ELSEVIER

Journal of Chromatography B, 771 (2002) 261–287

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

Review

Bioinformatics, functional genomics, and proteomics study of *Bacillus* sp.

Supachok Sinchaikul[a], Boonyaras Sookkheo[b], Supachai Topanuruk[a,c], Hsueh-Fen Juan[a], Suree Phutrakul[c], Shui-Tein Chen[a,*]

[a]*Institute of Biological Chemistry, Academia Sinica, 128 Yen Chiu Yuan Rd., Sec II, Nankang, Taipei, Taiwan 11529*
[b]*Chemistry Unit, School of Science, Rangsit University, Patumthani, Thailand 12000*
[c]*Department of Chemistry, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand 50200*

Abstract

The ability of bioinformatics to characterize genomic and proteomic sequences from bacteria *Bacillus* sp. for prediction of genes and proteins has been evaluated. Genomics coupling with proteomics, which is relied on integration of the significant advances recently achieved in two-dimensional (2-D) electrophoretic separation of proteins and mass spectrometry (MS), are now important and high throughput techniques for qualifying and analyzing gene and protein expression, discovering new gene or protein products, and understanding of gene and protein functions including post-genomic study. In addition, the bioinformatics of *Bacillus* sp. is embraced into many databases that will facilitate to rapidly search the information of *Bacillus* sp. in both genomics and proteomics. It is also possible to highlight sites for post-translational modifications based on the specific protein sequence motifs that play important roles in the structure, activity and compartmentalization of proteins. Moreover, the secreted proteins from *Bacillus* sp. are interesting and widely used in many applications especially biomedical applications that are the highly advantages for their potential therapeutic values. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Reviews; Bioformatics; Proteomics; Functional genomics; *Bacillus*

Contents

1. Introduction 262
2. Genome and differentially expressed genes of *Bacillus* sp. under different conditions 262
   2.1. Functional classification of gene products 263
   2.2. Differentially expressed genes under different conditions 263
3. Proteomics analysis of *Bacillus* sp. proteins 265
   3.1. Secrete to extracellular 268
   3.2. On the cell membrane 271
   3.3. In the cytosol 271

*Corresponding author. Tel.: +886-2-278-55981, ext. 7071; fax: +886-2-278-83473.
E-mail address: bcchen@gate.sinica.edu.tw (S.-T. Chen).

1570-0232/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S1570-0232(02)00054-5

4. Post-translational modification of gene product in *Bacillus* sp. 273
5. Bioinformatics to link differential display genes and proteomic identified proteins 277
6. Biomedical applications of *Bacillus* sp. 278
7. Conclusions 285
Acknowledgements 285
References 286

1. Introduction

Over the past few years, up to 56 microorganisms have had their genomes completely sequenced, with another 170 or in progress (http://www.tigr.org). The genome of *Bacillus* sp. has been found only two strains that were *B. subtilis* and *B. halodurans* [1,2]. It does have the other strains that have not been yet completely determined their genomes such as *B. stearothermophilus* strain 10, *B. cereus* and *B. licheniformis*. In the generally characteristics, the members of the bacterial genus *Bacillus* sp. more than 368 strains are Gram-positive, endospore forming, rod-shaped bacteria that grow aerobically [3]. They exhibit a wide range of physiological abilities that allow them to live in every natural environment. In addition, they are important organisms for researches and industrial applications such as they are used in many medical, pharmaceutical, agricultural, and industrial processes that take advantage of their wide range of physiological characteristics and their ability to produce a host of enzymes, antibodies and other metabolites. Due to more availability and applications of *Bacillus* sp., the study of genome is increasingly important and need to know the biological system of these bacteria, which intends to understand the gene function, metabolic pathway and proteomics including protein structure. For this review, it is discovered the bioinformatics, functional genomics and proteomics study of *Bacillus* sp. It intends to be a literature search for *Bacillus* strains that provides a guide to choosing the most efficient way to analyze a new sequence or to collect information on a gene or protein of interest by applying current publicly available databases and Web servers, and also help to analyze genomic and proteomic data clearly. In addition, the known functional genomics will underlie both a better understanding of cellular function and a more efficient application of biotechnology. Moreover, the thermophilic bacteria *B. stearothermophilus* P1 and TLS33 isolating from a hot spring in Chiang Mai, Thailand, are another strains that used to study the proteomic analysis in both extracellular and cytosol, and present as the first report. It is very interesting to their identified proteins and functions that will be useful for the other strains of *Bacillus* sp. to get more information.

2. Genome and differentially expressed genes of *Bacillus* sp. under different conditions

The outcome of genomics is the discipline of functional genomics whereby one uses DNA-based technologies to make inferences about organism structure and behavior. The complete genomic nucleotide sequence of the spore-forming bacterium *Bacillus* sp. has a few strains and the well-known strain is *B. subtilis*. The competition of the *B. subtilis* genomic nucleotide sequence has revealed 345 small polypeptide open-reading frames (of 85 codons or less) and comprises a total of approximately 4100 protein-coding genes with 4.2 Mbp in size and 42% unknown function [1]. Of these protein-coding genes, 53% are represented once, while a quarter of the genome corresponds to several gene families that have been greatly expanded by gene duplication, the largest family containing 77 putative ATP-binding transport proteins. Otherwise, the complete genome sequence of the alkaliphilic bacterium *B. halodurans* C-125 has been known since 2000 year [2]. Its 4 202 353-bp genome contains 4066 predicted protein-coding sequences (CDSs), 2141 (52.7%) of which have functional assignments, 1182 (29%) of which are conserved CDSs with unknown function and 743 (18.3%) of which have no match to any protein databases. Among the total CDSs, 8.8% match sequences of proteins found only in *B. subtilis* and 66.7% are widely conserved in comparison with the proteins of various organisms, including *B. subtilis*.

Although the bacterial genus *Bacillus* sp. has many strains but the genetic and physical maps may be different; for example, the assigned positions of the genes on the physical map of *B. halodurans* C-125 were compared with those on the genetic and physical maps of *B. subtilis* and showed that the positions of some putative genes of *B. halodurans* C-125 differed completely from *B. subtilis* orthologues, suggesting that genome organization is not conserved between these two strains, although the size and GC content of the C-125 chromosome are almost identical to those of *B. subtilis* (4.2 Mb and 43.7 mol%, respectively) [4]. In addition, the comparison of completely genomes between *B. subtilis and B. halodurans* also shows in Table 1.

2.1. Functional classification of gene products

Genes were classified, which based on the representation of cells as Turing machines in which one distinguishes between the machine and the program using BLAST2P software running against a composite protein databank compound of SWISS PROT, TREMBL and *B. subtilis* proteins [1]. It is well known that 58% of the *B. subtilis* proteins having known a function but up to 42% of the gene products have never been predicted by similarity to proteins of known function: 16% of the proteins are similar to unknown proteins from *B. subtilis* and some other organism; and 26% of the proteins are not significantly similar to any other proteins in databanks. The example of the functional classification devised for *B. subtilis* protein coding genes is shown in Table 2, which based on the one hand on the distinction between the machinery of the chemical reactions taking place in the cell [5].

2.2. Differentially expressed genes under different conditions

Most of *Bacillus* sp. initiates a series of transitional responses that are designed to maintain or restore growth. These include the induction of macromolecular hydrolases such as proteases, lipases and polysaccharidases, chemotaxis and mobility, and competence (i.e., the ability to take up DNA from the environment). If these response fail to reestablish growth, sporulation is induced. The ability of *Bacillus* sp. to form highly resistant endospores imparts an enormous competitive advantage in environments

Table 1
Comparison of the general features of the *B. subtilis* and *B. halodurans* genomes

Genome features	*B. subtilis*	*B. halodurans*
Size (bp)	4 214 814	4 202 353
G+C content (mol%)		
Total genome	43.5	43.7
Coding region	44.3	44.4
Non-coding region	39.3	39.8
Open reading frames		
Percent of genome (coding)	87	85
Protein coding genes	4104	4066
Conserved with function assigned	2379	2144
Conserved with unknown function	668	1182
Non-conserved	1053	743
Percent AUG initiation codons	78	78
Percent GUG initiation codons	9	12
Percent UUG initiation codons	13	10
Insertion elements		
PBSX prophage-related protein	27	2
Transposase and related protein	10	112
RNA elements		
Stable RNA (percent of genome)	1.27	1.02
16S, 23S and 5S rRNA	10	8
tRNA	86	78

Table 2
Functional classification of *B. subtilis* protein coding genes

1. Cell envelope and cellular processes	867 (21%)
1.1 Cell wall	93
1.2 Transport/binding proteins and lipoproteins	381
1.3 Sensors (signal transduction)	38
1.4 Membrane bioenergetics (electron transport chain and ATP synthase)	78
1.5 Mobility and chemotaxis	55
1.6 Protein secretion	18
1.7 Cell division	21
1.8 Sporulation	139
1.9 Germination	23
1.10 Transformation/competence	21
2. Intermediary metabolism	742 (18%)
2.1 Metabolism of carbohydrates and related molecules	261
2.1.1 Specific pathways	214
2.1.2 Main glycolytic pathways	28
2.1.3 TCA cycle	19
2.2 Metabolism of amino acids and related molecules	205
2.3 Metabolism of nucleotides and nucleic acids	83
2.4 Metabolism of lipid	77
2.5 Metabolism of coenzymes and prosthetic groups	99
2.6 Metabolism of phosphate	9
2.7 Metabolism of sulfur	8
3. Information pathways	482 (12%)
3.1 DNA replication	22
3.2 DNA restriction/modification and repair	39
3.3 DNA recombination	17
3.4 DNA packaging and segregation	10
3.5 RNA synthesis	244
3.5.1 Initiation	19
3.5.2 Regulation	213
3.5.3 Elongation	8
3.5.4 Termination	4
3.6 RNA modification	19
3.7 Protein synthesis	96
3.7.1 Ribosomal proteins	56
3.7.2 Aminoacyl-tRNA synthetases	25
3.7.3 Initiation	6
3.7.4 Elongation	6
3.7.5 Termination	3
3.8 Protein modification	27
3.9 Protein folding	8
4. Other functions	289 (7%)
4.1 Adaptation to atypical conditions	72
4.2 Detoxification	68
4.3 Antibiotic production	30
4.4 Phage-related functions	83
4.5 Transposon and IS	10
4.6 Miscellaneous	26
5. Similar to unknown proteins	667 (16%)
5.1 From *B. subtilis*	177
5.2 From other organisms	490
6. No similarity	1053 (26%)

such as soil, where long periods of drought and nutrient deprivation are common. At least 4% of the *B. subtilis* genome is dedicated to the processes of sporulation, germination and outgrowth.

The *Bacillus* sp. genome encodes more than 40 temperature-shock and general stress proteins, which presumably contribute to this organism's ability to survive shorter periods of adversity. It encodes numerous ''heat shock inducible'' genes, including chaperones and proteases [6]. Proteins induced in response to lower temperatures include cold acclimatization proteins, for those synthesized continuously at low temperature, and cold shock proteins (Csp), for those synthesized in response to a sudden temperature downshift [7]. The adaptation of bacteria to highly alkaline environments was interesting and also studied in *Bacillus* sp. because many alkaliphilic *Bacillus* strains produce unique alkaline enzymes such as xylanases, proteases and amylases [8]. One of the most commonly encountered is likely to be osmotic stress resulting from frequent wetting and drying of its habitat [9]. For example, the specific and general stress proteins in *B. subtilis* in which response to heat shock, salt and ethanol stress, and glucose and phosphate starvation was analyzed using 2-D PAGE analysis [10]. The induction of at least 42 general stress proteins absolutely required the alternative sigma factor σ^B except in oxidative stress condition. However, at least seven stress proteins, among them ClpC, ClpP, Sod, AhpC and AhpF, remained stress inducible in a *sigB* mutant. Such a detailed analysis also permitted the description of subgroups of general stress proteins, which are subject to additional regulatory circuits, indicating a very thorough fine-tuning of this complex response. The relative synthesis rate of the general stress proteins constituted up to 40% of the total protein synthesis of stressed cell and thereby emphasizes the importance of the stress regulation. Moreover, the oxidative stress stimulation in *B. subtilis* was also found more than 50 expressed proteins by proteomic approach using 2-D PAGE [11]. Not only the cytoplasmic stress response systems direct the upregulation genes encoding heat shock proteins, molecular chaperones and proteases but the extracytoplasmic secretion stress system also regulates the genes for the survival of severe secretion stress. In recently, a novel two-component regulatory system in *B. subtilis*, named CssR–CssS, was required for the cell to survive to severe secretion stress caused by a combination of high-level production of the α-amylase AmyQ and reduced levels of the extracytoplasmic folding factor PrsA [12].

