5-2003

Determining the Influence of the Extracellular Proteinase from *Brevibacterium linens* on the Metabolism of *Lactococcus lactis* spp. *lactis* Using Functional Genomics

Yi Xie
Utah State University

DETERMINING THE INFLUENCE OF THE EXTRACELLULAR PROTEINASE FROM *BREVIBACTERIUM LINENS* ON THE METABOLISM OF *LACTOCOCCUS LACTIS* SPP. *LACTIS* USING FUNCTIONAL GENOMICS

by

Yi Xie

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Nutrition and Food Sciences

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

2003

ABSTRACT

Determining the Influence of the Extracellular Proteinase from *Brevibacterium linens* on the Metabolism of *Lactococcus lactis* spp. *lactis* Using Functional Genomics

by

Yi Xie, Doctor of Philosophy

Utah State University, 2003

Major Professor: Dr. Bart C. Weimer
Department: Nutrition and Food Sciences

Since the catabolism of amino acids in cheese results in the formation of most volatile flavor compounds, a proper intracellular pool of amino acids must be established in order to produce a desirable flavor production in cheese. Generation of this pool of amino acids requires complex interactions among casein and its derivatives, proteolytic enzymes, and transport systems in the associated bacteria, including lactococci. In this project, we hypothesized that casein hydrolysis by the extracellular proteinases of *Brevibacterium linens* BL2 modulates the expression profile of proteolytic related genes in *Lactococcus lactis* spp. *lactis* IL1403.

In order to monitor the global gene regulation patterns in *L. lactis* ssp. *lactis* IL1403, a high-throughput gene expression tool was needed to study the gene expression profiles on a genomic scale. In this project, we developed a novel oligonucleotide-based filter DNA array protocol for this purpose. The success of this oligonucleotide-based DNA array was dependent on technical innovations including polyI tailing, indirect high-

density biotin labeling, careful probe design, and integrated computational data analysis. The utility and validity of this protocol were demonstrated by profiling the expression of 375 metabolically related genes in *L. lactis* ssp. *lactis* IL1403 during heat, acid, and osmotic stresses.

Subsequently the DNA macroarray was used to profile the gene expression changes of *L. lactis* spp. *lactis* IL1403 growing in a peptide-limited medium, in a casitone-based peptide-rich medium, and in a casein hydrolyte by *B. linens* BL2 proteolytic enzymes. *L. lactis* ssp. *lactis* IL1403 experienced nitrogen starvation even with an abundance of peptide resources because of lack of expression of peptide transporter genes. Conversely, a peptide pool generated by *B. linens* BL2 proteolytic activities was sufficient to sustain the growth of *L. lactis* ssp. *lactis* IL1403. The repression of the peptide transporter and other peptidase genes of *L. lactis* ssp. *lactis* IL1403 was relieved in this medium. Interestingly, the Opt system, a di-tripeptide transporter, was used as a primary peptide transporter, instead of the Opp system whose genes were not actively transcripted in IL1403.

We also conducted additional experiments to further describe the protease in *B. linens* BL2 responsible for the peptide pool generation. This enzyme was secreted as a non-active zymogen and matured into the active protease. Both proteolysis and maturation processes were regulated. Collectively, this work demonstrated that a unique protease of *B. linens* BL2 generated a pool of peptides transportable by *L. lactis* IL1403 and induced changes in gene expression in *L. lactis* IL1403. Consequently, this body of work demonstrated the hypothesis to be true.

(195 pages)

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to Dr. Bart C. Weimer for his guidance, support, and patience. In addition, I also want to thank him for encouraging me to start the bioinformatics research while I was under his guidance.

I would like to acknowledge and thank my committee members, Drs. Stephen Clyde, Adele Cutler, Donald McMahon, and Gregory Podgorski, for their time and efforts in finalizing my dissertation and helping me graduate. I am especially thankful to Dr. Adele Cutler for her introducing me to the statistical computing world.

I want express my gratitude and sincere thanks to my college, Dr. Lan-szu Chou, for his support and useful discussion during this project. I also thank Paul Joseph for his assistance on peptide profiling on reverse phase HPLC.

Finally, I would like to express my love and many thanks to Ji Lu, my wife, and Adabelle Xie, my lovely daughter, who have had to endure these endless time of stresses and struggles during the project. Without their constant support and encouragement this project would never have been finished. Special thanks to my parents, Huaiqing Xie and Fengqing Zhang, for their encouragement and support. Thanks to the rest of my family and my wife's family for their support while I completed this endeavor.

Yi Xie

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, NOTATIONS, DEFINITIONS

Abbreviation key

AA = amino acid

AAA = aromatic amino acid

Ala = alanine

Arg = arginine

Asn = asparagine

Asp = asparatic acid

ATP = an adenosine-derived nucleotide, the primary energy form inside the cells

BCAA = branch chain amino acids

cDNA = complementary DNA

CDM = chemically defined medium

CEP = cell envelope proteinase

Cys = cysteine

dNTP = a mixture of dATP, dTTP, dCTP, and dGTP

EDTA = ethylenediamine tetraacetic acid

FWER = family-wise error rate

FDR = false discovery rate

Gln = glutamine

Glu = glutamine

Gly = glycine

His = histidine

HPLC = high-performance liquid chromatography

Ile = isoleucine

Leu = leucine

Lys = lysine

Met = methionine

MWCO = molecular weight cut-off

NaCl = sodium chloride

Opp = oligopeptide permease system

Opt = a di/tri-peptide transport system, annotated as oligopeptide permease system

ORF = open reading frame

PCR = polymerase chain reaction

Phe = phenylalanine

PMF = proton motive force

Pro = proline

PTS = phosphotransferase system

SAGE = serial analysis of gene expression

Ser = serine

SDS = sodium dodecyl sulfate

SDS-PAGE = sodium dodecyl sulfate polyacrylamide gel electrophoresis

SVM = support vector machine analysis

TFA = trifluoroacetic acid

THF = tetrahydrofuran

Thr = threonine

Trp = tryptophan

TSB = tryptic soy broth

Tyr = tyrosine

USU = Utah State University

Val = valine

CHAPTER 1

INTRODUCTION

In recent years, low-fat dairy products have become increasingly popular. The dairy industry responded by making low fat cheeses, but the reduction in fat content caused consumer acceptance problems with flavor and texture. Fat influences many chemical and physical properties of cheese (2), including flavor, body and texture, therefore lower fat often means a less flavorful, harder texture, bitter cheese. Flavor adjunct bacteria are successfully used in low fat Cheddar cheese to improve flavor and overcome some defects. *Brevibacterium linens* is one of the possible flavor adjunct bacteria (3). This organism produces extracellular proteinases that are thought to be one of the major contributing factors to the flavor and texture improvement in low fat cheese. However, there are limited studies focusing on the influence of exogenous proteinase on the cellular physiology of starter cultures because of the inability to fully explore the metabolic interaction between starter cultures and the flavor adjuncts.

