การศึกษาลักษณะของโปรติเนส ยีนโปรติเนส และจีโนมของ *VIRGIBACILLUS* SP. SK37 ที่แยกได้จากน้ำปลาไทย

นายเอกราช พรมเหมา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาตรดุษฎีบัณฑิต

สาขาวิชาเทคโนโลยีอาหาร

มหาวิทยาลัยเทคโนโลยีสุรนารี

ปีการศึกษา 2553

CHARACTERIZATION OF PROTEINASES, PROTEINASE GENES, AND GENOME OF *VIRGIBACILLUS* SP. SK37 ISOLATED FROM THAI FISH SAUCE

Ekkarat Phrommao

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Food Technology

Suranaree University of Technology

Academic Year 2010

CHARACTERIZATION OF PROTEINASES, PROTEINASE GENES, AND GENOME OF *VIRGIBACILLUS* SP. SK37 ISOLATED FROM THAI FISH SAUCE

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

Thesis Examining Committee

_____

(Asst. Prof. Dr. Mariena Ketudat-Cairns)
Chairperson

_____

(Assoc. Prof. Dr. Jirawat Yongsawatdigul)
Member (Thesis Advisor)

_____

(Assoc. Prof. Dr. Montarop Yamabhai)
Member

_____

(Prof. Dr. James L. Steele)
Member

_____

(Asst. Prof. Dr. Sureelak Rodtong)
Member

_____

(Prof. Dr. Sukit Limpijumnong)
Vice Rector for Academic Affairs

_____

(Asst. Prof. Dr. Suwayd Ningsanond)
Dean of Institute of Agricultural Technology

เอกราช พรมเหมา : การศึกษาลักษณะของโปรติเนส ยีนโปรติเนส และจีโนมของ *VIRGIBACILLUS* SP. SK37 ที่แยกได้จากน้ำปลาไทย (CHARACTERIZATION OF PROTEINASES, PROTEINASE GENES, AND GENOME OF *VIRGIBACILLUS* SP. SK37 ISOLATED FROM THAI FISH SAUCE) อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร. จิรวัฒน์ ยงสวัสดิกุล, 248 หน้า.

Virgibacillus sp. SK37 เป็นแบคทีเรียแกรมบวกที่ชอบเกลือปานกลางซึ่งแยกได้จากน้ำปลาไทย แบคทีเรียสายพันธุ์นี้มีศักยภาพในการใช้เป็นกล้าเชื้อสำหรับการผลิตน้ำปลาเนื่องจากมีความสามารถสูงในการผลิตโปรติเนส วัตถุประสงค์ของการศึกษานี้คือเพื่อระบุ ศึกษาลักษณะ และโคลนโปรติเนสจากแบคทีเรียสายพันธุ์นี้ รวมถึงการวิเคราะห์จีโนมและยีนที่มีส่วนร่วมในการปรับตัวในแหล่งอาศัยที่มีเกลือและโปรตีนสูง

โปรติเนสที่ปลดปล่อยออกมาหกชนิดจากทั้งหมดสิบสองชนิด เป็นเอนไซม์ที่คล้ายสับทิลิซิน โปรติเนสสามชนิดจากหกชนิดดังกล่าวมีขนาดมวลโมเลกุลเท่ากับ 19, 34, และ 44 กิโลดาลตัน เมื่อผ่านการทำให้บริสุทธิ์บางส่วนแสดงกิจกรรมที่เหมาะสมที่พีเอช 8 และอุณหภูมิ 55 ถึง 65 องศาเซลเซียส โซเดียมคลอไรด์และแคลเซียมคลอไรด์สามารถกระตุ้นกิจกรรมของเอนไซม์โดยให้ค่ากิจกรรมสูงสุดที่ความเข้มข้นร้อยละ 30 และ 10 มิลลิโมลาร์ ตามลำดับ การตรวจวิเคราะห์เปปไทด์ที่ถูกย่อยโดยทริปซินด้วยเทคนิคลายพิมพ์มวลเปปไทด์และเทคนิคการหาลำดับเปปไทด์บางส่วนด้วยเทคนิค MALDI-TOF และ Tandem Mass พบว่าทั้งสามเอนไซม์เป็นชนิดใหม่และมีความคล้ายคลึงกับบาซิลโลเปปทิเดสเอฟ (bacillopeptidase F) คลังจีโนมของเชื้อ SK37 สร้างใน *E. coli* DH10B และคัดเลือกเพื่อค้นหายีนโปรติเนส ซึ่งพบหนึ่งยีนที่ถอดรหัสได้เอนไซม์ที่คล้ายสับทิลิซินและมีความเหมือนกับ AprX จากบาซิลลัสสายพันธุ์อื่นๆ เอนไซม์รีคอมบิแนนท์ซึ่งมีชื่อว่า AprX-SK37 แสดงกิจกรรมที่พีเอช 9.5 อุณหภูมิ 55 องศาเซลเซียส และโซเดียมคลอไรด์เข้มข้น 1.0 โมลาร์ เอนไซม์นี้จำเป็นต้องมีแคลเซียมเพื่อแสดงกิจกรรมโดยระดับความเข้มข้นที่เหมาะสมคือ 15 มิลลิโมลาร์ เอนไซม์มีความต้านทานต่อไฮโดรเจนเปอร์ออกไซด์ที่ความเข้มข้นสูงถึงร้อยละ 5 การวิเคราะห์ความสัมพันธ์ทางพันธุกรรมบ่งชี้ว่า AprX-SK37 เป็นวงศ์ใหม่ในวงศ์ใหญ่ (superfamily) ของซับทิเลส และมีความเสถียรต่อการเกิดออกซิเดชั่นและสามารถทนเกลือในระดับปานกลาง จีโนมทั้งหมดของสายพันธุ์ SK37 ประกอบด้วยโครโมโซมขนาด 3.8 เมกะเบส และสามพลาสมิดขนาด 4.1, 4.4, และ 12.1 กิโลเบส และประกอบด้วยลำดับรหัสพันธุกรรมของโปรตีน 3,823 ชนิด การวิเคราะห์จีโนมแสดงการนำเข้ากรดอะมิโนอย่างมีประสิทธิภาพและวิถีการสังเคราะห์ของสารให้กลิ่นรสในน้ำปลาจากอนุพันธ์ของกรดอะมิโน ผลการวิเคราะห์จีโนมแสดงถึงความสามารถใน

การปรับตัวในสิ่งแวดล้อมที่มีเกลือสูงของสายพันธุ์ SK37 เนื่องจากมียีนเพื่อการควบคุมและนำเข้าโพแทสเซียม โซเดียม และสารควบคุมความดันออสโมติก (osmoprotectants) การขาดยีน *hpr* และ *scoC* ที่ผลิตโปรตีนที่หน่วงการแสดงออกของโปรติเนสอาจอธิบายได้ถึงความสามารถในการผลิตโปรติเนสที่สูงของสายพันธุ์ SK37 จีโนมของ SK37 แสดงถึงยีนร่วมบรรพบุรุษ (orthologs) ที่คล้ายกับบาซิลลัสสายพันธุ์อื่นๆ โดยพบค่าความเหมือนสูงสุดในในแบคทีเรียแกรมบวกที่ชอบเกลือสายพันธุ์ *Oceanobacillus iheyensis* ความรู้เกี่ยวกับคุณลักษณะของโปรติเนสและจีโนมของ SK37 มีความสำคัญต่อความเข้าใจในสิ่งมีชีวิตในสิ่งแวดล้อมที่มีเกลือและโปรตีนสูงและการนำไปใช้ในการหมักน้ำปลา

สาขาวิชาเทคโนโลยีอาหาร ลายมือชื่อนักศึกษา_____

ปีการศึกษา 2553 ลายมือชื่ออาจารย์ที่ปรึกษา_____

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม_____

EKKARAT PHROMMAO : CHARACTERIZATION OF PROTEINASES, PROTEINASE GENES, AND GENOME OF *VIRGIBACILLUS* SP. SK37 ISOLATED FROM THAI FISH SAUCE. THESIS ADVISOR : ASSOC. PROF. JIRAWAT YONGSAWATDIGUL, Ph.D., 248 PP.

SUBTILISIN-LIKE PROTEINASES/HALOTOLERANCE/BACILLOPEPTIDASE F/APRX/GENOMIC LIBRARY CONSTRUCTION/GENOME SEQUENCE/ *VIRGIBACILLUS*/ FISH SAUCE

Virgibacillus sp. SK37 is a Gram-positive moderately halophilic bacterium isolated from Thai fish sauce. This bacterium has the potential for use as a starter culture for fish sauce production as it shows high proteinase production. Objectives of this study were to identify, characterize, and clone proteinases from this strain. Complete genome sequence of this bacterium was determined and genes involved in adaptation to high saline and protein-rich niche of fish sauce were analyzed.

Six out of twelve extracellular proteinases were identified as subtilisin-like enzymes, three of which with molecular mass of 19, 34, and 44 kDa were partially-purified. All three enzymes showed optimum activity at pH 8 and at 55-60ºC. These enzymes were activated by NaCl and $CaCl_2$ showing maximal activity at 30% NaCl and 100 mM $CaCl_2$. Peptide mass fingerprint and *de novo* peptide sequencing of tryptic digested peptides analyzed by MALDI-TOF and tandem mass spectrometry, respectively, suggested that all three enzymes were novel and homologous to bacillopeptidase F. Genomic library of SK37 was constructed in *E. coli* DH10B and screened in order to identify proteinase gene. One gene encoding a subtilisin-like

enzyme with a high homology to AprX of *Bacillus* species was obtained. The recombinant enzyme designated as AprX-SK37 exhibited activity at pH 9.5, 55°C, and 1.0 M NaCl. The enzyme was absolutely calcium-dependent showing optimum activity at 15 mM $CaCl_2$. The enzyme had a resistance to H_2O_2 up to 5%. Phylogenetic analysis suggested that AprX-SK37 belongs to a novel family of subtilases superfamily with oxidant stable and moderately halotolerant properties. The genome of SK37 consists of 3.8 Mbp of a circular chromosome and 4.1-, 4.4-, and 12.1-kbp plasmids with 3,823 protein coding sequences. Analysis of the genome sequence shows efficient uptake and utilization of amino acids and pathways for the synthesis of fish sauce flavors from amino acid derivatives. The genome reveals that SK37 is well equipped for adaptation to high saline environments by possessing genes for regulation and uptake of potassium, sodium, and osmoprotectants. The lacks of *hpr* and *scoC* genes, encoding proteinase repressors, may explain the high proteinase production ability of SK37. Genome of strain SK37 shows highly conserved orthologous genes to *Bacillus* species with highest similarity to Gram positive halophilic *Oceanobacillus iheyensis*. Knowledge about proteinase characteristics and genome sequence of SK37 is critical for understanding life in saline and protein-rich environments and applications in fish sauce fermentation.

School of Food Technology

Academic Year 2010

Student's Signature _____

Advisor's Signature _____

Co-advisor's Signature _____

ACKNOWLEDGMENTS

For six years of seemingly endless journey, I have worked and met with a great number of wonderful people whose contribution in assorted ways made this thesis possible. It is my pleasure to convey my gratitude to them all in my humble acknowledgement.

In the first place I would like to show my gratitude to my advisor, Assoc. Prof. Jirawat Yongsawatdigul. He has made available his support in a number of ways from the very early stage of this thesis and throughout the work. His truly scientist intuition has made him as an oasis of idea and passion for his students and exceptionally inspired me and enrich my growth to be a scientist I want to be. I am indebted to him more than he knows. I am grateful to acknowledge my co-advisor, Assoc. Prof. Montarop Yamabhai. Her involvement in this thesis has triggered and nourished my molecular biology practical skills that I will benefit from, for a long time to come. I would like to thank Asst. Prof. Dr. Mariena Ketudat-Cairns for accepting to be a member of my thesis defense. I gratefully thank Asst. Prof. Sureelak Rodtong for her crucial advises and constructive comments on this thesis. It is a pleasure to thank Prof. James L. Steele for accepting to be the collaborator on this thesis. It is an honor for me to work with him. I am benefited by his outstanding works and his particular skill in genome sequencing. Without his contribution, this part of my thesis would not have been existed. I hope to keep up our collaboration in the future. I am indebted to my

co-worker, Elena Lara. Again, without her, the genome sequencing project would not have been done on time. There was a lot of fun and joyful times working with her.

I am indebted to many of my colleagues both from the Protein Research Unit in SUT and members in Department of Food Science at University of Wisconsin-Madison to support my research and private life. There are too many to name them all here but they will always be in my mind. I owe my deepest gratitude to Nicole Denison (NiKi). She is one of wonderful people who ever came into my life. I adore her like my biological mother as her motherhood with relentless loves and cares make my life in the U.S.A. far beyond from the word “Comfort”. She made me feel living in Madison is living in my hometown.

Furthermore, I would like to thank the Royal Golden Jubilee Ph.D. program. My life has been changed tremendously ever since I was awarded this scholarship. It was an extraordinary fortune that ever appeared in my life.

My parents deserve special dearest mention for their support and prayers. Their caring and gently loves define the perfect meaning of parenthood. Even though both of them finished only primary school level for their highest education, their hard work for the most of their life has proven me and my siblings their intelligence which out-values a piece of paper of diploma certificate. Only one thing they asked for us was hard work and to focus only on school life. Dear Mom and Dad, “you did your works and I (now) did mine”.

This is not the end of the road but rather only the walk before the run.

Ekkarat Phrommao

CONTENTS

Page

CONTENTS (Continued)

Page

CONTENTS (Continued)

Page

CONTENTS (Continued)

Page

CONTENTS (Continued)

Page

CONTENTS (Continued)

Page

CONTENTS (Continued)

Page

CONTENTS (Continued)

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF FIGURES (Continued)

Figure **Page**

LIST OF FIGURES (Continued)

Figure **Page**

LIST OF ABBREVIATIONS

A	=	Adenine
Ala (A)	=	Alanine
Arg (R)	=	Arginine
Asn (N)	=	Asparagine
Asp (D)	=	Aspartic acid
β-ME	=	β-Mercaptoethanol
C	=	Cytosine
°C	=	Degree celsius
cDNA	=	complimentary DNA
CDS	=	Protein coding sequence
CID	=	Collision induced dissociation
CIP	=	Calf intestinal phosphatase
COG	=	Clusters of orthologous group
CRISPR	=	Clustered regularly interspaced short palindromic repeats
Cys (C)	=	Cysteine
2D	=	2-Dimentional
DEAE	=	Diethylaminoethyl
DTT	=	Dithiothreitol
E-64	=	*Trans*-epoxysuccinyl-L-leucylamido-(4-guanidine)-butane
EDTA	=	Ethylenediaminetetraacetic acid

LIST OF ABBREVIATIONS (Continued)

ESI	=	Electrospray ionization
EST	=	Expressed sequence tag
G	=	Guanine
Glu (E)	=	Glutamic acid
Gln (Q)	=	Glutamine
Gly (G)	=	Glycine
h	=	Hour
His (H)	=	Histidine
IAA	=	Iodoacetic acid
IEF	=	Isoelectric focusing
Ile (I)	=	Isoleucine
kDa	=	kilo Dalton (10^3 Dalton)
Leu (L)	=	Leucine
Lys (K)	=	Lysine
M	=	mol l^{-1}
mA	=	Milliampare
MALDI	=	Matrix-assisted laser desorption/ionization
Met (M)	=	Methionine
μg	=	Microgram (10^{-6} gram)
μl	=	Microliter (10^{-6} l)
μM	=	Micromolar (10^{-6} mol l^{-1})

LIST OF ABBREVIATIONS (Continued)

mg	=	Milligram (10^{-3} gram)
min	=	Minute
ml	=	Milliliter (10^{-3} l)
mM	=	Millimolar (10^{-3} mol l^{-1})
MM	=	Molecular mass
MS	=	Mass spectrometry
MS/MS	=	Tandem mass
MWCO	=	Molecular weight cutoff
m/z	=	Mass per charge ratio
NEM	=	N-Ethlymaleimide
Ni-NTA	=	Nickel nitrilotriacetic acid
nm	=	Nanometer (10^{-9} meter)
PAGE	=	Polyacrylamide gel electrophoresis
Phe (F)	=	Phenylalanine
p*I*	=	Isoelectric point
PMF	=	Peptide mass fingerprint
PMSF	=	Phenylmethanesulfonyl fluoride
pNA	=	*p*-Nitroanilide
Pro (P)	=	Proline
Q	=	Quaternary ammonium
s	=	Second
SDS	=	Sodium dodecyl sulfate

LIST OF ABBREVIATIONS (Continued)

Ser (S)	=	Serine
SNP	=	Single nucleotide polymorphism
Suc	=	Succinyl
T	=	Thymine
TCA	=	Trichloroacetic acid
Thr (T)	=	Threonine
TLCK	=	*N*-Tosyl-L-lysine chloromethyl ketone
TOF	=	Time-to-flight
TPCK	=	*N*-Tosyl-L-phenylalanine chloromethyl ketone
Tris	=	Tris(hydroxymethyl)aminomethane
Trp (W)	=	Tryptophan
Tyr (Y)	=	Tyrosine
U	=	Unit activity
V	=	Volt
v/v	=	Volume by volume
Val (V)	=	Valine
w/v	=	Weight by volume
xg	=	Gravitational acceleration

CHAPTER I

INTRODUCTION

1.1 Introduction

Fermented salted sauces from fish and soy have long been consumed as a condiment by Asian people. Fish sauce provides stronger and more complicated umami taste than that of soy sauce (Sanceda, Suzuki, and Kurata, 2003). This product is popular not only in Asia but currently outspreads worldwide. Thailand is one of the largest fish sauce producers with annual production of more than 400,000 metric tons (Bovornreungroj, 2005). In 2008, approximately 38,800 metric tons of Thai fish sauce were exported globally and accounted to about one billion Thai baht, while the demand of this product has been increasing over years (Food intelligent center, National food institute, Thailand, http://fic.nfi.or.th/th/thaifood/product52-condiment.asp). Traditional Thai fish sauce production is a natural fermentation of anchovy (*Stolephorus* sp.) at high salt content of 28- 30%) (Yongsawatdigul, Rodtong, and Raksakulthai, 2007). The fermentation mainly relies on fish protein hydrolysis by the action of bacterial proteinases and endogenous fish proteinases. It is a slow process and usually takes 1-1.5 year for complete hydrolysis and flavor development to yield a premium quality fish sauce. This extremely long fermentation time is a major barrier for the growth of fish sauce industry.

There have been a number of attempts focusing on the acceleration of fish sauce fermentation using acids and alkaline (Gildberg and Shi, 1994; Thongthai and

Suntinanalert, 1991), enzymes (Beddow and Ardeshir, 1979; Fu, You, and Kim, 2008; Gildberg, Hermes, and Orejana, 1984), raising temperature (Lopetcharat and Park, 2002), and microorganisms (Owens and Mendoza, 1985; Greig and Esterlla, 1988; Caveesuk, Smith, and Simson, 1993). However, these processes are considered to be harsh and costly, while the failure of retaining of original characteristics of fish sauce was often observed. Accordingly, these measures have never been practiced at industrial scale. Since it appears that different kinds of peptides produced by the hydrolysis of fish proteins as well as bacteria per se may play a role in forming the distinct flavor and aroma that, in turn, determine the quality of the fish sauce (Beddows and Ardeshir 1979), while the hydrolytic process is considered to be the main point for increasing the rate of fermentation, therefore, for over 20 years, numerous studies have been focusing on using proteinase-producing bacteria isolated from fish sauce to accelerate the fermentation process. In Thai fish sauce, *Halobacterim* sp., *Halococcus* sp., (Thongthai et al., 1991), *Halobacillus thailandensis* (Chaiyanan, Maugel, Huq, Robb, and Colwell, 1999), *Filobacillus* sp RF2-5 (Hiraga, Namwong, Tanasupawat, Takada, and Oda, 2005), and *Halobacillis* sp. SR5-3 (Namwong, Hiraga, Takada, Tsunemi, Tanasupawat, and Oda, 2006) were isolated and their extracellular proteinases were characterized. The secreted enzymes were mainly serine proteinases giving a high catalytic activity at high salt concentrations up to 30% NaCl.

In Vietnamese fish sauce, anchovy sauce, and fermented fish paste, *Bacillus* species. have been isolated (To, Tanaka, and Nagano, 1997; Lee, Kim, Ahn, Lee, Kim, and Chung, 1989; Nagano and To, 2000; Tran and Nagano, 2002; Kim, Nishiyama, Mura, Tokue, and Arai, 2004; Uchida, Kondo, Yamashita, Tanaka, and

Nagano, 2004; Kim et al., 2004; Setyorini, Takenaka, Murakami, and Aoki, 2006). These bacteria regularly excrete a large amount of serine proteinases exhibiting halophilic activity. Only trypsin-like serine proteinases from *B. subtilis* JM-3 was a partial halophile, which was dramatically inhibited at high salt content with 1.3% remaining activity at 30% (v/w) NaCl (Kim and Kim, 2005).

Bacteria isolated from fish sauce and related salted products are usually members of moderately or extremely halohpilic bacteria/archaea exhibiting optimum growth between 3% to saturate NaCl (Kushner, 1978; Johnson, Lanthier, and Gochnauer, 1986; Mongodin et al., 2005). Proteinases produced from this group of microorganisms mostly inherit halophilic property showing optimum catalytic activity at a wide range of NaCl content. For centuries, enzymes have been used as processing aids in various food industrial applications (Whitaker, 1994). Proteinases are one of the largest selling enzymes accounting about 60% of the total enzyme market worldwide (Rao, Tanksale, Ghatge, and Deshpande, 1998), 89% of which has been used in detergent formulation (Gupta, Beg, and Lorenz, 2002), others are used in a number of industrial applications, including peptide synthesis, protein processing, feed, leather treatment, dairy, chill proofing, meat tenderization, and fermented sauces (Maugh, 1984; Shimogaki, Takeuchi, Nishino, Ohdera, Kudo, and Ohba, 1991). With respect to halophilic proteinases, they have found their way to suit in various industrial processes, such as in fermented food, bioreactor systems, biotechnological processes, production of biopolymers, pharmaceuticals, and degradation of toxic compounds (Ventosa and Nieto, 1995; Ventosa, Nieto, and Oren, 1998; Hough and Danson, 1999; Margesin and Schiner, 2001; Van den Berg, 2003; Gomes and Steiner, 2004). For example, a SDS-resistant alkaline proteinase from *B. subtilis* CN2 isolated

from Vietnamese fish sauce showed potential in detergent application. (Uchida et al., 2004).

Investigation of active proteinases in biological samples is a challenge for proteomic study. Proteinases are a complex group of enzymes and vary enormously in their biochemical and catalytic properties. Therefore, studies of their nucleotide and amino acid sequence homology are important for a variety of purposes and have become routine in computational molecular biology (Argos, 1987). A number of molecular tools have been applied for identification of proteolytic enzymes in microorganisms. Identification techniques of bacterial proteinases include proteomic based methods (*i.e.* Mass spectrometry) (Simpson, 2003), Polymerase Chain Reaction (PCR)-based cloning from degenerated primers (Shi, Tang, Huang, Gan, Tang, and Shen, 2006; Okuda, Sumitomo, Takimura, Ogawa, Saeki, and Kawai, 2004; Kim et al., 2004), DNA probe hybridization (Maciver, McHale, Saul, and Bergquist, 1994; Miyaji, Otta, Nakagawa, Watanabe, Niimura, and Tomizuka, 2006), and genomic library construction (Tran, Wu, and Wong, 1991; Zabolotskaya, Demidyuk, Akimkina, and Kostrov, 2004; Sareen, Bornscheuer, and Mishra, 2005; Tang, Shen, Lakay, Shao, Wang, and Prior, 2004; Maciver et al., 1994). Mass spectroscopy and genomic library construction are more advantageous than others since no primary genetic information is required and more than one active enzyme could be obtained from one experiment. Recently, whole genome sequencing has become a promising technology since not only all proteinase genes contained in a bacterial genome can be simultaneously identified but also comprehensive understandings of genetics, biochemistry, and molecular biology of bacteria are obtained. The whole genome sequencing projects of various food grade lactic acid bacteria for dairy industry,

particularly cheese and yogurt (Mayo, van Sinderen, and Ventura, 2008), as well as *Aspergillus oryzae* for various fermentation industries such as soy sauce, sake, bean curd, and vinegar production (Masayuki, Kiyoshi, Toshihiro, Toshitaka, and Goro, 2005), have been employed. These accomplishments have undoubtedly revolutionized fermentation technology of the respective industries.

Nawong (2006) isolated and screened proteinase-producing microorganisms from Thai fish sauce samples. From a total of 308 bacterial isolates, three strains showing the highest proteinase activity were obtained, including *Virgibacillus* sp. SK33, *Virgibacillus* sp. SK37, and *Staphylococcus* sp. SK1-1-5. When using these strains as a starter culture for fish sauce fermentation, total amino acid content of 4-month-old fish sauce were comparable to that of 12-month-traditionally-fermented sample. Besides, both SK33 and SK37 appeared to decrease up to 50% of histamine content (an indicator for biogenic amines found mostly in fermented food and causes immune system disorder and allergies in human) (Silla Santos, 1996; Yongsawatdigul et al. 2007). *Virgibacillus* sp. SK37 (GenBank accession number DQ910840) is Gram positive, endospore-forming, rod-shaped, aerobic bacterium, and grows at a wide pH range of 4 to 11 (optimum 6.5 to 7.5), 20 to 45°C (optimum 30 to 40°C) and 0 to 25% NaCl (optimum 5 to 10%) (Nawong, 2006). With respect to proteinase production, this strain secreted at least 10 proteinases with molecular mass (MM) of 198, 98, 84, 81, 67, 63, 50, 38, 23, and 18 kDa, which is higher than that of *Virgibacillus* sp. SK33 (totally 6 proteinases) (Sinsuwan, Rodtong, and Yongsawatdigul, 2007; 2008a, 2010). Moreover, an active cell-bound proteinase of *Virgibacillus* sp. SK37 was reported (Sinsuwan et al., 2008b) Proteinases in both extracellular and cell-bound fractions were halotolerant showing activity at up to 20% NaCl. As this bacterium and its

proteinases are promising to be developed as a starter culture or a processing aid, respectively, for fish sauce production or other high saline systems, the knowledge of its halotolerant proteinases, proteolytic system, and genetic information should be established.

1.2 Research objectives

The objectives of this study were:

1. To purify, characterize, and identify halotolerant proteinases produced by *Virgibacillus* sp. SK37
2. To clone, overexpress, and characterize a selected proteinase gene isolated from *Virgibacillus* sp. SK37 genome.
3. To establish a comprehensive understanding of genetic, biochemistry, and molecular biology of *Virgibacillus* sp. SK37 based on its whole genome sequence

1.3 Research hypotheses

Proteinases produced from *Virgibacillus* sp. SK37 are halotolerant exhibiting activity in a wide range of NaCl content. Genetic information is important for proteinase classification, which would lead to discovery of novel enzymes produced from this bacterium. All genes encoding proteolytic enzymes of *Virgibacillus* sp. SK37 could be identified by whole genome sequencing. This information could also unravel other genes that are significant for adaptation to a high saline protein-rich environment like fish sauce.

1.4 Scope of the study

Major extracellular proteinases with halotolerant activity were purified and determined their biochemical properties. Protein identification was investigated using mass spectrometry. Genomic library of *Virgibacillus* sp. SK37 was constructed in an *E. coli* system in order to identify proteinase gene. The obtained gene was sequenced, cloned, and overexpressed, and the recombinant enzyme was biochemically characterized. Protein classification was carried out using its nucleotide and deduced amino acid sequences together with the biochemical characteristics. Whole genome sequence of *Virgibacillus* sp. SK37 was determined. Its genome sequence was analyzed using a combination of various sources of bioinformatics tools. Obtained information was used to establish the complex relationship between genetic and molecular biology of this bacterium.

1.5 Expected results

Halotolerant property of proteinases from *Virgibacillus* sp. SK37 is suitable for protein degradation in high saline environments. These enzymes could be used as a processing aid to shorten fish sauce fermentation time. Protein identification and classification of the proteinases based on their amino acid and nucleotide sequences will enable more understandings on their genetic and biochemical properties. Their genetic information could be used as a material for further study of halotolerant property of this group of enzymes in molecular level. Complete genome sequence of *Virgibacillus* sp. SK37 will be a powerful tool for investigation the intricate relationship of proteolytic system and other genomic industrial relevant characteristics of this strain. This information can be used to predict biosynthetic and degradation

capabilities and other industrial significances which will be a starting point for starter culture technology of *Virgibacillus* sp. SK37. This will be the keystone of fish sauce industry research in order to achieve quality consistency and desired characteristics of finished product.

1.6 References

Argos, P. (1987). A sensitive procedure to compare amino acid sequences. **J. Mol. Biol.** 193: 385-396.

Beddows, C. G., and Ardeshir, A. G. (1979). The production of soluble fish protein solution for use in fish sauce manufacture: The use of added enzymes. **J. Food Tech.** 14: 603-612.

Bovornreungroj, P. (2005). **Selection of extremely halophilic bacteria producing halophilic proteolytic enzyme for fish sauce production**, Ph.D. Dissertation, Kasetsart University, Bangkok.

Chaiyanan, S., Maugel, T., Huq, A., Robb, F. T., and Colwell, R. R. (1999). Polyphasic taxonomy of a novel halobacillus, *Halobacillus thailandensis* sp. nov. isolated from fish sauce. **Syst. Appl. Microbiol.** 22: 360-365.

Fu, X. T., You, S. G., and Kim, S. M. (2008). Characterization of a salt-tolerant acid proteinase produced by *Bacillus megaterium* KLP-98 and its protential as a fermentation starter for the manufacture of fish sauce. **J. Food Biochem.** 32: 279-298.

Gildberg, A., Espejo-Hermes, J., and Magno-OreJana, F. (1984). Acceleration of autolysis during fish sauce fermentation by adding acid and reducing the salt content. **J. Sci. Food Agric.** 35: 1363-1369.

Gildberg, A., and Shi, X. Q. (1994). Recovery of tryptic enzymes from fish sauce. **Process Biochem.** 29: 151-155.

Gomes, J., and Steiner, W. (2004). The biocatalytic potential of extremophiles and extremozymes. **Food Technol. Biotechnol.** 42: 223-235.

Greig, R. W., and Estrella, D. C. (1988). A study of the acceleration of fish sauce production using enzyme. **Food science and technology in industrial development, Proceeding of the food conference** (Vol. 1). Bangkok, Thailand.

Gupta, R., Beg, Q. K., and Lorenz, P. (2002). Bacterial alkaline proteases: molecular approaches and industrial applications. **Appl. Microbiol. Biotechnol.** 59:(1) 15-32.

Gupta, A., Joseph, B., Mani, A., and Thomas, G. (2008). Biosynthesis and properties of an extracellular thermostable serine alkaline proteinase from *Virgibacillus pantothenticus*. **World. J. Microbiol. Biotechnol.** 24: 237-243.

Hiraga, K., Y., N., Namwong, S., Tanasupawat, S., Takada, K., and Oda, K. (2005). Purification and characterization of serine proteinase from a halophilic bacterium, *Filobacillus* sp. RF2-5. **Biosci. Biotechnol. Biochem.** 69:(1) 38-44.

Hough, W. D., and Danson, J. M. (1999). Extremozymes. **Curr. Opin. Chem. Biol.** 3: 39-46.

Johnson, K. G., Lanthier, P. H., and Gochnauer, M. B. (1986). Studies of two strains of *Actinopolyspora halophila*, an extremely halophilic actinomycete. **Arch. Microbiol.** 143: 370-378.

Kim, M., Nishiyama, Y., Mura, K., Tokue, C., and Arai, S. (2004). Gene cloning and characterization of a *Bacillus vietnamensis* Metalloprotease. **Biosci. Biotechnol. Biochem.** 68:(7) 1533-1540.

Kim, W. J., and Kim, S. M. (2005). Purification and characterization of *Bacillus subtilis* JM-3 protease from anchovy sauce. **J. Food Biochem.** 29: 591-610.

Kushner, D. J. (1978). Life in high salt and solute concentrations: halophilic bacteria. In D. J. Kushner (ed.). **Microbial life in extreme environments** (pp. 317-368). London: Academic Press, Ltd.

Lee, E., Kim, J., Ahn, C., Lee, K., Kim, M., and Chung, B. (1989). Processing conditions of accelerated anchovy sauce extracts. **J. Kor. Soc. Food Nutr.** 18: 167-173.

Lopetcharat, K., and Park, J. W. (2002). Characteristics of fish sauce made from Pacific whiting and surimi by-products during fermentation stage. **J. Food Sci.** 67: 511-516.

Maciver, B., McHale, R. H., Saul, D. J., and Bergquist, P. L. (1994). Cloning and sequencing of a serine proteinase gene from a thermophilic *Bacillus* species and its expression in *Escherichia coli*. **Appl. Environ. Microbiol.** 60:(11) 3981-3988.

Margesin, R., and Schiner, F. (2001). Potential of halotolerant and halophilic microorganisms for biotechnology. **Extremophiles.** 5: 73-83.

Masayuki, M., Kiyoshi, A., S., M., Toshihiro, T., Toshitaka, K., and Goro, T. (2005). Genome sequencing and analysis of *Aspergillus oryzae*. **Nature.** 438: 1157-1161.

Maugh, T. H. (1984). A renewed interest in immobilized enzymes. **Science.** 223: 474-476.

Mayo, B., van Sinderen, D., and Ventura, M. (2008). Genome analysis of food grade lactic acid-producing bacteria: from basics to applications. **Curr. Genomics.** 9:(3) 169-183.

Miyaji, T., Otta, Y., Nakagawa, T., Watanabe, T., Niimura, Y., and Tomizuka, N. (2006). Purification and molecular characterization of subtilisin-like alkaline protease BPP-A from *Bacillus pumilus* strain MS-1. **Lett. Appl. Microbiol.** 42:(3) 242-247.

Mongodin, E. F., Nelson, K. E., Daugherty, S., Deboy, R. T., Wister, J., Khouri, H., et al. (2005). The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. **Proc. Natl. Acad. Sci. U S A**.102:(50) 18147-18152.

Nagano, H., and To, K. A. (2000). Purification of collagenase and specificity of its related enzyme from *Bacillus subtilis* FS-2. **Biosci. Biotechnol. Biochem.** 64: 181-183.

Namwong, S., Hiraga, K., Takada, K., Tsunemi, M., Tanasupawat, S., and Oda, K. (2006). A halophilic serine proteinase from *Halobacillus* sp. SR5-3 isolated from fish sauce: purification and characterization. **Biosci. Biotechnol. Biochem.** 70:(6) 1395-1401.

Nawong, S. (2006). **Isolation, selection and identification of proteinase-producing bacteria from fish sauce fermentation to be used as starter cultures.** M.S. thesis, Suranaree University of Technology, Nakhon-Ratchasima.

Okuda, M., Sumitomo, N., Takimura, Y., Ogawa, A., Saeki, K., and Kawai, S. (2004). A new subtilisin family: nucleotide and deduced amino acid sequences of new high-molecular-mass alkaline proteases from *Bacillus* spp. **Extremophiles.** 8:(3) 229-235.

Owens, J. D., and Mendoza, L. S. (1985). Enzymically hydrolysed and bacterially fermented fishery products. **Int. J. Food Sci. Tech.** 20:(3) 273-293.

Rao, M. B., Tanksale, A. M., Ghatge, M. S., and Deshpande, V. V. (1998). Molecular and biotechnological aspects of microbial proteases. **Microbiol. Mol. Biol. Rev.** 62:(3) 597-635.

Sanceda, N., Suzuki, E., and Kurata, T. (2003). Quality and sensory acceptance of fish sauce partially substituting sodium chloride or natural salt with potassium chloride during the fermentation process. **Int. J. Food Sci. Tech.** 38: 435-443.

Sareen, R., Bornscheuer, U. T., and Mishra, P. (2005). Cloning, functional expression and characterization of an alkaline protease from *Bacillus licheniformis*. **Biotechnol. Letters.** 27: 1901-1907.

Setyorini, E., Takenaka, S., Murakami, S., and Aoki, K. (2006). Purification and characterization of two novel halotolerant extracellular proteases from *Bacillus subtilis* strain FP-133. **Biosci. Biotechnol. Biochem.** 70:(2) 433-440.

Shi, W., Tang, X. F., Huang, Y., Gan, F., Tang, B., and Shen, P. (2006). An extracellular halophilic protease SptA from a halophilic archaeon *Natrinema* sp. J7: gene cloning, expression and characterization. **Extremophiles.** 10:(6) 599-606.

Shimogaki, H., Takeuchi, K., Nishino, T., Ohdera, M., Kudo, T., and Ohba, K. (1991). Purification and properties of a novel surface-active agent- and

alkaline-resistant protease from *Bacillus* sp. **J. Agric. Biol. Chem.** 55:(9) 2251-2258.

Silla Santos, M. H. (1996). Biogenic amines: their importance in foods. **Int. J. Food Microbiol.** 29:(2-3) 213-231.

Simpson, R. J. (2003). **Proteins and proteomics: a laboratory manual**. (pp. 163-180). New York: Cold Spring Harbor Laboratory Press.

Sinsuwan, S., Rodtong, S., and Yongsawatdigul, J. (2007). NaCl-activated extracellular proteinase from *Virgibacillus* sp. SK37 isolated from fish sauce fermentation. **J. Food Sci.** 72:(5) 264-269.

Sinsuwan, S., Rodtong, S., and Yongsawatdigul, J. (2008a). Production and characterization of NaCl-activated proteinases from *Virgibacillus* sp. SK33 isolated from fish sauce fermentation. **Process Biochem.** 43: 185–192.

Sinsuwan, S., Rodtong, S., and Yongsawatdigul, J. (2008b). Characterization of Ca^{2+}-activated cell-bound proteinase from *Virgibacillus* sp. SK37 isolated from fish sauce fermentation. **Food Sci. Technol.** 41:(10) 2166-2174.

Sinsuwan, S., Rodtong, S., and Yongsawatdigul, J. (2010). A NaCl-stable serine proteinase from *Virgibacillus* sp. SK33 isolated from Thai fish sauce. **Food Chem.** 119: 573-579.

Stepanov, V. M., Rudenskaya, G. N., Revina, L. P., Gryaznova, Y. B., Lysogorskaya, E. N., Filippova, I., et al. (1992). A serine proteinase of an archaebacterium, *Halobacterium mediterranei*. A homologue of eubacterial subtilisins. **Biochem. J.** 285:(Pt 1) 281-286.

Tang, X. M., Shen, W., Lakay, F. M., Shao, W. L., Wang, Z. X., and Prior, B. A. (2004). Cloning and over-expression of an alkaline protease from *Bacillus licheniformis*. **Biotechnol. Lett.** 26:(12) 975-979.

Thongthai, C., and Suntinanalert, P. (1991). **Halophiles in Thai fish sauce (Nam Pla)**. New York: Plenum Press.

To, K. A., Tanaka, T., and Nagano, H. (1997). Isolation of collagenase producing bacteria from traditional foods. **J. Home Econ. Jpn.** 12: 1083-1087.

Tran, L., Wu, X. C., and Wong, S. L. (1991). Cloning and expression of a novel protease gene encoding an extracellular neutral protease from *Bacillus subtilis*. **J. Bacteriol.** 173:(20) 6364-6372.

Tran, L. H., and Nagano, H. (2002). Isolation and characteristics of *Bacillus subtilis* CN2 and its collagenase production. **J. Food Sci.** 67:(3) 1184-1187.

Uchida, H., Kondo, D., Yamashita, S., Tanaka, T., Tran, L. H., and Nagano, H. (2004). Purification and properties of a protease produced by *Bacillus subtilis* CN2 isolated from a Vietnamese fish sauce. **J. Microbiol. Biotechnol.** 20: 579-582.

Van den Berg, B. (2003). Extremophiles as a source for novel enzymes. **Curr. Opin. Microbiol.** 6: 213-218.

Ventosa, A., and Nieto, J. J. (1995). Biotechnological applications and potentialities of halophilic microorganisms. **J. Microbiol. Biotechnol.** 11: 85–94.

Ventosa, A., Nieto, J. J., and Oren, A. (1998). Biology of moderately halophilic aerobic bacteria. **Microbiol. Mol. Biol. Rev.** 62:(2) 504-544.

Whitaker, J. R. (1994). **Introduction: In Principles of Enzymology for the Food Sciences** (pp. 1-27). New York: Marcel Dekker, Inc.

Yongsawatdigul, J., Rodtong, S., and Raksakulthai, N. (2007). Acceleration of Thai fish sauce fermentation using proteinases and bacterial starter cultures. **J. Food Sci.** 72:(9) 382-390.

Zabolotskaya, M. V., Demidyuk, I. V., Akimkina, T. V., and Kostrov, S. V. (2004). A novel neutral protease from *Thermoactinomyces* species 27a: sequencing of the gene, purification, and characterization of the enzyme. **Protein J.** 23:(7) 483-492.

CHAPTER II

LITERATURE REVIEWS

2.1 Microbiology of fish sauce fermentation

2.1.1 Fish sauce fermentation

Fish sauces are generally prepared by mixing 2 to 6 parts of fresh anchovy (*Stolephorus spp.*), mackerel (*Ristrelliger spp.*), herring (*Clupea spp.*), or other small fish, depending on recipe and country of producers, with 1 part of solar salt (Wilaipan, 1990). In industrial scale, large containers, for example an underground concrete tank, are normally used for Thai fish sauce production. Traditionally, fish and salt are placed in layers of salt at the bottom, then fish and another layer of salt on top. During the first week, brine will reach to the top of the layers, which at this point, a piece of bamboo mat or other similar non-toxic objects are usually used to weight down fish flesh. In the loosely covered containers, natural fermentation is allowed to progress at ambient temperature (30-40°C) for 12 to 18 months (Lopetcharat and Park, 2002; Owens and Mendoza, 1985) where fish proteins will be liquefied or digested by the action of endogenous fish proteinases and microbial proteinases. After fermentation time attains, liquid fraction is siphoned out, filtered to remove residual solid fraction, and subsequently transferred into a ripening tank. The ripening step usually takes 2-4 weeks and then a clear brown liquid with salty taste and distinct aroma and flavor, called “fish sauce” is obtained (Wilaipan, 1990). A premium grade fish sauce typically contains >20 g/L total nitrogen (16 g/L from amino acids), 25-

28% of NaCl, 0.2-0.7% ammonium, and pH 5.1-5.8 (Park et al., 2001; Saisithi, Kasemsarn, Liston, and Dollar, 1966).

2.1.2 Halophilic bacteria in fish sauce

2.1.2.1 Definition of halophilic bacteria

In microorganisms, there is no common definition for "halophilic" or "salt-loving" to explain the requirement of salt for growth since there is a high variation for minimum and maximum of salt level that often depends on the medium composition and growth temperature. Kushner et al. (1978) and Johnson, Lanthier, and Gochnauer (1986) firstly classified halophiles based on the level of salt requirement for growth. Moderate halophiles are defined as organisms growing optimally between 3% and 15% salt (Kushner, 1978). Bacteria able to grow in the absence of salt as well as in the presence of relatively high salt concentrations are designed as halotolerant (or extremely halotolerant if growth extends above 15%). Bacteria that require at least 12% salt (optimal growth at 20%) are considered as extreme halophiles (Johnson et al., 1986). Recently, Oren (2008) proposed a simple definition for halophiles as microorganisms that grow optimally at 50 g/l of NaCl (0.85 M NaCl) or higher, and tolerate at least 100 g/l of NaCl (1.7 M NaCl).

Halophiles have been identified from all three kingdom including archaea, bacteria, and eukaryote. Based on phylogenetic tree analysis of rRNA sequences, halophlies and non-halophiles are often grouped together in the same phylogenetic clade regardless to their salt requirement and tolerance (Oren, 2008). This indicates that microbial classification of halophiles using rRNA sequences is not applicable. However, genome and proteome comparative analysis provide common characteristics within halophiles that transcend the phylogenetic relationship of

species (Paul, Bag, Das, Harvill, and Dutta, 2008). The details of these characteristics are further described in subheading 2.3.3.2 and 2.5.2.

2.1.2.2 Microbial halophilic adaptation

All halophilic microorganisms need to osmotically balance their cytoplasm with their surrounding medium. Fundamentally, there are two different strategies for this process. Firstly, the accumulation of potassium and soldium in cytoplasm at equal molar concentrations to medium is called "high-salt-in strategy" (Oren, 2008). Extremely halophilic archaea such as family *Halobacteriaceae* and species such as *Halobacterium salinarum* and *Haloarcula marismortui* is a popular model. Genome of *Halobacterium* sp. NRC-1 contains multiple transport systems, which are responsible for permeability of K^+ and Na^+ ions across the cell membrane. These include ATP-driven three subunit K^+ transport system (KdpABC), TrkAH, a low-affinity K^+ transporter driven by the membrane potential, and NhaC proteins, an active Na^+ efflux. These key protein machineries are used to maintain the ionic distribution of the cells (Ng et al., 2000). The intracellular proteins of microorganism using the "high-salt-in strategy" have been extensively adapted to maintain their conformation and activity at salt approaching saturation (Lanyi, 1974). Proteome of these organisms is highly acidic, and therefore most proteins denature when expose to low salt (see subheading 2.3.3.2 for more details). This explains inability of such organism to survive in low salt media.

The second strategy is based on the biosynthesis and/or accumulation of organic osmotic solutes, called "compatible-solute-in strategy" (Oren, 2008). Salt is exclude from cells as much as possible, while high concentration of organic "compatible" solutes does not interfere with the activity of cytoplasmic proteins.

Thus, only few adaptation of proteome is needed. This strategy has extensively been found in halophilic prokaryote and eukaryote. Most compatible solutes are uncharged or zwitterionic including amino acids or amino acid derivatives such as glutamine, glutamate, proline, ecotine, choline, betaine, and glycine betaine, as well as sugar and sugar alcohols (Galinski, 1995; Robert, 2005). The biosynthesis of glutamate in microorganisms can be accomplished by the action of either glutamate synthase (GOGAT) or glutamate dehydrogenase (GDH) encoded from gene *gltAB1B2* and *adh1h2*, respectively, while glutamine is synthesized by the action of glutamine synthase (GS) produced from gene *glnA1A2* (Suam and Müller, 2007).The use of proline as compatible solute is found essentially in *Firmicutes* (Kulhmann and Bremer, 2002; Suam and Müller, 2007), where the biosynthesis is regulated by an operon *proHJA* that encodes for proline biosynthesis enzymes including pyrroline-5-carboxylate reductase (ProH), glutamate 5- kinase (ProJ), and glutamate 5-semialdehyde dehydrogenase (ProA) (Belitsky, Brill, Bremer, and Sonenshein, 2001). In many bacteria, ecotine is used as a dominant compatible solute (Bursy, Pierik, Pica, and Bremer, 2007; Galinski, 1995). Ecotine is a cyclic amino acid derivative biosynthesized from aspartate semialdehyde and the preduction is controlled by operon *ectABC*. Choline and glycine betaine are used only when they can be taken up from the medium (Oren, 2008). Only few prokaryotes are capable of *de novo* synthesis of glycine betaine *e.g. Actonobacteria* and *Proteobacteria* where the biosynthesis pathway is unknown (Oren, 2008). Betaine can be synthesized when choline is available by the action of choline dehydrogenase and betaine-aldehyde dehydrogenase encoded from operon *betAB*. Most of bacilli contain high affinity substrate specific transport systems (OpuE) for osmoprotectant uptake. The OpuE

transporter is a single component classified in a member of the Na^+/solute symporter family (SSF) (Kempf and Bremer, 1998), *e.g.* symporters specified for glutamate and proline. Active transporters are also used for scavenging environmental glycine betaine and choline. These include ABC-type proline/glycine betaine transport system, choline-glycine betaine transporter, and choline/carnitine/betaine transporter. These transporters are believed to play an important role for halotolerant ability of *Oceanobacillus iheyensis* (Takami, Takaki, and Uchiyama, 2002). In other moderately halophilic bacteria, *Halobacillus halophilus* produces glutamine and glutamate as osmolytes in low salinity (1 M NaCl) and switches to proline as the dominant solute at higher NaCl (2 to 3 M NaCl) (Suam and Müller, 2007). The production of glutamine and glutamate is chloride ions dependent. At notable salinity, the accumulation of intracellular Cl^- will activate the transcription of *glnA2* and later increase glutamine synthase activity, and thus increase cellular glutamine and glutamate (glutamine can be converted to glutamate by the action of glutamate synthase controlled by *gltB*) (Suam and Müller, 2007). At higher salinity, higher molar concentration of glutamate is used as a "second messenger" to trigger proline production (Suam and Müller, 2007). This "compatible-solute-in strategy" provides microorganisms' ability and flexibility to adapt to a surprisingly broad salt concentration range, where the concentration of solutes gives osmotic equilibrium with the external salt concentration without destructive effects to intracellular proteins (Galinski, 1995; Ventosa, Nieto, and Oren, 1998). However, the "high-salt-in strategy' is energetically less costly to the cell than the biosynthesis of large amounts of organic osmotic solutes, which energy is still needed to prevent salts from reaching the cytoplasm (Oren, 1999). Some prokaryotic kingdoms show a pronounced adaptation to the strategy of archaeal

"high-salt-in" rather than the accumulation of organic solutes, including *Halanaerobiales* (from *Firmicutes*) (Oren, 1986, 2006) and *Salinibacter ruber* (from *Bacteroidetes*) (Mongodin et al., 2005). In contrast, *Halobacillus halophilus* accumulates high molar concentrations of solute as well as a high concentration of intracellular chloride. This suggested that the bacterium may employ an intermediate strategy of salt-in-cytoplasm and the use of compatible solutes, so called "hybrid strategy" (salt-in and compatible-solute-in) (Suam and Müller, 2008).This counterproductive dual strategy is used to explain osmoprotection ability of *Halobacillus halophilus* and assumed to be common to all moderately halophilic organisms. On the other hand, *Halomonas elongate* has optimized its metabolism to minimize the energetic cost of osmotic adaptation (Maskow and Babel, 2001). This led to an assumption that at high salinity the metabolic types of prokaryotes may balance between the amount of energy available to the cells and the cost of production of organic solutes needed to provide osmotic balance (Oren, 2008). In deeper details, those genes responsible for halophilic ability of microorganisms are possibly related to a group of metabolic regulation genes. In *Bacillus* species, genes involving in central metabolic pathways are concurrently regulated by global transcriptional regulators including CcpA, CodY, and TnrA, which determine the rate of expression of the central metabolic genes and the flux of metabolites (Sonenshein et al., 2007). The CcpA and CodY are global regulators of carbon and nitrogen metabolism pathways whereas ThrA is a global regulator that mediates positive or negative expression of genes involving in nitrogen compounds metabolism (Fujita, 2009; Molle et a.l, 2003; Wray, Ferson, Rohrer, and Fisher, 1996). To date, however, evidence of the cooperation between genes involving in halophilic ability and these

global metabolic regulators has never been reported.

2.1.2.3 History of halophilic bacteria isolation in fish sauce

The extremely high concentration of NaCl in fish sauce can inhibit the growth of non-halophiles including spoilage microorganisms and most pathogens, however contributes to the propagation of halophilic bacteria. In fish sauce related research, most studies concentrated on isolation and identification of the microorganisms present in various stages of fermentation. Non-halotophilic proteolytic bacteria, such as *Psudomonas* sp. and *Bacillus* species, were dominant during the first month of fermentation, but were not detected after one month (Chaiyanan, 2000). In the first two months, halotolerant *Bacillus* species were found to be predominant and then a sharp increase of *Stapylococcus*, coryeform bacteria, and *Micrococcus* sp. became common in the third months (Chaiyanan, 2000). Nawong (2006) screened proteinase producing bacterial isolates from 1 to 12 month-old fish sauce. These isolates were identified as *Virgibacillus*, *Halomonas*, *Bacillus firmus*, *Brevibacteium*, *Corynebacterium*, and *Staphylococcus*, which were dominate in fish sauce.

Other microorganisms found during fish sauce fermentation included moderately halophilic bacteria, such as *Lentibacillus salicampi*, *L. jurispiscarius* (Thongsanit, Tanasupawat, Keeratipibul, and Jatikavanich, 2002), *L. halophilus* (Tanasupawat, Pakdeeto, Namwong, Thawai, Kudo, and Itoh , 2006), *Filobacillus* sp. RF2-5 (Hiraga et al., 2005), *Halobacillus* sp. SR5-3 (Namwong, Hiraga, Takada, Tsunemi, Tanasupawat, and Oda, 2006), *Halobacillus thailandensis* (Chaiyanan, Maugel, Hug, Robb, and Colwall, 1999), *Tetragenococcus halophilus* and *T. muriaticus* (Tongsanit et al., 2002) and extremely halophilic archaea such as *Halobacterium salinarum*, *Halococcus thailandensis* (Thongthai, McGenity,

Suntinanalert, and Grant, 1992; Namwong, Tanasupawat, Visessanguan, Kudo, and Itoh, 2007), *Chromohalobacter salexigens* JCM 8878, *Halobacterium saccharolyticus* (Tanasupawat, Namwong, Kudo, and Itoh, 2009) and *Natrinema gari* sp. nov (Tapinkae et al., 2008). The ultimate goal of bacterial isolation in fish sauce is to develop starter cultures of the appropriate microorganisms.

Yongsawatdigul, Rodtong, and Raksakulthai (2007) successfully applied three proteinase producing strains of moderately halophilic bacteria, including *Virgibacillus* sp. SK33, *Virgibacillus* sp. SK37, and *Staphylococcus* sp. SK1-1-5 as starter cultures in combination with commercial proteolytic enzymes for fish sauce fermentation in 4 months. In archaea, Akolkar, Duri, and Desai (2010) proposed a proteinase-producing halophile, *Halobacterium* sp. SP1(1) isolated from salt pan situated in Kandla, India for acceleration of fish sauce fermentation. This strain provided higher protein and nitrogen contains as well as better flavor and aroma aspects in final product compared to the control containing natural microflora of *Halobacterium* sp. F1 and F2.

2.2 Description of *Virgibacillus*

2.2.1 Morphology and physiology

The genus *Virgibacillus* was firstly reclassified from *Bacillus pantothenticus* and later combined with genera *Salibacillus* (Heyndrickx, Lebbe, Kersters, De Vos, Forsyth, and Logan, 1998; Heyrman et al., 2003). To date, over 195 of bacterial strains in this genus have been proposed to the Genbank database, where only 18 strains were validly described species. Out of 18 species, 9 of which were isolated from saline environments such as saline lakes, saltern ponds, desert, hypersaline soils, and salted food. Table 2.1 shows some physiological characteristics of *Virgibacillus*

species. *Virgibacillus* are Gram-positive rods (0.3-0.7 x 2.0-0.6 μm), and single, pair or long chain of cells arrangement. Their colonies are small, circular, low convex, and slightly transparent-to-opaque. Cell wall peptidoglycan of *Virgibacillus* contains *meso*-diaminopimelic acid. Major fatty acids in cell membrane are mostly anteiso-$C_{15:0}$ followed by iso-$C_{15:0}$. The G+C content is 36-43 mol %, compared to 32-69 mol % in *Bacillus* species (Heyrman et al., 2003). They are aerobic, catalase-positive, motile, and spore forming. *Virgibacillus* grow at pH 6.0-10.0 (optimum at 7.5-8.0), and 10-55°C (optimum at 30-40°C). Members of this genus are moderately halophilic bacteria able to grow at 0-25% (w/v) NaCl with optimum of 5-10% NaCl (Table 2.1).

2.2.2 Proteinases from *Virgibacillus*

Virgibacillus show ability to produce a variety of extracellular hydrolytic enzymes. A high production of proteinases from bacteria in this genus has been reported. Rohban, Amoozegar, and Ventosa (2009) isolated a number of moderately and extremely halophilic bacteria from a hypersaline lake in Iran. Of which, some isolates were identified as *Virgibacillus* based on 16S rRNA sequence analysis. These strains were able to produce a great variety of hydrolytic enzyme such as amylase, proteinase, cellulase, and inulinase. Chamroensaksri, Akaracharanya, Visessanguan, and Tanasupawat (2008) reported a halophilic proteolytic bacterium, *Virgibacillus marismortui* NB2-1 isolated from Pla-ra, fermented fish in Thailand. This bacterium secreted five proteinases with molecular masses (MMs) ranging from 17 to 35 kDa. The enzymes were alkaline serine proteinases and showed optimum catalytic activity at pH 10, 50°C and 5% NaCl. Gupta, Joseph, Mani, and Thomas (2008) reported a newly identified serine proteinase from *Virgibacillus pantotheticus* MTCC 6729 from chicken meat samples. The enzyme was alkaline and thermo-stable and compatible

Table 2.1 Physiological characteristics of bacteria in genus *Virgibacillus*

No.	Species	Growth range (Optimum)			%GC	Pigment	habitats	References
		pH	Temp (°C)	NaCl (%)				
1	*Virgibacillus arcitcus*	nd	0-30 (25)	0-20 (5)	38.2	-	Permafrost core from Canadian high Arctic	Niederberger et al., 2009
2	*Virgibacillus carmonnensis*	nd	10-40 (25-30)	5-10	38.9	Pink	Damaged mural, Carmana, Spain	Heyrman et al., 2003
3	*Virgibacillus chiguensis*	nd	15-55 (40)	5-10	37.3	-	Saltern, Taiwan	Wang et al., 2008
4	*Virgibacillus dokdonensis*	nd	15-50 (37)	4-5	36.7	-	Korean island	Yoon et al., 2005
5	*Virgibacillus halodenitrificans*	5.8-9.6 (7.4-7.5)	15-45 (38)	2-23 (3-7)	38-39	-	Solar saltern	Yoon et al., 2004
6	*Virgibacillus halophilus*	5-10	5-45	0-18	42.6	Yellow	Soil, Japan	An et al., 2007
7	*Virgibacillus kekensis*	6-10 (7)	10-50 (37)	0-25 (10)	41.8	-	Salt lake, China	Chen et al., 2008
8	*Virgibacillus koreensis*	5.5-9.0 (7)	10-45 (25)	0-20 (5-10)	41	-	Salt field, Korea	Lee et al., 2006
9	*Virgibacillus litoralis*	6.0-10.0 (8)	10-45 (30)	2-25 (5-10)	40.2	-	Saline soil, China	Chen et al., 2009
10	*Virgibacillus marismortui*	nd	15-50	5-25 (10)	39.0-42.8	-	Dead sea	Arahal et al., 1999
11	*Virgibacillus necropolis*	nd	10-40 (25-35)	5-10	37.3	-	Damaged mural, Carmana, Spain	Heyrman et al., 2004
12	*Virgibacillus olivae*	4-8	20-45	0-20 (5)	33.4	Yellow	Waste water	Quesada et al., 2007
13	*Virgibacillus pantothenticus*	nd	15-50 (37)	4-10	36.9-38.3	-	Soil	Heyndrickx et al., 1999
14	*Virgibacillus proomii*	nd	15-50 (37)	nd	36.8-37.0	-	Soli	Heyndrickx et al., 1999
15	*Virgibacillus salarius*	nd	10-50 (30-35)	0.5-25 (7-10)	37.3	-	Salt lake, Sahara	Hua et al., 2008
16	*Virgibacillus salexigens*	6-11 (7.5)	15-45 (37)	7-20 (10)	36.3-39.5	-	Solar saltern	Garabito et al., 1997
17	*Virgibacillus siammensis*	7-8 (7)	15-40 (37)	1-20 (5)	38	Red	Thai fermented fish	Tanasupawat et al., 2010
18	*Virgibacillus zhanjiangensis*	6.0-10.0 (7.5)	10-45 (30)	1-15 (4-7)	39.5	Cream	Sea water	Peng et al., 2009

Note: nd = no data available, "-" = not observed.

with commercial detergents showing potentials as a detergent additive.

Sinsuwan, Rodtong, and Yongsawatdigul (2007, 2008a, 2008b, 2010a, 2010b) characterized proteinases of *Virgibacillus* sp. SK33 and SK37 isolated from one-month-old Thai fish sauce. In SK33, The crude extracellular enzyme showed a subtilisin-like alkaline serine proteinase characteristic exhibiting optimum catalytic condition at 50°C and pH 8-11 with MM of 19-56 kDa. An enzyme at 19 and 32 kDa were purified and showed NaCl and $CaCl_2$-activated activities. In addition, the 32 kDa enzyme exhibited a high stability towards various organic solvents at concentrations up to 25% (v/v), suggesting potential uses in high saline and aqueous-organic solvent systems.

In SK37, crude extracellular and cell-bound proteinases were subtilisin-like serine proteinases and showed maximal activity at 65°C and pH 7-9.5. All enzymes were halotolerant showing NaCl-activated characteristic and stable up to 25% NaCl. $CaCl_2$ could activate enzyme activities in both extracellular and cell-bound fractions. These proteinases from both strains of SK33 and SK37 and the bacteria are promising enzymes and strains, respectively, to be developed for use in fish sauce fermentation.

2.3 Characteristics of proteinases produced from halophilic bacteria

2.3.1 Definition of proteinases

Proteinases are one of the largest commercial enzymes accounting about 60% of the total enzyme market worldwide (Rao, Tanksale, Ghatge, and Deshpande, 1998). They are used in a number of industrial applications including detergent formulation, peptide synthesis, protein processing, feed, leather treatment, dairy, chill proofing, meat tenderization, fermented sauces, and pharmaceutical products (Shimogaki et al.,

1991; Rao et al., 1998). Proteinases are defined as an enzyme that hydrolyzes peptide bonds. In general, they are divided into two main groups consisting of endopeptidases (proteinases), which cleave internal peptide bonds of proteins and exopeptidases, which cleave terminal peptide bonds either at N-terminus or C-terminus. Based on the nature of the active site, proteinases can be classified as aspartic proteinases (*e.g.* cathepsins D and E, pepsin, and renin), cysteine proteinases (*e.g.* cathepsins B, L, S, K, Q, papain, bromelain, and ficin), metalloproteinases (*e.g.* gelatinases A and B, collagenase, termoase, neutrase, and thermolysin), serine proteinases (*e.g.* plasmin, trypsin, chymotrypsin, and subtillisin), and threonine proteinases (*e.g.* proteasome) (Beynon and Bond, 2001).

