in Microbiology*, **5**, 355-359.
- Lawrence, J.G. (2002) Shared strategies in gene organization among prokaryotes and eukaryotes. *Cell*, **110**, 407-413.
- Lawrence, J.G. and Roth, J.R. (1996) Selfish Operons: Horizontal Transfer May Drive the Evolution of Gene Clusters. *Genetics*, **143**, 1843-1860.
- Lee, S., Kim, I., Marekov, L., O'Keefe, E., Parry, D. and Steinert, P. (1993) The structure of human trichohyalin. Potential multiple roles as a functional EF-hand-like calcium-binding protein, a cornified cell envelope precursor, and an intermediate filament-associated (cross-linking) protein. *J. Biol. Chem.*, **268**, 12164-12176.
- Lehesjoki, A.E. (2003) Molecular background of progressive myoclonus epilepsy. *EMBO J.*, **22(14)**, 3473-3478.
- Lehrer, R.I. and Ganz, T. (1996) Endogenous vertebrate antibiotics. Defensins, protegrins, and other cysteine-rich antimicrobial peptides. *Ann N Y Acad Sci*, **797**, 228-239.
- Lehrer, R.I. and Ganz, T. (2002) Cathelicidins: a family of endogenous antimicrobial peptides. *Curr Opin Hematol*, **9**, 18-22.
- Liu, C., Gelius, E., Liu, G., Steiner, H. and Dziarski, R. (2000) Mammalian Peptidoglycan Recognition Protein Binds Peptidoglycan with High Affinity, Is Expressed in Neutrophils, and Inhibits Bacterial Growth. *J. Biol. Chem.*, **275**, 24490-24499.
- Liu, L., Roberts, A.A. and Ganz, T. (2003) By IL-1 Signaling, monocyte-derived cells dramatically enhance the epidermal antimicrobial response to lipopolysaccharide. *Journal of Immunology*, **170**, 575-580.
- Lomas, G., Massey, S., Pantrini, S. and Ritchie, N. (2002) Taking part in the MRC crash trial. *Emerg Nurse*, **10**, 34-37.
- Lozovskaya, E.R., Hartl, D.L. and Petrov, D.A. (1995) Genomic regulation of transposable elements in *Drosophila*. *Curr Opin Genet Dev*, **5**, 768-773.
- Lu, S.M., Lu, W., Qasim, M.A., Anderson, S., Apostol, I., Ardelt, W., Bigler, T., Chiang, Y.W., Cook, J., James, M.N.G., Kato, I., Kelly, C., Kohr, W., Komiyama, T., Lin, T.-Y., Ogawa, M., Otlewski, J., Park, S.-J., Qasim, S., Ranjbar, M., Tashiro, M., Warne, N., Whatley, H., Wieczorek, A., Wieczorek, M., Wilusz, T., Wynn, R., Zhang, W. and Laskowski, M., Jr. (2001) Predicting the reactivity of proteins from their sequence alone: Kazal family of protein inhibitors of serine proteinases. *PNAS*, **98**, 1410-1415.
- Madsen, P., Rasmussen, H.H., Leffers, H., Honore, B., Dejgaard, K., Olsen, E., Kiil, J., Walbum, E., Andersen, A.H., Basse, B. and al., e. (1991) Molecular cloning, occurrence, and expression of a novel partially secreted protein "psoriasin" that is highly up-regulated in psoriatic skin. *J Invest Dermatol*, **97**, 701-712.
- Magert, H.-J., Standker, L., Kreutzmann, P., Zucht, H.-D., Reinecke, M., Sommerhoff, C.P., Fritz, H. and Forssmann, W.-G. (1999) LEKTI, a Novel 15-Domain Type of Human Serine Proteinase Inhibitor. *J. Biol. Chem.*, **274**, 21499-21502.
- Makino, T., Takaishi, M., Morohashi, M. and Huh, N.-h. (2001) Hornerin, a Novel Profilaggrin-like Protein and Differentiation-specific Marker Isolated from Mouse