Publication of the *Lactococcus lactis* spp. *lactis* IL1403 genome marks a paradigm shift in lactic acid bacteria research (1). Instead of focusing on a single gene or single cellular subsystem, it now is possible to study gene functions in across the full genome. In this proposal, we investigated the characteristics of extracellular proteinases from *Brevibacterium linens* BL2 and their influence on the global gene expression profile of *Lactococcus lactis* spp. *lactis* IL1403.

References

1. **Bolotin, A., P. Wincker, S. Mauger, O. Jaillon, K. Malarme, J. Weissenbach, S. D. Ehrlich, and A. Sorokin.** 2001. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. Genome Res. **11:**731-53.

2. **Simard, R. E.** 1991. Evaluation of low fat cheese problems. p. 37-41. *In* Center for Dairy Research (ed.). Cheese Research and Technology Conference, Madison, Wisconsin.

3. **Weimer, B., B. Dias, U. Madhavi, J. Broadbent, C. Brennand, J. Jaegi, M. Johnson, F. Milani, J. Steele, and D. V. Sission.** 1997. Influence of NaCl and pH on intracellular enzymes that influence Cheddar cheese ripening. Le Lait. **77:**383.

CHAPTER 2

LITERATURE REVIEW

Scientific investigation of cheese flavor began during the early part of 20th century. Initially, it was believed that cheese flavor might be due to a single compound (60). However, the "component balance theory" has become more and more popular. This theory proposes that cheese flavor is due to the correct balance and concentration of a wide range of sapid and aromatic compounds (2). During the last 100 years, extensive research on the flavors of several cheese varieties has been done.

Cheese flavor is inseparable from the biochemistry and microbiology of cheese ripening. There are at least 300 different volatile and nonvolatile compounds implicated in cheese flavor (86). All of these compounds originate from the action of indigenous and added enzymes in milk components and bacteria in the starter and non-starter organisms. The flavor compound production in cheese essentially involves three major biochemical processes: proteolysis, lipolysis and carbohydrate metabolism. Proteolysis is very important among these processes because it contributes to cheese flavor and texture in mature cheese (32).

General Aspects of Proteolysis in Cheese

Proteolysis in hard cheeses starts with hydrolysis of κ-casein by the action of chymosin (E.C. 3.4.23.4). This destabilizes the casein micelle and initiates the coagulation of milk. After the initial cheese making, residual coagulant and native milk proteinases such as plasmin, and perhaps cathepsin D (E.C. 3.4.23.5) are responsible for

the further casein hydrolysis. This results in the formation of large and intermediate-sized peptides (33). These peptides are subsequently degraded to small peptides by residual coagulant and proteolytic enzymes from the starter and non-starter microflora (32). Some small peptides may be hydrolyzed further to free amino acids by aminopeptidases released from lysed bacteria. More commonly, the small peptides are transported into the cell via various transport systems and digested into free amino acids and metabolized inside the bacteria (32, 50, 89) (Figure 2-1).

This general outline of proteolysis varies substantially among different cheese varieties due to differences in cheese manufacturing practices. Even in the same cheese varieties with different flavor adjunct bacteria, proteolysis can be significantly different (33). In Mozzarella, Swiss and other high-cook temperature varieties, the coagulant is almost completely denatured by the high cooking temperature. The contribution of plasmin to the initial hydrolysis of caseins in these cheese varieties is more pronounced in Cheddar and Dutch cheese varieties. In mold or bacterial surface-ripened cheese varieties, the enzymes from adjunct starters strongly influence proteolysis due to their strong extracellular proteinase and peptidase activity.

Flavor Compound Formation and Proteolysis

Intact casein has little or no taste. Intermediate-sized peptides also have no major contribution to cheese flavor, but short peptides can have important flavor characteristics, such as bitterness (62). The formation of bitter peptides during proteolysis of casein molecules is well-studied (1). Casein micelles conceal the majority of hydrophobic side chains, therefore they have no taste because of the lack of interaction with the taste buds.

When caseins are degraded by proteinases, peptides with different size and hydrophobicity are formed. Larger peptides are still able to mask their hydrophobic side chains to some extent by hydrophobic interaction, where U-shaped peptides or clusters of peptides are formed. But for the small peptides with short length (20 amino acid residue) and high hydrophobicity (Q-value > 1,300 cal/mol), they become bitter because of the inability to hide their internal hydrophobic amino acid side chains (62). For example, bitter peptide β-casein f193-209, a standard used to measure bitterness, contains eleven hydrophobic amino acid residues and has a Q-value of 1,760 cal/mol (47).

The activity and specificity of the proteinases can change dramatically in the cheese-ripening environments, compared with their optimal condition. For example, cleaving the Leu_{192}-Tyr_{193} bond in β-casein appears to be very sensitive to NaCl concentration. In cheese-like conditions (4% NaCl, pH5.2, and no sugar), the rate of β-casein cleavage into β-I-casein (f1-192) and f193-209 by chymosin decreases due to the inaccessibility of hydrophobic cleavage sites at this specific environment (47).

Free amino acids have little or no direct contribution to cheese flavor. It is the steps of further catabolism of amino acids that result in the formation of most volatile flavor compounds (86). The first stage in amino acid catabolism generally involves decarboxylation, deamination, transamination, desulphuration, or perhaps hydrolysis of the amino acid side chains. The second stage involves conversion of the resulting compounds (amines and α-keto acids), as well as amino acids themselves, to aldehydes, primarily by the action of deaminases on amines. The final level of amino acid catabolism is the reduction of the aldehydes to alcohols, or their oxidation to acids. Sulfur-containing amino acids undergo extensive conversion, leading to the formation of

a number of compounds, including methanethiol and other sulfur derivatives that provide desirable flavor in Cheddar cheese (94).

Chemical modification reactions on amino acid side-chain also occur in cheese. Hydrolysis can release ammonia from Asn or Gln or by the partial hydrolysis of the guanidino group from Arg. In addition to ammonia, some off-flavor compounds can also be produced during the side chain modification of tyrosine, phenylalanine and tryptophan (21, 37, 85).

Proteolysis and its products have little direct contribution to flavor formation in cheese, except bitterness. But as an upstream process, the specificity of proteolysis has a direct influence on the composition of peptide pool available for bacterial uptake system in the cheese matrix. Hence, the specificity of proteolysis can indirectly alter the internal amino acid pool inside of starter and non-starter cultures, which in turn have an impact on the generation of flavor compounds during cheese ripening.