In addition, the differentially expressed genes of *Bacillus* sp. can be found in substrate utilization such as carbohydrate, ammonium, opines and inorganic phosphate. For example, the ability of *Bacillus* sp. to secrete a wide variety of extracellular macromolecular depolymerases is a major factor contributing to their colonization of soil [13]. Genes encoding secreted amylases, arabinases, chitonases, mannanases, cellulases and xylanasess are evident in the genome sequence. Proteases and lipases are also frequently encountered, both intracellular and extracellular, the latter allowing proteins to provide sources of both carbon and nitrogen.

3. Proteomics analysis of *Bacillus* sp. proteins

Proteomics analysis is concerned with the global changes in protein expression as most commonly accomplished by a combination of two-dimensional gel electrophoresis (2-DE) to separate and visualize proteins and mass spectrometry (MS) for protein identification. Emerging technologies for proteome analysis also include microarray and microfluidic devices. There are called proteomics, which is the large-scale analysis of proteins, is contributed greatly to understanding gene function in post-genomic era even through the protein chemistry. Important technical advances related on 2-DE and protein MS have increased sensitivity, reproducibility and throughput of proteome analysis while creating an integrated technology.

In addition, artificial neural network were trained on the prediction of the subcellular location of bacterial proteins. A cross-validated average prediction accuracy of 93% was reached for distinction between cytoplasmic and non-cytoplasmic proteins, based on the analysis of protein amino acid composition [14]. The greatest differences in residue frequencies between cytoplasmic and non-cytoplasmic sequences are found for Glu, Leu, Ile and Arg (predominance in cytoplasmic sequences), and for Ser, Asn, Gly, Thr and Ala (predominance in

Table 3
Extracellular proteins of *B. subtilis* 168; the proteins were identified by mass spectrometry

Proteins	Function/similarity
Metabolism of carbohydrates	
AbnA	Arabinan-endo 1,5-α-L-arabinase
AmyE^{M}	α-Amylase
BglC	Endo-1,4-β-glucanase, cellulase
BglS	Endo-β-1,3-1,4 glucanase
CitH	Malate dehydrogenase
Csn^{H}	Chitosanase
Eno	Enolase
PdhB	Pyruvate dehydrogenase (E1 β subunit)
PdhD	Pyruvate dehydrogenase (E3 subunit)
Pel^{H}	Pectate lyase
PelB	Pectate lyase
XynA	Endo-1,4-β-xylanase
XynD$^{lipo\ 2\ TM\ H}$	Endo-1,4-β-xylanase
YnfF	Endo-xylanase
YvgN	Similar to plant-metabolite dehydrogenase
YvpA	Pectate lyase
YwjH	Similar to transaldolase (pentose phosphate)
YxlA	Arabinan-endo 1,5-α-L-arabinase
Metabolism of amino acids	
RocA	Pyrroline-5 carboxylate dehydrogenase
RocF	Arginase
Metabolism of proteins	
AprE	Serine alkaline protease (subtilin E)
Bpr	Bacillopeptidase F
Epr	Minor extracellular serine protease
Ggt^{TM}	γ-Glutamyltranspeptidase
Mpr	Extracellular metalloprotease
NprE	Extracellular neutral metalloprotease
Vpr^{H}	Extracellular serine protease
WprA$^{RR\ W\ 3H}$	Cell wall-associated protein precursor
YwaD	Aminopeptidase
Metabolism of nucleotides and nucleic acids	
YfkN$^{RR\ TM\ pst}$	2′,3′-Cyclic-nucleotide 2′-phosphodiesterase
YhcR$^{RR\ TM}$	5′-Nucleotidase
Yurl	Ribonuclease
Metabolism of lipids	
LlpA^{RR}	Lipase
Metabolism of phosphate	
GlpQ^{pst}	Glycerophosphoryl diester phosphodiesterase
PhoA^{pst}	Alkaline phosphatase A
PhoB^{pst}	Alkaline phosphatase III
PhoD$^{RR\ pst}$	Phosphodiesterase/alkaline phosphatase D
Metabolism of the cell wall	
CwlC$^{W\ ex}$	*N*-Acetylmuramoyl-L-alanine amidase
LytD^{W}	*N*-Acetylglucosaminidase (major autolysin)
PbpA$^{TM\ 3}$	Penicillin-binding protein 2A
PbpC$^{lipo\ lgt}$	Penicillin-binding protein 3
PbpX^{RR}	Penicillin-binding protein
WapA$^{RRW\#\ 4\ H}$	Cell wall-associated protein precursor
Todj$^{lipo\ lgt}$	D-Alanyl-D-alanine carboxypeptidase
YvcE^{W}	Cell wall-binding protein
YwtD$^{W\ H}$	Similar to murein hydrolase
Mobility and chemotaxis	
Hag$^{dual\ H}$	Flagellin protein
FlgK^{ex}	Flagellar hook-associated protein 1 (HAP1)
FliD^{ex}	Flagellar hook-associated protein 2 (HAP2)

Table 3. Continued

Proteins	Function/similarity
Protein synthesis (Elongation)	
Fus, Ef-G	Elongation factor G
Protein folding	
GroEL	Class I heat-shock protein (chaperonin)
Transport/binding proteins and lipoproteins	
FeuA$^{lipo\ lgt}$	Iron-uptake system (binding protein)
FhuD$^{lipo\ lgt}$	Ferrichrome ABC transporter (binding protein)
MntA^{lipo}	Manganese ABC transporter (binding protein)
MsmE$^{lipo\ lgt}$	Multiple sugar ABC transporter (binding protein)
OppA^{lipo}	Oligopeptide ABC transporter (binding protein)
OpuAC$^{lipo\ pst\ lgt}$	Glycine betaine ABC transporter(binding protein)
PstS$^{lipo\ pst}$	Phosphate ABC transporter (binding protein)
RbsB$^{lipo\ lgt}$	Ribose ABC transporter (binding protein)
YcdH$^{lipo\ pst}$	Zinc ABC transporter (binding protein)
YclQ$^{lipo\ H}$	Ferrichrome ABC transporter (binding protein)
YdhF$^{RR\ lipo\ pst}$	Similar to unknown proteins from *B. subtilis*
YflY$^{lipo\ lgt}$	Iron (III) didtrate ABC transporter (binding protein)
YfmC^{lipo}	Ferrichrome ABC transporter (binding protein)
YqlX$^{lipo\ pst}$	Amino acid ABC transporter (binding protein)
YrpE$^{lipo\ pst\ lgt}$	Similar to unknown proteins from *B. subtilis*
YusA$^{lipo\ lgt}$	Probably part of the S box regulon
YxeB$^{lipo\ lgt}$	ABC transporter (binding protein)
Sporulation	
TasA^{H}	Translocation-dependent antimicrobial spore component
RNA synthesis/regulation	
Ywt^{F}	Transcriptional regulator
Detoxification	
KatA^{H}	Vegetative catalase 1
PenP^{H}	β-Lactamase precursor
SodA^{H}	Superoxide dismutase
Ybxl	Similar to β-lactamase
YceD	Similar to tellurium resistance protein
Phage-related functions	
XepA^{ex}	PBSX prophage lytic exoenzyme
XkdG$^{ex\ H}$	PBSX prophage gene
XkdK^{ex}	PBSX prophage gene
KkdM^{ex}	PBSX prophage gene
XlyA$^{W\ ex}$	*N*-Acetylmuramoyl-L-alanine amidase
Similar to unknown proteins	
YbdN	Unknown
Yfnl$^{TM\#H}$	Probable transmembrane glycoprotein
YlqBH	Unknown
YncM$^{\#H}$	Similar to unknown proteins from *B. subtilis*
YolA	Unknown
YoaW	Unknown
YrpD$^{lipo\ 2\ \#}$	Similar to unknown proteins from *B. subtilis*
YuaB	Unknown
YvgO	Unknown
YweA$^{3\ \#}$	Similar to unknown proteins from *B. subtilis*
YwoF	Unknown
YxaK^{H}	Similar to unknown proteins from *B. subtilis*
YxkC$^{3\ H}$	Unknown

periplasmic and extracellular sequences). On average, these differences are more significant for extracellular proteins than for periplasmic proteins. Periplasmic proteins seem to contain a higher amount of Pro, Lys, Cys and Met compared to extracellular sequences.

3.1. Secrete to extracellular

It is well known that most species of *Bacillus* strains have a high capacity to secrete proteins into the culture medium such as *B. subtilis*, *B. licheniformis*, *B. amyloliquefaciens*, *B. firmus*, *B. stearothermophilus*, *B. catenulatus* and *B. thermoleovolans*. However, the properties of their proteins vary from strain to strain. In *B. subtilis* genome, several genes encoding proteins of the major secretion pathway have been identified: *secA*, *secD*, *secE*, *secF*, *secY*, *ffh* and *ftsY*. In addition, the proteomic approaches were used to identify the extracellular complement of the *B. subtilis* secretome and showed that using different growth conditions and a hypersecreting mutant, ~200 extracellular proteins were visualized by 2-D PAGE of which 82 identified proteins were identified by MALDI-TOF mass spectrometry [15]. The 82 identified extracellular proteins include 50 proteins to which a function had been assigned previously (Table 3). In addition, the secreted proteins from xylan-induced *B. firmus* K-1 were analyzed using 2-D PAGE and found that five stimulatory secreted proteins containing xylose isomerase, xylanases, GltC regulatory protein and 3-dehydroquinate dehydratase were expressed under substrate induction using xylan [16]. Moreover, the study of extracellular proteins from *B. stearothermophilus* using 2-D PAGE that have never been reported also showed in this review. In our previously reports, the bacterium *B. stearothermophilus* strain P1 and TLS33 have been known that they could secrete the extracellular lipase and proteases, respectively [17,18]. It is therefore studied the extracellular proteins using proteomic approach and the preliminary result of 2-D gel comparing between extracellular and intracellular proteins showed the different protein patterns (Figs. 1 and 2). Otherwise, it also found the extracellular superoxide dismutase (SOD) with the molecular mass approximately 28 kDa and p*I* 5.8 [19]. Furthermore, the substrate-induced lipase from *B. stearothermophilus* P1 using olive oil induction not only was studied the proteomic analysis using 2-D PAGE but it was also analyzed using Agilent 2100 Bioanalyzer and Protein 200 LabChip kit (AgilentTechnologies, USA), which is the new technique for protein sizing and chip-based separation. The analyzed proteins under substrate induction were shown the high-resolution separation on

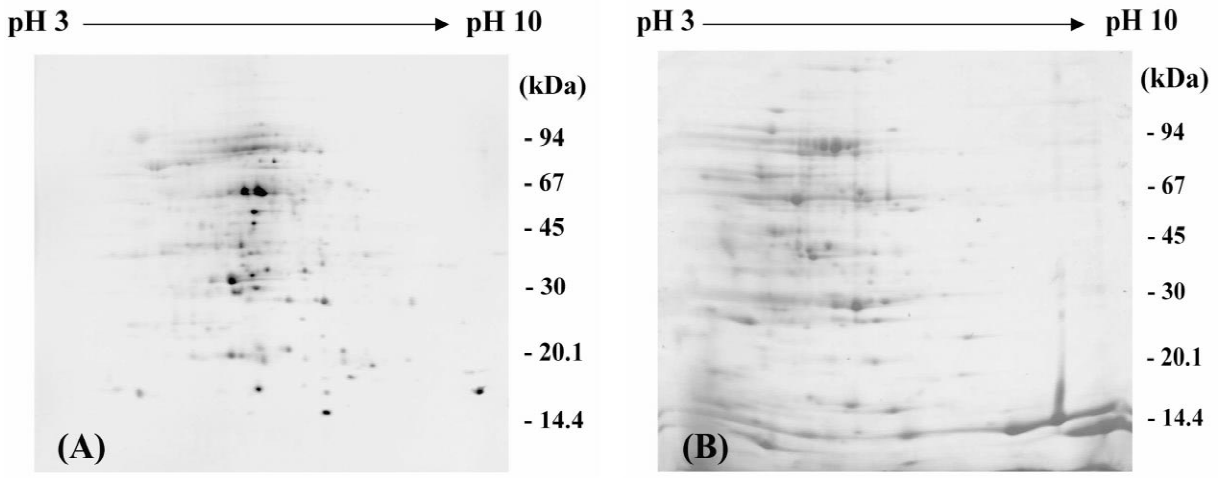

Fig. 1. Comparison of the cytosolic proteins (A) and extracellular proteins (B) from a thermophilic bacterium *B. stearothermophilus* TLS33 by 2-D PAGE analysis using a wide pH range of IPG strip 3–10, 18 cm.