- derived from histone H2A in the catfish, *Parasilurus asotus*. *FEBS Letters*, **437**, 258-262.
- Pearson, D.J., Dale, B.A. and Presland, R.B. (2002) Functional Analysis of the Profilaggrin N-Terminal Peptide: Identification of Domains that Regulate Nuclear and Cytoplasmic Distribution. *J Invest Dermatol*, **119**, 661-669.
- Peress, N., Perillo, E. and Zucker, S. (1995) Localization of tissue inhibitor of matrix metalloproteinases in Alzheimer's disease and normal brain. *J Neuropathol Exp Neurol*, **54**, 16-22.
- Peschel, A. (2002) How do bacteria resist human antimicrobial peptides? *Trends in Microbiology*, **10**, 179-186.
- Pietas, A., Schluns, K., Marenholz, I., Schafer, B.W., Heizmann, C.W. and Petersen, I. (2002) Molecular cloning and characterization of the human S100A14 gene encoding a novel member of the S100 family. *Genomics*, **79**, 513-522.
- Pilpel, Y., Sudarsanam, P. and Church, G.M. (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat Genet*, **29**, 153-159.
- Pognonec, P., Boulukos, K.E., Gesquiere, J.C., Stehelin, D. and Ghysdael, J. (1988) Mitogenic stimulation of thymocytes results in the calcium-dependent phosphorylation of c-ets-1 proteins. *EMBO J.*, **7**, 977-983.
- Polakowska, R.R.M., Bartlett, R., Goldsmith, L.A. and Haake, A.R. (1994) Apoptosis in human skin development: morphogenesis, periderm, and stem cells. *Dev Dyn*, **199**, 176-188.
- Pollock, J.J., Denepitiya, L., MacKay, B.J. and Iacono, V.J. (1984) Fungistatic and fungicidal activity of human parotid salivary histidine-rich polypeptides on *Candida albicans*. *Infect Immun*, **44**, 702-707.
- Presland, R.B., Boggess, D., Lewis, S.P., Hull, C., Fleckman, P. and Sundberg, J.P. (2000) Loss of Normal Profilaggrin and Filaggrin in Flaky Tail (*ft/ft*) Mice: an Animal Model for the Filaggrin-Deficient Skin Disease Ichthyosis Vulgaris. *J Invest Dermatol*, **115**, 1072-1081.
- Presland, R.B., Kuechle, M.K., Lewis, S.P., Fleckman, P. and Dale, B.A. (2001) Regulated expression of human filaggrin in keratinocytes results in cytoskeletal disruption, loss of cell-cell adhesion, and cell cycle arrest. *Exp Cell Res*, **270**, 199-213.
- Puente, X.S. and Lopez-Otin, C. (2004) A Genomic Analysis of Rat Proteases and Protease Inhibitors. *Genome Res.*, **14**, 609-622.
- Rawlings, N.D., Tolle, D.P. and Barrett, A.J. (2004) MEROPS: the peptidase database. *Nucl. Acids Res.*, **32**, D160-164.
- Regnier, M., Caron, D., Reichert, U. and Schaefer, H. (1993) Barrier function of human skin and human reconstructed epidermis. *J Pharm Sci*, **82**, 404-407.
- Resing, K., al-Alawi, N., Blomquist, C., Fleckman, P. and Dale, B. (1993) Independent regulation of two cytoplasmic processing stages of the intermediate filament-associated protein filaggrin and role of Ca<sup>2+</sup> in the second stage. *J. Biol. Chem.*, **268**, 25139-25145.
- Resing, K.A., Thulin, C., Whiting, K., Al-Alawi, N. and Mostad, S. (1995) Characterization of Profilaggrin Endoproteinase 1. *J. Biol. Chem.*, **270**, 28193-28198.
- Rheinwald, J.G. and Green, H. (1975) Serial cultivation of strains of human epidermal keratinocytes: the formation of keratinizing colonies from single cells. *Cell*, **6**, 331-343.
- Ritchie, B.C. (2003) Protease inhibitors in the treatment of hereditary angioedema.

- Transfusion and Apheresis Science*, **29**, 259-267.
- Roder, B.L. and Gutschik, E. (1989) In-vitro activity of ciprofloxacin combined with either fusidic acid or rifampicin against *Staphylococcus aureus*. *J Antimicrob Chemother*, **23**, 347-352.
- Ross, S.A., McCaffery, P.J., Drager, U.C. and De Luca, L.M. (2000) Retinoids in Embryonal Development. *Physiol. Rev.*, **80**, 1021-1054.
- Rothnagel, J. and Rogers, G. (1986) Trichohyalin, an intermediate filament-associated protein of the hair follicle. *J. Cell Biol.*, **102**, 1419-1429.
- Schafer, B.W. and Heizmann, C.W. (1996) The S100 family of EF-hand calcium-binding proteins: functions and pathology. *Trends in Biochemical Sciences*, **21**, 134-140.
- Schechter, I. and Berger, A. (1968) On the active site of proteases. 3. Mapping the active site of papain; specific peptide inhibitors of papain. *Biochem Biophys Res Commun*, **32**, 898-902.
- Schibli, D.J., Epanand, R.F., Vgel, H.J. and Epanand, R.M. (2002) Tryptophan-rich antimicrobial peptides: comparative properties and membrane interactions. *Biochem Cell Biol*, **80**, 667-677.
- Schitteck, B., Hipfel, R., Sauer, B., Bauer, J., Kalbacher, H., Stevanovic, S., Schirle, M., Schroeder, K., Blin, N., Meier, F., Rassner, G. and Garbe, C. (2001) Dermcidin: a novel human antibiotic peptide secreted by sweat glands. *Nature Immun.*, **2**, 1133-1137.
- Schroder, J.M. (1999) Epithelial antimicrobial peptides: innate local host response elements. *Cell Mol Life Sci*, **56**, 32-46.
- Selsted, M.E., Szklarek, D. and Lehrer, R.I. (1984) Purification and antibacterial activity of antimicrobial peptides of rabbit granulocytes. *Infect Immun*, **45**, 150-154.
- Simmaco, M., Mignogna, G. and Barra, D. (1998) Antimicrobial peptides from amphibian skin: what do they tell us? *Biopolymers*, **47**, 435-450.
- Snel, B., Bork, P. and Huynen, M. (2000) Genome evolution: gene fusion versus gene fission. *Trends in Genetics*, **16**, 9-11.
- Spitznagel, J.K. Origins and Development of Peptide Antibiotic Research. In: Antibacterial peptide protocols. Shafer, W.M. (ed.), Humana press, Totowa, New Jersey. pp 1-14, 1997.
- Steinbakk, M., Naess-Andresen, C.F., Lingaas, E., Dale, I., Brandtzaeg, P. and Fagerhol, M.K. (1990) Antimicrobial actions of calcium binding leucocyte L1 protein, calprotectin. *Lancet*, **336**, 763-765.
- Steinberg, D.A. and Lehrer, R.I. (1997) Designer assays for antimicrobial peptides. Disputing the "one-size-fits-all" theory. *Methods Mol Biol*, **78**, 169-186.
- Steiner, H., Hultmark, D., Engstrom, A., Bennich, H. and Boman, H.G. (1981) Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature*, **292**, 246-248.
- Steinert, P.M. and Marekov, L.N. (1999) Initiation of Assembly of the Cell Envelope Barrier Structure of Stratified Squamous Epithelia. *Mol. Biol. Cell*, **10**, 4247-4261.
- Steinert, P.M., Parry, D.A.D. and Marekov, L.N. (2003) Trichohyalin Mechanically Strengthens the Hair Follicle: multiple cross-bridging roles in the inner root sheath. *J. Biol. Chem.*, **278**, 41409-41419.
- Steinman, H.A. (1996) "Hidden" allergens in foods. *J Allergy Clin Immunol*, **98**, 241-250.
- Swartzendruber, D.C., Wertz, P.W., Madison, K.C. and Downing, D.T. (1987) Evidence that the corneocyte has a chemically bound lipid envelope. *J Invest Dermatol*, **88**, 709-713.