Proteolytic Enzymes in Cheese Production

Proteolytic enzymes present during the cheese making process commonly come from the following sources: milk, coagulant, starter bacteria, non-starter lactic acid bacteria and adjunct microorganisms. They can also originate from the addition of exogenous proteinases. The following is a discussion of the proteolytic enzymes involved in cheese production and aging.

Proteinases from Coagulant and Indigenous Proteinases

Chymosin (E.C. 3.4.23.4), the major proteinase in traditional rennet used for cheese making, is an aspartyl proteinase (two aspartic acid residues in the active site) of gastric origin, secreted by the neonates of various mammalian species. The principal role of chymosin in cheese making is to coagulate the milk by specifically cleaving the micelle-stabilizing protein, κ-casein at Phe_{105}-Met_{106}. About 6% of the chymosin added to cheese milk is retained in the curd of Cheddar and similar cheeses (26). Residual chymosin plays a major role in the initial proteolysis in cheese (32). Although chymosin is quite specific for κ-casein's Phe_{105}-Met_{106}, it is also active on α_{s1}-, and β-casein. The primary site of chymosin action on α_{s1}- and β- casein are Leu_{192}-Tyr_{193} and Phe_{23}-Phe_{24}, respectively, although there is evidence that the sites may not be available in high salt environment, such as cheese-like conditions (47). α_{s2}-Casein appears to be relatively resistant to proteolysis by chymosin.

Plasmin (fibrinolysin, E.C. 3.4.21.7), the principal indigenous proteinase in milk, is a trypsin-like proteinase with pH optimum at about 7.5 and is very specific for peptide bonds involving lysyl residue (95). In milk, plasmin hydrolyzes β- and α_{s2}-casein quite easily, but α_{s1}- and κ-casein are relatively resistant to plasmin activity (26). Under normal circumstances, the proteolysis of milk protein is prevented by the presence of complex plasmin inhibitor networks. During cheese making, most plasmin inhibitors are lost in the whey, thereby relieving the inhibition of plasmin (26). Although plasmin is optimally active in alkaline pH, it has an important role in β-casein degradation in acidic pH environments such as cheese ripening (26) and may be used to accelerate cheese

ripening (27). The advantage of using plasmin as a cheese ripening accelerant is that since it is indigenous to milk, plasmin is evenly distributed throughout the milk. Plasmin has a narrow specificity for casein, which limits the possible detrimental flavors generated from casein hydrolysis by other proteolytic enzymes.

Cathepsin D (E.C. 3.4.23.5), another indigenous proteinase in bovine milk, is an aspartic proteinase and active in acidic pH. It exists in four different forms, including pro-, pseudo-, single-chained, and two-chained cathepsin D with procathepsin D being the dominant form (52). Most of the cleavage sites for cathepsin D and chymosin are identical. It is interesting to note that cleavage of the Leu_{192}-Tyr_{193} bond of β-casein by both enzymes produces the C-terminal bitter peptide Tyr_{193}-Val_{209}. This bitter peptide cannot be cleaved by chymosin, but can be further hydrolyzed by cathepsin D at Phe_{205}-Pro_{206} (51). Procathepsin D is considered to be a whey protein, and therefore it is expected that only a very small amount of this proteinase is incorporated into cheese. This limits its use in cheese making.

Proteinases from coagulant and milk itself have great potential to be used in accelerating cheese ripening. They generally have a narrow specificity toward casein, well-controlled proteolytic activity, and an even distribution throughout the milk. On the other hand, they are expensive for industrial usage due to purification cost and they are difficult to retain in the cheese curd during the manufacturing processes.

Proteolytic and Peptidolytic Enzymes from Starter Culture

Lactococci, the starter culture used in hard cheese manufacture, are fastidious organisms with multiple amino acid auxotrophies, and as a result their growth is

dependent on an efficient system of protein degradation and transport of amino acids and small peptides. Generally, the proteolysis system in lactococci includes i) a cell envelope proteinase (CEP or lactocepin, EC 3.4.21.96), ii) transport systems for amino acids and peptides, and iii) intracellular peptidases (50).

The CEP is the first enzyme of the pathway of casein degradation in lactococcal bacteria. This enzyme is an exocellular proteinase with its C-terminal anchored to the cell membrane and its catalytic domains exposed outside the cell wall (50). As a key enzyme in proteolytic degradation in cheese, it has been studied extensively, both genetically and biochemically. There is also evidence that the specificity of CEP is closely linked to the accumulation of bitter peptides in cheese matrix (8).

Genetic studies show that the CEP proteinase and related genes in *L. lactis* are located on plasmids ranging in size from 14 to 100 kb. Proteinase gene organization in *L. lactis* is highly conserved and consisted of two oppositely transcribed genes (87). The smaller open reading frame is the *prtM* gene that encodes a maturation protein, while the larger open reading frame is *prtP*, the structural gene for the proteinase. The *prtP* and *prtM* genes are separated by a 0.3 kb AT-rich promoter region. In the case of the *L. lactis* ssp. *cremoris* Wg2 and SK11 genes, the promoters of *prtM* and *prtP* overlap and transcription initiation occurs at adjacent sequences in a region with an unusual rotational symmetry. Active proteinase production in *L. lactis* is dependent on the simultaneous expression of *prtM* and *prtP* genes (48).

The nucleotide sequences of the *prtP* genes from different lactococcal strains show a remarkably high degree of similarity that exceeds 95%. PrtP proteins are synthesized as the pre-pro form and composed of eight domains (78) (Figure 2-2). The pre-

propeptide (PP domain in Figure 2-2) has 188 amino acid residues and includes a signal peptide (pre-peptide) and maturation peptide (pro-peptide). In *L. lactis*, pre-peptide includes a sec-dependent 33 residue signal sequence and a consensus signal peptidase I cleavage site. A number of studies demonstrate that this signal peptide is sufficient to direct the secretion of heterogeneous proteins in *L. lactis* (15). The maturation of PrtP proteinase relies on the action of the membrane-anchored protein PrtM (38). During or after the secretion of PrtP, maturation protein PrtM acts as intermolecular chaperone and helps pro-PrtP to fold properly. PrtP cleaves its own pro-peptide region to become active (39).