2.3.2 Subtilisin-like proteinases

2.3.2.1 Active site, activity, and specificity

Subtilisin-like proteinases or known as subtilases, are members of family S8 which is the second largest family of serine peptidases (the largest family is chymotrypsin family S1) (Rawlings, Morton, Kok, Kong, and Barrett, 2008). Enzyme active site of family S8 contains Asp, His, and Ser as a catalytic triad. The common motifs of the active site residues frequently are Asp-Thr/Ser-Gly, His-Gly-Thr-His, and Gly-Thr-Ser-Met-Ala-Xaa-Pro, while Asp-Asp-Gly, His-Gly-Thr-Arg and Gly-Thr-Ser-Ala/Val-Ala/Ser-Pro are particularly found in kexin, a member of subtilases (Rawlings et al., 2008). Even though the location of Asp_{32}, His_{64}, and Ser_{221} in subtilases are sequentially far apart (Seizen and Luanissen, 1997), they converge in the 3D structure to form the catalytic triad and function as charge-relay network. This characteristic is explained by, at physiological pH, the negatively charged carboxylate side-chain of Asp_{32} forms hydrogen-bond to a nitrogen-bonded proton on His_{64}

imidazole ring. Other nitrogen on His_{64} forms another hydrogen-bond to the O-H proton of Ser_{221} creating partially negative oxygen atom. This results the oxygen atom of Ser to be more nucleophilic and thus is susceptible to attack incoming substrates (*i.e.* peptide bonds), which is also assisted by a neighboring carboxyamide side-chain of Asn_{155}. Most of subtilases are endopeptidases, while only some of them are exopeptidases *e.g.* TPP-II, releasing peptides from N-terminus of protein substrates. Subtilases are mostly active at neutral to alkali pH, while a number of enzymes are thermostable. Most members of subtilases are nonspecific peptidase. In general, the specificity of subtilases largely determined by the enzyme's S_1-S_4 binding sites, which interact with the P_1-P_4 residue side chains of peptide substrates, while the S_1 and S_4 dominate the substrate preference of the enzyme (Gron, Meldal, and Breddam, 1992). The S_1 and S_4 binding sites are large and hydrophobic, which explains the broad specificity of subtilases with a preference for aromatic or large nonpolar P_1 and P_4 substrate residues (Seizen and Luanissen, 1997; Gron et al., 1992). Kexin and furin is an exception, which preferably cleave residues after dibasic amino acids (Lys-Arg or Arg-Arg) or multiple basic residues (Rawlings et al., 2008). This is because acidic residues are found in the S_1 and S_2 sites of the enzymes, and leads to a specificity for dibasic residues (Ballinger, Tom, and Wells, 1995). Bovine casein and Suc-Ala-Ala-Pro-Phe-pNA are often used as a protein and synthetic substrate, respectively. Most of subtilases are inhibited by serine peptidase inhibitors such as phenylmethane sulfonylfluoride (PMSF) and diisopropyl fluorophosphates (DFP). Many members of subtilases bind calcium ions for activity and stability, thus inhibition from metal chelating agents such as ethylenediaminetetraacetic acid (EDTA) and ethyleneglycoltetraacetic acid (EGTA) is detected and often misidentified as a

metallopeptidase. Other inhibitors for chymotrypsin such as turkey ovomucoid third domain, eglin C, and barley inhibitor CI-1A also inhibit activity of subtilases (Rawlings et al., 2008).

2.3.2.2 Molecular structure

Even though the catalytic triad of subtilases is likely related to chymotrypsin, subtilases are clearly unrelated to chymotrypsin in term of structural features. A typical tertiary structure of subtilases consists of a seven-stranded β-sheet sandwiched between two layers of helices (Seizen and Luanissen, 1997; Rawlings et al., 2008). Most members of subtilases are not multidomain proteins. However, several variants contain large insertions and deletions sequences. Some large insertions can be up to 65 residues as found in pyrolysin with weak homology and often unique in single enzyme, whereas sequence conservation in large inserts is frequently observed in thermitase (Rawlings et al., 2008). Most deletions are reported in lantibiotic leader peptidases which include large N- and C-terminal deletions. Besides, large N-terminals extensions with no sequence homology are particularly found in kexin and their function is unknown. However, some usual insertions may play a special role in an extremoenzyme. The insertion sequences of a thermophilic subtilase of *Bacillus* sp. WF146 have been proven to be responsible for thermostability (Wu et al., 2004). Subtilases contain multiple calcium-binding sites for stability. Bond calcium ions tie together high mobility loops on protein's surface which are highly susceptible for autoproteolysis. Thus, the enhanced rigid structure upon calcium binding reflects higher thermostability and resistance to autolysis. Calcium binding sites in these enzymes could be classified into four calcium sites based on their dissociation constants (K_d). The geometry of the calcium ligand comprises of a coordination of

peptide carbonyl oxygen from hydrophobic or nucleophilic residues and bidentate carboxylate from aspartic or gultamic acid. In thermitase, calcium site 1 (Ca1) comprises one carboxyl oxygen ligand of threonine (amino acid no. 67) and three aspartic acids (no. 57, 60, and 62). It is an extremely strength calcium site showing K_d at $< 10^{-10}$ M (Gros, Betzel, Dauter, Wilson, and Hol, 1989; Gros, Kalk, and Hol, 1991; Teplyakov et al., 1990). Ca2 and Ca3 are moderate and weak calcium sites exhibiting K_d at 10^{-10} and 10^{-4} M, respectively. These sites are present in thermitase, subtilisin BPN′, and subtilisin Carlsberg, while only Ca3 is found in proteinase K with a distinct K_d of 10^{-7} M. In BPN′, the Ca2 site is coordinated by a carbonyl oxygen loop (no. 75-84), one aspartic acid (no. 41) and a bidentate carboxylate from N-terminal glutamic acid (no. 2) (Bott, Ultsch, Kossiakoff, Graycar, Katz, and Power, 1988). The Ca3 in BPN′ comprises carbonyl oxygen at amino acid no. 169-174, two carboxylates from glutamic acid (no. 195) and aspartic acid (no. 197). This corresponds to carbonyl oxygen at amino acid no. 173-178 and one carboxylate ligand from aspartic acid (no. 200) in thermitase (Gros et al., 1989, 1991; Teplyakov et al., 1990). The Ca4 is a very weak calcium binding site ($K_d < 10^{-4}$ M) and found only in proteinase K (Betzel, Pal, and Saenger, 1988). It contains carbonyl oxygen from threonine (no. 16) and aspartic acid (no. 240). Disulfide bridges can enhance the thermal stability of a protein. In subtilases, seven commonly naturally occurring disulfide bonds have been reported including Cys no. 27-118 and 175-247 in proteinase K (Betzel et al., 1988), 61-98 and 163-195 in aqualysin (Kwon, Terada, Matsuzawa, and Ohta, 1988), 53-100 and 171-131 in *Dichelobacter* basic proteinase (Lilley, Stewart, and Kortt, 1992) and 47- 59 in *Bacillus* subtilisin S41 (Davail, Feller, Narinx, and Garday, 1994). Exceptionally, extracellular subtilases from gram-positive bacteria rarely contain disulfides since Cys

residues are very rare in the primary structure of these enzymes (Seizen and Luanissan, 1997).

2.3.2.3 Subtilase classification

Multiple amino acid sequence alignment of hundreds of subtilases shows that all members have "structurally conserved region" at their catalytic domains (Seizen and Luanissen, 1997). Based on amino acid sequence similarity, subtilases are classified into six families, including:

Subtilisin family (S8A) Members in this family are exclusively found in microorganisms mainly from *Bacillus* species. This is the largest family and can be classified into six subfamilies, which are described in-depth below.

True subtilisins Two original members which have long been identified and characterized are subtilisin Carlsberg from *Bacillus licheniformis* and subtilisin Novo or bacterial proteinase Nagase (BPN′) from *Bacillus amyloliquefaciens* (Rao et al., 1998). Subtilisins are monomeric proteins of 26.9-27.5 kDa with basidic isoelectric point (p*I*) of 7.8-11 and exhibit catalytic activity at pH 8.2-8.6, stability at pH 7-10, and optimum temperature around 60°C (Betzel, Klupsch, Papendorf, Hastrup, Branner, and Wilson, 1992; Graycer, Ballinger, and Wells, 2004). A number of enzymes in this subfamily are highly commercial valuable mainly used in detergents *e.g.* subtilisin Carlsberg (Rao et al., 1998).

High alkaline proteinases The members in this subfamily are inhibited by DFP but not by tosyl-L-phenylalanine chloromethyl ketone (TPCK) or tosyl-lysyl chloromethyl ketone (TLCK). The optimal pH is around pH 10, and p*I* is around pH 9. Their MMs are in the range of 15-30 kDa (Rao et al., 1998). Examples of members in this subfamily are esperase from *B. lentus*, ALP1 from *Bacillus* sp. NKS-21, PB92

from *B. alcalophilus* PB82, elastase YaB from *B. subtilis* YaB, subtilisin Sendai from *Bacillus* sp. G-825-6, and M-proteinase from *Bacillus* sp. KSM-K16 (Seizen and Luanissen, 1997; Kobayashi et al., 1995).

Intracellular proteinases Based on amino acid sequence comparison, intracellular serine proteinases (Isps) and extracellular subtilisins in *B. subtilis* putatively share the same ancestral protein as the result of ancestral gene duplication (Strongin, Izotova, Abramov, and Stepanov, 1978). Isp-I from *B. subtilis* has MM of 30 kDa and acidic p*I* at 4.3 and shows low activity against protein substrates. A higher composition of Gln, Glu, Lys, and Phe residues than extracellular subtilisins was observed (Stepanov et al., 1977). The requirement of calcium ions for enzyme activity is varied, for example, Isp-I from *B. megateruim* was absolutely calcium-dependent while its homologous enzyme from *B. subtilis* was not (Kucerova, Hlavacek, Vachova, Mlichova, and Chaloupka, 2001). Other examples of this subfamily are Isp-Q from *Bacillus* sp. NKS21, Isp enzymes from *Bacillus* sp. 221, and *B. polymyxa*, and a non *Bacillus* enzyme, tiap from *Thermoactinomyces* sp. (Seizen and Luanissen, 1997).

High molecular mass subtilisins (HMSs) Members in this subfamily are produced from alkaliphilic *Bacillus* species. The enzymes show open reading frames of 2.3-2.4 kbp which encode 798-808 amino acids with a high sequence identity of 60-95% to each other (Ogawa et al., 2003; Okuda et al., 2004). Purified enzymes have MM at 72 kDa and acidic p*I* at 4.2-4.7. Optimum catalytic activity was reported at pH 10.5-11, 40-45°C. The enzymes are thermolabile but can be enhanced by calcium. Aliphatic amino acid Leu was found as a preference at P_1 site of substrates (Ogawa et al., 2003). Examples of HMSs are SZ proteinase from *Bacillus*

sp. strain KSMP9860, SF proteinase from *Bacillus* sp. strain KSM-9865, NV proteinase from *Bacillus* sp. strain NCIB12289, HK proteinase from *Bacillus* sp. strain D-6, YK proteinase from *Bacillus* sp. strain Y, SD proteinase from *Bacillus* sp. strain SD521 (Okuda et al., 2004), FT proteinase from *Bacillus* sp. strain KSM-KP43 (Ogawa et al., 2003), Vpr from *B.* subtilis (Sloma, Rufo, Theriault, Dwyer, Wilson, and Pero 1991), and Bha from *B.* halodurans C-125 (Takami et al., 2000).

Phylogenetically intermediate subtilisins (PISs) A small subfamily which currently comprises of two members, including LD1 from alkaliphilic *Bacillus* sp. KSM-LD1 (Saeki et al., 2003) and ALTP from extremely alkaliphilic *Alkaliphilus transvaalensis* (Kobayashi et al., 2007) with a high sequence identity of 64%. The enzymes have MM of 30 kDa and optimum activity and stability at pH 10-12 and 60-70°C. Calcium has no effect on enzyme thermal stability. The LD1 showed moderate resistance to oxidation by H_2O_2 and strong resistance to detergents, which is promising for detergent formulation (Saeki et al., 2003).

Oxidant stable proteinases (OSPs) This subfamily was firstly proposed by Saeki et al. (2000) with an attractive property of high stability and oxidative resistance to oxidizing agents. Members of the OSPs have shown a wide application in detergent and bleaching industries. Most OSPs are produced from alkaliphilic *Bacillus* species including E-1, LP-Ya, and SD-521 from *B. cohnii* DSM 6307, NP-1 from *B. halmapalus* DSM8723, and KP-43 from *Bacillus* sp. KSM-KP43 (Saeki et al., 2000, 2002). The enzymes show optimum condition at pH 10.5 and 65°C and are calcium-independent since EDTA and EGTA has no inhibitory effect. The MM and p*I* of OSPs are at 43 kDa and basic p*I* of 8.9-9.1 (Saeki, Ozaki, Kobayashi, and Ito, 2007).All members have a great sequence identity of 88% and contain unique of C-

terminal extension of ~160 amino acids, which showed similarity to the integral segment of a membrane-associated subtilisin (Beja et al., 2000; Saeki et al., 2007).

Thermitase family (S8B) This is a thermophilic proteinase family found only in microorganisms, including thermophiles such as *Thermoactinomyces* and thermophilic *Streptomyces* and some halophiles. The enzymes bind three calcium ions for their thermal stability. One of them is an extremely strong bound-Ca^{2+} which cannot be removed without protein denaturation (Briedigkeit and Frommel, 1989).

Proteinase K family (S8C) Proteinase K was the first member of this family which was produced from *Engyodontium album*. However, this is a large family of extracellular endopeptidases found only in fungi, yeasts, and gram-negative bacteria. This family has a high sequence similarity (>37% identity, and >55% sequence identity among bacterial origin) (Seizen and Luanissen, 1997). The enzymes preferably cleave after the carboxyl group of aliphatic and aromatic amino acids of peptides. Only a few enzymes have a large C-terminal extension, while minor insertions and deletions are common.

Lantibiotic peptidase family (S8D) This is a small family found only in Gram-positive bacteria. They are intracellular peptidases with a specific role of cleaving leader peptides from precursors of lantibiotics, which is a group of antimicrobial peptides (Sahl, Jack, and Bierbaum, 1995). A member of this family, llnisp, contains a C-terminal extension for membrane anchor. This family has numerous insertions/deletions which result in low sequence similarity with each other and other subtilases (Seizen and Luanissen, 1997).

Kexin family (S8E) This composes a large group of kexin, furin, and proprotein convertases (PCs), which involve in activation of peptide hormones,

growth factors, viral proteins, etc. (Barr, 1991; van de Ven et al., 1993). Almost members are from eukaryote with high sequence homology (>40% identity) (Seizen and Luanissen, 1997). Some members including endopeptidase R, T, and K from yeast *Tritirachium* and cuticle-degrading peptidase from *Metarhizium,* require the activation of thiol group for activity since the presence of Cys near active site His (Seizen and Luanissen, 1997).

Pyrolysin family (S8F) Members of this family are from varied origin with low sequence homology (most <37% identity) (Seizen and Luanissen, 1997). However, they usually contain homologous large insertions and/or long C-terminal extensions. Most members are N-terminal tripeptidyl peptidases releasing tripeptides from peptide substrates.

2.3.2.4 Biological functions and applications

Subtilases are widely distributed in various organisms, including bacteria, archaea, eucaryotes, and viruses (Siezen, de Vos, Leunissen, and Dijkstra, 1991). Most subtilases are secreted while some of which are intracellular or attached at cell wall *e.g.* lantibiotic leader peptidases. Some members are implicated to phatogenesis, for example, a peptidase from *Dichelobacter* in ovine foot rot (Kortt and Stewart, 2004) and C5a peptidase from pathogenic streptococci in destroying the complement chemotaxin C5a (Cleary and Matsuka, 2004). In mammals, Site-1 proteinase catalyzes the first step in the proteolytic activation of sterol regulatory element-binding protein (Sakai et al., 1998), cytoplasmic TPP-II has a role in the processing of antigenic peptides and the degradation of cholecystokinin (Warburton and Bernardini, 2002), and furin located in the mammalian *trans*-Golgi network and endosome membranes controls a number of proproteins. In eukaryote, kexin from yeast is responsible for

alpha-mating factor and killer toxin precursors. In bacteria, Gram-positive bacilli usually produce a large amount of subtilases. In *B. subtilis*, major extracellular proteinases are alkaline serine proteinases encoded from genes *aprE*, *bpr*, *epr*, *vpr*, and *wprA* (Kunst et al., 1997), while major intracellular proteinases *e.g.* ISP-1 and ISP-4 are expressed from gene *isp-1*and *isp-4*, respectively (Koide, Nakamura, Uozumi, and Beppu, 1986; Sheehan and Switzer, 1990, 1991). All of these enzymes are members of subtilisin family S8A. These enzymes are associated with nitrogen source scavenging and sporulation (Burnett, Shankweiler, and Hageman, 1986; Lee et al., 2004; Piggot and Coote, 1976; Wong, Price, Goldfarb, and Doi, 1984). Subtilisins from bacilli have long been used in various industrial applications, particularly in detergent formulation. They are dominantly from family S8A including subtilisin Carlsberg from *B. licheniformis* and M-proteinase from alkaliphilic *B. clausii* KSM-K16 (Rao et al., 1998; Saeki et al., 2007). The M-proteinase exhibited MM of 28 kDa and showed catalytic activity at pH 12.3 and 55°C and was highly stable to various surfactants (Kobayashi et al., 1995). The enzymes in OSP family have shown to have wide application in detergent and bleaching industries and recently become one of the most attractive enzymes. For example, the KP-43 from *Bacillus* sp. strain KSM-KP43 has a MM of 43 kDa and its optimum activity was observed at pH 11-12 and 65°C in the presence of calcium (Saeki et al., 2007). The enzyme was not inhibited by EDTA and EGTA and highly resistant to chemical oxidants. Crystal structures of M-proteinase and KP43 showed general α/β barrel subtilisin-like fold. However, a unique salt bright Arg_{19}-Glu_{217}-Arg_{275} (BPN′ numbering) of M-proteinase, which was not found in other subtilisins S8A, may contribute to higher thermo-stability in alkali pH than other subtilisins (Saeki et al., 2007). It has been well known that general

subtilases are inactivated by hydrogen peroxide due to the oxidation of methionine (Met) locating next to the catalytic serine (Met_{222} in subtilisin BPN′ numbering), which prevents the formation of tetrahedral intermediate during transition state of hydrolytic reaction (Bott et al., 1988; Stauffer and Etson, 1969) . However, the detailed mechanism of oxidant stability is still unclear since this Met residue is also conserved in all OSPs. The sulfoxide of this Met_{256} (KP43 numbering) (oxidized form of Met) in the vicinity of catalytic Ser was detected in the crystal structure of oxidized KP43 (Nonaka, Fujihashi, Kita, Saeki, Ito, and Miki, 2001). A possible mechanism is due to longer distance between Met_{256} and oxyanion hole and the presence of nearby $Met_{251,}$ resulting in lower oxidation rate of the Met_{256} (Saeki et al., 2007). Curiously, OSPs contains many insertions with low sequence similarity to other subtilisins. In case of KP43, a large C-terminal extension of ~160 amino acid residues exhibiting jelly roll β-barrel structure was detected which is absent in other subtilisins and does not share similarity to any sequences reported to date (Nonaka et al., 2001). This insertion may contribute the overall structural integrity and reflects stability of the enzymes in the harsh condition of detergents.

2.3.3 Halotolerant and halophilic proteinases

2.3.3.1 General biochemical characteristics

Proteinases produced by extremely halophilic bacteria usually perform their function *in vivo* and *in vitro* at 3–4.5 M NaCl, and lose activity rapidly when expose to low salt concentrations (Lanyi, 1974). Some haloneutrophilic bacteria known to produce extracellular halophilic proteinases have been isolated, characterized and showed maximum activity at neutral pH (Izotova et al., 1983; Kamekura and Seno, 1990; Norberg and von Hofsten, 1969; Reddy, Jayalaksmi, and Sreeramulu, 2003;

Schmitt, Rdest, and Goebel, 1990; Stepanov et al., 1992). Haloalkaliphilic bacteria require not only high salt concentration but also high pH (10–11) for growth (Tindall, Ross, and Grant, 1984). Their extracellular enzymes are active at salt concentrations in the range of 0.2–5.2 M (Gimenez, Studdert, Sanchez, and De Castro, 2000; Studdert, De Castro, Seitz, and Sanchez, 1997; Ventosa and Nieto, 1995; Yu, 1991). In contrast, halotolerant proteinases produced from moderately halophilic or halotolerant bacteria can function over a wide range of salinity. Unlike halophilic proteinases from extremely halophilic bacteria which need at least 12.5% (w/v) NaCl for catalytic activity (Ventosa et al., 1998), the halotolerant proteinases could retain their activity in either absence or presence of NaCl. At nearly saturated NaCl concentration, the enzyme activity can be either activated, or partially inhibited, or not affected by NaCl.

2.3.3.2 Mechanism of halophilicity of proteins

Halophilic archaea usually regulate cellular osmotic pressure by accumulating intracellular K^+ and adapting their intracellular and extracellular proteins to be active in 4 M KCl or <5 M NaCl (Danson and Hough, 1997). Whereas, halophilic eubacteria and eukaryote accumulate intracellularly compatible organic solutes, such as sugars, amino acids, and ectoines providing most of the osmotic balance (Ventosa and Nieto, 1995), therefore, their intracellular enzymes lack of halophilic property. The key structural features of halophilic proteins include: (i) a high content (up to 20% of total residues) of acidic residues, such as aspartic and glutamic acid on the protein's surface. This reduces surface tension at the protein surface by interacting with hydrate counterions reflecting the prevention of protein precipitation (Lanyi, 1974); (ii) extensive ion-pair networks by increasing acidic residue content; (iii) low proportion

of large hydrophobic amino acids (especially tyrosine, tryptophan, and phenylalanine), lower lysine content (often replaced by arginine), the increase in small hydrophobic residues (including glycine, alanine, and valine), and a decrease in aliphatic amino acids (Mardern, Pfister, and Zaccai, 1995). These help prevent hydrophobic interaction between the enzymes as a result of salting-out effect of salt (Bohm and Jaenicke, 1994; Sellek and Chaudhuri, 1999). Protein-solvent interaction is also important for haloadaptation. Most of the negative charges on the protein surface are surrounded by water molecules. Crystal structure of ferredoxin, a halophilic protein from *Haloarcula marsimortui* showed an average of 2 to 6 of water molecule around carboxylate of its surface acidic amino acid residues, compared to the average of 1.9 per residue of bulk amino acids (Frolow, Harel, Sussman, Mevarech, and Shoham, 1996). These bound-surface water molecules provide extensive hydrogen bonding. Besides, protein conformation of halophiles shows tendency of lower helix and higher coil structure (Paul et al., 2008). However, halophilic proteins need at least 1 M NaCl/KCl for proper folding and activity (Ventosa et al., 1998). This is mainly because charge repulsion of acidic residues clustered on the protein surface in a low salt condition (Madern et al., 2000). This is a highly distinguishable feature defined for halophilic proteins, which is clearly different from haloterant proteins. Besides a high stability at high salt concentrations, the hatolerant proteins are still active and stable in the absence of salt (Inouye, Kuzuya, and Tonomura, 1998). The salt-tolerant characteristic of proteins from moderately halophilic bacteria is discriminated into three categories: (i) intracellular enzymes, which are expected to present in low salt intracellular environment; (ii) membrane-bound enzymes where they are exposed to extracellular medium; and (iii)

extracellular enzymes which are needed to be active and stable in high salinity (Ventosa et al., 1998). Even though osmotic solutes generally provide osmotic balance between inside and outside the cells of moderately halophilic bacteria, however, the apparent NaCl concentrations can be often fairly high showing at 0.13 to 3.52 M with an average of 1.37 M, depending on the degree of salinity in the growth media (Ventosa et al., 1998). Therefore, halophilic characteristics of intracellular and membrane-bound proteins as well as extracellular proteins from this group of organisms are expected to be similar to those of halophilic archaea. However, determination of abundance of amino acids showed that the ratios of acidic amino acids and basic amino acids of bulk proteins from moderately halophilic bacteria were in between those values from nonhalophile, *e.g. E. coli*, and halophilic archaea (Ventosa et al., 1998). This implies a hybride characteristic in term of primary structure of the halotolent proteins. It should be noted that this observation is rather tendency as the abundance of hydrophobic amino acids of moderately halophilic bacteria is comparable to those of nonhalophile (Ventosa et al., 1998). Conversely, the recent study demonstrated that the key of halophilicity was exclusively related to a decrease in the accessible surface area reflected from surface residuals rather than total protein charge (Tadeo, Lopez-Mendez, Trigueros, Lain, Castano, and Millet, 2009). Crystal structure of a halotolerant carbonic anhydrase from salt-tolerant *Dunaliella salina* showed uniformly negative surface charges. This characteristic well-resembles to the highly acidic halophilic proteins, excepting that a lower surface electrostatic potential was detected in the halotolerant protein (Premkumar et al., 2005). However, the evolutionary relationship and knowledge of amino acids composition and structural investigations with respect to saline adaptability in

halotolerant proteins have yet to be elucidated.

2.3.4 Halophilic and halotolerant proteinases in fish sauce

Since protein hydrolytic process is considered to be the major point for accelerating the rate of fermentation, proteinases have become one of major efforts in fish sauce research. During salting process, fish tissue is preliminarily digested by endogenous fish proteinases (Haard and Simpson, 1994; Siringan, Raksakulthai, and Yongsawatdigul, 2006a) which are leached out from viscera and muscles caused by osmotic effect. Major endogenous enzymes include pepsin, trypsin, chymotrypsin, carboxypeptidase and aminopeptidase in vicera and cathepsins, calpain, neutral trypsin-like and alkaline proteinases from fish muscles (Haard, 1994). The hydrolytic products then serve as nutrients for microflora to grow and to begin fermentation. Even though the presence of endogenous fish proteinases' activity was reported throughout 12 months of fermentation (Siringan, Raksakulthai, and Yongsawatdigul, 2006b), the activity was limited due to high salt content (25-30%) of fish sauce (Noda, Vo-Van, kusakabe, and Murakami, 1982). On the other hand, a number of extremely and moderately halophilic bacteria isolated from fish sauce samples have diverted researcher's attention. Their ability to grow in hypertonic saline (above 30% NaCl) and their proteolytic enzymes showing ability to hydrolyze proteins at high salt concentration are presumed as the main source in changing fish proteins into liquid. Thus, halotolerant/halophilic proteinases, which are active at high salt concentrations, produced from microfloral halophilic bacteria are believed to play the major role during fermentation process.

Identification of specific microorganisms giving high proteinases production in the fermentation process has been reported. These include halotolerant or halophilic

archaea/eubacteria from various genera. In Thai fish sauce, Tongthai and Suntinanalert (1991) isolated two archeal strains including *Halobacterium* sp. and *Halococcus* sp. which produce extracellular proteinases hydrolyzing gelatin and caseins. These proteinases were salt activated and required high salt concentrations (~25%) for optimal activity. Chaiyanan et al. (1999) isolated a novel bacterial strain, *Halobacillus thailandensis* sp. nov., from Thai fish sauce. This strain secreted two serine proteinases with MMs of 100 and 17 kDa, and a metalloproteinase with MM of 42 kDa. Hiraga et al. (2005) purified and characterized a proteinase secreted by *Filobacillus* sp RF2-5 isolated from Thai fish sauce. This enzyme was moderately thermophilic and halo-alkaliphilic exhibiting NaCl-activated activity and pH optimum at 10-11. Based on N-terminal sequence, it showed a high identity towards *B. subtilis* serine proteinase with unique substrate specificity. Namwong et al. (2006) characterized a serine proteinase produced by *Halobacillus* sp. SR5-3 isolated from Thai fish sauce. This enzyme exhibited MM of 43 kDa and was a haloalkaliphilic proteinase. However, this enzyme was not stable at low salt conditions (0-5% salt).

Besides, *Bacillus* species were often reported as a microflora in fish sauce and other salted fermented food. Lee, Chun, Kho, and Chun (1998) firstly reported an extracellular leucine aminopeptidase with MM of 58 kDa from *Bacillus* sp. N2 isolated from anchovy sauce. In Vietnamese fish sauces, To, Tanaka, and Nagano (1997) and Nagano and To (2000) isolated a gram-positive rod bacterium, *B. subtilis* FS2, which produced a collagenolytic enzyme. This enzyme showed MM of 125 kDa and exhibited maximum activity at pH 9.0 and 50°C. The enzyme degraded casein liberating hydrophilic and hydrophobic amino acids such as Asn, Gly, Val, and Ile. Tran and Nagano (2002) isolated a halotolerant *B. subtilis* CN2 from Vietnamese fish

sauce. This strain secreted a proteinase with similar activity as the collagenase from *B. subtilis* FS2. The purified enzyme was characterized by Uchida et al. (2004) and was classified as alkaline serine proteinase with MM of 27.6 kDa. Interestingly, this enzyme exhibited SDS resistance which was suitable for industrial detergent formulation. Kim, Nishiyama, Mura, Tokue, and Arai (2004) purified and characterized a metalloproteinase from *B. vietnamensi*s 11-4. The enzyme was a neural proteinase showing optimum pH at 7.5 and was a salt-activated enzyme (halotolerant). Amino acid sequence of this enzyme showed homology towards thermolysis of *B. thermoproteolyticus*.

In other salted fermented food, Kim and Kim (2005) isolated *B. subtilis* JM-3 from anchovy sauce. This strain produced a trypsin-like serine proteinase with MM of 17.1 kDa and optimum pH at 5.5. The enzyme was partially-halotolerant as it was dramatically inhibited at high salt concentrations (1.3% remaining activity at 30% NaCl). Setyorini, Takenaka, Murakami, and Aoki (2006b) purified two novel salt activated halotolerant extracellular proteinases, namely expro-I and expro-II, from *B. subtilis* FP-133 isolated from fermented fish paste. The expro-I was a neutural-serine proteinase showing optimum pH at 7.5 and MM of 29 kDa while the expro-II was a metalloproteinase with MM of 34 kDa. Setyorini, Kim, Takenaka, Murakami, and Aoki (2006a) characterized an intracellular aminopeptidase from a halotolerant strain, *B. subtilis* FP-133, isolated from a fish paste. The enzyme contained four subunits each of 14 kDa (total 59 kDa) and was halotolerant showing optimum activity at 5% NaCl. Recently, Toyokawa et al. (2010) isolated a thermotolerant bacterium *B. licheniformis* RKK-04 from Thai fish sauce. This strain produced a subtilisin-like alkaline serine proteinase with MM of 31 kDa and optimum catalytic conditions at pH

10.0 and 50°C. The enzyme was halotolerant showing 60% activity at 30% NaCl compared to its activity in the absence of NaCl. It should to be noted that proteinases produced from fish sauce and other related fish products microflora were mostly serine- and metallo-proteinases, while halophilic or halotolerant activity of the enzymes is not an absolute characteristic inherited from their producers.

2.4 Microbial proteinase research approaches

2.4.1 Genetic-based approaches

More than 50% of the industrially important enzymes are produced from genetically engineered microorganisms (Hodgson, 1994). In fact, gene cloning and manipulation give benefits by enabling (i) enzyme overproduction in order to commercialize with high yield at a low-cost method, (ii) details of the primary structure of the protein, and (iii) protein engineering to locate the active-site residues and/or to alter the enzyme properties to suit its commercial applications (Rao et al., 1998). Moreover, because proteinases are a complex group of enzymes and vary enormously in their physicochemical and catalytic properties, study of nucleic acid and amino acid sequence homology is important for a variety of purposes and has therefore become routine in computational molecular biology (Argos, 1987).

Gene identification and cloning of halotolerant/halophilic proteinases by PCR-based methods using degenerated primers have been reported. Accordingly, cloning of halophilic proteinases from *Natrinema* sp. J7 (Shi, Tang, Huang, Gan, Tang, and Shen, 2006), *Bacillus* sp. (Okuda et al., 2004) was achieved using primers designed by DNA homology search, and from *B. vietnamensis* (Kim et al., 2004) using Inverse Polymerase Chain Reaction (IPCR) method, where the primers were degenerated by

N-terminal and internal amino acid sequences of the enzyme. Extracellular proteinases' gene identification using genomic DNA expression library was also reported. The main advantage of this approach is that only active proteinase genes in active/inactive genes pool in a bacterial genome are theoretically obtained. *Bacillus subtilis* is often used as an expression host (Tran, Wu, and Wong, 1991; Zabolotskaya, Demidyuk, Akimkina, and Kostrov, 2004), while using *E. coli* as a host for constructing the genomic library of proteinases from Gram-positive bacteria was also reported (Maciver, McHale, Saul, and Bergquist, 1994; Sareen, Bornscheuer, and Mishra, 2005; Tang et al., 2004). The screening system of this method is based on the formation of halo (clear zone) around bacterial colonies on casein-containing agar plates, indicating proteolytic degradation of casein by secreted proteinases from transformants. Hence, bacillus promoters and signal peptides are customarily compatible in *E. coli*. However, the accumulation of inactive enzymes in inclusion body of *E. coli* has been reported in cloning of mature enzyme (excluding of signal and pre-sequences) of an extracellular alkaline serine proteinase of *B. pumilus* (Pan, Huang, and Zhang, 2004).

Another gene identification method of *Bacillus* proteinases is southern hybridization using DNA probes (Maciver et al., 1994; Miyaji, Otta, Nakagawa, Watanabe, Niimura, and Tomizuka, 2006). The probes were designed based on sequence homology and used to hybridize DNA fragments containing interested proteinase gene after genomic DNA was restrictively or randomly digested. Obtained data provided information of a complete set of gene including its promoter, ribosomal binding site and transcriptional terminator site, which are informative for further steps of gene manipulation.

2.4.2 Proteomic-based approaches

2.4.2.1 Gel-based proteomic approaches

The most popular method for analyzing heterogeneous mixtures of protein is sodium dodecylsulfate gel electrophoresis (SDS-PAGE). This technique exhibits remarkable resolving power and is possible to determine the MM of proteins. Based on the gel electrophoresis, proteinases can be detected directly using a technique called "*Zymography"* or "*In-gel activity assay"*. This method is based on the incorporation of enzyme substrate, such as gelatin, casein, and fibrin into a gel slab (Kelleher and Juliano, 1984; Lantz and Ciborowski, 1994; Garcia-Carreno, Dimes, and Haard, 1993). The acrylamide gel is prepared conventionally, excepting that a selected protein substrate is co-polymerized or conjugated to the acrylamide polymer using glutaraldehyde. Based on the hypothesis that most proteinases exhibit overlapping protein substrate specificity, the location of proteinases on the gel can be determined as clear zones (where the protein substrates are digested) against dark background after incubating the gel in an appropriate reaction buffer followed by staining the gel.

Recently, using a peptide substrate labeled with a chromo- or fluorescent-tag to detect proteinase activities after the SDS-PAGE has been reported (Hou, Chen, Chen, and Lin, 1999; Williams, McGrath, and Mangel, 2000; Yasothornsrikul and Hook, 2000; Zhao and Russell, 2003). Specific aminoacyl or peptidyl p-nitroanilides or other chromatic tags are either copolymerized with acrylamide or added into an appropriate reaction buffer. Proteinase activity can be monitored chromatically when the chromatic product are liberated by proteinase action. Based on high specificity of the peptide substrates, not only the detection of MM and proteinase activity but also

proteinase classification can be effectively achieved in a single experiment. Besides, proteinases with different MMs sharing activity towards the same substrate can be detected and quantified on a single gel. In addition, proteinases in extremely low quantity as in picogram range can be detected by this method (Kleiner and Stetler-Stevenson, 1994).

1-dimentional gel electrophoresis is possible to resolve proteins in a sample containing less than approximately 100 components as in required resolution. In contrast, 2-dimensional (2D) PAGE is the only technique available todate for the theoretically simultaneous separation of 10,000 of proteins and practically detection of 1,500-2,000 proteins on one gel plate (O'Farrell, 1975; Walker, 2002). The 2D-electrophoresis has become popular in proteomic studies in recent years (Ong and Pandey, 2001). It is a method combining isolectric focusing (IEF) in the first dimension and SDS-PAGE in the second dimension. In the first dimension step, the IEF, proteins which are amphoteric molecules carrying a positive, negative, or zero net charge, depending on the pH of their surroundings are separated according to their p*I*. This can be used to separate proteins with an only one unit of charge difference (O'Farrell, 1975). Following the IEF, the second-dimension separation is carried out by standard SDS-PAGE which individual proteins are separated according to their MM. As the proteins are separated by IEF in the first dimension and by SDS-PAGE in the second dimension, distinct characters including p*I* and MM of proteins are simultaneously determined. Moreover, protein identification can be directly achieved after 2D SDS-PAGE using Mass spectrometry (MS). Protein identification by 2D SDS-PAGE coupled with MS has become popular in the past decade. Details of this approach are described below.

2.4.2.2 Non-gel based proteomic approache by Mass spectrometry

Mass spectrometry (MS) has become a classical method for identification of interested proteins in solution or after electrophoresis. Recently, MS has essentially replaced other classical techniques for protein identification, for example, N-terminal sequencing based on Edmen degradation. This technique is much more sensitive and can cope with protein mixture offering much higher sample throughput. Current MS-based proteomic strategies rely primarily on digestion of electrophoretically separated proteins into peptides rather than the direct use of intact proteins. Trypsin is normally used as a sequence-specific proteinase for generating the peptides. It cleaves peptide chains exclusively at the carboxyl side of the amino acids lysine or arginine (Olsen, Ong, and Mann, 2004). The main advantage of this approach is that the tryptic peptides can be effectively recovered from acrylamide gels compared to intact proteins . Polypeptides can be transferred into gas phase by the ionization techniques including Electrospray ionization (ESI) (Fenn, Mann, Meng, Wong, and Whitehouse, 1989) and matrix assisted laser desorption/ionization (MALDI) (Karas and Hillenkamp, 1988). A small number of peptides usually yield sufficient information to permit protein identification by peptide mass fingerprinting (PMF) and partial amino acid sequencing. The principles of these procedures are described below.

2.4.2.2.1 Peptide mass fingerprint (PMF) using MALDI-TOF

MALDI is a soft ionization method, allowing for analysis of large biomolecules that tend to fragment when conventional ionization methods are used (Karas and Hillenkamp, 1988). The ionization process is triggered by laser beam. A matrix is used to protect the sample from being destroyed by direct laser beam during ionization. The sample is mixed and co-crystallized with a selected matrix, for

example, α- cyano-4-hydroxycinnamic acid for peptide smaller than 10,000 Da and sinapinic acid for higher mass peptides (Hillenkamp, Karas, Beavis, and Chait, 1991). When a pulsed laser beam hits the matrix, partially vaporized matrix is desorbed from the surface and carries intact sample into gas phase.

Protons are exchanged between analytes and matrix during expansion of the MALDI plume, resulting in formation of charged analytes (Karas, Gluckmann, and Schafer, 2000). For PMF analysis, MALDI are typically coupled with time-of-flight (TOF) mass analyzer. MALDI-TOF MS is a method which the velocity of the ion depends on ions-mass-to-charge ratio (m/z) of the particle (heavier particles reach with lower speeds) (Weickhardt, Moritz, and Grotemeyer, 1996). This technique provides high sensitivity from femtomole to the low picomole with mass resolution range of 3,000 to 6,000 fwhm (full width at half maximum intensity) (Whittal and Li, 1995). Protein identification by PMF is based on the principle that absolute masses of peptides obtained from a specific proteolysis of a protein can be accurately measured, where a unique mass fingerprint pattern of each individual protein is often obtained (Fabris et al., 1995). Trypsin is usually used as a specific proteinase to generate peptides (Olsen et al., 2004). Due to the high specific activity of trypsin, the masses of tryptic peptides can be theoretically predicted from any entry in a protein sequence database.

The protein can be identified by comparing experimentally obtained peptide masses with those obtained theoretically. The protein query is ranked according to the number of peptide matches and scoring algorithms are used to evaluate the mass accuracy and percentage of the protein sequence coverage (Pappin , Hojrup, and Bleasby, 1993). Mass accuracy is an important parameter for protein identification

that determines the level of confidence for the match, thus higher mass accuracy could ensure the accurate results (Perkins et al., 1999).The typical procedure for protein identification after gel electrophoresis by PMF using MALDI-TOF is shown in Figure 2.1a,b.

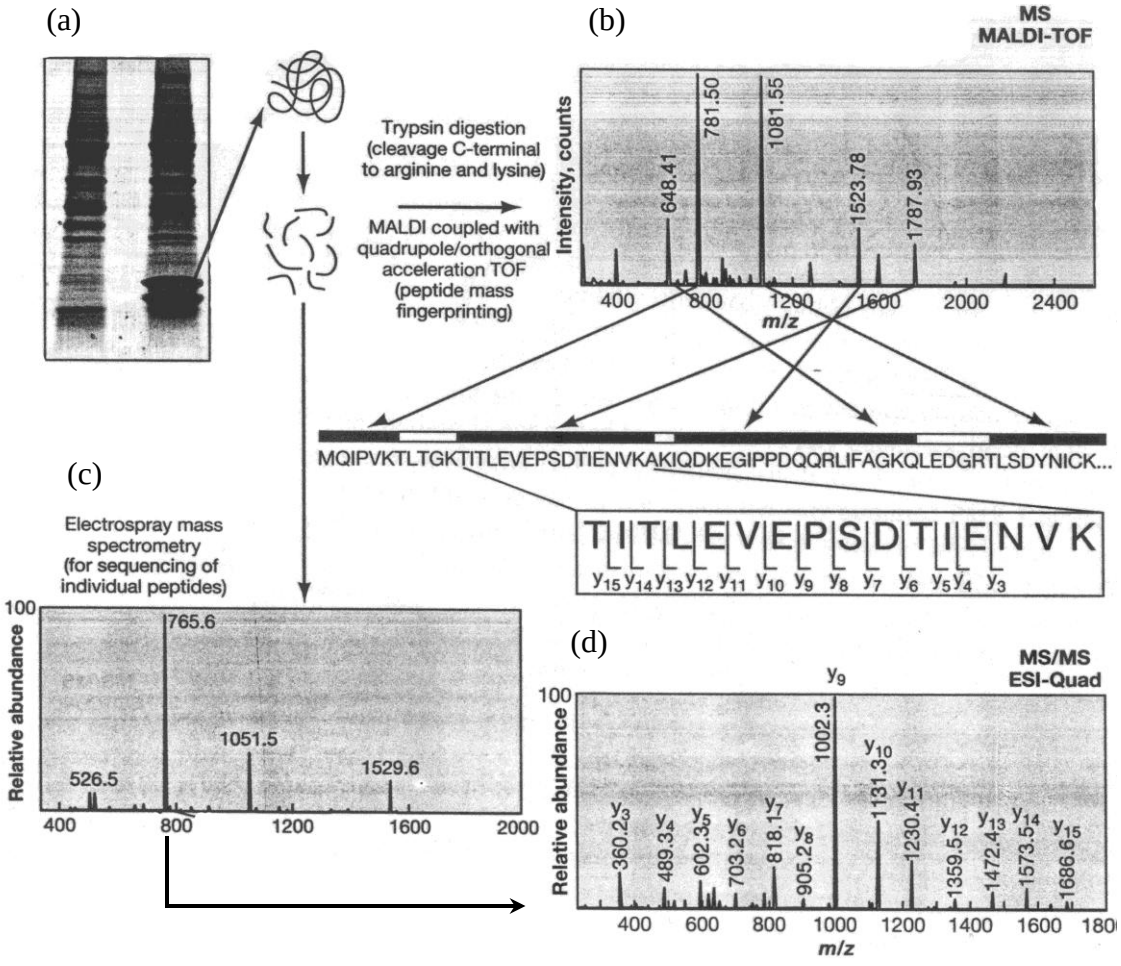


Figure 2.1 A strategy for mass spectrometric identification of proteins. (a) After gel electrophoresis, the protein bands of interest are excised and subjected to trypsin digestion. (b) Analysis of an aliquot of the tryptic peptide mixture using a MALDI-TOF. The resultant spectrum represents a peptide mass

(Figure 2.1 Continued)

fingerprint (PMF) of a protein. This information permits the identification of the protein by matching of the experimentally obtained peptide masses against theoretically predicted peptides of proteins in publicly available database. Tryptic peptides derived from human ubiquitin (sequence below) are shown as an example. (c) Analysis of another aliquot of the peptide mixture using an electrospray ion-trap mass spectrometer, which usually coupled on-line to a RP-HPLC. Peptide masses at a given of time obtained from the first mass analyzer. One of these ions is selected to enter the collision cell and then the ion is fragmented. The fragmented ions are separated in the second mass analyzer before being scanned out to the detector. (d) From the resultant MS/MS spectrum, amino acid sequence information can be derived using an algorithm that correlates the experimental spectrum with those in a database. In this example, the fragment ions in this spectrum correspond to amino acid residues TITLEVEPSDTIENVK. Protein can be identified from such peptide sequence information.

Modified from: Zugaro et al., 1998 and Simpson, 2003

2.4.2.2.2 Partial amino acid sequencing using Tandem mass spectrometry

Tandem mass spectrometry (MS/MS) is a technique which is coupled with several mass analyzers. These mass analyzers are designed to select, collide, and separate ions which provide extensive benefits for structural determination (Canas, Lopez-Ferrer, Ramos-Fernandez, Camafeita, and Calvo, 2006). In tandem mode, after

the masses of mixture peptides are analyzed in the first mass analyzer, one of the peptides are isolated and further fragmented by collision-induced dissociation (CID), by collision with inert gas such as argon and nitrogen in a collision cell. The resulting fragments are subsequently analyzed in the second mass analyzer in order to sequence the peptide. The fragments obtained by this method are derived from the amino and carboxyl terminus of the peptide and are designated as *b* and *y* ions, respectively. The sequence of peptides determined from the fragmentation in MS/MS is called *de novo* peptide sequencing. As peptides derived from tryptic digestion contain either arginine or lysine residue at C-terminus and may also be doubly charged or triply charged (Canas et al., 2006), the *de novo* sequencing is typically performed using these ions since they are fragmented easily and provide structural information at lower activation energy. CID spectra always contain overlapping *b*- and *y*-series ions, multiple internal ions from the same peptide, and immonium ions, so obtained spectrum is redundancy. The immonium ions are derived from double cleavage of the peptide backbone, N- and C- terminal to the amino acid residues. These ions can indicate or confirm the presence of individual amino acids. The redundancy information makes CID spectra a rich source of sequence information that is highly specific to an individual peptide. Because peptide fragmentation by the CID process is in a predictable manner, sequences in the database can be used to predict an expected pattern and thus used to match the experimentally observed MS/MS spectrum. Therefore, amino acid sequences of the peptide are resolved and then partial amino acid sequences of the tested proteins can be obtained. Even though a complete set of *b*- and *y*- ions confirms the entire peptide sequence, not all fragment ions are present at detectable levels and

thus it is often possible to interpret only part of the sequence with confidence. More description for partial amino acid sequencing is illustrated in Figure 2.1c,d.

2.4.2.2.3 PMF versus partial amino acid sequencing

The accumulation of nucleotide and protein sequence collection in various databases allows the rapid characterization of proteins via mass spectrometry. Table 2.2 shows a comprehensive list of PMF and MS/MS search engines available to date. A potent disadvantage of protein identification by MS approach is that, for the best result, proteins of interest are needed to be well-documented in term of amino acid sequence or degenerated nucleotide sequence and their primary structure is the direct product from transcription (Yates, Eng, McCormack, and Schieltz, 1995). The main disadvantage of PMS is that tryptic peptides derived from complex protein mixtures are not suitable, because it is unclear that all peptide masses observed are originally from the same protein species. In addition, this method is not suitable for searching of expressed sequence tag (EST) databases (Simpson, 2003). The EST databases are deduced protein sequence libraries of a complete genome sequence whose their six-frame translations are encoded. These databases are an alternative option for protein identification when the complementary DNA (cDNA) libraries of an interested organism have not been determined. Since the ESTs are normally present only portions from gene-encoding sequences, they may be too short to cover a sufficient number of peptides observed in the experimentally obtained PMF. Therefore, it is more difficult to identify because the peptide masses may correlate to the theoretical digest of an unrelated protein leading to a false identification (Simpson, 2003). Advantages of partial amino acid sequencing approach compared to PMF are that a protein in a complex mixture can be identified from the CID spectrum of a single

peptide and matches one or more tandem mass spectra to peptide sequences in the same protein, this provides a high level of confidence in the identification process (Simpson, 2003).

Table 2.2 A comprehensive list of PMF and MS/MS databases

Approach	Program name	Web Site/e-mail	Availability
Peptide mass fingerprinting	FRAGFIT	ckw@gene.com	C-language source code, available via e-mail
	MassSearch	http://cbrg.inf.ethz.ch	Automated sever at ETH, Zurich
	PeptideSearch	http://195.41.108.38	Free Web searches
	MOWSE (molecular weight search)	http://srs.hgmp.mrc.ac.uk/cgi-bin/mowse	Free Web searches
	MS-Fit	http://prospector.ucsf.edu	Free Web searches
	Mascot	http://www.matrix-science.com	Free Web searches
	Pepldent2	http://expasy.ch/tools/peptident.html	Free Web searches
	ProteinLynx Global SERVER	http://www.micromass.co.uk	Commercial product
	ProFound	http://prowl.rockefeller.edu	Free Web searches
	PeptideMass	http://expasy.org/tools/	Free Web searches
Peptide ion fragment spectra (MS/MS)	SEQUEST	http://www.thermoquest.com	Commercial product
	MS-Tag	http://prospector.ucsf.edu	Free Web searches
	PepFrag	http://prowl.rockefeller.edu	Free Web searches
	Mascot	http://matrix-science.com	Free Web searches
	ProteinLynx Global SERVER	http://www.micromass.co.uk	Commercial product
	MassFrag FRAGFIT Find Protein	Jeanluc.Verschelde@rug.ac.be	Free Web searches

Modified from: Simpson, 2003

2.5 Whole genome sequencing

2.5.1 Genome sequencing technologies

Whole genome sequencing or complete genome sequencing is a method for determination of complete DNA sequences containing in an organism. Besides chromosomal DNA, this also includes plasmids for single cell microorganisms, mitochondria for mammals, and chloroplast for plants (Markowitz et al., 2010). An overview of current whole genome sequencing technologies is shown in Figure 2.2.

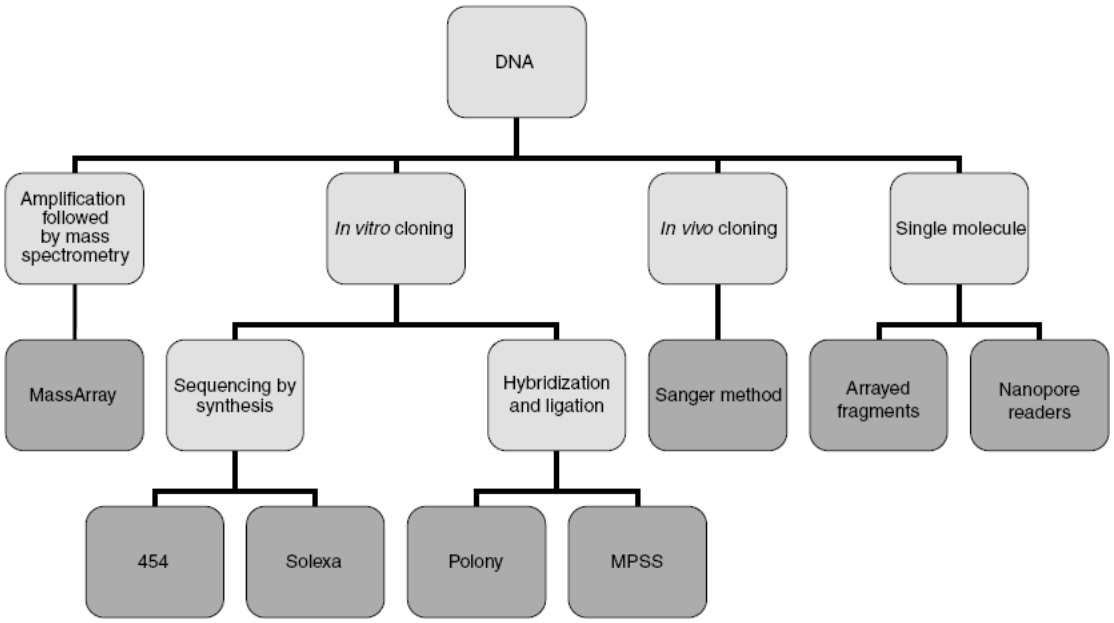


Figure 2.2 Diagram of current technologies for genome sequencing.

Modified from: Hall et al. (2007)

The sequencing techniques can be categorized into four main strategies including amplification and mass spectrometry, *in vitro* cloning, *in vivo* cloning, and single-molecule approaches. Sequencing by mass spectrometry and the single-

molecule approaches are still very under the developmental stage (Hall, 2007). However, the mass spectrometry methods, for example MassArray method, are usually used for single nucleotide polymorphism (SNP) analysis (Jurinke, van den Boom, Cantor, and Koster, 2002). *In vivo* cloning followed by Sanger sequencing is a classical methods for genome sequencing. This method is based on synthesizing DNA on a single stranded template in the presence of chain terminators (Sanger, Nicklen, and Coulson, 1977). The chain terminators are randomly incorporated into the synthesizing DNA and generate a range of different fragment sizes corresponding to the positions of the terminators. In the older methods, each reaction has a different base as a terminator (one for each base: G, A, T and C), while the size of each fragment can be identified by running on a gel. The current method is based on the use of different colored fluorescent dyes to label terminators (Prober et al., 1987), so that all of the terminators are incorporated in a single reaction and the size of each fragment is separated on capillaries, giving the average length of a sequencing read at 850·bp (Madabhushi, 1998). Genomic DNA is firstly fragmented by restriction nuclease digestion or mechanical shearing or a combination of these approaches. Subsequently, DNA fragments are conventionally cloned, transformed, and then a statistically significant number of randomly selected transformants is subjected to sequencing (Anderson, 1981; Roach, Boysen, Wang, and Hood, 1995). This approach was later considered as laborious and time consuming which was no longer applicable when most of sequencing projects had gone too far into more complicated genomes, like human genome. In addition, one of the main drawbacks this approach is that there are inherent biases against certain stretches of DNA, such as some DNA fragments contain regions that do not replicate well in *E. coli* or that code for toxic compounds

(Hall, 2007). Therefore, the genome sequencing by Sanger technique is currently replaced by *in vitro* cloning, or called "Next-generation sequencing" (NGS) technologies.

The NGS technologies are a high-throughput sequencing technique that parallelizes sequencing process, producing thousands or millions of sequences at once (Hall, 2007). Examples of such well-known and most currently promising technologies of NGS include 454 pyrosequencing (Margulies et al., 2005), the Solexa system (Bennett, Barnes, Cox, Davies, and Brown, 2005), and the SOLiD system (or Polony) (Shendure et al., 2005). Although these sequencing platforms are different in term of engineering configurations and sequencing chemistries, their technical paradigms can be described below:

454 pyrosequencing This technique is currently developed by ROCHE/454 LIFE SCIENCES (http://www.454.com). It is derived from the combination of single-molecule emulsion PCR and pyrosequencing (Margulies et al., 2005). The pyrosequencing is a sequencing approach based on chemiluminescent detection of pyrophosphate released during polymerase mediated deoxynucleoside triphosphate (dNTP) incorporation (Nyren, Pettersson, and Uhlen, 1993). For sequencing, genomic DNA is fragmented by nebulization or sonication, generating a library of template DNA with several hundred base pairs in length. Fragments in the library are sequentially end-repaired, ligated to adapter oligonucleotides, diluted to single-molecule concentration, denatured, and hybridized to individual beads containing sequences complementary to adapter oligonucleotides. The next step is the single-molecule emulsion polymerase chain reaction (emPCR) (Tawfik and Griffiths, 1998). The DNA library fragments captured on beads, enzyme, and PCR reagents in a water

mixture are injected into small, cylindrical plastic containers containing synthetic oil, creating the water mixture with the beads in oil droplets, called a water-in-oil emulsion. Typically, most droplets that contain DNA will contain only one DNA fragment. The water mixture includes an enzyme that causes the single and isolated DNA fragment in each droplet to be amplified into millions of copies of DNA. After amplification, the emulsion is disrupted and the beads are separated by limiting dilution. Then the beads are deposited into individual "picotiter-plate" wells and combined with sequencing enzymes. The "picotiter plate" well is made from a fused fiber-optic bundle containing approximately $3.4x10^6$ picoliter-scale sequencing-reaction wells (Voelkerding, Dames, and Durtschi, 2009). The picotiter plate functions as a flow cell in which iterative pyrosequencing is performed by successive flow of the 4 dNTPs. As a nucleotide incorporates in a clonally amplified template in a well, it releases pyrophosphate which produces localized luminescence. The luminescent signal is transmitted through the fiber-optic plate and recorded on a charge-coupled device camera. With the flow of each dNTP reagent, wells are sequentially imaged, analyzed for their signal-to-noise ratio, filtered according to quality criteria, and subsequently algorithmically translated into a linear sequence output. A single run often generates approximately $1x10^6$ sequence reads, with read lengths of ~400 bases yielding up to 500 million base pairs (Mb) of sequence (Voelkerding et al., 2009).

Solexa/Illumina This technique is acquired by Illumina (http://www.Illumina.com) in 2006. The principal of this approach is to sequence single DNA molecules attached to microspheres (Bennett et al., 2005). A flow cell consists of an optically transparent slide with 8 individual lanes on the surfaces, of

which are bound oligonucleotide anchors. Template genomic DNA is fragmented into several hundred base pairs in length and subsequently end-repaired to generate 5′-phosphorylated blunt ends. Klenow fragment is used to add a single A base to the 3′ end of the blunt phosphorylated DNA fragments. This increases ligation efficiency of the DNA fragments and oligonucleotide adapters since the adaptors have an overhang of a single T base at their 3′ end. The flow-cell anchors are complementary to the adapter oligonucleotides. The single-stranded template DNA is added to the flow cell under limiting-dilution conditions and immobilized by hybridization to the anchors. Unlike emPCR, DNA templates are amplified in the flow cell in a distinct geometry called "bridge" amplification. This is explained by the captured DNA strands arch over and hybridize to an adjacent anchor oligonucleotide. The single-molecule DNA template is converted to a clonally amplified arching "cluster" via multiple amplification cycles, wherein each cluster contains approximately 1000 clonal molecules (Voelkerding et al., 2009). For sequencing, the clusters are denatured followed by a chemical cleavage reaction and washing, which leave only forward strands for single-end sequencing. The forward strands are sequenced by hybridizing a primer complementary to the adapter sequences, followed by addition of polymerase and a mixture of 4 differently colored fluorescent reversible dye terminators. The terminators are incorporated into in each strand in a clonal cluster if their sequences are complementary. After incorporation, excess are washed away and the fluorescence is recorded from the clusters. To begin the next round, the reversible dye terminators are unblocked, the fluorescent labels are cleaved and washed away, and the next iterative sequencing cycle is performed. This process requires approximately 2.5 days for the read lengths of 36 bases (Voelkerding et al., 2009). With respect to

approximately 50 x10^6 separate clusters can be generated per flow cell, the overall sequence output is ~1 billion base pairs (Gb) per analytical run (Bentley et al., 2008).

SOLiD/APPLIED BIOSYSTEMS The SOLiD (Supported Oligonucleotide Ligation and Detection) system is distributed by Applied Biosystems (http://www.solid.appliedbiosystems.com). This technology is based on ligation of short-read sequencing (Shendure et al., 2005). The preparation of library of DNA template shares similarities with the 454 technology, where DNA fragments are ligated to oligonucleotide adapters, attached to beads, and clonally amplified by emPCR. After clonally amplification, beads with template are immobilized onto a derivitized-glass flow-cell surface. Sequencing is initiated by annealing a primer oligonucleotide complementary to the adapter at the adapter template junction. During the first "ligation sequencing" step, the primer is oriented to provide a 5′ phosphate group for ligation to interrogation probes instead of providing a 3′ hydroxyl group. The interrogation probe is an octamer consisting of (in the 3′-to-5′ direction) 2 probe specific bases followed by 6 degenerate bases, where one of 4 fluorescent labels linked to the 5′ end. The 2 probe specific bases consist of one of 16 possible 2-base combinations (for example TT, GT, and so forth) (Voelkerding et al., 2009). In the first ligation-sequencing step, interrogation probes representing the 16 possible 2-bases compete for annealing to the template sequences. After annealing, a ligation step is performed by thermostable ligase, followed by washing step to removal unbound probe. Optical intensity of fluorescence signals are collected upon the cleavage of the ligated probes and another washing step is employed to remove the fluor and regenerate the 5′ phosphate group. In the subsequent sequencing steps, interrogation probes are ligated to the 5′ phosphate group of the preceding pentamer.

One round of reaction is referred to seven cycles of ligation performed to extend the first primer. The synthesized strand is then denatured, and a new sequencing primer offset by 1 base in the adapter sequence (n-1) is annealed. Five rounds total are performed, each time with a new primer with a successive offset (n-2, n-3, and so on). By this sequencing manner, each template nucleotide is sequenced twice and 6 days of instrument run are needed to generate sequence read lengths of 35 bases (Voelkerding et al., 2009). By Placing 2 flow-cell slides in the instrument per analytical run, a combined output of 4 Gb of sequence or greater can be produced (Voelkerding et al., 2009).

These techniques are a high-throughput and cost-effective sequencing approach. Since the first NGS platform was commercially lunched in 2005 (Voelkerding et al., 2009), the number of successfully sequenced genomes has dramatically gone up ever since 2006 (Figure 2.3). This is a promising technology for genetic-based research in the near foreseeable future and could be also a disruptive or compliment technology towards any other sequencing technologies *e.g.* DNA array (Hurd and Nelson, 2009; Teng and Xiao, 2009).

2.5.2 Genomic overview of *Bacillus* species

According to the Integrated Microbial Genomes (IMG) database, more than 6115 genomic sequences have been publicly available to date (Markowitz et al., 2010). These genomes were from different biota including eukaryotes, bacteria, viruses, and archaea. Due to the significance in term of clinical and biotechnological uses, viruses and bacteria are the largest groups of microorganisms whose their genomic world has been extensively investigated (Figure 2.4a). However, out of 9.4 million of total gene counts, genes with bacterial origin take up more than 87.3% of

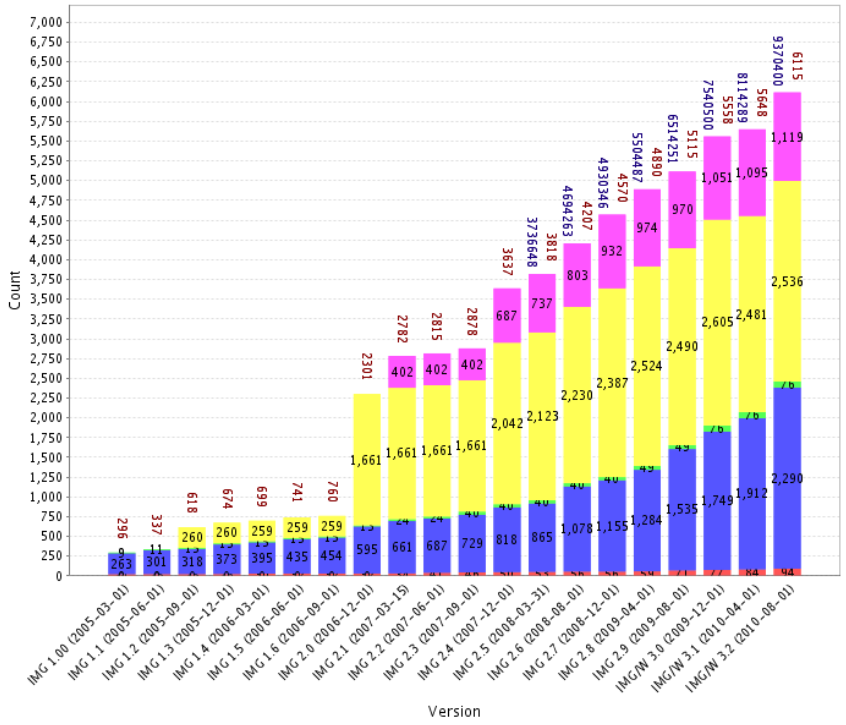


Figure 2.3 Numbers of genes and complete genomes submitted into Integrated Microbial Gnome (IMG) institute during the past five years. Blue and red numbers indicate total gene and complete genome in each column, respectively. Black numbers in red, blue, green, yellow and pink bars represent genome count in archaea, bacteria, eukaryote, viruses, and plasmid, respectively.