- Sybert, V.P., Dale, B.A. and Holbrook, K.A. (1985) Ichthyosis vulgaris: identification of a defect in synthesis of filaggrin correlated with an absence of keratohyaline granules. *J. Invest. Derm.*, **84**, 191-194.
- Takeda, K. and Akira, S. (2003) Toll receptors and pathogen resistance. *Cell. Microbio.*, **5**, 143-153.
- Tamames, J., Casari, G., Ouzounis, C. and Valencia, A. (1997) Conserved Clusters of Functionally Related Genes in Two Bacterial Genomes. *J. Mol. Evol.*, **44**, 66-73.
- Tang, Y.-Q., Yuan, J., Osapay, G., Osapay, K., Tran, D., Miller, C.J., Ouellette, A.J. and Selsted, M.E. (1999) A Cyclic Antimicrobial Peptide Produced in Primate Leukocytes by the Ligation of Two Truncated -Defensins. *Science*, **286**, 498-502.
- Tarcsa, E., Marekov, L.N., Andreoli, J., Idler, W.W., Candi, E., Chung, S.-I. and Steinert, P.M. (1997) The Fate of Trichohyalin. Sequential post-translational modifications by peptidyl-arginine deiminase and transglutaminases. *J. Biol. Chem.*, **272**, 27893-27901.
- Tesfaigzi, J., Carlson, D.M. (1999) Expression, regulation, and function of the SPR family of proteins. *Cell Biochem Biophys*, **30**, 243-265.
- Timmers, H.T.M., Meyers, R.E. and Sharp, P.A. (1992) Composition of transcription factor B-TFIID. *Proc. Natl. Acad. Sci. USA*, **89**, 8140-8144.
- Tossi, A. and Sandri, L. (2002) Molecular diversity in gene-encoded, cationic antimicrobial polypeptides. *Curr Pharm Des*, **8**, 743-761.
- Tsai, H. and Bobek, L.A. (1998) Human salivary histatins: promising anti-fungal therapeutic agents. *Crit Rev Oral Biol Med*, **9**, 480-497.
- Underhill, D.M. and Ozinsky, A. (2002) PHAGOCYTOSIS OF MICROBES: Complexity in Action. *Annual Review of Immunology*, **20**, 825-852.
- Vansant, G. and Reynolds, W. (1995) The Consensus Sequence of a Major Alu Subfamily Contains a Functional Retinoic Acid Response Element. *PNAS*, **92**, 8229-8233.
- Velasco, G., Geelen, M.J.H., del Pulgar, T.G. and Guzman, M. (1998) Malonyl-CoA-independent Acute Control of Hepatic Carnitine Palmitoyltransferase I Activity. Role of Ca<sup>2+</sup>/Calmodulin-dependant protein kinase II and cytoskeletal components. *J. Biol. Chem.*, **273**, 21497-21504.
- Verrijzer, C.P., Alkema, M.J., van Weperen, W.W., Van Leeuwen, H.C., Strating, M.J. and van der Vliet, P.C. (1992) The DNA binding specificity of the bipartite POU domain and its subdomains. *EMBO J.*, **11**, 4993-5003.
- Vizioli, J. and Salzet, M. (2002) Antimicrobial peptides from animals: focus on invertebrates. *Trends in Pharmacological Sciences*, **23**, 494-496.
- Volz, A., Korge, B.P., Compton, J.G., Ziegler, A., Steinert, P.M. and Mischke, D. (1993) Physical mapping of a functional cluster of epidermal differentiation genes on chromosome 1q21. *Genomics*, **18**, 92-99.
- Walley, A.J., Chavanas, S., Moffatt, M.F., Esnouf, R.M., Ubhi, B., Lawrence, R., Wong, K., Abecasis, G.R., Jones, E.Y., Harper, J.I., Hovnanian, A. and Cookson, W.O. (2001) Gene polymorphism in Netherton and common atopic disease. *Nat Genet*, **29**, 175-178.
- Wernicke, D., Seyfert, C., Hinzmann, B. and Gromnica-Ihle, E. (1996) Cloning of collagenase 3 from the synovial membrane and its expression in rheumatoid arthritis and osteoarthritis. *J Rheumatol*, **23**, 590-595.
- Wilkinson, M.M., Busuttil, A., Hayward, C., Brock, D.J., Dorin, J.R. and Van Heyningen, V. (1988) Expression pattern of two related cystic fibrosis-associated calcium-binding