The catalytic domain is at the N-terminal part of the mature proteinase. It has approximately 500 residues, and shows significant similarity to serine proteinases of the subtilisin family, also known as subtilases (79) (PR domain in Figure 2-2). The *L. lactis* spp. *cremoris* SK11 proteinase has four essential residues Asp_{30}, His_{94}, Ser_{33}, and Asn_{196} in its active site. Based on sequence alignment and comparison of the three-dimensional structure of subtilisin, residues at 131, 138, 142, 144, and 166 of SK11 are part of the substrate-binding pocket of this proteinase. In support of this view, substitutions at these residues change the specificity of the proteinase in lactococci (Table 2-1).

Between the catalytic domain and the C-terminal spacer region, there is another region with approximately 1000 residues that is characteristic for the proteinases of lactic acid bacteria and not found in other subtilases (A and B domains in Figure 2-2). The function of this region is not entirely known although residues 747 and 748 in A-domain is involved in substrate binding, possibly in electrostatic interactions with substrate caseins (Table 2-1).

The C-terminus of the lactococcal proteinase contains a conserved membrane anchor sequence that is thought to act as a stop-transfer sequence (AN domain in Figure 2-2). This membrane anchor sequence is highly conserved in all exo-proteins of Gram-positive bacteria and consists of a conserved hexapeptide with the sequence of Leu-Pro-X-Thr-Gly-Glu preceding a hydrophobic, membrane-spanning segment of about 18 residues followed by a 7-amino-acid charged tail (29). The spacer domains (H and W domains in Figure 2-2) are located just preceding membrane anchor AN domain. Its function is to cross the cell wall and expose the catalytic domain of the proteinase to the outside of the lactic acid bacteria.

The specificity of lactococcal protease has an important implication in cheese manufacture. It dictates not only the potential for bitter peptide generation, but also the composition of extracellular peptide pool in the cheese matrix. Lactococcal proteinases were initially divided into P_I and P_{III} specificity classes on the basis of their caseinolytic specificity. Later, it was found that the proteinase types can be further classified into seven specificity groups, *a-h* (25, 50). This new classification system is based on the peptide profiles generated during the hydrolysis of α_{s1}-casein fraction f1-23 (Table 2-1). These differences in casein or peptide hydrolysis are also reflected in single residue polymorphism in the primary sequences of the proteinases (Table 2-1). The specificity of PrtP proteinase linked directly to the bitterness problem in cheese. In reduced-fat Cheddar cheeses, bitterness scores of cheese made with starter containing PrtP proteinases are significant higher than cheese made with proteinase-negative starter. Among three proteinase specificity types examined, starter with proteinase belonging to group *h* produced the bitterest low-fat cheese (8).

Except for the proteinase PrtP and the endopeptidase NisP, all other lactococcal proteolytic-related components are membrane or intracellular proteins (50). It appears that PrtP is the only extracellular enzyme participating in the degradation of casein to transportable-size peptide fragments. However, limited lysis of cells could also release intracellular peptidases that may participate in the extracellular proteolysis process. After casein degradation, lactococcal peptide transport systems are responsible for peptide uptake (50). Only a small fraction of peptides generated by PrtP are actually used probably due to the specificity of Opp system. Casein hydrolysis is not a rate-limiting step for the lactococcal proteolytic system (49).

The most important peptide transport system in lactococci is the oligopeptide transport system (Opp system). The Opp system translocates oligopeptides from 4-18 amino acid residues into growing cells at the expense of ATP (18). Although most of the peptides derived from casein hydrolysis are transported into the cell via Opp system, a small portion of them are transported using the di/tri-peptide transport systems (DtpT). The proton motive force (PMF) is the primary driving force for DtpT peptide transport systems in lactococci.

Three different types of amino acid transport systems operate in lactococci (50): i) The PMF-driven transporter couples amino acid transport to the proton motive force. ii) The phosphate-bond-linked amino acid transporter is driven by the high-energy phosphate bond of ATP or an ATP-derived metabolite. iii) The exchange transporter is powered by the exchange between concentration gradients of two different amino acids (Table 2-2).

Once di/tri- peptides and oligopeptides are translocated inside the cell, the intracellular peptidases hydrolyze them into amino acids. Recently, comprehensive transcriptional profiling of the 16 genes in the proteolytic system of *L. lactis* ssp. *cremoris* MG1363 have been examined (36). This study reveals that a rich peptide pool severely represses the transcription of some of proteolytic genes, including peptidase genes *pepC*, *pepN*, *pepX*, and *pepO1*, but not other peptidase genes. This indicates that *pepC*, *pepN*, *pepX*, and *pepO1* in lactococci are likely involved in peptidolysis of translocated peptides. The other peptidase genes, such as *pepF2*, *pepM*, *pepP*, and *pepT*, are not regulated by the peptide source and more likely involved in cellular functions, such as protein maturation and cellular peptide recycling.

Although the exocellular proteinase PrtP alone is able to sustain growth and survival of *L. lactis* in milk and cheese, extracellular proteinases with different specificities can supply additional transportable peptides. Changing the composition of peptide and amino acid pools inside the cheese matrix by other proteinase can accelerate cheese-ripening (33).

Proteinases from Starter Adjunct Bacteria and Other Sources

Proteinases from Bacillus subtilis

The neutral proteinase from *Bacillus subtilis* has been studied for its ability to hydrolyze casein and accelerate Cheddar cheese ripening (33). The bacterial neutral proteinases have relatively low thermostability when the pH is moved only modestly away from the optimum. This allows for increased control of proteolysis during cheese

ripening. Proteinase N (Amano Pharmaceutical), a neutral proteinase, exhibits high affinity for cleavage at hydrophobic amino acids, and can generate hydrolysates with significantly lower bitterness.

In cheese ripening studies, Neutrase, another commercial neutral proteinase from *B. subtilis*, is also used extensively. Compared to acid and alkaline proteinases, cheese made with adding Neutrase has enhanced flavor development at a low enzyme level, but caused bitterness at higher enzyme levels (54). By combining Neutrase with lactococcal cell-free extract, containing mainly aminopeptidases, proteolysis is enhanced significantly, but the cheese flavor improvement is still not satisfied (55). On the other hand, when cheese is made with lactococci containing a plasmid carrying the *B. subtilis* neutral proteinase gene, it shows accelerated ripening (even though the flavor is not preferred) and judged to be >1 year old (61).