From: Markowitz et al., 2010 (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi)

all sequences in the databank (Figure 2.4b). Thus, the knowledge of bacterial genomes and their coding products are currently most comprehensively established in diverse dimensions, such as gene, genome, and function, along with the expanding of sequencing technologies. Regarding to halophilic bacteria, in archaea, family *Haloteriaceae* contains most halophiles, of which, the genome sequence of *Halobac-*

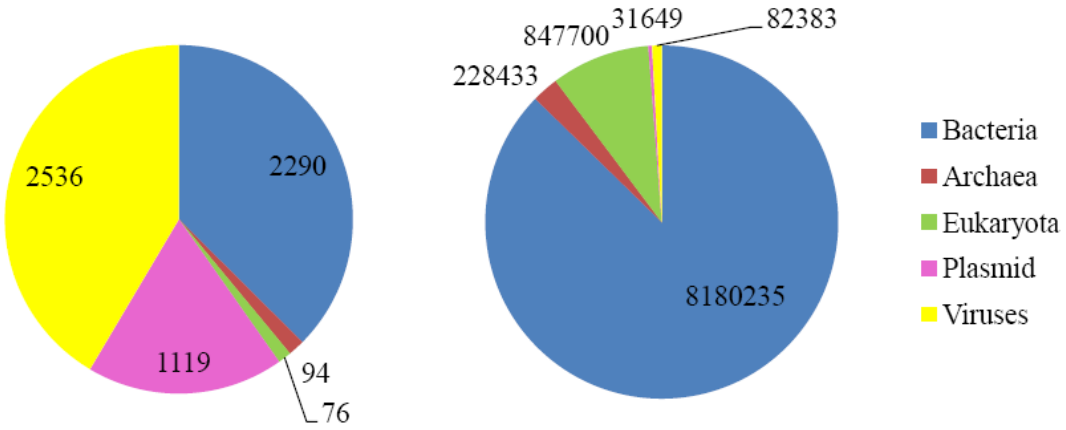


Figure 2.4 Genome (a) and gene count (b) in each group of organisms including genetic material (plasmid) retrieved from the most updated IMG statistics database (01-08-2010).

Modified from: Markowitz et al., 2010 (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi)

terium sp. NRC-1 is known and well-studied (Ng et al., 2000). In bacterial kingdoms, halophiles have been reported in *Cyanobacteria*, *Proteobacteria*, *Actinobacteria*, *Spirochaetes*, *Firmicutes*, and *Bacteroidetes* (Oren, 2008). However, only a few genome sequences have been investigated, including *Geobacillus kaustophilus* and *Oceanibacillus iheyensis* from *Firmicutes* (Takami et al., 2000; 2002) and *Salinibacter ruber* from *Bacteroidetes* (Mongodin et al., 2005). Unfortunately, genomic information of bacteria in genus *Virgibacillus*, a group of halophiles, has never been reported. The genome of halophilic species is characterized by a common characteristic of dinucleotide abundance profiles of halophilic genomes, which is unique from those of non-halophiles. Halophiles also exhibit similar patterns of synonymous codon usage, indicating convergent evolution of halophilic species regardless of their long-term evolutionary history (Paul et al., 2008).

In *Bacillus* species, 121 genomes have been determined. Of which, 38 sequences are complete genomes (called finished genomes), while others are incomplete or unfinished genomes (called draft genomes). Table 2.3 shows genomic statistics of some selected finished genomes. These bacteria were isolated from different sources including soil, food, human intestine, freshwater and marine. According to Table 2.3, the genomic sizes were found in between 3.6 and 5.9 Mbp with average % DNA coding number and %G+C of 86.1 and 41.6, respectively. Maximal number of protein coding sequences (CDSs) was found in *B. cereus* (5878 CDSs) compared to an average of 4384 CDSs. Herein, CDSs with function prediction is at 65.4% of the average CDSs, where 24.3% of which are related to enzymes. In addition, 71.5% of the average CDSs were found to link to cluster of orthologous groups (COGs). Bacteria in *Bacillus* species contain average rRNA operon (5S, 16S, and 23S rRNA subunits) and tRNAs of 9.8 and 85.9 genes, respectively. Plasmids were also detected in most species with an amount ranging from 1 to 4 plasmids per bacterium. This indicates the significance of these transferable elements in horizontal gene transfer events in the evolution of this group of organisms. Almost half (49.7%) of the average CDSs are members of secreteome (22.1 and 27.6% of CDSs contain signal peptide for secretion and transmembrane domains, respectively). It should be noted that more than 34.6% of the average CDSs from all genomes are proteins with unknown function, which are mysterious and need efforts to unravel their biological significances and also other novel perspective applications.

Table 3. Genome statistics of *Bacillus* species

	Bac amy FZ2	Bac ant Anr	Bac cer 107	Bac cla KS6	Bac hal C-5	Bac lic (N)	Bac meg DS	Bac pse OF4	Bac pum SA2	Bac sel ML0	Bac sub 168	Bac thu 977	Bac wei KB4	Geo kau HT6	Geo the NG2	Oce ihe HT1
Geographic isolation	soil	soil	soil, daily	soil	fresh water, soil	soil	soil	soil	soil	fresh water	soil	soil	soil	marine	fresh water, oil field	marine
Physnotype	mesophile, protease producer	phatogen	food poisoning	alkaliphile, prebiotic, protease producer	protease producer	Phatogen, protease producer	mesophile	alkaliphile, mesophile	mesophile, biomass degradator	alkaliphile, mesophile, halotolerant	mesophile, protease producer	mesophile	mesophile	thermophile , halolerant	thermophile	alkaiphile, halotolerant
DNA, total number of bases	3,918,589	5,503,926	5,432,290	4,303,871	4,202,352	4,222,334	5,097,447	4,249,248	3,704,465	3,592,487	4,215,606	5,314,794	5,872,743	3,592,666	3,608,012	3,630,528
%DNA coding number of bases	89.5	81.6	87.0	87.3	86.5	88.3	83.0	85.9	88.9	87.4	89.2	85.6	84.2	86.6	86.2	85.7
%DNA G+C number of bases	46.5	35.2	35.5	44.8	43.7	46.2	38.1	39.9	41.3	48.7	43.5	35.4	35.5	52.0	48.9	35.7
Plasmid count	0	2	1	0	0	0	0	2	0	0	0	1	4	1	1	0
Genes total number	3814	5853	6126	4261	4239	4420	5248	4434	3823	3420	4354	5452	5983	3714	3642	3664
Protein coding genes	3696	5619	5878	4105	4066	4196	5100	4335	3729	3326	4176	5197	5831	3540	3471	3500
Pseudo Genes	4	2	34	9	0	44	1	0	48	71	0	0	178	0	26	0
RNA genes	118	234	248	156	173	224	148	99	94	94	178	255	152	174	171	164
rRNA genes	30	33	36	22	25	21	33	22	21	22	30	42	42	27	30	22
5S rRNA	10	11	12	8	9	7	11	8	7	8	10	14	14	9	10	8
16S rRNA	10	11	12	7	8	7	11	7	7	7	10	14	14	9	10	7
18S rRNA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23S rRNA	10	11	12	7	8	7	11	7	7	7	10	14	14	9	10	7
28S rRNA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tRNA genes	87	95	97	75	78	72	115	77	70	68	86	105	108	87	87	69
Other RNA genes	1	106	115	59	70	131	0	0	3	4	62	108	2	60	54	73
Protein coding genes with function prediction	3,338 (90.3)	3,201 (57.0)	3,403 (57.9)	2,781 (67.7)	2,175 (53.5)	3,203 (76.3)	3,457 (67.8)	2,382 (54.9)	2,589 (69.4)	2,439 (73.3)	3,261 (78.1)	3,360 (64.7)	3,860 (66.2)	1,944 (54.9)	2,373 (68.4)	1,957 (55.9)
Protein coding genes without function prediction	358 (9.7)	2,418 (43.0)	2,475 (42.1)	1,324 (32.3)	1,891 (46.5)	993 (23.7)	1,643 (32.2)	1,953 (45.1)	1,140 (30.6)	887 (26.7)	915 (21.9)	1,837 (35.3)	1,971 (33.8)	1,596 (45.1)	1,098 (31.6)	1,543 (44.1)
Protein coding genes connected to KEGG pathways	1,137 (30.8)	1,264 (22.5)	1,374 (23.4)	1,316 (32.1)	1,210 (29.8)	1,244 (29.6)	1,335 (26.2)	1,090 (25.1)	1,104 (29.6)	995 (29.9)	1,204 (28.8)	1,400 (26.9)	1,388 (23.8)	1,094 (30.9)	1,120 (32.3)	1,126 (32.2)
not connected to KEGG pathways	2,559 (69.2)	4,355 (77.5)	4,504 (76.6)	2,789 (67.9)	2,856 (70.2)	2,952 (70.4)	3,765 (73.8)	3,245 (74.9)	2,625 (70.4)	2,331 (70.1)	2,972 (71.2)	3,797 (73.1)	4,443 (76.2)	2,446 (69.1)	2,351 (67.7)	2,374 (67.8)
Protein coding genes connected to SwissProt Protein Product	488 (13.2)	707 (12.6)	582 (9.9)	496 (12.1)	726 (17.9)	500 (11.9)	0	0	422 (11.3)	0	4,170 (99.9)	563 (10.3)	434 (7.4)	501 (14.2)	406 (11.7)	547 (15.6)
not connected to SwissProt Protein Product	3,208 (86.8)	4,912 (87.4)	5,296 (90.1)	3,609 (87.9)	3,340 (82.1)	3,696 (88.1)	5,100 (100)	4,335 (100)	3,307 (88.7)	3,326 (100)	6 (0.1)	4,634 (89.2)	5,397 (92.6)	3,039 (85.8)	3,065 (88.3)	2,953 (84.4)
Protein coding genes with enzymes	1,002 (27.1)	1,117 (19.9)	1,215 (20.7)	1,090 (26.6)	1,022 (25.1)	1,075 (25.6)	1,152 (22.6)	938 (21.6)	964 (25.9)	848 (25.5)	1,067 (25.6)	1,233 (23.7)	1,236 (21.1)	955 (27.0)	961 (27.7)	951 (27.2)
Protein coding genes with COGs	2,845 (77.0)	3,488 (62.1)	3675 (62.5)	3,172 (77.3)	3,088 (75.9)	3,091 (73.7)	3,655 (71.7)	2,883 (66.5)	2,746 (73.6)	2,618 (78.7)	3,052 (73.1)	3,608 (69.4)	3,778 (64.8)	2,594 (73.3)	2,593 (74.7)	2,669 (76.3)
Protein coding genes coding signal peptides	777 (21.0)	1,131 (20.1)	1,172 (19.9)	868 (21.1)	803 (19.7)	887 (21.1)	1,464 (28.7)	1,285 (29.6)	787 (21.1)	973 (29.3)	1,205 (28.9)	1,113 (21.4)	1,210 (20.8)	631 (17.8)	622 (17.9)	693 (19.8)
Protein coding genes coding transmembrane proteins	998 (27.0)	1,646 (29.3)	1,722 (29.3)	1,059 (25.8)	1,099 (27.0)	1,158 (27.6)	1,468 (28.8)	1,238 (28.6)	1,053 (28.2)	883 (26.5)	1,157 (27.7)	1,583 (30.5)	1,678 (28.8)	859 (24.3)	887 (25.6)	1,049 (30.0)

Note: 1) Bacterial genomes include of *Bacillus amyloliquefaciens* FZB42 (Bac amy FZ2), *B. anthracis* Ames Ancestor (Bac ant Anr), *B. cereus* ATCC 10987 (Bac cer 107), *B. clausii* KSM-K16 (Bac cla KS6), *B. halodurans* C-125 (Bac hal C-5), *B. licheniformis* ATCC 14580 (Novozymes) (Bac lic (N)), *B. mycoides* DSM2048 (Bac meg DS), *B. pseudofirmus* OF4 (Bac pse OF4), *B. pumilus* SAFR-032 (Bac pum SA2), *B. selenitireducens* MLS10 (Bac sel ML0), *B. subtilis* subsp. subtilis str. 168 (Bac sub 168), *B. thuringiensis* sv konkukian 97-27 (Bac thu 977), *B. weihenstephanensis* KBAB4 (Bac wei KB4), *Geobacillus kaustophilus* HTA426 (Geo kau HT6), *Geobacillus thermodenitrificans* NG80-2 (Geo the NG2), *Oceanobacillus iheyensis* HTE831 (Oce ihe HT1).

2) Numbers in parenthesis are percentage of total protein codding genes presenting in respective genomes

From: Markowitz et al, 2010 (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi)

2.6 References

Akolkar, A. V., Durai, D., and Desai, A. J. (2010). *Halobacterium* sp. SP1(1) as a starter culture for accelerating fish sauce fermentation. **J. Appl. Microbiol.** 109:(1) 44-53.

An, S. Y., Asahara, M., Goto, K., Kasai, H., and Yokota, A. (2007). *Virgibacillus halophilus* sp. nov., spore-forming bacteria isolated from soil in Japan. **Int. J. Syst. Evol. Microbiol.** 57:(Pt 7) 1607-1611.

Anderson, S. (1981). Shotgun DNA sequencing using cloned DNase I-generated fragments. **Nucleic. Acids. Res.** 9:(13) 3015-3027.

Arahal, D. R., Marquez, M. C., Volcani, B. E., Schleifer, K. H., and Ventosa, A. (1999). *Bacillus marismortui* sp. nov., a new moderately halophilic species from the Dead Sea. **Int. J. Syst. Bacteriol.** 49:(Pt 2) 521-530.

Argos, P. (1987). A sensitive procedure to compare amino acid sequences. **J. Mol. Biol.** 193: 385-396.

Ballinger, M. D., Tom, J., and Wells, J. A. (1995). Designing subtilisin BPN' to cleave substrates containing dibasic residues. **Biochemistry.** 34: 13312-13319.

Barr, P. J. (1991). Mammalian subtilisins, The long-sough dibasic processing endopeptidases. **Cell.** 66: 1-3.

Beja, O., Suzuki, M. T., Koonin, E. V., Aravind, L., Hadd, A., Nguyen, L. P., et al. (2000). Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. **Environ. Microbiol.** 2:(5) 516-529.

Belitsky, B. R., Brill, J., Bremer, E., and Sonenshein, A. L. (2001). Multiple genes for the last step of proline biosynthesis in *Bacillus subtilis*. **J. Bacteriol.** 183:(14) 4389-4392.

Bennett, S. T., Barnes, C., Cox, A., Davies, L., and Brown, C. (2005). Toward the 1,000 dollars human genome. **Pharmacogenomics.** 6:(4) 373-382.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. **Nature.** 456:(7218) 53-59.

Betzel, C., Klupsch, S., Papendorf, G., Hastrup, S., Branner, S., and Wilson, K. S. (1992). Crystal structure of the alkaline proteinase Savinase from *Bacillus lentus* at 1.4 A resolution. **J. Mol. Biol.** 223:(2) 427-445.

Betzel, C., Pal, G. P., and Saenger, W. (1988). Three-dimensional structure of proteinase K at 0.15-nm resolution. **Eur. J. Biochem.** 178:(1) 155-171.

Beynon, R. J., and Bond, J. S. (2001). **Proteolytic enzymes a practical approach** (2 ed.). New York: Oxford university press.

Bohm, G., and Jaenicke, R. (1994). A structure-based model for the halophilic adaptation of dihydrofolate reductase from *Halobacterium volcanii*. **Protein Eng.** 7:(2) 213-220.

Bott, R., Ultsch, M., Kossiakoff, A., Graycar, T., Katz, B., and Power, S. (1988). The three-dimensional structure of *Bacillus amyloliquefaciens* subtilisin at 1.8 A and an analysis of the structural consequences of peroxide inactivation. **J. Biol. Chem.** 263:(16) 7895-7906.

Briedigkeit, L., and Frommel, C. (1989). Calcium ion binding by thermitase. **FEB.** 253:(1,2) 83-87.

Burnett, T. J., Shankweiler, G. W., and Hageman, J. H. (1986). Activation of intracellular serine proteinase in *Bacillus subtilis* cells during sporulation. **J. Bacteriol.** 165:(1) 139-145.

Bursy, J., Pierik, A. J., Pica, N., and Bremer, E. (2007). Osmotically induced synthesis of the compatible solute hydroxyectoine is mediated by an evolutionarily conserved ectoine hydroxylase. **J. Biol. Chem.** 282:(43) 31147-31155.

Canas, B., Lopez-Ferrer, D., Ramos-Fernandez, A., Camafeita, E., and Calvo, E. (2006). Mass spectrometry technologies for proteomics Brief. Funct. Genomic. **Proteomics.** 4: 295-320.

Chaiyanan, S. (2000). **Selection of bacterial strains for improvement of fish sauce fermentation.** PhD. thesis, University of Maryland Biotechnology Institute, Maryland.

Chaiyanan, S., Maugel, T., Huq, A., Robb, F. T., and Colwell, R. R. (1999). Polyphasic taxonomy of a novel halobacillus, *Halobacillus thailandensis* sp. nov. isolated from fish sauce. **System. Appl. Microbiol.** 22: 360-365.

Chamroensaksri, N., Akaracharanya, A., Visessanguan, W., and Tanasupawat, S. (2008). Characterization of halophilic bacterium NB2-1 from Pla-Ra and its protease production. **J. Food Biochem.** 32: 536-555.

Chen, Y. G., Cui, X. L., Fritze, D., Chai, L. H., Schumann, P., Wen, M. L., et al. (2008). *Virgibacillus kekensis* sp. nov., a moderately halophilic bacterium isolated from a salt lake in China. **Int. J. Syst. Evol. Microbiol.** 58:(Pt 3) 647-653.

Chen, Y.-G., Liu, Z.-X., Peng, D.-J., Zhang, Y.-Q., Wang, Y.-X., Tang, S.-K., et al. (2009). *Virgibacillus litoralis* sp. nov., a moderately halophilic bacterium isolated from saline soil. **Anton. Leeuw.** 96, 323-329.

Cleary, P. P., and Matsuka, Y. V. (2004). C5a peptidase. In A. J. Barrett, N. D. Rawlings, and J. F. Woessner (eds.). **Handbook of Proteolytic Enzymes** (pp. 1808-1810). London: Elsevier.

Danson, M. J., and Hough, D. W. (1997). The structural basis of protein halophilicity. **Comp. Biochem. Physiol.** 117.

Davail, S., Feller, G., Narinx, E., and Garday, C. (1994). Clod adaptation of proteins. **J. Biol. Chem.** 268: 21826-21834.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989). Electrospray ionisation for mass spectrometry of large biomolecules. **Science.** 246: 64-71.

Frolow, F., Harel, M., Sussman, J. L., Mevarech, M., and Shoham, M. (1996). Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe-2S ferredoxin. **Nat. Struct. Biol. 3**:(5) 452-458.

Fujita, Y. (2009). Carbon catabolite control of the metabolic network in *Bacillus subtilis*. **Biosci Biotechnol Biochem.** 73:(2) 245-259.

Galinski, E. A. (1995). Osmoadaptation in bacteria. **Adv. Microb. Physiol.** 37: 272-328.

Garabito, M. J., Arahal, D. R., Mellado, E., Marquez, M. C., and Ventosa, A. (1997). *Bacillus salexigens* sp. nov., a new moderately halophilic *Bacillus* species. **Int. J. Syst. Bacteriol.** 47:(3) 735-741.

Garcia-Carreno, F. L., Dimes, L. E., and Haard, N. F. (1993). Substrate-gel electrophoresis for composition and molecular weight of proteinases or proteinaceous proteinase inhibitors. **Anal. Biochem.** 214:(1) 65-69.

Gimenez, M. I., Studdert, C. A., Sanchez, J. J., and De Castro, R. E. (2000). Extracellular protease of *Natrialba magadii*: purification and biochemical characterization. **Extremophiles.** 4:(3) 181-188.

Graycer, T. P., Ballinger, M. D., and Wells, J. A. (2004). Subtilisins. In A. J. Barrett, N. D. Rawlings, and J. F.Woessner (eds.) **Handbook of Proteolytic Enzymes** (pp. 1786-1792). London: Elsevier.

Gron, H., Meldal, M., and Breddam, K. (1992). Extensive comparison of the substrate preferences of two subtilisins as determined with peptide substrates which are based on the principle of intramolecular quenching. **Biochemistry.** 31:(26) 6011-6018.

Gros, P., Betzel, C., Dauter, Z., Wilson, K. S., and Hol, W. G. (1989). Molecular dynamics refinement of a thermitase-eglin-c complex at 1.98 A resolution and comparison of two crystal forms that differ in calcium content. **J. Mol. Biol.** 210:(2) 347-367.

Gros, P., Kalk, K. H., and Hol, W. G. (1991). Calcium binding to thermitase. Crystallographic studies of thermitase at 0, 5, and 100 mM calcium. **J. Biol. Chem.** 266:(5) 2953-2961.

Gupta, A., Joseph, B., Mani, A., and Thomas, G. (2008). Biosynthesis and properties of an extracellular thermostable serine alkaline proteinase from *Virgibacillus pantothenticus*. **World. J. Microbiol. Biotechnol.** 24: 237-243.

Haard, N. F. (1994). Protein hydrolysis in seafoods. In F. Shahidiand, and J. R. Botta (eds.). **Seafoods chemistry, processing technology, and quality** (pp. 10–26). London: Chapman and Hall.

Haard, N. F., and Simpson, B. K. (1994). Proteases from aquatic organisms and their uses in the seafood industry. In Martin, A. M. (ed.). **Fisheries processing: biotechnological applications** (pp. 132-154). London: Chapman and Hall.

Hall, N. (2007). Advanced sequencing technologies and their wider impact in microbiology. **J. Exp. Biol.** 210:(Pt 9) 1518-1525.

Heyndrickx, M., Lebbe, L., Kersters, K., De Vos, P., Forsyth, G., and Logan, N. A. (1998). *Virgibacillus*: a new genus to accommodate *Bacillus pantothenticus* (Proom and Knight 1950). Emended description of *Virgibacillus* pantothenticus. **Int. J. Syst. Bacteriol.** 48: 99-106.

Heyndrickx, M., Lebbe, L., Kersters, K., Hoste, B., De Wachtert, R., De Vos, P., et al. (1999). Proposal of *Virgibacillus proomii* sp. nov. and emended description of *Virgibacillus pantothenticus* (Proom and Knight 1950) Heydrickx et al. 1998. **Int. J. Syst. Bacteriol.** 49: 1083-1090.

Heyrman, J., Logan, N. A., Busse, H. J., Balcaen, A., Lebbe, L., Rodriguez-Diaz, M., et al. (2003). *Virgibacillus carmonensis* sp. nov., *Virgibacillus necropolis* sp. nov. and *Virgibacillus picturae* sp. nov., three novel species isolated from deteriorated mural paintings, transfer of the species of the genus salibacillus to *Virgibacillus*, as *Virgibacillus marismortui* comb. nov. and *Virgibacillus salexigens* comb. nov., and emended description of the genus *Virgibacillus*. **Int. J. Syst. Evol. Microbiol.** 53:(Pt 2) 501-511.

Hillenkamp, F., Karas, M., Beavis, R. C., and Chait, B. T. (1991). Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. **Anal. Chem.** 63:(24) 1193A-1203A.

Hiraga, K. Y. N., Namwong, S., Tanasupawat, S., Takada, K., and Oda, K. (2005). Purification and characterization of serine proteinase from a halophilic bacterium, *Filobacillus* sp. RF2-5. **Biosci. Biotechnol. Biochem.** 69:(1) 38-44.

Hodgson, J. (1994). The changing bulk catalysis market: recombinant DNA techniques have changed bulk enzyme production dramatically. **Bio/Technology.** 12: 789-790.

Hou, W. C., Chen, H. J., Chen, T. E., and Lin, Y. H. (1999). Detection of protease activities using specific aminoacyl or peptidyl p-nitroanilides after sodium dodecyl sulfate - polyacrylamide gel electrophoresis and its applications. **Electrophoresis.** 20:(3) 486-490.

Hua, N. P., Hamza-Chaffai, A., Vreeland, R. H., Isoda, H., and Naganuma, T. (2008). *Virgibacillus salarius* sp. nov., a halophilic bacterium isolated from a Saharan salt lake. **Int. J. Syst. Evol. Microbiol.** 58:(Pt 10) 2409-2414.

Hurd, P. J., and Nelson, C. J. (2009). Advantages of next-generation sequencing versus the microarray in epigenetic research. **Brief. Funct. Genomic. Proteomic.** 8:(3) 174-183.

Inouye, K., Kuzuya, K., and Tonomura, B. (1998). Sodium chloride enhances markedly the thermal stability of thermolysin as well as its catalytic activity. **Biochim. Biophys. Acta.** 1388:(1) 209-214.

Izotova, L. S., Strongin, A. Y., Chekulaeva, L. N., Sterkin, V. E., Ostoslavskaya, V. I., Lyublinskaya, L. A., et al. (1983). Purification and properties of serine protease from *Halobacterium halobium*. **J. Bacteriol.** 155:(2) 826-830.

Johnson, K. G., Lanthier, P. H., and Gochnauer, M. B. (1986). Studies of two strains of *Actinopolyspora halophila*, an extremely halophilic actinomycete. **Arch. Microbiol.** 143: 370-378.

Jurinke, C., van den Boom, D., Cantor, C. R., and Koster, H. (2002). The use of MassARRAY technology for high throughput genotyping. **Adv. Biochem. Eng. Biotechnol.** 77: 57-74.

Kamekura, M., and Seno, Y. (1990). A halophilic extracellular protease from a halophilic archaebacterium strain 172p1. **Biochem. Cell. Biol.** 68: 352-359.

Karas, M., Gluckmann, M., and Schafer, J. (2000). Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. **J Mass Spectrom.** 35:(1) 1-12.

Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. **Anal. Chem.** 60:(20) 2299-2301.

Kelleher, P. J., and Juliano, R. L. (1984). Detection of proteases in polyacrylamide gels containing covalently bound substrates. **Anal. Biochem.** 136:(2) 470-475.

Kempf, B. and Bremer, E. (1998). Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. **Arch. Microbiol.** 170:(5) 319-330.

Kim, M., Nishiyama, Y., Mura, K., Tokue, C., and Arai, S. (2004). Gene cloning and characterization of a *Bacillus vietnamensis* Metalloprotease. **Biosci. Biotechnol. Biochem.** 68:(7) 1533-1540.

Kim, W. J., and Kim, S. M. (2005). Purification and characterization of *Bacillus subtilis* JM-3 protease from anchovy sauce. **J. Food Biochem.** 29: 591-610.

Kleiner, D. E. and Stetler-Stevenson, W. G. (1994). Quantitative zymography: detection of picogram quantities of gelatinases. **Anal. Biochem.** 218:(2) 325-329.

Kobayashi, T., Hakamada, Y., Adachi, S., Hitomi, J., Yoshimatsu, T., Koike, K., et al. (1995). Purification and properties of an alkaline protease from alkalophilic *Bacillus* sp. KSM-K16. **Appl. Microbiol. Biotechnol.** 43:(3) 473-481.

Kobayashi, T., Lu, J., Li, Z., Hung, V. S., Kurata, A., Hatada, Y., et al. (2007). Extremely high alkaline protease from a deep-subsurface bacterium, *Alkaliphilus transvaalensis*. **Appl. Microbiol. Biotechnol.** 75:(1) 71-80.

Koide, Y., Nakamura, A., Uozumi, T., and Beppu, T. (1986). Cloning and sequencing of the major intracellular serine protease gene of *Bacillus subtilis*. **J. Bacteriol.** 167:(1) 110-116.

Kortt, A. A., and Stewart, D. J. (2004). Dichelobacter (sheep foot-rot) basic serine proteinase. In A. J. Barrett, N. D. Rawlings, and J. F. Woessner (eds.). **Handbook of Proteolytic Enzymes** (pp. 1811-1813). London: Elsevier.

Kucerova, H., Hlavacek, O., Vachova, L., Mlichova, S., and Chaloupka, J. (2001). Differences in the regulation of the intracellular Ca^{2+}-dependent serine proteinase activity between *Bacillus subtilis* and *B. megaterium*. **Curr. Microbiol.** 42:(3) 178-183.

Kuhlmann, A. U., and Bremer, E. (2002). Osmotically regulated synthesis of the compatible solute ectoine in *Bacillus pasteurii* and related *Bacillus* spp. **Appl. Environ. Microbiol.** 68:(2) 772-783.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., et al. (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. **Nature.** 390:(6657) 249-256.

Kushner, D. J. (1978). Life in high salt and solute concentrations: halophilic bacteria. In D. J. Kushner (ed.). **Microbial life in extreme environments** (pp. 317-368). London: Academic Press, Ltd.

Kwon, S. T., Terada, I., Matsuzawa, H., and Ohta, T. (1988). Nucleotide sequence of the gene for aqualysin I (a thermophilic alkaline serine protease) of *Thermus aquaticus* YT-1 and characteristics of the deduced primary structure of the enzyme. **Eur. J. Biochem.** 173:(3) 491-497.

Lantz, M. S., and Ciborowski, P. (1994). Zymographic techniques for detection and characterization of microbial proteases. **Methods Enzymol.** 235: 563-594.

Lanyi, J. K. (1974). Salt-dependent properties of proteins from extremely halophilic bacteria. **Bacteriol. Rev.** 38:(3) 272-290.

Lee, A. Y., Goo Park, S., Kho, C. W., Young Park, S., Cho, S., Lee, S. C., et al. (2004). Identification of the degradome of Isp-1, a major intracellular serine protease of *Bacillus subtilis*, by two-dimensional gel electrophoresis and matrix- assisted laser desorption/ionization-time of flight analysis. **Proteomics.** 4:(11) 3437-3445.

Lee, G.-D., Chun, S.-S., Kho, Y.-H., and Chun, H.-K. (1998). Purification and properties of an extracellular leucine aminopeptidase from the *Bacillus* sp. N2. **J. Appl. Microbiol.** 83:(3) 561-566.

Lee, J. S., Lim, J. M., Lee, K. C., Lee, J. C., Park, Y. H., and Kim, C. J. (2006). *Virgibacillus koreensis* sp. nov., a novel bacterium from a salt field, and

transfer of *Virgibacillus picturae* to the genus *Oceanobacillus* as *Oceanobacillus picturae* comb. nov. with emended descriptions. **Int. J. Syst. Evol. Microbiol.** 56:(Pt 1) 251-257.

Lilley, G. G., Stewart, D. J., and Kortt, A. A. (1992). Amino acid and DNA sequences of an extracellular basic protease of *Dichelobacter nodosus* show that it is a member of the subtilisin family of proteases. **Eur. J. Biochem.** 210:(1) 13-21.

Lopetcharat, K., and Park, J. W. (2002). Characteristics of fish sauce made from pacific whiting and surimi by-products during fermentation stage. **J. Food Sci.** 67: 511-516.

Maciver, B., McHale, R. H., Saul, D. J., and Bergquist, P. L. (1994). Cloning and sequencing of a serine proteinase gene from a thermophilic *Bacillus* species and its expression in *Escherichia coli*. **Appl. Environ. Microbiol.** 60:(11) 3981-3988.

Madabhushi, R. S. (1998). Separation of 4-color DNA sequencing extension products in noncovalently coated capillaries using low viscosity polymer solutions. **Electrophoresis.** 19:(2) 224-230.

Madern, D., Pfister, C., and Zaccai, G. (1995). Mutation at a single acidic amino acid enhances the halophilic behaviour of malate dehydrogenase from Haloarcula marismortui in physiological salts. **Eur. J. Biochem.** 230:(3) 1088-1095.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. **Nature.** 437:(7057) 376-380.

Markowitz, V. M., Chen, I. M., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., et al. (2010). The integrated microbial genomes system: an expanding

comparative analysis resource. **Nucleic Acids Res.** 38(Database issue), D382-390.

Maskow, T., and Babel, W. (2001). Calorimetrically obtained information about the efficiency of ectoine synthesis from glucose in *Halomonas elongata*. **Biochim. Biophys. Acta.** 1527:(1-2) 4-10.

Miyaji, T., Otta, Y., Nakagawa, T., Watanabe, T., Niimura, Y., and Tomizuka, N. (2006). Purification and molecular characterization of subtilisin-like alkaline protease BPP-A from *Bacillus pumilus* strain MS-1. **Lett. Appl. Microbiol.** 42:(3) 242-247.

Molle, V., Nakaura, Y., Shivers, R. P., Yamaguchi, H., Losick, R., Fujita, Y., et al. (2003). Additional targets of the *Bacillus subtilis* global regulator CodY identified by chromatin immunoprecipitation and genome-wide transcript analysis. **J. Bacteriol.** 185:(6) 1911-1922.

Mongodin, E. F., Nelson, K. E., Daugherty, S., Deboy, R. T., Wister, J., Khouri, H., et al. (2005). The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. **Proc. Natl. Acad. Sci. U S A**.102:(50) 18147-18152.

Nagano, H., and To, K. A. (2000). Purification of collagenase and specificity of its related enzyme from *Bacillus subtilis* FS-2. **Biosci. Biotechnol. Biochem.** 64: 181-183.

Namwong, S., Hiraga, K., Takada, K., Tsunemi, M., Tanasupawat, S., and Oda, K. (2006). A halophilic serine proteinase from *Halobacillus* sp. SR5-3 isolated from fish sauce: purification and characterization. **Biosci. Biotechnol. Biochem.** 70:(6) 1395-1401.

Namwong, S., Tanasupawat, S., Visessanguan, W., Kudo, T., and Itoh, T. (2007). *Halococcus thailandensis* sp. nov., from fish sauce in Thailand. **Int. J. Syst. Evol. Microbiol.** 57: 2199–2203.

Nawong, S. (2006). **Isolation, selection, and identification of proteinase-producing bacteria from fish sauce fermentation to be used as starter cultures.** M.S. thesis, Suranaree University of Technology, Nakhon-Ratchasima.

Ng, W. V., Kennedy, S. P., Mahairas, G. G., Berquist, B., Pan, M., Shukla, H. D., et al. (2000). Genome sequence of *Halobacterium* species NRC-1. **Proc. Natl. Acad. Sci. U S A.** 97:(22) 12176-12181.

Niederberger, T. D., Steven, B., Charvet, S., Barbier, B., and Whytent, L. G. (2009). *Virgibacillus arcticus* sp. nov., a moderately halophilic, endospore-forming bacterium from permafrost in the Canadian high Arctic. **J. Syst. Evol. Microbiol.** 59: 2219-2225.

Noda, M., Vo-Van, T., Kusakabe, I., and Murakami, K. (1982). Substrate specificity and salt inhibition of five proteinases isolated from the pyloric caeca and stomach of sardine. **Agric. Biol. Chem.** 46:(1565-1569).

Nonaka, T., Fujihashi, M., Kita, A., Saeki, K., Ito, S., and Miki, K. (2001). Crystallization and preliminary X-ray diffraction studies of a novel alkaline serine protease (KP-43) from alkaliphilic *Bacillus* sp. strain KSM-KP43. **Acta. Crystallogr. D Biol. Crystallogr.** 57:(Pt 5) 717-718.

Norberg, P., and von Hofsten, B. (1969). Proteolytic enzymes from extremely halophilic bacteria. **J. Gen. Microbiol.** 55:(2) 251-256.

Nyren, P., Pettersson, B., and Uhlen, M. (1993). Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. **Anal. Biochem.** 208:(1) 171-175.

O'Farrel, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. **J. Biol. Chem.** 250: 4007-4021.

Ogawa, A., Sumitomo, N., Okuda, M., Saeki, K., Kawai, S., Kobayashi, T., et al. (2003). Nucleotide and deduced amino acid sequences of a high-molecular-mass subtilisin from an alkaliphilic *Bacillus* isolate. **Biochim. Biophys. Acta.** 1624:(1-3) 109-114.

Okuda, M., Sumitomo, N., Takimura, Y., Ogawa, A., Saeki, K., Kawai, S., et al. (2004). A new subtilisin family: nucleotide and deduced amino acid sequences of new high-molecular-mass alkaline proteases from *Bacillus* spp. **Extremophiles.** 8:(3) 229-235.

Olsen, J. V., Ong, S. E., and Mann, M. (2004). Trypsin cleaves exclusively C-terminal to arginine and lysine residues. **Mol. Cell. Proteomics.** 3:(6) 608-614.

Ong, S. E., and Pandey, A. (2001). An evaluation of the use of two-dimensional gel electrophoresis in proteomics. **Biomol. Eng.** 18:(5) 195-205.

Oren, A. (1986). Intracellular salt concentrations of the anaerobic halophilic eubacteria *Haloanaerobium praevalens* and *Halobacteroides halobius*. **Can. J. Microbiol.** 32: 4-9.

Oren, A. (1999). Bioenergetic aspects of halophilism. **Microbiol. Mol. Biol. Rev.** 63:(2) 334-348.

Oren, A. (2006). The order *Haloanaerobiales*. In M. Dworkin, S. Falkow, E. Rosenberg, K. -H. Schleifer, and E. Stackebrandt (eds.). **The Prokaryotes A**

Handbook on the Biology of Bacteria (3 ed., Vol. 4, pp. 804-817). New York: Springer.

Oren, A. (2008). Microbial life at high salt concentrations: phylogenetic and metabolic diversity. **Saline Systems.** 4: 2.

Owens, J. D., and Mendoza, L. S. (1985). Enzymically hydrolysed and bacterially fermented fishery products. **Int. J. Food Sci. Tech.** 20:(3) 273-293.

Pan, J., Huang, Q., and Zhang, Y. (2004). Gene cloning and expression of an alkaline serine protease with dehairing function from *Bacillus pumilus*. **Curr. Microbiol.** 49:(3) 165-169.

Pappin, D. J., Hojrup, P., and Bleasby, A. J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. **Curr. Biol.** 3:(6) 327-332.

Park, J. N., Fukumoto, Y., Fujita, E., Tanaka, T., Washio, T., Otsuka, S., et al. (2001). Chemical composition of fish sauces produced in Southeast and East Asian countries. **J. Food. Comp. Anal.** 14: 113-125.

Paul, S., Bag, S. K., Das, S., Harvill, E. T., and Dutta, C. (2008). Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. **Genome Biol.** 9:(4) R70.

Peng, Q. Z., Chen, J., Zhang, Y. Q., Chen, Q. H., Peng, D. J., Cui, X. L., et al. (2009). *Virgibacillus zhanjiangensis* sp. nov., a marine bacterium isolated from sea water. **Anton. Leeuw.** 96: 645-652.

Piggot, P. J., and Coote, J. G. (1976). Genetic aspects of bacterial endospore formation. **Bacteriol. Rev.** 40:(4) 908-962.

Premkumar, L., Greenblatt, H. M., Bageshwar, U. K., Savchenko, T., Gokhman, I., Sussman, J. L., et al. (2005). Three-dimensional structure of a halotolerant

algal carbonic anhydrase predicts halotolerance of a mammalian homolog. **Proc. Natl. Acad. Sci. U S A.** 102:(21) 7493-7498.

Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., et al. (1987). A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. **Science.** 238:(4825) 336-341.

Quesada, T., Aguilera, M., Morillo, J. A., Ramos-Cormenzana, A., and Monteoliva-Sanchez, M. (2007). *Virgibacillus olivae* sp. nov., isolated from waste wash-water from processing of Spanish-style green olives. **Int. J. Syst. Evol. Microbiol.** 57:(Pt 5) 906-910.

Rao, M. B., Tanksale, A. M., Ghatge, M. S., and Deshpande, V. V. (1998). Molecular and biotechnological aspects of microbial proteases. **Microbiol. Mol. Biol. Rev.** 62:(3) 597-635.

Rawlings, N. D., Morton, F. R., Kok, C. Y., Kong, J., and Barrett, A. J. (2008). MEROPS: the peptidase database. **Nucleic Acids Res.** 36:(Database issue) 320-325.

Reddy, C. S., Jayalaksmi, S. K., and Sreeramulu, K. (2003). Protease from *Halobacterium salinarum*. **J. Ind. Microbiol. Biotechnol.** 43: 253-256.

Roach, J. C., Boysen, C., Wang, K., and Hood, L. (1995). Pairwise end sequencing: a unified approach to genomic mapping and sequencing. **Genomics.** 26:(2) 345-353.

Roberts, M. F. (2005). Organic compatible solutes of halotolerant and halophilic microorganisms. **Saline Systems.** 1: 5.

Rohban, R., Amoozegar, M. A., and Ventosa, A. (2009). Screening and isolation of halophilic bacteria producing extracellular hydrolyses from Howz Soltan Lake, Iran. **J. Ind. Microbiol. Biotechnol.** 36:(3) 333-340.

Saeki, K., Hitomi, J., Okuda, M., Hatada, Y., Kageyama, Y., Takaiwa, M., et al. (2002). A novel species of alkaliphilic *Bacillus* that produces an oxidatively stable alkaline serine protease. **Extremophiles.** 6:(1) 65-72.

Saeki, K., Magallones, M. V., Takimura, Y., Hatada, Y., Kobayashi, T., Kawai, S., et al. (2003). Nucleotide and deduced amino acid sequences of a new subtilisin from an alkaliphilic *Bacillus* isolate. **Curr. Microbiol.** 47:(4) 337-340.

Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H., et al. (2000). Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic *Bacillus* spp.: enzymatic properties, sequences, and evolutionary relationships. **Biochem. Biophys. Res. Commun.** 279:(2) 313-319.

Saeki, K., Ozaki, K., Kobayashi, T., and Ito, S. (2007). Detergent alkaline proteases: enzymatic properties, genes, and crystal structures. **J. Biosci. Bioeng.** 103:(6) 501-508.

Sahl, H. G., Jack, R. W., and Bierbaum, G. (1995). Biosynthesis and biological activities of lantibiotics with unique post-translational modifications. **Eur. J. Biochem.** 230:(3) 827-853.

Saisithi, P., Kasemsarn, B.-O., Liston, J., and Dollar, A. M. (1996). Microbiology and chemistry of fermented fish. **J. Food Sci.** 31:(1) 105-110.

Sakai, J., Rawson, R. B., Espenshade, P. J., Cheng, D., Seegmiller, A. C., Goldstein, J. L., et al. (1998). Molecular identification of the sterol-regulated luminal

protease that cleaves SREBPs and controls lipid composition of animal cells. **Mol. Cell.** 2:(4) 505-514.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. **Proc. Natl. Acad. Sci. U S A.** 74:(12) 5463-5467.

Sareen, R., Bornscheuer, U. T., and Mishra, P. (2005). Cloning, functional expression and characterization of an alkaline protease from *Bacillus licheniformis*. **Biotech. Lett.** 27: 1901-1907.

Saum, S. H., and Muller, V. (2007). Salinity-dependent switching of osmolyte strategies in a moderately halophilic bacterium: glutamate induces proline biosynthesis in *Halobacillus halophilus*. **J. Bacteriol.** 189:(19) 6968-6975.

Schmitt, W., Rdest, U., and Goebel, W. (1990). Efficient high performance liquid chromatographic system for the purification of Halobacterial serine protease. **J. Chromatogr.** 521: 211-220.

Sellek, G. A., and Chaudhuri. (1999). Biocatalysis in organic media using enzymes from extremophiles. **Enzyme Microb. Tech.** 25: 471-482.

Setyorini, E., Kim, Y. J., Takenaka, S., Murakami, S., and Aoki, K. (2006a). Purification and characterization of a halotolerant intracellular protease from *Bacillus subtilis* strain FP-133. **J. Basic Microbiol.** 46:(4) 294-304.

Setyorini, E., Takenaka, S., Murakami, S., and Aoki, K. (2006b). Purification and characterization of two novel halotolerant extracellular proteases from *Bacillus subtilis* strain FP-133. **Biosci. Biotechnol. Biochem.** 70:(2) 433-440.

Sheehan, S. M., and Switzer, R. L. (1990). Intracellular serine protease 1 of *Bacillus subtilis* is formed in vivo as an unprocessed, active protease in stationary cells. **J. Bacteriol.** 172:(1) 473-476.

Sheehan, S. M., and Switzer, R. L. (1991). Intracellular serine protease-4, a new intracellular serine protease activity from *Bacillus subtilis*. **Arch. Microbiol.** 156:(3) 186-191.

Shendure, J., Porreca, G. J., Reppas, N. B., Lin, X., McCutcheon, J. P., Rosenbaum, A. M., et al. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. **Science.** 309:(5741) 1728-1732.

Shi, W., Tang, X. F., Huang, Y., Gan, F., Tang, B., and Shen, P. (2006). An extracellular halophilic protease SptA from a halophilic archaeon *Natrinema* sp. J7: gene cloning, expression and characterization. **Extremophiles.** 10:(6) 599-606.

Shimogaki, H., Takeuchi, K., Nishino, T., Ohdera, M., Kudo, T., Ohba, K., et al. (1991). Purification and properties of a novel surface-active agent- and alkaline-resistant protease from *Bacillus* sp. Y. **Agric. Biol. Chem.** 55:(9) 2251-2258.

Siezen, R. J., de Vos, W. M., Leunissen, J. A. M., and Dijkstra, B. W. (1991). Homology modeling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteases. **Protein Eng.** 4: 719-739.

Siezen, R. J., and Leunissen, J. A. (1997). Subtilases: the superfamily of subtilisin-like serine proteases. **Protein Sci.** 6:(3) 501-523.

Simpson, R. J. (2003). **Proteins and proteomics: a laboratory manual**. New York: Cold Spring Harbor Laboratory Press.

Sinsuwan, S., Rodtong, S., and Yongsawatdigul, J. (2007). NaCl-activated extracellular proteinase from *Virgibacillus* sp. SK37 isolated from fish sauce fermentation. **J. Food Sci.** 72:(5) 264-269.

Siringan, P., Raksakulthai, N., and Yongsawatdigul, J. (2006a). Autolytic activity and biochemical characteristics of endogenous proteinases in Indian anchovy (*Stolephorus indicus*). **Food Chem.** 98: 678-684.

Siringan, P., Raksakulthai, N., and Yongsawatdigul, J. (2006b). Source and changes of proteinase activities of Indian anchovy (*Stolephorus indicus*) during fish sauce fermentation. **J. Sci. Food Agric.** 86: 1070-1976.

Sloma, A., Rufo, G. A., Jr., Theriault, K. A., Dwyer, M., Wilson, S. W., and Pero, J. (1991). Cloning and characterization of the gene for an additional extracellular serine protease of *Bacillus subtilis*. **J. Bacteriol.** 173:(21) 6889-6895.

Sonenshein, A. L. (2007). Control of key metabolic intersections in *Bacillus subtilis*. **Nat. Rev. Microbiol.** 5:(12) 917-927.

Stauffer, C. E., and Etson, D. (1969). The effect on subtilisin activity of oxidizing a methionine residue. **J. Biol. Chem.** 244:(19) 5333-5338.

Stepanov, V. M., Rudenskaya, G. N., Revina, L. P., Gryaznova, Y. B., Lysogorskaya, E. N., Filippova, I., et al. (1992). A serine proteinase of an archaebacterium, *Halobacterium mediterranei*. A homologue of eubacterial subtilisins. **Biochem. J.** 285:(Pt 1) 281-286.

Stepanov, V. M., Strongin, A. Y., Izotova, L. S., Abramov, Z. T., Lyublinskaya, L. A., Ermakova, L. M., et al. (1977). Intracellular serine protease from *Bacillus subtilis*. Structural comparison with extracellular serine proteases-subtilisins. **Biochem. Biophysic. Res. Comm.** 77: 298-305.

Strongin, A. Y., Izotova, L. S., Abramov, D. I., and Stepanov, V. M. (1978). Two related structural genes coding two homologous serine proteases in *Bacillus subtilis* genome. **Molec. Gen. Genet.** 159: 337-339.

Studdert, C. A., De Castro, R. E., Seitz, K. H., and Sanchez, J. J. (1997). Detection and preliminary characterization of extracellular proteolytic activities of the haloalkaliphilic archaeon *Natronococcus occultus*. **Arch. Microbiol.** 168:(6) 532-535.

Tadeo, X., Lopez-Mendez, B., Trigueros, T., Lain, A., Castano, D., and Millet, O. (2009). Structural basis for the aminoacid composition of proteins from halophilic archea. **PLoS Biol.** 7:(12) e1000257.

Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., et al. (2000). Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. **Nucleic Acids Res.** 28:(21) 4317-4331.

Takami, H., Takaki, Y., and Uchiyama, I. (2002). Genome sequence of *Oceanobacillus iheyensis* isolated from the Iheya Ridge and its unexpected adaptive capabilities to extreme environments. **Nucleic Acids Res.** 30:(18) 3927-3935.

Tanasupawat, S., Chamroensaksri, N., Kudo, T., and Itoh, T. (2010). Identification of moderately halophilic bacteria from Thai fermented fish (pla-ra) and proposal of *Virgibacillus siamensis* sp. nov. **J. Gen. Appl. Microbiol.** 56:(5) 369-379.

Tanasupawat, S., Namwong, S., Kudo, T., and Itoh, T. (2009). Identification of halophilic bacteria from fish sauce (nam-pla) in Thailand. **J. Culture Collections.** 6: 69-75.

Tanasupawat, S., Pakdeeto, A., Namwong, S., Thawai, C., Kudo, T., and Itoh, T. (2006). *Lentibacillus halophilus* sp. nov., from fish sauce in Thailand. **Int. J. Syst Evol. Microbiol.** 56:(Pt 8) 1859-1863.

Tang, X. M., Shen, W., Lakay, F. M., Shao, W. L., Wang, Z. X., Prior, B. A., et al. (2004). Cloning and over-expression of an alkaline protease from *Bacillus licheniformis*. **Biotechnol. Lett.** 26:(12) 975-979.

Tapingkae, W., Tanasupawat, S., Itoh, T., Parkin, K. L., Benjakul, S., Visessanguan, W., et al. (2008). *Natrinema gari* sp. nov., a halophilic archaeon isolated from fish sauce in Thailand. **Int. J. Syst. Evol. Microbiol.** 58:(Pt 10) 2378-2383.

Tawfik, D. S., and Griffiths, A. D. (1998). Man-made cell-like compartments for molecular evolution. **Nat. Biotechnol.** 16:(7) 652-656.

Teng, X., and Xiao, H. (2009). Perspectives of DNA microarry and next-generation DNA sequencing technologies. **Sci. China C. Life Sci.** 52:(1) 7-16.

Teplyakov, A. V., Kuranova, I. P., Harutyunyan, E. H., Vainshtein, B. K., Frommel, C., Hohne, W. E., et al. (1990). Crystal structure of thermitase at 1.4 A resolution. **J. Mol. Biol.** 214:(1) 261-279.

Thongsanit, J., Tanasupawat, S., Keeratipibul, S., and Jatikavanich, S. (2002). Characterization and identification of *Tetragenococcus halophilus* and *Tetragenococcus muriaticus* strains from fish sauce (nam-pla). **Jpn. J. Lactic Acid Bacteria.** 13: 46–52.

Thongthai, C., McGenity, T. J., Suntinanalert, P., and Grant, W. D. (1992). Isolation and characterization of an extremely halophilic archaeobacterium from traditional fermented Thai fish sauce (nam-pla). **Lett. Appl. Microbiol.** 14: 111–114.

Thongthai, C., and Suntinanalert, P. (1991). **Halophiles in Thai fish sauce (Nam Pla)**. New York: Plenum Press.

Tindall, B. J., Ross, H. N. M., and Grant, W. D. (1984). *Natronobacterium* gen. nov. and *Natronococcus* gen. nov. Two new genera of haloalkaliphilic archaebacteria. **Syst. Appl. Microbiol.** 5: 41-57.

To, K. A., Tanaka, T., and Nagano, H. (1997). Isolation of collagenase producing bacteria from traditional foods. **J. Home Econ. Jpn.** 12: 1083-1087.

Toyokawa, Y., Takahara, H., Reungsang, A., Fukuta, M., Hachimine, Y., Tachibana, S., et al. (2010). Purification and characterization of a halotolerant serine proteinase from thermotolerant *Bacillus licheniformis* RKK-04 isolated from Thai fish sauce. **Appl. Microbiol. Biotechnol.** 86:(6) 1867-1875.

Tran, L., Wu, X. C., and Wong, S. L. (1991). Cloning and expression of a novel protease gene encoding an extracellular neutral protease from *Bacillus subtilis*. **J. Bacteriol.** 173:(20) 6364-6372.

Tran, L. H., and Nagano, H. (2002). Isolation and characteristics of *Bacillus subtilis* CN2 and its collagenase production. **J. Food Sci.** 67(3): 1184-1187.

Uchida, H., Kondo, D., Yamashita, S., Tanaka, T., Tran, L. H., Nagano, H., et al. (2004). Purification and properties of a protease produced by *Bacillus subtilis* CN2 isolated from a Vietnamese fish sauce. **J. Microbiol. Biotechnol.** 20: 579-582.

van de Ven, W. J. M. (1993). Structure and function of eukaryotic proprotein processing enzymes of the subtilisin family of serine proteases. **Crit. Rev. Oncogesis.** 4: 115-136.

Ventosa, A., and Nieto, J. J. (1995). Biotechnological applications and potentialities of halophilic microorganisms. **World J. Microbiol. Biotechnol.** 11: 85–94.

Ventosa, A., Nieto, J. J., and Oren, A. (1998). Biology of moderately halophilic aerobic bacteria. **Microbiol. Mol. Biol. Rev.** 62:(2) 504-544.

Voelkerding, K. V., Dames, S. A., and Durtschi, J. D. (2009). Next-generation sequencing: from basic research to diagnostics. **Clin. Chem.** 55:(4) 641-658.

Walker, J. M. (2002). **The protein protocols handbook**. New York: Human Press.

Warburton, M. J., and Bernardini, F. (2002). Tripeptidyl peptidase-I is essential for the degradation of sulphated cholecystokinin-8 (CCK-8S) by mouse brain lysosomes. **Neurosci. Lett.** 331:(2) 99-102.

Wang, C. Y., Chang, C. C., Ng, C. C., Chen, T. W., and Shyu, Y. T. (2008). *Virgibacillus chiguensis* sp. nov., a novel halophilic bacterium isolated from Chigu, a previously commercial saltern located in southern Taiwan. **Int. J. Syst. Evol. Microbiol.** 58:(Pt 2) 341-345.

Weickhardt, C., Moritz, F., and Grotemeyer, J. (1996). Time-of-flight mass spectrometry: state-of-the-art in chemical analysis and molecular science. **Mass Spectrom. Rev.** 15: 139-162.

Whittal, R. M., and Li, L. (1995). High-resolution matrix-assisted laser desorption/ionization in a linear time-of-flight mass spectrometer. **Anal. Chem.** 67:(13) 1950-1954.

Wilaipan, P. (1990). **Halophilic bacteria producing lipase in fish sauce.** M.S. thesis, Chulalongkorn University, Bangkok.

Williams, J., McGrath, W. J., and Mangel, W. F. (2000). Sensitive method to identify and characterize proteinases in situ after SDS-PAGE. **Biotechniques.** 29:(5) 1108-1113.

Wong, S. L., Price, C. W., Goldfarb, D. S., and Doi, R. H. (1984). The subtilisin E gene of *Bacillus subtilis* is transcribed from a sigma 37 promoter in vivo. **Proc. Natl. Acad. Sci. U S A.** 81:(4) 1184-1188.

Wray, L. V., Jr., Ferson, A. E., Rohrer, K., and Fisher, S. H. (1996). TnrA, a transcription factor required for global nitrogen regulation in *Bacillus subtilis*. **Proc. Natl. Acad. Sci. U S A.** 93:(17) 8841-8845.

Wu, J., Bian, Y., Tang, B., Chen, X., Shen, P., and Peng, Z. (2004). Cloning and analysis of WF146 protease, a novel thermophilic subtilisin-like protease with four inserted surface loops. **FEMS Microbiol. Lett.** 230:(2) 251-258.

Yasothornsrikul, S., and Hook, V. Y. (2000). Detection of proteolytic activity by fluorescent zymogram in-gel assays. **Biotechniques.** 28:(6) 1166-1168: 1170, 1172-1163.

Yates, J. R., 3rd, Eng, J. K., McCormack, A. L., and Schieltz, D. (1995). Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. **Anal. Chem.** 67:(8) 1426-1436.

Yoon, J., Oh, T., and Park, Y. (2004). Transfer of *Bacillus halodenitrificans* Denariaz et al. 1989 to the genus *Virgibacillus* as *Virgibacillus halodenitrificans* comb. nov. **Int. J. Syst. Evol. Microbiol.** 54: 2163-2167.

Yoon, J. H., Kang, S. J., Lee, S. Y., Lee, M. H., and Oh, T. K. (2005). *Virgibacillus dokdonensis* sp. nov., isolated from a Korean island, Dokdo, located at the edge of the East Sea in Korea. **Int. J. Syst. Evol. Microbiol.** 55:(Pt 5) 1833-1837.

Yu, X. T. (1991). Protease of haloalkaliphiles. **In Superbugs** (ed.) Horikoshi K. and Grant, W. D. (pp. 76-83). New York: Springer Verlag.

Zabolotskaya, M. V., Demidyuk, I. V., Akimkina, T. V., and Kostrov, S. V. (2004). A novel neutral protease from *Thermoactinomyces* species 27a: sequencing of the gene, purification, and characterization of the enzyme. **Protein J.** 23:(7) 483-492.

Zhao, Z., and Russell, P. J. (2003). Trypsin activity assay in substrate-specific one- and two-dimensional gels: a powerful method to separate and characterize novel proteases in active form in biological samples. **Electrophoresis.** 24:(18) 3284-3288.

CHAPTER III

IDENTIFICATION OF NOVEL HALOTOLERANT BACILLOPEPTIDASE F-LIKE PROTEINASES FROM A MODERATELY HALOPHILIC BACTERIUM, *VIRGIBACILLUS* SP. SK37

3.1 Abstract

Virgibacillus sp. SK37 isolated from Thai fish sauce produced numerous NaCl-activated subtilisin-like proteinases. Our objectives were to purify, characterize and identify these extracellular proteinases. Three major subtilisin-like enzymes including 19, 34 and 44 kDa were partially purified and showed maximum activity at pH 8, 55–60°C, 25–30% NaCl and 70–100 mM $CaCl_2$. Enzymes showed stability at 0–30% NaCl and <20 mM $CaCl_2$ and were completely inhibited by phenylmethanesulphonyl fluoride but not by ethylenediaminetetraacetic acid. The isoelectric points based on 2D electrophoresis of the 19, 34 and 44 kDa proteinases were at 3.6, 5.2 and 3.8, respectively. Peptide mass fingerprint and *de novo* peptide homology analysis of tryptic peptides using MALDI-TOF and LC–MS⁄MS, respectively, suggested that all three enzymes were novel and homologous to bacillopeptidase F. This is the first report on bacillopeptidase F-like proteinases in genus *Virgibacillus* with a distinct halotolerant feature. They showed potential to be a

processing aid for food and biotechnological applications, particularly in high salt condition.

Keywords: bacillopeptidase F, fish sauce, halotolerant enzymes, moderately halophilic bacteria, Na-activated proteinases, subtilisin-like proteinases, *Virgibacillus* sp.

3.2 Introduction

Subtilisins are one of the most valuable industrial enzymes that have been extensively studied in terms of molecular structure, protein engineering and their application (Gupta, Beg, and Lorenz, 2002; Rao, Tanksale, Ghatge, and Deshpande, 1998). They are a member of subtilase superfamily widely distributed in various organisms including, archaea, bacteria, eukaryotes, and viruses (Siezen, de Vos, Leunissen, and Dijkstra, 1991). Gram-positive bacilli are well-known sources for high productivity of subtilisins (Rawlings, Morton, Kok, Kong, and Barrett, 2008). Unlike nonhalophiles, moderately and extremely halophilic bacteria growing optimally at between 5 and 25% NaCl are able to produce proteinases exhibiting optimal activity in high ionic strength environments (Kushner, 1978; Ventosa, Nieto, and Oren, 1998). Extracellular proteinases from moderately halophilic bacteria have been characterized in details, including *Bacillus clausii* I-52 (Joo, Kumar, Park, Paik, and Chang, 2003), *Halobacillus* sp. SR5-3 (Namwong, Hiraga, Takada, Tsunemi, Tanasupawat, and Oda, 2006) and *Bacillus subtilis* FP-133 (Setyorini, Takenaka, Murakami, and Aoki, 2006). *Virgibacillus* is one of the genera of which, to date, 331 isolates have been reported in the GenBank database. This genus is generally isolated from salt lakes, saltern ponds, hypersaline soils, and salted foods. They are Gram-positive, endospore-forming,

catalase- and oxidase-positive, motile, rod-shaped, aerobic bacteria and grow optimally at between 5 and 25% NaCl, at pH 6.0–10.0 (optimum 7.5–8.0) and 10–55°C (optimum 30–40°C) (Carrasco, Ma´rquez, and Ventosa, 2009; Chen et al., 2009; Hua, Hamza, Vreeland, Isoda, and Naganuma, 2008; Peng et al., 2009; Wang, Chang, Chen, and Shyu, 2008; Yoon, Kang, Jung, Lee, Oh, and Oh, 2010). Like other Gram-positive bacilli, *Virgibacillus* showed the ability to produce a great variety of extracellular hydrolytic enzymes. A high production of extracellular proteinases from bacteria in this genus has been reported (Chamroensaksri, Akaracharanya, Visessanguan, and Tanasupawat, 2008; Gupta, Joseph, Mani, and Thomas, 2008; Rohban, Amoozegar, and Ventosa, 2009; Sinsuwan et al., 2007, 2008a, 2008b 2010). However, information regarding biochemical and molecular characteristics as well as the proteomics of proteinases from *Virgibacillus* is still very limited.

Recently, a moderately halophilic bacterium, namely *Virgibacillus* sp. SK37, (GenBank/NCBI no. DQ910840), was isolated from 1-month-old Thai fish sauce mashes. It is able to grow at a wide pH range of4–11, at 20–45°C and 0–25% NaCl. Extracellular proteinases produced from this strain exhibited a high level of subtilisin-like proteolytic activity at high NaCl contents (up to 20% NaCl) (Sinsuwan et al., 2007). This strain has the potential to be developed as a starter culture for fish sauce fermentation because it shortened fish sauce fermentation from 12 to 4 months (Yongsawatdigul, Rodtong, and Raksakulthai, 2007). The Objectives in this study were to identify and characterize subtilisin-like proteinases secreted from *Virgibacillus* sp. SK37. As the genomic data of bacteria in genus *Virgibacillus* are unavailable, we proposed a combination of proteomic approaches with two-dimensional electrophoresis and mass spectroscopy to identify the selected

proteinases. We found major subtilisin-like enzymes showing homologous to bacillopeptidase F that has never been reported in genus *Virgibacillus*. A protein purification scheme for major proteinases was established. Enzyme identification and homology search using peptide mass fingerprints (PMFs) and partial amino acid sequences obtained from matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) and tandem mass spectrometry, respectively, were carried out.