- proteins in normal and abnormal tissues. *J. Cell Sci.*, **91**, 221-230.
- Williamson M.P. , M.D., Wuthrich K. (1984) Secondary structure in the solution conformation of the proteinase inhibitor IIA from bull seminal plasma by nuclear magnetic resonance. *J. Mol. Biol.*, **173**, 341- 359.
- Witt, H., Luck, W., Hennies, H.C., Classen, M., Kage, A., Lass, U., Landt, O. and Becker, M. (2000) Mutations in the gene encoding the serine protease inhibitor, Kazal type 1 are associated with chronic pancreatitis. *Nat Genet*, **25**, 213-216.
- Xia, L., Stoll, S.W., Liebert, M., Ethier, S.P., Carey, T., Esclamado, R., Carroll, W., Johnson, T.M. and Elder, J.T. (1997) CaN19 expression in benign and malignant hyperplasias of the skin and oral mucosa: Evidence for a role in regenerative differentiation. *Cancer Res*, **57**, 3055-3062.
- Xu, S.-j., Wang, Y.-p., Roe, B. and Pearson, W.R. (1998) Characterization of the Human Class Mu Glutathione S-Transferase Gene Cluster and the GSTM1 Deletion. *J. Biol. Chem.*, **273**, 3517-3527.
- Xu, Z., Wang, M.-R., Xu, X., Cai, Y., Han, Y.-L., Wu, K.-M., Wang, J., Chen, B.-S., Wang, X.-Q. and Wu, M. (2000) Novel Human Esophagus-Specific Gene C1orf10: cDNA Cloning, Gene Structure, and Frequent Loss of Expression in Esophageal Cancer. *Genomics*, **69**, 322-330.
- Yagui-Beltran, A., Craig, A.L., Lawrie, L., Thompson, D., Pospisilova, S., Johnston, D., Kernohan, N., Hopwood, D., Dillon, J.F. and Hupp, T.R. (2001) The human oesophageal squamous epithelium exhibits a novel type of heat shock protein response. *Eur J Biochem*, **268**, 5343-5355.
- Zasloff, M. (1987) Magainins, a class of antimicrobial peptides from *Xenopus* skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor. *Proc Natl Acad Sci U S A.*, **84**, 5449-5453.
- Zasloff, M. (1992) Antibiotic peptides as mediators of innate immunity. *Current Opinion in Immunology*, **4.**, 3-7.
- Zasloff, M. (2002) Antimicrobial peptides of multicellular organisms. *Nature*, **415**, 389-395.
- Zeya, H.I. and Spitznagel, J.K. (1969) Cationic protein-bearing granules of polymorphonuclear leukocytes: separation from enzyme-rich granules. *Science*, **163**, 1069-1071.
- Zheng, X.L. and Zheng, A.L. (2002) Genomic organization and regulation of three cecropin genes in *Anopheles gambiae*. *Insect Mol Biol*, **11**, 517-525.
- Zhou, Y. and Mishra, B. (2005) Quantifying the mechanisms for segmental duplications in mammalian genomes by statistical analysis and modeling. *PNAS*, **102**, 4051-4056.
- Zimmer, D.B., Chessher, J. and Song, W. (1996) Nucleotide homologies in genes encoding members of the S100 protein family. *Biochim Biophys Acta*, **1313**, 229-238.
- Zimmer, D.B., Wright, S.P. and Weber, D.J. (2003) Molecular mechanisms of S100-target protein interactions. *Microsc Res Tech*, **60**, 552-559.
- Zmasek, C. and Eddy, S. (2002) RIO: Analyzing proteomes by automated phylogenomics using resampled inference of orthologs. *BMC Bioinformatics*, **3**, 14.
- Zmasek, C.M. and Eddy, S.R. (2001) A simple algorithm to infer gene duplication and speciation events on a gene tree. *Bioinformatics*, **17**, 821-828.