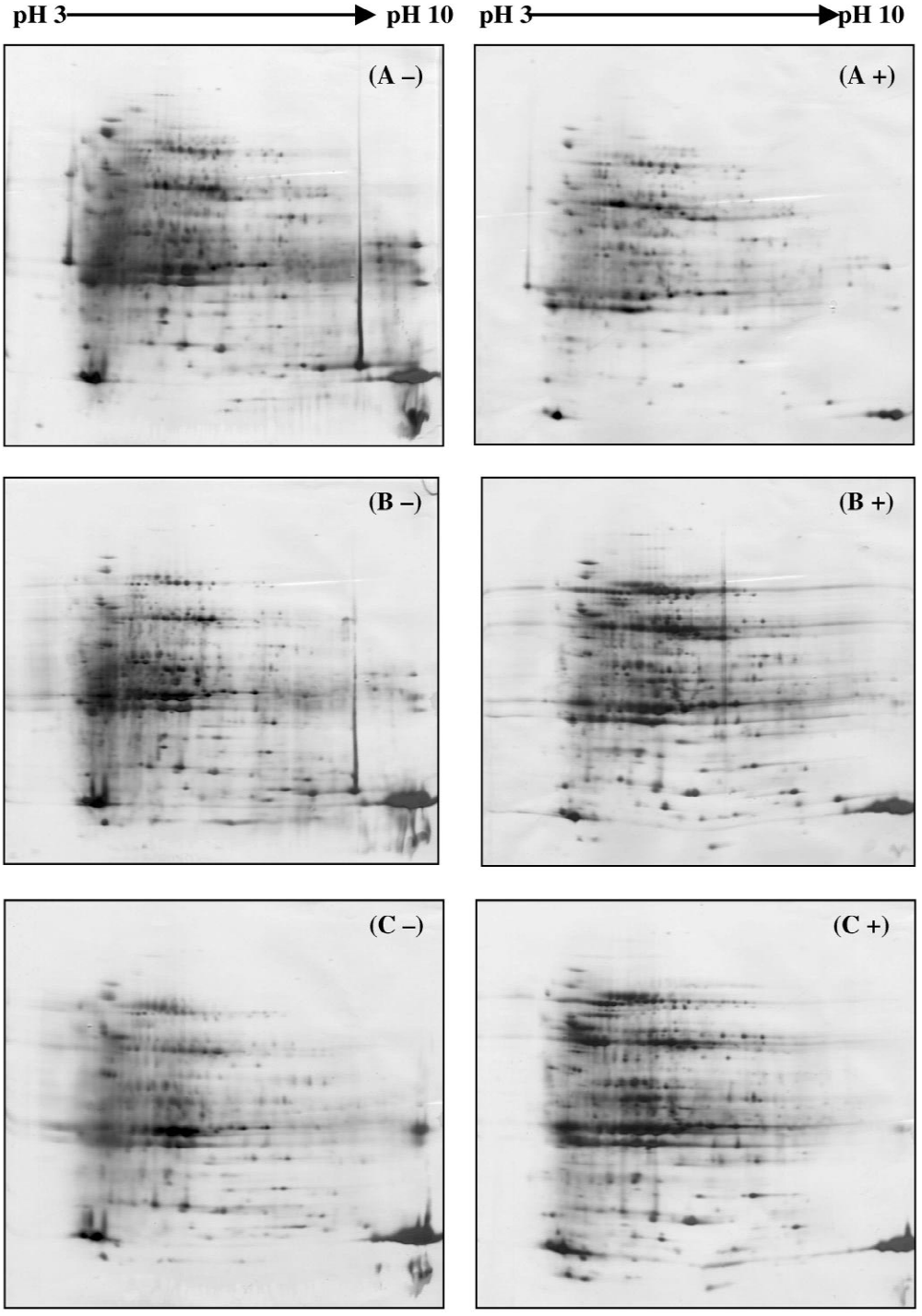

Fig. 2. 2-D PAGE gels of the extracellular proteins after culturing the thermophilic bacterium *B. stearothermophilus* P1 in the presence (+) and absence (—) of 0.5% (v/v) olive oil at 65 °C for 12 (A), 24 (B), 36 (C), 48 (D), 72 (E) and 96 h (F). Dying with silver staining.

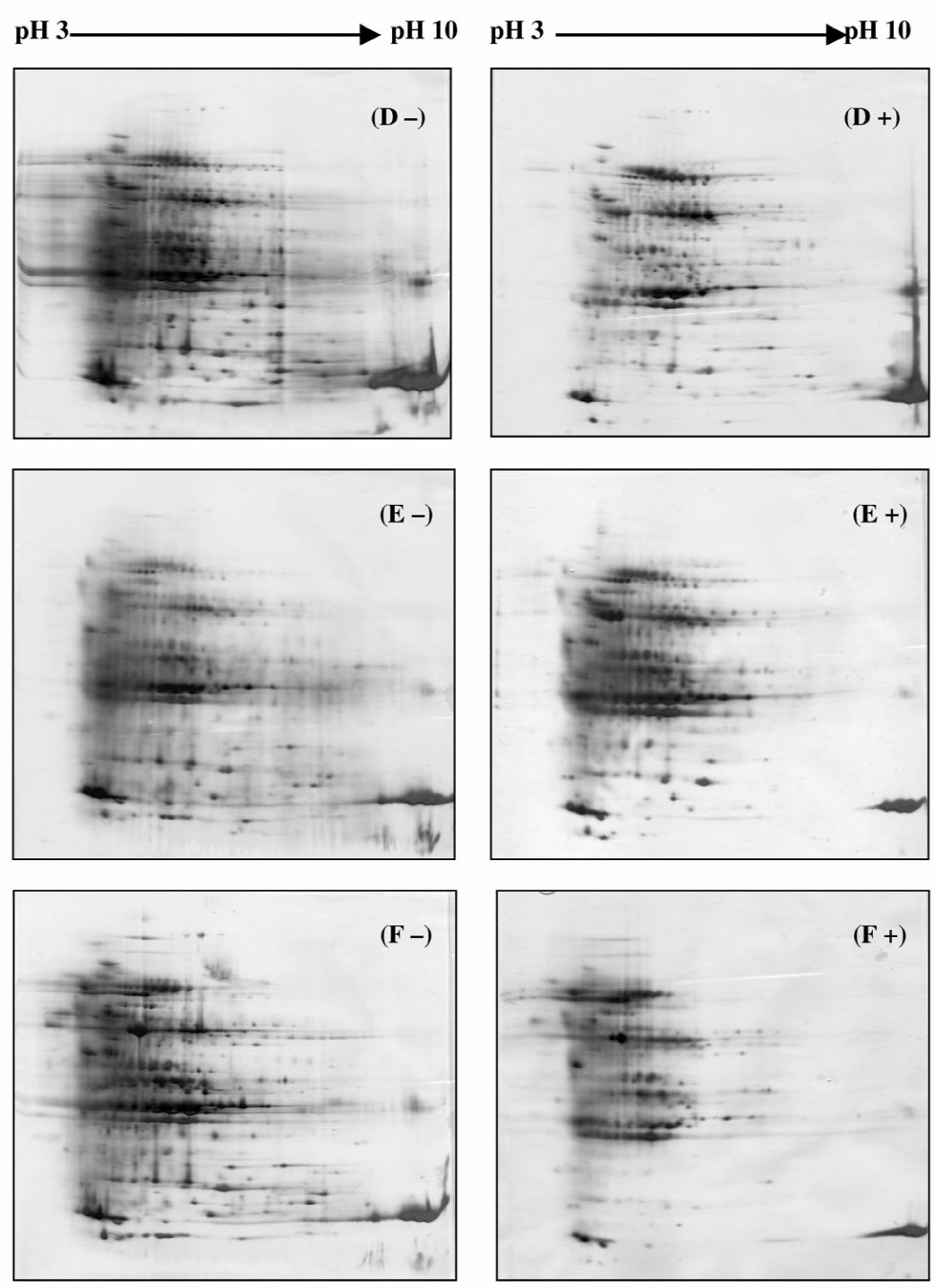

Fig. 2. (*continued*)

SDS–PAGE with the different protein patterns (Fig. 3) and reviewed in real-time as separation on different electropherograms (Fig. 4).

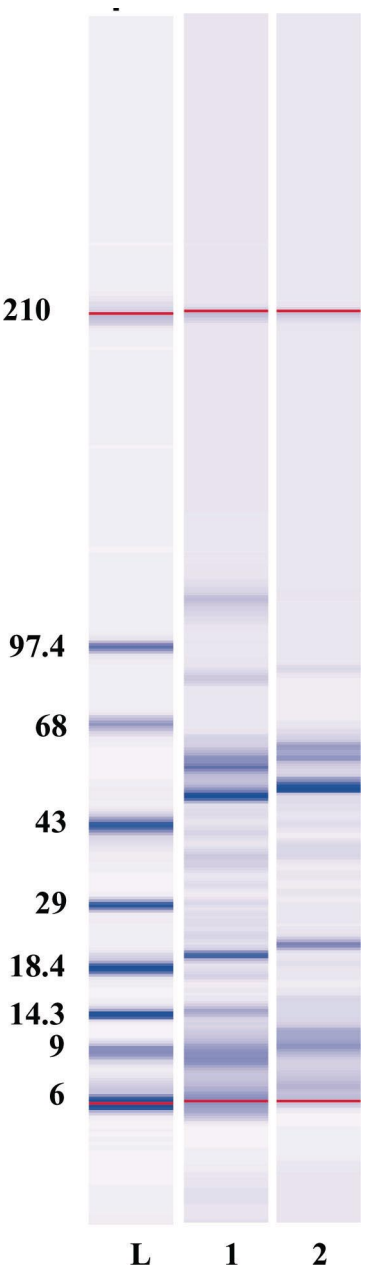

Fig. 3. Analysis of olive oil-induced lipase from *B. stearothermophilus* P1 with a linear gel, gradient gel, and the Protein 200 assay using Agilent 2100 Bioanalyzer. Labels: (L) protein marker (kDa); (1) 0.5% (v/v) olive oil induction; (2) non-induction.

3.2. On the cell membrane

The membrane proteins are found in a few strains of *Bacillus* sp., for example, the membrane proteinase from *B. cereus* [20], δ-endotoxin and insecticidal crystal protein Cry1C from *B. thuringiensis* [21,22], and penicillinase from *B. licheniformis* [23]. Most Gram-positive *Bacillus* sp. are found to be extracellular proteins and have a few proteins in membrane portions. The membrane proteins in *Bacillus* sp. have rarely been identified using 2-D PAGE analysis. Otherwise, the membrane-associated or -bound proteins are difficult to underrepresent on 2-D gels because they are generally low abundant, alkaline p*I* and poorly soluble in aqueous media used for isoelectric focusing. However, it has also found in the genome of *B. subtilis* with 29.2% membrane proteins and the proteomic approach using 2-D PAGE found some membrane proteins such as YfnI, LytC, MurG and Ffh [24,25]. Furthermore, the hydropathy profile alignment can be introduced as a tool to search for structural homologs of membrane proteins and the *B. subtilis* genome was found 109 sequence homologs of membrane proteins [26].

3.3. In the cytosol

Bacillus species have also been used to produce a number of intracellular proteins on an industrial scale, including glucose isomerase [27], thermostable enzymes such as glucokinases [28], and a variety of restriction endonucleases such as *Bam*HI, *Bci*I, *Bgl*II and *Bst*I [29]. The proteomic analysis of cytosolic proteins in *Bacillus* sp. had a few study; for example, the biofilm proteome of whole cell proteins from *B. cereus* strain 5 was investigated using 2-D PAGE and 10 proteins were synthesized as result of surface attachment of which four were unique to biofilm profile [30], the cytosolic proteins of *B. subtilis* 168 was also identified up to 50 proteins using 2-D PAGE [31,32], and the cold shock protein in a thermophilic bacterium *B. stearothermophilus* P1 was also found with the molecular mass approximately 7.3 and p*I* 4.5 using 2-D PAGE (Fig. 5) and the comparison of downshift temperatures from 65 to 37 °C and 25 °C by ImageMaster 2D Elite software showed the different protein patterns (Fig. 6) [33]. In addition, the new protein analysis technique of