3.3 Materials and methods

3.3.1 Materials

DEAE-Sephacel and Source Q were purchased from Amersham Pharmacia Biotech (GE Healthcare, Uppsala, Sweden). Succinyl-Ala-Ala-Pro-Phe-amino-7-methylcoumarin (Suc-AAPF-AMC) was obtained from Bachem A.G. (Bubendorf, Switzerland). All other reagents were of analytical grade.

3.3.2 Culture medium and conditions

Virgibacillus sp. SK37 was cultured in a modified halophilic medium containing 5% NaCl as described by Sinsuwan et al. (2008a). The cultures were aerobically incubated under constant shaking at 100 rev min^{-1} at 35°C for 2 days providing late exponential growth phase and maximal proteinase activity.

3.3.3 Partial purification of proteinases

Proteinase preparation was carried out at 4°C. The cultured broth was centrifuged at 10,000xg for 15 min. Supernatants were filtered through a 0.45-μm membrane (Whatman, Kent, UK) and divided equally into two portions. Ammonium sulfate was added to the first portion to attain 80% saturation. The precipitate was collected by centrifugation, and resultant precipitate was dissolved in 2 ml of 50 mM

Tris-HCl pH 8.0 and dialysed against the same buffer for 48 h with 12-h intervals of buffer exchange. This fraction was referred to as 'Total extracellular proteinases'.

The second portion was fractionated with 60-80% ammonium sulfate that captured all major enzymes in the crude fraction. The pellet was collected by centrifugation, dissolved in 2 ml of 50 mM Tris-HCl pH 8.0 and dialysed against the same buffer at 4°C for 48 h with 12-h intervals of buffer exchange. Protein purification was carried out using ÄKTA^{TM} FPLC with UNICORN^{TM} software version 3.2 (GE healthcare, Uppsala, Sweden). The dialysate was loaded onto a DEAE-Sephacel column (60 ml) equilibrated with buffer A (50 mM Tris–HCl pH 8.0). The unbound proteins were washed with three column volumes of buffer A. Subsequently, the bound proteins were eluted with a linear gradient of buffer B (1.0 M NaCl in mM Tris-HCl, pH 8.0) from 0 to 100% within five-column volumes at a flow rate of 1.0 ml min^{-1}. Fractions of 5 ml were collected and assayed for proteinase activity using Suc-AAPF-AMC as a substrate. The proteinase purity of each fraction was also evaluated by casein zymography. At this point, an enzyme with a molecular mass of 44 kDa was obtained. Other fractions containing proteinase activity were pooled and dialysed against buffer C (15 mM sodium acetate buffer, pH 5.0) overnight at 4°C and subsequently loaded onto a Source Q column (30 ml) previously equilibrated with buffer C. The unbound proteins were washed with three-column volumes of buffer C. The bound proteins were eluted with a linear gradient of buffer D (1.0 M NaCl in 15 mM sodium acetate buffer, pH 5.0) from 0 to 50% within three column volumes at a constant flow rate of 2.5 ml min^{-1}. All eluates were collected in a 3-ml fraction and analyzed as previously described. Fractions containing enzymes with molecular mass of 19 and 34 kDa were pooled separately. Three partially

purified enzyme fractions including 19-, 34-, and 44-kDa proteinase were dialyzed against buffer A and kept at 4°C until use.

3.3.4 Proteinase assay

Proteinase activity was assayed using Suc-AAPF-AMC as a substrate. The reaction mixture (1 ml) contained 50 μl of enzyme solution, 100 μl of 10 μM Suc-AAPF-AMC and 850 μl of McIlvaine buffer (0.1 M trisodium citrate in 0.2 M trisodium phosphate), pH 8.0. The reaction was incubated at 60°C for 5 min and terminated by adding 1.5 ml of a stopping solution (30% butanol, 35% methanol and 35% deionized water). Fluorescence intensity was measured at excitation and emission wavelength of 380 and 460 nm, respectively, using a spectrofluorometer (RF-1501 Shimadzu, Kyoto, Japan). One unit of activity was defined as 1 nmole of liberated AMC per min. Protein concentration was determined by the Bradford method using bovine serum albumin as a standard.

3.3.5 Electrophoresis and zymographic assay

Nonreducing SDS-PAGE was performed in 4 and 12%T of stacking and separating gel, respectively (Laemmli, 1970). Proteins were detected by either silver staining (Heukeshoven and Dernick, 1985) or colloidal Coomassie Brilliant Blue G250 (Neuhoff, Stamm, and Eibl, 1985). Standard molecular weight markers (Broad range; Bio-Rad, Hercules, CA, USA) were used. Casein zymography was carried as previously described by Sinsuwan et al. (2007). In-gel proteolytic reaction was carried out in McIlvaine buffer pH 8.0 at 60°C for 20 min. After destaining step, clear zones indicating the presence of proteinase activity were detected. For substrate-gel zymography, Suc-AAPF-AMC was used as a substrate. After electrophoresis, the gel was washed with 2.5% Triton X-100 for 1 h at 4°C and immersed in 50 mM Tris–

HCl, pH 8.0, 1 M NaCl, and 0.1 mM Suc-AAPF-AMC at 4°C for 20 min. Subsequently, the gel was incubated at 60°C for 10 min to initiate proteolytic reaction. Fluorescence was detected using a Gel Doc™ XR system (Bio-Rad).

3.3.6 Biochemical characteristics

Proteinase assay for each biochemical characteristic was carried out according to the method described earlier, unless otherwise specified. All values are means of triplicate experiments with %CV lesser than 15%.

3.3.6.1 Optimum pH and pH stability

Proteinase activity of each proteinase was measured at various pH values ranging from 3 to 11 in McIlvaine buffer at 60°C. For pH stability, 25 μl of enzyme solution was added to 175 μl McIlvaine buffer at various pHs and incubated at 30°C for 2 h. Subsequently, all reactions were adjusted to pH 8 ± 0.5 by adding 790 μl Tris buffer saline I (TBS I: 1 M NaCl in 1 M Tris–HCl, pH 8.0). It should be noted that TBS I contains NaCl at the same concentration (1 M) as McIlvain buffer allowing comparable enzyme activity to that in McIlvain buffer. Proteinase activity assay was initiated by adding 10 μl of 100 μM substrate and further incubated at 60°C for 5 min. Remaining activity was calculated by taking reactions prepared under the same condition but without incubation as 100%.

3.3.6.2 Optimum temperature and thermal stability

The effect of temperature on enzyme activity was measured at 25–80°C in McIlvain buffer pH 8.0. To investigate thermal stability, each proteinase was incubated at various temperatures (30, 40, 50, 60, 70 and 80°C) for 2 h in McIlvain buffer pH 8.0. Residual activity was measured at 60°C and compared to a sample without incubation.

3.3.6.3 Effects of NaCl and $CaCl_2$ on enzyme activity and stability

Proteinase activity at various NaCl concentrations (0-30% w⁄v) was measured. Reaction mixtures contained 50 mM Tris–HCl pH 8.0 instead of McIlvain buffer as the latter contained Na^+. The reactions were terminated by adding 1.5 ml of deionized water and immediately incubating at 90°C for 15 min. This was to avoid phase separation of NaCl when a butanol-containing solution was used. The activity assay was carried out at the optimum pH and temperature of each enzyme. Relative activity of each enzyme was defined as the percentage of activity with respect to the maximum proteinase activity detected in the assay.

The effect of $CaCl_2$ on proteinase activity was assayed at 0–100 mM $CaCl_2$ in 50 mM Tris-HCl, pH 8.0. The activity assay was carried out at the optimum pH and temperature of each enzyme and expressed in the same manner as described for NaCl. To investigate the effect of Na^+ and Ca^{2+} on proteinase stability, the purified enzymes were incubated at various concentrations of either NaCl [0–30% (w ⁄ v)] or $CaCl_2$ (0-100 mM) in 50 mM Tris-HCl, pH 8.0 at 30°C for 24 h. When the incubation time was reached, residual activity was measured at the optimum pH and temperature of each enzyme. The relative residual activity was calculated using the activity of the sample without incubation at the respective NaCl or $CaCl_2$ concentrations as 100%.

3.3.6.4 Effect of metal ions on enzyme activity

The effect of metal ions including Cu^{2+}, Cd^{2+}, Mn^{2+}, Hg^{2+}, Zn^{2+} and Fe^{3+} (all in chloride salts except for ferric sulfate) on proteinase activity was investigated. The proteinase activity assay was carried out using Tris buffer saline II (TBS II: 1 M NaCl in 50 mM Tris-HCl, pH 8.0) instead of McIlvaine buffer to avoid the chelating effect of phosphate. Reactions were performed at the optimum pH and temperature of each

partially purified proteinase. The reaction without the ion was taken as 100%.

3.3.6.5 Effects of proteinase inhibitors and denaturants on enzyme activity

The effect of various inhibitors and denaturants on proteinase activity was determined using various substances, including N-tosyl-l-lysine chloromethyl ketone (TLCK), N-tosyl-l-phenylalanine chloromethyl ketone (TPCK), phenylmethane-sulphonyl fluoride (PMSF), ethylenediaminetetraacetic acid (EDTA), l-histidine, bestatin, pepstatin A, iodoacetic acid, N ethylmaleimide, β-mercaptoethanol, SDS, Triton X-100, urea and H_2O_2 in McIlvain buffer pH 8.0. Proteinase activity was assayed at the optimum pH and temperature of each enzyme. Relative activity was calculated by taking the reaction without any substances as 100%.

3.3.7 2D electrophoresis

Eluted fractions from a DEAE-Sephacel column exhibiting high subtilisin-like proteinase activity were pooled and concentrated using ultrafiltration (NANOSEP 10K MWCO; Pall Filtron, Northborough, MA, USA) and subjected to 2D electrophoresis (GE healthcare, Piscataway, NJ, USA) according to the manufacturer's instructions. Twenty micrograms of protein was immobilized in an Immobiline Dry Strip (IPG strip, pH 3–10 l, 7 cm; Amersham) under a constant voltage of 50 V, at 20°C for 12 h. Isoelectric focusing was performed using the Ettan IPGphor II electrophoresis (GE Healthcare). Subsequently, the IPG strip was treated with dithiothreitol (DTT) as per the manufacture's protocol (GE healthcare, USA). The second dimension was performed on a 12.5% SDS-polyacrylamide gel using the miniVE vertical electrophoresis system (GE healthcare, USA). The gel was stained with colloidal Coomassie Brilliant Blue G250 as mentioned previously. Protein spots were excised and subjected to in-gel tryptic digestion for further mass analysis.

3.3.8 Mass spectrometry

MALDI-TOF MS and LC–MS/MS were conducted at the Proteomics Service Center of the Genome Institute, Thailand. Results obtained from MALDI-TOF were further analyzed for PMFs. Protein identifications were analyzed using MASCOT (www.matrixscience.com). The search criteria were fixed by modification carboxy-amidomethylation of cysteine, variable modification methionine oxidation and the accuracy of the experimental to theoretical p*I* and mass. Bacteria (Eubacteria) sequences were limited for matching peptide mass fingerprints in the NCBInr 20080718 database. A protein score >78 is considered to be significant ($P < 0.05$).

As we found that the mass of the obtained peptides did not match any available peptides in the database, amino acid sequences of each peptide were then determined by *de novo* sequencing using Peak software ver. 2.0. The obtained peptide sequences were searched for protein homology using Blastp (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The search strategy was focused only on serine proteinases produced from Gram-positive bacteria, and any peptide queries giving high homology (>60% sequence identity) towards this group of enzymes were selected.

3.4 Results

3.4.1 Determination of proteinase homogeneity

Virgibacillus sp. SK37 secreted proteins with different mass ranging from 19 to 55 kDa on a nonreducing SDS-PAGE (Figure 3.1a). At least 11 proteinases with mass of approximately 19, 22, 24, 28, 34, 36, 40, 43, 44, 45, and 55 kDa were detected on a casein zymogram (Figure 3.1b). Based on protein band intensity on the

SDS-PAGE, the major extracellular proteinases exhibited mass of 19, 34, 43, and 44 kDa. A similar result was obtained when the sample was run in a reducing electrophoresis condition (data not shown). Six UV luminous bands on the synthetic substrate (Suc-AAPF-AMC) zymogram indicated the presence of subtilisin-like proteinases with mass of 19, 28, 34, 36, 44, and 55 kDa (Figure 3.1c). Enzymes with mass of 28, 36, and 55 kDa were considered to be minor as they showed faint protein band intensity on SDS-PAGE (Figure 3.1a).

3.4.2 Proteinase purification

Purification approach was intentionally designed to capture 19-, 34-, and 44-kDa proteinase as they were major proteinases secreted by *Virgibacillus* sp. SK37. The 44-kDa proteinase was eluted at fraction 52–70 using DEAE-Sephacel chromatography (Figure 3.2a). Fractions 37–46 were pooled and subsequently loaded onto a Source Q column. Our preliminary experiments suggested that the use of acidic buffer (15 mM sodium acetate, pH 5.0) with a strong anion-exchanger was able to eliminate other contaminant proteinases, and also effectively fractionated the 19- and 34-kDa proteinase (Figure 3.2b).

High protein purity with a single band of all three enzymes was observed on the silver-stained SDS-PAGE (Figure 3.2c). However, casein zymogram illustrated a very faint proteinase activity at 34 kDa in 19-kDa proteinase fraction, at 24 kDa in 34-kDa proteinase fraction, and at 55 kDa in 44-kDa proteinase fraction. It should be noted that 24-kDa enzyme was unlikely to interfere in the enzyme assay because it did not hydrolyze Suc-AAPF-AMC.

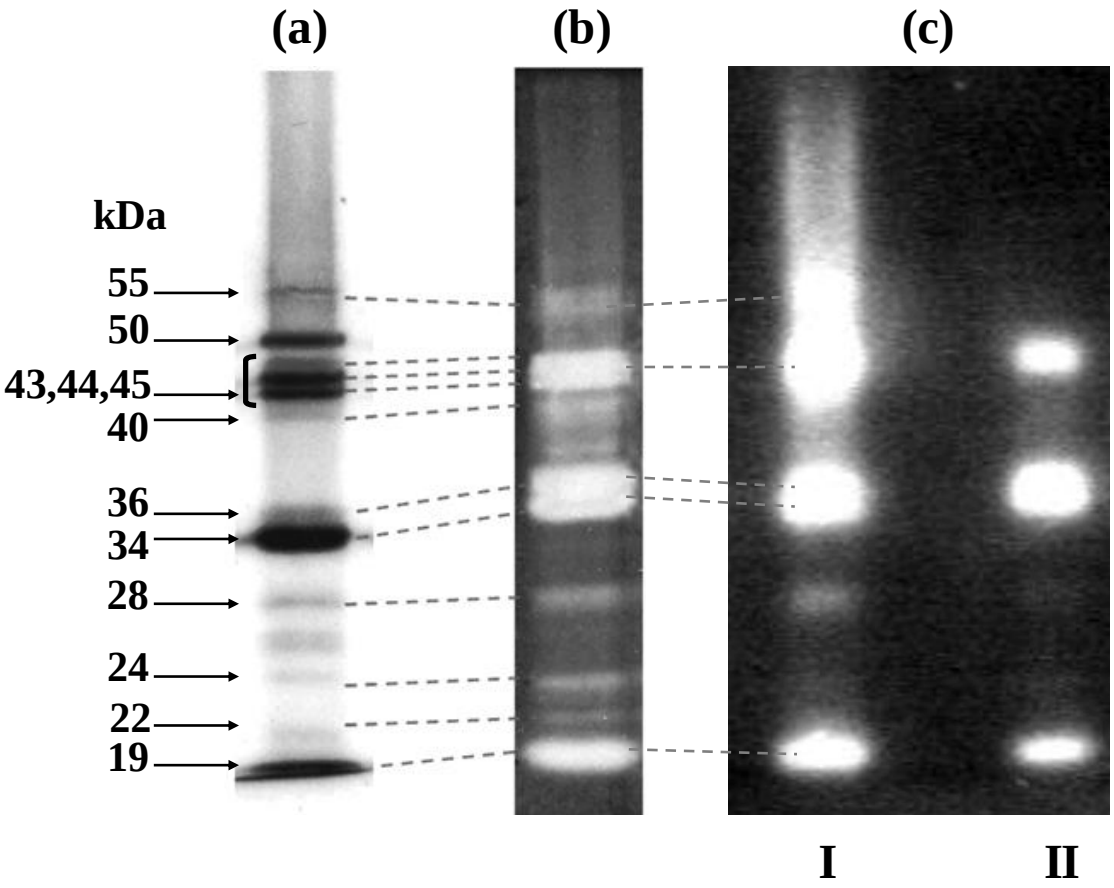


Figure 3.1 Nonreducing SDS-PAGE (12%T) stained with colloidal Coomassie Brilliant Blue (a), casein zymography (b) and substrate (Suc-AAPF-AMC)-gel zymography (c) of total extracellular proteinases fraction. Dashed lines indicate relative positions of protein bands in each gel. In (c), the enzyme was loaded at 0.5 (I) and 0.1 μg (II) and the subtilisin-like proteinases were located as luminous bands under UV light.

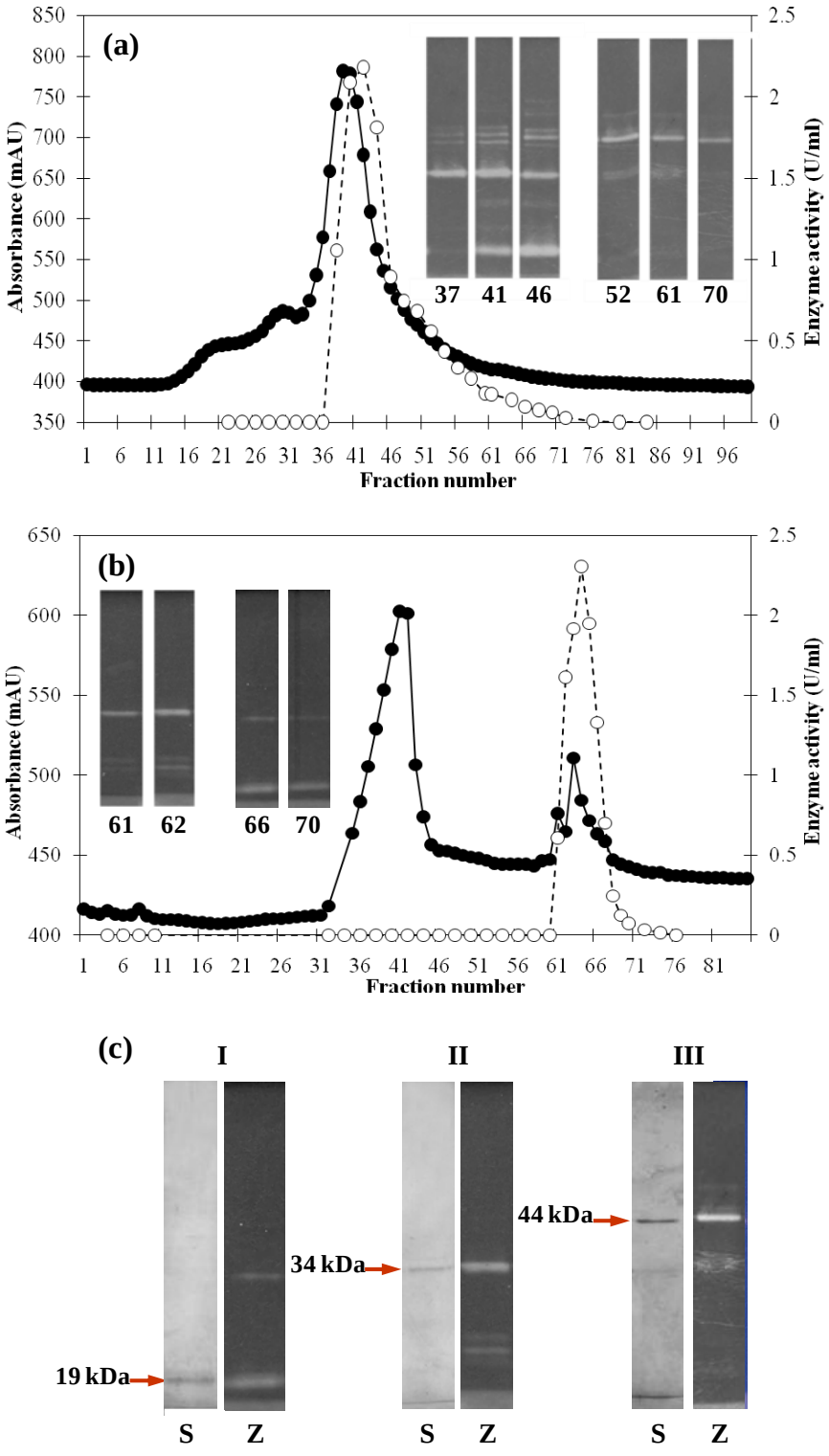


Figure 3.2 Purification of extracellular enzymes in 60–80% ammonium sulfate fraction by DEAE-Sephacel (a) and Source Q (b) column. The protein

(Figure 3.2 Continued)

elution profile (solid lines) and proteinase activity using Suc-AAPF-AMC (dashed lines) and casein zymography (inset panels) of each eluting fraction are shown. Nonreducing SDS-PAGE stained by silver (S) and casein zymography (Z) of partially purified proteinase fractions including 19- (I), 34- (II), and 44-kDa (III) proteinase (c).

3.4.3 Biochemical characteristics of partially purified proteinases

3.4.3.1 pH and temperature optimum and stability

All three enzymes (19, 34, and 44 kDa) showed catalytic activity in a wide pH range at 5–10 with maximal activity at pH 8 (Figure 3.3a). The proteolytic activity was stable (>50% remaining activity) at pH 5.5–11, pH 6.5–10.5, and pH 5.5–11 for 19-, 34-, and 44-kDa proteinase, respectively. The optimum temperature was 55°C for the 19-kDa proteinase and 55–60°C for 34- and 44-kDa proteinase (Figure 3.3b). All enzymes showed stability at temperatures lower than 50°C.

3.4.3.2 Effects of NaCl and $CaCl_2$ on proteinase activity and stability

All three enzymes exhibited a NaCl-dependent characteristic (Figure 3.3c). The activities sharply increased with NaCl and reached the maximum at 20–30% NaCl. In the absence of NaCl, the activity of the 19-, 34-, and 44-kDa proteinases was only 4.1, 4.6, and 3.4%, respectively, of its activity at 30% NaCl. The distinct feature of these proteinases was their stability in a wide range of NaCl (0–30%). All enzymes were activated by calcium ions (Figure 3.3d). Activities steeply increased at 20 mM $CaCl_2$ and gradually increased thereafter up to 100 mM. All enzymes were stable at $CaCl_2$ < 20 mM. Stability of the 19-, 34-, and 44-kDa proteinase decreased to about

50, 35, and 45%, respectively, after incubation in 100 mM $CaCl_2$ for 24 h. High concentrations of Ca^{2+} are likely to induce structural destabilization caused by excessive ionic interaction between the enzyme surface and calcium molecules, promoting enzyme autolysis and a decreased stability.

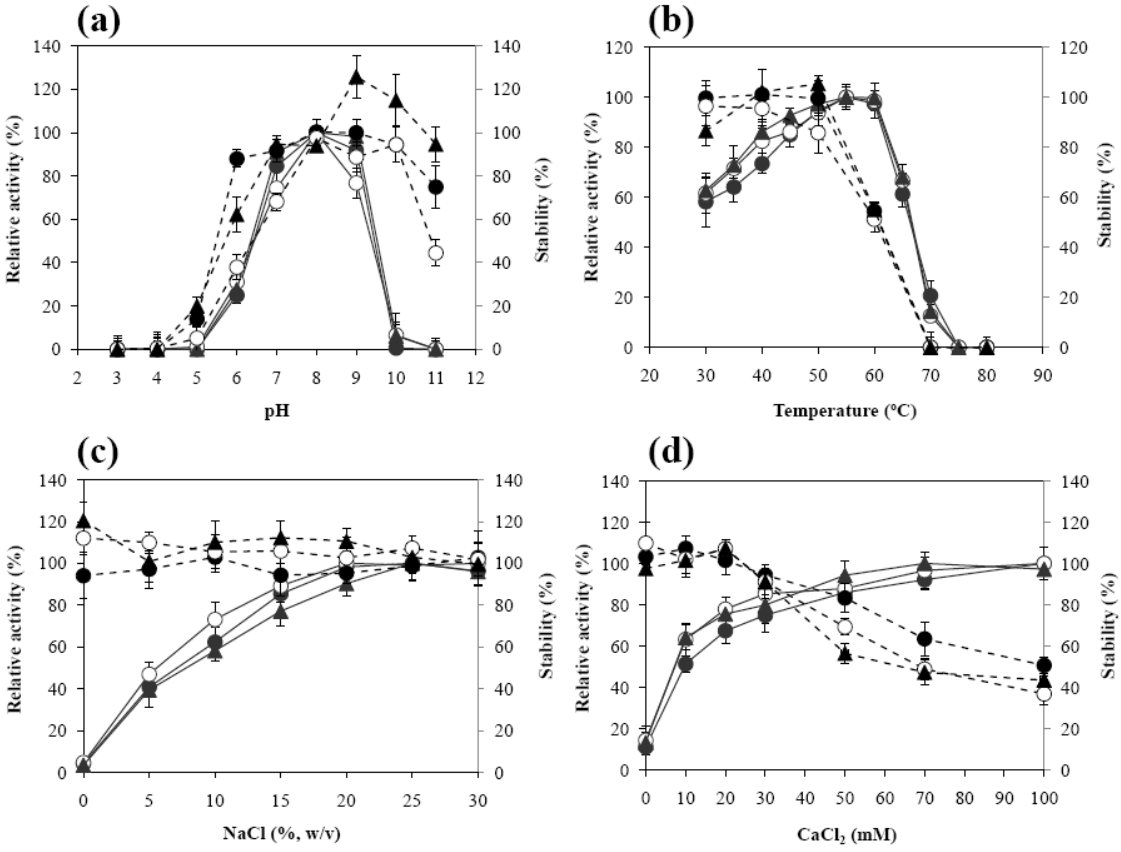


Figure 3.3 Effects of pH (a), temperature (b), NaCl (c) and $CaCl_2$ (d) on activity (solid lines) and stability (dashed lines) of 19- (●), 34- (Δ), and 44-kDa (▲) proteinase.

3.4.3.3 Effects of proteinase inhibitors, denaturants, and ions

All three proteinases were completely inhibited by PMSF and slightly inhibited by TPCK and TLCK, suggesting a serine proteinase characteristic (Table

3.1). Activities of all enzymes were retained in the presence of other inhibitors. Denaturants, namely SDS, Triton X-100, H_2O_2, and urea, markedly inhibited activity of all three proteinases. However, a reducing agent like β-mercaptoethanol did not affect the activity (Table 3.1). These results indicated that hydrogen bonds and hydrophobic interactions rather than disulfide bonds mainly involved in enzyme structural stabilization. Most metal ions tested (Cu^{2+}, Cd^{2+}, Mn^{2+}, and Hg^{2+}) did not affect enzyme activities in the presence of 1 M NaCl (Table 3.2). Only Zn^{2+} and Fe^{3+} strongly inhibited activity of all enzymes, particularly the 34-kDa proteinase.

3.4.4 Isoelectric point determination and mass spectrometry

After DEAE-Sephacel chromatography, proteins with mass of 19, 34, 44, and 50 kDa were detected on SDS-PAGE, while proteinase activity was only found at 19, 34, and 44 kDa on the casein zymogram (Figure 3.4a,b). The p*I* values of the 19, 34, 44, and 50 kDa proteins were found in the acidic pH range at approximately pH 3.6, 5.2, 3.8, and 3.4, respectively (Figure 3.4c). Based on PMF analysis, only a 50-kDa protein was identified as a putative transposase (accession no. 18466600) ($P < 0.05$), while 19-, 34- and, 44-kDa enzyme did not significantly matched with any known proteins ($P > 0.05$). Therefore, these three spots were further analysed. Amino acids in each protein spot were partially sequenced by tandem mass spectrometry. Homology search of de novo peptides of all three proteinases exhibited limited homology towards bacillopeptidase F enzymes (bpFs) from various strains. Only one peptide of the 34-kDa proteinase coincidentally matched other serine proteinase (protease do [EDT73458.1]). Therefore, their *de novo* peptides were partially multiple-aligned with

Table 3.1 Effects of inhibitors and denaturants on the activity of three partially purified proteinases

	Characteristic	Concentration	Relative activity (%)		
			19 kDa	34 kDa	44 kDa
No substance			100	100	100
Inhibitors					
PMSF	All serine proteinases	1 mM	0	0	0
TLCK	Trypsin-like specific	100 μM	69	86	84
TPCK	Chymotrypsin/subtilisin-like specific	100 μM	47	64	69
EDTA	Metal chelator	10 mM	84	91	104
L-Histidine	Metal chelator	10 mM	77	91	90
Pepstatin A	Acid proteinase specific	10 mM	90	89	95
N-Ethylmaleimide	SH-blocking	1 mM	88	94	98
Iodoacetic acid	SH-blocking	1 mM	82	94	98
Bestatin	Amino peptidase specific	10 μM	82	93	110
Denaturants					
SDS	Ionic detergent	2%	0	0	0
Triton X-100	Non-ionic detergent	2%	27	30	31
β-ME	Reducing agent	10 mM	81	94	95
H_2O_2	Oxidizing agent	5%	1	0	0
Urea	H-bond interferring agent	4 M	1	3	1

Note: The enzymes were preincubated in various inhibitors and denaturants for 30 min in McIlvaine buffer, pH 8.0 and proteinase activities were measured at 55°C for 19-kDa proteinase, and 60°C for 34- and 44-kDa proteinase in McIlvaine buffer, pH 8.0. Reactions without substances were taken as 100%.

those of bpFs and some subtilisins from subtilase family S8A (Figure 3.5). All peptide queries from the three enzymes exhibited relatively high homology towards conserved sequence regions of the template sequences aligned (shaded boxes). Three enzymes showed higher homology towards bpFs than subtilisins as most peptide queries corresponded to homologous sequence regions of the bpFs.

Table 3.2 Effects of various metal ions (1 mM) on the activity of three partially purified proteinases

Ions	Concentration (mM)	Relative activity (%)		
		19 kDa	34 kDa	45 kDa
Cu^{2+}	1	93	86	91
Cd^{2+}	1	102	94	100
Mn^{2+}	1	104	89	111
Hg^{2+}	1	100	94	98
Zn^{2+}	1	82	29	68
Fe^{3+}	1	60	28	62

Note: 1) TBS, Tris buffer saline.

2) Proteinase activities were measured in TBS buffer II (50 mM Tris–HCl in 1 M NaCl pH 8.0) containing given ions at 55°C for 19-kDa proteinase and 60°C for 34-, and 44-kDa proteinase. Reactions without ions were taken as 100%.

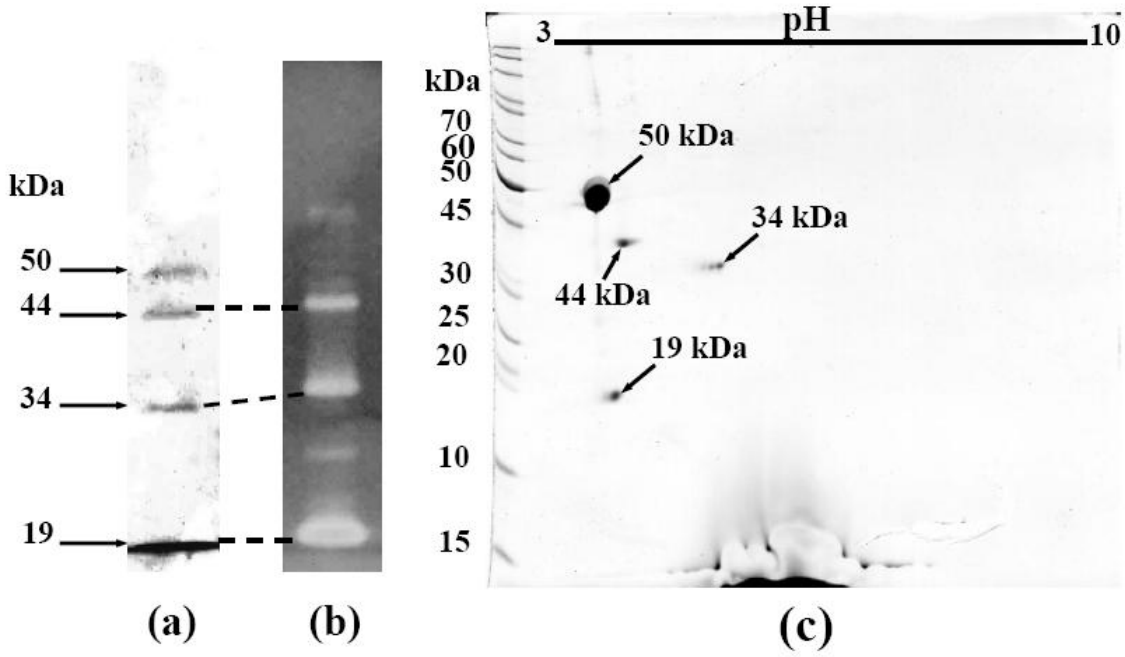


Figure 3.4 Fractions showing a high subtilisin-like activity eluted from a DEAE-Sephacel column were pooled and analysed by nonreducing SDS-PAGE (a) and casein zymography (b). Dashed lines indicate protein bands exhibiting proteinase activity. Isoelectric points of the pooled fraction were analysed using 2D-PAGE with a linear gradient pH from 3 to 10 in the first dimension and 12.5% separating polyacrylamide gel in the second dimension (c). Mass of standard proteins was marked on the left side of the gel. Molecular mass of each protein spot on the 2D gel was indicated.

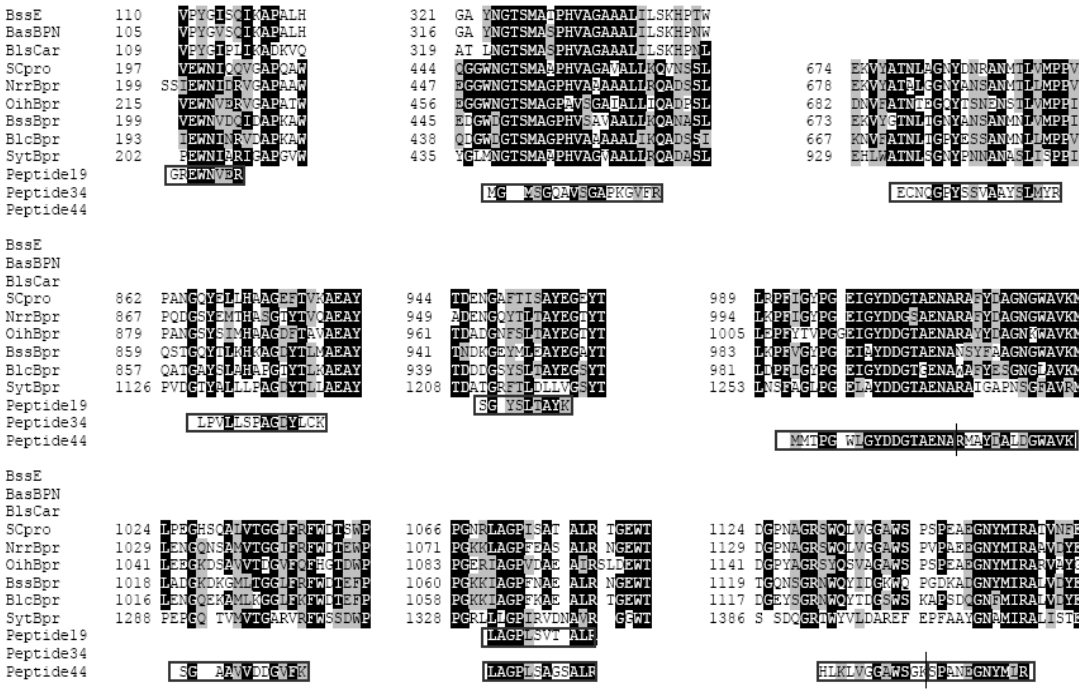


Figure 3.5 Multiple-amino acid alignment of selected serine proteinases in subtilases subfamily S8A with those of selected peptides from proteinases with mass of 19, 34, and 44 kDa obtained from LC–MS⁄MS partial amino acid sequencing. BssE = subtilisin E (AAA22742.1) from *Bacillus subtilis* ssp. subtilis str. 168; BasBPN = subtilisin BPN′ (CAB56500.1) from *B. amyloliquefaciens*; BlsCar = subtilisin Carlsberg (CAB56500.1) from *B. licheniformis*; SCpro = SC protease (BAD99234.1) from *Bacillus* sp. KSM-LD; NrrBpr = bacillopeptidase F (EAR64665.1) from *Bacillus* sp. NRRL B-14911; OihBpr = bacillopeptidase F (BAC14287.1) from *Oceanobacillus iheyensis* HTE831; BssBpr = bacillopeptidase F (AAA62679.1) from *B. subtilis* ssp. subtilis str. 168; BlcBpr = bacillopeptidase F (AAU23286.1) from *B. licheniformis* ATCC 14580; SytBpr = bacillopeptidase-like protein (BAD39703.1) from *Symbiobacterium thermophilum* IAM 14863; and Peptide19, Peptide34, and Pep-

(Figure 3.5 Continued)

tide44 = de novo peptides from 19-, 34- and 44-kDa proteinase, respectively. Numbers on the left side of each alignment indicate the amino acid number of each aligning template. The length of each peptide query is boxed. Amino acid identity and similarity were set at a threshold of 50% and are shaded in a black and grey box, respectively. Only regions of template sequences showing homology to the peptide queries are shown.

3.5 Discussion

In this study, we have shown that at least six out of 11 extracellular proteolytic enzymes secreted by *Virgibacillus* sp. SK37 were subtilisin-like. All three major proteinases with mass of 19, 34, and 44 kDa exhibited similar pH optimum at pH 8 and were thermophilic with optimum at 55–60°C. These characteristics are in agreement with typical subtilisins (Rao et al., 1998). Sensitivity towards H_2O_2 of these enzymes was presumably caused by the oxidation of methionine located next to the active site serine found in most subtilases (Bott, Ultsch, Kossiakoff, Graycar, Katz, and Power, 1988; Stauffer and Etson, 1969). Inactivation of proteinase by Zn^{2+} was also reported in subtilisin-like enzymes produced from *Streptomyces albogriseolus* (Suzuki, Taguchi, Yamada, Kojima, Miura, and Momose, 1997) and *Salinivibrio costicola* (Lama, Romano, Calandrelli, Nicolaus, and Gambacorta, 2005). The catalytic activity of all three enzymes was not inactivated in the absence of NaCl and showed appreciable stability both in the absence and presence of saturated NaCl (30%). Most extracellular serine proteinases produced from extremely and moderately

halophilic bacteria exhibit a halophilic character, which is inactive and irreversibly denatured at relatively low NaCl concentrations (Kim and Kim, 2005; Namwong et al., 2006; Setyorini et al., 2006; Uchida et al. 2004). Convergent evolution analysis of the amino acid compositions of proteins exhibiting the halophilic character reveals a higher content of acidic residues and a lower content of basic residues than those of nonhalophiles (Lanyi 1974; Ventosa et al., 1998). Accordingly, the enzymes in this study, *i.e.* 19-, 34-, and 44-kDa, should be classified as halotolerant enzymes that are capable of retaining activity at a wide range of NaCl concentrations but require no NaCl for their activity and stability. Besides PMSF, Ca^{2+} chelators such as EDTA and ethylene glycol tetraacetic acid (EGTA) inhibited many members of subtilases (Rawlings et al., 2008). Crystal structures of subtilisin BPN´ (Bott et al., 1988) and subtilisin Carlsberg (Bode, Papamokos, and Musil, 1987) revealed that they contain conserved Ca^{2+}-binding sites that are essential for enzyme stability. However, no inhibitory effect of EDTA on the activity of three enzymes was detected. Moreover, Ca^{2+} was found to be an activator for all enzymes. All three enzymes retained their activity even in the absence of $CaCl_2$, indicating that Ca^{2+} ion was not necessary to maintain proteinase stability. Such character has also been reported in subtilases subfamily S8A produced from various strains of alkaliphilic *Bacillus* sp. (Joo et al., 2003; Kobayashi et al., 1995; Kumar and Takagi, 1999; Nonaka, Fujihashi, Kita, Saeki, Ito, and Miki, 2001; Okuda et al., 2004; Saeki et al., 2000).

The PMF analysis suggested that all three proteinases were putative novel. The relatively low protein scores and the nonproteolytic-based function of top-matched proteins (data not shown) implied distant evolutionary relationships with other bacterial serine proteinases. According to amino acid sequence homology analysis, 3

of 13 (23%), 4 of 14 (28.6%) and 7 of 14 (50%) *de novo* peptides from the 19-, 34-, and 44-kDa enzymes, respectively, showed remarkable homology towards bpFs. Other peptide queries showed some homology to a limited amino acid sequence of unrelated proteins (data not shown). Multiple sequence alignments among subtilases unraveled residues of large or unusual deletions and insertions (Siezen and Leunissen, 1997). The high variations of number, position and sequence homology within these regions allowed the implication of divergent rather than convergent evolution. Most of the insertion sequences were located in turns and external loops within enzyme structures and were proven to play an important role in some extraordinary characteristics of an extremozyme, such as thermostability of a thermophilic subtilase (Wu, Bian, Tang, Chen, Shen, and Peng, 2004). Multiple alignments of the enzyme templates used in this study (Figure 3.5) revealed six insertions showing additional residues within catalytic domains, and N- and C-terminal extensions (data not shown). As all three enzymes in this study exhibited a distinctive halotolerant characteristic with no requirement of Ca^{2+} for their stability, it is thought that the variation in homology among query peptides mainly originated from the nature of the insertion sequences, presumably in their primary structure.

The bpF is a member of the subtilase subfamily S8A and has been reported in genera *Choroflexus*, *Roseiflexus*, *Herpetosiphon*, *Symbiobacterium*, *Salinispora*, *Oceanobacillus*, *Bacillus* and *Syntrophobacter*, where complete coding genes are available (Rawlings et al., 2008). They typically exhibit mass of 30–90 kDa, optimum pH around 8.1 and are inhibited by PMSF but not by EDTA, reducing agents or several metal ions (Hageman, 2004). The bpF shows an acidic p*I* around 3.9–5.4 (Hageman, 2004), which is different from other enzymes in the subfamily S8A

showing basidic p*I* of 7.8–11 (Graycer, Ballinger, and Wells, 2004). Recent studies have indicated that bpFs produced from *B. subtilis* ‘natto’ and *B. licheniformis* KJ-31 exhibited fibrinolytic and antithrombotic effects, which are attractive for health promoting biomaterials and other medical applications (Hwang, Choi, Kim, Park, and Cha, 2007; Omura, Hitosugi, Zhu, Ikeda, Maeda, and Tokudome, 2005). On the other hand, bpFs produced from moderately halophilic bacteria has been rarely reported. Based on N-terminal sequence similarity, a bpF-like enzyme secreted from *Filobacillus* sp. RF2-5, a moderately halophilic bacterium isolated from Thai fish sauce, was reported (Hiraga, Nishikata, Namwong, Tanasupawat, Takada, and Oda, 2005). It showed molecular mass of 49 kDa, optimum catalytic conditions at pH 10–11 (stable at pH 5–10), 60°C (stable at lower 50°C), 15–25% NaCl (stable up to 25% NaCl) and was strongly inhibited by PMSF but not by TLCK, TPCK and EDTA. Besides molecular mass, these characteristics are in good agreement with biochemical properties of three proteinases from *Virgibacillus* sp. SK37.

Even though all three proteinases were found to be related to members of bpF, the relationship between these three enzymes is questionable. Generally, bpF is expressed from one copy of bpr gene in premature form (Sloma, Rufo, Rudolph, Sullivan, Theriault, and Pero, 1990). This form is further processed to cleave off its pre-, pro-, and signal peptide located in both N- and C-terminal of the sequence and yields several isoforms that differ from each other in size without detectable loss in catalytic activity (Sloma et al., 1990). Some of these isoforms were reported with mass of 36 and 33 kDa (p*I* 4.4), 50 kDa (p*I* 5.4), 48, 50, 68 and 80 kDa (p*I* 5.4), 47 kDa (p*I* 4.5), 65, 74 kDa and 88 kDa (p*I* 3.9) (Hageman, 2004). These are in good agreement with the results obtained from 34- and 44-kDa proteinase, but not with

those from a 19-kDa counterpart. Nevertheless, whether these three enzymes are isoforms of the bpF or are completely distinct from each other in terms of gene origin needs further investigation. In addition, it is of interest to consider the role of these three enzymes because they are produced in a large amount during the exponential phase of *Virgibacillus* sp. SK37. In contrast, bpFs in bacilli accounts for very few percentages of total proteinase activity and is produced mainly in the stationary phase and is believed not to play important biological functions (Boyer and Carlton, 1968). Our study might suggest the importance of these bpF-like enzymes for the adaptability of *Virgibacillus* sp. SK37 in hypersaline environments.

3.6 Conclusions

This is the first report of the identification of bpF-like proteinases in *Virgibacillus*. The halotolerant activity of these three enzymes might simply explain the ability and adaptability of the utilization of nitrogen sources over a wide range of salt concentrations in *Virgibacillus* sp. SK37, and possibly in other moderate halophiles. These proteinases might be useful as a catalyse of a reaction containing high salt concentration. In addition, these enzymes are a good candidate for further investigation of the mechanism of halophilicity at the molecular level.

3.7 References

Bode, W., Papamokos, E., and Musil, D. (1987). The highresolution X-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin c, an elastase inhibitor from the leech Hirudo medicinalis structural analysis, subtilisin structure and interface geometry. **Eur. J. Biochem.**166: 673-692.

Bott, R., Ultsch, M., Kossiakoff, A., Graycar, T., Katz, B., and Power, S. (1988). The three-dimensional structure of *Bacillus amyloliquefaciens* subtilisin at 1.8 A and an analysis of the structural consequences of peroxide inactivation. **J. Biol. Chem.** 263: 7895-7906.

Boyer, H. W., and Carlton, B. C. (1968) Production of two proteolytic enzymes by a transformable strain of *Bacillus subtilis*. **Arch. Biochem. Biophys.** 128: 422–455.

Carrasco, I. J., Ma´rquez, M. C., and Ventosa, A. (2009). *Virgibacillus salinus* sp. nov., a moderately halophilic bacterium from sediment of a saline lake. **Int. J. Syst. Evol. Microbiol.** 59: 3068-3073.

Chamroensaksri, N., Akaracharanya, A., Visessanguan, W., and Tanasupawat, S. (2008). Characterization of halophilic bacterium NB2-1 from Pla-Ra and its protease production. **J. Food Biochem.** 32: 536-555.

Chen, Y. G., Cui, X. L., Wang, Y. X., Zhang, Y. Q., Tang, S. K., Li, W. J., Liu, Z. X., and Wen, M. L. (2009). *Virgibacillus sediminis* sp. nov., a moderately halophilic bacterium isolated from a salt lake in China. **Int. J. Syst. Evol. Microbiol.** 59: 2058-2063.

Graycer, T. P., Ballinger, M. D., and Wells, J. A. (2004). Subtilisins. In A. J. Barrett, N. D. Rawlings, and J. F. Woessner (eds.). **Handbook of Proteolytic Enzymes** (pp. 1786-1792). London: Elsevier.

Gupta, R., Beg, Q. K., and Lorenz, P. (2002). Bacterial alkaline proteases: molecular approaches and industrial applications. **Appl. Microbiol. Biotechnol.** 59: 15-32.

Gupta, A., Joseph, B., Mani, A., and Thomas, G. (2008). Biosynthesis and properties of an extracellular thermostable serine alkaline proteinase from *Virgibacillus pantothenticus.* **World J. Microbiol. Biotechnol.** 24: 237-243.

Hageman, J. H. (2004). Bacillopeptidase. In A. J. Barrett, N. D. Rawlings, and J. F. Woessner (eds.). **Handbook of Proteolytic Enzymes** (pp. 1795-1796). London: Elsevier.

Heukeshoven, J., and Dernick, R. (1985). Simplified method for silver staining of proteins in polyacrylamide gels and the mechanism of silver staining. **Electrophoresis.** 6: 103-112.

Hiraga, K., Nishikata, Y., Namwong, S., Tanasupawat, S., Takada, K., and Oda, K. (2005). Purification and characterization of serine proteinase from a halophilic bacterium, *Filobacillus* sp. RF2-5. **Biosci. Biotechnol. Biochem.** 69: 38-44.

Hua, N. P., Hamza, C. A., Vreeland, R. H., Isoda, H., and Naganuma, T. (2008). *Virgibacillus salarius* sp. nov., a halophilic bacterium isolated from a Saharan salt lake. **Int. J. Syst. Evol. Microbiol.** 58: 2409-2414.

Hwang, K. J., Choi, K. H., Kim, M. J., Park, C. S., and Cha, J. (2007). Purification and characterization of a new fibrinolytic enzyme of *Bacillus licheniformis* KJ-31, isolated from Korean traditional Jeot-gal. **J. Microbiol. Biotechnol.** 17: 1469-1476.

Joo, H. S., Kumar, C. G., Park, G. C., Paik, S. R., and Chang, C. S. (2003). Oxidant and SDS-stable alkaline protease from *Bacillus clausii* I-52: production and some properties. **J. Appl. Microbiol.** 95: 267-272.

Kim, W. J., and Kim, S. M. (2005). Purification and characterization of *Bacillus subtilis* JM-3 protease from anchovy sauce. **J. Food Biochem.** 29: 591-610.

Kobayashi, T., Hakamada, Y., Adachi, S., Hitomi, J., Yoshimatsu, T., Koike, K., Kawai, S., and Ito, S. (1995). Purification and properties of an alkaline protease from alkalophilic *Bacillus* sp. KSM-K16. **Appl. Microbiol. Biotechnol.** 43: 473-481.

Kumar, C. G., and Takagi, H. (1999). Microbial alkaline proteinases: from a bioindustrial viewpoint. **Biotechnol. Adv.** 17: 561-594.

Kushner, D. J. (1978). Life in high salt and solute concentrations: halophilic bacteria. In D. J. Kushner (ed.). **Microbial Life in Extreme Environments** (pp. 317-368). London: Academic Press, Ltd.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. **Nature.** 277: 680-685.

Lama, L., Romano, I., Calandrelli, V., Nicolaus, B., and Gambacorta, A. (2005). Purification and characterization of a protease produced by an aerobic haloalkaliphilic species belonging to the *Salinivibrio* genus. **Res. Microbiol.** 156: 478-484.

Lanyi, J. K. (1974). Salt-dependent properties of proteins from extremely halophilic bacteria. **Bacteriol. Rev.** 38: 272-290.

Namwong, S., Hiraga, K., Takada, K., Tsunemi, M., Tanasupawat, S., and Oda, K. (2006). A halophilic serine proteinase from *Halobacillus* sp. SR5-3 isolated from fish sauce: purification and characterization. **Biosci. Biotechnol. Biochem.** 70: 1395-1401.

Neuhoff, V., Stamm, R., and Eibl, H. (1985). Clear background and highly sensitive protein staining with Coomassie Blue dyes in polyacrylamide gels: a systematic analysis. **Electrophoresis.** 6: 427-448.

Nonaka, T., Fujihashi, M., Kita, A., Saeki, K., Ito, S., and Miki, K. (2001). Crystallization and preliminary X-ray diffraction studies of a novel alkaline serine protease (KP-43) from alkaliphilic *Bacillus* sp. strain KSM-KP43. **Acta. Crystallogr.** 57: 717-718.

Okuda, M., Sumitomo, N., Takimura, Y., Ogawa, A., Saeki, K., Kawai, S., Kobayashi, T., and Ito, S. (2004). A new subtilisin family: nucleotide and deduced amino acid sequences of new high-molecular-mass alkaline proteases from *Bacillus* spp. **Extremophiles.** 8: 229-235.

Omura, K., Hitosugi, M., Zhu, X., Ikeda, M., Maeda, H., and Tokudome, S. (2005). A newly derived protein from *Bacillus subtilis* natto with both antithrombotic and fibrinolytic effects. **J. Pharm. Sci**. 99: 247-251.

Peng, Q. Z., Chen, J., Zhang, Y. Q., Chen, Q. H., Peng, D. J., Cui, X. L., Li, W. J., and Chen, Y. G. (2009). *Virgibacillus zhanjiangensis* sp. nov., a marine bacterium isolated from sea water. **Anton Leeuw.** 96: 645-652.

Rao, M. B., Tanksale, A. M., Ghatge, M. S., and Deshpande, V. V. (1998). Molecular and biotechnological aspects of microbial proteases. **Microbiol. Mol. Biol. Rev.** 62: 597–635.

Rawlings, N. D., Morton, F. R., Kok, C. Y., Kong, J., and Barrett, A. J. (2008). MEROPS: the peptidase database. **Nucleic Acids Res.** 36: 320-325.

Rohban, R., Amoozegar, M. A., and Ventosa, A. (2009). Screening and isolation of halophilic bacteria producing extracellular hydrolyses from Howz Soltan Lake, Iran. **J. Ind. Microbiol. Biotechnol.** 36: 333-340.

Saeki, K., Mitsuyoshi, O., Hatada, Y., Kobayashi, T., Ito, S., Takami, H., and Horikoshi, K. (2000). Novel oxidatively stable subtilisin-like serine proteases

from alkaliphilic *Bacillus* spp.: enzymatic properties, sequences, and evolutionary relationships. **Biochem. Biophys. Res. Commun.** 279: 313-319.

Setyorini, E., Takenaka, S., Murakami, S., and Aoki, K. (2006). Purification and characterization of two novel halotolerant extracellular proteases from *Bacillus subtilis* strain FP-133. **Biosci. Biotechnol. Biochem.** 70: 433–440.

Siezen, R. J., and Leunissen, J. A. M. (1997). Subtilases: the superfamily of subtilisin-like serine proteases. **Protein Sci.** 6: 501-523.

Siezen, R. J., de Vos, W. M., Leunissen, J. A. M., and Dijkstra,B. W. (1991). Homology modeling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteases. **Protein Eng.** 4: 719-739.

Sinsuwan, S., Rodtong, S., and Yongsawatdigul, J. (2007). NaCl activated extracellular proteinase from *Virgibacillus* sp. SK37 isolated from fish sauce fermentation. **J. Food Sci.** 72: 264-269.

Sinsuwan, S., Rodtong, S., and Yongsawatdigul, J. (2008a). Production and characterization of NaCl-activated proteinases from *Virgibacillus* sp. SK33 isolated from fish sauce fermentation. **Process Biochem.** 43: 185-192.

Sinsuwan, S., Rodtong, S., and Yongsawatdigul, J. (2008b). Characterization of Ca^{2+}-activated cell-bound proteinase from *Virgibacillus* sp. SK37 isolated from fish sauce fermentation. **LWT-Food Sci. Technol.** 4: 2166-2174.

Sinsuwan, S., Rodtong, S., and Yongsawatdigul, J. (2010). A NaCl-stable serine proteinase from *Virgibacillus* sp. SK33 isolated from Thai fish sauce. **Food Chem.** 119: 573-579.

Sloma, A., Rufo, G. A. Jr, Rudolph, C. F., Sullivan, B. J., Theriault, K. A., and Pero, J. (1990). Bacillopeptidase F of *Bacillus subtilis*: purification of the protein and cloning of the gene. **J. Bacteriol.** 172: 1470-1477.

Stauffer, C. E., and Etson, D. (1969). The effect on subtilisin activity of oxidizing a methionine residue. **J. Biol. Chem.** 244: 5333-5338

Suzuki, M., Taguchi, S., Yamada, S., Kojima, S., Miura, K. I., and Momose, H. (1997). A novel member of the subtilisin like protease family from *Streptomyces albogrisolus*. **J. Bacteriol**. 179: 430-438.

Uchida, H., Kondo, D., Yamashita, S., Tanaka, T., Tran, L. H., Nagano, H., and Uwajima, T. (2004). Purification and properties of a protease produced by *Bacillus subtilis* CN2 isolated from a Vietnamese fish sauce. **J. Microbiol. Biotechnol.** 20: 579-582.

Ventosa, A. J., Nieto, J., and Oren, A. (1998). Biology of moderately halophilic aerobic bacteria. **Microbiol. Mol. Biol. Rev.** 62: 504-554.

Wang, C. Y., Chang, C. C., Ng, C. C., Chen, T. W., and Shyu, Y. T. (2008). *Virgibacillus chiguensis* sp. nov., a novel halophilic bacterium isolated from Chigu, a previously commercial saltern located in southern Taiwan. **Int. J. Syst. Evol. Microbiol.** 58: 341-345.

Wu, J., Bian, Y., Tang, B., Chen, X., Shen, P., and Peng, Z. (2004). Cloning and analysis of WF146 protease, a novel thermophilic subtilisin-like protease with four inserted surface loops. **FEMS Microbiol. Lett.** 230: 251-258.

Yongsawatdigul, J., Rodtong, S., and Raksakulthai, N. (2007). Acceleration of Thai fish sauce fermentation using proteinases and bacterial starter cultures. **J. Food Sci.** 72: 382-390.

Yoon, J. H., Kang, S. J., Jung, Y. T., Lee, K. C., Oh, H. W., and Oh, T. K. (2010). *Virgibacillus byunsanensis* sp. nov., isolated from a marine solar saltern. **Int. J. Syst. Evol. Microbikol.** 60: 291-295.

CHAPTER IV

CLONING, EXPRESSION, AND CHARACTERIZATION OF A NOVEL SUBTILASE WITH NaCl-ACTIVATED AND OXIDANT-STABLE ACTIVITY FROM *VIRGIBACILLUS* SP. SK37

4.1 Abstract

A gene encoding a novel member of subtilase superfamily was isolated from *Virgibacillus* sp. SK37, a proteinase-producing bacterium isolated from Thai fish sauce fermentation. The gene was cloned by activity-based screening of a genomic DNA expression library on LB agar plates containing 1 mM IPTG and 3% skim milk. Out of 100,000 clones screened, all six isolated positive clones comprised one overlapping open reading frame of 45% identity to *aprX* gene from *Bacillus* sp. This gene, designated *aprX-sk37* was cloned into pET21d(+) and over-expressed in *E. coli* BL21(DE3). The enzyme product, designated AprX-SK37, was purified by immobilized metal ion affinity chromatography to apparent homogeneity and characterized. The recombinant AprX-SK37 enzyme showed optimal catalytic conditions at pH 9.5 and 55°C, based on azocasein assay containing 5 mM $CaCl_2$. Maximum catalytic activity was found at 1 M NaCl with residual activity of 30% at 3 M NaCl. Thermal stability of the enzyme was also enhanced by 1 M NaCl. The enzyme was absolutely calcium-dependent, with optimal concentration of $CaCl_2$ at 15

mM. Inhibitory effects by phenylmethanesulfonyl fluoride and ethylenediaminetetra-acetic acid indicated that this enzyme is a metal-dependent serine proteinase. The enzyme activity was sensitive towards reducing agents, urea, and SDS, but relatively stable up to 5% of H_2O_2. Phylogenetic analysis suggested that AprX-SK37 belongs to a novel family of subtilase superfamily. The name of this new family was proposed as alkaline serine proteinase-X (AprX). The stability towards H_2O_2 and moderately halo- and thermo-tolerant properties of this enzyme are attractive for various biotechnological applications.

Keywords: AprX, subtilase, subtilisin-like proteinase, oxidant-stable, halotolerant, NaCl activated, *Virgibacillus* sp.

4.2 Introduction

Subtilases from *Bacillus* spp. are one of the most commercially valuable enzymes and have been extensively studied in terms of both biological properties and applications (Gupta, Beg, and Lorenz, 2002; Kumar and Takagi, 1999; Rao, Tanksale, Ghatge, and Deshpande et al., 1998). Subtilase is the superfamily of subtilisin-like serine proteinases (Siezen and Leunissen, 1997) in the clan SB of serine peptidases, family S8, according to the MEROPS database (Rawlings, Barrett, and Bateman, 2010). It is one of the biggest clans of the serine peptidases. Subtilases are ubiquitously distributed in various organisms, including bacteria, archaea, eukaryotes, and viruses (Siezen, de Vos, Leunissen, and Dijkstra, 1991). Based on amino acid sequence similarity, they are classified into six families, *i.e.*, subtilisins, thermitase, proteinase K, lantibiotic peptidase, kexin, and pyrolysin (Siezen and Leunissen, 1997). Subtilisins are further classified into six subfamilies namely, true subtilisins,

high-alkaline proteinases, intracellular proteinases (Siezen and Leunissen, 1997), phylogenetically intermediate subtilisins (PISs) (Kobayashi et al., 1995; Saeki et al., 2003), high molecular mass subtilisins (HMSs) (Ogawa et al., 2003; Okuda et al., 2004), and oxidant-stable proteinases (OSPs) (Saeki et al., 2000, 2002). The later subfamily-OSPs has been shown to have a wide application in detergent and bleaching industries and recently become one of the most attractive enzymes in various biotechnological applications. (Gupta, Gupta, Saxena, and Khan, 1999; Haddar et al, 2009; Jaouadi, Ellouz-Chaabouni, Rhimi, and Bejar, 2008; Joo, Kumar, Park, Paik, and Chang, 2003; Saeki et at., 2000, 2002, 2003; Saeki, Ozaki, Kobayashi, and Ito, 2007).

Despite the tremendous amount of information on microbial proteinases, a search for novel proteinase with unique property is still of interest for both basic and applied aspects of these highly complex class of enzymes (Gupta et al., 2002). In addition to a classical method of screening for a new microorganism harboring interesting proteinase from various environments (Gupta et al., 2002), different molecular biology techniques, such as directed evolution (Ness et al., 1999; Zhao and Arnold, 1999), site directed mutagenesis (Maurer, 2004) and metagenomic analysis (Waschkowitz, Rockstroh, and Daniel, 2009.), have been used to engineer numerous proteinases with improved or novel properties. By applying DNA technology, it is essential that the gene of the proteinase of interest be cloned, engineered and expressed in an efficient host.

In this research, a genomic library of *Virgibacillus* sp. SK37, a moderately halophilic bacterium isolated from Thai fish sauce (Nawong, 2006) was constructed and screened for heterologous proteinase expression in *Escherichia coli*. After an

extensive screening, one open reading frame encoding a gene similar to *aprX* was obtained. This gene was cloned, over-expressed in *E. coli* and purified by immobilized metal ion affinity chromatography (IMAC). Biochemical characterization of the purified enzyme revealed that this novel proteinase was relatively oxidant-stable and moderately halo- and thermo-tolerant. This is the first report on the biochemical characterization of recombinant AprX-related enzyme. In addition, evolutionary relationship of this enzyme with other AprXs from different strains and other enzymes in the subtilases superfamily was studied.