5.1 Nucleotide and amino acid sequences of human THHL1 cDNA. The full-length nucleotide sequence of human THHL1 cDNA (3603 bp) was verified by direct cDNA cloning and sequencing and aligned with the predicted open reading frame amino acid sequence. The underlined red nucleotides indicate the canonical poly(A) addition signal. The exon/intron boundaries were determined by the BLAT Search. Sequence data have been deposited under the GenBank™ accession no. AY456639.

```

1 acagcccgcgtgccccacgtttcactgctctcgcttgtgaactag...cctctcctgagtttgtaaaag
66 atg<u>cctcagc</u>tctctgagaatgtcctcctgtgtaattgagacatccacaaatagccagtgggacagtaacggggcaacactgactggcagagagctgaaacaa
M P Q L L R N V L C V I E T F H K Y A S E D S N G A T L T G R E L K Q

171 ctcatccaggggaggtttggggaacttttttcag...cctctggtccttcctgctgtggaaaaaaatccaatctctgaaatattgacagtaat
L I Q G E F G D F F Q P C V L H A V E K N S N L L N I D S N
261 ggcatcatcagttttgatgaatttggcttggcaatctcaacttggtagaacctctgttatcttgacataaaatcattactaagttctgaaactaagacaggtgact
G I I S P D E F V L A I F N L L N L C Y L D I K S L L S S E L R Q V T
366 aaaccagagaggagaagctagatgatgtggtgtcaggcaaccaccggagatggctcagtgagacagtggaacttcaccaactcaagaaaagagatgcttctt
K P E K E K L D D V D V Q A T T G D G Q W T V G T S P T Q E K R M L P
471 tcaggaatggcatcatctcagctcctcctgaagaaagtggagcagtggaaataacagagtggaaccatggagagagccaagactccaacttccagga
S G M A S S S Q L I P E E S G A V G N N R V D P W R E A K T H N F P G
576 gaagcattgaaacacatgatcctaaagaacaaacacctggaaagagatgacaaagtcaggaagtggctcaagatatacaaacacagaaagacaaatgaaggccaa
E A S E H N D P K N K H L E G D E Q S Q E V A Q D I Q T T E D N E G Q
681 ctttaagcaaatagccaatggcaggtcaaaaagaccagcagctccacagagaggaaggacaagataaggagatctccagggaaggagatgaaccagccaga
L K T N K P M A G S K K T S S P T E R K G Q D K E I S Q E G D E P A R
786 gagcaaatgtttccaagatgaagagaccagtttggagaacaggaaggaacttggcaaccacaaagtccaccacaaaagaaagcaaacagaccatgtgaagat
E Q S V S K I R D Q F G E Q E G N L A T Q S S P P K E A T C R P C E D
891 caggaagttagaacagaaaaggaaaaacactcctaataatcagaagaccaccctcagaagagagatgagccagttcacagcatgctgacctgccagaacaagct
Q E V R T E K E K H S N I Q E P P L Q R E D E P S S Q H A D L P E Q A
996 gctgccaggtcaccatctcagacacagaaaatcaactgattccaaggatgtctgtagaaatgtttgacactcaagaaccaggaaggatgctgaccagacaccagct
A A R S P S Q T Q K S T D S K D V C R M F D T Q E P G K D A D Q T P A
1101 aaaacaaagaatttgggtgaactcaggatgatggcagaaatctgagacccaagaaaaaagaatgtgaaacaaaggacctgccagtcacatggtagcagaat
K T K N L G E P E D Y G R T S E T T Q E K E C E T K D L P V Q Y G S R N
1206 ggttcagaaaactctgacatgagagatgaaaaggaagagaaggaggtcctgaggccactggaacagcagggcagaagaaagcagtgacagaaactcggccacta
G S E T S D M R D E R K E R R G P E A H G T A G Q K E R D R K T R P L
1311 gtctctggaaacccaacacagatgaggaagtatcaggaactccaaggattatcaaaatcaaaagatgctgaaaaaggtctctgagacacaaatctcaagctcagaa
V L E T Q T Q D G K Y Q E L Q G L S K S K D A E K G S E T Q Y L S S E
1416 ggaggagatcagactcaccctgaaacttgaaaggaacagcagctctcaggaagagagagagagagacacacacaaagaaagcagcagcaggaagcatttggaaacgcaaaaac
G G D Q T H P E L E G T A V S G E E A E H T K E G T A E A F V N S K N
1521 gcacctgcagcagaaggcactgggggcaagagaagaaacacaaagattagcaccacttgagaagcagctctgtaggagaaaaactaggttccacaagactcat
A P R A E R T L G A R E R T Q D L A P L E K Q S V G E N T R V T K T H
1626 gaccaaccagttgaggaggaggtggttaccagggggaggacctgagtcaccatccacacagagtgatgaggggtctctctgaaactcccaacagcctggcttca
D Q P Y V E E E D G Y Q G E D P E S P F T Q S D E G S S E T P N S L A S
1731 gaggaaggcaatagcagctcagagacaggtgaactgctgtgcaaggggactccagagctcaaggggaccaacatggagagctgtgcaagaggtcacaataat
E E G N S S S E T G E L P V Q G D S Q S Q G D Q H G E S V Q G G H N N
1836 aaccagatcaccagagcaggaacacactggtgagaaaaacagagctctggaggcagtggtaccagcagctcagagagagagatgtacagctccacagaggaaccag
N P D T C Q R Q G T P G E K N R A L E A V V P A V R G E D V Q L T E D Q
1941 gaacagcctgccaagggagaacacaagaatcaaggccagggacccaagggccagctgtggagcccaatggagacccagagcagcaggaatccacagca
E Q P A R G E H K N Q G P G T K G P G A A V E P N G H P E A Q E S T A
2046 ggagatgaaaatagaaaactcctggaaaatagagatccagagtgccctggatgaagactcactgaccagcttccctatmgcagctcctggaaaaggagatagc
G D E N R K S L E I E I T G A L D E D F T D Q L S L M Q L P G K G D S
2151 agaaatgaataaaggctcaggcccaagtagcaaaagagaaggaagagcaacagagccccaagatactctgttagaaagtctagatgaggacaattcagcc
R N E L K V Q G P S S K E E K G R A T E A Q N T L L E S L D E D N S A
2256 tccctcaagatacaacttgaacaaaggaaacctgtaacatcagaggaggaagatgaaagtcccagaagctggcaggagaaaggtggtagc aaaaagttccagcc
S L K A I E R T E L T K E P V T S E E E A G E S P Q E L A G E G D Q K S P A
2361 aagaaagagcacaatctcagctccctggccaagtctgaaaagcagatgcagagagacaaagagccctgtctgtggagaggggtgagctctatccagtcca
K K E H N S S V P W S S L E K Q M Q R D Q E P C S V E R G A V Y S S P
2466 ctataccagtagctacagggagaagatactgcagcaaacaaatgtaacccaagagggagcctcaaaaagcaagtcagatagccagggcatcagggccagagctttgc
L Y Q Y L Q E K I L Q Q T N V T Q E E H Q K Q V Q I A Q A S G P E L C
2571 agtgtatccctcactagtgagatctcagatgttctctgtcttttcaactacagccaagcacaacacatataccaggggacttccactgtatgagagctctgct
S V S L T S E I S D C S V F P N Y S Q A S Q P Y T R G L P L D E S P A
2676 ggtgcacaggaacaaccagctcccagggccttggaaagataagcaaggtcaccctcagagagagaggtggtactacaagggaggccaagcaccacaaaagcaatga
G A Q E T P A P Q A L E D K Q G H P Q R E R L V L Q R E A S T T K Q *
2781 atcattatcactcacaatgcccccaatctctctcccaatcacctcaatctcactcagctgacacatcagatgcagatbaacaatctttaaattctctctaa
2886 attctcacaacaaacaaat<u>aa</u>aaatatacagcagtggggacaccacataaagaaacatctgagcaaacaaatcagctcagtaaaactctaccacaaatgctaa
2991 ctccaagtgcacaaatggcctttgtcatgagacagtttatctcaatgtaacctgacatataaaaatggaaaaatctctccatctctttagcaggtgacagtttt
3096 ccaactgattcattgaatttaaaatagagcagtaggaaatataagactcgaataatactcataatcataaaaatataatcattatataatgactactctgc
3201 ttcatgacattctbagaagtatctcaagtctatctcaaaatcttctcaaaatatttagacaatgccaacttaattatgattttctagcactctgact
3306 aaagtaaatctgactctcaggtgcagctgaaatctctcagctgactgactctctcaatctctcaatctctcaatctctcaatctctcaatctctcaatctctc
3411 ttacacaaactttctatagcttttggctca<u>aa</u>aaacaaatcaagaaataatgatgagagaaaacacaaatcaaaagtatgttggatcttataagaagctctgttaa
3516 ccaaccatccagtgagctactaaggtttacctatggtttgatgagggtatcaaaatgaaatgata<u>aa</u>aaacaaactgatgat 3603
    
```