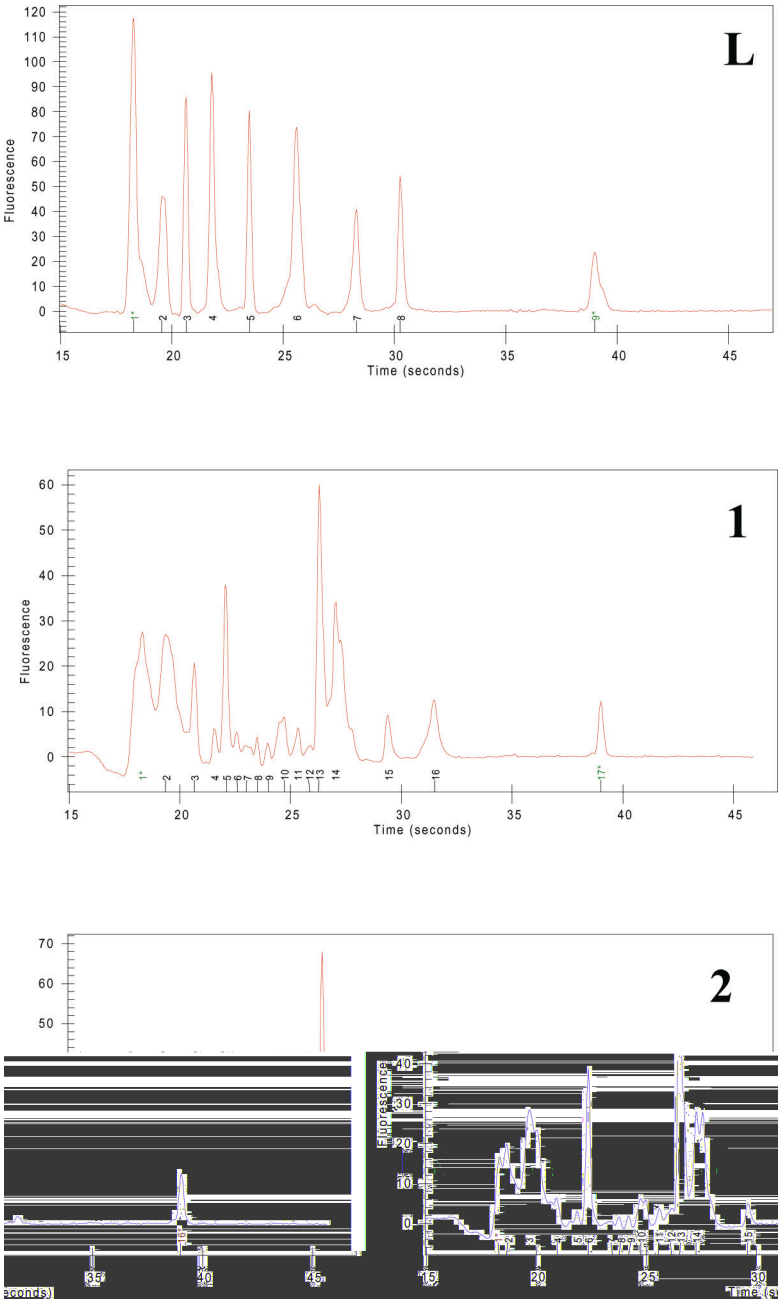

Fig. 4. Electropherogram of olive oil-induced lipase from *B. stearothermophilus* P1 demonstrating the resolution of the Protein 200 assay using Agilent 2100 Bioanalyzer. Labels: (L) protein marker (kDa); (1) 0.5% (v/v) olive oil induction; (2) non-induction.

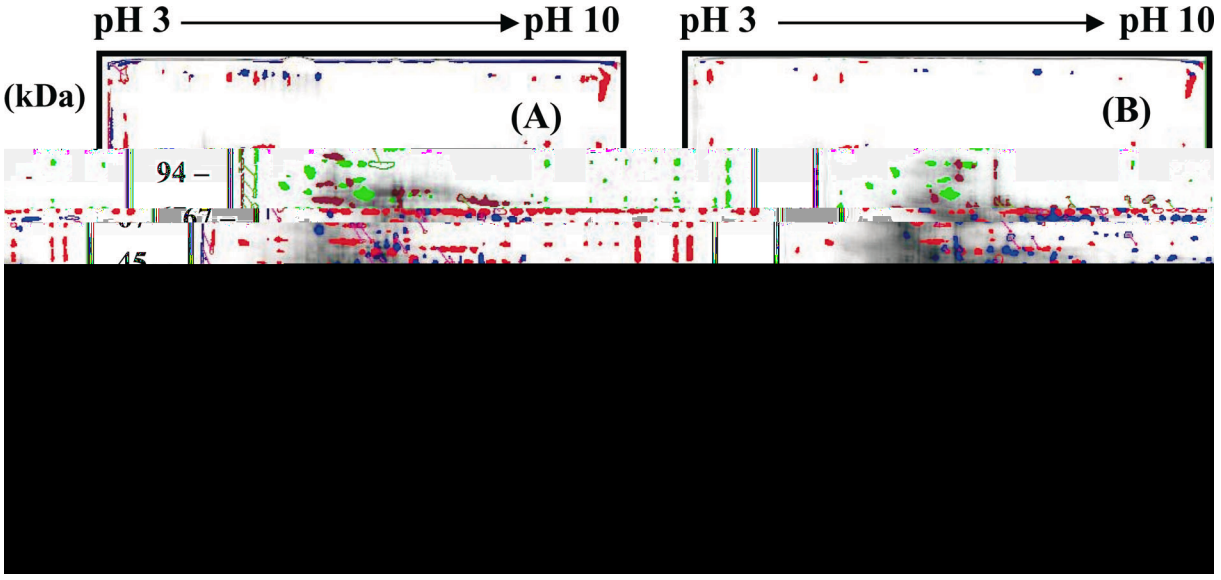

Fig. 6. 2-D Gel image comparison of the protein contents from *B. stearothermophilus* P1 after temperature downshifts from 65 to 37 °C (A) and 25 °C (B) by ImageMaster 2D Elite software. The same analysis parameters show the sensitivity of 8475, operator size of 31, noise factor of 5, background factor of 1 and split level of 7. Labels: blue color spot, protein contents at 37 and 25 °C; red color spot, protein contents at 65 °C as a reference; no color spot with the line, similar protein spot.

proteins are frequently glycosylated, whereas intracellular signaling molecules can be basically or inducible phosphorylated. It is thought that at any one time approximately one-third of all proteins in eukaryotic cells are phosphorylated. Although it is more difficult to detect and analyze post-translational modification containing peptides, post-translational modifications are frequently discovered during large-scale proteomic experiments. In the study of post-translational modifications of *Bacillus* sp., it still has a few reports; for example, Lentz et al. [35] studied the modification of fatty acid specificity of cytochrome P450 BM-3 (CYP102) from *B. megaterium* by directed evolution using a validated assay. The CYP102 enzyme catalyzing the subterminal hydroxylation of fatty acids with a chain length of 12–22 carbons was mutated variants towards caprylic, capric and lauric acid by site-specific random mutagenesis at position 87 of the mutant F87A(LARV). The best mutants, F87V(LAR) and F87V(LARV), showed a higher catalytic activity toward ω-(*p*-nitrophenoxy)decanoic acid (10-*p*-NCA) than F87A(LARV). In addition, they proved capable of hydroxylating ω-(*p*-nitrophenoxy)octanoic acid (8-*p*-NCA) which the wild-type enzyme is unable to do. Both variants catalyzed hydroxylation of capric acid, which is not a substrate for the wild-type, with a conversion rate of up to 57%. The chain length specificity of the mutants in fatty acid hydroxylation processes shows a good correlation with their activity towards *p*-NCA pseudosubstrates. The *p*-NCA assay therefore, allows high-throughput screening of large mutant libraries for the identification of enzyme variants with the desired catalytic activity towards fatty acids as the neutral substrates. In another study by Lessard et al. [36], the gene encoding the dihydrolipoyl acetyltransferase (E2) and dihydrolipoyl dehydrogenase (E3) components of the pyruvate dehydrogenase (PDH) multienzyme complex from *B. stearothermophilus* were overexpressed in *E. coli* and used the post-translational modification for improvement of stoichiometry of the subunit interaction in assembly in vitro. The lipoyl domain of the recombinant E2 protein was found to be unlipoylated, but it could be correctly post-translationally modified in vitro using a recombinant lipoate protein ligase from *E. coli*. The lipoylated E2 component was able to bind recombinant E1 and E3

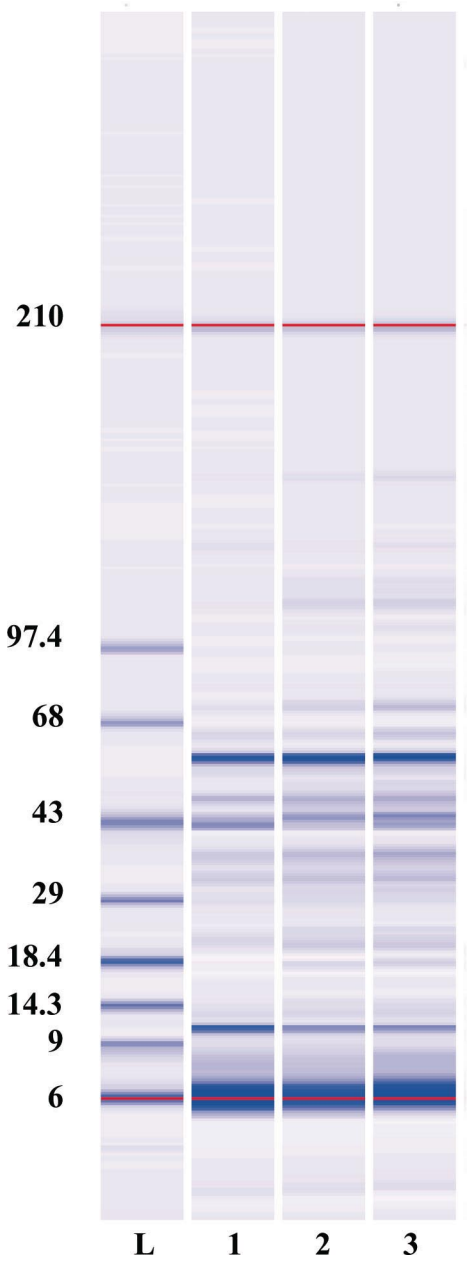

Fig. 7. Analysis of cold shock-stress proteins from *B. stearothermophilus* P1 after downshift temperatures from 65 to 37 °C and 25 °C with a linear gel, gradient gel, and the Protein 200 assay using Agilent 2100 Bioanalyzer. Labels: (L) protein marker (kDa); (1) control at 65 °C; (2) downshift temperature at 37 °C; (3) downshift temperature at 25 °C.

components in vitro to generate a PDH complex with a catalytic activity comparable with that of the wild-type enzyme. Therefore, the use of post-translational modification could be assembly of a fully active PDH complex from recombinant proteins in vitro. On the other hand, Schreier et al. [37] also studied the *B. subtilis glnR* mutants defective in regulation by post-translational modification. The *B. subtilis glnR* gene encoded a 135-amino acid repressor, GlnR, regulates *glnRA* transcription in response to nitrogen levels in the growth medium. Two glnR mutants unable to repress under nitrogen excess conditions were mutated at Gly72 that lie within α-helix-turn-α-helix (HTH) motif by site-directed mutagenesis and affected regulation, suggesting that amino acids within the putative HTH region were critical for GlnR function and might be involved in DNA binding. The mutations within the C-terminal region of GlnR at amino acid 110, 116, 123 and 129 were found to affect regulation because the production of truncated proteins were constitutively repressed and the substitution of Asp129 with Asn also led to loss of repression. In addition, the study of Sakamoto et al. [38] reported that the *meso*-diaminopimelate dehydrogenase gene from *B. sphaericus* was expressed in *E. coli* and the N-terminal amino acid of both the enzyme from native and transformant were serine, indicating that the N-terminal methionine is removed by post-translational modification. Even though the post-translational modifications cannot currently be studied at high throughput, certain categories such as phosphorylation, acylation or alkylation also influence protein structure and function, and the known protein sequence specificity.