4.3 Materials and Methods

4.3.1 Materials

Restriction endonucleases, T4 DNA ligase, and calf intestinal phosphatase (CIP) were purchased from New England Biolabs (New England Biolabs, Ipswich, MA). *Pfu* polymerase was obtained from Promega (Promega, Madison, WI). Skim milk was purchased from Merk (MerkKGoA, Darmstadt, Germany). Azocasein and subtilisin A (subtilisin Carlsberg) were purchased from Sigma (Sigma-Aldrich, St. Louis, MO). All other reagents for molecular and biochemical analysis were of molecular and analytical grade, respectively.

4.3.2 Bacterial strains, plasmids, and, growth conditions

Virgibacillus sp. SK37 (GenBank/NCBI no. DQ910840) was cultured at 35°C in a modified halophilic medium (1% yeast extract, 0.3% trisodium citrate, 0.2% potassium chloride, 2.5% magnesium sulfate, and 5% NaCl). The pH was adjusted to 7.0 by NaOH. *Escherichia coli* strain DH5α and DH10B (Mega DH10B^{TM} T1^{R}electrocomp^{TM}, Invitrogen, CA) were used as host strains for cloning and genomic

DNA expression, respectively. *E. coli* BL21(DE3) (Novagen, Madison, WI) was used for enzyme expression. They were grown aerobically at 37°C in Luria-Bertani (LB) medium. Plasmid vector pUC19 (New England Biolabs, Ipswich, MA) and pET21d+ (Novagen, Madison, WI) were used for genomic library construction and over expression, respectively.

4.3.3 General molecular biology techniques

Cloning, DNA manipulation, and agarose gel electrophoresis were done as previously described (Sambrook, Maniatis, and Fritsch, 2001). Genomic DNA was isolated from *Virgibacillus* sp. SK37 using a Wizard® genomic extraction kit (Promega, Madison, WI). Plasmid was isolated with Qiagen plasmid preparation kit (Qiagen, Hilden, Germany). PCR products were separated by gel electrophoresis, and bands of the expected size were isolated from the gel using Qiagen QIAquick gel extraction kit (Qiagen, Hilden, Germany). Automated DNA sequencing was carried out by Macrogen (Seoul, Korea), using primer-walking technique with pUC/M13 forward/reverse primers for pUC19, and T7/SP6 promoter primers for pET21d+ as initial primers, respectively. Enzymes and reagents were used according to the manufacturer's instructions unless otherwise stated.

4.3.4 Genomic DNA library construction

The genomic DNA of *Virgibacillus* sp. SK37 was partially digested by incubating with 10 units of *Bfu*CI per 150 μg of genomic DNA at 37°C for 20 min. Fragments ranging from 3 to 10 kilobase pairs (kb) were isolated from the agarose gel. The plasmid pUC19 was digested with *Bam*HI and 5′-phosphate groups were removed with calf intestinal phosphatase (CIP, New England Biolabs) prior to gel purification using Qiagen QIAquick gel extraction kit (Qiagen, Hilden, Germany).

The genomic *Bfu*CI fragments were ligated with dephosphorylated linearized pUC19 by T4 DNA ligase at a ratio of 3:1. This ligation mixture was precipitated by isopropanol and re-dissolved in 30 µl nuclease-free water prior to transform into competent *E. coli* DH10B by electroporation (2 mm cuvette, 2500 V, Electroporator 2510, Eppendorf AG, Hamburg, Germany). The transformants were then spread on LB_{Amp} plates containing 2% (w/v) glucose ($LB_{Amp+Glu}$). The plates were incubated for approx. 14 h at 37°C. Colonies growing on the $LB_{Amp+Glu}$ plates were scraped from the agar surface and transferred into LB broth containing 15% glycerol and kept at -70°C until use. A small aliquot of the transformants were spread onto LB_{Amp} plate supplemented with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and 0.2% (w/v) bromo-chloro-indolyl-galactopyranoside (X-gal) ($LB_{Amp+IPTG+X\text{-}gal}$) for enumeration and determination of percentage of blue white colonies, representing DNA library size and % background.

4.3.5 Screening of proteinase-producing clones

The clones from genomic DNA library were spread onto LB_{Amp} plates supplemented with 1 mM IPTG and 3% (w/v) skim milk powder. The clones were grown at 37°C for 18 h. Positive colonies showing proteinase activity as indicated by the formation of transparent halo zones around the colonies were selected as positive clones after continual incubation of the plates at 30°C for additional 7 h.

4.3.6 DNA sequence analysis

Plasmid from each proteinase-positive clone was isolated by Qiagen plasmid preparation kit. Both strands of DNA insert were sequenced by automated DNA sequencing (Macrogen, Korea). The full length DNA insert of each positive-clone, open reading frames (ORFs), and homology of putative coding sequences were

analyzed using tools provided by the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov). Bacterial promoter and transcription termination site of proteinase coding gene were predicted by Softberry Bioinformatics tools (http://linux1.softberry.com/berry.phtml). Isoelectric point (p*I*) and molecular mass (MM) of deduced amino acid sequence were calculated using software from ExPASy Proteomics server (http://ca.expasy.org/tools/pi_tool.html).

4.3.7 Cloning, over-expression, and purification of the recombinant AprX-SK37

Putative proteinase gene was amplified by PCR using forward (5′-CTG TGC TCT AGA GCA AAG TAA AAA CAA CTA TTC TAT AAG GAG GAA G-3′) and reverse (5′-CTG TGC CTC GAG ATT CAC TTT CTG CAC CTG CTG TAC G-3′) primers (the *Xba*I and *Xho*I sites are underlined). Thirty-five cycles of PCR were performed in a thermal cycler (MJ Research PCT-200 Thermal Cycler DNA Engine, GMI Inc, Ramsey, MI) using *Pfu* DNA polymerase. The amplified 1356 bp DNA product flanked by *Xba*I and *Xho*I restriction site was cloned into the corresponding restriction sites of pET21d+ vector for the expression of C-terminally 6xHis tagged enzyme. The integrity of the construct, designated pET-AprX-SK37, was confirmed by agarose gel electrophoresis followed by automated double-strand DNA sequencing. To express the enzyme, the pET-AprX-SK37 vector was transformed into *E. coli* BL21(DE3). A single colony of *E. coli* BL21(DE3) harboring pET-AprX-SK37 (BL21(DE3)/pET-AprX-SK37) was grown at 37°C in LB_{Amp} until the OD_{600} reached 0.6-0.7, then 0.1 mM IPTG was added to induce the gene expression. After further incubation at 16°C for 18 h, the cells were harvested and resuspended in lysis buffer (30 mM Tris-HCl, pH 8.0, 0.3 M NaCl, and 10 mM imidazole). Lysozyme was

added to a final concentration of 1 mg/ml and incubated on ice for 30 min. Lysozyme-treated cells were broken by sonication on ice for 2 min with a 10 s bursts at 300 W and a 10 s cooling period between each burst. The lysate was clarified by centrifugation at 10,000 xg for 30 min at 4°C. The recombinant 6xHis-tagged enzyme in the cleared lysate was purified using nickel nitrilotriacetic acid (Ni-NTA) metal affinity chromatography (Qiagen, Hilden, Germany), according to the manufacturer's protocol with some modifications. After the cleared lysate was incubated with Ni-NTA at 4°C for 4 h with rotary shaking, the slurry was loaded into a column and washed with 10 bed volumes of lysis buffer followed by each 10 bed volumes of wash buffers (30 mM Tris-HCl, pH 8.0, 1.0 M NaCl) containing 50 and 80 mM imidazole, respectively. The recombinant enzyme was recovered with an elusion buffer (30 mM Tris-HCl, pH 8.0, 1.0 M NaCl and 250 mM imidazole). Purified enzyme was desalted against 30 mM Trsi-HCl, pH 8.0 in 0.1 M NaCl using HiTrap^{TM} Desalting column (GE healthcare, Uppsala, Sweden), mixed with an equal volume of glycerol (50% glycerol final in 15 mM Tris-HCl, pH 8.0) and kept at -20°C until use. Protein concentration and purity of the purified protein were analyzed by Bradford method using bovine serum albumin (BSA) as the standard (Bradford, 1976), and SDS-PAGE stained with Coomassie Brilliant Blue, respectively.

4.3.8 Determination of proteinase activity

Proteinase activity was determined using azocasein as a substrate. Standard assay reaction in 500 µl contained 5 mg/ml azocasein, 100 mM Tris-glycine buffer, pH 9.5, 5 mM $CaCl_2$, and 20 µg/ml of purified recombinant proteinase. The mixture was incubated at 55°C for 30 min with vigorous shaking. The reaction was stopped by adding 150 µl of 20% trichloroacetic acid (TCA) and kept on ice for 30 min, followed

by centrifugation at 12,500 xg for 30 min. Then, the supernatant (250 µl) was mixed with equal volume of 2.5 M NaOH and the absorbance at 405 nm was measured. Specific activity (U/mg) was defined as an increase of 0.01 OD_{405} units per minute per milligram of enzyme. Proteinase activity was also analyzed by native-PAGE casein zymography. Purified proteinase in a loading buffer without SDS and β-mercaptoethanol were electrophoresed through a 12% polyacrylamide under native condition. Subsequently, the gel was submerged in 2% (w/v) casein solution for 30 min at 4°C with gently shaking. The gel was then transferred into reaction buffer containing 50 mM Tris-glycine, pH 9.5, 250 mM NaCl, and 5 mM $CaCl_2$ and further incubated at 55°C for 90 min, followed by staining with Coomassie Brilliant Blue. The proteinase activity was determined as a clear band against a blue background.

4.3.9 Effects of pH and temperature on proteinase activity

The effect of pH on proteolytic activity was examined in the presence of various 100 mM buffers, *i.e.*, sodium acetate (pH 5.0-6.0), Tris-Maleate (pH 6.0-7.5), Tris-HCl (pH 7.5-9.0), Tris-glycine (pH 8.5-9.5), and glycine (pH 9.0-11.0), at 55°C and 5 mM $CaCl_2$. To study the effect of temperature, the standard assay was applied using Tris-glycine buffer, pH 9.5 at 25-90°C. Thermostability of the proteinase was analyzed by incubating the enzyme at various temperatures (25-90°C) for 120 min and residual activity was measured at standard assay condition (55°C, pH 9.5 and 5 mM $CaCl_2$) and calculated as a percentage of activity of the enzyme prior to incubation.

4.3.10 Effects of NaCl and $CaCl_2$ on proteinase activity

Proteinase activity at various concentrations of either NaCl (0 to 4.0 M) or $CaCl_2$ (0 to100 mM) was measured at 55°C and pH 9.5. Subtilisin A was used as a representative of the subtilisin S8A family, and its catalytic activity was determined at

its optimum condition at 60°C, in 100 mM Tris-HCl, pH 8.0 in appropriate concentrations of NaCl and $CaCl_2$. The $CaCl_2$ concentration was fixed at 5 mM in all NaCl assays. Maximum activity of each proteinase was defined as 100%. In enzyme stability studies, the enzymes were pre-incubated in the presence of various concentrations of either NaCl (0 to 4 M) or $CaCl_2$ (0 to 100 mM) at 25°C for 24 h at pH 9.5 or 8.0 for the purified recombinant AprX-SK37 and subtilisin A, respectively. Residual activities were measured at 55 and 60°C for recombinant AprX-SK37 and subtilisin A, respectively, and expressed as relative activity compared to the reaction containing appropriate concentrations of NaCl or $CaCl_2$ without pre-incubation.

4.3.11 Effects of NaCl and $CaCl_2$ on thermostability

The effects of NaCl and $CaCl_2$ on the thermostability of the recombinant AprX-SK37 was determined by monitoring enzyme activity at a moderate (40°C) and high (55°C) inactivation temperatures according to the results from the thermal stability assays. The enzyme was prepared in 100 mM Tris-glycine pH 9.5 containing; (I) 5 mM $CaCl_2$: (II) 5 mM $CaCl_2$ and 1 M NaCl: (III) 15 mM $CaCl_2$: (IV) 15 mM $CaCl_2$ and 1 M NaCl. The reactions were pre-incubated at either 40 or 55°C for up to 4 h and aliquots were withdrawn at various time intervals to measure residual activities at standard assay condition.

4.3.12 Effects of proteinase inhibitors and denaturants

The effects of various proteinase inhibitors and protein denaturants on proteinase activity were determined using various substances, namely leupeptin, soy-bean trypsin inhibitor (SBTI), L-1-chloro-3-[-4-tosylamido]-7-amino-2-heptanone (TLCK), L-1-chloro-3-[-4-tosylamida]-4- phenyl-2-butane (TPCK), phenylmethane-sulfonyl fluoride (PMSF), ethylenediaminetetraacetic acid (EDTA), imidazole, L-

histidine, bestatin, pepstatin A, iodoacetic acid (IAA), NEM, β-mercaptoethanol (β-ME), dithiothreitol (DTT), E64, urea, and sodium dodecyl sulfate (SDS), under the standard assay condition. Relative activity was calculated using the activity of the enzyme in the absence of these substances as 100%.

4.3.13 Stability towards oxidizing agent

Hydrogen peroxide (H_2O_2) was used as a representative of oxidizing agent. Proteinase activity was measured in the presence of H_2O_2 at concentrations ranging from 0 to 5% (w/v). Enzyme activity was performed before and after pre-incubation the reaction in the absence of substrate for 30 min at 25°C. The reaction in the absence of H_2O_2 and without pre-incubation was taken as 100%.

4.3.14 Amino acid sequence alignment and construction of the phylo-genetic tree

Amino acid sequences of proteinases similar to the enzyme in this study were retrieved by BLAST search (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Multiple amino acid sequence alignment was done by ClustalX program (Thompson, Gibson, Plewniak, Jeanmougin, and Higgins, 1997) using sequences of AprX-related enzymes retrieved from SwissProt database (http://www.isb-sib.ch/services/databases). Phylogenetic tree was constructed from sequences of enzymes representing each members of subtilase superfamily (Siezen and Leunissen, 1997) and oxidant-stable proteinases (OSPs), obtained from extensive search of the Pubmed (http://www.ncbi.nlm.nih.gov/pubmed/) and Swiss-Prot (http://www.isbsib.ch/services/databases) databases. The phylogenetic tree was constructed based on amino acid similarity within the catalytic domain as described by Saeki et al. (2000), using *MEGA* version 4 software (Tamura, Dudley, Nei, and Kumar, 2007) and was inferred

by the bootstrapping neighbor-joining method (Saitou and Nei, 1987), of which sites involving gaps were excluded from the analysis. Bacterial 16S rRNA sequences of OSPs producers were analyzed for evolutionary distance by a similar method.

4.3.15 Genbank accession number

The accession number our recently isolated gene from the *Virgibacillus* sp. SK37 is HM587897 and the protein identification number is ADM26217.

4.4 Results

4.4.1 Genomic DNA library construction and screening

Screening of *Virgibacillus* sp. SK37 genomic DNA expression library in *E. coli* DH10B were done on LB-skim milk agar plates as described in materials and methods section. Enumeration analysis determined the size of the genomic library to be 1.56 $x10^5$ clones with 15% background (blue colonies harboring self-ligated plasmid), yielding 1.33 $x10^5$ clones with inserts. When an average size of 4.4 $x10^6$ bp of bacilli genomic DNA was considered (calculated from whole genome sizes available from the NCBI database), this library theoretically covered the genome size at least 91 and 302 times (depth of coverage) when the DNA fragments of 3,000 and 10,000 bps was taken into account, respectively. Out of approximately 100,000 clones screened (accounting for more than 68 depth of coverage), six clones showed clear halo zones, indicating the presence of proteolytic activity. The recombinant plasmids were individually extracted from these six clones and their DNA inserts were analyzed. Restriction digestion analysis showed that among these 6 clones, there were four different DNA inserts as judged by size, i.e., ~ 3,300 bp (1 clone), ~3,800 bp (1 clone), ~4,300 bp (3 clones), and ~5,300 bp (1 clone). Remarkably, nucleotide

```
                                          -35          P1          -10
-225  ACAAAACTATGCTATGCCTTATAAAACCACCTATATATTTGTGTATTTTTCACCTTTAATGAAGGAAATTTTTTT  -151
                                                                  -35
-150  GCTAACTAATCTAAAAGCAATAGTTAACTGAATTTTGGGATATTTGACTTATTGTAATTTACCAGAAAAGTAGAT   -76
        Ps     -10                                                    RBS
 -75  AAGCTTTAGAGCAAGTAGTGCATGCAGTCGAATAAAGCAAAGTAAAAACAACTATTCTATAAGGAGGAAGTAAGG   -1

   1  M  T  E  R  K  Q  V  W  F  E  E  A  N  S  R  L  D  P  G  L  V  G  Q  L  M     25
   1  ATGACGGAGAGGAAACAAGTATGGTTTGAAGAAGCCAATTCCAGACTAGATCCAGGATTAGTAGGACAGTTGATG    75

  26  K  K  R  K  E  D  P  N  E  T  S  E  D  T  L  P  V  I  V  K  V  Y  Q  N  C     50
  76  AAAAAAAGAAAAGAAGACCCAAATGAAACAAGTGAGGACACGCTACCAGTAATTGTGAAAGTTTACCAGAATTGC   150

  51  T  K  D  M  K  E  D  L  L  K  T  C  E  G  D  S  C  N  T  L  N  D  D  M  E     75
 151  ACAAAGGATATGAAGGAAGATTTGTTAAAAACATGTGAGGGAGATTCTTGTAATACGTTAAATGATGATATGGAA   225

  76  I  L  H  S  L  Y  G  D  L  T  P  K  K  I  R  E  L  K  N  H  E  A  V  E  R    100
 226  ATTTTACACTCGTTGTATGGAGATTTAACACCCAAAAAGATCAGAGAACTAAAAAATCATGAGGCAGTGGAACGT   300

 101  I  F  Y  D  R  D  V  T  A  F  L  D  V  A  T  K  E  I  N  A  V  E  V  Q  Q    125
 301  ATTTTTTATGACCGAGATGTAACGGCGTTTCTTGATGTGGCTACAAAAGAAATAAATGCAGTAGAAGTACAACAA   375

 126  D  L  G  L  T  G  K  D  I  T  I  A  V  I  D  S  G  V  F  P  H  P  D  L  T    150
 376  GATCTTGGTTTAACTGGGAAAGATATTACTATCGCTGTCATTGATTCCGGAGTATTTCCGCATCCTGATTTAACA   450

 151  K  P  E  N  R  I  V  A  F  K  D  F  V  N  K  Q  E  E  P  Y  D  D  N  G  H    175
 451  AAACCGGAGAATAGAATTGTAGCTTTCAAAGATTTTGTTAATAAGCAGGAAGAACCTTATGATGATAATGGCCAT   525

 176  G  T  H  C  C  G  D  A  A  G  N  G  H  H  S  N  G  K  Y  T  G  P  A  P  A    200
 526  GGGACCCACTGTTGTGGAGACGCTGCAGGAAATGGACACCATTCTAATGGAAAGTATACGGGCCCAGCACCGGCT   600

 201  A  S  I  V  G  V  K  V  L  N  E  K  G  G  G  K  L  S  T  I  I  R  G  I  E    225
 601  GCTTCAATTGTTGGGGTTAAAGTCTTAAATGAAAAAGGCGGCGGGAAGCTTTCCACTATTATACGAGGGATAGAA   675

 226  W  C  M  K  H  R  E  K  Y  G  I  R  I  I  S  L  S  S  L  G  A  E  A  Y  E  S    250
 676  TGGTGTATGAAGCATCGAGAAAAGTATGGTATACGAATTATTTCTCTATCTCTAGGTGCTGAAGCATATGAATCT   750

 251  Y  R  D  D  P  L  T  Q  A  T  Q  K  A  W  H  S  G  I  V  V  C  A  A  A  G    275
 751  TACCGTGACGACCCACTCACCCAGGCAACCCAAAAAGCCTGGCATAGCGGTATAGTTGTATGTGCTGCAGCCGGA   825

 276  N  D  G  P  S  R  S  T  I  S  T  P  A  I  D  P  F  I  I  T  V  G  S  A  D    300
 826  AACGACGGTCCATCACGCAGCACCATTAGTACGCCTGCCATTGATCCATTCATTATTACAGTTGGTTCTGCGGAT   900

 301  D  Q  N  T  V  T  R  S  D  A  V  I  S  K  F  S  S  R  G  P  T  I  D  E  L    325
 901  GATCAAAACACAGTAACTCGATCAGATGCTGTTATTTCCAAATTTTCGAGTCGTGGTCCAACCATTGATGAATTA   975

 326  V  K  P  D  I  Y  A  P  G  S  N  I  I  S  L  L  S  P  G  S  A  L  E  K  Q    350
 976  GTCAAACCAGATATTTACGCACCTGGTTCAAATATTATTTCTCTATTATCTCCCGGATCGGCGCTAGAAAAACAA  1050

 351  I  P  E  R  V  I  D  E  N  Y  V  S  L  S  G  T  S  M  A  T  P  I  C  A  G    375
1051  ATACCTGAGCGTGTAATAGATGAAAATTATGTATCACTATCTGGCACTTCGATGGCAACACCTATTTGTGCAGGA  1125

 376  V  I  A  L  M  L  E  A  N  P  Q  L  S  P  N  D  I  K  S  I  L  Q  A  T  S    400
1126  GTTATCGCTCTAATGCTTGAGGCAAATCCTCAGCTAAGTCCAAATGACATAAAAAGTATTTTACAAGCAACATCA  1200

 401  Q  P  T  L  A  D  K  W  G  Y  I  H  A  K  T  A  V  E  M  A  K  D  Y  V  Q    425
1201  CAACCAACACTTGCAGATAAATGGGGATATATACATGCTAAGACTGCCGTAGAGATGGCAAAAGATTACGTACAG  1275

 426  Q  V  Q  K  V  N  *                                                          431
1276  CAGGTGCAGAAAGTGAATTGATTGAATGAGGTGGTTGAGGAGCCCTCTCATCTTGGGGAGCGTTCTCTCATTTGA  1350

1351  ATGGTGGAGTCTCTCATCTTGGAGCCAGGTTCTCTCATTTGGCATTTGATTATGTCCGCTTCAAAATTTGTTACC  1425

1426  CGTCCATTAAAAAAGAACATCTCCTACGGGGAGATGTTCTTCCTTTATTTTCCAAAGCTGTGCTCTGCTAATGCT  1500

1501  GTCACAATTGCATCTTGAATTTGTGGCATGGAGCCTTGGATATAATTGTTCATCATAATGGAGAAGATTAACTTT  1575
1576  TCCCCATCCTTGTTTGTTACATAACCCGATAATGTATTTACTCCGGTTAAGGAACCGGTTT                 1636
```

Figure 4.1 (see next page)

Figure 4.1 Nucleotide and deduced amino acid sequences of the gene *aprX-sk37*. Translation starts at position 1. The *aprX-sk37* gene comprises a 1293 bp ORF encoding 431 amino acid residues, indicated as single-letters above the nucleotide sequence. The nucleotide and predicted amino acid are numbered on the right and left sides. The predicted -35 and -10 regions of the putative short and long promoters (P_S and P_L) are presented in bold. The putative ribosomal binding site (RBS) and the transcription terminator sequences are underlined. Residues involving in the hypothetical catalytic triad, Asp_{140}, His_{178}, and Ser_{367}, are boxed. Stop codon is indicated by an asterisk.

sequence analysis of these clones indicated that they all shared nucleotide sequences overlapping each other, resulting in an open reading frame (ORF) of only one putative gene.

4.4.2 Nucleotide and deduced amino acid sequences analysis

The nucleotide and deduced amino acid sequences of the isolated ORF and its flanking sequences are shown in Figure 4.1. The ORF begins with ATG start codon at nucleotide position 1 and terminates with TGA termination codon at nucleotide position 1,294. A sequence with good consensus to the canonical ribosomal binding-site (AAGGAG) was found 10 bp upstream from the ATG start codon. Two putative promoters (P_L and P_S) showing moderate consensus to σ_A promoter corresponding to the -10 and -35 regions were located at nucleotide position -166 and -189 for P_L and -63 and -81 for P_S, respectively (Figure 4.1). An inverted-repeat sequence from nucleotide 1,440 to 1,465 is found 143 bp downstream of the termination codon. The

ORF in the nucleotide sequence encodes 431 amino acids with a calculated molecular mass of 47,100 Da, and predicted p*I* of 5.2. The deduced amino acid sequence exhibited resemblance to members of subtilase superfamily (subtilisin-like proteinases) with 30 to 50% identity as determined by BLAST search. In addition, the coding enzyme shows a limited similarity to AprXs from *Bacillus* sp. SG-1 (ZP_01859218), *B. licheniformis* ATCC 14580 (YP_079346), and *B. amyloliquefaciens* FZB42 (YP_001421301) with 45, 43, and 42% identity, respectively. Apr-related enzymes are alkaline serine proteinases encoded from *apr* gene found mostly in class bacilli. To date, a number of *apr* genes have been identified such as *aprA* (Alkaline proteinase from *B. thuringiensis*), *aprE* (Subtilisin E from *B. subtilis*; AprE from *B. amyloliquefaciens*; M-proteinase from *B. clausii*; Elastase from *B. cereus*), *aprN* (Subtilisin Nattokinase from *B. subtilis* subsp. "natto"), *aprJ* (Subtilisin J from *B. stearothermophilus*), *aprQ* (Subtilisin ALP from *Bacillus* sp.), *aprM* (AprM from *Bacillus* sp.), *aprS* (Subtilisin Sendi from *Bacillus* sp. G-825-6), and *aprX* (AprX in *B. subtilis*). Herein, this recently isolated gene from the *Virgibacillus* sp. SK37 was designated as *aprX-sk37* and the encoded enzyme as AprX-SK37. Amino acid sequence identities between the AprX-SK37 and other members of subtilases superfamily were: 25% to subtilisin E (P04189), 26% to thermitase (P04072), 19% to proteinase K (P06873), 13% to lantibiotic leader peptidase (Q48854), 15% to kexin (P13134), and 15% to pyrolysin (P72186), which represent the subtilisins (S8A), thermitase (S8B), proteinase K (S8C), lantibiotic peptidase (S8D), kexin (S8E), and pyrolysin families (S8F), respectively. Multiple amino acid sequence alignment of Apr-related enzymes obtained from SwissProt database and the AprX-SK37 is shown in Figure 4.2. The similarity to other AprX

enzymes starts from residue 130 of the AprX-SK37 and extends nearly to the C-terminus of the protein. Aspartic, histidine, and serine in the catalytic triad and asparagine in the oxyanion hole of subtilases are fully conserved in the AprX-SK37 sequence (Asp_{140}, His_{178}, Ser_{367}, and Asn_{276}, respectively, according to AprX-SK37 numbering). All common secondary structure elements of enzymes belonging to the subtilases (Siezen and Leunissen, 1997) could be found in the catalytic core of the AprX-SK37. When alignment gaps of at least three amino acids are considered, among AprX sequences, AprX-SK37 shows unique insertion and deletion regions. The deletions are between Lys_{29} and Glu_{30} (12 residues), Gly_{236}, and Ile_{237} (3 residues); while the insertion is found from Pro_{152} to Asn_{154} (3 residues) according to the AprX-SK37 numbering. The AprX-SK37 lacks a canonical signal sequence for membrane translocation (signal peptide), indicating an intracellular location as suggested by sub-cellular prediction servers of SignalP 3.0 (Emanuelsson, Brunak, von Heijne, and Nielsen, 2007) and ProtCompB (Softberry Bioinformatics tools: http://linux1.softberry.com/berry.phtml). This prediction agreed with all known AprXs, which lack canonical signal sequences for secretion. However, the N-terminal domain (residue 1-129) of AprX-SK37 showed low similarity to those of AprXs from *B. subtilis* (13% identity) (O31788), *B. amyloliquefaciens* (13% identity) (A7Z4Z4), *B. licheniformis* (14% identity) (Q65IP4), *B. halodurans* (15% identity) (Q9KBJ7), *B. pumilus* (18% identity) (A8FDI1), as well as those of the intracellular serine proteinases from *B. subtilis* (6% identity with Isp-1) (P11018) and *Bacillus* sp. NKS-21 (7% identity with Isp-Q) (Q45621).

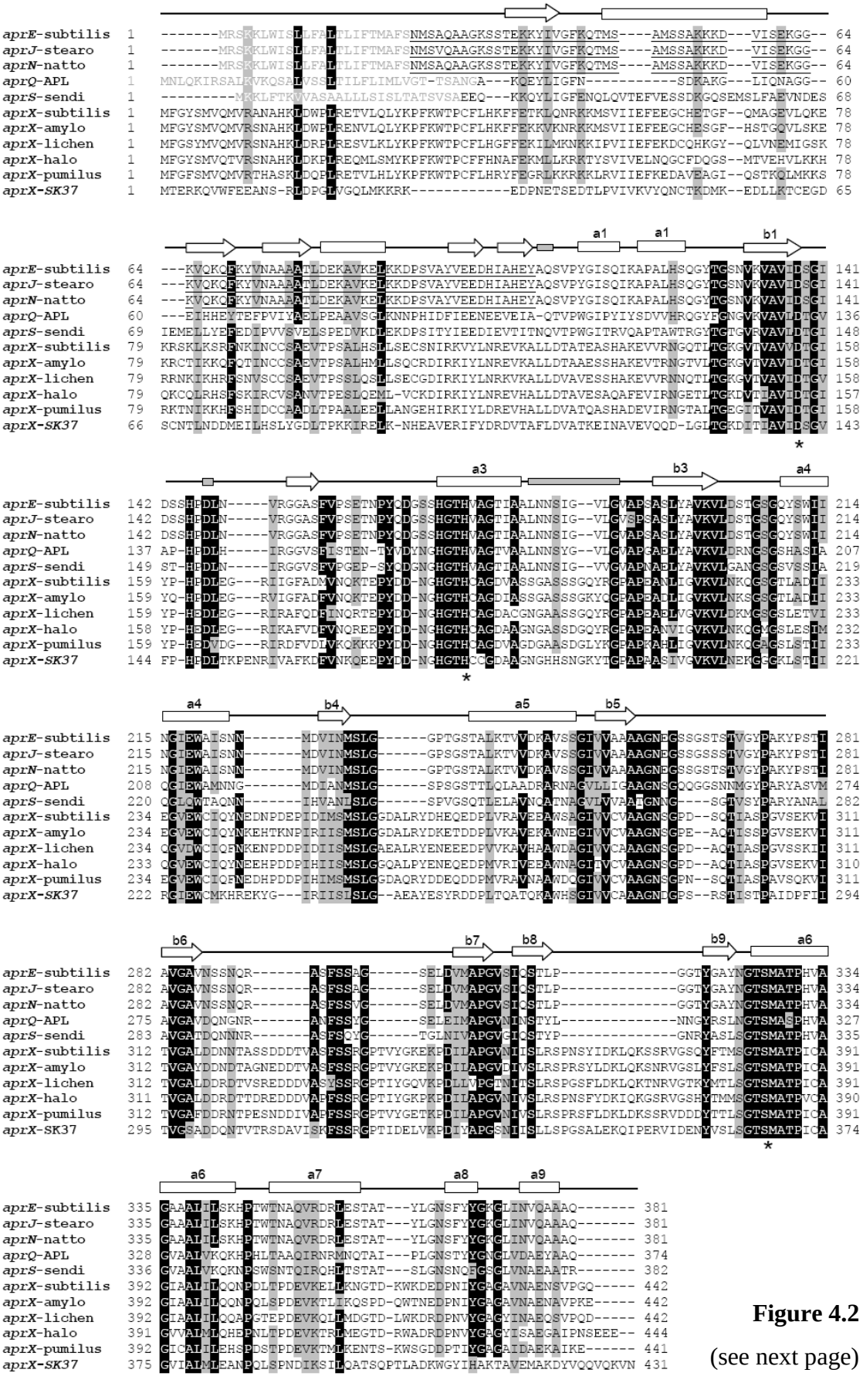


Figure 4.2

(see next page)

Figure 4.2 Multiple amino acid sequence alignment of AprX-SK37 with selected Apr-related subtilisin-like serine proteinases. Sequence aligned: subtilisin E (*aprE*-subtilis) (P04189) from *Bacillus subtilis*; Subtilisin J (*aprJ*-stearo) (P29142) from *B. stearothermophilus*; subtilisin NAT (*aprN*-natto) (P35835) from *B. subtilis* subsp. "natto"; subtilisin ALP (*aprQ*-APL) (Q45523) from *Bacillus.* sp. GN; subtilisin Sendai (*aprS*-sendi) (Q45522) from *Bacillus* sp. G-825-6; AprX (*aprX*-subtilis) (O31788) from *B.* subtilis; AprX (*aprX*-amylo) (A7Z4Z4) from *B. amyloliquefaciens*; AprX (*aprX*-lichen) (Q65IP4) from *B. licheniformis*; intracellular alkaline serine proteinase (*aprX*-halo) (Q9KBJ7) from *B. halodurans*; S8A subfamily proteinase (*aprX*-pumilus) (A8FDI1) from *B. pumilus*; alkaline serine proteinase X (*aprX-sk37*) from *Virgibacillus* sp. SK37. Secondary structure elements of the AprE are indicated above the alignment; common structural elements are shown as: rectangular for α-helices, arrow for β-sheet strands, and filled thin bar for calcium binding domain (Subbian, Yabuta, and Shinde, 2004). Identical and similar (>75%) amino acid residues among all enzymes are shaded in black and gray, respectively. Conserved amino acid residues involved in the active side (Asp, His, and Ser) are indicated by asterisks below the alignments. Signal and pro-peptide are marked by a gray-letter type and underline, respectively.

4.4.3 Over-expression and purification of recombinant *Virgibacillus* AprX-SK37

To express recombinant *Virgibacillus* AprX-SK37, the *aprX-sk37* gene (nucleotide no. -39 to 1293, according to Figure 4.1) was amplified by PCR and cloned in to the *Xba*I and *Xho*I restriction sites of pET21d(+) vector such that the ribosomal binding site (RBS) and the start codon of the vector were replaced by those of the native gene. Notably, the RBS of the expression vector and *aprX-sk37* gene are identical (AAGGAG). The recombinant plasmid, pET-AprX-SK37 was expressed using *E. coli* BL21(DE3) as an expression host. *E. coli* BL21/pET-AprX-SK37 were cultured as described in materials and methods section. To optimize the induction condition, samples from various fractions including inclusion body, cytoplasm, periplasmic space, and culture supernatant at various time points after induction with different concentration of IPTG were analyzed by SDS-PAGE and activity assay. An *E. coli* BL21 carrying empty pET21d+ vector was used as a control, of which no significant enzymatic activity could be detected.

Upon induction with 0.1 mM IPTG, the enzyme could only be found in the cytoplasmic fraction as soluble protein. No enzyme could be detected in periplasmic extract or culture supernatant (data not shown). Neither lowering the temperature nor varying the concentration of IPTG dramatically affected expression level of the enzyme as judged by SDS-PAGE (data not shown). Routinely, ~16 mg of purified recombinant AprX-SK37 could be obtained from 1-liter culture. The C-terminal 6xHis-tagged enzyme could be purified from cleared cell lysate by one step affinity chromatography using Ni-NTA resin in native condition to apparent homogeneity as shown in Figure 4.3 (lane 1 and 2). The purified recombinant AprX-SK37 showed a

molecular mass of 46 kDa on SDS-PAGE, which corresponded well to the theoretical mass of 47 kDa. In the native PAGE analysis, proteolytic activity could be observed

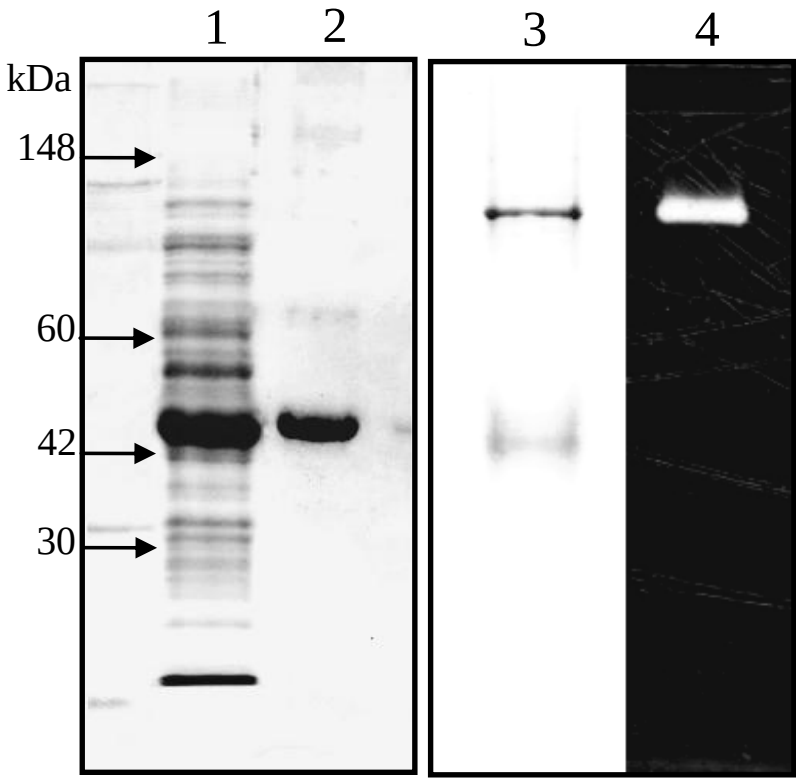


Figure 4.3 SDS-PAGE and casein-zymography of recombinant AprX-SK37. Samples were electrophoresed on a 10-15% SDS polyacrylamide gel (left panel) and 12% native polyacrylamide gel (right panel). After electrophoresis, the gel was stained with Coomassie Brilliant Blue (lanes 1-3) or detected for proteinase activity (lane 4). Lane 1, whole cell lysate obtained from *E. coli* BL21/pET-AprX-SK37; Lane 2, purified recombinant AprX-SK37; Lane 3, purified recombinant AprX-SK37 run in native polyacrylamide gel; Lane 4, purified recombinant AprX-SK37 in casein-zymogram gel. Standard molecular weights for SDS-PAGE were marked by arrows.

by casein-zymogram at the corresponding position on the Coomassie stained gel (Figure 4.3, lane 3 and 4). Zymographic assay in denaturation condition was impracticable due to enzyme sensitivity towards SDS, even though an in-gel refolding step was performed (data not shown).

4.4.4 Effects of pH and temperature

The optimal pH for caseinolytic activity of recombinant AprX-SK37 is 9.5. The enzyme is more active at pH 8.5 and 9.0 when using Tris-HCl buffer than Tris-glycine buffer. However, recombinant AprX-SK37 showed a maximal activity at pH 9.5 in the Tris-glycine buffer (Figure 4.4a). No activity was found at pH lower than 7.5, indicating an alkaline proteinase characteristic. Caseinolytic activity was maximum around 55°C (Figure 4.4b). The enzyme appeared to be activated when pre-incubated for 2 h at 25-30°C prior to the assay. Whereas, after 2 h of pre-incubation time, 50% and 0% residual activity was detected at 47°C and more than 50°C, respectively. Relative activities of the enzymes at various NaCl or $CaCl_2$ concentrations in 100 mM Tris-glycine, pH 9.5 at 55ºC for recombinant AprX-SK37, and in 100 mMTris-HCl, pH 8.0 at 60°C for subtilisin A under standard assay conditions were shown in Figure 4.4c,d. Enzyme stability was assayed by pre-incubating the enzymes at different concentrations of NaCl or $CaCl_2$ at 25ºC for 24 h and expressed as relative activities compared to the reaction at the corresponding concentrations of NaCl or $CaCl_2$ without pre-incubation.

4.4.5 Effects of NaCl and $CaCl_2$

Proteinases produced from *Virgibacillus* sp. SK37 have been shown to exhibit NaCl- and $CaCl_2$-activated characteristic (Sinsuwan, Rodtong, and Yongsawatdigul, 2007, 2008). Therefore, the effect of NaCl and $CaCl_2$ on the activity of recombinant

AprX-SK37 was studied. As demonstrated in Figure 4.4c the activity of recombinant AprX-SK37 increased slightly when NaCl was added, and showed maximum activity at 1 M. The activity decreased to 30% at 3 M NaCl. On the contrary, subtilisin A, which was used as a control, did not show NaCl-activated characteristic as indicated by a decrease in activity with increasing concentration of NaCl as shown in Figure 4.4c (square, solid line). Both recombinant AprX-SK37 and subtilisin A appeared to require at least 0.25 M NaCl to retain their original activities after incubation at 25°C for 24 h, as illustrated by Figure 4.4c (dashed lines). When the recombinant AprX-SK37 enzyme was first pre-incubated with 0.5-3 M NaCl for 24 h at 25°C and pH 9.5, the activity was enhanced up to 125% (Figure 4.4c, diamond, dashed lines). These NaCl-enhancing stability effects could not be observed in subtilisin A (Figure 4.4c, square, dashed lines). In addition to NaCl, recombinant AprX-SK37 also showed a $CaCl_2$-dependent characteristic as shown in Figure 4.4d (diamond, solid line). The enzyme was inactive in the absence of Ca^{2+} and required 15 mM $CaCl_2$ for its maximal activity. This is opposite to the Ca^{2+} independent characteristic of subtilisin A (Figure 4.4d, square, solid line). However, the activities of both recombinant AprX-SK37 and subtilisin A gradually decreased when the $CaCl_2$ concentration was higher than 30 mM. The remaining activities of and subtilisin A, when $CaCl_2$ reached 100 mM, were 66 and 73%, respectively. Both recombinant AprX-SK37 and subtilisin A showed moderate stability (50% remaining activity) when pre-incubated for 24 h at 25°C at their optimal pHs in the presence of all tested concentration of $CaCl_2$ as demonstrated in Figure 4.4d (dashed line).

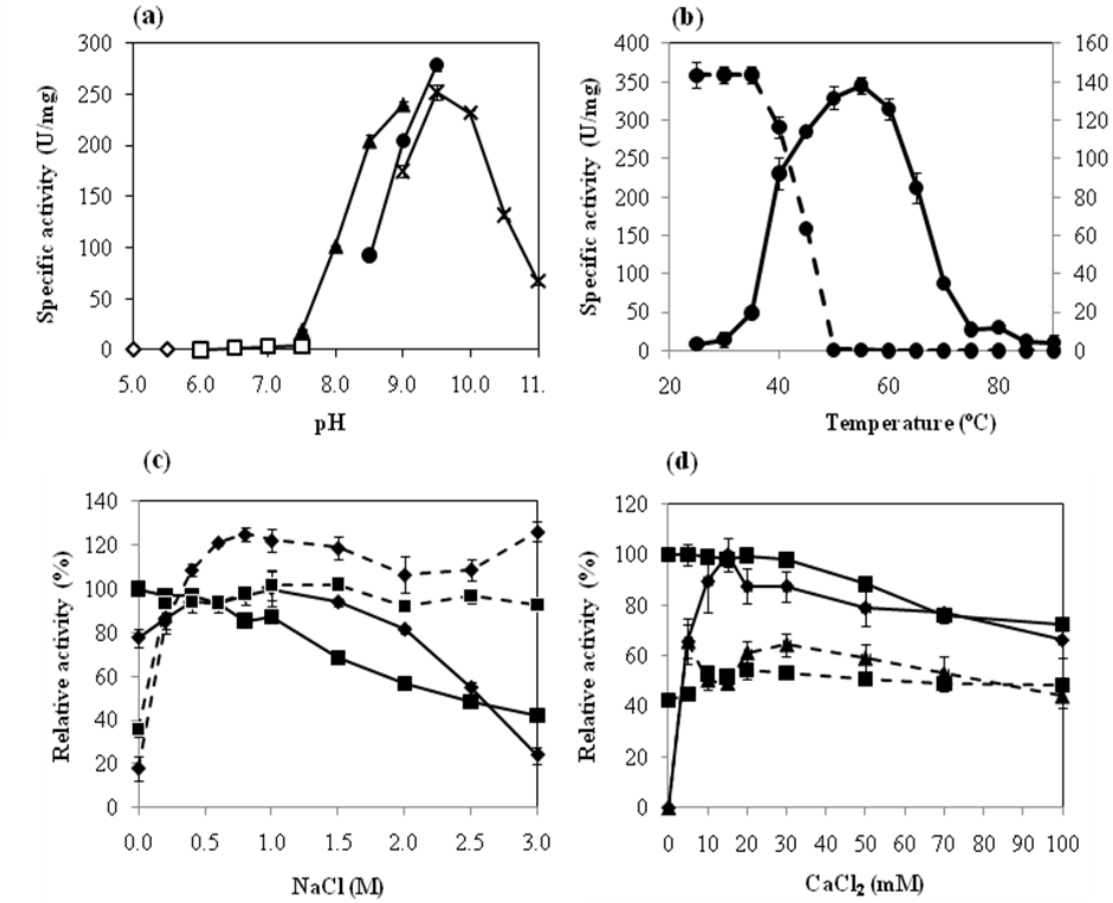


Figure 4.4 Effects of pH (a), temperature (b), NaCl (c), and $CaCl_2$ (d) on the activity (solid lines) and stability (dashed lines) of recombinant AprX-SK37. In (**a**), pH profile was carried out in acetate (-◊-), Tris-maleate (-□-), Tris-HCl (-▲-), Tris-glycine (-●-), and glycine (-×-) buffer. In (**b**) temperature profile and thermo stability of recombinant AprX-SK37 are illustrated by solid line and dashed line, respectively. The thermal stability was determined by pre-incubating the enzyme at various temperatures for 2 h and the remaining activity was measured at 55°C under standard assay conditions and reported as relative activity compared to a reaction assayed at 55°C under standard assay conditions without pre-incubation. In (c) and

(Figure 4.4 Continued)

(d), the effect of sodium and calcium on recombinant AprX-SK37 (-♦-) and subtilisin A (-■-) activities were shown.

4.4.6 Effects of reducing agents and various inhibitors

The effects of various inhibitors, reducing agents and detergent on recombinant AprX-SK37 activity were reported in Table 4.1. A marked inhibition by 10 mM PMSF suggested that this enzyme is a serine proteinase (Table 4.1). However, no inhibition by other serine proteinase inhibitors, *i.e.* TLCK, TPCK, leupeptin, and SBTI could be observed. A strong inhibition by EDTA indicated a metal dependent character of this enzyme. However, bestatin, a metalloproteinase inhibitor selective for aminopeptidases, had no inhibitory effect. The enzyme was very sensitive to detergent (SDS), and reducing agents such as DTT and β-ME, whereas the inhibitor E-64, which is an active-site titrant of cysteine proteinases, as well as IAA and NEM, did not affect enzyme activity (Table 4.1).

4.4.7 Thermal stability

Previous report has suggested that mono and divalent cations contribute to the stability of the native state and increase the activation energy of unfolding of subtilisin BPN′ (Alexander, Ruan, and Bryan, 2001), therefore the thermostability of the recombinant AprX-SK37 in the presence of various concentrations of Ca^{2+} and Na^{+} cations was investigated. At 40°C, the enzyme activity could be activated by calcium. In the standard reaction condition (5 mM $CaCl_2$) the enzyme activity was activated after incubation for 30 min and the optimal activity was retained for up to 2 h (Figure

Table 4.1 Effects of proteinase inhibitors and denaturants on recombinant AprX-SK37 activity using azocasein as a substrate[a].

Substance	Characteristic	Concentration	Remaining activity (%)
Leupeptin	Serine proteinase inhibitor	10 mM	100.7
STI		0.4 mg/ml	100.9
TLCK		10 mM	99.4
TPCK		10 mM	99.9
PMSF		1mM	89.2
		10mM	1.0
EDTA	Metalloproteinase	0.5mM	78.1
	inhibitor	1mM	64.2
		5mM	1.4
Imidazole		0.5 M	82.4
L-Histidine		0.5 M	108.1
IAA	Cysteine proteinase	0.1 M	98.0
NEM	inhibitor	0.1 M	114.2
E64		1 mM	99.2
	Aspartic proteinase		
Pepstatin A	inhibitor	1 mM	102.1
Bestatin	Amino peptidase inhibotor	2 mM	102.6
Urea	H-bond breaker	1M	20.8
		2M	0.0
SDS	Detergent	0.5%	0.0
		1.0%	0.0
DTT	Reducing agents	1mM	38.5
		10mM	2.3
		20mM	0.1
β-ME		1mM	64.7
		10mM	41.9
		20mM	14.1

[a] The enzymatic activity was determined at pH 9.5 and 55°C in the presence of 5 mg/ml substrate and 5 mM $CaCl_2$. The experiment was performed in triplicate, and standard errors from the means were within 15% of the reported values.

4.5a, diamond). When the concentration of $CaCl_2$ was increased to 15 mM (Figure 4.5a, triangle), the highest enzyme activation could be observed from 0.5-2 h. The activity declined, when the enzyme was incubated longer than 2 h in both 5 and 15 mM $CaCl_2$. Moreover, when 1 M NaCl was added in the presence of either 5 (Fig. 4.5a, square) or 15 mM (Fig. 4.5a, cross) $CaCl_2$, the stability of the enzyme was increased up to 4 h. (Figure 4.5a). The same phenomenon but at a higher rate of thermal inactivation was observed at 55°C (Figure 4.5b). The enzyme activity in the presence of only $CaCl_2$ (either 5 or 15 mM) decreased rapidly after the first 1 h of incubation and became almost completely inactivated after 2 h (Figure 4.5b, diamond and triangle). In contrast, addition of 1 M NaCl could significantly lower the rate of thermal inactivation as compared to those of without NaCl. The residual activity of the reaction containing 1 M NaCl plus 5 or 15 mM $CaCl_2$ at 2 h was approximately 45 and 60%, of their original activity, respectively (Figure 4.5b, cross and square). These observations suggested that NaCl played an important role in exerting enzyme thermal stability while $CaCl_2$ was only responsible for activating enzyme activity.

4.4.8 Effect of oxidizing agent

Stability against oxidizing agent is one of the most attractive properties of bacterial proteinases. Therefore, H_2O_2 was used as the representative of oxidizing agent to test recombinant AprX-SK37 oxidant stability. As shown in Figure 4.6, the enzyme was fully active in the presence of as much as 5% H_2O_2 (black dot). However, the enzyme activity gradually decreased when it was pre-incubated for 30 min with increasing concentration of H_2O_2 (white dot). No activity was found in a reaction pre-incubated with 10% H_2O_2 (data not shown).

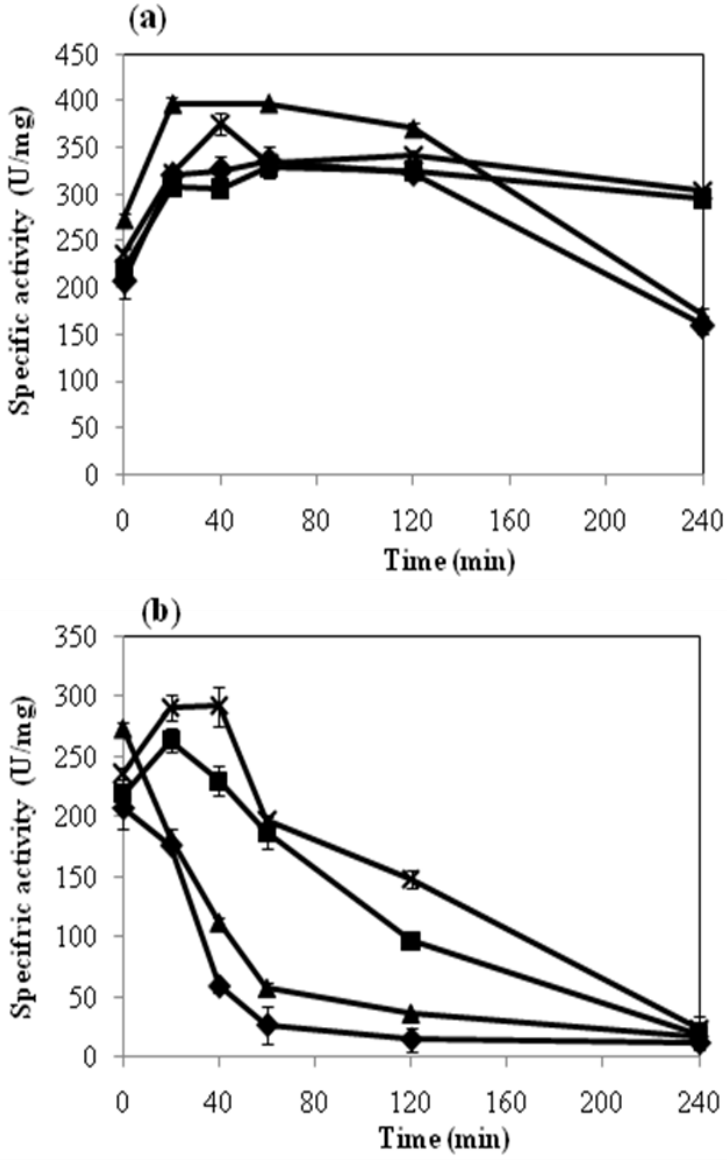


Figure 4.5 Effects of NaCl and $CaCl_2$ on thermostability of recombinant AprX-SK37. The enzyme was pre-incubated in either 40 (a) or 55 °C (b) in 100 mM Tris-Glycine pH 9.5 containing; 5 mM $CaCl_2$ (-♦-): 5 mM $CaCl_2$ and 1 M NaCl (-■-): 15 mM $CaCl_2$ (-▲-): 15 mM $CaCl_2$ and 1 M NaCl (-×-). Residual enzyme activities were measured at 55°C under standard assay conditions.

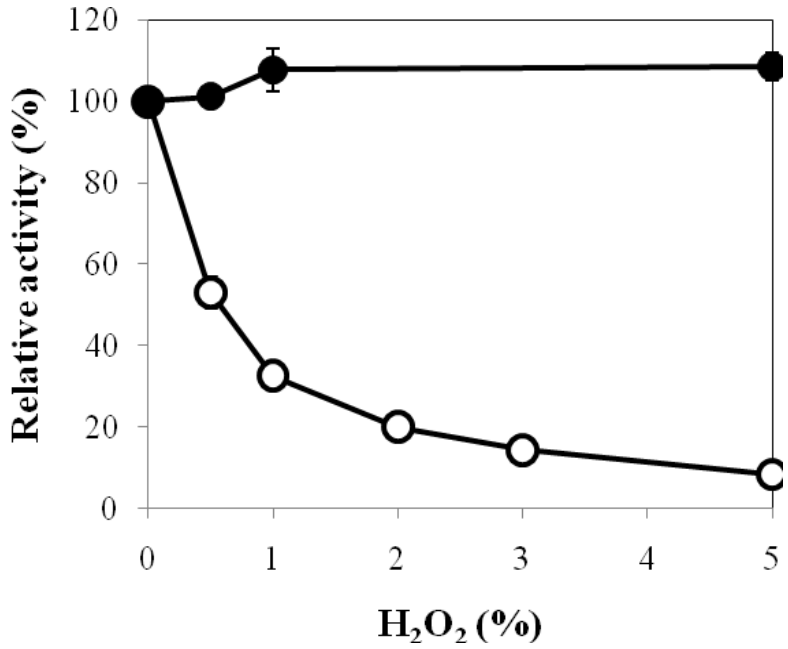


Figure 4.6 Stability of recombinant AprX-SK37 towards H_2O_2. The activity was determined at various H_2O_2 concentrations after pre-incubation for 0 (-●-) and 30 (-○-) min. The activity without H_2O_2 was taken as 100%.

4.4.9 Phylogenetic tree analysis

The predicted amino acid sequence of AprX-SK37 was aligned and compared with those of well-characterized proteinases in subtilase superfamily retrieved from MEROPS (Rawlings et al., 2010) and SwissPort (Jain et al., 2009) databases. Oxidant stable proteinases (OSPs) from various strains were also included in the alignment (indicated by black dots). The phylogenetic tree was inferred using the neighbor-joining algorithm, as shown in Figure 4.7. The subtilase superfamily was phylogenetically grouped into six clusters, *i.e.* subtilisins, thermitase, proteinase K, lantibiotic peptidases, kexin, and pyrolysins (indicated by solid square brackets), which represent subtiliase family S8A, B, C, D, E, and F, respectively, (Siezen and

Leunissen, 1997). From our analysis, the AprX-SK37 was not categorized into any of these families, but shared the same cluster with other AprXs, which was positioned apart from other subtilase families (dashed square bracket). Notably, within the AprX cluster, the AprX-SK37 is relatively distant from other AprXs. Most OSPs that are produced from alkaliphilic bacilli are clustered together in the tree (dashed square bracket), *i.e.* E-1, LP-Ya, and SD-521 from *B. cohnii* DSM 6307, NP-1 from *B. halmapalus* DSM8723, KP-43 from *Bacillus* sp. KSM-KP43 (Saeki et al., 2007). However, some of OSPs such as sapB (AM748727) from *B. pumilus* CBS (Jaouadi et al., 2008), LD1 (AB085752) from *Bacillus* sp. KSM-LD (Saeki et al., 2003), and I-52 from *B. clausii* I-52 (Chang, Joo, and Choi, 2006; Joo et al., 2003), were not located in the OSPs cluster, but appeared to be members of the subtilisin family.

Distribution of enzymes exhibiting oxidant stability across the subtilase phylogenic tree as illustrated in Figure 4.7 is a curiosity and prompted us to investigate the evolutionary distances among the OSP producing microorganisms. Herein, 16S rRNA sequences of *B. licheniformis* DSM13 (subtilisin Carlsberg producer) and *B. clausii* KSM-K16 (M-proteinase producer) were also included in the alignment as non-OSPs producer strains. The phylogenetically aligned 16S rRNA gene sequences were grouped into four clusters (Figure 4.8).

The evolutionary distances of E-1, LP-Ya, SD-521, NP-1, and KP-43 producers are consistent with previous report (Saeki et al., 2007). The LD1 producer was located into the same cluster as E-1, LP-Ya, and SD-521 producers, whereas sapB was grouped in the same cluster as non-OSPs producers. Interestingly, the AprX-SK37 producer was distantly related to other strains in the tree.

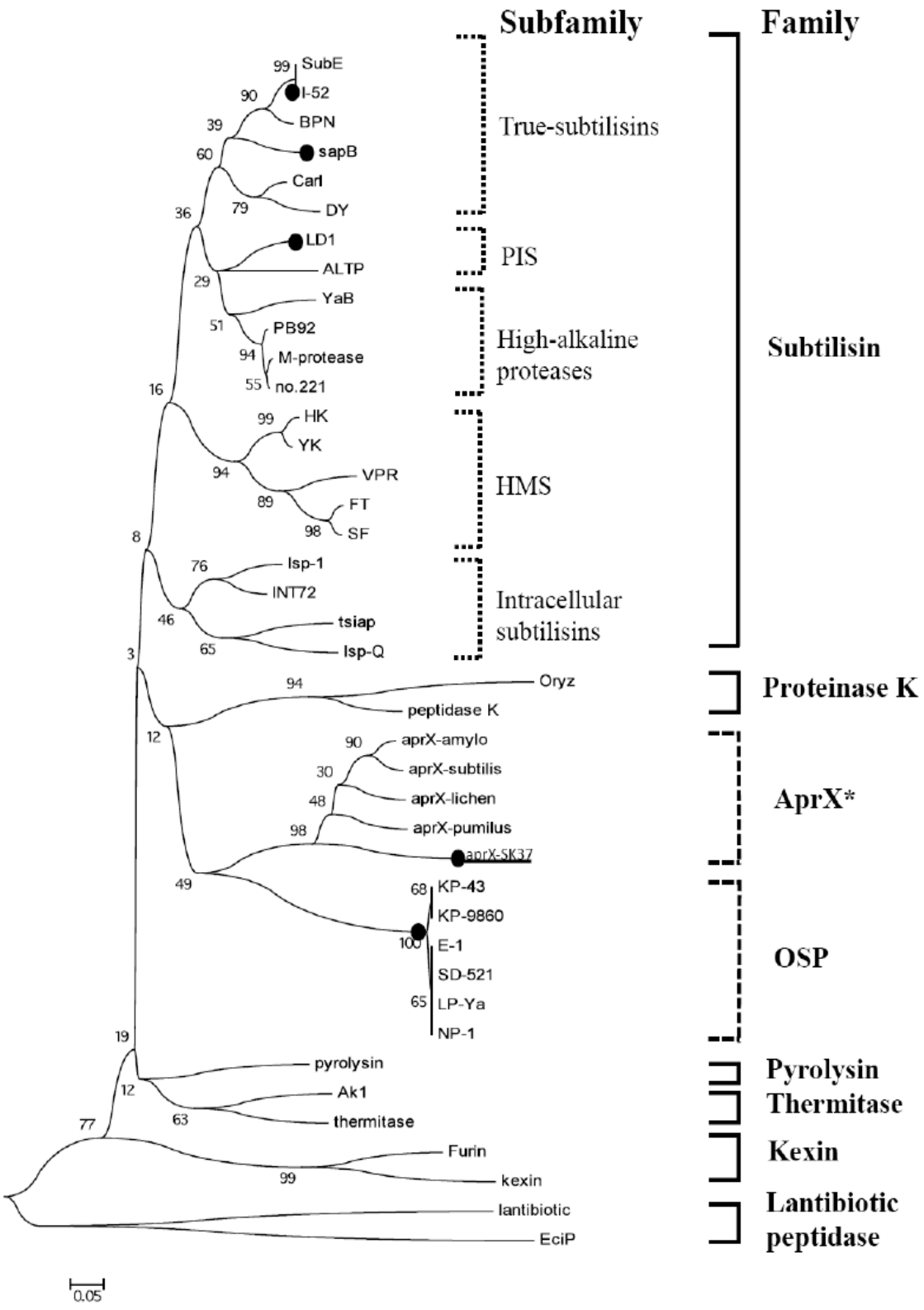


Figure 4.7 (see next page)

Figure 4.7 Phylogenetic tree of enzymes in subtilase superfamily based on amino acid sequence similarity in catalytic domains. The bootstrap consensus tree inferred from 1,000 replicates (Felsenstein, 1985) was taken to represent the evolutionary history of the sequence analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated sequence clustered together in the bootstrap test (1,000 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 66 positions in the final dataset. Sequences aligned: subtilisin E (SubE) (P04189) from *Bacillus subtilis* 168; I-52 from *B. clausii* I-52 (Chang et al., 2006); BPN´ (P00782) from *B. amyloliquefaciens*; sapB (CAO03040) from *B. pumilus*; Carlsberg (Carl) (P00780) from *B. licheniformis*; DY (P00781) from *B. subtilis* DY; ALTP (BAF34115) from *Alkaliphilus transvaalensis*; LD1 (BAD02409) from *Bacillus sp.* strain KSM-LD1; YaB (P20724) from *Bacillus sp.* strain YaB; PB92 (P27693) from *B. alkalophilus* PB92; M-proteinase (Q99405) from *B. clausii* KSM-K16; no. 221 (P41362) from *Bacillus* sp. strain 221; HK (BAD36786) from *Bacillus* sp. strain D6; YK (BAD36788) from *Bacillus* sp. strain Y; VPR (AAA22881) from *B.*

(Figure 4.7 Continued)

subtilis; FT (BAD21124) from *Bacillus* sp. strain KSMKP43; SF (BAD21125) from *Bacillus* sp. strain KSM-9865; Isp-1 (AAA22557) from *B. subtilis*; INT72 (P29139) from *B. polymyxa* 72; tsipa (BAA13418) from *Thermo-actinomyces* sp.; Isp-Q (BAA07142) from *Bacillus* sp. strain NKS-21; KP-43 (BAB55674) from *Bacillus* sp. strain KSM-KP43; KP- 9860 (BAB21266) from *Bacillus* sp. strain KSM-KP9860; E-1 (BAB21265) from *Bacillus* sp. strain D 6; SD- 521 (AB046402) from *Bacillus* sp. strain SD-521; LP-Ya (AB046404) from *Bacillus* sp. strain Y; NP-1 (BAB21269) from *Bacillus* sp. strain NV1; Oryzin (Oryz) (P35211) from *Aspergillusflavus*; Peptidase K (P06873) from *Engyodontium album*; pyrolysin (P72186) *Pyrococcus furiosus*; Thermitase 12 (P04072) from *Thermoactinomyces vulgaris*; subtilisin AK1 (AK1) from *Geobacillus stearothermophilus*; kexin (P13134) from *Saccharomyces cervisiae*; furin (P09958) from *Homo sapiens*; LasPlantibiotic (lantibiotic) (Q48854) from *Lactobacillus sakei*; EciPlantibiotic (EciP) (O54221) from *Staphylococcus epidermidis*. AprX–SK37 (HM587897). Subtilase superfamily is classified as recommended by Siezen et al (1997) (square brackets). Subtilisin family is subdivided into true subtilisins, phylogenetically intermediate subtilisins between true subtilisins and highalkaline proteinase (PIS), high-alkaline proteinases, high-molecular-mass subtilisins (HMS), and intracellular subtilisins (Saeki et al., 2007) (dotted square brackets). The AprX-SK37 from *Virgibacillus* sp. SK37 in this study (underlined) is a member of the novel AprX family. Black dots indicate oxidant stable proteinases (OSPs). Newly classified alkaline serine proteinase-

(Figure 4.7 Continued)

X (AprX) family (dashed square bracket), proposed in this study is marked with a star.