**5.2 Nucleotide and amino acid sequences of human repetin cDNA.** The full-length nucleotide sequence of human hornerin cDNA (3569 bp) was verified by direct cDNA cloning and sequencing and aligned with the predicted open reading frame amino acid sequence. The underlined red nucleotides indicate the canonical poly(A) addition signal. The exon/intron boundaries were determined by the BLAT Search. Sequence data have been deposited under the GenBank<sup>TM</sup> accession no. AY396742.

```

1 ATCCTCTCTGACTCCTGAATCTTCTGCTCCTCTCCTCAAGCAG... Intron I(1274 bp) ...GTTCCACCCGTACTTGTCAAA
66 ATGGCTCAACTCCTGAATAGCATACTCAGTGTGATTGACGTATCCACAAATATGCCAAGGGGAATGGGGACTGTGCTTACTATGCAAGGAAAGAGTTGAAACAA
M A Q L L N S I L S V I D V F H K Y A K G N G D C A L L C K E E L K Q
171 CTGCTCTTGGCTGAGTTTGGAGACATCCTCCAG... Intron II(701 bp) ...AGACCAAATGACCCAGAGACTGTGGAAACCATCTTGAACCTCTTAGATCAAGAC
L L L A E F G D I L Q R P N D P E T V E T I L N L L D Q D
258 CGAGATGGACATATTGATTTTCATGAGTACCTCTTGTGGTGTCCAGTTGGTCCAAAGCCTGTATCATTAAGCTAGACAATAAGTACATGGAGCCAGGACCTCA
R D G H I D F H E Y L L L V F Q L V Q A C Y H K L D N K S H G G R T S
363 CAGCAAGAAAGGGGGCAGGAAGGAGCACAAGACTGTAAAGTTCCAGGAAACACAGGCAGACAACACAGACAGAGCCAGGAAAGAAAGGCAGAACTCCCAACCAC
Q Q E R G Q E G A Q D C K F P G N T G R Q H R Q R H E E E R Q N S H H
468 AGTCAGCCTGAGAGACAAGACGGAGATTCCCAACATGGTCCAGCTGAGAGACAAGACAGAGATTCCCAACATGGTCCAGTCTGAGAAACAAGACAGAGATTCCCAAC
S Q P E R Q D G D S H H G Q P E R Q D R D S H H G Q S E K Q D R D S H
573 CACAGTCAACCTGAGAGACAAGACAGAGATTCTCAACCAATCAGTCTGAGAGACAAGACAGGATTTCCAGCTTTGATCAGTCAGAGACAAGTCAAGACTCC
H S Q Q P E R D S H H N Q S E R Q D K D F S F D Q S E R Q S Q D S
678 AGCTCTGGTAAAAAGTGTAGTCAACAATCTACCAGTGGCCAGGCTAAATGGCAGGGACATATCTTTGCCTTAAATCGGTGAAAAACCAATTCAGGATTTCTCAT
S S G K K V S H K S T S G Q A K W Q G H I F A L N R C E K P I Q D S H
783 TATGGTCAAGTCAAGACATACACAACATCTGAAACACTTGGACAAAGCCTCTCACTTTAACAGACAAATCAACAGAAATCAGGCTCTTATTTGGTGGACAGTCT
Y G Q S E R H T Q Q S E T L G Q A S H F N Q T N Q Q K S G S Y C G Q S
888 GAGAGGCTAGGTCAGGAATTAGGCTGTGGTCAGACAGACAGACAGGCCAGAGTTCCCACTACGGTCAGACGGACAGACAAGACCAAGATTTATCATTATGGTCAG
E R L G Q E L G C G Q T D R Q G Q S S H Y G Q T D R Q D Q S Y H Y G Q
993 ACAGACAGACAAGGCCAGAGTTCCCACTACAGTCAAGCCGACAGACAAGGCCAGAGTTCCCACTACAGTCAAGCCAGACAGACAAGGTCAGAGTTCCCACTATGGT
T D R Q G Q S S H Y S Q T D R Q G Q S S H Y S Q P D R Q G Q S S H Y G
1098 CAAATGGACAGAAAGGCCAGTGTATCATTATGATCAGACAAACAGACAAGGCCAGGCTTCCCACTACAGTCAACAAACAGACAAGGTCAGAGTTCCCACTAT
Q M D R K G A Q C Y H Y D Q T N R Q G Q G Q S S H Y S Q P N R Q G Q S S H Y
1203 GGTCCAGCAGACACAAGATCAGAGTTTCTCACTATGGTCAGACAGACAGACAAGCCAAAGTCTCACTATGGTCAGACAGAGACAGGCCAGAGTTCCCACT
G Q P D T Q D Q S S H Y G Q T D R Q D Q S S H Y G Q T E R Q G Q S S H
1308 TACAGTCAAGTGGACCCAGACAGGCCAGGTTCTCACTACGGTCAAGCAGACAGACAAGGCCAGAGTTCCCACTATGGTCAGCCAGACAGACAAGGCCAGAAATCC
Y S Q M D R Q G Q G S H Y G Q T D R Q G Q S S H Y G Q P D R Q G Q N S
1413 CACTATGGTCAGACAGACAAGGCCAGAGTTCCCACTATGGTCAGACAGACAGACAAGGCCAGAGTTCCCACTACAGTCAAGCCAGACACAAGGCCAGAGT