Extracellular Proteinase from
Brevibacterium linens

Flavor adjuncts are microorganisms that are added to the cheese matrix to enhance the flavor of the final product. *Brevibacterium linens* is a starter culture used in the production of smear-ripened cheese varieties (46). Addition of *B. linens* to low fat Cheddar cheese can improve cheese flavor and texture without causing defects that are normally observed with the use of exogenous proteinase mixture (80, 92). With *B. linens* as an adjunct, low fat Cheddar cheese is found to be more flavorful and softer compared to cheese made with *Lactobacillus helveticus* and *Lactobacillus casei* (92). This suggests that there are some unique properties of brevibacteria that contribute a "Cheddar-type" flavor to low fat Cheddar cheese. In addition to improving cheese flavor, there is also

evidence that the low fat cheese made with *B. linens* does not increase bitter peptide accumulation in low fat cheese comparing other adjunct bacteria. This indicated that proteolytic activities from *B. linens* played some important roles during cheese ripening.

Proteinase production, isolation, and characterization in *B. linens* were studied by a number of investigators (34, 41, 45, 83). It is shown that intact casein in the growth medium favors production of the extracellular proteinase while glucose hinders its production (34). In a comparison of seven different media, *B. linens* was found to produce the highest proteinase activity in a medium composed of peptone, yeast extract, NaCl, K_2HPO_4, and casein (83).

Although it is agreed that the extracellular proteinase of *B. linens* is a serine proteinase, there are considerable differences existing in other properties. For example, pH and temperature optima are reported that range from 5.0 to 11.0 and 25 to 55°C, respectively (34, 41, 45, 83). Only one extracellular proteinase from *B. linens* ATCC 9174 has been purified into uniform. The purified proteinase has pH and temperature optima of 8.5 and 50°C, respectively, and a molecular weight of 56 kDa as determined by SDS-PAGE and 126 kDa by gel filtration (66). This indicates that the native enzyme exists as a dimer. Mg^{2+} and Ca^{2+} enhance the activity of proteinase, while Hg^{2+}, Fe^{2+}, and Zn^{2+} cause strong inhibition.

The specificity of extracellular proteinase from *B. linens* ATCC 9174 on α_{s1}- and β-casein has been studied in details (67, 68). This proteinase has a broad specificity at P1 and P1′ positions similar to lactococcal CEP (Figure 2-3). On P2, P3, P4 and P2′, P3′, and P4′ positions (Figure 2-3), the *B. linens* proteinase shows a preference for hydrophobic residues. This is different from lactococcal CEP, which has a preference for

charged residues at these positions (24). The different preferences in substrate binding are also reflected in α_{s1}- and β-casein hydrolysis patterns of these two enzymes. Among the eighteen major cleavage sites on α_{s1}-casein by the extracellular proteinase from *B. linens* ATCC 9174 (67), only His_8-Gln_9, Phe_{23}-Phe_{24}, Phe_{150}-Arg_{151}, Ser_{161}-Gly_{162}, and Leu_{169}-Gly_{170} are shared by some or all types of lactococcal CEP (50). On the other hand, among the fourteen major cleavage sites on β-casein by the extracellular proteinase from *B. linens* ATCC 9174 (68), only Ser_{18}-Ser_{19}, Gln_{72}-Asn_{73}, Leu_{139}-Leu_{140}, and Ser_{142}-Trp_{143} are not common to some or all types of lactococcal CEP (50). Although the best substrate for the extracellular proteinase is casein, it also shows some activity towards hemoglobin and albumin (83).

The *B. linens* proteinase is produced in a cyclic fashion during cell growth (91). The dimeric nature of the proteinase (66) may allow allosteric control in the cheese matrix and potentially limit excessive proteolysis during cheese ripening. This is not possible for monomer proteinase like lactococcal CEP. Indeed, Cheddar cheese made with *B. linens* proteinases became over-ripened, because the enzyme was added independently without a controlling mechanism so that proteolysis became unmanageable(42, 43).

Unlike lactococci, several extracellular aminopeptidases exist in *B. linens* (31). Combined with a commercial neutral proteinase, aminopeptidases from *B. linens* have successfully accelerated Cheddar cheese ripening without production of bitter flavors (42).

Proteinases from Other Sources

There are many other bacterial and fungal proteinases used in cheese flavor improvement and acceleration of cheese ripening (33). In Cheddar cheese-making trials with an added proteinase-peptidase mixture from a *Pseudomonas* (53), a low level of enzyme addition accelerated flavor development, while larger amounts caused bitterness and other off-flavors. Fungal proteinases, such as Proteinase 11 from *Aspergillus oryzae* give detectable flavor enhancement, but a high level of bitter peptides were produced with increasing enzyme levels (28).

Although the relationship between proteolysis and flavor in cheese has been studied extensively, there is still no model to clearly describe this relationship at a whole system level. Most of the current studies focus on a subsystem or a single component in this complex system. Recent publication of the *L. lactis* spp. *lactis* IL1403 genome gives us a new opportunity to understand the role of proteolysis in flavor formation at the whole system level (7). With function genomics and high-throughput tools like DNA arrays, we are able to study, for the first time, how the extracellular proteinases with different specificity influence the status of genetic network of proteolytic system and metabolic influx in the starter culture. This will illuminate how we can engineer to enhance flavor production during cheese manufacturing (93).

Functional Genomics

Since the first autonomous living organism, *Haemophilus influenzae*, was sequenced in 1995 (30), large-scale genome sequencing projects have progressed at an explosive rate. The publication of the draft version of the human genome marks the

beginning of the post-genomic era in biological research. To date, there are about 510 genome sequencing projects around the world funded by both public and private sources, and more than half of them focus on prokaryotic organisms (4).

Lactococcus lactis spp. *lactis* IL1403 was the first lactic acid bacterium whose genome has been sequenced (6). Its physical and genetic maps were established as early as 1992 (56). The first draft version of genome sequences from this bacteria was published in 1999 (6), and a fully annotated genome finally became available to the public in 2001 (7). In the genome of *L. lactis* spp. *lactis* IL1403, there are 2,310 open reading frames (ORF), more than half of them (1,482 ORFs) with known functions assigned (7). In addition to *L. lactis*, the Lactic Acid Bacteria Genome Consortium (LABGC) recently initiated a systematic study of the genomes from a full range of lactic acid bacteria, including lactococci, lactobacilli, oenococci, and bifidobacteria (93).

Introduction of Functional Genomics

The complete sequence of a genome does not necessarily lead to a clear picture of the biological activity of organisms. This is mainly because the molecular function of many computationally predicted genes needs to be confirmed experimentally. For example, about 94 % (2,162 ORFs) of *L. lactis* spp. *lactis* IL1403 predicted ORFs are experimentally uncharacterized, except by homology searches for potential orthologues. Therefore, there is an urgent need to develop high-throughput tools that use complete genome sequences to characterize functions for genes and their products. The studies that aim at characterizing genes or their protein functions on a genome-wide scale are collectively called "functional genomics."