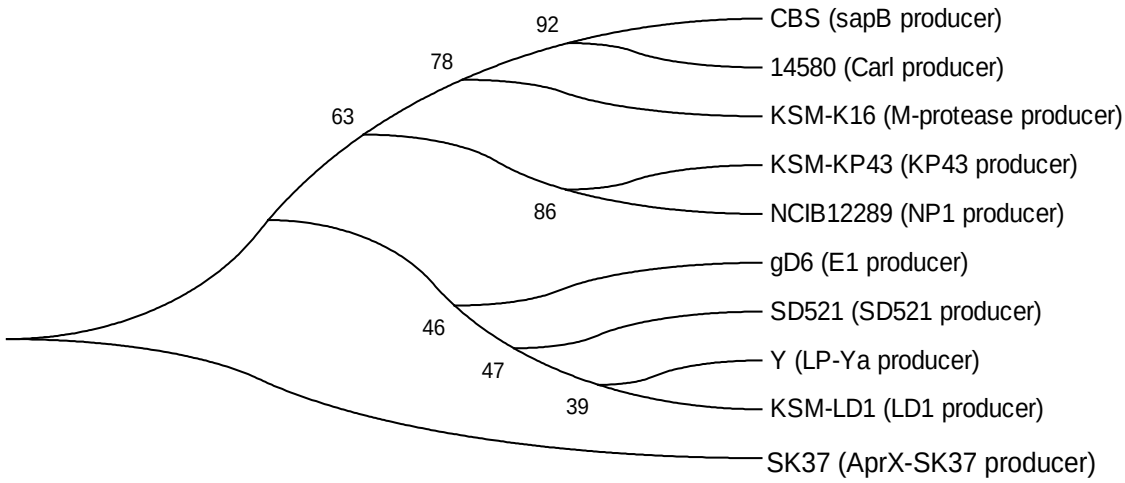


Figure 4.8 Phylogenetic tree of 16S rRNA sequences of OSPs producing bacilli. The tree was constructed by the method described in Fig 7. *Bacillus licheniformis* ATCC 1480 (14580) (163119169: sequence 9909-11453) (carl-producer) and *B. clausii* KSM-K16 (KSM-16) (56961782: sequence 11675-13227) (M-proteinase-producer) were selected as non-OSPs producer strains. Sequences aligned; *Bacillus* sp. KSM-KP43 (KSM-KP43) (AB055093), *Bacillus* sp. NCIB12289 (NCIB12289) (AB055097), *Bacillus* sp. SD521 (SD521) (AB055096), *Bacillus* sp. Y (Y) (AB055095), *Bacillus* sp. D6 (gD6) (AB05594), *Bacillus* sp. KSM-LD1 (KSM-LD1) (AB212915), *B. pumilus* strain CBS (CBS) (EF113957), *Virgibacillus* sp. SK37 (SK37) (DQ910840).

4.5 Discussion

This is the first report on cloning, expression and biochemical characterization of bacterial AprX proteinase. Previous report (Valbuzzi, Ferrari, and Albertini, 1999) of this enzyme from *B. subtilis* focused on bioinformatic analysis and regulation of gene expression, which indicated that this enzyme is an intracellular serine proteinase, which is expressed in stationary phase and not essential for either growth or sporulation. The AprX-SK37 enzyme in this study was derived from a moderately halophilic bacterium, *Virgibacillus* sp. SK37, which was isolated from Thai fish sauce that was saturated with 26-28% of NaCl. Therefore, the NaCl-activated property of this enzyme was not unexpected as it has been shown that the cell membrane of halophilic bacteria is incapable to completely prevent the internal increment of sodium (Padan and Krulwich, 2000). Intracellular Na^+ concentrations of moderately halophilic bacteria growing in media containing 2-4 M NaCl were found between 0.13 and 3.52 M with an average of 1.37 M (Ventosa, Nieto, and Oren, 1998). Our enzyme can be activated and stabilized by 1 M NaCl (5.8% w/v) and tolerate up to 3 M NaCl (17.4% w/v). A similar result was found in a moderately halophile, *B. subtilis* FP-133, that produced a halotolerant intracellular cysteine aminopeptidase exhibiting optimal activity at approx. 1 M NaCl (Setyorini, Kim, Takenaka, Murakami, and Aoki, 2006). For AprX-SK37, to the best of our knowledge, this is the first report on halotolerant intracellular serine proteinase. DNA sequence analysis of *aprX-sk37* indicated that the gene is regulated by two putative σ_A promoters (P_L and P_S), which have been suggested to function as primary and secondary promoters of *B. subtilis aprX*, respectively (Valbuzzi et al., 1999). The open reading frame (ORF) of *Virgibacillus aprX-sk37* is slightly smaller than that of *B. subtilis aprX* and lacks the alternative

start condon TTG, at position 76-78, Met/Leu_{26} in AprX numbering of *B. subtilis*. The predicted MM and p*I* of recombinant AprX-SK37 are in the same range as those of AprXs from *Bacillus* species (MM of 42.2-48.9 kDa and p*I* at 4.9-6.1). The key different between AprX-SK37 and other AprXs and some other intracellular proteinases is its N-terminal domain, which showed no similarity to those of known proteins in the database. It has been suggested that the N-terminal domain of subtilases acts as an intramolecular chaperone or pro-peptide that mediate enzyme folding (Eder, Rheinnecker, and Fersht, 1993; Takagi and Takahashi, 2003; Shinde, Liu, and Inouye, 1997); nevertheless the role of this domain on AprX-SK37 activity is unclear. The localization of AprX from *B. subtilis* in cytoplasm has been previously reported (Kodama et al., 2007). The recombinant AprX-SK37 in this study is found in cytoplasm, however the precise localization of the wild-type enzyme is unknown. The enzyme appeared to be active without cleavage of any pro-peptide. Its putative function as chaperone for correct enzyme folding or as a regulatory domain remains to be explored.

The benefit of isolating novel enzyme from a DNA expression library instead of PCR cloning using degenerated primers is that the gene encoding active enzyme can be readily obtained. This advantage is very important especially for *Bacillus* subtilases, which often form inactive insoluble aggregates when produced in *E. coli* expression systems (Chiang, Chen, Chao, and Tzen, 2005; Choi, Chang, Jae Maeng, and Kim, 2004; Zhang, Xiao, Peng, Wang, Bai, and Zhang, 2005). The recombinant AprX-SK37 appeared to be located as active soluble protein in the cytosolic compartment, none of them were found as inactive insoluble inclusion bodies or secreted into the periplasm or culture media. The fact that the enzyme was first

identified from *E. coli* colonies that showed halo zone during genomic DNA library screening could be the result of membrane leakage after prolonged growth for more than 24 h. Despite a thorough screening of the library that covered approx. 70 depth of the library, no other proteinase genes could be isolated. This could be because other enzymes formed inactive aggregate or the enzyme activity required higher sodium or calcium, which was not included in the screening condition. The recombinant AprX-SK37 could be highly expressed using pET system. This system has been shown to be efficient for the expression of various *Bacillus* hydrolytic enzymes (Yamabhai, Emrat, Sukasem, Pesatcha, Jaruseranee, and Buranabanyat, 2008). The canonical ribosomal binding site in this vector is identical to that of *Virgibacillus aprX-sk37*. The high level of expression could be because of synonymous codon usage, which leads to fast translation rate in *E. coli* expression system (Ermolaeva, 2001).

Biochemical characterization revealed that recombinant AprX-SK37 is a Ca^{2+} dependent alkaline serine proteinase and moderate thermophile with optimal pH and temperature at 9.5 and 55°C, respectively. The enzyme requires calcium to function with maximum activity at 15 mM $CaCl_2$. Calcium-dependent characteristic is well known for members of subtilases superfamily, of which conserve calcium binding sites are essential for enzyme stability and/or activity (Siezen and Leunissen, 1997). In nature, calcium may play a physiological role in *aprX-sk37* gene regulation as it has been shown that a massive accumulation of calcium during sporulation is found in most bacilli in order to respond to adverse environmental changes (Rosen, 1987) and in *B. subtilis*, AprX is expressed in late stationary phase under the control of various transition state regulatory genes (Valbuzzi et al., 1999). The high sensitivity of recombinant AprX-SK37 to reducing agent suggested that disulfide bonds are

essential for structural stability. Primary amino acid sequence revealed that AprX-SK37 contains 8 cysteines (Cys). This rich in Cys is unique for enzymes in the AprX family since Cys is usually absent from subtilases in Gram-positive bacteria (Siezen and Leunissen, 1997; Valbuzzi et al., 1999). In general, subtilases are inactivated by hydrogen peroxide due to the oxidation of methionine (Met), located next to the catalytic serine (Met_{222} in subtilisin BPN´ numbering), preventing the formation of tetrahedral intermediate during transition state of their hydrolytic reaction (Bott, Ultsch, Kossiakoff, Graycar, Katz, and Power, 1988; Stauffer and Etson, 1969). The AprX-SK37 possesses Met_{368} next to catalytic Ser_{367}, similar to other subtiliases (Figure 4.1) and other OSPs (Haddar et al., 2009; Jaouadi et al., 2008; Joo et al., 2003; Saeki et al., 2000, 2002, 2003). The actual mechanism of oxidant stability is still unclear. In the study of KP-43, an OSP enzyme, Nonaka, Fujihashi, Kita, Saeki, Ito, and Miki (2001) found that the oxidation of Met altered enzyme substrate specificity rather than inactivated the enzyme. In addition, according to KP-43 crystal structure, Saeki et al. (2007) proposed an assumption that the rate of Met-oxidation in KP-43 may be lower than those of other subtilases, probably due to longer distance between the Met residue in the catalytic vicinity and the oxyanion hole or the presence of neighbor Met. Hence, the oxidant-stable ability is strongly dependent on enzyme conformation and structural integrity, which is intrinsic to each enzyme. The oxidant stability as well as halotolerant and moderately thermophilic properties of recombinant AprX-SK37 should make this novel enzyme attractive for biotechnological applications in the future.

Phylogenetic analysis revealed that AprX-SK37 belongs to the superfamily of subtilases or subtilisin-like proteinases. This result is consistent with the family

analysis of *B. subtilis* AprX (Valbuzzi et al., 1999). Moreover, our analysis indicated that the AprXs are clustered in the distinct group, of which AprX-SK37 is relatively different from other AprXs. This could be because recombinant AprX-SK37 possesses a unique characteristic, *i.e.* oxidant-stable and moderately salt-activated, which is not found in other AprXs. However, according to our phylogenetic tree, AprX-SK37 was not categorized into the OSPs group. Notably, phylogenetic analysis of OSPs reports little evolutionary relationship among this group of enzymes. Our phylogenetic analysis indicated that most of OSPs does not belong to any of the subtilisin families previously described in bacteria or fungi, but appears to belong to a different cluster (designated OSP in Figure 4.7). This finding is similar to the phylogenetic analysis of a subtilisin-like proteinase gene family in the grass endophytic fungus *Epichloë festucae* (Bryant, Schardl, Hesse, and Scott, 2009). Hence, OSPs should not be classified as the subfamily of subtilisins as previously suggested (Saeki et al., 2000, 2002, 2007). Phylogenetic analysis of various OSPs producers suggested that the sources of OSPs encompass bacilli from diverse evolutionary faith, which could be the result of different mechanisms of oxidant stabilization.

4.6 Conclusions

Taken together the results of biochemical characteristics and phylogenetic analysis as discussed above, a new family of subtilases superfamily, namely alkaline serine proteinase-X (AprX) or subitlase S8G was proposed, according to subtilase classification of Siezen and Leunissen (1997). In addition, the classification of OSPs has to be re-defined.

4.7 References

Alexander, P. A., Ruan, B., and Bryan, P. N. (2001). Cation-dependent stability of subtilisin. **Biochemistry.** 40: 10634-10639.

Bott, R., Ultsch, M., Kossiakoff, A., Graycar, T., Katz, B., and Power, S. (1988). The three-dimensional structure of *Bacillus amyloliquefaciens* subtilisin at 1.8 A and an 548 analysis of the structural consequences of peroxide inactivation. **J. Biol. Chem.** 263: 7895-7906.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. **Anal. Biochem.** 72: 248-254.

Bryant, M. K., Schardl, C. L., Hesse, U., and Scott, B. (2009). Evolution of a subtilisin-like proteinase gene family in the grass endophytic fungus *Epichloe festucae*. **BMC Evol. Biol.** 9**:**168.

Chang, C., Joo, H. S., and Choi, J. W. August (2006). ***Bacillus* sp. I-52, a detergent- and oxidant-stable haloalkalophilc proteinase therefrom, a gene encoding the same and a use thereof**. PCT patent WO 2006/088325.

Chiang, C. J., Chen, H. C., Chao, Y. P., and Tzen, J. T. (2005). Efficient system of artificial oil bodies for functional expression and purification of recombinant nattokinase in *Escherichia coli*. **J. Agric. Food Chem.** 53: 4799-4804.

Choi, N. S., Chang, K. T., Jae Maeng, P., and Kim, S. H. (2004). Cloning, expression, and fibrin(ogen)olytic properties of a subtilisin DJ-4 gene from *Bacillus* sp. DJ-4. **FEMS Microbiol. Lett.** 236: 325-331.

Eder J., Rheinnecker, M, and Fersht, A. R. (1993). Folding of subtilisin BPN´: Role of the pro-sequence. **J. Mol. Biol.** 233: 293-304

Ermolaeva, M. D. (2001). Synonymous codon usage in bacteria. **Curr. Issues Mol. Biol.** 3: 91-97.

Gupta, R., Beg, Q. K., and Lorenz, P. (2002). Bacterial alkaline proteinases: molecular approaches and industrial applications. **Appl. Microbiol. Biotechnol.** 59: 15-32.

Gupta, R., Gupta, K., Saxena, R. K., and Khan, S. (1999). Bleach-stable, alkaline proteinase from *Bacillus* sp. **Biotechnol. Lett.** 21: 135-138.

Haddar, A., Agrebi, R., Bougatef, A., Hmidet, N., Sellami-Kamoun, A., and Nasri, M. (2009). Two detergent stable alkaline serine-proteinases from *Bacillus mojavensis* A21: purification, characterization and potential application as a laundry detergent additive. **Bioresour. Technol.** 100: 3366-3373.

Jain, E., Bairoch, A., Duvaud, S., Phan, I., Redaschi, N., Suzek, B. E., Martin, M. J., McGarvey, P., and Gasteiger, E. (2009). Infrastructure for the life sciences: design and implementation of the UniProt website. **BMC Bioinformatics.** 10**:** 136.

Jaouadi, B., Ellouz-Chaabouni, S., Rhimi, M., and Bejar, S. (2008). Biochemical and molecular characterization of a detergent-stable serine alkaline proteinase from *Bacillus pumilus* CBS with high catalytic efficiency. **Biochimie.** 90: 1291-1305.

Joo, H. S., Kumar, C. G., Park, G. C., Paik, S. R., and Chang, C. S. (2003). Oxidant and SDS-stable alkaline proteinase from *Bacillus clausii* I-52: production and some properties. **J. Appl. Microbiol.** 95: 267-272.

Kobayashi, T., Hakamada, Y., Adachi, S., Hitomi, J., Yoshimatsu, T., Koike, K., Kawai, S., and Ito, S. (1995). Purification and properties of an alkaline

proteinase from alkalophilic *Bacillus* sp. KSM-K16. **Appl. Microbiol. Biotechnol.** 43: 473-481.

Kodama, T., Endo, K., Sawada, K., Ara, K., Ozaki, K., Kakeshita, H., Yamane, K., and Sekiguchi, J. (2007). *Bacillus subtilis* AprX involved in degradation of a heterologous protein during the late stationary growth phase. **J. Biosci. Bioeng.** 104: 135-143.

Kumar, C. G., and Takagi, H. (1999). Microbial alkaline proteinases: from a bioindustrial viewpoint. **Biotechnol. Adv.** 17: 561-594.

Maurer, K. H. (2004). Detergent proteinases. **Curr. Opin. Biotechnol.** 15: 330-334.

Nawong, S. (2006). **Isolation, selection, and identification of proteinase-producing bacteria from fish sauce fermentation to be used as starter cultures.** M.S. thesis. Suranaree University of Technology, Nakhon-Ratchasima.

Ness, J. E., Welch, M., Giver, L., Bueno, M., Cherry, J. R., Borchert, T. V., Stemmer, W. P., and Minshull, J. (1999). DNA shuffling of subgenomic sequences of subtilisin. **Nat. Biotechnol.** 17: 893-896.

Nonaka, T., Fujihashi, M., Kita, A., Saeki, K., Ito, S., and Miki, K. (2001). Crystallization and preliminary X-ray diffraction studies of a novel alkaline serine proteinase (KP-43) from alkaliphilic *Bacillus* sp. strain KSM-KP43. **Acta. Crystallogr. D. Biol. Crystallogr.** 57: 717-718.

Ogawa, A., Sumitomo, N., Okuda, M., Saeki, K., Kawai, S., Kobayashi, T., and Ito, S. (2003). Nucleotide and deduced amino acid sequences of a high-molecular-mass subtilisin from an alkaliphilic *Bacillus* isolate. **Biochim. Biophys. Acta.** 1624: 109-114.

Okuda, M., Sumitomo, N., Takimura,Y., Ogawa, A., Saeki, K., Kawai, S., Kobayashi, T., and Ito, S. (2004). A new subtilisin family: nucleotide and deduced amino acid sequences of new high-molecular-mass alkaline proteinases from *Bacillus* spp. **Extremophiles.** 8: 229-235.

Padan, E., and Krulwich, T. A. (2000). **Sodium stress.** ASM Press, Washington DC.

Rao, M. B., Tanksale, A. M., Ghatge, M. S., and Deshpande, V. V. (1998). Molecular and biotechnological aspects of microbial proteinases. **Microbiol. Mol. Biol. Rev.** 62: 597-635.

Rawlings, N. D., Barrett, A. J., and Bateman, A. (2010). MEROPS: the peptidase database. **Nucleic Acids Res.** 38: 227-233.

Rosen, B. P. (1987). Bacterial calcium transport. **BBA-REV Biomembranes.** 906: 101-110.

Saeki, K., Hitomi, J., Okuda, M., Hatada, Y., Kageyama, Y., Takaiwa, M., Kubota, H., Hagihara, H., Kobayashi, T., Kawai, S., and Ito, S. (2002). A novel species of alkaliphilic *Bacillus* that produces an oxidatively stable alkaline serine proteinase. **Extremophiles.** 6: 65-72.

Saeki, K., Magallones, M. V., Takimura, Y., Hatada, Y., Kobayashi, T., Kawai, S. and Ito, S. (2003). Nucleotide and deduced amino acid sequences of a new subtilisin from an alkaliphilic *Bacillus* isolate. **Curr. Microbiol.** 47: 337-340.

Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H., and Horikoshi, K. (2000). Novel oxidatively stable subtilisin-like serine proteinases from alkaliphilic *Bacillus* spp.: enzymatic properties, sequences, and evolutionary relationships. **Biochem. Biophys. Res. Commun.** 279: 313-319.

Saeki, K., Ozaki, K., Kobayashi, T., and Ito, S. (2007). Detergent alkaline proteinases: enzymatic properties, genes, and crystal structures. **J. Biosci. Bioeng.** 103: 501-508.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Mol. Biol. Evol.** 4: 406-425.

Sambrook, J., Maniatis, T., and Fritsch, E. F. (2001). **Molecular cloning: a Laboratory Manual,** vol. 1-3. Cold Spring Harbor Laboratory Press, New York.

Setyorini, E., Kim, Y. J., Takenaka, S., Murakami, S., and Aoki, K. (2006) Purification and characterization of a halotolerant intracellular protease from *Bacillus subtilis* strain FP-133. **J. Basic Microbiol.** 4: 294-304

Shinde, U. P., Liu, J. J., and Inouye, M. (1997). Protein memory through altered folding mediated by intramolecular chaperones. **Nature.** 389: 520-522.

Siezen, R. J., de Vos, W. M., Leunissen, J. A., and Dijkstra, B. W. (1991). Homology modeling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. **Protein. Eng.** 4: 719-737.

Siezen, R. J., and Leunissen J. A. (1997). Subtilases: the superfamily of subtilisin-like serine proteinases. **Protein Sci.** 6: 501-523.

Sinsuwan, S., Rodtong, S., and Yongsawatdigul, J. (2008). Characterization of Ca^{2+}-activated cell-bound proteinase from *Virgibacillus* sp. SK37 isolated from fish sauce fermentation. **LWT-Food Sci. Technol.** 41: 2166-2174.

Sinsuwan, S., Rodtong, S. and Yongsawatdigul, J. (2007). NaCl-activated extracellular proteinase from *Virgibacillus* sp. SK37 isolated from fish sauce fermentation. **J. Food Sci.** 72: 264-269.

Stauffer, C. E., and Etson, D. (1969). The effect on subtilisin activity of oxidizing a methionine residue. **J. Biol. Chem.** 244: 5333-5338.

Takagi, H., and Takahashi, M. (2003). A new approach for alteration of protease functions: pro-sequence engineering. **Appl. Microbiol. Biotechnol.** 63: 1-9.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. **Mol. Biol. Evol.** 24: 1596-1599.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. **Nucleic Acids Res.** 25: 4876-4882.

Valbuzzi, A., Ferrari, E., and Albertini, A. M. (1999). A novel member of the subtilisin-like proteinase family from *Bacillus subtilis*. **Microbiology.** 145 (Pt 11):3121-3127.

Ventosa, A., Nieto, J. J., and Oren, A. (1998). Biology of moderately halophilic aerobic bacteria. **Microbiol. Mol. Biol. Rev.** 62:504-544.

Waschkowitz, T., Rockstroh, S., and Daniel, R. (2009). Isolation and characterization of metalloproteinases with a novel domain structure by construction and screening of metagenomic libraries. **Appl. Environ. Microbiol.** 75: 2506-2516.

Yamabhai, M., Emrat, S., Sukasem, S., Pesatcha, P., Jaruseranee, N., and Buranabanyat, B. (2008). Secretion of recombinant *Bacillus* hydrolytic enzymes using *Escherichia coli* expression systems. **J. Biotechnol.** 133: 50-57.

Zhang, R. H., Xiao, L., Peng,Y., Wang, H. Y., Bai, F, and Zhang, Y. Z. (2005). Gene expression and characteristics of a novel fibrinolytic enzyme (subtilisin DFE) in *Escherichia coli*. **Lett. Appl. Microbiol.** 41:190-195.

Zhao, H., and Arnold, F. H. (1999). Directed evolution converts subtilisin E into a functional equivalent of thermitase. **Protein Eng.** 12: 47-53.

CHAPTER V

COMPLETE GENOME SEQUENCE OF *VIRGIBACILLUS* SP. SK37, A MODERATELY HALOPHILIC BACTERIUM ISOLATED FROM THAI FISH SAUCE FERMENTATION

5.1 Abstract

Virgibacillus sp. SK37 (SK37) is a Gram-positive moderately halophilic bacterium isolated from Thai fish sauce containing 26-28% NaCl. This bacterium has a potential to use as a starter culture for fish sauce production. The objectives of this study were to determine the complete genome sequence and analyze genes involved in adaptation to a saline protein-rich niche of fish sauce. The genome consists of chromosome (3.8 Mbp) and three plasmids (4.1, 4.4, and 12.1 Kbp) comprising of 3,823 protein coding sequences. Genes responsible for central dogma's general transfer, transpoases, and insertion elements shows homologous to *Bacillus* species. No intact prophage was found indicating least importance in lateral gene transfer. Analysis of the genome sequence shows efficient uptake and utilization of amino acids and pathways for the biosynthesis of fish sauce flavors from amino acid derivatives. The genome of SK37 possesses genes for regulation and uptake of inorganic ions and osmoprotectants for adaptation to high saline environments. Phylogenetic analysis of CcpA, a global nutrient regulator, suggests relatedness in

adaptive nutrient stress response among halophilic and alkaliphilic bacteria. The lack of genes encoding Hpr and ScoC, the proteinase repressors, in SK37 genome may explain the high proteinase production ability of this bacterium. Comparative genomic analysis indicates that SK37 shows highly conserved orthologs to *Bacillus* species, with higher similarity to the Gram positive moderately halophilic *Oceanobacillus iheyensis*. This first genome sequence of *Virgibacillus* will be useful in understanding life in saline protein-rich environments and applications in fish sauce fermentation industry.

Keywords: Moderately halophilic bacterium, Genome sequencing, high saline, protein-rich niche, haloadaptation, *Virgibacillus*, Fish sauce

5.2 Introduction

During the past decade a large number of extreme and moderate halophiles have been isolated from hypersaline environments. These microorganisms are divided according to their ability to grow in the presence of NaCl. Moderate halophiles are defined as organism growing optimally between 3% and 15% NaCl (Kushner, 1978). Bacteria able to grow in the absence of NaCl, as well as in the presence of relatively high NaCl concentration are designated as halotolerant (or extremely halotolerant if growth extents above 15%). Bacteria that require at least 12% salt (optimal growth at 20%) are designated as extreme halophiles (Johnson, Lanthier, and Gochnauer, 1986). Mechanisms of salt adaptation has typically been studied in extremely halophilic bacteria and halophilic archaeas, the adaptations that have been examined include tolerance to high concentration of NaCl in the cytoplasm, specialized salt-requiring proteins, and proton and ion pumping systems (Kushner, 1978; Ventosa, Nieto, and

Oren, 1998). Recently, research has been initiated on moderately halophilic organisms due to their ability to grow over a wide range of NaCl concentrations. These bacteria hold promise in various industrial processes containing vastly different NaCl concentrations such as in fermented foods, bioreactor systems, biotechnological processes, production of biopolymers, pharmaceuticals, and degradation of toxic compounds. (Gomes and Steiner, 2004; Hough and Danson, 1999; Margesin and Schiner, 2001; Van den Berg, 2003; Ventosa, and Nieto, 1995; Ventosa et al., 1998).

Moderately halophilic bacteria constitute a heterogeneous physiological group of microorganisms belonging to different genera. However, most species are Gram-negative or Gram-positive aerobic or facultative anaerobic bacteria (Ventosa et al., 1998). They have been isolated from a wide range of habitats including saline lakes, saltern ponds, desert and hypersaline soil, and salted foods (Ventosa et al., 1998). In salted foods, these organisms have been isolated from salted and dried fish, fish intestines, fish sauces, and soy sauce mashes. It is interesting to note these ecological niches are protein-rich, which suggests that these organisms have evolved an efficient proteolytic enzyme system for obtaining amino acids from the environmental proteins and energy from amino acids. Fish sauce is an excellent environment to isolate halophilic proteolytic bacteria and study their ecology. Fish sauce is produced via a natural fermentation of anchovies in concentrated brine (25-28% NaCl) at pH 5.5-6.0. Hydrolysis of fish protein is accomplished by the joint action of bacterial and endogenous fish proteolytic enzymes (Haard and Simpson, 1994; Yongsawatdigul, Rodtong, and Raksakulthai, 2007). The product is widely consumed as a condiment in Southeast and East Asian countries. It is a clear, brown, salty liquid that contains peptides, amino acids, calcium, and vitamin B and provides a strong and complicated

umami taste, even stronger than that provided by soy sauce (Park et al., 2001). The fermentation process is dependent upon proteolytic enzymes and the high NaCl content significantly reduces proteolytic activity, hence it is a long fermentation taking approximately 12 months to complete (Saisithi, 1994). The long fermentation time adds significant cost to fish sauce manufacture and limits the growth of this industry. A number of research groups have isolated halophilic bacteria from fish sauce with high proteolytic activity, with the intent of developing starter cultures for this fermentation. The bacteria isolated include *Halobacillus thailandensis* (Chaiyanan, Maugel, Huq, Robb, and Colwell, 1999), *Filobacillus* sp RF2-5 (Hiraga et al., 2005), and *Halobacillis* sp. SR5-3 (Namwong, Hiraga, Takada, Tsunemi, Tanasupawat, and Oda, 2006) from Thai fish sauces as well as *Bacillus* species from Vietnamese and Korean fish sauce (Kim, Nishiyama, Mura, Tokue, and Arai, 2004; Nagano and To, 2000; Setyorini, Takenaka, Murakami, and Aoki, 2006; Tran and Nagano, 2002; Uchida et al., 2004). However, the genetics and physiology of these organisms remains poorly defined. Nawong (2006) has isolated 308 microorganisms from Thai fish sauce and began screening these organisms for industrially relevant phenotypes. This screening resulted in the identification of a moderately halophilic, highly proteolytic bacterium, designated *Virgibacillus* sp. SK37 (GenBank/NCBI no. DQ910840) (Nawong, 2006). This organism is a Gram-positive, endospore-forming, rod-shaped, aerobic bacterium, which grows over a wide range of pHs (range pH 4 to 11; optimum pH 6.5-7.5), temperatures (range 20 to 45°C; optimum 30-40 °C) and NaCl concentrations (range 0 to 25% NaCl; optimum 5-10% NaCl). This strain has been successfully used as a starter culture for the production of fish sauce, resulting in 4 months of total fermentation time, as determined by amino acid profile

(Yongsawatdigul et al., 2007). Proteinases produced from this strain have been characterized and determined to be halotolerant and retain appreciable activity at high NaCl concentrations (Phrommao, Rodtong, and Yongsawatdigul, 2010; Sinsuwan, Rodtong, and Yongsawatdigul, 2007, 2008). In this work, the complete genome sequence of *Virgibacillus* sp. SK37, the first report of genome sequence from Genus *Virgibacillus*, was determined. Additionally, a comparative analysis of this genome with non-halophilic, moderately halophilic, and extremely halophilic bacteria is provided, with a special emphasis on genes thought to encode enzymes relevant to the fish sauce fermentation.

5.3 Materials and Methods

5.3.1 DNA sequencing and genome assembly

Genomic DNA of *Virgibacillus* sp. SK37 was prepared using MasterPure^{TM} Gram positive DNA purification kit (Epicentre Biotechnologies, Madison, WI) and its quality was confirmed by spectrometry and agarose gel electrophoresis. The DNA was submitted for genome sequencing at the Innovation centre (McGrill University and Génome Guébec, Guebéc, Canada) using 454 pyrosequencing GS-FLX platform (454 Life Science, Roche Diagnostics Crop.). About 97% of the reads (369 bp of median read length) were assembled into 50 contigs with sizes ranging from 5.8 to 590 Kbp (comprising 3.79 Mbp in total) resulting in 40X coverage of genomic size. Independently, an optical map was generated from the *Virgibacillus* sp. SK37 genome at OpGen (Gaithersburge, MD) using NcoI. The assembled contigs were ordered using the optical map, and the remaining gaps were closed using standard PCR and DNA sequencing methods. During assembly, the positions of rRNA operons and all

repeated DNA regions in the genome were verified by PCR amplification across repeat and the size of the product was confirmed via agarose gel electrophoresis. The plasmid contigs were confirmed to be circular by PCR amplification and DNA sequencing.

5.3.2 Genome analysis

The complete genome sequence was concurrently annotated by RAST (http://rast.nmpdr.org/) (Aziz et al., 2008) and IMG (http://img.jgi.doe.gov/cgi-bin /pub/main.cgi) pipeline servers without manual curation. The annotated data from IMG was used as the primary source for functional annotation of protein products, computation, comparative analysis, and orthologous gene grouping, while the RAST annotation was used for confirmation. Genomes of related *Bacillus* species and their physiological and biogeographical descriptions were obtained from the IMG database and are listed in supporting material Table 1 (please see Appendix). All gene product names used in this study are derived from the homologs of *Bacillus subtilis* subsp. subtilis str 168 as presented in the NCBI database (http://www.ncbi.nlm.nih.gov), unless otherwise noted. Central metabolic pathways were predicted using the KEGG database; while other pathways, such as volatile compounds biosynthesis, were constructed according to information obtained from relevant publications (indicated in text). Other general genome analysis and visualization were conducted in Artemis (Rutherford et al., 2000) and Unipro UGENE (http://ugene.unipro.ru/index.html).

5.3.3 Compositional difference of amino acids

To evaluate differences in amino acid usage in non-halophiles, moderate halophilies, and extreme halophiles, protein sequences containing signal peptide for secretion were selected from complete genome sequences of *Bacillus* species in the

IMG database (supporting material Table 1). Herein, *Salinibacter ruber* (Mongodin et al., 2005) and *Halobacterium* sp. NRC-1 (Ng et al., 2000) were selected as representatives for extreme halophiles. Amino acid composition was calculated by Protein composition tool kits (http://www.vivo.colostate.edu/molkit/procomp/index .html) and the abundance of each amino acid was converted into percentage of frequency (%F). The %F values of each amino acid of the bacteria were divided by the respective %F of *E. coli* as a well-defined reference for standardization (Fukuchi, Yoshimune, Wakayama, Moriguchi, and Nishikawa, 2003) and expressed as F-ratio. Standardized F-ratio of <1.0 and >1.0 indicate under-representation and over-representation of amino acids of the bacteria and are presented in the Heat-map format using red or green colored blocks, respectively.

5.3.4 Phylogenetic tree construction

Deduced Amino acid sequences of CcpA, CodY, and GlnR from *ccpA*, *codY*, and *glnR* genes of related *Bacillus* species and their 16S rRNA (supporting material Table 1) were obtained from the IMG and NCBI databases. To construct the phylogenetic trees, each group of sequences, in turn, was multiple-aligned by Clustal*W* (Thompson, Gibson, Plewniak, Jeanmougin, and Higgins, 1997) and the phylogenetic tree was generated by MEGA*4* (Tamura, Dudley, Nei, and Kumar, 2007). The trees were inferred by bootstrapping neighbor-jointing method with competed gap deletion.

5.4 Results and discussions

5.4.1 General features

The genome of *Virgibacillus* sp. SK37 (SK37) is composed of a single circular chromosome (3,821,074 bp) and three plasmids: pSK37064 (4,094 bp), pSK37387 (4,494 bp), and pSK37042 (12,068 bp). The presence of these three plasmids was confirmed by agarose gel electrophoresis (data not shown). Average G+C contents of chromosome and pSK37042 are comparable at 37.6 and 38.2%, respectively, while significantly lower G+C contents of 30.9 and 34.2% were found in pSK37387 and pSK37064, respectively (Figure 5.1 and Table 5.1). 3,823 protein coding sequences (CDSs) were identified covering 84.6% of the genome. These CDSs have an average length of 816 bp with the density of 1.023 genes per kb. In the chromosome, 55.2 and 44.8% CDSs are transcribed on the positive and negative strand, respectively, in the direction of replication (Figure 5.1), which has been observed in various low G+C Gram-positive bacterial genomes (Boekhorst, Wels, Kleerebezem, and Siezen, 2006). Of which, 2,887 (74.1%) were assigned predicted biological functions, 2,685 (68.9%) based upon clusters of orthologous groups (COGs), 621 (15.9%) are proteins in paralog clusters, and 1045 (26.8%) are proteins connected to KEGG pathways.

5.4.2 Replication origin and terminus

The origin of chromosome replication (*oriC*) in bacteria contains a conserved organization, a sharp change of G-C skew, and a cluster of four or more DnaA boxes (McLean, Wolfe, and Devine, 1998; Moriya, Imai, Hassan, and Ogasawara, 1999; Yoshikawa and Wake, 1993). On the basis of G-C skew (G-C/G+C) and genome organization, the likely SK37's replication origin was identified. The genome organization of *oriC* flanking regions exhibits co-localization of *gidA*, *trmE*, *yidC*, *rnpA*, *rpmH*, *dnaA*, *dnaN*, *recF*, *gyrB*, and *gyrA*; these genes are mainly homologous to *Bacillus* species which are oriented in the same manner (Yen, Lin, Hung, Choy,

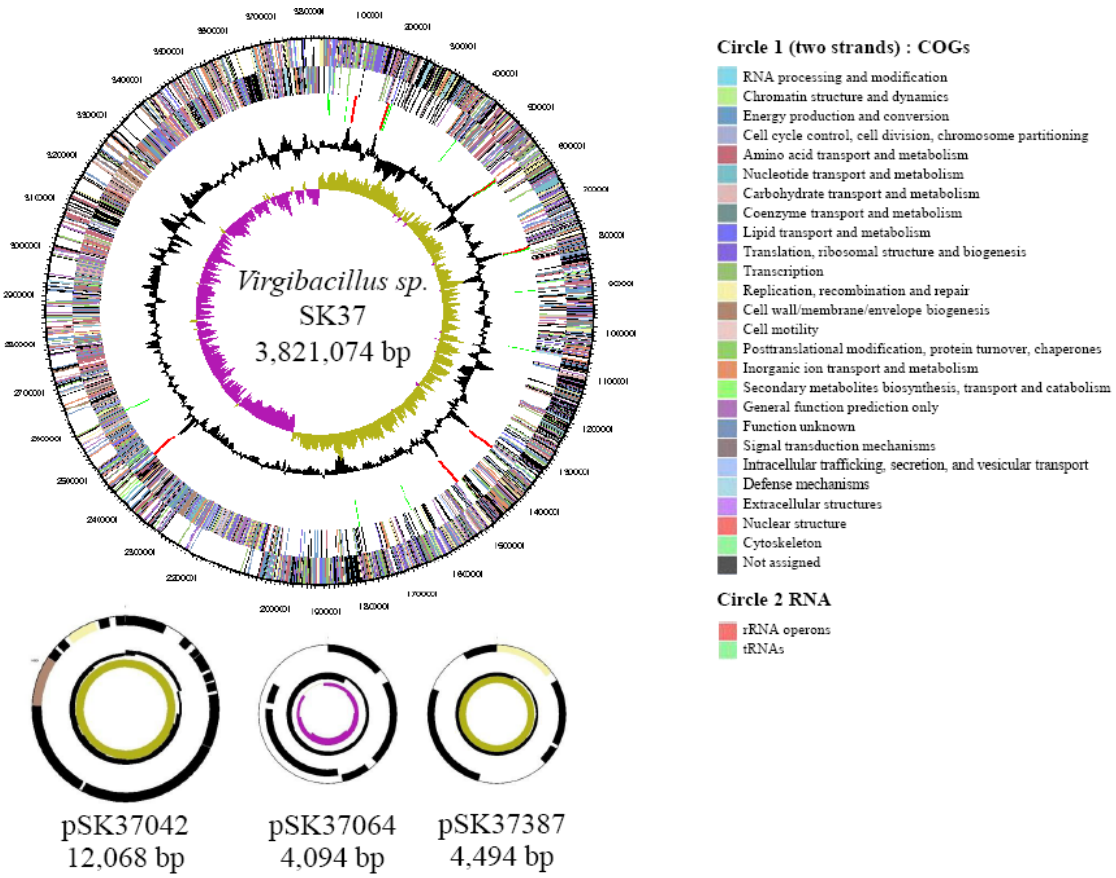


Figure 5.1 Genome atlas of *Virgibacillus* sp. SK37 and its plasmids. The color coding of the genomic feature represents different COG categories. The last two inner cycles in each atlas represent %G+C content and GC screw, respectively.

Weng, and Tseng, 2002). Moreover, 11 and 4 of DnaA boxes with >80% of conserved TTATCCACA sequences were found in the upstream and downstream noncoding regions, respectively, flanking the *dnaA* gene. Finally, a sharp transition of the G-C skew was detected immediately upstream of the *dnaA* gene. On the basis of these observations, the first nucleotide of *dnaA* was assigned as the first base of the SK37 genome. Bacterial circular chromosomes replicate in a bidirectional manner and meet

at the terminus region (*terC*), approximately opposite *oriC* (Bachmann, 1990; Kuempel, Henson, Dircks, Tecklenburg, and Lim, 1991). At the end of replication, the XerC and XerD recombinases convert the circular chromosome dimers into monomers at the *dif* site (Blakely, Colloms, May, Burke, and Sherratt, 1991; Clerget, 1991; Kuempel et al., 1991). The *terC* site was located at 2.05 Mb by the sharp G-C skew transition (Figure 5.1a). This region is located approximately diametrically opposite to *oriC* (190°). A *dif*-like sequence with 92% sequence identity to *B. subtilis* was found in the SK37 genome, starting at nucleotide no. 2,048,267. Additionally, XerC- and XerD-like coding sequences were also detected in SK37. In *B. subtilis*, a replication termination protein (*rtp*) also plays a role in termination of DNA replication (Kaul, Mohanty, Sahoo, Patel, Khan, and Bastia, 1994). It forms two dimers and binds two inverted repeat region (IRI and IRII) at the termination site (Carrigan, Haarsma, Smith, and Wake, 1987). However, *rtp* was not detected in SK37, the absence of *rtp* has also been reported in *B. halodurans*, *B. licheniformis*, and *Oceanobacillus iheyensis*. It is likely that these organisms lost *rtp* after diverging from *B. subtilis* and contain an orthologs having the same function as *rtp*.

In general, bacterial plasmids replicons are divided into two types: rolling-circle and theta-type replicons (Alpert, Crutz-Le Coq, Malleret, and Zagorec, 2003; Ehrlich et al., 1991; Meijer et al., 1998). Both replicons require a replication initiation protein (*rep*) (Sorvig, Skaugen, Naterstad, Eijsink, and Axelsson, 2005). The theta-type plasmids are generally larger and have a relatively limited host-range, while plasmids containing a rolling-circle replicon have a relatively broad host range and are frequently exchanged between bacterial species (Sorvig et al., 2005). All three plasmids in SK37 genome show no overall similarity to other sequenced plasmids.

Table 5.1 General features of the *Virgibacillus* sp. SK37 genome

General features	Chromosome	Plasmids		
		pSK37387	pSK37064	pSK37042
Total size, bp	3,821,074	4,494	4,094	12,068
DNA coding, %	84.6	69.23	74.96	89.59
G+C content,%	37.63	30.89	34.2	38.17
Gene total number	4,010	5	5	27
Protein coding gene	3,957	5	5	27
with function prediction	2,867 (71.5%)	3	2	4
without function prediction	1,090 (27.2%)	2	3	23
RNA coding gene				
RNA operons	**7**	**0**	**0**	**0**
16S-23S-5S	4	-	-	-
16S-23S-5S-tRNA	1	-	-	-
tRNA-16S-23S-5S-tRNA	1	-	-	-
16S-tRNA-23S-5S-tRNA	1	-	-	-
tRNAs	**53**	**0**	**0**	**0**
Insertion elements				
Phage-associated genes	**14**	**0**	**0**	**0**
PBSX	2	-	-	-
Others	12	-	-	-
Transpoase genes	**18**	**0**	**0**	**0**
IS*652*	8	-	-	-
IS*605*	4	-	-	-
IS*654*	3	-	-	-
IS*657**	1	-	-	-
IS*66*	1	-	-	-
IS*21*	1	-	-	-
Recombinases	**8**	**1**	**1**	**1**
Tyrosine superfamily	3	-	-	-
Serine superfamily	4	-	-	-
Phage origin	1	-	-	1
Others	0	1	1	-

*Interrupted gene

Based upon nucleotide homology, the products encoded by *rep* (called Rep) from pSK37064 and pSK37387 have highest similarity to RepL (Firmicute plasmid replication protein) and Rep1 families, respectively. These plasmids seem to replicate via rolling-circle mechanism since their homologs are rolling-circle replication

proteins from *Amoeba proteus* plasmid pAP3.9 (76% identity) (ACE62910) and *B. mycoides* DX plasmid pBMYdx (71% identity) (CAB88026) for pSK37064, and from *Marinococcus halophilus* plasmid pLP1 (72% identity) (AAC45761) for pSK37387. In addition, each plasmid also contains a single recombinase-coding sequence in the Mob-Pre family, which is responsible for conjugative plasmid mobilization (Priebe and Lacks, 1989). Their MoB-like enzymes are homologous to relevant proteins from *Geobacillus stearothermophilus* plasmid pTB19 (50% identity) (AAA98307) for pSK37064 and from *B. cereus* H3081.97 plasmid pH308197-11 (38% identity) (ACI30479) for pSK37387. The average G+C content (32.5% for the plasmids vs. 37.6% for the chromosome) and divergent replicon, suggest these plasmids were obtained by horizontal gene transfer.

In pSK37042, no obvious *rep* gene was observed, however, a *repA*-like sequence with high homology to enzymes in copG family was observed. Enzymes in this family are involved in the regulation of plasmid copy number by controlling the synthesis of the RepB initiator protein (del Solar, Acebo, and Espinosa, 1995; del Solar and Espinosa, 1992). However, a *repB*-related gene was not detected in the SK37 genome. Interestingly, one hypothetical protein in this plasmid is related to a phage-related Rep. BLAST search suggested that the first 102 deduced amino acids (35% sequence coverage) of this protein are homologous to putative phage replication proteins from *Staphylococcus aureus* (ACY10755) and *Staphylococcus* phage phiMR25 (BAG48116) both with 46% identity. Alternately, pSK37042 may possess partitioning genes in its replicon and use host replication machinery for replication. This suggests a limited host range for this plasmid and implies a close stable

relationship between the plasmid and host (van Passel, Bart, Luyf, van Kampen, and van der Ende, 2006).

5.4.3 Transcription and translation

Genes encoding five subunits (α, β, β′, ω, and δ) of the bacterial RNA polymerase were identified in the SK37 genome. The multiunit RNA polymerase binds to dissociable initiation sigma (σ) factors to initiate transcription (Gross et al., 1998). Fifteen σ factors belonging to σ^{70} (14 genes) and σ^{54} families (one gene) were detected in the SK37 genome. Function of the σ^{70} family members was inferred by homology to *B. subtilis* σ^{70} family members. All σ factors required for spore formation (σ^{E}, σ^{F}, σ^{G}, σ^{H}, and σ^{K}), one for flagellar biosynthesis (σ^{28}), and one for expression of stress responses (σ^{B}) are present in SK37. However, only one σ factor (σ^{W}), compared to 7 in *B. subtilis*, for expression of extracytoplasmic/extreme function (ECF) was observed (Helmann, 2002). Interestingly, five σ^{70} factor genes of the SK37 are unique compared to other bacilli and their functions are unknown. Homology analysis indicated that 3 of these have moderate homology to σ^{A} (for general transcription, heat shock responses, and stationary phase degradative enzyme synthesis), σ^{D} (flagellar mobility, chemotaxis, and cell wall-related function control), and genes in ECF group (which are related to cell surface transport controls and extracellular stress signal responses) (Helmann, 2002). Hence, these functions are coordinated with various stress responses and therefore they may be important in regulation of genes responsible for salt tolerance in SK37.

The SK37 genome contains seven copies of rRNA operon which have only a very limited number of sequence polymorphisms. A total of 53 tRNA species were identified representing all 20 amino acids. They are organized into 8 clusters plus 7

single genes. Out of the 8 tRNA cluster, 5 are associated with rRNA operons (Table 5.1). The same pattern of gene organization was found in *B. subtilis* and *B. halodurans* (16S-23S-5S-tRNA) (Takami et al., 2000) and in *Oceanobacillus iheyensis* HT (tRNA-16S-23S-5S-tRNA and 16S-tRNA-23S-5S-tRNA) (Takami, Takaki, and Uchiyama, 2002). Aminoacyl tRNA synthetases are present for all amino acids except asparagine and glutamine. It is likely that Asn- and Gln-tRNA are produced by transamidation of misacylated Asp- and Glu-tRNA, respectively, by the activity of aspartyl/glutamyl-tRNA amidotransferase (Freist, Gauss, Ibba, and Soll, 1997). Homologs of *gatA*, *gatB*, and *gatC* genes encoding three intact domains of this enzyme were detected in the SK37 genome.

5.4.4 Insertion elements, prophage, and atypical regions

The SK37 genome contains 18 genes encoding putative transpoases (Tpases) and inactivated derivatives, including 17 complete copies and one interrupted copy, categorized into 6 families (Table 5.1). Most of these Tpases have sequence similarity to other *Bacillus* Tpases. The predominant Tpases in the SK37 genome (IS*652* and IS*605*) are also the dominant species in some other moderately halophilic bacteria, such as *Oceanobacillus iheyensis* (10 of 21 genes) and *B. selenitireducens* (16 of 34 genes), indicating that they may have been distributed by horizontal gene transfer between moderately halophilic organisms. A Tpase element in IS*21* family was determined to have homology with IS*21-558* from *Staphylococcus aureus* (61-71% identity). It contains overlapping CDSs of *istAS* and *istBS* and imperfect terminal inverted repeats (Kehrenberg and Schwarz, 2006). Among *Bacillus* species, this IS element had been reported only in *Geobacillus thermodenitrificans* (with 23-45% identity). Two classes of transposon-like coding-regions were found, including

Tn554- and Tn*7*-like elements. The Tn554-like element (at nucleotide no. 295,717) is 3.7 kb and has three core transposition genes (*tnpABC*) (Bastos and Murphy, 1988; Murphy, Huwyler, and de Freire Bastos Mdo, 1985) resembling Tn554-like element in *B. cereus* (62-75% identity). However, unlike *B. cereus*, a cadmium-resistant operon was found at 3′ end of *tnpC*. Two divergent (58-64% sequence similarity between these two copies) Tn*7*-like elements are present in the SK37 genome, each are 6.4 kb and they are located approximately 56.5 kb apart (nucleotide no. 262,957 and no. 325,888, designated Tn*7*-copy1, and -copy2, respectively). The core genes constituting the Tn*7* family are present (*tnsABCD*). In general, Tn*7* requires GlmS for translocation and the *glmS* is usually located nearby the Tn*7* element (Bainton, Kubo, Feng, and Craig, 1993; Oppon, Sarnovsky, Craig, and Rawlings, 1998; Rogers, Ekaterinaki, Nimmo, and Sherratt, 1986). However, no *glmS* homolog was detected in SK37. Additionally, Tn*7* elements typically contain Tn*7*L/R flanking sequences, these sequences were not observed in either Tn*7*-copy1 or Tn*7*-copy2. These results suggest that Tn*7*-copy1 and Tn*7*-copy2 are inactive.

The SK37 genome was analyzed for prophage-related sequences by identification of homologs to the large subunit of a highly conserved prophage terminase (*ygaT*) and analysis of proximal regions (Casjens, 2003; Rey et al., 2004) (Table 5.1). In the SK37 chromosome, *ygaT* was detected at nucleotide no. 1,831,025 and no. 3,638,262 and proximal regions of 23.4 kb and 17.7 kb, respectively, encoding phage related proteins (Table 5.1), but no intact prophages were identified. The phage-related CDSs included phage tail protein, phage portal protein SPP1 Gp6-like, phage minor capsid protein2, phage tail tape measure protein (TP901 family), putative prophage psiM100 structural protein, capsid protein gpC, SA bacteriophage

11 (Mu50B) element, and recombinases (*xerD* and DNA invertase Pin homolog). An operon encoding membrane proteins (*yukEDCA*, and *yueB*) that serve as a bacteriophage SPP1 receptor were detected at nucleotide no. 813,330, this receptor is present in a variety of Gram-positive species (Sao-Jose, Baptista, and Santos, 2004). Additionally, non-clustered phage remnants such as phage PBSX-like, prophage λ-Ba04 elements, and other putative phage-associated genes were identified dispersedly along the chromosome.

Analysis of satellite DNA (micro/minisatellites and tandem repeats) was conducted to explore possible remnants of horizontal gene transfer in the SK37 genome, 7 satellite (excluding microsatellites) regions having 12-23 bp repeats with 2-3 repeats per region were found randomly along the genome, it is likely these regions are the result of IS and/or phage integration. No obvious CRISPRs were detected.

5.4.5 Cell envelop component, protein export, and transportation system

Prokaryotic cell envelop composition is directly related to the environment(s) in which a given organism inhabits. *Virgibacillus* sp. SK37 is capable of growth over a wide range of pHs (range from pH 4 to pH 11; optimum pH 6.5 to pH 7.5) and temperatures (range from 20 to 45°C; optimum 30-40°C). Alkaliphilic *Bacillus* species (*i.e. B. halodurans* and *Oceanobacillus iheyensis*) contain acidic polymers such as teichronopeptide, a copolymer of glutamate and polyglucuronic acid, which contributes to pH-homeostatis during their growth in alkaline pH environments (Takami et al., 2000; Takami et al., 2002). The main composition in cell wall proteoglycan of the SK37 and other *Virgibacillus* species is *meso*-diaminopimelic acid, which is common present in most bacteria (Nawong, 2006). When compared to neutrophilic *B. subtilis*, the SK37 genome contains all essential genes for

peptidoglycan biosynthesis and four (*tagD, tagH, tagG*, and *ispD*) of the 15 genes required for teichoic acid biosynthesis in *B. subtilis*, but lacks all the genes for teichuronic acid and teichronopeptide biosynthesis. The lack of core genes for the biosynthesis of teichoic acid was also reported in thermophilic *Geobacillus kaustophilus* and some of mesophilic *Bacillus* species (Takami et al., 2004). Interestingly, the genes required for isopentenyl diphosphate (IPP) biosynthesis via the mevalonate pathway (*pksG*, *Mva1*, *Mvak1*, *Mvak2*, and *mvaD*) were detected in SK37. IPP is the precursor of all isoprenoids in all organisms. SK37 colonies and cell envelop fractions are light-pink, suggesting that SK37 is capable of pigment biosynthesis (data not showed), something that has not previously been reported in *Bacillu*s (Laurent, 2006). Additionally, SK37 genome contains a gene (*idi*) encoding isopentenyl-diphosphate δ-isomerase, the presence of which has not previously been reported in *Bacillus* species. This enzyme converts IPP into dimethylally diphosphate, a precursor for long isoprenoid chain production in archaeal membrane lipid synthesis (Payandeh, Fujihashi, Gillon, and Pai, 2006). Other archaeal membrane synthesis associated genes are also present in the SK37 genome, including dimethylally-transferase, geranyltransferase, geranylgeranyl pyrophosphate synthase, and (s)-3-o-geranlygeranlyglyceryl phosphate synthase which mediate the production of (s)-2,3-di-o-geranlygeranlyglyceryl phosphate and its variants, key membrane lipid components in archaea.

In Gram-positive bacteria, most secreted proteins are synthesized with a N-terminal signal peptide and utilize the general secretory (Sec) pathway. Excluding *secB*, genes encoding all components of a functional Sec-protein complex (*secAEGDFY*, *yajC*, *yidC*, *ftsY*, and *ffh*) are present in SK37. Additionally, two signal

peptidase genes (*spaseI* and *spaseII*) encoding enzymes for the removal of signal peptides are present in SK37. In addition to the Sec pathway, bacteria can utilize the twin-arginine translocation (Tat) pathway (require two consecutive arginines in the signal sequence) for protein export. SK37, like most of *Bacillus* species, contains two *tat* gene orthologs (*tatA* and *tatC*), but lacks genes (*tatB*, *tatE*, and *tatD*) encoding three other components of the Tat pathway. Curiously, the SK37 genome contains 4 and 2 paralogs of *tatA* and *tatC*, respectively, suggesting that the Tat pathway may be functional in SK37and that other components of this system have not been identified. Lipoproteins are a diverse class of secreted bacterial proteins anchored to the cell surface. SK37 contains *lspA*, which encodes a lipoprotein signal peptidase, indicating that this bacterium has a pathway for the secretion of lipoproteins. Moreover, like other Gram-positive bacteria, SK37 contains sortase encoding genes (two copies), a key enzyme for membrane protein translocation via the sorting pathway. Overall, SK37 has a repertoire of pathways for the secretion of proteins similar to that described in other *Bacillus* species.

There are 1,056 and 1,097 CDSs encoding signal peptides and transmembrane domains, respectively, in the SK37 genome. These CDSs account for 55.25% of the total SK37 CDSs, indicating that the SK37 secreteome is extremely important to the biology of this organism. Of the CDSs that comprise the SK37 secretome, 394 CDSs are defined as functional transportation systems. According to COG analysis, these transporters are associated with nucleotides (8 genes), coenzymes (7), inorganic ions (110), amino acids (78), carbohydrates (68), organic solutes (14), defense mechanisms (23), and general functions (42) (Figure 5.2 and supporting material Table 2). Of which, 187 CDSs are members of the superfamily of ABC transporters associated

primarily with inorganic ions transport (63) and amino acid uptake (55). Since the ABC transport systems are believed to be extremely important in Gram-positive bacteria (Takami et al., 2000), these results suggest ABC transport systems have significant roles in SK37's ability to tolerate NaCl and obtain nutrients. Among the small-ion transporters, most has counterparts in other *Bacillus* species. Genes for exporting heavy metals (cobalt, molybdate, chromate, cadmium, and tellurite) and other toxic compounds (antimicrobial peptides, and polyketide antibiotics) are present. For cadmium resistant genes, they are organized in an operon containing a cadmium efflux system accessory protein, cadmium-transporting ATPase, and cadmium resistance protein (permease subunit) which have orthologs in *B. psudofirmus*. Interestingly, the *B. psudofirmus* operon is also associated with Tn*554*, but near *tnpA* rather than *tnpC* as was described above for SK37.

5.4.6 General metabolism

Virgibacillus sp SK37 is a moderately halophilic chemoorganotroph, which was isolated from 1-month-old Thai fish sauce containing anchovy mash in 26-28% NaCl (w/v). This bacterium is routinely cultured in a modified halophilic medium containing 1% yeast extract, 0.3% trisodium citrate, 0.2% potassium chloride, 2.5% magnesium sulfate, 5% NaCl at pH 7. However, when a fish broth prepared from anchovy tissue extract was used, cell growth from 0-30% NaCl was observed (Nawong, 2006). Supplementation of fish broth with D-glucose (3% w/v) did not affect cell growth, when compared to a non-glucose supplemented fish broth (Nawong, 2006). This is in agreement with API 50 CHB/E (BioMérieux, Marcy l'Etoile, France) (Nawong, 2006) results which indicate that SK37 utilizes only a very limited number of carbohydrates (*i.e.* salicin and cellobiose) and was capable of

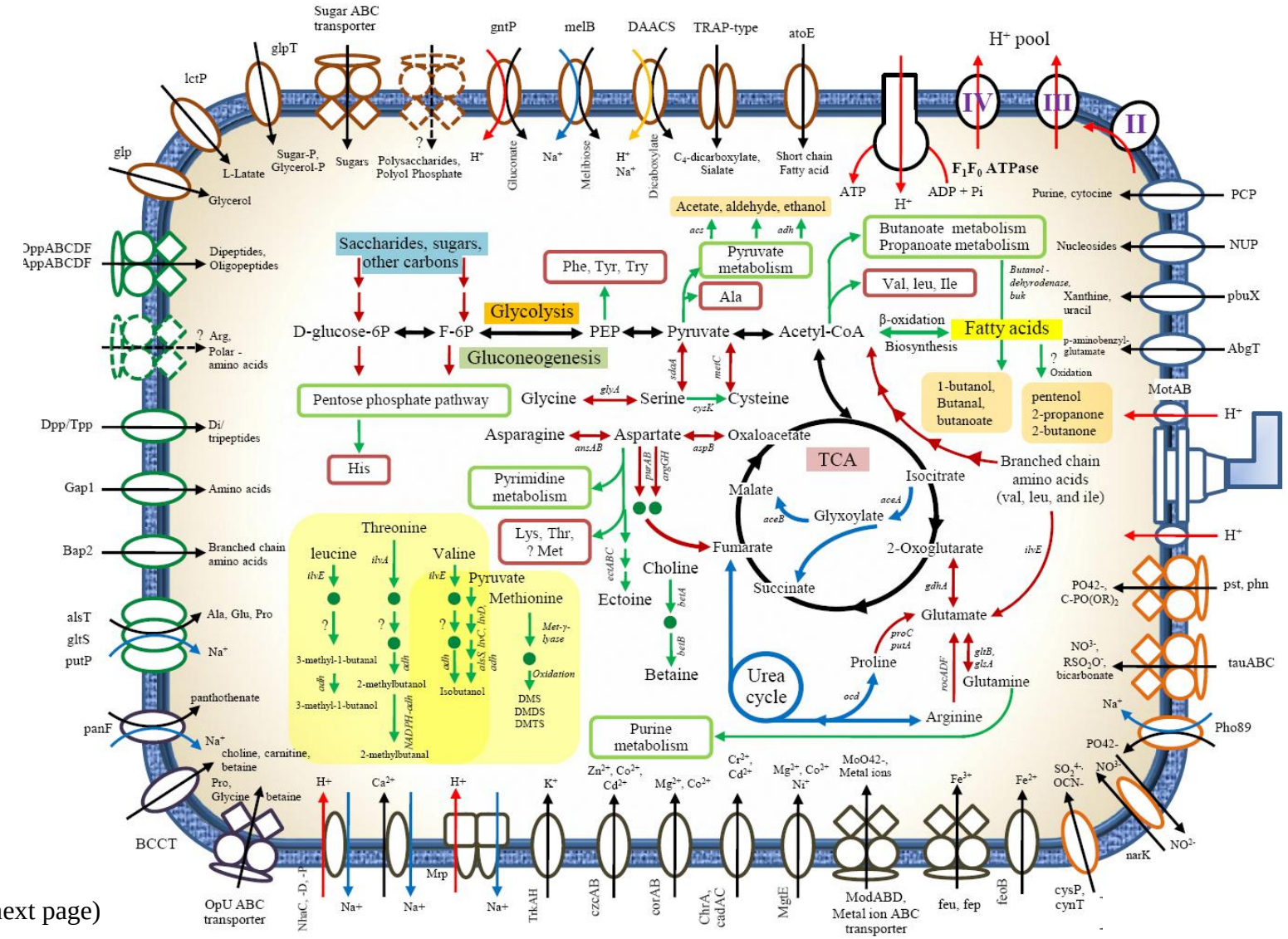


Figure 5.2 (see next page)

Figure 5.2 Central metabolism, organic ions, inorganic ions, and solute transports of the *Virgibacillus* sp. SK37. Overview of cell transport systems is grouped by their category: amino acids and peptides (green), carbohydrates, carboxylates and derivatives (brown), nucleotides and coenzymes (dark blue), compatible solutes (purple), cations (grey), anions (orange) and also respiratory chain (black). Regarding to membrane translocation system, H^+ (red arrow), Na^+ (blue arrow), and H^+/Na^+ (orange arrow) recycling are shown. KEGG database was used to construct central metabolic pathways where amino acid catabolism, central metabolism, and other related metabolism are demonstrated in red, black, and blue arrows, respectively. Biosynthesis pathways are showed in green arrows while their metabolic intermediates are indicated by green dots (see texts for details). Putative fish sauce flavor generation pathways are in light-yellow boxes. Pease see supporting material Table 2 and texts for all gene abbreviations and the detail of mechanisms.

hydrolyzing gelatin, but not starch or lipids. These observations indicate that SK37prefers peptides and amino acids as a carbon and energy source and is capable of growth on these substrates at near NaCl saturation. Genes encoding proteins involved in protein hydrolysis and amino acid catabolism in the SK37 genome were identified based upon their annotation. This analysis resulted in the identification of 65 genes encoding putative endo- and exopeptidases (Figure 5.3); 31 of these genes contain either a signal peptide (11 genes) or transmembrane domains (20 genes), suggesting that they are active on substrates outside the cytoplasmic membrane (Figure 5.3 and

supporting material Table 3.). Transporters associated with free amino acid, di-/tripeptide, and oligopeptide uptake are present (supporting material Table 2.) and likely are responsible for the transport of the products of extracellular protein hydrolysis. The remaining 34 genes encode a variety of endo- and exopeptidases, most likely involved in the intracellular hydrolysis of transported peptides. These results indicate that SK37 expresses a complex proteolytic enzyme system comprised of extracelluar endoproteinases and peptidases, a variety of transporters, and intracellular endo- and exopeptidases that provides the cell with amino acids to meet its nitrogen and carbon requirements.