H Y G Q T D R Q G Q S S H Y G Q T D R Q G Q S S H Y S Q P D R Q G Q S
1518 TCCCACTATGGTAAAGTACAGACAAGACCAGAGTTATCATTATGGTCAGCCAGACAGACAAGGCCAAAGTCCCACTATGGTCAGACAGACAGACAAGGCCAG
S H Y G K I D R Q D Q S Y H Y G Q P D G Q G Q S S H Y G Q T D R Q G Q
1623 AGTTTCCACTATGGTCAGCCAGACAGACAAGGCCAGAGTTCCCACTACAGTCAAGTGGACAGACAAGGCCAGAGTTCCCACTATGGTCAGACAGACAGACAAGGC
S F H Y G Q P D R Q G Q S S H Y S Q M D R Q G Q S S H Y G Q T D R Q G
1728 CAGAGTTCCCACTACGGTCAGACAGACAAGGCCAGAGTTATCATTATGGTCAGACAGACAGACAAGGCCAGAGTTCCCACTATTTCAATTCAGACTGGG
Q S S H Y G Q T D R Q G Q S Y H Y G Q T D R Q G Q S S H Y I Q S Q T G
1833 GAAATACAGGGCAAAATAGTACTTCCAAAGGACTGAAGGAAACAGAAAAGCCTCTTATGTTGAAACAATCAGGAAAGTCAAGGGAGCTAAGTCAACAGACTCCA
E I Q G Q N K Y F Q G T E G T R K A S Y V E Q S G R S G R L S Q Q T P
1938 GGACAGAAAGGTTACCAAAACAGGCCAGGGAATCCAGTCTAAGGACTCAGCAGACAAGGCCACCAAGGTATGGGAGCTGAAGAGGATAGCCAACTACCAAA
G Q E G Y Q N Q G Q G F Q G N Q D R S Q Q N G H Q V W E P E E D S Q H H Q
2043 CACAAACTCTTAGCACAAATCCAAACAAGAAAGAACCTTTGTGCAAAAAGGAGAGACTGGCAATCATGCAAGTGTAGCAGGGCCACAGACAGGCCACAGCCAGG
H K L L A Q I Q Q E R P L C H K G R D W Q S C S S E Q G H R Q A Q T R
2148 CAGAGTCAAGTGGTGGGGCTGAGCCACTGGGCAAGGAGAGCAGGGCCATCAAACTTGGGATAGACAAGCCATGAGAGTCAAGGAGGTCATGTGGGACACAG
Q S H G E L S H W A E E E Q G H Q T W D R H S H E S Q E G P C G T Q
2253 GACAGGCGAAACCCATAAAGATGAGCAGAACCCATCAGAGAGGAGACAGACAAGCCATGAAACATGAGCAGAGCCATCAGACAGGACAGGCAACCCATGAAGAC
D R R T H K D E Q N H Q R R D R Q T H E H E Q S H Q R R D R Q T H E D
2358 AAGCAGAAACCGTACAGACAGAGCAGGCAACCCATGAAGACAGCAGAAACCATCAGACAGGACAGGCAAACTCATGAAGAGGATCAAAACCATCAGCAACAA
K Q N R Q R R D R Q T H E D E Q N H Q R R D R Q T H E E D Q N H Q Q Q
2463 CATAATGACAAAATCATGAGGAAAGAGAGGTATCAAGGATCTCAGAAATCAAAAATCCCAAGTACCCACAGAACTGTCTTAAACAGAAAATTTCCACATG
H N R Q N Y E E K E R Y Q G S Q N Q K S Q V T H R S C P N R E K F H M
2568 AAAGAGGATGACAGAGCCAGGCTCACAGAAAAGACACACTGAAACCTCTTCTATCCAAACCCAGAGCAGTGGGAGGCCCAACCCAGAGAAACAGCGTGGTCAC
K E D D Q S Q G S Q K R H T E P S F Y P T Q S S G R P Q T R E Q R G H
2673 CCTGCCAAAGGGAGCTACTGTTCCCAACCCCCCTATGACTATGTGCAAGAGCAGAAATCCTACCCATACTAGGCTATGTGAGCACCAGAGGTAAGCAACACAAA
P A K G A T A T V P N P P Y Y Q E Q K S Y P Y *
2778 GCCAAAAAGAAAACCTATATGTCGAGTTTCACTTAAATTCGATTTAAGAGATTGAGAAATGATTTTCTTCCCTCTATCCTCTGCTGATCTCCCTGCAAGGATGTT
2883 TTTGAGTACAGCATTCACGTAAGGGCTTTTGGTCTTTGAGCAGCAATCTAGGGTTTTCTGATTCAGTGAATC TAGGTGGTAAATGGGTGAAATAATCATA
2988 TCTTAATGTTGATCAGTTTGGGTATTTAAATCATTTCTCTGCTTGCCTCTGTGATGATGAGACCTGGTTGAATCA TTGAAAGGTAAGAACCTTCTCCCTAAAGT
3093 TCAGAAACAGAAAAGTATTTCTGAAGATTACATACATCCAAAGGATTATACTACATCTCCACTCCTCAAGGAAACCTAGATTTGGTGGAGCTTTGGCTCCTGTGG
3198 ATAAACATCAGGAATTTCAAGCCTCATAGTCTAGACTAGAATGAAACTCAGTTCAAGGCTGCTACAAAGCCGAGAAATCAGAGTGCATCTCTGATGTTTACTG
3303 CATCTCCCATGCCAGGCTCTTATTTCCACACCCAGGGACTGTCCTAATGTCCTATCACTAGAGCTGAATGGACCCATTCCAATTCAAATACCTTTCTAGTTTCTG
3408 ATTATATAAATAAAGAGAAAGTATATGTCATATGTTTGCAAAATCCAGGGGATCAAGACTATGATTCATAGCTCCTAAGACCTGAAAACCTCTTCTCAAGAAA
3513 TCFATGTGCAGATTATGAAATGTAATTAATAATCTAATAAATAATTGATGAGCAATA 3569