Furthermore, the post-translational modification databases and resources can be found on the Internet with websites:

(1) DSDBASE, disulfide database derived from 3D data (http://www.ncbs.res.in/~faculty/mini/dsdbase/dsdbase.html),
(2) GlycoSuiteDB, database of glycan structures (http://www.glycosuite.com),
(3) *O*-GlycBase, *O*-glycosylated protein database (http://www.cbs.dtu.dk/databases/OGLYCBASE),

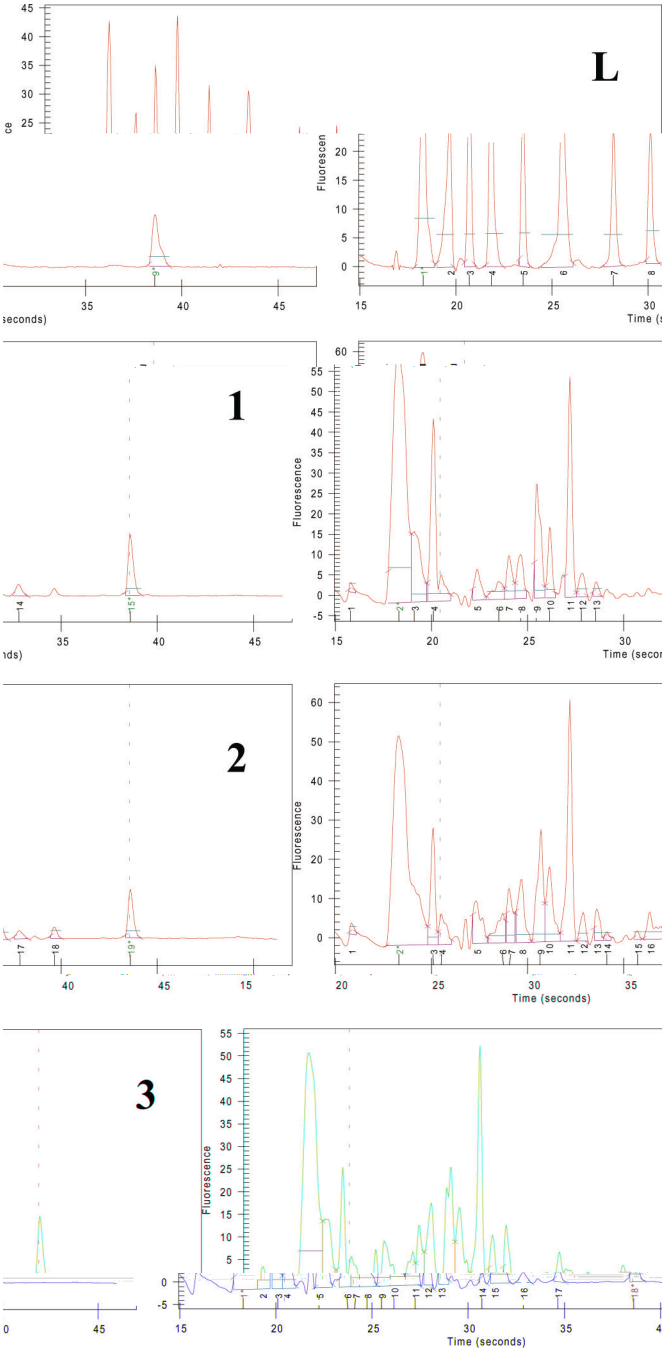

Fig. 8. Electropherogram of cold shock-stress proteins from *B. stearothermophilus* P1 after downshift temperatures from 65 to 37 °C and 25 °C demonstrating the resolution of the Protein 200 assay using Agilent 2100 Bioanalyzer. Labels: (L) protein marker (kDa); (1) control at 65 °C; (2) downshift temperature at 37 °C; (3) downshift temperature at 25 °C.

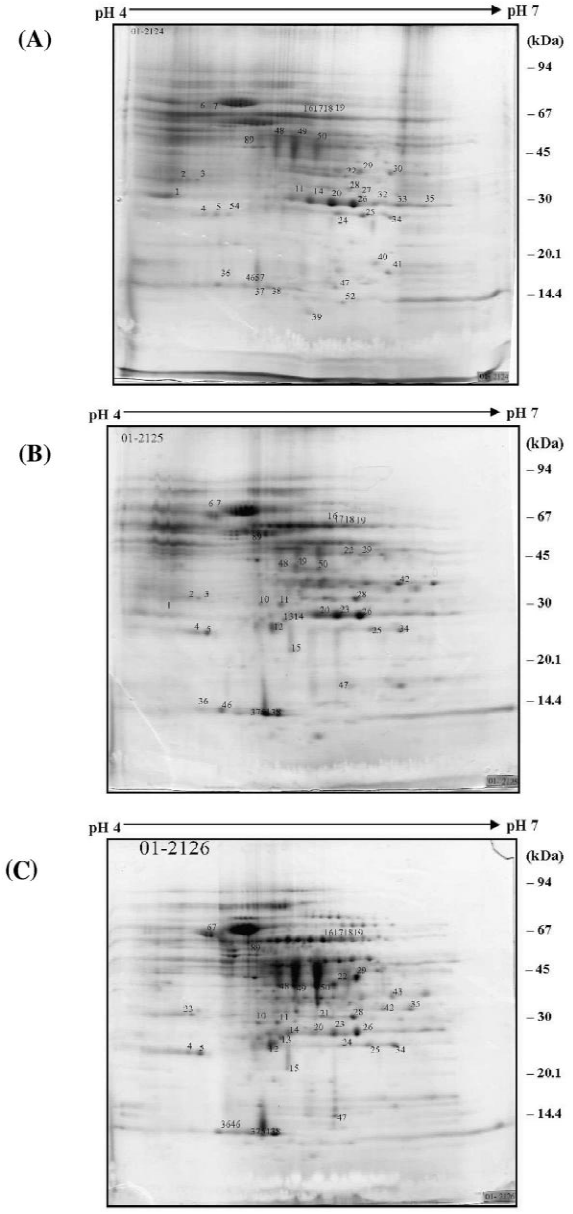

Fig. 9. The cytosolic proteins in *B. stearothermophilus* TLS33 after downshift temperatures to 37 °C (B) and 25 °C (C) analyzed by 2-D PAGE using a narrow pH range of IPG strip 4–7, 18 cm. The control is used the cell extract at 65 °C (A).

(4) PhosphoBase, phosphorylation site database (http://www.cbs.dtu.dk/database/PhosphoBase), and
(5) RESID, database of amino acid modifications (http://www-nbrf.georgetown.edu/pirwww/dbinfo/resid.html)

5. Bioinformatics to link differential display genes and proteomic identified proteins

Bioinformatics are now important and very useful for searching the unknown sequence including gene function. An up-to-date list of the databases of *Bacillus* sp. has several websites and shows in Table 5. Some of these are available on the Internet with links to many provided from the ExPASy proteomics server including protein identification and characterization programs (http://www.expasy.ch/tools.html). These allow not only identification of proteins but further characterization ranging from the calculation of basic physiochemical properties to the prediction of potential post-translational modifications and three-dimensional structures. Annotated protein and two-dimensional electrophoresis databases is the bioinformatic core of proteome research. Swiss-Prot is a typical example of such as annotated database. Many proteome projects are now underway, resulting in the generation of two-dimensional electrophoresis database that are accessible on the Internet and can be browsed with interactive software and integrated with in-house results. SubtiList database is a one server that provides a complete dataset of DNA and protein sequences derived from the paradigm strain *B. subtilis* 168, linked to the relevant annotations and functional assignments, and available at http://genolist.pasteur.fr/SubtiList. It allows one to easily browse through these data and retrieve information, using various criteria (such as gene names, location, keyword, functional category, etc.). This server supplementing with EMBL/GenBank/DDBJ databanks is also used for identification of the cytosolic proteins in *B. stearothermophilus* TLS33 showing the identified protein and gene including accession numbers, description and functional category (Table 6). Another database of Clusters of Orthologous Groups of protein (COG) is a new database search and represents an attempt on a phylogenetic classification of proteins from complete genome (http://www.ncbi.nlm.nih. gov/COG) [39,40]. It is to serve as a platform for functional annotation of newly sequence genomes and for

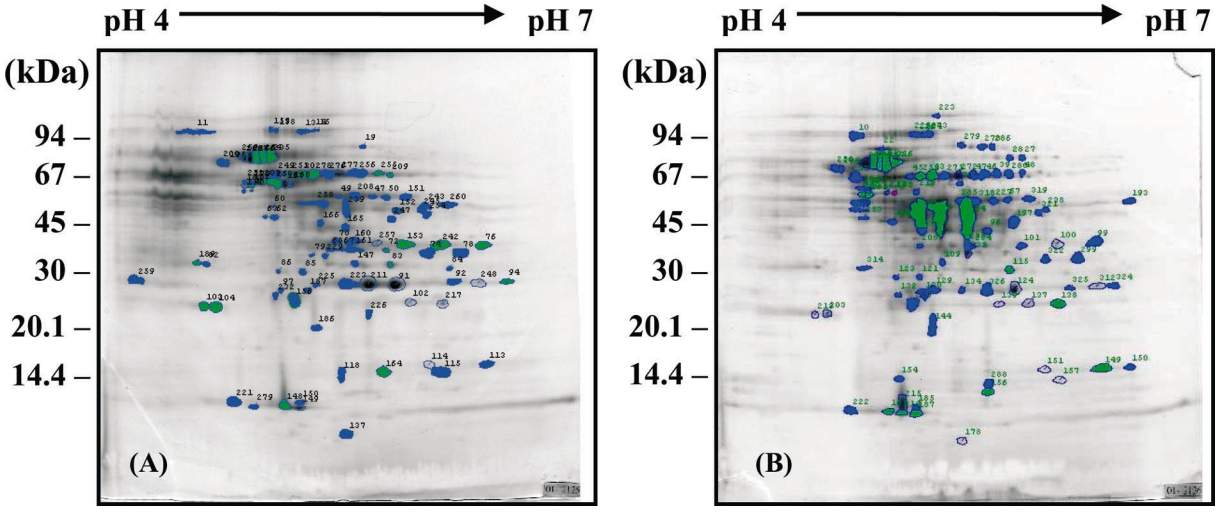

Fig. 10. 2-D Gel image comparison of the cytosolic proteins from *B. stearothermophilus* TLS33 after downshift temperatures from 65 to 37 °C (A) and 25 °C (B) by ImageMaster 2D Elite software. Labels: blue color spots, decreased protein spots; red color spot, increased protein spots; no color spot, no change of proteins.

studies on genome evolution. The protein sequences encoded in two complete genomes of *B. subtilis* and *B. halodurans* were compared by COG database using COGNITOR program (Table 7). It can be used to search the BLAST protein homologs and identified the protein functions. Not only the COG database, the proteome analysis of *B. subtilis* and *B. halodurans* is also identified and shown in the Proteome Analysis@EBI database from website http://www.ebi.ac.uk/proteome that the SWISS-PROT/TrEMBL complete non-redundant proteome sets are constructed by selecting entries from SPTR, which is a comprehensive protein sequence database consisting of SWISS-PROT, TrEMBL and TrEMBLnew (Table 8) [41]. Otherwise, the InterPro comparative analysis of the proteome of *B. subtilis* with another available proteome of *B. halodurans* or *E. coli* can be investigated and shown the functional classification of organism using Gene Ontology (GO) with the general statistics. Therefore, it is very useful and can be used as a bioinformatic tool. Furthermore, we have also developed Bulk Gene Search System (BGSS) that is able to correlated gene identifying from cDNA microarray with their corresponding protein functions. The developed database search of BGSS, which is composed of the UniGene, LocusLink and Proteome databases, can be easily used to find the associated protein functions and related information in the worldwide databases by applying only accession numbers [42]. The BGSS database of *Bacillus* derived from Swiss-Prot and SubtiList database is now available at http://bio-.clkao.org/perl/genequery2.pl that is very useful for analyzing the whole *Bacillus* genomes in the nearly future.

6. Biomedical applications of *Bacillus* sp.

In the present, the effect of genome and proteome based approaches on biomedical research has not yet been achieved. However, the exciting progress is being made, and brief overviews of several biomedical areas are given below to illustrate the potential of this approach. The role of *Bacillus* sp. in biomedical applications is higher important coupled with knowing genome of some species as shown in Table 9. For example, Shih et al. [43] reported the chemistry and biosynthesis of poly-γ-glutamic acid (γ-PGA) produced by various strains of *Bacillus* sp. such as *B. anthracis*, *B. natto*, *B. subtilis*, *B. anthracis* and *B. licheniformis*. γ-PGA is an unusual anionic, naturally occurring homo-polyamide that is made of D- and L-glutamic acid units connected by

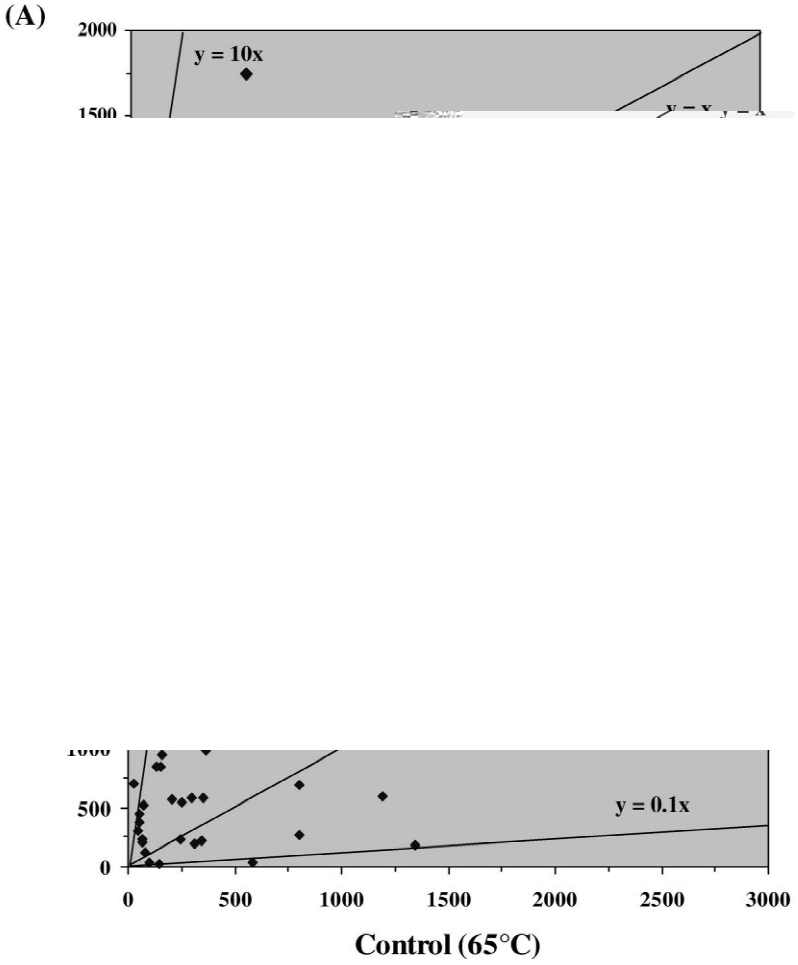

Fig. 11. The relationships of cytosolic proteins from *B. stearothermophilus* TLS33 between control at 65 °C and downshift temperature at 37 °C (A) or 25 °C (B).