A large number of SK37 pathways for the catabolism of amino acids are present in SK37, many of which yield either pyruvate or compounds that can enter the tricarboxylic acid cycle (TCA). The catabolism of arginine and serine is of particular importance as this organism has been shown to utilize these amino acids in a laboratory scale fish sauce fermentation (Yongsawatdigul et al., 2007). The most likely metabolic pathway for the catabolism of serine is its conversion to pyruvate by 2-serine dehydratase (*sdaA*) and this enzyme is present in the SK37 genome. Two pathways are present for the catabolism of arginine. Genes encoding all the enzymes required to convert arginine to fumurate via the urea cycle are present, the fumurate then would likely enter the TCA cycle and the genes encoding the TCA cycle are also present in SK37. Alternatively, under aerobic conditions, arginine may be converted to glutamate by the arginine deiminase pathway (Maghnouj, de Sousa Cabral, Stalon, and Vander Wauven, 1998), and then glutamate could be converted to 2-oxoglutarate by glutamate dehydrogenase (*gdhA*) and then enter the TCA cycle. All genes encoding enzymes involved in this cascade, *i.e. rocADF* for arginine deiminase

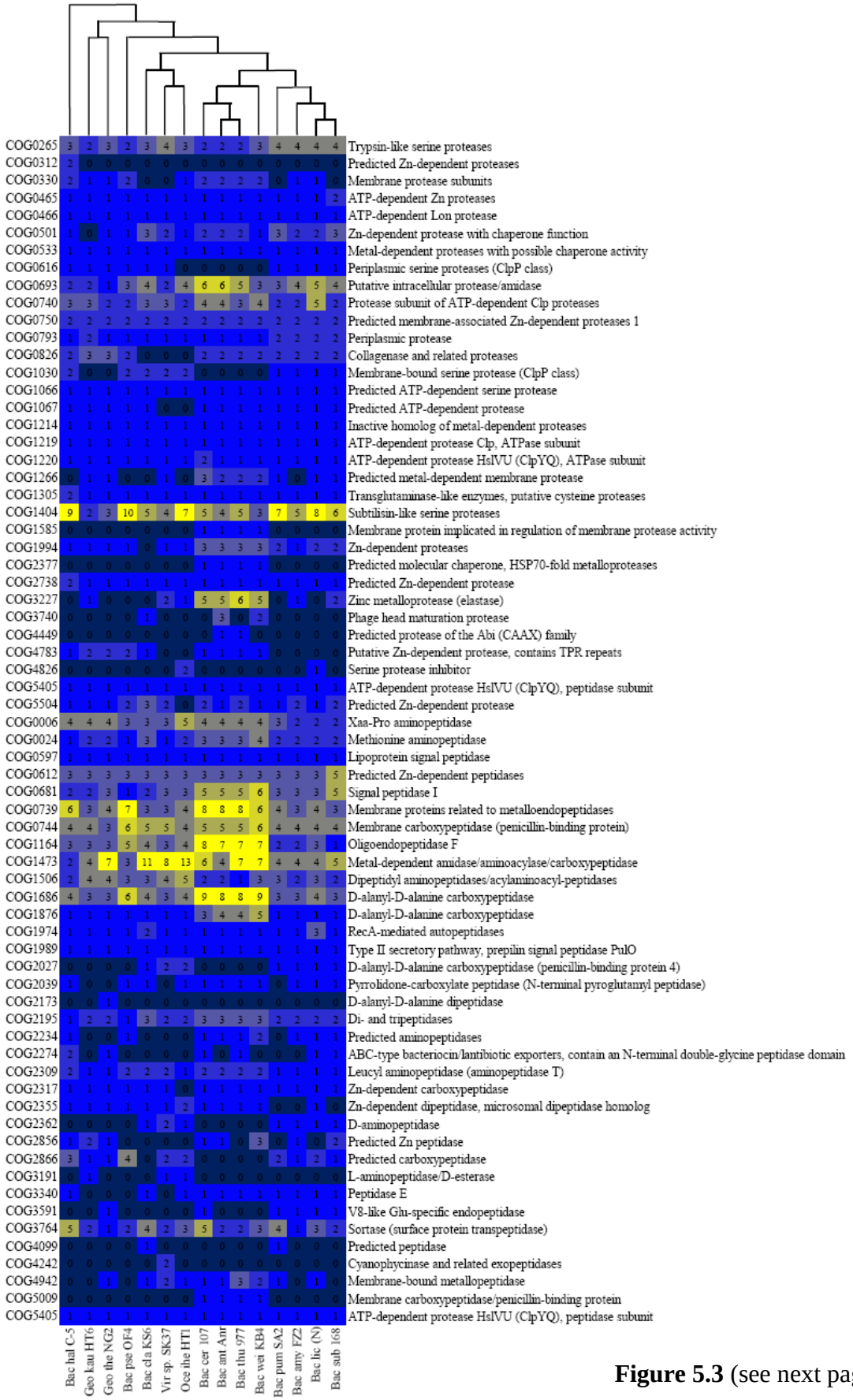


Figure 5.3 (see next page)

Figure 5.3 Gene abundant profile of proteinases and peptidases in 14 *Bacillus* species and *Virgibacillus* sp. SK37. Bacterial clustering was suggested by COG functional profile Hierarchical clustering. Identification numbers of each group of enzymes were based on COG category system. Bacterial name acronyms are indicated in supporting material Table 1.

pathway and *gdhA*, are present in SK37. Additionally, analysis of the SK37 genome suggested that glycine, cysteine, aspartate, asparagine, glutamine, proline, and branched chain amino acids may also be utilized. Glycine and cysteine may be catabolized to pyruvate by glycine hydroxymethyltransferase (*glyA*) and cystathionine betalyase (*metC*), respectively. Asparagine could be interchanged to aspartate by L-asparaginase (*ansAB*) and then aspartate enters the TCA via oxaloacetate by apspatate aminotransferase (*aspB*) or fumarate by either adenylosuccinate or L-arginino-succinate intermediate by the action of *purAB* and *argGH*, respectively. Glutamine may be interconverted to glutamate by glutamate synthase (*gltB*) or glutaminase (*glsA*). Proline could enter the urea cycle by ornithine cyclodeaminase (*ocd*) which is later converted to fumarate and then can enter the TCA cycle. In addition, this amino acid may be converted to glutamate via an L-1-pyrroline-5-carboxylate intermediate by pyrroline-5-carboxylate reductase (*proC*) and pyrroline-5-carboxylate dehydro-genase (*putA*). Branched chain amino acids, such as valine, leucine, and isoleucine, may be converted to glutamate by a branched chain amino acid aminotransferase (*ilvE*) or converted to acetyl-CoA or succinyl-CoA, which can then enter the TCA cycle. Amino acid catabolism is the primary source of energy for SK37 and hence it

was not surprising to observe a large array of genes of encoding enzymes involved in these pathways.

Although experimental evidence suggests that SK37 is unable to metabolize sugars and carbohydrates (Nawong, 2006), analysis of its genome revealed that all central pathways for carbohydrate metabolism are present, except for the Entner-Doudoroff pathway. Additionally, genes encoding α-amylase (*amyA*) and oligo-1,6-glucosidase for degrading starch and dextrin were found. Moreover, transporters associated with ABC-type and sugar-phosphate transport systems were detected. Though, the ABC-type carbohydrate transport systems detected lacked of the periplasmic component. In bacteria, phosphoenolpyruvate (PEP)-dependent phospho-transferase system (PTS) is used as the main carbohydrate uptake (Warner and Lolkema, 2003) which mediates the sugar uptake and transfer of phosphoryl group from PEP to the sugar molecule through a cascade of phosphotransfer proteins including EI, HPr, Hpr kinase, and sugar specific EII components, where the resultant phosphorelated sugar enters glycolysis for energy production (Warner and Lolkema, 2003). Genes encoding all conserved components of the PEP-PTS system are present in the SK37 genome. The sugar specific EII components, detected included the Glc family (glucose, N-acetyl-D-glucosamine, N-acetyl-muramic acid, sucrose, maltose, D-glucosamine, arbutin, and salicin), Lac family (lactose and cellobiose), Fru family (fructose and manitol), Man family (mannose), and a sugar alcohol (galactitol). In addition, genes encoding β-fructofuranosidase (*sacA*) and α-glucosidase (*malZ*) for sucrose and maltose hydrolysis were detected, as well as the enzymes, fructokinase (*scrk*) and glucokinase (*glk*), which are required to convert the monomers to derivatives capable of entering. Curiously, a CDS encoding a L-lactate permease was

detected (with annotation bit score = 280, lower than cutoff = 300), even though neither L-lactate dehyrogease (*ldh*) nor propionate CoA-transferase, enzymes involved in L-lactate metabolism to pyruvate and propanoate metabolism, respectively, were found. The importance of these enzymes involved in carbohydrate metabolism in the ecology and physiology of *Virgibacillus* is unclear, however it is possible that the conditions required for their expression have not been identified or that they are non-functional remnants of physiological traits undergoing genome decay as *Virgibacillus* evolves to the fish sauce environment.

Similar to the observations concerning carbohydrate metabolism, although no evidence exists that SK37 can utilize fatty acids, key enzymes for lipid utilization, such as lipases and the fatty acid β-oxidation pathway are present. However, these lipases lack signal sequences, suggesting an intracellular location for these enzymes; hence these enzymes are most likely involved in membrane lipid biosynthesis and modifications. Predicted transport systems related to lipid (2 genes of ABC-type) and short chain fatty acid uptake (1 gene) were detected. In addition, an entire set of enzymes, except acetoacetate decarboxylase (*adc*), involving in synthesis and degradation of ketone bodies, the byproducts of fat catabolism, are present. The absence of *adc* is not unexpected as most *Bacillus* species lack this gene.

Based on genomic properties, gene required for synthesis purine and pyrimidine nucleotides from aspartate and glutamine intermediates, pentose phosphate pathway from glycolysis derivatives, fatty acids from pyruvate metabolism, and all amino acids excepting methionine are present. The inability to biosynthesis methionine was predicted due to the SK37 genome lacks of cystathionine-γ-synthase (*metB*) in methionine synthesis via aspartate metabolism, although all other genes in

this pathway were detected. In addition, no genes encoding enzymes in methionine savage pathway were found. Interestingly, core enzymes of the glyoxylate cycle are found, suggesting the alternate use of fatty acids for carbohydrate synthesis. Regarding to vitamin and cofactor biosynthesis, SK37 contains genes necessary for synthesis of thiamine, riboflavin, $NAD^+/NADP^+$, and pantothenate, but lacks competed synthetic systems of pyridoxal, ubiquinone, folate, and biotin. In addition, one of each gene of pantothenate and cobalamin transporters is detected. Three sets of ABC-type nitrate/sulfonate/bicarbonate transport system and each one of genes encoding sulfate permease (MFS superfamily) and to nitrate/nitrite transporter were detected. However, the SK37 presumably could not metabolize sulfur in all forms (sulfur, sulfate, and sulfite) excepting hydrogen sulfite (H_2S) which may be used to synthesis L-cysteine via O-acetyl-L-serine by cysteine synthase A (*cysK*). Nitrate may be converted to nitrite and then reduced to nitric oxide by nitrate reductase (narI) and nitrite reductase, respectively. However, completed denitrification cycle to obtain molecular nitrogen (N_2) or ammonia seems to be impossible due to lack of key enzymes such as formate-dependent nitrite reductase (*nrf*) and nitrous oxide reductase (*nos*). Although whether the SK37 possesses nonortholog of these enzymes or its capability to use sulfurs and other non-protein nitrogen sources have not been conformed experimentally yet. Besides, ammonia, a primary source for nitrogenous compound biosynthesis, may be assimilated from amino acids via deaminase-related enzymes commonly found in bacteria. Moreover, one set of genes encoding tripartite ATP-independent periplasmic (TRAP) transporter is present. This TRAP-type transporter is associated with C_4-dicarboxylate uptake, mostly succinate, fumarate, malate, and also C_4-dicarboxylic amino acid aspartate, which can be metabolized and

serve as carbon and energy source in the TCA cycle (Janausch, Zientz, Tran, Kroger, and Unden, 2002).

5.4.7 Respiration and fermentation

Virgibacillus sp SK37 is a facultative aerobe, growing slowly in liquid media without aeration. In fish sauce production, fermentation is usually carried out in a cement tank buried in the ground and kept at ambient temperature. Neither agitation nor aeration are applied during the process until complete fermentation, usually takes about 12 to 18 months. The survival of the SK37 in laboratory scale fish sauce fermentation was found at least up to 4 months (Yongsawatdigul et al., 2007), suggesting the growth of this strain in oxygen-limited condition. Genome analysis reveals an aerobic respiratory electron transport chain and basic set which are conserved in most *Bacillus* species, including bacterial succinate dehydrogenase (complex II), cytochrome *c* reductase (complex III), cytochrome *c* oxidase (complex IV), and cytochrome *bd* complex (Figure 5.2). It should be noted that the *bd*-type has high affinity to O_2 for operating under low O_2 concentration (Janausch et al., 2002). However, the SK37 genome lacks of all enzymes in NADH dehydrogenase (complex I), which is also absent in other aerobic bacilli. Besides, one operon of compete F_1F_0-ATP synthase (complex V) with gene order of a-c-b-δ-α-γ-β-ε-subunit, found in other marine bacteria such as *Oceanbacillus iheyensis* and *Bacillus* sp. SG-1was observed. In addition, the operon for a flagella-specific ATP synthase is also conserved between the SK37 and other *Bacillus* species. None of genes related to fumarate reductases (*fdrABCD*) was found, indicating the lack of fumarate respiration, an anaerobic energy metabolism found mostly in anaerobic or facultative anaerobic bacteria (Kroger, Geisler, Lemma, Thesis, and Longer, 1992; Lancaster et al., 2005).

Fermentation is one of importance mechanism for energy generation in bacteria when deprived of oxygen (Salas, Garcia-Gonzalez, and Aparicio, 2006). Several genes of typical pyruvate fermentation pathways were found, such as acetyl-CoA syntethase (*acs*), aldehyde dehydrogenase, and alcohol dehydrogenase (*adh*), yielding acetic acid, acetyldehyde, and ethanol, respectively. However, L-lactate dehydrogenase (*ldh*) was not found, indicating no lactic acid production. It has been shown that ethanol was the predominant compound in fish sauce fermented by SK37 (Yongsawatdigul et al., 2007). Volatile fatty acids, such as butanoic acid and 3-methylbutanolic acid, had been reported to contribute the flavor of fish sauce (Park et al., 2001), which, however, did not accumulate significantly in the SK37-fermented fish sauce sample. Based on genomic information, this observation is not unexpected since the lipolytic oxidative pathways seem not to dominate in this strain. Evidently, an applicable amount of 1-penten-3-ol (pentenol), a fatty acid derivative giving mild green flavor, was detected. In plans, this compound is derived from polyunsaturated fatty acids through the lipoxygenase pathway as described in the review of Salas et al. (2006), which however are poorly understood in bacteria. Other ketones from fatty acid derivatives, including 2-propanone and 2-butanone most likely are formed from auto-oxidation of short chain fatty acids. Most of alcohols and aldehydes based volatile compounds are derivatives of amino acid catabolism. The degradation of leucine, threonine, and valine via α-ketoisocaproate, 2-keto-3-methylvalerate, and 2-ketoisovalerate, respectively, generates 3-methyl-1-butanal and 3-methyl-1-butanol (from leucine), 2-methylbutanol (from threonine), and 2-methyl-1-propanol (isobutanol) (from valine). These reactions require two key enzymes, including 2-keto acid decarboxylase (*kdc*) and alcohol dehydrogenase (*adh*) (Atsumi, Hanai, and Liao,

2008; Cann and Liao, 2010; Marilley and Casey, 2004). The 2-methylbutanol could be further converted to 2-methylbutanal by the action of board specificity NADPH-dependent alcohol dehydrogenase (Larroy, Fernandez, Gonzalez, Pares, and Biosca, 2002), which is also present in the genome. Excepting *adh*, the *kdc* was not found in the SK37 genome as well as most of *Bacillus* species, suggesting the presence of a orthologous enzyme with the same function. Alternatively, isobutanol may be produced from *kdc*-independent synthetic pathway using pyruvate as a precursor (Atsumi, Li, and Liao, 2009). All necessary genes including acetolactate synthase (*alsS*), ketol-acid reductoisomerase (*ilvC*), and dihydroxy-acid dehydratase (*ilvD*), are present. Some compounds such as 1-butanol, butanal, and butanoic acid may be derived from the butanoate metabolism using acetyl-CoA as a precursor (Nielsen, Leonard, Yoon, Tseng, Yuan, and Prather, 2009). Genes encoding key enzymes such as NADPH-dependent butanol dehydrogenase and butyrate kinase (*buk*) were detected. Methionine may be converted to methanethiol via demethiolation reaction catalyzed by methionine-γ-lyase, which is present in the SK37 genome. The methanethiol could be subsequently auto-oxidized to dimethyl sulfide (DMS), dimethyl disulfide (DMDS), and dimethyl trisulfide (DMTS) (McSweeney and Sousa, 2000). These compounds are believed to be the distinctive fish sauce odor (Fukami et al., 2002; Peralta, Shimoda, and Osajima, 1996). It should be noted that all of the volatile compounds discussed are based on experimental observation (Yongsawatdigul et al., 2007).

Biogenic amines accumulate during fish sauce fermentation from the metabolism of microorganisms (Yongsawatdigul et al, 2004). Histamine, one of most abundant biogenic amines, is used to indicate the freshness of row fish for the

fermentation and also reflects the final quality of fish sauce. According the Canadian Food Inspection Agency (CFIA), the maximum allowable histamine in commercial fish sauce is 20 mg/100 g (CFIA 2003). Using the SK37 as a starter culture for fish sauce fermentation has been showed to lower the level of biogenic amines in the final product (Yongsawatdigul et al., 2007). In addition to histamine, other biogenic amines, including cadavarine, tyramine, tryptamine, putrescine, spermine, and spermidine have been detected in food related materials (Silla Santos, 1996). Histamine, cadavarine, tyramine, and tryptamine are derivatives of decarcoxylated histidine, lysine, tyrosine, and tryptophan, respectively. Lysine decarboxylase (*cadA*) was detected in the SK37 genome, however histidine decarboxylase (*hdc*), tyrosine decarboxlyase, and aromatic-L-amino acid decarboxylase (*ddc*) were not detected. This is in agreement with the experimental observation that .The lysine decarboxylase may show a limited activity in fish sauce fermentation as very low amount of cadavarine was detected in the SK37 fermented fish sauce (Yongsawatdigul et al, 2004). In contrast, arginine can be converted to ornithine in the urea cycle, and then this compound is further catabolized into, subsequently, putrescine, spermidine, and, spermine by the action of ornithine decarboxylase (*speCF*), sermidine synthase (*speE*) and spermine synthase (*sms*), respectively (Silla Santos, 1996). The SK37 genome apparently lacks of *speCF* and *sms* genes but not *speE*. In addition, arginine may be alternately catabolized to putrescine via agmatine intermediate by arginine decarboxylase (*speA*) and agmatinase (*speB*). Even though all of these genes were detected the SK37 genome, they seem to show limited activity since these three biogenic amines, *i.e.* putrescine, spermidine, and spermine, were not detected in the experimental fermentation (Yongsawatdigul et al., 2007). In addition, even though the

SK37 seems to impair the production of biogenic amines but the quality of row fish is still very important as SK37 may uptakes putrescine and spermine to its cytoplasm and processes these amines afterward. In addition, an ABC-type spermidine/putrecine transport system was detected, although it is an uncompleted set (1 gene of ATPase component) and its activity is still questionable.

5.4.8 Halophilic adaptation

Osmolarity response in saline environments of moderately and extremely halophilic bacteria has been reported. The adaptive response is characterized by two distinct modes, including inorganic ion recycling called "high-salt-in strategy" and compatible solute accumulation called "compatible-solute-in strategy" (Oren, 2008). Like haloarchaea, *Salinibacter ruber* accumulates intracellularly high concentrations of K^+ and Na^+ in order to allow growth at NaCl higher than 1.7 M (Oren, Heldal, Norland, and Galinski, 2002). In contrast, *B. subtilis* responds a sudden increase of external osmolarity by rapidly uptake of K^+ and subsequently accumulates compatible solutes, especially proline (Takami et al., 2002). On the other hand, *Halobacillus halophilus* mainly accumulates glutamate and glutamine at moderate salinity (1 M NaCl) and switches to proline at higher salinity (2-3 M NaCl) (Saum and Muller, 2008). The $Na^+/K^+/H^+$ cycle plays a major role in the inorganic ion recycling in a wide range of organisms. The SK37 grows at a wide pH range of 4 to 11 (optimum 6.5-7.5) and 0 to 25% NaCl (optimum 5-10%). Its genome employs 5 genes related to K^+ transporters (4 genes in TrkHA transport mechanism [3 copies of *trkA*, and 1 copy of *trkH*], and 1 gene in K+/H^+ antiporter in CAP1 family), compared to total 3 genes in *B. clausii*, 4 genes in *Oceanobacillus iheyensis*, although 7 genes in *B. subtilis*, 10 genes in *B. psudofirmus* and 14 genes *B. thuringiensis*. In addition, *S. ruber* genome

contains 4 and 2 copies of *trkA* and *trkH*, respectively (Mongodin et al., 2005). These transporters could play a role on initial rapid uptake of K^+. However, unlike other most bacilli, high affinity K^+ transporters such as ABC-type and Kef-type were not detected. Besides, SK37 genome possesses 47 genes associated with Na^+ cycle, which is more than any other *Bacillus* species compared. Of these are genes related to Na^+/H^+ antiporter which are directly responsible for pH and Na^+ homeostasis (Padan, Venturi, Gerchman, and Dover, 2001), such as a complete set of Mnh-type (6 genes, *mrpABCEFG*), NhaC-type (5 copies), NhaD-type (2 copies), and NhaP (4 copies). These Na^+/H^+ specific transporters would cooperate together to facilitate the growth of this strain in wide pH and NaCl ranges. Interestingly, 2 copies of gene encoding Na^+/Ca^{2+} antiporter were detected, which have only *Ocenobacillus iheyensis* counterpart, but not found in other *Bacillus* species. In eukaryotic cells, this transporter is thought to maintain low Ca^{2+} in the cells but could reverse exchanger's direction when intracellular level of Na^+ rise beyond a critical point (Bindokas and Miller, 1995; Nicoll, Longoni, and Philipson, 1990). Since, in most cases, the stoichiochemistry of exchanging is 1 Ca^{2+} per 3 Na^+ (Yu and Choi, 1997), these could be an effective mechanism for controlling intracellular Na^+ in high salinity, although there is no experimental evidence for such function in halophilic bacteria. Na^+ re-entry via the Na^+/solute symporter is the second major of the Na^+ cycle (Krulwich, Ito, and Guffanti, 2001). Osmoprotectants or compatible solutes are osmolytes, such as amino acids and their derivatives and sugars and sugar alcohols, which help organisms survive under extreme osmotic stress by using these molecules to osmotically balance their cytoplasm (Lang, 2007; Oren, 2008). Strikingly, the SK37 genome contains the highest number of genes associated with high affinity substrate specific transport

systems (OpuE) for osmoprotectant uptake. The OpuE transporter is a single component classified in a member of the Na^+/solute symporter family (SSF) (Kempf and Bremer, 1998). In the SK37 genome, this includes Na^+/solute symporters specified for alanine (4 copies), glutamate (1 copy), proline (7 copies), and also panthathenate (1 copy), another member of SSF. These genes showed homologous and comparable abundant to *Oceanobacillus iheyensis*. The SK37 genome also employs transporters for scavenging glycine betaine and choline, which are compatible solute amino acid derivatives, and proline via ABC-type proline/glycine betaine transport system (each 3 copies of ATPase, permease, and periplasmic components), choline-glycine betaine transporter (2 copies), and choline/carnitine /betalaine transporter (6 copies). It should be noted that the abundance of the last two transporters are a principal difference between moderately halophilic SK37 as well as *Oceanobacillus iheyensis* against other non-halophilic *Bacillus* species (Takami et al., 2002). Moreover, betaine may be synthesized when choline is available by the action of choline dehydrogenase (*betA*) and betaine-aldehyde dehydrogenase (*betB*), while ectoine, a compatible solute found in most bacteria (Bursy, Pierik, Pica, and Bremer, 2007; Galinski and Truper, 1994), may be *de novo* synthesized from aspartate where complete sets of enzymes involving these pathways (*betAB* and *ectABC*) are present in the SK37 genome. Biosynthesis of ectoine in other *Virgibacillus* sp. has been reported (Kuhlmann, Bursy, Gimpel, Hoffmann, and Bremer, 2008).

It has been reported that there is unique difference in term of amino acid composition of proteins between non-halophilic and halophilic bacteria (Fukuchi et al., 2003). This strategy helps proteins of halophiles maintain proper conformation and biological function at high salt concentration. Proteinases produced from this

strain have been characterized and showed a halophilic character which is capable to express appreciable activity at high salt concentration (Phrommao et al., 2010; Sinsuwan et al., 2007, 2008). However, unlike enzymes from extremely halophiles, the proteinases from the SK37 could retain its activity event in the absence of NaCl where halophilic enzymes are irreversibly denatured (De Castro, Maupin-Furlow, Gimenez, Herrera Seitz, and Sanchez, 2006; Ventosa et al., 1998). This observation brought to investigate the amino acid compositional difference of proteins in non-halophile, moderately halophilic, and extremely halophilic bacteria. It should be noted that only protein sequences containing signal peptide for secretion were used in this analysis since most moderate halophiles use the "compatible-solute-in strategy" to keep low salt concentration inside the cells, whereas their extracellular enzymes function in extreme salt environments. The relative abundance of aspartate, lysine, asparagines, alanine, isoleucine, phenylalanine, and threonine are significant of extreme halophiles against non-halophiles (Figure 5.4). This is in agreement with other reports showing that halophiles adapt more acidic residues and use less basic and hydrophobic residues in order to maintain protein conformation by electrostatic interactions (Fukuchi et al., 2003; Oren, 2008). However, amino acid composition of non-halophiles and moderate halophiles are fairly consistent, suggesting the amino acid compositional adaptation of extreme halophiles may not be applied in the moderate halophiles. The amino acid compositions of the extracellular proteinase of SK37 and its homologs from non-halophiles and other moderate halophiles were compared and showed the same pattern (data not showed). This may support the hypothesis of "dragging effects" explaining that the high level of deviation of aspartate and basic and hydrophobic residues in extremely halophiles is mainly from

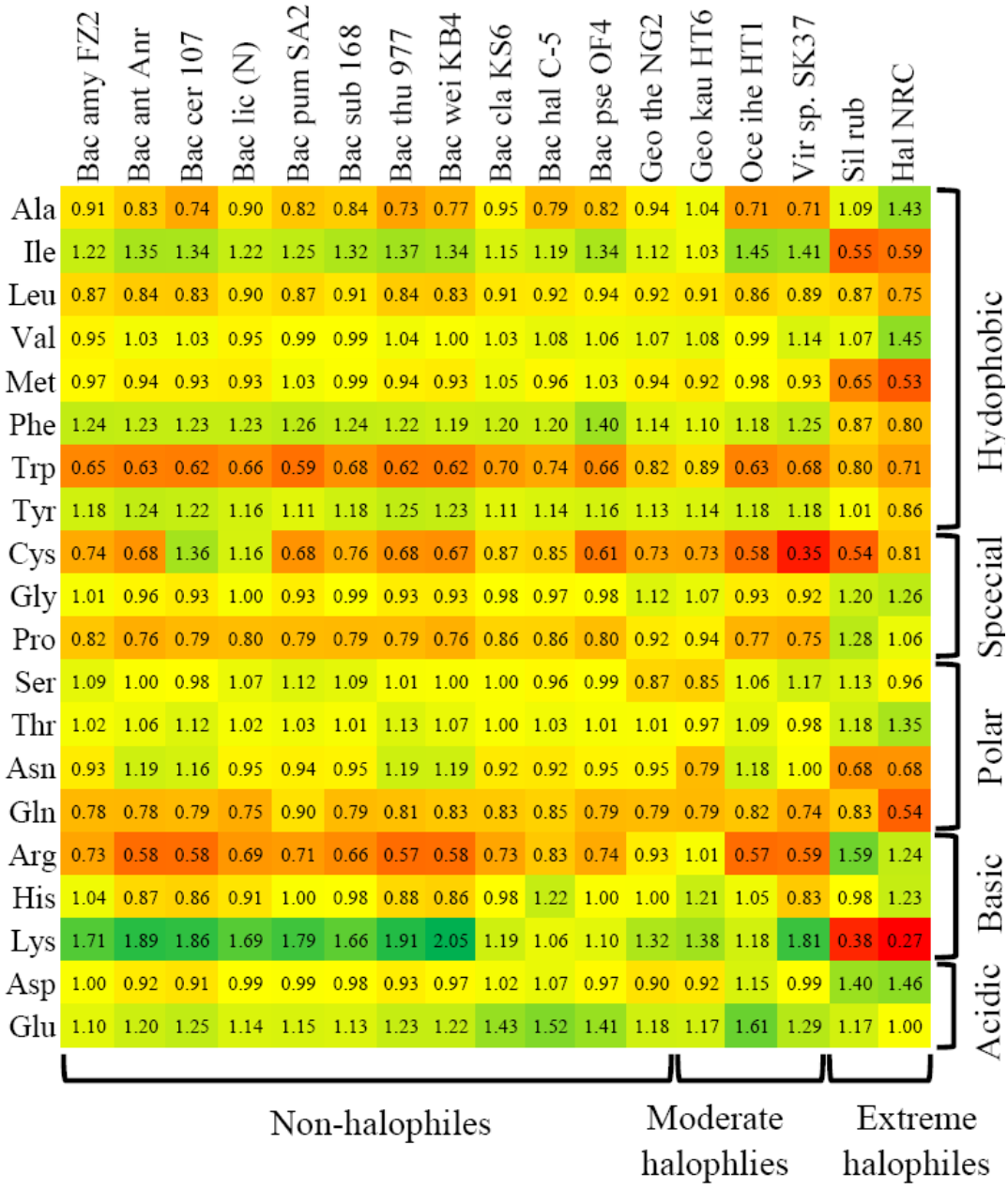


Figure 5.4 Compositional difference of amino acids in non-halophiles, moderate halophiles, and extreme halophiles. Genomic sequences of *Salinibacter ruber* (Sal rub) and *Halobacterium* sp. NRC-1 (Hal NRC) were selected as representative for extreme halophiles. Only proteins with signal peptide for secretion were selected for analysis. The amino acid profiles

(Figure 5.4 Continued)

of all bacteria were standardized by respective proteins of *E. coli* and showed as standardized values in blocks where their under-representation and over-representation in the organisms are shown in green and red colors, respectively. Please see material and method for details.

the effect of dinucleotide compositional shift of extremely high G+C content (~68%) in these bacteria (Fukuchi et al., 2003), whereas other non-halophiles and moderate halophiles have lower G+C contents. In addition, this also agreed with an assumption suggested by (Saum and Muller, 2008) that the amino acid compositional adaptation is rather a tendency than a rule as the knowledge of osmoregulatory of most halophiles have not yet well established

5.4.9 Adaptation to a protein-rich niche

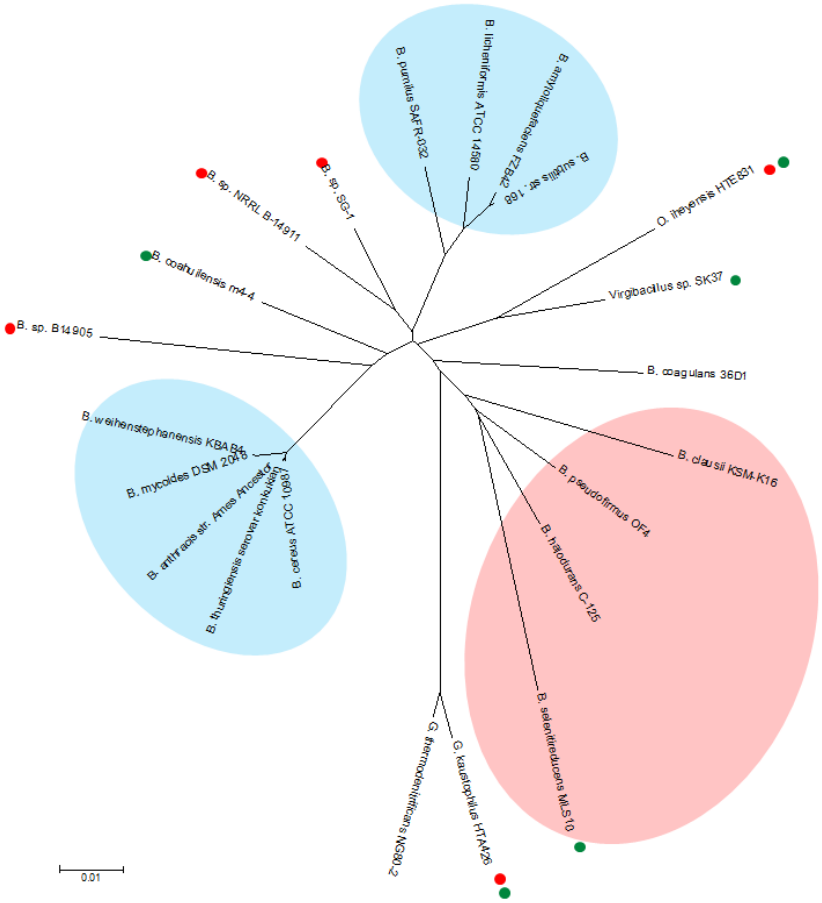
Although experimental observations suggested that SK37 prefers amino acids and proteins as its primary source of energy and lacks the ability to metabolize various carbohydrates and lipids, genes encoding a number of potential metabolic pathways for the use of carbohydrates and lipids were detected in the SK37 genome. The most likely explanation for this contradiction is that the expressions of genes encoding enzymes for carbohydrate and lipid metabolism are tightly regulated. Unlike the rich media regularly used to culture bacteria in the laboratory, nutrients in bacterial natural habitats can be present in extremely dilute concentration, inaccessible, or transiently available (Sonenshein, 2007). Therefore, bacteria have to develop a sophisticated system to regulate pathways involved in central metabolism. Specifically, genes

required for the utilization of an abundant readily utilized substrate are expressed, while expression of genes required for the utilization of other alternative energy sources are limited. In *B. subtilis*, global transcriptional regulators, such as CcpA, CodY, and TnrA, are used to control and coordinate the expression of genes involved in central metabolism and hence metabolic flux (Sonenshein, 2007). CcpA and CodY are global regulators of carbon and nitrogen metabolism pathways, whereas ThrA is a global regulator that mediates positive or negative expression of genes involving in nitrogen compound metabolism (Fujita, 2009; Molle et al., 2003; Wray, Ferson, Rohrer, and Fisher, 1996). In *B. subtilis*, CcpA and CodY are each responsible for the regulation of more than 100 transcription units (Molle et al., 2003; Sonenshein, 2007) and they function to control metabolic pathways resulting in pyruvate formation (Kim, Roux, and Sonenshein, 2002). The expression of TnrA is under the control of GlnR, which is also an autorepressor and negative regulator of the urease (Brandenburg, Wray, Beier, Jarmer, Saxild, and Fisher, 2002). The link between carbon and nitrogen metabolism is 2-oxoglutarate, a metabolite in TCA cycle and a *de novo* carbon skeleton for the biosynthesis of most important nitrogen-containing compounds, glutamate and glutamine (Sonenshein, 2007). In *E. coli*, glutamine is used to supply nitrogen for the synthesis of 25% of nitrogen-containing compounds in the cell (Reitzer, 2003). This carbon-nitrogen metabolic intersection is regulated by at least six proteins including GltC, TnrA, RocG, RocR, CcpA, and CodY (Sonenshein, 2007). The SK37 genome contains all of these genes excepting RocR and TnrA. RocR is an operon activator for synthesis 2-oxoglutarate from glutamate. This also reflects the metabolism of other amino acids that can convert to glutamate, such as glutamine, arginine, proline, and histidine (Sonenshein, 2007). The SK37 might employ other

activator with similar function as all other genes of this pathway including *rocG* and *rocABC* were found. On the other hand, TnrA is a repressor for glutamine syntethase and its activity is inactivated if glutamine is available (Wray et al., 1996; Wray, Zalieckas, and Fisher, 2001) whereas GlnR is activated by glutamate syntethase and glutamine (Fisher, 1999; Schreier, Brown, Hirschi, Nomellini, and Sonenshein, 1989). Interesting, other organism that lacks TnrA uses GlnR to repress glutamine and glutamate synthesis (Kloosterman et al., 2006). If this character was applied to the SK37, the lack of TnrA and the repression activity of GlnR would seem to make sense since this bacterium might require to catabolite glutamine for energy generation and drive the metabolic flux, together with other amino acids degradation, to produce central metabolites rather than *de novo* amino acid synthesis.

In addition, evolutionary relationship of the global regulators (CcpA, CodY, and TnrA) among *Bacillus* species was analyzed along with their 16S rRNA phylogenetic tree. Since Gram-positive bacteria that lack of TnrA (including the SK37) are believed to use GlnR to carry out some global function of TnrA, GlnR was used instead of TnrA for this analysis (Sonenshein, 2007). In this analysis, *Bacillus* species were preliminarily selected regarding to their three major different features including their physiology (acidophile, neutral, and alkaliphile), origins (terrestrial and marine), and halotophilic ability (see material and method for details). Evidently, terrestrial neutrophile was the major group that was clustered commonly in the same phylogenetic clade (with two subgroups) in all trees. In contrast, a few phylogenetically coherent groups consisting in marine and halophilic features were found (Figure 5.5). With respect to halotophilic ability, halophiles and non-halophiles are often group together in the phylogenetic tree (Oren, 2008). However, unlike other

(a)



(b)

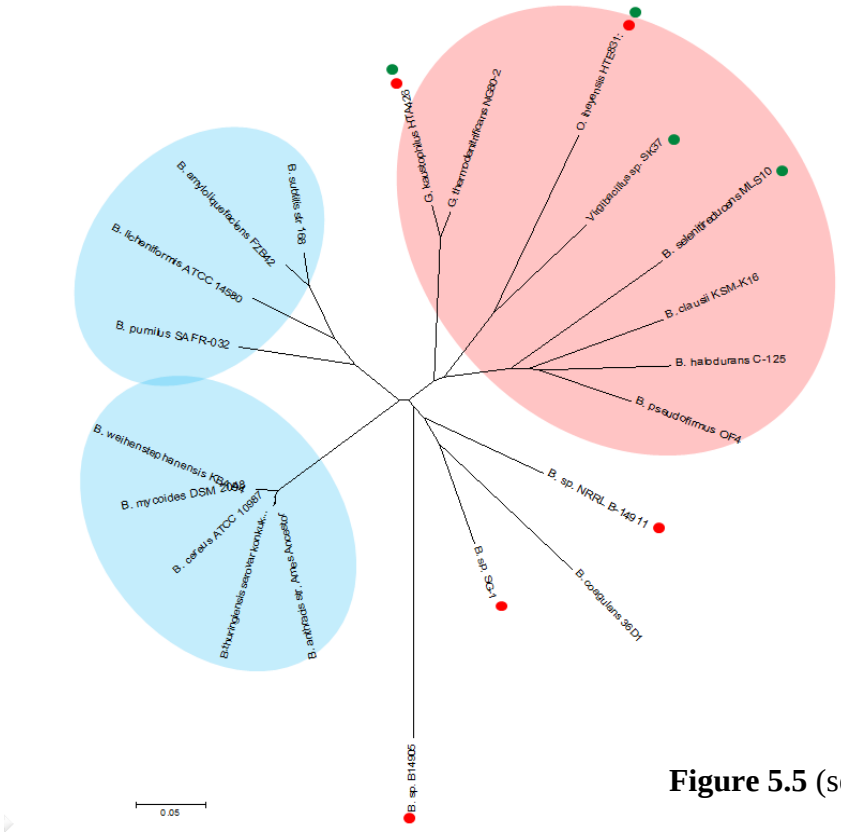


Figure 5.5 (see page 210)

(c)

(d)

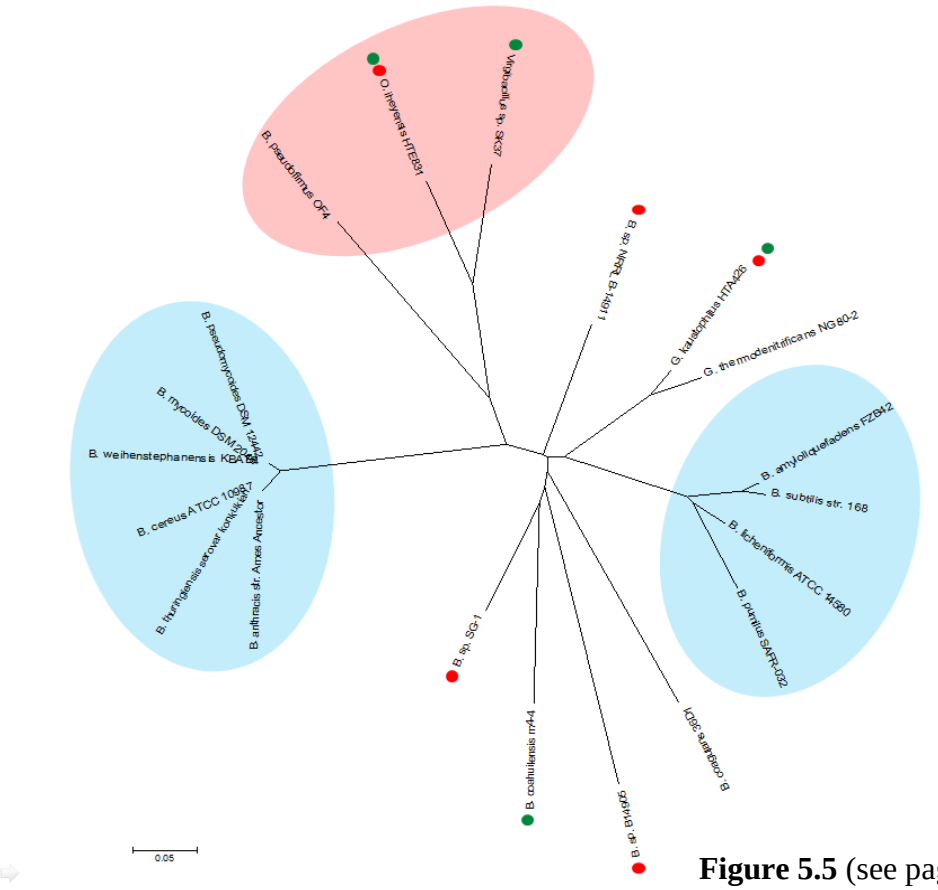


Figure 5.5 (see page 210)

Figure 5.5 Evolution phylogenetic trees for nucleotide alignment of 16S rRNA (a) and amino acid alignments of CcpA (b), CodY (c), and GlnR (d) from *Bacillus* species. In each alignment, strains that lack of respective sequence were excluded from the analysis. The trees were constructed by Neighbor-joining method with 1,000 random bootstrap replications and completed deletion. Terrestrial-neutrophiles and alkaliphiles are showed in blue and pink cycles, respectively. Marine and halophilic bacteria are marked by red and green dots, respectively.

trees obtained from 16S rRNA, CodY, and GlnR, bacterial clustering by CcpA protein apparently showed the most relatedness regarding to the halophilic and alkaliphilic properties of bacteria (Figure 5.5b). This observation suggests that halophilic and alkaliphilic bacteria were evolved concomitantly by mean of CcpA. As previous mentioned, CcpA is one of the most important global regulators for carbon-nitrogen metabolism that bacteria used to cope with environmental changing in term of nutrient availability. In fact, CcpA could play functions differently in different bacteria by mean of operon specificity. For example, while most of bacteria use CcpA to control their sugar uptake, some *Lactobacillus* species use this protein as an activator for peptidases expression (Morel, Frot-Coutaz, Aubel, Portalier, and Atlan, 1999; Morel, Lamarque, Bissardon, Atlan, and Galinier, 2001; Schick, Weber, Klein, and Henrich, 1999; Stucky, Schick, Klein, Henrich, and Plapp, 1996). SK37 and other halophilic and alkaliphilic bacteria may use this protein as a key regulatory response for their inche-associated evolution. In addition, this also raises a question as how halotophiles and alkaliphiles are evolutionarily related since they seem to use the same mechanism

for their nutrient stress responses.

A high production of extracellular proteinases is one of most important characters of bacteria growing in a protein-rich environment since the degradation products of these enzymes could supply the cell with enormous nitrogen source. SK37 produces at least 12 extracellular endoproteinases differentiated by molecular masses (Phrommao et al., 2010). In fact, this strain was one of the highest proteinase producer isolated from Thai fish sauce samples (Nawong, 2006). In *B. subtilis,* the expression of many hydrolytic enzymes is repressed in the exponential phase of growth by a group of proteins known as "transition state regulators" (Hoch, 1993; Strauch, 1993). With respect to proteinase production, negative regulation of extracellular serine and metalloproteinases, which are major extracellular proteinases in most Gram-positive bacilli, are controlled by three proteins, including a repressor Hpr, and corepressors ScoC and SinR (Kallio, Fagelson, Hoch, and Strauch, 1991; Kodgire, Dixit, and Rao, 2006; Koide, Perego, and Hoch, 1999). Gene deletion of these proteins in *B. subtilis* resulted a mutant with hyper production of proteinases (Kodgire et al., 2006; Perego and Hoch, 1988). Strikingly, CDSs encoding Hpr and ScoC related proteins are absence in the SK37 genome. Excepting to *Oceanobacillus iheyensis*, this is unique to all *Bacillus* species compared. Although there are no evident of proteinase production in *Oceanobacillus iheyensis*, genomic comparison showed that almost proteinase genes in SK37 have counterparts to this bacterium while a high proteinase production in SK37 was experimentally observed (Phrommao et al., 2010; Sinsuwan et al., 2007, 2008). This observation suggests that (I) the SK37 has evolved specifically to elevate the level of proteinase production by spontaneous deletion of these key regulatory proteins or (II) it uses another specific regulatory circuit for proteinase production,

which is potentially different from other *Bacillus* by mean of regulatory machinery. Either SK37 deploys one or both of these assumptions the result is common as to thrive on optimum growth in a protein-rich niche.

5.4.10 Comparative genome analysis

Besides the phylogenetic tree based on 16S rRNA sequences shown in Figure 5.5a, SK37 genome was compared to other complete genome sequences of *Bacillus* related species listed in the IMG database (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi) suggested by genome clustering analysis by COG functional profile Hierarchical clustering (supporting material Figure 1). SK37 was mostly related to *Oceanobacillus iheyenisis*, a marine bacterium. Although all *Bacillus* related species are from the same phylogenetic clade (Priest, 1993), they show an adaptive capability to their own environmental origin. Therefore, complete genome sequences of 14 *Bacillus* related strains were selected for comparative analysis based on orthologous genes.

The comparison was further divided into six groups (one strain could fall into more than one group when it showed overlapped characters) based on; their origins, including terrestrial and marine; their physiology, such as thermophiles, alkaliphiles, and halophiles and their phenotype, such as proteinase producers (Figure 5.6a). Based on COGs, 996 orthologs were conserved in all *Bacillus* compared. These orthologs represent 41.0-52.9% of each genome. The major functions of unique orthologs (orthologs that conserve only in each own group) in terrestrial, marine-halophiles (members in marine and halophilic characters appeared to be the same group of bacteria), and alkaliphilic bacteria are associated with translation, energy production, secondary metabolite catabolism, amino acid metabolism, and unknown function. In

contrast, the unique orthologs in thermophiles were mostly related to cobalamin biosynthesis (12 out of 25). With respect to proteinase producers, one of their three unique orthologs is a molecular chaperone in HSP90 family. In eukaryotic cells, this protein is one of the most abundant cellular proteins whose function is related to protein folding and stabilization and refolding of denaturated protein after stress (Sreedhar, Kalmar, Csermely, and Shen, 2004). In *E coli*, the expression of this related protein depended on growth environment, especially in a large pool of amino acids and proteins (Mason, Dunner, Indra, and Colangelo, 1999), implying the relevance to protein turnover of bacteria. In addition, there are 11 orthologs shared only between the SK37 and *Oceanobacillus iheyensis*. Most of which are related to DNA recombinant and repair, unknown function, and enzymes in the mevalonate pathway, as previously described.

Moreover, there are 19 orthologs (22 genes) of the SK37 which were unique to other *Bacillus* species. Most of them are associated with membrane proteins, unknown function, and energy production. Interestingly, some enzymes are related to protein and carbon source metabolisms. Regarding to protein catabolism, we found genes encoding phosphoserine phosphatase (*serB*) (1 copy) and high affinity L-asparaginase II (*ansB*) (2 copies) which convert phosphoserine and asparagine into serine and aspartate, which could further enter TCA cycle for energy production as previously described. The *serB* and *ansB* showed high homology to marine origin of *Bacillus* sp. SG-1(55% identity) and *Geobacillus* sp. G11MC16 (68% identity), respectively, suggesting a horizontal gene transfer event. In addition, 2 copies of gene for cyanophycinase, a peptidase degrading cyanophycin, were detected. The cyanophycin is the branched multi-L-arginyl-poly-aspartic acid which is a reserved protein material

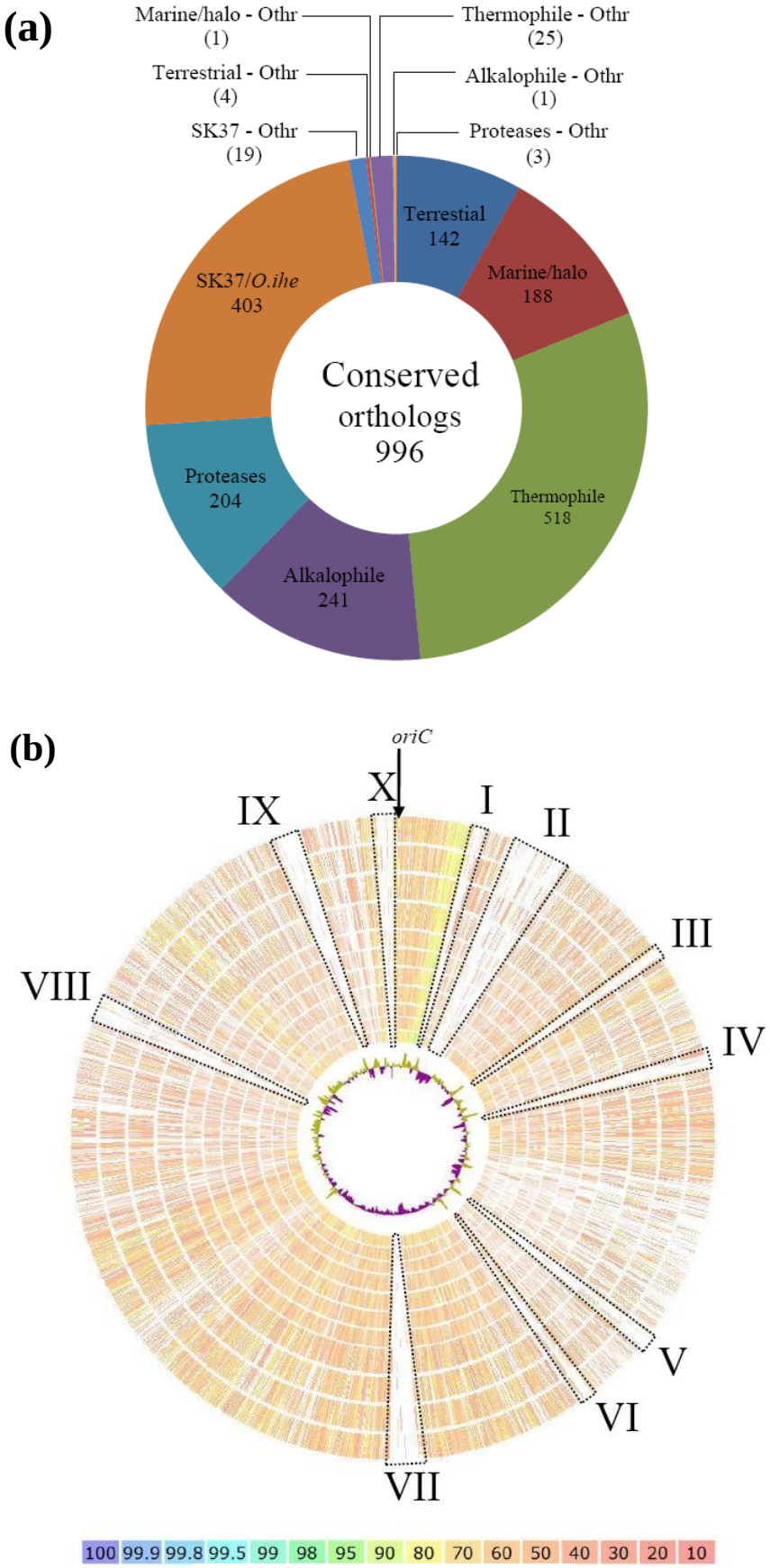


Figure 5.6 (see next page)

Figure 5.6 Summary of orthology relationships among proteins in genomes of *Bacillus* species. (a) Bacteria were categorized based on their physiology, origin, and proteinase production listed in supporting material Table 1. The group name "othr" means that these orthologous sequences are conserved only in their own group and not present in any other genomes compared. (b) Pairwise analysis of orthologous sequences between SK37 and other bacilli genomes. Form outer to inner cycles: *B. amyloliquefaciens* FZB42, *B. anthracis* Ames Ancestor, *B. cereus* ATCC 10987, *B. clausii* KSM-K16, *B. halodurans* C-125, *B. licheniformis* ATCC 14580 (Novozymes), *B. subtilis* subsp. subtilis str. 168, *B. pumilus* SAFR-032, *G. kaustophilus* HTA426, *G. thermodenitrificans* NG80-2, and *O. iheyensis* HTE831. Colors representing %sequence identity are indicated. Genomic islands (I to X), which are the regions on SK37 genome that did not show orthologous sequences to other bacilli compared are marked. The *oriC* and %G+C content of SK37 genome are arrowed and demonstrated in the inner cycle, respectively.

found only in cyanobacteria while its hydrolysis is highly specific to cyanophycinase (Richter, Hejazi, Kraft, Ziegler, and Lockau, 1999). Since the hydrolysis products *i.e.* arginine and aspartate seem to be utilized by the SK37, the presence of this enzyme together with *serB* and *ansB*, may confers high efficiency of this strain for using amino acids as a main source of energy. For other carbon source metabolism, one copy of neuraminidase (sialidase) for degradation of sialic acid, a nine-carbon keto sugar acid found ubiquitously in bacterial and eukaryotic protein and cell surface

(Vimr, Kalivoda, Deszo, and Steenbergen, 2004), was detected. In fact, other core enzymes of sialic acid metabolism (*nanAEK* and *nagAB*) are also present. The final product of this pathway is fructose-6-phosphate which enters glycolysis for energy generation (Severi, Hood, and Thomas, 2007), suggesting the capability of using this glycan molecule as a nutrient as found in other non *Bacillus* bacteria (Vimr et al., 2004).

The pairwise comparison of the SK37 genome against other selected *Bacillus* related species was performed. Average sequence identity of orthologous sequences was found at 40-60% (Figure 6b). About 57-69 genes of the SK37 showed more than sequence 80% sequence identity to all bacterial strain compared (78 genes showed more than 90% identity to *Oceanobacillus iheyensis*), where their major functions are related to ribosomal proteins and other proteins in DNA transcription and translation clustered in between nt. 133-152 kb from the *oriC*. There are at least 10 well-defined genomic islands on the SK37 genome and their locations correspond to remarkably shift of G+C content (Figure 5.6b). Of which, genomic island II and VII appeared in considerable sizes (~99.8 kb for II and ~34.0 kb for VII). The island II is associated with transposons (2 copies of Tn*4*-like and 1 copy of Tn*554*) while the island VII is related to phage remnants, as previously described. In contrast, other islands mostly contain hypothetical proteins (>90%) and minor genes of transcriptional regulators, restriction enzymes, transpoases, recombinases, and other non-clustered phage remnants. It should be noted that total genes that presence in all of these genomic islands account to 9.0% (343 genes) of total CDSs in the SK37 genome. In addition, according to Dotplot analysis, the physical distribution of common genes between the SK37 and *Oceanobacillus iheyensis* is largely collinear, compared to other bacilli,

suggesting conserved gene organization between these two genomes (supporting material Figure 2). However, the direction of linearity inverses at region around 135-225°, which may be the action of insertion elements since this region is in between genomic island IV and a transpoase encoding gene.

5.5 Conclusions

SK37 is the first strain in genus *Virgibacillus* that its genomic sequence has been determined. Along with other moderately halophilic bacteria which their whole genomic sequences have been investigated, including *Ocenobacillus iheyensis* and *Geobacillus kaustophilus*, this third genomic information of SK37 will provide better understanding of the mechanism of adaptation to extreme environments through their transport systems, adaptive responses, regulatory networks, and metabolisms. In addition, amino acid catabolic pathways responsible for fish sauce flavor compounds and energy production which are industrial relevant characteristics have been predicted. In term of industrial perspective, as the SK37 has been successfully utilized as a starter culture for fish sauce production, the information of genomic sequence should facilitate the insightful understanding as how this bacterium implements its complex enzyme system to efficiently accelerate the fermentation time.

5.6 References

Alpert, C. A., Crutz-Le Coq, A. M., Malleret, C., and Zagorec, M. (2003). Characterization of a theta-type plasmid from *Lactobacillus sakei*: a potential basis for low-copy-number vectors in lactobacilli. **Appl. Environ. Microbiol.** 69:(9) 5574-5584.

Atsumi, S., Hanai, T., and Liao, J. C. (2008). Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. **Nature.** 451:(7174) 86-89.

Atsumi, S., Li, Z., and Liao, J. C. (2009). Acetolactate synthase from *Bacillus subtilis* serves as a 2-ketoisovalerate decarboxylase for isobutanol biosynthesis in *Escherichia coli*. **Appl. Environ. Microbiol.** 75:(19) 6306-6311.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. **BMC Genomics.** 9: 75.

Bachmann, B. J. (1990). Linkage map of *Escherichia coli* K-12, edition 8. **Microbiol. Rev.** 54:(2) 130-197.

Bainton, R. J., Kubo, K. M., Feng, J. N., and Craig, N. L. (1993). Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified in vitro system. **Cell.** 72:(6) 931-943.

Bastos, M. C., and Murphy, E. (1988). Transposon Tn554 encodes three products required for transposition. **Embo. J.** 7:(9) 2935-2941.

Bindokas, V. P., and Miller, R. J. (1995). Excitotoxic degeneration is initiated at non-random sites in cultured rat cerebellar neurons. **J. Neurosci.** 15:(11) 6999-7011.

Blakely, G., Colloms, S., May, G., Burke, M., and Sherratt, D. (1991). *Escherichia coli* XerC recombinase is required for chromosomal segregation at cell division. **New Biol.** 3:(8) 789-798.

Boekhorst, J., Wels, M., Kleerebezem, M., and Siezen, R. J. (2006). The predicted secretome of *Lactobacillus plantarum* WCFS1 sheds light on interactions with its environment. **Microbiology.** 152:(Pt 11) 3175-3183.

Brandenburg, J. L., Wray, L. V., Jr., Beier, L., Jarmer, H., Saxild, H. H., and Fisher, S. H. (2002). Roles of PucR, GlnR, and TnrA in regulating expression of the *Bacillus subtilis* ure P3 promoter. **J. Bacteriol.** 184:(21) 6060-6064.

Bursy, J., Pierik, A. J., Pica, N., and Bremer, E. (2007). Osmotically induced synthesis of the compatible solute hydroxyectoine is mediated by an evolutionarily conserved ectoine hydroxylase. **J. Biol. Chem.** 282:(43) 31147-31155.

Cann, A. F., and Liao, J. C. (2010). Pentanol isomer synthesis in engineered microorganisms. **Appl. Microbiol. Biotechnol.** 85:(4) 893-899.

Carrigan, C. M., Haarsma, J. A., Smith, M. T., and Wake, R. G. (1987). Sequence features of the replication terminus of the *Bacilus subtilis* chromosome. **Nucleic Acids Res.** 15: 8501-8509.

Casjens, S. (2003). Prophages and bacterial genomics: what have we learned so far? **Mol. Microbiol.** 49:(2) 277-300.

Chaiyanan, S., Maugel, T., Huq, A., Robb, F. T., and Colwell, R. R. (1999). Polyphasic taxonomy of a novel halobacillus, *Halobacillus thailandensis* sp. nov. isolated from fish sauce. **System Appl. Microbiol.** 22: 360-365.