```



5.4 Nucleotide and amino acid sequences of human ifapsorin cDNA. The full-length nucleotide sequence of human ifapsorin cDNA (9117 bp) was verified by direct cDNA cloning and sequencing and aligned with the predicted open reading frame amino acid sequence. The underlined red nucleotides indicate the canonical poly(A) addition signal. The exon/intron boundaries were determined by the BLAT Search. Sequence data have been deposited under the GenBank™ accession no. AY827490.

Diagram showing Exon I (52 bp), Intron I (1043 bp), Exon II (160 bp), Intron II (1099 bp), and Exon III (3400 bp). Below is the nucleotide sequence (lines 124-910) with the amino acid sequence in red above it. The poly(A) addition signal is underlined in red at the end of the sequence.

## Erklärung

---

Hiermit erkläre ich, dass die vorgelegte Dissertation von mir selbständig verfasst wurde und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt wurden. Die Stellen der Arbeit, einschließlich Abbildungen, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, sind in jedem Einzelfall als Entlehnung kenntlich gemacht. Diese Dissertation hat noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen und wurde, abgesehen von unten angegebenen Teilpublikationen, noch nicht veröffentlicht. Eine solche Veröffentlichung wird von mir auch nicht vor Abschluss des promotions-verfahrens vorgenommen werden. Die Bestimmungen der Promotions-ordnungen sind mir bekannt.

Kiel, den 20.05.2005

Zhihong Wu

Teilpublikationen der vorliegenden Arbeit

Wu Z. and J.-M. Schröder (2004): Identification and expression of two novel splice variants of human p38 $\alpha$ . *J Invest Dermatol.* 123(2):A18. (Abstract)

## Acknowledgment

---

I would like to express my deepest appreciation to Professor Jens-Michael Schröder for his constant support and encouragement. Without his help, this work would not be possible.

I would like to express my heartfelt appreciation Dr Jürgen Harder for his patient, friendly, and unfailing support over the past 2.5 years.

I wish to express my sincere gratitude to Professor Thomas C. G.Bosch for his interest in my work.

I appreciate Dr Joachim Bartels's careful attention to my work. Always willing to tackle and discuss the hard problems, he has been of tremendous assistance with the intellectual progress of this research.

I would like to thank PD. Dr. Regine Gläser and PD. Dr. Daguang Cai for my thesis correction.

I would like to express my appreciation to Bente Rudolph and Eske Voß, who are always willing to help me whenever I need. I would also like to thank Yinghong He for help with my data collection. Dr. Ulf Meyer-Hoffert, Dr. Michael Schunck, Dr. Las Schwichtenberg have also been a source of support and friendship. I am most grateful to them for their very patient and thorough instruction on cell culture and data analysis.

I am deeply indebted to all my colleagues in the Department Dermatology: especially Stephe for help with mass spectrometry analyses; Jutta Quitzau and Claudia Mehrens for help with HPLC analyses; Marlies Brandt for help with SDS-PAGE and western blot analyses. I would also like to thank Frau Sylvia Voß (Department of Medical Microbiology) for help with antimicrobial analyses.

I acknowledge the financial support by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 617.

Finally, I am indebted to my family for encouraging me to pursue this degree. I dedicate this thesis to my wife and my two lovely daughters.

## Curriculum Vitae

---

**Name:** Zhihong Wu  
**Nationality:** Chinese  
**Gender:** male  
**Place of birth:** Zhejiang, China  
**Date of birth:** September 8, 1966

### Education

Sept. 1983 - July 1987      Zhejiang University, China  
(Bachelor's degree)  
Sept. 1991 - July 1994      Zhejiang University, China  
(Master in Plant Pathology)  
Sept. 1998 - Oct. 2000      Zhejiang University, China  
(PhD candidate, Microbiology)  
Dec. 2002 -                      Department of Dermatology, University of Kiel  
(PhD candidate, Cell Biology)

### Scientific Experiences

Oct. 1999 – Nov. 2000      Institute of Biotechnology  
Zhejiang University (research assistant)  
Nov. 2000 - May 2002      Institute of Plant Breeding  
University of Kiel (visiting scientist)  
June 2001 - Nov. 2002      Institute of Plant Breeding  
University of Kiel (assistant researcher)  
Dec. 2002 -                      Department of Dermatology  
University of Kiel (assistant researcher)

### Work Experiences

July. 1994 – Oct. 1999      The International Program Office  
Zhejiang University