There are several strategies currently available for conducting functional genomics studies, targeting different levels of the genome and its products, such as mRNA, protein, protein function and regulatory networks (Figure 2-4). One of the most popular functional genomic tools is the DNA array or gene chip. This technique monitors mRNA expression in a particular organism or a tissue and analyzes the variation of gene expression profile during development or under particular experimental condition (59, 76). In lactic acid bacteria research, researchers have applied a small-scale DNA array during the investigation of arginine and sugar metabolism in *L. lactis* spp. *lactis* ML3 and IL1403 (12, 23).

DNA Array Technology

One of the most difficult challenges in research on the genome scale is the sheer amount of information stored in the genome sequence (93). The biological information encoded in the genome is far more complex than any physical or artificial system encountered before. It is this complexity that makes traditional tools of experimental biology obsolete. These traditional tools have to be applied in parallel on a large scale to increase the throughput and meet the challenge of genomic research.

Along with serial analysis of gene expression (SAGE) (88), the DNA array is one of the most powerful and versatile tools designed specifically for functional genomics (76). This technique is a high-throughput version of the reverse dot blot, where each DNA probe is immobilized at known coordinates on the solid media, and then labeled targets are hybridized on to them. The hybridization signal provides information on the presence and abundance of mRNA species in the original sample (Figure 2-5). The strength of

DNA arrays lies in its high throughput nature. In a single experiment, DNA arrays profile the expression of thousands of genes in a particular organism or tissue in response to a variety of conditions (77). In addition to gene expression profiling, DNA arrays have also been used in comparative genomic studies to identify conserved and non-conserved genes, such as pathogenic-related genes in *Helicobacter pylori* (73).

Overview of DNA Arrays

The nomenclature of DNA arrays is still evolving and therefore can be very confusing. Most people use terminology like DNA microarray, gene chip or oligo array interchangeably, although they all have specific meanings. For example, gene chip is a trademark for the Affymetrix DNA microarray, and should not be used as a general term to describe other DNA arrays. The next section is an overview of the major differences between DNA arrays.

Pitch size. DNA arrays can be classified into micro- and macroarrays based on the pitch size (i.e., the horizontal and vertical distance between two spots in the array). On a glass-based DNA microarray, each spot normally has a pitch size smaller than a micron with overall spot density as high as 10,000 features/cm^2. Affymetrix, the leading commercial vendor in DNA microarray technology, can achieve an even higher density with photolithography and combinatorial chemistry techniques (58). Currently, Affymetrix's newest GeneChip®, Human Genome U133, has 1,000,000 unique oligonucleotides immobilized on two 1 cm^2 silicon surfaces (www.affymetrix.com).

DNA macroarrays have larger pitch sizes around 1-2 mm with low to medium feature density. Most DNA macroarrays have less than 10,000 features. It should be

noted that this density usually sufficient to achieve single gene resolution for the majority of prokaryotic organisms. For example, *E. coli* K12 has 4,288 ORFs (5), which means that for single probe per gene resolution, a DNA array will require less than 5,000 features. In comparison to microarrays, DNA macroarrays tend to be quite large, which in fact is an advantage for signal detection. The data acquisition step in a DNA microarray normally requires specialized instruments, while for DNA macroarrays, an imager or a standard desktop scanner is usually sufficient.

Probe types. The DNA molecules attached to the surface are known as the probes. According to the type of probes in each feature, DNA arrays can be classified as oligonucleotide-based or complementary DNA (cDNA)-based arrays.

Oligonucleotide-based DNA arrays have many advantages over cDNA-based arrays. In cDNA arrays, cross-hybridization between genes with high sequence similarity, such as paralogues, is problematic. It can lead to false positive results that severely limit the biological relevance of cDNA arrays. For example, the expression levels of the alcohol dehydrogenase genes 1 and 2 (adh1 and adh2) can be differentiated using oligonucleotide-based arrays (Motorola Codelink™ Expression chip) (19). But the expression profiles of these genes are indistinguishable in cDNA arrays because of the high level of sequence homology (88%) (17). With careful attention to oligonucleotide probe design, cross-hybridization can be reduced or even eliminated in oligonucleotide-based arrays. Adoption of an oligonucleotide as the probe also makes it possible to normalize probe selection criteria so that more stringent hybridization conditions can be used to further reduce possible false positive signals. Oligonucleotide probes also avoid other problems with cDNA production during cDNA array fabrication, such as no PCR

product (i.e. no probe), or contaminated PCR products (i.e. wrong probe). These issues require extensive quality control steps and usually involve in the high costs associated with sequence verification.

Immobilization media. DNA arrays can be made on glass slides, nylon membranes, silicon, or other solid media. Combined with a fluorescent detection system, glass slides are the most popular medium used in DNA array fabrication. The medium selection has a great impact on the immobilization chemistry and detection method because of different physical and chemical properties of the surface. For example, porous immobilization media like nylon membranes have higher DNA binding capacities than a flat-surface glass slides. On the other hand, probe droplets during array printing tend to diffuse on a nylon membrane, but not on a glass slide due to the surface tension of glass slides.

Array use. The use of DNA array involves several steps, starting with total RNA or polyA RNA extraction, followed by labeling, hybridization, and detection (Figure 2-5). Total RNA extraction starts with cell lysis to release RNA and other cellular components, including RNase, an enzyme that degrades RNA molecules. To preserve cellular messager RNA (mRNA), it is extremely important to prevent RNA degradation during the extraction process. This can be achieved by treatment with strong protein denaturants, such as guanidinium thiocynate or hot phenol solutions. By relying on unique physical and chemical properties of RNA, total RNA can be easily separated from other cellular components via filter-based (RNeasy® from Qiagen Inc.) or solvent partitioning (Trizol® from Life Technologies).

Because the majority of total RNA (~98%) is ribosomal RNA (rRNA), mRNA enrichment leads to higher sensitivity because of less interference. In the eukaryotic cell, mRNA molecules have poly(A) tails and can be enriched with binding to oligo(dT) affinity resin, such as oligo(dT) coated magnetic beads. Most bacterial mRNA lacks a poly(A) tail, its enrichment has to be achieved with alternative strategies. For example, bacterial mRNA can be enriched by removal of the ribosomal RNA. This involves reverse transcription of rRNA into DNA with rRNA-specific primers, followed by digestion of rRNA:DNA hybrid molecules with RNaseH and DNase I (70). It is also possible to polyadenylate mRNA molecules with sequence-independent poly(A) polymerase I to add the poly(A) tails artificially, and then remove the rRNA molecules with oligo(dT) chromatography (96).