amide linkages between α-amino and γ-carboxylic acid groups. Due to suitable properties of γ-PGA such as water-soluble, biodegradable, edible and non-toxic toward humans and the environment, their potential applications have been of interest in the past few years in a broad range of industrial fields such as fields of food, cosmetics, medicine and water treatments. γ-PGA and its derivatives also offer a wide range of unique applications including being used as thickener, cryoprotectant, bitterness relieving agent, sustained release materials, humectant, drug carrier, biodegradable fibers, biological adhesive, highly water-absorbable hydrogels, biopolymer flocculants, heavy metal absorber and animal feed additives. Most applications of γ-PGA and its derivatives in medicine are used as drug carrier and biological adhesives. For example, polyglutamic pacitaxel (PG-TXL) or Taxol is an anticancer agent that exhibited markedly antitumor activity against murine tumors and human tumor xenografts [44,45], prostaglandin E_1 (PGE_1) is clinically used to treat peripheral vascular disturbance, in addition, it is also effective on fulminant or subfulminant viral hepatitis by means of its cytoprotective activity [46], the cross-linking of gelatin and poly-(L-glutamic acid) has been shown to be promising as a surgical adhesive and hemostatic agent [47–50], the cured gel or gelatin–PGA aqueous solutions is lowly biodeg-

Table 4
Protein identification of cytosolic proteins in *B. stearothermophilus* TLS33 after downshift temperatures to 37 and 25 °C by MS-FIT software (http://prospector.ucsf.edu/ucsfbin3.4/msfit.cgi) and SwissProt and TrEMBL (http://www.expasy.ch/sprot)

Spot no.	App. M_w (Da)/p*I*	MOWSE score	Matched	Theoretical M_w (Da)/p*I*	Accession no.	Protein name
1	35051/4.37	207	9/29 (31%)	18225.3/8.96	P26380	PTS system, fructose-specific IIB component (EIIB-FRU) (fructose-permease IIB component) (phosphotransferase enzyme II, B component) (P18)
2	39191/4.55	4.32	4/14 (28%)	43843.1/6.55	P49782	Stage III sporulation protein AE
3	39191/4.62	13.2	5/17 (29%)	55725.3/6.31	P21879	Inosine-5′-monophosphate dehydrogenase (IMP dehydrogenase) (IMPDH) (IMPD) (superoxide-inducible protein 12) (SOI12)
4	29901/4.76	15.9	4/23 (17%)	26413.8/9.11	P39147	COMF operon protein 3
5	78100/4.64	–	–	–	–	Unknown
6	78100/4.64	118	4/27 (14%)	13536.7/5.71	P39740	Flagellar protein flit
7	78250/4.55	8.66	26/47 (55%)	79078.9/8.97	P39814	DNA topoisomerase I (ω-protein) (relaxing enzyme) (untwisting enzyme) (swivelase)
8	78700/4.44	2.72	5/18 (27%)	78304.4/6.07	P13484	Probable poly (glycerol-phosphate) α-glucosyltransferase (teichoic acid biosynthesis protein E)
9	78100/4.64	3.74	29/40 (72%)	78622.2/5.40	P17889	Translation initiation factor IF-2
10	36067/5.1	43.1	6/30 (20%)	38120.6/5.01	P54518	Putative peptidase in GCVT-SPOIIIAA intergenic region
11	34803/5.14	31.8	4/30 (13%)	14349.6/6.30	P42411	Anti-σB factor RSBT
12	31533/5.2	71.5	11/32 (34%)	29759.9/5.24	Q45499	Extragenic suppressor protein SUHB homolog
13	33103/5.3	–	–	–	–	Unknown
14	49187/5.53	51.6	5/22 (22%)	28560.1/6.56	P50733	Hypothetical 28.6-kDa protein in RECQ-CMK intergenic region precursor
15	49187/5.53	51.6	11/37 (29%)	56309.1/5.34	P39773	2,3-bisphosphoglycerate-independent phosphoglycerate mutase (phosphoglyceromutase) (BPG-independent PGAM) (vegetative protein 107) (VEG107)
16	69284/5.49	5.57	26/46 (56%)	68587.5/5.17	P37949	GTP-binding protein LEPA
17	69284/5.61	157	12/38 (31%)	56309.1/5.34	P39773	2,3-bisphosphoglycerate-independent phosphoglycerate mutase (phosphoglyceromutase) (BPG-independent PGAM) (vegetative protein 107) (VEG107)
18	69486/5.72	–	–	–	–	Unknown
19	69284/5.81	7.31	4/11 (36%)	20182.6/5.75	P54154	Peptide methionine sulfoxide reductase (protein-methionine-s-oxide reductase) (peptide MET(O) reductase)
20	686284/5.55	29.6	11/28 (39%)	38639.1/5.39	P50863	MRP protein homolog
21	68614/5.59	137	12/31 (38%)	58175.0/4.95	P50848	Hypothetical 58.2-kDa protein in KDGT-XPT intergenic region
22	39144/5.66	64.2	14/30 (46%)	46609.5/5.52	P25995	Dihydroorotase (dhoase)
23	41924/5.63	186	8/27 (29%)	20126.3/5.43	Q06795	Transcription antitermination protein NUSG
24	33229/5.67	16	4/29 (13%)	29955.0/6.46	P49938	Ferrichrome transport ATP-binding protein FHUC
25	30331/5.85	4.12	5/16 (31%)	32510.1/8.77	P24327	Protein export protein PRSA precursor
26	29802/6.16	84.1	6/24 (25%)	23303.0/5.78	O34565	Amidotransferase HISH
27	35465/5.84	84.1	6/24 (25%)	23303.0/5.78	O34565	Amidotransferase HISH
28	3485/5.75	8.46	6/36 (16%)	35012.1/6.47	O34504	Hypothetical 35.0-kDa protein in RAPJ-OPUAA intergenic region
29	37121/4.55	15.6	18/36 (50%)	60763.5/6.42	P40408	Hypothetical transcriptional regulator in FEUA-SIGW intergenic region (ORF3)
30	78250/6.07	–	–	–	–	Unknown
31	41178/5.85	23.6	8/30 (26%)	43247.7/6.51	O34992	Glycine betaine/carnitine/choline transport ATP-binding protein OPUCA
32	32081/6.01	2.09	6/21 (28%)	99563.2/4.94	P39793	Penicillin-binding protein 1A/1B (PBP1) [includes: penicillin-insensitive transglycosylase (peptidoglycan TGASE); penicillin-sensitive transpeptidase (DD-transpeptidase)]
33	32484/6.14	5.76	4/37 (10%)	30244.1/9.15	P39785	PBSX phage terminase small subunit
34	29604/6.05	1.86	4/18 (22%)	26949.8/8.79	P54537	Probable amino acid ABC transporter ATP-binding protein in BMRU-ANSR intergenic region
35	32236/6.3	303	4/26 (15%)	19119.6/4.62	P40405	Hypothetical 19.1-kDa protein in SIGD-RPSB intergenic region precursor (ORFC)
36	18139/4.6	48.4	4/33 (12%)	14573.7/5.79	P17622	RIBT protein
37	18200/4.77	21.1	4/26 (15%)	14538.0/5.54	P39910	PAL-related lipoprotein precursor
38	17648/5.15	91.7	4/33 (12%)	6189.2/4.93	2897794	(D49467) Unnamed protein product

Table 4. Continued

Spot no.	App. M_w (Da)/p*I*	MOWSE score	Matched	Theoretical M_w (Da)/p*I*	Accession no.	Protein name
39	12806/5.44	48.8	4/24 (16%)	11124.6/5.23	P21468	30S Ribosomal protein S6 (BS9)
40	22031/5.95	125	5/28 (17%)	26974.8/5.58	P94498	Phosphoadenosine phosphosulfate reductase (PAPS Reductase, Thioredoxin-dependent) (padops Reductase) (3′-phosphoadenylylsulfate reductase) (PAPS sulfotransferase)
41	20546/6.05	–	–	–	–	Unknown
42	28549/6.14	47.6	11/27 (40%)	73207.7/5.01	P54496	Hypothetical 73.2-kDa protein in SODA-COMGA intergenic region
43	40764/6.32	34.1	13/23 (56%)	43836.3/4.84	P08495	Aspartokinase 2 (aspartokinase II) (aspartate kinase 2) [contains: aspartokinase II α subunit; aspartokinase II β subunit]
44	43986/5.25	202	17/36 (47%)	45333.3/6.12	P45742	Hypothetical 45.3-kDa protein in PRKA-CSPB intergenic region (ORF4)
45	16284/5.35	17	6/29 (20%)	17895.3/5.37	O07636	Hypothetical 17.9-kDa protein in NPRE-PYCA intergenic region
46	18200/4.77	–	–	–	–	Unknown
47	17587/5.64	5.69	4/10 (40%)	23805.8/4.78	P49785	Stage III sporulation protein AH
48	61667/5.2	103	9/28 (32%)	16231.7/8.38	P39807	Hypothetical 16.2-kDa protein in COMF-FLGM intergenic region
49	58758/5.34	181	9/32 (28%)	29728.7/5.31	P39650	Prespore-specific transcriptional activator RSFA
50	58515/5.52	240	21/35 (60%)	50527.5/5.04	P80860	Glucose-6-phosphate isomerase (GPI) (phosphoglucose isomerase) (PGI) (phosphohexose isomerase) (PHI) (vegetative protein 54) (VEG54)
51	20249/5.07	–	–	–	–	Unknown
52	14461/5.7	129	4/33 (12%)	21023.0/6.59	P37569	Hypothetical 21.0-kDa protein in LYSS-MECB intergenic region
53	18445/5.87	22.1	5/23 (21%)	19145.1/5.79	O31704	Molybdopterin-guanine dinucleotide biosynthesis protein B

Table 5
Database searches of *Bacillus* sp.

Server	Web sites
General	
1. LBSN	http://www.bacterio.cict.fr/b/bacillus.html
2. DSMZ	http://www.dsmz.de/bactnom/nam0379.htm
Genome project	
1. Entrez	http://www.ncbi.nlm.nih.gov/Entrez
2. *B. stearothermophilus* Genome Sequencing–Strain 10	http://www.genome.ou.edu/bstearo.html
3. *B. subtilis* Genome Sequencing Project	http://genolist.pasteur.fr/SubtiList/help/project.html
4. *B. halodurans* Genome Project–JAMSTEC	http://www.sun01.hydra.mki.co.jp:8093/jamstec/micrHome.html
Genome analysis (DNA sequence data)	
1. DDBJ	http://www.ddbj.nig.ac.jp
2. EMBL	http://www.ebi.ac.uk/embl/index.html
3. Genbank	http://www.ncbi.nlm.nih.gov/Genbank/Genbank/GenbankSearch.html
4. SubtiList	http://genolist.pasteur.fr/SubtiList
5. BSORF (JAFAN)	http://bacillus.genome.ad.jp/BSORF-DB.html
6. NRSub (Non-Redundant *B. subtilis* Database)	http://pbil.univ-lyon1.fr/nrsub/nrsub.html
7. Micado	http://locus.jouy.inra.fr/cgi-bin/genmic/madbase/progs/madbase.operl
8. PEDANT	http://pedant.gsf.de/index.html
9. EBI GeneQuiz	http://jura.ebi.ac.uk:8765/ext-genequiz/genomes/bs0005
10. GIB (DDBJ)	http://gib.genes.nig.ac.jp/Bs
11. TIGR	http://www.tigr.org/tdb/mdb/mdb/mdbcomplete.html
Proteome analysis (Protein sequence data)	
1. Proteome Analysis@EBI	http://www.ebi.ac.uk/proteome/BACSU (*B. subtilis*), or BACHD (*B. halodurans*)
2. COGs	http://www.ncbi.nlm.nih.gov/COG
3. Protein Data Bank	http://www.rcsb.org/pdb
4. SwissProt	http://kr.expasy.org/sprot
5. Sub-2D	http://microbio2.biologie.uni-greifswald.de:8880