Clerget, M. (1991). Site-specific recombination promoted by a short DNA segment of plasmid R1 and by a homologous segment in the terminus region of the *Escherichia coli* chromosome. **New Biol.** 3:(8) 780-788.

De Castro, R. E., Maupin-Furlow, J. A., Gimenez, M. I., Herrera Seitz, M. K., and Sanchez, J. J. (2006). Haloarchaeal proteinases and proteolytic systems. **FEMS Microbiol Rev.** 30:(1) 17-35.

del Solar, G., Acebo, P., and Espinosa, M. (1995). Replication control of plasmid pLS1: efficient regulation of plasmid copy number is exerted by the combined action of two plasmid components, CopG and RNA II. **Mol. Microbiol.** 18:(5) 913-924.

del Solar, G., and Espinosa, M. (1992). The copy number of plasmid pLS1 is regulated by two trans-acting plasmid products: the antisense RNA II and the repressor protein, RepA. **Mol. Microbiol. 6**:(1) 83-94.

Ehrlich, S. D., Bruand, C., Sozhamannan, S., Dabert, P., Gros, M. F., Janniere, L., et al. (1991). Plasmid replication and structural stability in *Bacillus subtilis*. **Res. Microbiol.** 142:(7-8) 869-873.

Fisher, S. H. (1999). Regulation of nitrogen metabolism in *Bacillus subtilis*: vive la difference! **Mol. Microbiol.** 32:(2) 223-232.

Freist, W., Gauss, D. H., Ibba, M., and Soll, D. (1997). Glutaminyl-tRNA synthetase. **Biol. Chem.** 378:(10) 1103-1117.

Fujita, Y. (2009). Carbon catabolite control of the metabolic network in *Bacillus subtilis*. **Biosci. Biotechnol. Biochem.** 73:(2) 245-259.

Fukami, K., Ishiyama, S., Yaguramaki, H., Masuzawa, T., Nabeta, Y., Endo, K., et al. (2002). Identification of distinctive volatile compounds in fish sauce. **J. Agric. Food. Chem.** 50:(19) 5412-5416.

Fukuchi, S., Yoshimune, K., Wakayama, M., Moriguchi, M., and Nishikawa, K. (2003). Unique amino acid composition of proteins in halophilic bacteria. **J. Mol. Biol.** 327:(2) 347-357.

Galinski, E. A., and Truper, H. G. (1994). Microbial behavior in salt-stressed ecosystems. **FEMS Microbiol. Rev.** 15: 95–108.

Gomes, J., and Steiner, W. (2004). The biocatalytic potential of extremophiles and extremozymes. **Food Technol. Biotechnol.** 42: 223-235.

Gross, C. A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J., et al. (1998). The functional and regulatory roles of sigma factors in transcription. **Cold Spring Harb. Symp. Quant Biol.** 63: 141-155.

Haard, N. F., and Simpson, B. K. (1994). Proteinases from aquatic organisms and their uses in the seafood industry. In A. M. Martin (ed.). **Fisheries processing: biotechnological applications** (pp. 132-154). London: Chapman and Hall.

Helmann, J. D. (2002). The extracytoplasmic function (ECF) sigma factors. **Adv. Microb. Physiol.** 46: 47-110.

Hiraga, K. Y. N., Namwong, S., Tanasupawat, S., Takada, K., and Oda, K. (2005). Purification and characterization of serine proteinase from a halophilic bacterium, *Filobacillus* sp. RF2-5. **Biosci. Biotechnol. Biochem.** 69:(1) 38-44.

Hoch, J. A. (1993). spo0 genes, the phosphorelay and the initiation of sporulation. In A. L. Sonenshein, J. A. Hoch, and R. Losick (eds.). ***Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics** (pp. 747–756). Washington, D.C.: American Socitety for Microbiology.

Hough, W. D., and Danson, J. M. (1999). Extremozymes. **Curr. Opin. Chem. Biol.** 3: 39-46.

Janausch, I. G., Zientz, E., Tran, Q. H., Kroger, A., and Unden, G. (2002). C4-dicarboxylate carriers and sensors in bacteria. **Biochim. Biophys. Acta.** 1553:(1-2) 39-56.

Johnson, K. G., Lanthier, P. H., and Gochnauer, M. B. (1986). Studies of two strains of *Actinopolyspora halophila*, an extremely halophilic actinomycete. **Arch. Microbiol.** 143: 370-378.

Kallio, P. T., Fagelson, J. E., Hoch, J. A., and Strauch, M. A. (1991). The transition state regulator Hpr of *Bacillus subtilis* is a DNA-binding protein. **J. Biol. Chem.** 266:(20) 13411-13417.

Kaul, S., Mohanty, B. K., Sahoo, T., Patel, I., Khan, S. A., and Bastia, D. (1994). The replication terminator protein of the gram-positive bacterium *Bacillus subtilis* functions as a polar contrahelicase in gram-negative *Escherichia coli*. **Proc. Natl. Acad. Sci. U S A.** 91:(23) 11143-11147.

Kehrenberg, C., and Schwarz, S. (2006). Distribution of florfenicol resistance genes fexA and cfr among chloramphenicol-resistant Staphylococcus isolates. **Antimicrob. Agents Chemother.** 50:(4) 1156-1163.

Kempf, B., and Bremer, E. (1998). Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. **Arch. Microbiol.** 170:(5) 319-330.

Kim, H. J., Roux, A., and Sonenshein, A. L. (2002). Direct and indirect roles of CcpA in regulation of *Bacillus subtilis* Krebs cycle genes. **Mol. Microbiol.** 45:(1) 179-190.

Kim, M., Nishiyama, Y., Mura, K., Tokue, C., and Arai, S. (2004). Gene cloning and characterization of a *Bacillus vietnamensis* Metalloproteinase. **Biosci. Biotechnol. Biochem.** 68:(7) 1533-1540.

Kloosterman, T. G., Hendriksen, W. T., Bijlsma, J. J., Bootsma, H. J., van Hijum, S. A., Kok, J., et al. (2006). Regulation of glutamine and glutamate metabolism by GlnR and GlnA in *Streptococcus pneumoniae*. **J. Biol. Chem.** 281:(35) 25097-25109.

Kodgire, P., Dixit, M., and Rao, K. K. (2006). ScoC and SinR negatively regulate epr by corepression in *Bacillus subtilis*. **J. Bacteriol.** 188:(17) 6425-6428.

Koide, A., Perego, M., and Hoch, J. A. (1999). ScoC regulates peptide transport and sporulation initiation in *Bacillus subtilis*. **J. Bacteriol.** 181:(13) 4114-4117.

Kroger, A., Geisler, V., Lemma, E., Thesis, F., and Longer, R. (1992). *Bacterial fumarate* respiration. **Arch. Microbiol.** 158: 311-314.

Krulwich, T. A., Ito, M., and Guffanti, A. A. (2001). The Na(+)-dependence of alkaliphily in *Bacillus*. **Biochim. Biophys. Acta.** 1505:(1) 158-168.

Kuempel, P. L., Henson, J. M., Dircks, L., Tecklenburg, M., and Lim, D. F. (1991). dif, a recA-independent recombination site in the terminus region of the chromosome of *Escherichia coli*. **New Biol.** 3:(8) 799-811.

Kuhlmann, A. U., Bursy, J., Gimpel, S., Hoffmann, T., and Bremer, E. (2008). Synthesis of the compatible solute ectoine in *Virgibacillus pantothenticus* is triggered by high salinity and low growth temperature. **Appl. Environ. Microbiol.** 74:(14) 4560-4563.

Kushner, D. J. (1978). Life in high salt and solute concentrations: halophilic bacteria. In D. J. Kushner (ed.). **Microbial life in extreme environments** (pp. 317-368). London: Academic Press, Ltd.

Lancaster, C. R., Sauer, U. S., Gross, R., Haas, A. H., Graf, J., Schwalbe, H., et al. (2005). Experimental support for the "E pathway hypothesis" of coupled transmembrane e- and H^+ transfer in dihemic quinol:fumarate reductase. **Proc. Natl. Acad. Sci. U S A.** 102:(52) 18860-18865.

Lang, F. (2007). Mechanisms and significance of cell volume regulation. **J. Am. Coll. Nutr.** 26:(5 Suppl) 613S-623S.

Larroy, C., Fernandez, M. R., Gonzalez, E., Pares, X., and Biosca, J. A. (2002). Characterization of the *Saccharomyces cerevisiae* YMR318C (ADH6) gene product as a broad specificity NADPH-dependent alcohol dehydrogenase: relevance in aldehyde reduction. **Biochem. J.** 361:(Pt 1) 163-172.

Laurent, D. (2006). Microbial Production of Food Grade Pigments. **Food Technol. Biotechnol.** 44:(3) 313-321.

Maghnouj, A., de Sousa Cabral, T. F., Stalon, V., and Vander Wauven, C. (1998). The arcABDC gene cluster, encoding the arginine deiminase pathway of *Bacillus licheniformis*, and its activation by the arginine repressor argR. **J. Bacteriol.** 180:(24) 6468-6475.

Margesin, R., and Schiner, F. (2001). Potential of halotolerant and halophilic microorganisms for biotechnology. **Extremophiles.** 5: 73-83.

Marilley, L., and Casey, M. G. (2004). Flavours of cheese products: metabolic pathways, analytical tools and identification of producing strains. **Int. J. Food Microbiol.** 90:(2) 139-159.

Mason, C. A., Dunner, J., Indra, P., and Colangelo, T. (1999). Heat-induced expression and chemically induced expression of the *Escherichia coli* stress protein HtpG are affected by the growth environment. **Appl. Environ. Microbiol.** 65:(8) 3433-3440.

McLean, M. J., Wolfe, K. H., and Devine, K. M. (1998). Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. **J. Mol. Evol.** 47:(6) 691-696.

McSweeney, P. L. H., and Sousa, M. J. (2000). Biochemical pathways for the production of flavour compounds in cheeses during ripening: A review. **INRA, EDP Sciences.** 80: 293-324.

Meijer, W. J., Wisman, G. B., Terpstra, P., Thorsted, P. B., Thomas, C. M., Holsappel, S., et al. (1998). Rolling-circle plasmids from *Bacillus subtilis*: complete nucleotide sequences and analyses of genes of pTA1015, pTA1040, pTA1050 and pTA1060, and comparisons with related plasmids from gram-positive bacteria. **FEMS Microbiol. Rev.** 21:(4) 337-368.

Molle, V., Nakaura, Y., Shivers, R. P., Yamaguchi, H., Losick, R., Fujita, Y., et al. (2003). Additional targets of the *Bacillus subtilis* global regulator CodY identified by chromatin immunoprecipitation and genome-wide transcript analysis. **J. Bacteriol.** 185:(6) 1911-1922.

Mongodin, E. F., Nelson, K. E., Daugherty, S., Deboy, R. T., Wister, J., Khouri, H., et al. (2005). The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. **Proc. Natl. Acad. Sci. U S A.** 102:(50) 18147-18152.

Morel, F., Frot-Coutaz, J., Aubel, D., Portalier, R., and Atlan, D. (1999). Characterization of a prolidase from *Lactobacillus delbrueckii* subsp. bulgaricus CNRZ 397 with an unusual regulation of biosynthesis. **Microbiology.** 145:(Pt 2) 437-446.

Morel, F., Lamarque, M., Bissardon, I., Atlan, D., and Galinier, A. (2001). Autoregulation of the biosynthesis of the CcpA-like protein, PepR1, in *Lactobacillus delbrueckii* subsp bulgaricus. **J. Mol. Microbiol. Biotechnol.** 3:(1) 63-66.

Moriya, S., Imai, Y., Hassan, A. K., and Ogasawara, N. (1999). Regulation of initiation of *Bacillus subtilis* chromosome replication. **Plasmid.** 41:(1) 17-29.

Murphy, E., Huwyler, L., and de Freire Bastos Mdo, C. (1985). Transposon Tn554: complete nucleotide sequence and isolation of transposition-defective and antibiotic-sensitive mutants. **Embo. J.** 4:(12) 3357-3365.

Nagano, H., and To, K. A. (2000). Purification of collagenase and specificity of its related enzyme from *Bacillus subtilis* FS-2. **Biosci. Biotechnol. Biochem.** 64: 181-183.

Namwong, S., Hiraga, K., Takada, K., Tsunemi, M., Tanasupawat, S., and Oda, K. (2006). A halophilic serine proteinase from *Halobacillus* sp. SR5-3 isolated from fish sauce: purification and characterization. **Biosci. Biotechnol. Biochem.** 70:(6) 1395-1401.

Nawong, S. (2006). **Isolation, selection and identification of proteinase-producing bacteria from fish sauce fermentation to be used as starter cultures**. M.S. thesis. Suranaree University of Technology, Naknon-Ratchasima.

Ng, W. V., Kennedy, S. P., Mahairas, G. G., Berquist, B., Pan, M., Shukla, H. D., et al. (2000). Genome sequence of *Halobacterium* species NRC-1. **Proc. Natl. Acad. Sci. U S A.** 97:(22) 12176-12181.

Nicoll, D. A., Longoni, S., and Philipson, K. D. (1990). Molecular cloning and functional expression of the cardiac sarcolemmal Na^{+}-Ca^{2+} exchanger. **Science.** 250:(4980) 562-565.

Nielsen, D. R., Leonard, E., Yoon, S. H., Tseng, H. C., Yuan, C., and Prather, K. L. (2009). Engineering alternative butanol production platforms in heterologous bacteria. **Metab. Eng.** 11:(4-5) 262-273.

Oppon, J. C., Sarnovsky, R. J., Craig, N. L., and Rawlings, D. E. (1998). A Tn7-like transposon is present in the glmUS region of the obligately chemoautolithotrophic bacterium *Thiobacillus ferrooxidans*. **J. Bacteriol.** 180:(11) 3007-3012.

Oren, A. (2008). Microbial life at high salt concentrations: phylogenetic and metabolic diversity. **Saline Systems.** 4: 2.

Oren, A., Heldal, M., Norland, S., and Galinski, E. A. (2002). Intracellular ion and organic solute concentrations of the extremely halophilic bacterium *Salinibacter ruber*. **Extremophiles.** 6:(6) 491-498.

Padan, E., Venturi, M., Gerchman, Y., and Dover, N. (2001). Na^{+}/H^{+} antiporters. **Biochimi. Biophys. Acta.** 1505: 144-157.

Park, J. N., Fukumoto, Y., Fujita, E., Tanaka, T., Washio, T., Otsuka, S., et al. (2001). Chemical composition of fish sauces produced in Southeast and East Asian countries. **J. Food. Comp. Anal.** 14: 113-125.

Payandeh, J., Fujihashi, M., Gillon, W., and Pai, E. F. (2006). The crystal structure of (S)-3-O-geranylgeranylglyceryl phosphate synthase reveals an ancient fold for an ancient enzyme. **J. Biol. Chem.** 281:(9) 6070-6078.

Peralta, R. R., Shimoda, M., and Osajima, Y. (1996). Further identification of volatile compounds in fish sauce. **J. Agric. Food Chem.** 44: 3606–3610.

Perego, M., and Hoch, J. A. (1988). Sequence analysis and regulation of the hpr locus, a regulatory gene for proteinase production and sporulation in *Bacillus subtilis*. **J. Bacteriol.** 170:(6) 2560-2567.

Phrommao, E., Rodtong, S., and Yongsawatdigul, J. (2010). Identification of novel halotolerant bacillopeptidase F-like proteinases from a moderately halophilic bacterium, *Virgibacillus* sp. SK37. **J. Appl. Microbiol.** 110:(1) 191-201.

Priebe, S. D., and Lacks, S. A. (1989). Region of the streptococcal plasmid pMV158 required for conjugative mobilization. **J. Bacteriol.** 171:(9) 4778-4784.

Priest, F. G. (1993). Systematics and ecology of *Bacillus*. In A. L. Sonenshein, J. A. Hoch, and R. Losick (eds.). ***Bacillus subtilis* and other Gram-positive bacteria: biochemistry, physiology, and molecular genetics** (pp. 3-16). Washington, D. C.: ASM Press.

Reitzer, L. (2003). Nitrogen assimilation and global regulation in *Escherichia coli*. **Annu. Rev. Microbiol.** 57: 155-176.

Rey, M. W., Ramaiya, P., Nelson, B. A., Brody-Karpin, S. D., Zaretsky, E. J., Tang, M., et al. (2004). Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species. **Genome Biol.** 5:(10) R77.

Richter, R., Hejazi, M., Kraft, R., Ziegler, K., and Lockau, W. (1999). Cyanophycinase, a peptidase degrading the cyanobacterial reserve material multi-L-arginyl-poly-L-aspartic acid (cyanophycin): molecular cloning of the gene of *Synechocystis* sp. PCC 6803, expression in *Escherichia coli*, and biochemical characterization of the purified enzyme. **Eur. J. Biochem.** 263:(1) 163-169.

Rogers, M., Ekaterinaki, N., Nimmo, E., and Sherratt, D. (1986). Analysis of Tn7 transposition. **Mol. Gen. Genet.** 205:(3) 550-556.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A., et al. (2000). Artemis: sequence visualization and annotation. **Bioinformatics.** 16:(10) 944-945.

Saisithi, P. (1994). Traditional fermented fish: fish sauce production. In A. M. Martin (ed.). **Fisheries processing biotechnological application** (pp. 111–131). London: ChapmanandHall.

Salas, J. J., Garcia-Gonzalez, D. L., and Aparicio, R. (2006). Volatile compound biosynthesis by green leaves from an *Arabidopsis thaliana* hydroperoxide lyase knockout mutant. **J. Agric. Food Chem.** 54:(21) 8199-8205.

Sao-Jose, C., Baptista, C., and Santos, M. A. (2004). *Bacillus subtilis* operon encoding a membrane receptor for bacteriophage SPP1. **J. Bacteriol.** 186:(24) 8337-8346.

Saum, S. H., and Muller, V. (2008). Regulation of osmoadaptation in the moderate halophile *Halobacillus halophilus*: chloride, glutamate and switching osmolyte strategies. **Saline Systems.** 4: 4.

Schick, J., Weber, B., Klein, J. R., and Henrich, B. (1999). PepR1, a CcpA-like transcription regulator of *Lactobacillus delbrueckii* subsp. lactis. **Microbiology.** 145:(Pt 11) 3147-3154.

Schreier, H. J., Brown, S. W., Hirschi, K. D., Nomellini, J. F., and Sonenshein, A. L. (1989). Regulation of *Bacillus subtilis* glutamine synthetase gene expression by the product of the glnR gene. **J. Mol. Biol.** 210: (1) 51-63.

Setyorini, E., Takenaka, S., Murakami, S., and Aoki, K. (2006). Purification and characterization of two novel halotolerant extracellular proteinases from *Bacillus subtilis* strain FP-133. **Biosci. Biotechnol. Biochem.** 70:(2) 433-440.

Severi, E., Hood, D. W., and Thomas, G. H. (2007). Sialic acid utilization by bacterial pathogens. **Microbiology.** 153:(Pt 9) 2817-2822.

Silla Santos, M. H. (1996). Biogenic amines: their importance in foods. **Int. J. Food Microbiol.** 29:(2-3) 213-231.

Sinsuwan, S., Rodtong, S., and Yongsawatdigul, J. (2007). NaCl-activated extracellular proteinase from *Virgibacillus* sp. SK37 isolated from fish sauce fermentation. **J. Food Sci.** 72:(5) 264-269.

Sinsuwan, S., Rodtong, S., and Yongsawatdigul, J. (2008). Characterization of Ca^{2+}-activated cell-bound proteinase from *Virgibacillus* sp. SK37 isolated from fish sauce fermentation. **LWT-Food Sci. Tech.** 41: 2166-2174.

Sonenshein, A. L. (2007). Control of key metabolic intersections in *Bacillus subtilis*. **Nat. Rev. Microbiol.** 5:(12) 917-927.

Sorvig, E., Skaugen, M., Naterstad, K., Eijsink, V. G., and Axelsson, L. (2005). Plasmid p256 from *Lactobacillus plantarum* represents a new type of replicon

in lactic acid bacteria, and contains a toxin-antitoxin-like plasmid maintenance system. **Microbiology.** 151:(Pt 2) 421-431.

Sreedhar, A. S., Kalmar, E., Csermely, P., and Shen, Y. F. (2004). Hsp90 isoforms: functions, expression and clinical importance. **FEBS Lett.** 562:(1-3) 11-15.

Strauch, M. A. (1993). AbrB, a transition state regulator. In A. L. Sonenshein, J. A. Hoch, and R. Losick (eds.). ***Bacillus subtilis*** **and other gram-positive bacteria: biochemistry, physiology, and molecular genetics** (pp. 757–764). Washington, D.C.: ASM Press.

Stucky, K., Schick, J., Klein, J. R., Henrich, B., and Plapp, R. (1996). Characterization of pepR1, a gene coding for a potential transcriptional regulator of *Lactobacillus delbrueckii* subsp. lactis DSM7290. **FEMS Microbiol. Lett.** 136:(1) 63-69.

Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., et al. (2000). Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. **Nucleic Acids Res.** 28:(21) 4317-4331.

Takami, H., Takaki, Y., Chee, G. J., Nishi, S., Shimamura, S., Suzuki, H., et al. (2004). Thermoadaptation trait revealed by the genome sequence of thermophilic *Geobacillus kaustophilus*. **Nucleic Acids Res.** 32:(21) 6292-6303.

Takami, H., Takaki, Y., and Uchiyama, I. (2002). Genome sequence of *Oceanobacillus iheyensis* isolated from the Iheya Ridge and its unexpected adaptive capabilities to extreme environments. **Nucleic Acids Res.** 30:(18) 3927-3935.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. **Mol. Biol. Evol.** 24:(8) 1596-1599.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. **Nucleic Acids Res.** 25:(24) 4876-4882.

Tran, L. H., and Nagano, H. (2002). Isolation and characteristics of *Bacillus subtilis* CN2 and its collagenase production. **J. Food Sci.** 67:(3) 1184-1187.

Uchida, H., Kondo, D., Yamashita, S., Tanaka, T., Tran, L. H., Nagano, H., et al. (2004). Purification and properties of a proteinase produced by *Bacillus subtilis* CN2 isolated from a Vietnamese fish sauce. **J. Microbiol. Biotechnol.** 20: 579-582.

Van den Berg, B. (2003). Extremophiles as a source for novel enzymes. **Curr. Opin. Microbiol.** 6: 213-218.

van Passel, M. W., Bart, A., Luyf, A. C., van Kampen, A. H., and van der Ende, A. (2006). Compositional discordance between prokaryotic plasmids and host chromosomes. **BMC Genomics.** 7: 26.

Ventosa, A., and Nieto, J. J. (1995). Biotechnological applications and potentialities of halophilic microorganisms. **World J. Microbiol. Biotechnol.** 11: 85–94.

Ventosa, A., Nieto, J. J., and Oren, A. (1998). Biology of moderately halophilic aerobic bacteria. **Microbiol. Mol. Biol. Rev.** 62:(2) 504-544.

Vimr, E. R., Kalivoda, K. A., Deszo, E. L., and Steenbergen, S. M. (2004). Diversity of microbial sialic acid metabolism. **Microbiol. Mol. Biol. Rev.** 68:(1) 132-153.

Warner, J. B., and Lolkema, J. S. (2003). CcpA-dependent carbon catabolite repression in bacteria. **Microbiol. Mol. Biol. Rev.** 67:(4) 475-490.

Wray, L. V., Jr., Ferson, A. E., Rohrer, K., and Fisher, S. H. (1996). TnrA, a transcription factor required for global nitrogen regulation in *Bacillus subtilis*. **Proc. Natl. Acad. Sci. U S A.** 93:(17) 8841-8845.

Wray, L. V., Jr., Zalieckas, J. M., and Fisher, S. H. (2001). *Bacillus subtilis* glutamine synthetase controls gene expression through a protein-protein interaction with transcription factor TnrA. **Cell.** 107:(4) 427-435.

Yen, M. R., Lin, N. T., Hung, C. H., Choy, K. T., Weng, S. F., and Tseng, Y. H. (2002). oriC region and replication termination site, dif, of the *Xanthomonas campestris* pv. campestris 17 chromosome. **Appl. Environ. Microbiol.** 68:(6) 2924-2933.

Yongsawatdigul, J., Rodtong, S., and Raksakulthai, N. (2007). Acceleration of Thai fish sauce fermentation using proteinases and bacterial starter cultures. **J. Food Sci.** 72:(9) 382-390.

Yoshikawa, H., and Wake, R. G. (1993). Initiation and termination of chromosome replication. In A. L. Sonenshein, J. A. Hoch, and R. Losick (eds.). ***Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics** (pp. 507–528). Washington, D.C.: ASM Press.

Yu, S. P., and Choi, D. W. (1997). Na^{+}-Ca^{2+} exchange currents in cortical neurons: concomitant forward and reverse operation and effect of glutamate. **Eur. J. Neurosci.** 9:(6) 1273-1281.

CHAPTER VI

SUMMARY

Virgibacillus sp. SK37 is a proteinase producer with high level of extracellular proteinases production. Three major extracellular subtilisin-like proteinases with MM of 19, 34, and 44 kDa were partially purified, characterized. These enzymes showed activity at pH 8 and 55-60°C and were activated by NaCl with optimum activity at 4.0 M NaCl. Analysis of protein identification by mass spectrometry using peptide mass fingerprinting (PMF) and partial amino acid sequencing techniques indicated that these enzymes are novel and related to bacillopeptidase F, an extracellular enzyme found in most *Bacillus* species.

An intracellular Ca^{2+}-dependent moderately halotolerant subtilisin-like enzyme showing high homology to alkaline proteinase-X (AprX) enzymes were obtained from activity-based screening of SK37's genomic library constructed in *E. coli* DH10B. It showed optimum catalytic activity at pH 9.5, 55°C and 1.0 M NaCl. Based on phylogenetic tree analysis, this enzyme is proposed to be in a novel family of the subtilase superfamily, called AprX. The concomitant of haloterant ability of all aforementioned enzymes indicates the adaptation of proteinases in SK37 against high saline environments throughout secreted proteins, cell-bound, and in cytoplasm. This observation might simply explain the ability and adaptability of the utilization of exogenous nitrogen sources over a wide range of salt concentrations of this bacterium.

The SK37 is the first strain in genus *Virgibacillus* that the genomic sequence has been determined. According to comparative genomic analysis, like other moderately halophilic bacteria, SK37 possesses a number of protein coding sequences (CDSs) associated with osmoregulatory systems, including inorganic ion transport systems and compatible solute synthesis and uptake, which are responsible for balancing cellular osmotic stress in a high saline environment. The lack of genes encoding *tnrA*, *hpr*, and *scoC* and the evolutionary analysis of CcpA in SK37 imply the different regulatory nutrient-stress responses compared to other *Bacillus*, which possibly help survive in a saline protein-rich environment of fish sauce. Amino acid catabolic pathways responsible for fish sauce flavor development and energy production are efficiently employed in this bacterium. Since SK37 is a moderately halophilic bacterium, its genomic information will shed light on how this bacterium adapts to hypersaline environments and could be a model organism in the scope of halophilic adaptability in bacteria.

APPENDIX

Supporting materials for Chapter V

Supporting material Table 1 General information of *Bacillus* related species used for comparative analysis in this study.

No.	Bacterial name	Acronym	Genome[a] sequence	%GC	Origin	Characteristics
1	*Bacillus amyloliquefaciens* FZB42	Bac amy FZ2	F	46.48	Soil	Mesophile, proteinase producer
2	*Bacillus anthracis* Ames Ancestor	Bac ant Anr	F	35.24	Soil	Pathogen
3	*Bacillus cereus* ATCC 10987	Bac cer 107	F	35.50	Soil, dairy	Food poisoning
4	*Bacillus clausii* KSM-K16	Bac cla KS6	F	44.75	Soil	Alkaliphile, proteinase producer
5	*Bacillus coagulans* 36D1	Bac coa 361	D	46.49	Intestinal microflora	Acidphile, bioenergy
6	*Bacillus coahuilensis* M4-4	Bac coa M44	D	38.05	Freshwater	Mesophile, halotolerance
7	*Bacillus halodurans* C-125	Bac hal C-5	F	49.69	Freshwater, soil	Proteinase producer
8	*Bacillus licheniformis* ATCC 14580 (Novozymes)	Bac lic (N)	F	46.20	Soil	Pathogen, proteinase producer
9	*Bacillus mycoides* DSM2048	Bac myc DS8	D	35.21	Soil	Mesophile
10	*Bacillus pseudofirmus* OF4	Bac pse OF4	F	39.86	Soil	Alkaliphile, mesophile
11	*Bacillus pseudomycoides* DSM12442	Bac pse DS2	D	35.37	Soil	Mesophile
12	*Bacillus pumilus* SAFR-032	Bac pum SA2	F	41.29	Soil	Mesophile, biomass degradator
13	*Bacillus selenitireducens* MLS10	Bac sel ML0	D	48.65	Freshwater	Alkaliphile, mesophile, halotolerance
14	*Bacillus* sp. B-14905	Bac sp. B15	D	37.56	Marine (surface)	Mesophile
15	*Bacillus* sp. NRRL B-14911	Bac sp. NR1	D	45.79	Marine (10 m depth)	Mesophile
16	*Bacillus* sp. SG-1	Bac sp. SG1	D	42.09	Marine (10 m depth)	Mesophile
17	*Bacillus subtilis* subsp. subtilis str. 168	Bac sub 168	F	43.51	Soil	Mesophile, proteinase producer
18	*Bacillus thuringiensis* sv konkukian 97-27	Bac thu 977	F	35.37	Soil	Mesophile
19	*Bacillus weihenstephanensis* KBAB4	Bac wei KB4	F	35.48	Soil	Mesophile
20	*Geobacillus kaustophilus* HTA426	Geo kau HT6	F	51.98	Marine (deep sea)	Thermophile, halolerance
21	*Geobacillus thermodenitrificans* NG80-2	Geo the NG2	F	48.46	Freshwater, oil field	Thermophile
22	*Oceanobacillus iheyensis* HTE831	Oce ihe HT1	F	35.68	Marine (deep sea)	Alkaiphile, halotolerance

[a] F = finished sequence, D = draft sequence.

Note: All sequences and information listed in this table were obtained from IMG database (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi)

Supporting material Table 2 Orthologous groups of proteins involving in transportation systems of SK37 compared to related *Bacillus* species

Function	COG	Name	Vir sp. SK37	Bac amy FZ2	Bac ant Anr	Bac cer 107	Bac cla KS6	Bac hal C-5	Bac lic (N)	Bac pse OF4	Bac pum SA2	Bac sub 168	Bac thu Ham	Bac thu 977	Bac wei KB4	Geo kau HT6	Geo the NG2	Oce ihe HT1
Nucleotide transports	COG1972	Nucleoside permease	1	3	7	7	1	1	3	1	1	3	7	7	8	0	0	1
	COG1457	Purine-cytosine permease and related proteins	3	1	0	0	2	1	1	1	0	1	0	0	0	0	0	0
	COG2233	Xanthine/uracil permeases	4	3	4	4	1	1	3	2	3	5	4	4	5	2	2	2
Coenzyme transports	COG1120	ABC-type cobalamin/Fe^{3+}-siderophores transport systems, ATPase components	4	3	5	5	3	5	3	3	5	4	5	5	5	3	3	3
	COG4145	Na+/panthothenate symporter	1	0	1	1	0	0	0	2	0	0	1	1	1	0	1	1
	COG2978	Putative p-aminobenzoyl-glutamate transporter	2	0	0	0	2	2	0	2	0	0	0	0	0	0	1	4
Inorganic ions transports	COG1122	ABC-type cobalt transport system, ATPase component	2	2	2	2	3	2	2	2	2	2	2	2	2	3	3	2
	COG0619	ABC-type cobalt transport system, permease component CbiQ and related transporters	2	2	3	3	2	1	2	1	3	2	3	3	3	3	3	2
	COG4604	ABC-type enterochelin transport system, ATPase component	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
	COG4607	ABC-type enterochelin transport system, periplasmic component	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
	COG4605	ABC-type enterochelin transport system, permease component	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
	COG4606	ABC-type enterochelin transport system, permease component	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
	COG1840	ABC-type Fe^{3+} transport system, periplasmic component	1	0	3	3	2	1	0	1	0	0	3	3	3	0	0	1
	COG1178	ABC-type Fe^{3+} transport system, permease component	1	0	2	2	2	1	0	1	0	0	2	2	2	0	0	1
	COG4594	ABC-type Fe^{3+}-citrate transport system, periplasmic component	1	1	1	0	0	0	1	0	0	2	1	1	1	0	0	0
	COG0614	ABC-type Fe^{3+}-hydroxamate transport system, periplasmic component	4	6	8	8	9	8	8	5	9	6	9	9	9	5	4	4
	COG0609	ABC-type Fe^{3+}-siderophore transport system, permease component	5	7	12	12	8	8	9	6	8	9	13	13	13	7	5	5
	COG1135	ABC-type metal ion transport system, ATPase component	2	1	3	3	3	1	2	2	1	1	3	2	3	1	2	4
	COG0803	ABC-type metal ion transport system, periplasmic component/surface adhesin	3	1	2	2	3	3	2	5	2	2	3	1	1	1	1	2
	COG1464	ABC-type metal ion transport system, periplasmic component/surface antigen	2	2	4	4	3	1	3	2	1	2	4	3	4	1	2	4
	COG2011	ABC-type metal ion transport system, permease component	2	1	3	3	3	1	2	2	1	1	3	2	3	1	2	4
	COG1121	ABC-type Mn^{2+}/Zn^{2+} transport systems, ATPase component	2	1	2	1	3	2	2	3	2	2	2	1	1	1	1	2
	COG1108	ABC-type Mn^{2+}/Zn^{2+} transport systems, permease components	2	1	2	1	4	2	2	4	2	3	2	1	1	1	1	3
	COG4148	ABC-type molybdate transport system, ATPase component	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Supporting material Table 2 (Continued)

Function	COG	Name	Vir sp. SK37	Bac amy FZ2	Bac ant Anr	Bac cer 107	Bac cla KS6	Bac hal C-5	Bac lic (N)	Bac pse OF4	Bac pum SA2	Bac sub 168	Bac thu Ham	Bac thu 977	Bac wei KB4	Geo kau HT6	Geo the NG2	Oce ihe HT1
Inorganic ions transports	COG4148	ABC-type molybdate transport system, ATPase component	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	COG0725	ABC-type molybdate transport system, periplasmic component	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0
	COG4149	ABC-type molybdate transport system, permease component	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	0
	COG1668	ABC-type Na^+ efflux pump, permease component	2	2	1	1	1	1	1	1	1	2	2	1	1	1	1	3
	COG1116	ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component	3	2	4	5	7	2	2	3	1	2	4	4	3	1	3	1
	COG0600	ABC-type nitrate/sulfonate/bicarbonate transport system, permease component	3	2	4	5	7	2	2	3	1	2	4	4	3	1	3	1
	COG0715	ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components	3	2	4	5	7	2	2	3	1	2	4	4	3	1	4	1
	COG1117	ABC-type phosphate transport system, ATPase component	2	3	1	1	2	1	2	2	2	3	1	1	1	1	1	2
	COG0226	ABC-type phosphate transport system, periplasmic component	1	1	2	2	1	1	1	1	1	1	2	2	2	2	1	1
	COG0573	ABC-type phosphate transport system, permease component	1	1	2	2	1	1	1	1	1	1	2	2	2	2	1	1
	COG0581	ABC-type phosphate transport system, permease component	1	1	2	2	1	1	1	1	1	1	2	2	2	2	1	1
	COG3638	ABC-type phosphate/phosphonate transport system, ATPase component	1	0	1	1	1	3	0	2	0	0	1	1	1	0	0	2
	COG3221	ABC-type phosphate/phosphonate transport system, periplasmic component	1	0	1	1	1	3	0	2	0	0	1	1	1	0	0	2
	COG3639	ABC-type phosphate/phosphonate transport system, permease component	2	0	2	2	2	6	0	3	0	0	2	2	2	0	0	4
	COG0530	Ca^{2+}/Na^+ antiporter	2	0	0	0	0	2	0	2	0	0	0	0	0	0	0	1
	COG0474	Cation transport ATPase	1	1	3	5	2	1	1	1	1	1	3	3	4	1	1	1
	COG2217	Cation transport ATPase	6	3	3	4	4	3	3	6	3	3	3	3	4	4	2	3
	COG2059	Chromate transport protein ChrA	3	2	1	1	3	3	3	3	2	2	1	2	1	3	2	5
	COG1230	Co/Zn/Cd efflux system component	1	1	1	2	1	0	1	0	0	1	1	1	1	3	2	1
	COG2807	Cyanate permease	3	1	0	0	2	0	1	2	2	1	0	0	1	1	1	2
	COG0471	Di- and tricarboxylate transporters	1	1	1	0	3	2	4	1	1	1	1	0	0	0	0	3
	COG0370	Fe^{2+} transport system protein B	1	0	1	1	0	1	3	0	1	1	3	3	3	4	3	1
	COG0735	Fe^{2+}/Zn^{2+} uptake regulation proteins	3	3	3	3	3	3	5	3	3	3	3	3	3	3	3	3
	COG0651	Formate hydrogenlyase subunit 3/Multisubunit Na^+/H^+ antiporter, MnhD subunit	1	1	0	0	2	1	1	1	1	1	0	0	0	0	1	2
	COG0569	K^+ transport systems, NAD-binding component	3	4	2	3	2	4	3	3	3	2	3	3	3	2	3	3

Supporting material Table 2 (Continued)

Function	COG	Name	Vir sp. SK37	Bac amy FZ2	Bac ant Anr	Bac cer 107	Bac cla KS6	Bac hal C-5	Bac Lic (N)	Bac pse OF4	Bac pum SA2	Bac sub 168	Bac thu Ham	Bac thu 977	Bac wei KB4	Geo kau HT6	Geo the NG2	Oce ihe HT1
Inorganic ions transports	COG2076	Membrane transporters of cations and cationic drugs	4	7	6	6	4	4	5	2	6	7	3	6	5	1	1	2
	COG2239	Mg/Co/Ni transporter MgtE (contains CBS domain)	1	1	0	0	1	4	1	3	1	1	0	0	0	0	1	1
	COG0598	Mg^{2+} and Co^{2+} transporters (corA)	1	2	2	3	0	0	2	0	1	2	3	3	3	3	3	0
	COG1009	Multisubunit Na^+/H^+ antiporter, MnhA subunit	1	1	2	2	2	1	2	1	1	2	2	2	1	2	3	2
	COG2111	Multisubunit Na^+/H^+ antiporter, MnhB subunit	1	2	0	0	2	1	1	1	1	1	0	0	0	0	2	3
	COG1006	Multisubunit Na^+/H^+ antiporter, MnhC subunit	1	1	0	0	2	1	1	1	1	1	0	0	0	0	1	2
	COG1863	Multisubunit Na+/H+ antiporter, MnhE subunit	1	1	0	0	2	1	1	1	1	1	0	0	0	0	1	2
	COG2212	Multisubunit Na^+/H^+ antiporter, MnhF subunit	1	1	0	0	2	1	1	1	1	1	0	0	0	0	1	2
	COG1320	Multisubunit Na+/H+ antiporter, MnhG subunit	1	1	0	0	2	1	1	1	1	1	0	0	0	0	1	2
	COG1757	Na^+/H^+ antiporter	5	4	4	4	4	2	3	4	1	2	4	4	4	2	1	3
	COG1055	Na^+/H^+ antiporter NhaD and related arsenite permeases	2	2	4	3	1	2	1	2	1	2	4	4	3	2	1	2
	COG1283	Na^+/phosphate symporter	1	1	1	1	2	2	2	5	1	1	1	1	1	1	1	1
	COG0025	NhaP-type Na^+/H^+ and K^+/H^+ antiporters	4	1	1	0	1	0	0	0	1	1	1	1	1	0	0	1
	COG3263	NhaP-type Na^+/H^+ and K^+/H^+ antiporters with a unique C-terminal domain	1	0	0	0	0	1	1	3	0	0	0	0	0	0	0	0
	COG2223	Nitrate/nitrite transporter	1	3	3	1	2	3	4	4	1	3	3	3	2	3	3	0
Inorganic ions transports	COG0053	Predicted Co/Zn/Cd cation transporters	2	5	2	2	1	2	1	2	2	3	2	2	2	2	2	1
	COG0428	Predicted divalent heavy-metal cations transporter	1	0	1	1	0	0	0	1	0	0	1	1	1	2	1	1
	COG2072	Predicted flavoprotein involved in K^+ transport	1	1	1	0	1	2	0	1	0	1	1	1	1	1	0	1
	COG4300	Predicted permease, cadmium resistance protein	1	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0
	COG4536	Putative Mg^{2+} and Co^{2+} transporter CorB	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
	COG0659	Sulfate permease and related transporters (MFS superfamily)	1	2	2	2	0	1	2	2	2	2	2	3	2	0	0	1
	COG1275	Tellurite resistance protein and related permeases	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	COG0168	Trk-type K^+ transport systems, membrane components	1	1	3	3	1	2	3	1	1	2	3	4	3	1	1	3
Amino acid transports	COG0765	ABC-type amino acid transport system, permease component	5	7	4	8	6	2	3	3	9	8	5	5	6	3	3	5
	COG0834	ABC-type amino acid transport/periplasmic component/domain	4	6	3	7	4	3	2	4	7	7	4	4	5	2	2	5

Supporting material Table 2 (Continued)

Function	COG	Name	Vir sp. SK37	Bac amy FZ2	Bac ant Anr	Bac cer 107	Bac cla KS6	Bac hal C-5	Bac lic (N)	Bac pse OF4	Bac pum SA2	Bac sub 168	Bac thu Ham	Bac thu 977	Bac wei KB4	Geo kau HT6	Geo the NG2	Oce ihe HT1
Amino acid transports	COG0747	ABC-type dipeptide transport system, periplasmic component	3	1	4	4	11	10	4	7	4	0	3	4	4	0	3	7
	COG0444	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	5	3	4	5	9	7	5	4	6	3	5	6	5	1	4	5
	COG0601	ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	6	3	7	8	10	7	6	4	6	3	8	8	8	2	6	5
	COG1173	ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	7	3	7	8	10	7	6	4	6	3	8	9	8	3	5	5
	COG4608	ABC-type oligopeptide transport system, ATPase component	3	2	6	5	9	6	6	3	2	2	5	6	2	1	3	5
	COG4166	ABC-type oligopeptide transport system, periplasmic component	8	3	11	12	1	0	2	0	2	2	12	14	12	1	1	1
	COG1126	ABC-type polar amino acid transport system, ATPase component	4	5	4	7	3	2	2	3	7	5	4	4	5	2	2	3
	COG4175	ABC-type proline/glycine betaine transport system, ATPase component	1	1	1	1	1	0	1	0	1	1	1	1	1	0	0	3
	COG4176	ABC-type proline/glycine betaine transport system, permease component	1	1	1	1	1	0	1	0	1	1	1	1	1	0	0	3
	COG1125	ABC-type proline/glycine betaine transport systems, ATPase components	2	2	1	1	2	0	1	1	2	2	1	1	1	1	1	1
	COG2113	ABC-type proline/glycine betaine transport systems, periplasmic components	3	1	1	2	3	0	1	0	1	1	1	1	1	0	0	4
	COG1174	ABC-type proline/glycine betaine transport systems, permease component	2	4	1	1	3	0	2	0	3	4	1	1	1	1	1	1
	COG3842	ABC-type spermidine/putrescine transport systems, ATPase components	1	0	2	3	3	1	1	1	1	0	3	3	3	2	2	2
	COG0531	Amino acid transporters	4	8	8	11	2	1	7	0	8	8	12	9	8	4	2	1
	COG1114	Branched-chain amino acid permeases	2	1	6	7	1	0	1	1	2	2	6	7	6	0	1	1
	COG3104	Dipeptide/tripeptide permease	1	1	3	3	0	0	1	0	1	1	3	3	5	2	1	0
	COG1115	Na^+/alanine symporter	4	2	6	6	3	5	6	3	3	4	7	6	6	1	1	3
	COG0786	Na^+/glutamate symporter	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2
	COG0591	Na^+/proline symporter	7	4	3	3	4	2	5	3	3	4	3	4	3	3	5	7
	COG1296	Predicted branched-chain amino acid permease (azaleucine resistance)	1	1	1	2	3	1	1	1	2	1	1	1	1	1	1	1
	COG1703	Putative periplasmic protein kinase ArgK and related GTPases of G3E family	1	0	0	0	0	2	0	1	0	0	0	0	0	2	2	0
Carbohydrate transports	COG1682	ABC-type polysaccharide/polyol phosphate export systems, permease component	1	1	2	1	1	2	1	1	1	1	1	1	1	0	1	1
	COG1134	ABC-type polysaccharide/polyol phosphate transport system, ATPase component	1	1	1	0	1	1	1	0	1	1	0	0	0	0	0	1
	COG1653	ABC-type sugar transport system, periplasmic component	6	3	1	2	26	16	8	2	3	6	1	1	2	3	10	8
	COG1879	ABC-type sugar transport system, periplasmic component	1	1	2	2	1	5	2	0	3	1	2	3	3	5	5	1
	COG0395	ABC-type sugar transport system, permease component	6	3	1	2	26	16	7	2	3	6	1	1	2	4	9	8

Supporting material Table 2 (Continued)

Function	COG	Name	Vir sp. SK37	Bac amy FZ2	Bac ant Anr	Bac cer 107	Bac cla KS6	Bac hal C-5	Bac lic (N)	Bac pse OF4	Bac pum SA2	Bac sub 168	Bac thu Ham	Bac thu 977	Bac wei KB4	Geo kau HT6	Geo the NG2	Oce ihe HT1
Carbohydrate transports	COG3839	ABC-type sugar transport systems, ATPase components	1	2	2	3	4	2	1	1	1	2	2	2	3	1	1	1
	COG1175	ABC-type sugar transport systems, permease components	6	3	2	3	22	13	7	3	3	6	2	2	3	4	10	7
	COG2814	Arabinose efflux permease	20	42	37	40	25	17	33	14	28	45	38	39	40	17	18	17
	COG0580	Carbohydrate transport, Glycerol uptake facilitator and related permeases (MIP Family)	2	1	2	3	1	1	1	1	1	1	2	2	2	1	1	1
	COG2610	H^+/gluconate symporter and related permeases	2	2	4	4	5	2	2	4	3	3	5	5	3	1	2	5
	COG1620	L-lactate permease	1	1	3	3	1	4	2	1	1	2	3	3	3	1	1	1
	COG1301	Na^+/H^+-dicarboxylate symporters	6	3	4	4	3	2	4	4	5	3	5	5	4	2	4	3
	COG2211	Na^+/melibiose symporter and related transporters	1	3	3	3	3	0	1	1	1	4	3	3	1	2	2	1
	COG0697	Permeases of the drug/metabolite transporter (DMT) superfamily	1	1	8	9	0	3	3	1	1	7	9	10	9	1	3	0
	COG2271	Sugar phosphate permease (glpT) glycerol-3-phosphate transporter	4	5	9	8	2	2	11	5	3	9	8	8	8	2	3	1
	COG1593	TRAP-type C4-dicarboxylate transport system, large permease component	2	0	0	0	9	3	1	7	1	0	0	0	0	1	1	6
	COG1638	TRAP-type C4-dicarboxylate transport system, periplasmic component	2	1	0	0	8	4	1	7	1	1	0	0	0	3	4	6
	COG3090	TRAP-type C4-dicarboxylate transport system, small permease component	1	0	0	0	8	3	1	5	0	0	0	0	0	1	2	5
	COG4666	TRAP-type uncharacterized transport system, fused permease components	2	0	0	0	1	3	1	2	0	0	0	0	0	0	1	2
	COG2358	TRAP-type uncharacterized transport system, periplasmic component	2	0	0	0	1	3	1	2	0	0	0	0	0	0	1	2
Cell wall/membrane/ envelope biogenesis	COG5386	Cell surface protein	1	0	6	5	2	2	0	0	1	0	5	5	6	0	0	0
	COG1732	ABC-type Periplasmic glycine betaine/choline-binding (lipo)protein	2	2	1	1	2	0	1	1	2	2	1	1	1	1	1	1
	COG1292	Choline-glycine betaine transporter	8	2	3	3	4	3	1	5	1	1	3	3	2	0	1	8
	COG1970	large conductance mechanosensitive channel protein	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1
	COG0845	Membrane-fusion protein	2	1	2	2	1	2	1	2	1	2	2	2	2	0	1	3
	COG3559	Putative exporter of polyketide antibiotics	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1
	COG0668	Small-conductance mechanosensitive channel	3	3	2	2	2	2	1	2	3	3	2	2	2	1	1	1
Defense mechanisms	COG1136	ABC-type antimicrobial peptide transport system, ATPase component	2	6	12	14	8	11	7	3	6	7	15	19	19	1	3	4
	COG1132	ABC-type multidrug transport system, ATPase and permease components	9	7	8	10	12	11	11	9	8	9	11	10	15	5	6	11
	COG1131	ABC-type multidrug transport system, ATPase component	12	16	22	23	21	18	19	14	11	14	24	23	24	12	9	17
Lipid metabolism	COG2031	Short chain fatty acids transporter	1	0	1	1	1	1	0	1	0	0	1	1	1	0	0	1

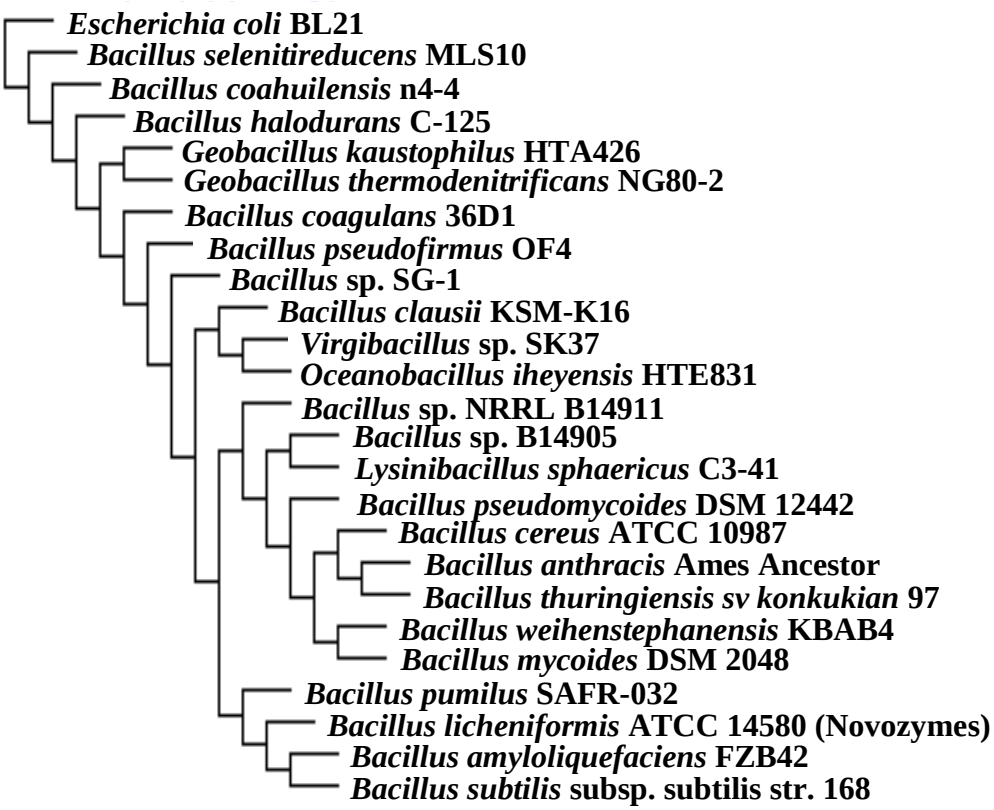
Supporting material Table 2 (Continued)

Function	COG	Name	Vir sp. SK37	Bac amy FZ2	Bac ant Anr	Bac cer 107	Bac cla KS6	Bac hal C-5	Bac lic (N)	Bac pse OF4	Bac pum SA2	Bac sub 168	Bac thu Ham	Bac thu 977	Bac wei KB4	Geo kau HT6	Geo the NG2	Oce ihe HT1
Intracellular trafficking	COG4473	Predicted ABC-type exoprotein transport system, permease component	1	2	3	3	1	1	1	1	1	2	3	3	3	1	1	2
General function prediction only	COG4988	ABC-type cytochrome bd biosynthesis, ATPase and permease components	2	1	1	1	0	1	1	1	1	1	1	1	2	0	0	1
	COG4987	ABC-type cytochrome bd biosynthesis, fused ATPase and permease components	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	1
	COG0755	ABC-type cytochrome c biogenesis, permease component	2	2	2	2	2	2	2	3	2	2	2	2	2	2	2	2
	COG1277	ABC-type multi-copper enzyme maturation, permease component	1	1	1	1	0	1	2	0	1	1	1	1	1	1	1	0
	COG1101	ABC-type uncharacterized transport system, ATPase component	1	0	0	0	1	0	0	0	0	0	0	0	0	2	2	0
	COG3127	ABC-type lysophospholipase L1 biosynthesis, permease component	2	0	1	3	1	0	0	0	1	0	2	3	2	0	0	0
	COG1123	ABC-type transport systems, contain duplicated ATPase	3	2	3	3	5	0	6	1	4	2	3	3	5	1	2	2
	COG2984	ABC-type uncharacterized transport system, periplasmic component	1	0	0	0	1	0	0	0	0	0	0	0	0	2	2	0
	COG1272	channel protein, hemolysin III family	1	1	2	2	1	1	1	1	1	1	2	2	2	0	0	2
	COG0073	EMAP domain	2	3	3	3	2	3	3	1	3	2	3	3	3	2	2	3
	COG2244	Membrane protein involved in the export of O-antigen and teichoic acid	5	6	8	7	6	5	6	5	6	6	8	8	9	6	4	4
	COG0733	Na^+-dependent transporters of the SNF family	3	2	4	4	3	3	3	5	0	2	4	4	4	0	0	2
	COG2270	Permeases of the major facilitator superfamily	1	1	3	2	1	1	1	0	2	1	2	2	4	0	0	1
	COG2409	Predicted drug exporters of the RND superfamily	2	2	2	3	0	0	2	1	2	2	3	3	3	1	2	0
	COG1033	Predicted exporters of the RND superfamily	1	0	0	1	1	1	0	1	0	0	0	0	0	0	0	1
	COG1823	Predicted Na^+/dicarboxylate symporter	1	1	1	1	1	0	1	0	0	1	1	1	1	0	1	1
	COG0385	Predicted Na^+-dependent transporter	1	2	1	1	2	2	2	1	2	2	1	1	0	1	2	1
	COG2056	Predicted permease	1	1	1	1	1	2	1	2	1	1	1	1	1	0	0	1
	COG5006	Predicted permease, DMT superfamily	2	3	4	5	3	2	4	4	1	3	5	6	3	0	1	2
	COG2962	Predicted permeases	3	2	4	4	1	1	1	2	1	1	4	4	4	2	1	1
	COG4147	Predicted symporter	1	0	1	1	1	1	1	2	1	1	1	1	1	1	1	3
	COG1811	Uncharacterized membrane protein, possible Na^+ channel or pump	1	0	1	1	0	1	0	1	0	1	1	1	1	1	1	1

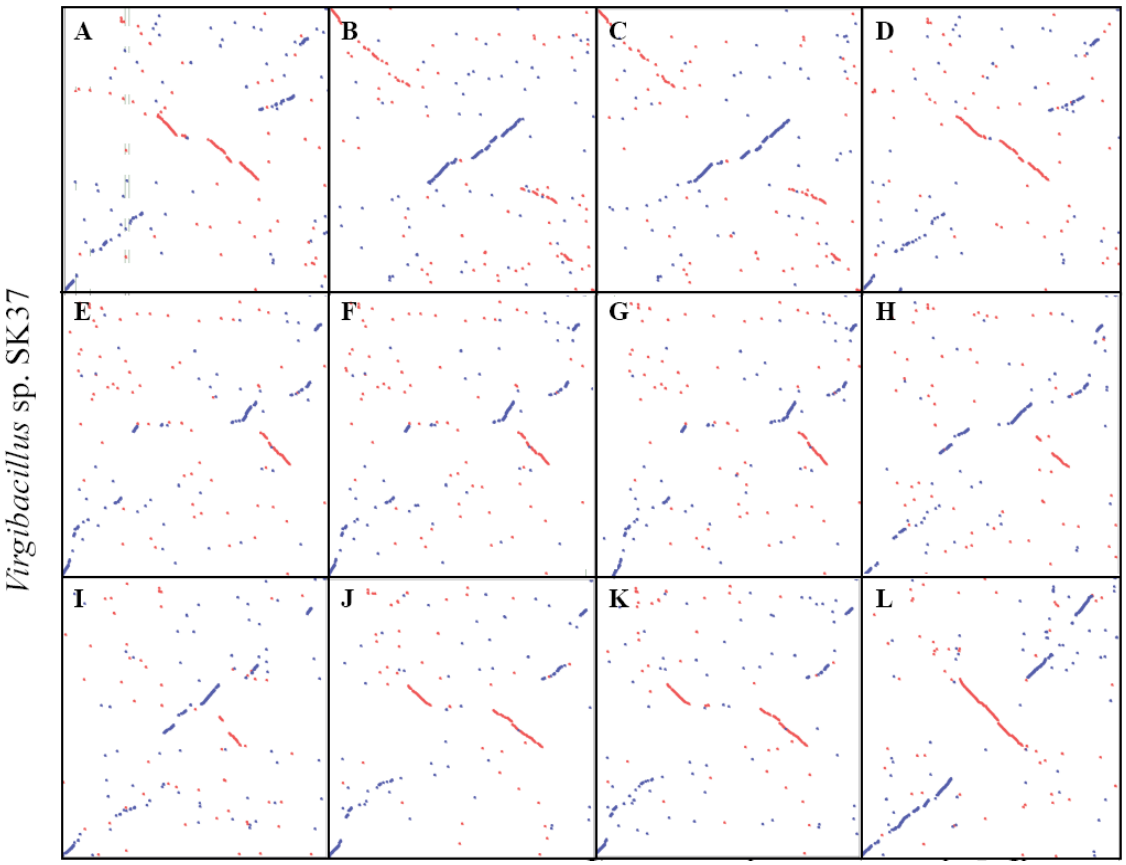
Supporting material Table 3 Orthologous groups of proteases that may involve in protein and peptide metabolism in the SK37 cells

Function ID	Name	Topology (Number of CDSs)					total
		cytoplasm	cell membrane	periplasmid	cell wall	extracelllular	
COG0006	Xaa-Pro aminopeptidase	3	-	-	-	-	3
COG0024	Methionine aminopeptidase	1	-	-	-	-	1
COG0260	Leucyl aminopeptidase	1	-	-	-	-	1
COG1164	Oligoendopeptidase F	3	-	-	-	-	3
COG1506	Dipeptidyl aminopeptidases/acylaminoacyl-peptidases	3	-	-	-	1	4
COG2195	Di- and tripeptidases	2	-	-	-	-	2
COG2309	Leucyl aminopeptidase (aminopeptidase T)	2	-	-	-	-	2
COG2317	Zn-dependent carboxypeptidase	1	-	-	-	-	1
COG2355	Zn-dependent dipeptidase, microsomal dipeptidase homolog	1	-	-	-	-	1
COG2866	Predicted carboxypeptidase	1	-	-	-	1	2
COG3191	L-aminopeptidase/D-esterase	1	-	-	-	-	1
COG3227	Zinc metalloprotease (elastase)	-	-	-	-	2	2
COG4942	Membrane-bound metallopeptidase	-	2	-	-	-	2
COG0739	Membrane proteins related to metalloendopeptidases	-	3	-	-	-	3
COG0744	Membrane carboxypeptidase (penicillin-binding protein)	-	5	-	-	-	5
COG0750	Predicted membrane-associated Zn-dependent proteases 1	-	2	-	-	-	2
COG0793	Periplasmic protease	-	-	1	-	-	1
COG0612	Predicted Zn-dependent peptidases	1	-	-	-	2	3
COG0693	Putative intracellular protease/amidase	2	-	-	-	-	2
COG1266	Predicted metal-dependent membrane protease	-	1	-	-	-	1
COG1473	Metal-dependent amidase/aminoacylase/carboxypeptidase	8	-	-	-	-	8
COG1994	Zn-dependent proteases	-	1	-	-	-	1
COG2738	Predicted Zn-dependent protease	-	1	-	-	-	1
COG0265	Trypsin-like serine proteases	-	-	2	-	2	4
COG0465	ATP-dependent Zn proteases	-	-	1	-	-	1
COG1404	Subtilisin-like serine proteases	1	-	-	1*	2	4
COG5504	Predicted Zn-dependent protease	1	-	-	-	1	2
COG4242	Cyanophycinase and related exopeptidases	2	-	-	-	-	2
Total		34	15	4	1	11	65

Note: Topology analysis was suggested by TMHMM (Moller et al, 2001) ans SignalP (Emaulesson et al, 2007)
* Protein that contains both transmembrane domain and signal peptide.



Supporting material Figure 1 Genome clustering analysis by COG functional profile Hierarchical clustering. *E. coli* BL21 genomic sequence was used as an outgroup.



Supporting material Figure 2 Dotplot analysis showing comparison of ortholog organization between SK37 and (A) *B. subtilis* subsp. subtilis str. 168, (B) *B. licheniformis* ATCC 14580 (Novozymes), (C) *B. amyloliquefaciens* FZB42, (D) *B. pumilus* SAFR-032, (E) *B. anthracis* Ames Ancestor, (F) *B. cereus* ATCC 10987, (G) *B. weihenstephanensis* KBAB4, (H) *B. halodurans* C-125, (I) *B. clausii* KSM-K16, (J) *G. kaustophilus* HTA426, (K) *G. thermodenitrificans* NG80-2, and (L) *O. iheyensis* HTE831. The *y*- and *x*-axes represent SK37 and compared genome, respectively, with all the same arbitrary scale of 360°. Blue and red dots represent the location of ortholog on positive and negative strand, respectively.

BIOGRAPHY

Ekkarat Phrommao was born in February 25th, 1982 in Prachinburi, Thailand. He studied for his high school diploma at Mattayom Watmaikrongthong School (1998-2000). In 2004, he received the degree of Bachelor of Science in Food Technology with first class honor from Suranaree University of Technology, Nakhon Ratchasima. He also graduated Summa cum laude and received the Golden Badge awarded to a student with the highest score. In 2005-2010, he received the Royal Golden Jubilee Scholarship from Thailand Research Fund to study for the degree of Doctor of Philosophy (Food Technology) at Suranaree University of Technology. During his graduate study, he obtained opportunities to present his research works including at RGJ-Ph.D. Congress IX (Pattaya, Chonburi, April 4-6th 2008) with awarded "Outstanding poster presentation" and IFT annual meeting and Food expo (Anaheim, CA, June 6-9th 2009; Chicago, IL, July 17-20th 2010) with awarded "2nd place in the finalist of poster presentation in Biotechnology Section". He also published his research work under the title of "Identification of novel halotolerant bacillopeptidase F-like proteinases from a moderately halophilic bacterium, *Virgibacillus* sp. SK37" in Journal of Applied Microbiology (volume 110, page 191-210) in 2010.