Once isolated, the target mRNA molecules need to be labeled with fluorescence dyes or other labels such as biotin before hybridization and detection (Figure 2-5). mRNA molecules can be labeled directly using chemical (44) or enzymatic (70) methods, or indirectly by converting them to cDNAs via reverse transcription (RT). The latter method is more common, because cDNA is more stable and therefore easy to handle than mRNA targets, although direct-labeled mRNA is more representative of the abundance of RNA species in the original sample (44).

In RT-based labeling reactions, oligo(dT)s are commonly used as primers for eukaryotic mRNA, while either random primers or a genome-specific primer pool can be used in prokaryotic RNA labeling. It is often argued that specific primer sets are better for priming reverse transcription because random primers have a stronger tendency for hot-spot priming and bias for longer RNA transcripts (16). An alternative indirect two-

step RT labeling method recently is gaining popularity (101). This method involves reverse transcription to incorporate primary amino groups into cDNA molecules, followed by dye-coupling a NHS-ester to the functional amino groups. This methods has many advantages over traditional direct RT labeling procedures, including high labeling efficiency, low cost and no dye bias (100).

Hybridization refers to a chemical process where the target nucleic acid molecules form stable hybrids with probes fixed on the immobilization media. Hybrid formation is dependent on the sequences of both targets and probes. Ideally, only completely complementary targets and probes will form a stable hybrid. In practice, target and probe with some mismatches can hybridize. The specificity of target and probe hybrid formation can be modulated with different hybridization and washing conditions. Hybridization with DNA macroarrays on nylon filters can be done either in polyethylene bags or in roller bottles inside hybridization ovens with a large hybridization volume (>1 ml). The majority of protocols published for DNA microarrays require much less hybridization solution (~10 μl). In this setup, the microarray is covered with a coverslip, and the hybridization solution forms a thin film over the printed probes. The primary advantage of micro-volume hybridization is the high target concentration, which leads to high sensitivity and fast hybridization kinetics.

The standard detection method with a glass slide DNA microarray is fluorescence-based. The images are scanned with a fluorescence laser scanning device, such as a confocal microscope (76) (Figure 2-5). This detection method has several advantages, such as a low background, short scanning time, and possibility of using multi-color fluorescence (13). Due to their high background fluorescence, nylon membranes are not

suitable for use with fluorescence-based detection. Instead, other types of detection methods, such as radioactive (90), chemiluminescent (64) or colorimetric (11) based methods, are used. Irrespective of the detection chemistry, after detection the resulting signals are recorded and acquired as intensity-dependent grayscale images. The intensities were extracted from each spot on the image and used to assign the biological activity for the correspondent genes.

Computational Tasks for DNA Array Design and Analysis

Design and analysis of DNA arrays are very complex tasks because of the massive amount of data and other constraints associated, such as noise and signal non-linearity. The success of DNA array technology requires the integration of molecular biology, computer science, statistics, and robotics. There are several key computational-related problems that exist in DNA array design and data analysis (63, 81), such as large-scale oligonucleotide primer or probe design, expression data normalization and statistical analysis.

Probe design. The oligonucleotide sequences have to be carefully designed and selected before array fabrication (57, 71). An effective oligonucleotide probe designed from a specific gene sequence has to meet the following three conditions: uniqueness, usefulness, and uniformity (the 3U Principle) (97). Uniqueness means that a probe must have only one complementary sequence in the genome, so that the potential cross-hybridization will be reduced or eliminated. The usefulness requirement dictates that the oligonucleotide must not have strong tendencies to form intra-molecular secondary structures, like hairpins, or inter-molecular secondary structures, like dimers, that will

reduce the ability of probes to detect its target. Uniformity, the last requirement, is specific for large-scale oligonucleotide design. It requires that several attributes of each oligonucleotide have to be calibrated in a narrow range, such as uniform thermodynamics, length, and GC percentages, so that hybridization conditions can be set as stringently as possible without the risk of uneven hybridization across the array.

Image analysis. Image processing plays an indispensable role in DNA array technology. DNA array images are composed of spots with variable grayscale intensities on known coordinates. To estimate the relative abundance of mRNA species, the numerical intensity value from each spot must be extracted properly. DNA array image processing can be divided into three stages: i) gridding, ii) spot segmentation, and iii) spot intensity extraction (98).

Gridding is a process to partition the entire image into a number of grid cells, so each grid cell contains only a single spot. This process is commonly implemented as a semi-automatic procedure, based on knowledge of the number and layout of spots and sub-grids. Spot segmentation is the process of drawing the boundaries between foreground and background pixels. There are essentially two kinds of segmentation algorithms in DNA array image analysis, pattern-dependent and pattern-independent algorithms (14). In DNA array analysis, pattern-dependent segmentation assumes the circular shape of a spot with or without the knowledge of the center position and diameter. Pattern-independent segmentation is designed under the assumption that all foreground pixels are connected, therefore, they can be located with thresholding or by seeded growing. Intensity extraction is a process that outputs the background-corrected intensities based on the statistics of foreground and background pixels. The major

difficulty at this stage lies in correctly estimating the background intensity for each spot. The choice of background has a great impact on the extracted intensity value (98).

Data normalization. After image analysis and feature extraction, extracted raw intensity data must be standardized between experiments and arrays. In the context of DNA array analysis, normalization is used interchangeably with terms like standardization and calibration. It can generally involve in multiplying a scaling factor (constant or variable) with raw intensity data. Although it seems like a straightforward process, normalization in DNA array analysis is quite difficult because of sample heterogeneity and signal non-linearity due to limited dynamic range of the recording instruments. The raw intensity data from gene expression experiments can be calibrated in several ways with or without controls, or with internal or external controls.

In gene expression experiments, raw intensity data can be calibrated on the basis of total or median signal intensities or on the basis of total signals of the set of housekeeping genes (13). This approach is risky because there is no single universal set of genes in bacteria whose gene expressions are constant under all conditions. Also, both total RNA and rRNA levels can change dramatically during certain physiological stages, so the expression of mRNA species will be misrepresented if the raw data are simply normalized to equal total signal intensities (40). Externally spiked control mRNAs offer a better option, since known concentration of mRNA molecules can be added.

It should be emphasized that there is an implicit assumption in the control-based normalization methods, that is, signals are linear across the full dynamic range of raw data intensities. This assumption does not always hold true because of the dynamic range limitations of detection instruments (65). For example, X-ray film has a dynamic range

of two orders of magnitude, while mRNA abundance levels have a dynamic range of more than four orders of magnitude. Obviously, there will be saturation in the high signal range and extensive variability in the low signal range.