Table 6
Protein and gene identification, description, and functional category of cytosolic proteins in *B. stearothermophilus* TLS33 after downshift temperatures at 37 and 25 °C by SubtiList databases, supplemented with EMBL/GenBank/DDBJ databanks

Spot no.	Protein ID/no.	Gene name/no.	Description	Functional category
1	PTFB_BACSU (P26380)	*levE* (X56098)	PTS fructose-specific enzyme IIB component	Transport/binding proteins and lipoproteins
2	S3AE_BACSU (P49782)	*spoIIIAE* (U35252)	Mutants block sporulation after engulfment	Sporulation
3	IMDH_BACSU (P21879)	*guaB* (X55669)	Inosine-monophosphate dehydrogenase	Metabolism of nucleotides and nucleic acids
4	CMF3_BACSU (P39147)	*comFC* (Z18629)	Late competence gene	Transformation/competence
5	Unknown	–	–	–
6	FLIT_BACSU (P39740)	*fliT* (Z31376)	Flagellar protein	Mobility and chemotaxis
7	TOP1_BACSU (P39814)	*topA* (L27797)	DNA topoisomerase I	DNA packing and segregation
8	TAGE_BACSU (P13484)	*tagE* (X15200)	UDP-glucose: polyglycerol phosphate glucosyltransferase	Cell wall
9	IF2_BACSU (P17889)	*infB* (M34836)	Initiation factor IF-2	Initiation
10	YQHT_BACSU (P54518)	*yqhT* (D84432)	Unknown; similar to Xaa-Pro dipeptidase	Protein modification
11	RSBT_BACSU (P42411)	*rsbT* (L35574)	Positive regulator of σB activity (switch protein/serine-threonine kinase)	Adaptation to atypical conditions
12	SUHB_BACSU (Q45499)	*suhB* (AF012285)	–	–
13	Unknown	–	–	–
14	YPBG_BACSU (P50733)	*ypbG* (L47648)	Unknown; similar to unknown proteins	From other organisms
15	PMGI_BACSU (P39773)	*pgm* (L29475)	Phosphoglycerate mutase	Main glycolytic pathways
16	LEPA_BACSU (P37949)	*lepA* (X91655)	GTP-binding protein	Elongation
17	PMGI_BACSU (P39773)	*pgm* (L29475)	Unknown	–
18	Unknown	–	–	–
19	MSRA_BACSU (P54154)	*msrA* (L77246)	Peptidyl methione sulfoxide reductase	Detoxification
20	MRP_BACSU (P50863)	*mrp* (X74737)	–	–
21	YPWA_BACSU (P50848)	*ypwA* (L47838)	Unknown; similar to carboxypeptidase	Metabolism of amino acids and related molecule
22	Unknown	–	–	–
23	NUSG_BACSU (Q06795)	*nusG* (D13303)	Transcription antitermination factor	Termination
24	FHUC_BACSU (P49938)	*fhuC* (X93092)	Ferrichrome ABC transporter (ATP-binding protein)	Transport/binding proteins and lipoproteins
25	PRSA_BACSU (P24327)	*prsA* (X57271)	Protein secretion (post-translocation molecular chaperone)	Protein secretion
26	Unknown	–	–	–
27	HIS5_BACSU (O34565)	*hisH* (AF017113)	Amidotransferase	Metabolism of amino acid and related molecules

Table 6. Continued

28	YCEB_BACSU (O34504)	*yceB* (AB000617)	Unknown; similar to unknown proteins	From other organisms
29	YBBB_BACSU (P40408)	*ybbB* (L19954)	Unknown; similar to transcriptional regulator (AraC/XylS family)	Regulation
30	Unknown	–	–	–
31	OPCA_BACSU (O34992)	*opuCA* (AF009352)	Glycine betaine/carnitine/choline ABC transporter (ATP-binding protein)	Transport/binding proteins and lipoproteins
32	PBPA_BACSU (P39793)	*ponA* (U11883)	Penicillin-binding proteins 1A/1B	Cell wall
33	XTMA_BACSU (P39793)	*xtmA* (Z70177)	PBSX terminase (small subunit)	Phage-related functions
34	YQIZ_BACSU (P54537)	*yqiZ* (D84432)	Unknown; similar to amino acid ABC transporter (ATP-binding protein)	Transport/binding proteins and lipoproteins
35	YLXL_BACSU (P40405)	*ylxL* (Z99112)	Unknown	–
36	RIBT_BACSU (P17622)	*ribT* (L09228)	Reductase	Metabolism of coenzymes and prosthetic groups
37	SLP_BACSU (P39910)	*slp* (M57435)	Small peptidoglycan-associated lipoprotein	Transport/binding proteins and lipoproteins
38	YVYF_BACSU (P39807)	*yvyF* (L14437)	Unknown; similar to flagellar protein	Mobility and chemotaxis
39	Unknown	–	–	–
40	CYH1_BACSU (P94498)	*cysH* (U76751)	Phosphoadenosine phosphosulfate	Metabolism of amino acids and related molecules
41	Unknown	–	–	–
42	YQGS_BACSU (P54496)	*yqgS* (D84432)	Unknown; similar to putative molybdate binding protein	From other organisms
43	AK2_BACSU (P08495)	*lysC* (J03294)	Aspartokinase II (α and β subunits)	Metabolism of amino acids and related molecules
44	YHBH_BACSU (P45742)	*yhbH* (Z99108)	Unknown; similar to unknown proteins	From other organisms
45	Unknown	–	–	–
46	Unknown	–	–	–
47	Unknown	–	–	–
48	YLAL_BACSU (O07636)	*ylaL* (Z97025)	Unknown; similar to unknown proteins	From *B. subtilis*
49	RSFA_BACSU (P39650)	*rsfA* (X73124)	Probable regulator of transcription of σF-dependent genes	Sporulation
50	G6PI_BACSU (P80860)	*pgi* (Z93936)	Glucose-6-phosphate	Main glycolytic pathways
51	YQBP_BACSU (P45932)	*yqbP* (D32216)	Unknown; similar to phage-related protein	Phage-related functions
52	YACH_BACSU (P37569)	*yacH* (D26185)	Unknown	–
53	MOBB_BACSU (O31704)	*mobB* (AF012285)	Molybdopterin-guanine dinucleotide biosynthesis	Metabolism of coenzymes and prosthetic groups

raded in the body without inducing any problematic inflammatory response, and a potential biological adhesive made from porcine collagen and poly-(L-glutamic acid) has been developed, which is superior to fibrin in sealing air leakage from the lungs [51]. Moreover, most drugs are metabolized in the liver via the cytochrome P-450 family of enzymes, encompassing a wide array of overlapping specificities for different substrates [52,53]. They also play a major role in metabolism that can lead to the production and removal of toxic species, and in some instances it is possible to correlate the ability of failure to remove such a toxin with specific P-450 or subgroup. Not only cytochrome P-450 was found in

Table 7
Comparison of the function systems of the *Bacillus* sp. in protein coding genes by COG database (http://www.ncbi.nlm.nih.gov/COG)

Genome functions	*B. subtilis*	*B. halodurans*	*B. stearothermophilus* TLS33[a]
Information storage and processing			
1. Translation, ribosomal structure and biogenesis	152	153	–
2. Transcription	272	269	–
3. DNA replication, recombination and repair	131	227	–
Cellular processes			
1. Cell division and chromosome partitioning	31	32	–
2. Posttranslational modification, protein turnover, chaperones	87	84	–
3. Cell envelope biogenesis, outer membrane	161	84	–
4. Cell motility and secretion	90	90	–
5. Inorganic ion transport and metabolism	148	146	–
6. Signal transduction mechanisms	122	135	–
Metabolism			
1. Energy production and conversion	164	158	2
2. Carbohydrate transport and metabolism	271	262	5
3. Amino acid transport and metabolism	293	284	15
4. Nucleotide transport and metabolism	82	73	5
5. Coenzyme metabolism	109	109	–
6. Lipid metabolism	84	89	3
7. Secondary metabolites biosynthesis, transport and catabolism	128	110	4
Poorly characterized			
1. General function prediction only	332	329	–
2. Function unknown	226	246	–
Not in COGs	1221	1155	–

[a] Some identifications.

human but it was also found in microorganism; for example, Hopkins et al. [54] and English et al. [55] reported the induction of cytochrome P-450_{BM-3} (CYP102) by barbiturates, peroxisome proliferators and non-steroidal anti-inflammatory drugs in *B. megaterium*, in which cytochrome *P*-450s are a superfamily of mono-oxygenases involved in the oxidation of a wide range of compounds, including steroids, fatty acids and xenobiotics such as drugs and environmental chemicals. The inducers interacted with the repressor protein Bm3R1 that also interacted with fatty acids, causing it to dissociate with the operator of the CYP102 gene and allowing transcription to occur. Otherwise, they also play a key part in various steps of primary and secondary metabolism, as well as in drug detoxification [56]. Schiavon et al. [57] used the uricase from *B. fastidiosus* (UC) conjugate to linear or branched poly(ethylene glycol) and poly(*N*-acryloymorpholine) for improved therapy with uricase and prevented gout disease or hyperuricemia and the uricase could catalyse the oxidation of uric acid, a final product of purine catabolism, to allantoin which is more easily excreted than the starting compound.

Table 8
Comparison of proteome analysis of *B. subtilis* and *B. halodurans* by Proteome Analysis@EBI database based on high-level terms of the Gene Ontology (GO) data that has been assigned to InterPro entries and shown the general statistics for the number of proteins in the proteome

GO Classification	*B. subtilis*	*B. halodurans*
Term		
Nucleic acid binding	475 (11.5%)	435 (10.8%)
DNA binding	370 (8.9%)	347 (8.6%)
DNA repair protein	22 (0.5%)	19 (0.4%)
DNA replication factor	6 (0.1%)	9 (0.2%)
Transcription factor	212 (5.1%)	217 (5.4%)
RNA binding	100 (2.4%)	87 (2.1%)
Structural protein of ribosome	60 (1.4%)	59 (1.4%)
Translation factor	12 (0.2%)	12 (0.2%)
Motor	6 (0.1%)	5 (0.1%)
Enzyme	1310 (31.7%)	1192 (29.7%)
Peptidase	92 (2.2%)	77 (1.9%)
Protein kinase	56 (1.3%)	63 (1.5%)
Protein phosphatase	20 (0.4%)	12 (0.2%)
Signal transducer	95 (2.3%)	115 (2.8%)
Receptor	5 (0.1%)	0 (0%)
Transmembrane receptor	5 (0.1%)	0 (0%)
Cell adhesion molecule	2 (0%)	4 (0%)
Structural protein	70 (1.6%)	70 (1.7%)
Transporter	372 (9.0%)	348 (8.6%)
Ion channel	8 (0.1%)	10 (0.2%)
Neurotransmitter transporter	2 (0%)	3 (0%)
Ligand binding or carrier	446 (10.8%)	458 (11.4%)
Electron transporter	97 (2.3%)	93 (2.3%)
Molecular function unknown	39(0.9%)	35 (0.8%)
Unclassified	**1409 (34.1%)**	**1472 (36.7%)**
Total	**4121 (100%)**	**4008 (100%)**

Tomita et al. [58] studied the stabilized enzymatic reagents for measuring glucose, creatine kinase and γ-glutamyltransferase (γ-GT) with thermostable enzymes from a thermophilic bacterium *B. stearothermophilus*. These enzymes are glucokinase (GlcK, EC 2.7.1.2) and alanine dehydrogenase (AlaDH, EC 1.4.1.1), which the long-term stability in using GlcK for measurement of glucose and creatine kinase (CK) activity, and in using AlaDH for γ-glutamyltransferase (γ-GT) activity. In addition, Yamashita et al. [59] found the Cry and Cyt proteins of *B. thuringiensis* exhibiting cytocidal activity against human leukemic T cells and uterus cervix cancer (HeLa) cells but not against normal T cells. Moreover, the *Bacillus* strains could be also used as a host in expression of proteins; for example, Heikkila et al. [60] studied the immunization with meningoccal class 1 outer membrane protein (BacP1) produced in *B. subilis* and reconstituted in the presence of Zwittergent or Triton X-100. The use of adjuvant improved the immunogenicity of the BacP1-detergent preparations and such liposomes were reproducibly immunogenic in mice and guinea pigs at a low dose indicating its immunizing properties. Otherwise, the superoxide dismutase (SOD) in *Bacillus* sp. has also useful in biomedical application such as it can alleviate a wide range of effects produced by exposure to O_2^--generating systems and it has an anti-inflammatory property which was discovered well before the protein was identified as an enzyme. SOD administered intravenously along with catalase, was found to protect against oxygen toxicity and reduced inflammation caused by irradiation. Orgotein is a sample of the generic name adopted for drug versions of highly purified SOD and was reported to have anti-inflammatory and anti-viral activities [61]. On the other hand, the lipase secreting from the same species *B. stearothermophilus* P1 could catalyze the esterification of ibuprofen, which is an important group of non-steroidal anti-inflammatory drugs and used to reduced swelling and inflammation, representing an effective way to prepare the prodrug [62].

7. Conclusions

Genome and proteome analysis of *Bacillus* sp. provide a powerful set of tools for study of functional genomics thoroughly to the biochemistry of proteins and the combination with bioinformatics is also increasingly useful for interpretation of comparative-genomic, functional genomic and proteomic results from many species. Therefore, the proteomics has unique and significant advantages as an important complement to a genomics approach, and applications for this technology are readily apparent.

Acknowledgements

We gratefully thank the National Taiwan Science Council and Academia Sinica, Taiwan for supporting a grant of the *Bacillus* research. This review is also part of the Royal Golden Jubilee Ph.D. project of Mr.

Table 9
Biomedical applications of *Bacillus* sp.

Proteins	Sources	Functions	Applications
1. Poly-γ-glutamic acid (γ-PGA) [43]	*B. anthracis* *B. natto* *B. subtilis* *B. thuringiensis*	Drug carrier or sustained release materials, curable biological adhesive and hemostatic or medical bonding kit suture thread	Use for gene therapy; cancer drug, substitutes for fibrin
2. Cytochrome P450 (CYP102) [54,55]	*B. megaterium*	Regulation of transcription and/or fatty acid hydroxylation, play a key part in various steps of primary and secondary metabolism, as well as in drug detoxification	Use in oxidation of a wide range of compounds including steroids, fatty acids and xerobiotics such as drugs and environmental Chemical
3. Uricase [56]	*B. fastidiosus*	Catalyse the oxidation of uric acid, a final product of purine catabolism, to allantoin which is more easily excreted than the starting compound	Use as therapeutic proteins
4. GlcK and AlaDH [58]	*B. stearothermophilus*	GlcK is used for measurement of glucose and creating kinase (CK) activity and AlaDH is used for measurement of γ-GT activity	Stabilize enzymatic reagents for measuring some components in biological fluids
5. Cry and Cyt [59]	*B. thuringiensis*	Exhibit cytocidal activity against human leukaemic T cells and uterus cervix cancer (HeLa) cells	Anti-cancer cell drug
6. BacP1 [60]	*B. subtilis*	Induce the synthesis of bactericidal antibodies reacting with P1 epitopes exposed on meningococcal surface	Use as meningococcal disease vaccine
7. Superoxide dismutase (SOD) [61]	*B. stearothermophilus*	Catalyze the dismutation od superoxide radicals to oxygen and hydrogen peroxide.	Anti-inflammatory drug and antiviral activities
8. Lipase [62]	*B. stearothermophilus*	Catalyze the esterification process	Synthesis of ester prodrug or or non-steroidal anti-inflammatory drugs such as ibuprofen and naproxen

Supachai Topanurak supported by the Thailand Research Fund (TRF), Thailand.

References

[1] F. Kunst, N. Ogasawara, I. Moszer, A.M. Albertini, G. Alloni, V. Azevedo, M.G. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S.C. Brignell, S. Born, S. Brouillet, C.V. Bruschi, B. Caldwell, V. Capuano, N.M. Carter, S.K. Choi, J.J. Codani, I.F. Connerton, A. Danchin et al., Nature 390 (1997) 249.

[2] H. Takami, K. Nakasone, Y. Takaki, G. Maeno, R. Sasaki, N. Masui, F. Fuji, C. Hirama, Y. Nakamura, N. Ogasawara, S. Kuhara, K. Horikoshi, Nucleic Acids Res. 28 (2000) 4317.

[3] F.G. Priest, M. Goodfellow, C. Todd, J. Gen. Microbiol. 134 (1988) 1847.

[4] H. Takami, K. Horikoshi, Extremophiles 4 (2000) 99.

[5] I. Moszer, FEBS Lett. 430 (1998) 28.

[6] M. Hecker, W. Schumann, U. Volker, Mol. Microbiol. 19 (1996) 417.

[7] P. Graumann, T.M. Wendrich, M.H.W. Weber, K. Schroder, M.A. Marahiel, Mol. Microbiol. 25 (1997) 741.

[8] H. Takami, K. Nakasone, C. Hirama, Y. Takaki, N. Masui, F. Fuji, Y. Nakamura, A. Inoue, Extremophiles 3 (1999) 21.

[9] J. Boch, B. Kempf, E. Bremer, J. Bacteriol. 178 (1996) 5121.
[10] J. Bernhardt, U. Volker, A. Volker, H. Antelmann, R. Schmid, H. Mach, M. Hecker, Microbiology 143 (1997) 999.
[11] J. Bernhardt, K. Buttner, C. Scharf, M. Hecker, Electrophoresis 20 (1999) 2225.
[12] H.L. Hyyrylainen, A. Bolhuis, E. Darmon, L. Muukkonen, P. Koski, M. Vitikainen, M. Sarvas, Z. Pragai, S. Bron, J.M. van Dijl, V.P. Kontinen, Mol. Microbiol. 41 (2001) 1159.
[13] A. Wipat, C.R. Harwood, FEMS Microbiol. Ecol. 28 (1999) 1.
[14] G. Schneider, Gene 237 (1999) 113.
[15] H. Antelmann, H. Tjalsma, B. Voigt, S. Ohlmeier, S. Bron, J.M. van Diji, M. Hecker, Genome Res. 11 (2001) 1484.
[16] P.W. Chu, M.N. Yap, C.Y. Wu, C.M. Huang, F.M. Pan, M.J. Tseng, S.T. Chen, Electrophoresis 21 (2000) 1740.
[17] B. Boonsinthai, S. Phutrakul, J. Sci. Fac. CMU 23 (1991) 9.
[18] B. Sookkheo, S. Sinchaikul, S. Phutrakul, S.T. Chen, Protein Expr. Purif. 20 (2000) 142.
[19] B. Sookkheo, S. Sinchaikul, H. Thannan, O. Thongprasong, S. Phutrakul, S.T. Chen, Proteomic (2002) in press.
[20] B. Fricke, T. Buchmann, S. Friebe, J. Chromatogr. A 715 (1995) 247.
[21] H. Ihara, T. Uemure, M. Masuhara, S. Ikawa, K. Sugimoto, A. Wadano, M. Himeno, Comp. Biochem. Phys. B 120 (1998) 197.
[22] M. Villalon, V. Vachon, R. Brousseau, J.L. Schwartz, R. Laprade, Biochim. Biophys. Acta 1368 (1998) 27.
[23] A. Helenius, M. Sarvas, K. Simons, Eur. J. Biochem. 116 (1981) 27.
[24] T.J. Stevens, I.T. Arkin, Proteins 39 (2000) 417.
[25] S. Ohlmeier, C. Scharf, M. Hecker, Electrophoresis 21 (2000) 3701.
[26] J.S. Lolkema, D.J. Slotboom, FEMS Microbiol. Rev. 22 (1998) 305.
[27] I. Diers, in: A.C.R. Dean, D.C. Ellwood, C.G.T. Evans, J. Melling (Eds.), Continuous Culture, Ellis Horwood, Chichester, 1976, p. 208.
[28] C.R. Goward, R. Hartwell, T. Atkinson, M.D. Scawen, Biochem. J. 237 (1986) 415.
[29] R.J. Sharp, M.D. Scawen, T. Atkinson, in: C.R. Harwood (Ed.), Biotechnology Handbooks, Vol. 2: Bacillus, Plenum, New York, 1989, p. 255.
[30] M.C. Oosthuizen, B. Steyn, D. Lindsay, V.S. Brozel, A. von Holy, FEMS Microbiol. Lett. 194 (2001) 47.
[31] K. Buttner, J. Bernhardt, C. Scharf, R. Schmid, U. Mader, C. Eymann, H. Antelmann, A. Volker, U. Volker, M. Hecker, Electrophoresis 22 (2001) 2908.
[32] R.D. Appel, A. Bairoch, J.C. Sanchez, J.R. Vargas, O. Golaz, C. Pasquali, D.F. Hochstrasser, Electrophoresis 17 (1996) 540.
[33] S. Sinchaikul, B. Sookkheo, S. Phutrakul, S.T. Chen, Proteomic (2001) submitted for publication.
[34] A.A. Gooley, N.H. Packer, The importance of protein co- and post-translational modifications in proteome projects, in: M.R. Wilkins, K.L. Williams, R.D. Appel, D.F. Hochstrasser (Eds.), Proteome Research: New Frontiers in Functional Genomics, Springer, Berlin, 1997, p. 65, Chapter 4.
[35] O. Lentz, Q.S. Li, U. Schwaneberg, S.L. Wahl, P. Fischer, R.D. Schmid, J. Mol. Catal. B 15 (2001) 123.
[36] I.A.D. Lessard, G.J. Domingo, A. Borges, R.N. Perham, Eur. J. Biochem. 258 (1998) 491.
[37] H.J. Schreier, C.A. Rostkowski, Gene 161 (1995) 51.
[38] S. Sakamoto, M. Seki, S. Nagata, H. Misono, J. Mol. Catal. B 12 (2001) 58.
[39] R.L. Tatusov, E.V. Koonin, D.J. Lipman, Science 278 (1997) 631.
[40] R.L. Tatusov, D.A. Natale, I.V. Garkavtsev, T.A. Tatusova, U.T. Shankavaram, B.S. Rao, B. Kiryutin, M.Y. Galperin, N.D. Fedorova, E.V. Koonin, Nucleic Acids Res. 29 (2001) 22.
[41] A. Bairoch, R. Apweiler, Nucleic Acids Res. 28 (2000) 45.
[42] H.F. Juan, C.L. Kao, C.Y. Kao, S.T. Chen, Bulk Gene Sercch System: a facile bioinformatic system to search and classify gene and protein functions from microarray, Bioinformatics (2002) submitted for publication.
[43] I.L. Shih, Y.T. Van, Biores. Technol. 79 (2001) 207.
[44] C. Li, D.F. Yu, A. Newman, F. Cabral, C. Stephens, N. Hunter, L. Milas, S. Wallace, Cancer Res. 58 (1998) 2404.
[45] C. Li, J.E. Price, L. Milas, N.R. Hunter, S. Ke, W. Tansey, C. Charnsagavej, S. Wallace, Clin. Cancer Res. 5 (1999) 891.
[46] M. Hashida, K. Akamutsu, M. Nishikawa, F. Yamashita, Y. Takakura, J. Control. Release 62 (1999) 253.
[47] Y. Otani, Y. Tabata, Y. Ikada, J. Biomed. Mater. Res. 31 (1996a) 157.
[48] Y. Otani, Y. Tabata, Y. Ikada, Biomaterials 19 (1996b) 1387.
[49] Y. Otani, Y. Tabata, Y. Ikada, Biomaterials 19 (1998a) 2167.
[50] Y. Otani, Y. Tabata, Y. Ikada, Biomaterials 19 (1998b) 2091.
[51] T. Sekine, T. Nakamura, Y. Shimizu, H. Ueda, K. Matsumoto, Y. Takimoto, T. Kiyotani, J. Biomed. Mater. Res. 35 (2000) 305.
[52] F.P. Guengerich, A. Parikh, Curr. Opin. Biotechnol. 8 (1997) 623.
[53] S. Rendic, F.J. Di Carlo, Drug. Metab. Rev. 29 (1997) 413.
[54] N.E. Hopkins, N. English, V. Hughes, C.W. Rowley, C.R. Wolf, W.L. Alworth, Biochem. Biophys. Res. Commun. 244 (1998) 868.
[55] N. English, V. Hughes, C.R. Wolf, Biochem. J. 316 (1996) 279.
[56] U. Schwaneberg, D. Appel, J. Schmitt, R.D. Schmid, J. Biotechnol. 84 (2000) 249.
[57] O. Schiavon, P. Caliceti, P. Ferruti, F.M. Veronese, Il Farmaco 55 (2000) 264.
[58] K. Tomita, K. Nomura, H. Kondo, K. Nagata, H. Tsubota, J. Pharm. Biomed. Anal. 13 (1995) 477.
[59] S. Yamashita, T. Akao, E. Mizuki, H. Saitoh, K. Higuchi, Y.S. Park, H.S. Kim, M. Ohba, Can. J. Microbiol. 46 (2000) 913.
[60] I.I. Heikkila, E. Wahlstrom, S. Muttilainen, M. Nurminen, H. Kayhty, M. Sarvas, P.H. Makela, Vaccine 14 (1996) 886.
[61] W.P. Michalski, J. Chromatogr. B 684 (1996) 59.
[62] S. Sinchaikul, B. Sookkheo, S. Phutrakul, F.M. Pan, S.T. Chen, Protein Expr. Purif. 22 (2001) 388.