To overcome the non-linearity in expression data, a robust normalization algorithm was developed based on local regression procedures, such as *lowess* (99). The assumption here is that expression intensities of most genes are not statistically different, therefore, local linearity is used as the basis of calibration. Localized linear regression relies on linearity in the data points with similar intensities. This approach makes a weaker assumption than global linearity. It is possible to further weaken this assumption for samples with substantially different gene expression patterns using an algorithm called the rank invariant selection algorithm (74). In this scheme, a subset of gene expression intensities with relatively constant expression profiles is selected to calculate the calibration curve, much like selection of a subset of internal controls that are distributed over the full dynamic range and are relatively invariant to different treatments.

Statistical significance of differentially expressed genes. Fundamentally, the purpose of genome-wide expression profiling is similar to single gene expression analysis – to determine differential expressions between the control and treatment. However, the method used in statistical analysis of genome-scale expression data is much different from single gene expression analysis because of the massive number of data, the experimental noise, and the potential for a non-normal data distribution. Without a firmly grounded analysis technique, biological conclusions are misleading at best and false at the worst. Genome-scale analysis offers a new opportunity to compare every

gene in the genome in exactly the same experimental conditions, and new statistical procedures have already been developed to try to overcome these problems.

Noisy data can come from various sources, such as non-uniform spotting and immobilization, inconsistent extraction efficiency and biological heterogeneity. It is not uncommon that DNA array experiments require ≥4 replicates to overcome lack of repeatability in DNA array data.

Multiple comparison is a series statistical problem in DNA array data analysis. It refers to conducting a large number of individual pair-wise tests, such as t-tests, which introduces a new source of error that complicates the biological conclusions produced by the array. The multiple comparison problem can be best illuminated with the following example. Supposedly, in an array with 1,000 spots, there are exactly 50 corresponding genes expressed that are truly different under control and treatment. If we set the Type I error rate at 0.05, we expect to find total $50 + 0.05X(1,000-50) = 97.5$ genes differentially expressed. That is, we have almost 50% false positives in the detected "differentially expressed" genes. Currently, there are two ways to handle this problem, i.e., control the family-wise error rate (FWER) (20) or control the false discovery rate (FDR) (84). Controlling FWER methods are generally more stringent than controlling FDR methods, and therefore require more replicates to discover statistically differentially expressed genes (81).

It is important to note that most assumptions of classical statistical tests are often not valid in DNA array data. For example, the t-statistic requires normal distributions and equal variance between the two samples, both of which are false for some DNA array data. To overcome this difficulty, most DNA array data analyses have to rely on non-

parametric methods or test statistics with weakened assumptions. For example, when data deviates from the normal distribution, permutation is a common statistical technique used to overcome this problem, because it can estimate the null distribution numerically from the existing data set (20, 84). DNA array technology has posed as much of a challenge to statistical research as to biological research. It has been an active research topic in the field of statistics in the last five years and new statistical procedures and methodologies are constantly showing up.

Classification of gene expression patterns: Machine learning, or statistical learning, is a new interdisciplinary field involving computer science, statistics and engineering. The goal of this field is to develop tools to extract important patterns and trends in massive amounts of data. From the beginning, machine learning has played a pivotal role in DNA array data analysis. This is primarily due to the high dimensionality in DNA array data. We have to depend on computational algorithms to reveal the intricate structures and patterns inside massive amounts of gene expression data. This pattern can be in the form of temporal or spatial association and interdependence among genes or in the form of differences between normal and disease samples. There are mainly two kinds of data mining methodologies, supervised and unsupervised (81). The primary difference between these methods is that the unsupervised method explores the data without prior class knowledge.

Clustering is a form of unsupervised machine learning and is the most popular multivariate technique that is used to find structure in DNA array data. This popularity probably has more to do with the conceptual simplicity and elegance of clustering algorithms. Clustering techniques are very different from traditional statistical methods.

Instead of testing existing hypotheses, clustering algorithms explore and discover the patterns in the data (class discovery). Various clustering techniques are applied to identify the patterns in gene expression data, such as hierarchical (22), self-organized map (82), and principal component analysis (69). Although cluster analysis is extremely powerful, researchers must take a great care when applying this technique to DNA array data analysis. Without proper biological considerations, clustering algorithms will be degraded into a numerical game, and will produce clusters even on unrelated data.

In DNA array data analysis, supervised learning, also called classification, can be used to predict the functions of unknown genes if they share an expression pattern similar to known class of genes (10). Classification algorithms have quickly gained popularity recently. For example, a Bayesian classification system was used to predict whether genes are co-expressed as an operon in *E. coli* (72). Several classification-based algorithms have been applied to DNA array data analysis, such as support vector machine analysis (SVM) (10). One of the most promising applications of classification lies in diagnostics, especially in complex diseases, such as cancer (class prediction) (35). DNA array technology has posed as much of a challenge to statistical research as to biological research. It has been an active research topic in the field of statistics in the last five years and new statistical procedures and methodologies are constantly showing up.

DNA array technology forms the core of functional genomics, which promises to revolutionize the way we design and conduct biological experiments in food systems, especially food microbiology (93). Historically, the complexity of food systems has hindered discovery of molecular mechanisms of bacterial metabolism in food. With high-throughput tools like DNA arrays, functional genomics holds great promise to

provide answers to age-old questions about the role of microbes in flavor, spoilage, and pathogenesis. The aim of this work is to merge functional genomics with food microbiology to examine the gene expression of lactococci. This will help us to understand the complex dynamic interactions among the gene regulatory network (transcriptome), protein functional network (proteome) and metabolites network (metabolome). It will have a direct impact on the improvement of flavor and texture during cheese manufacture and on the enhancement of starter cultures.

Hypothesis

Casein hydrolysis by the extracellular proteinases of *Brevibacterium linens* BL2 changes the expression profiles of proteolytic related metabolism genes in *Lactococcus lactis* spp. *lactis* IL1403.

Objectives

1. Develop an oligonucleotide-based DNA array technique suitable for monitoring gene expression in *L. lactis* spp. *lactis* IL1403.
2. Establish statistical and computational procedures that are suitable to analyze the data generated with the DNA arrays.
3. Demonstrate and utilize this technique in: a) profiling the stress responses of metabolic-related genes in *L. lactis* spp. *lactis* IL1403; b) determining the effect of the extracellular proteolytic activities of *B. linens* BL2 on the growth and gene expression profiles in *L. lactis* IL1403.

Preamble

The chapters in this dissertation do not match the objectives as they are stated. Some of the objectives were combined, while other objectives were split. The description below defines what chapter meets each objective.

To monitor the global gene regulation patterns in *L. lactis* ssp. *lactis* IL1403, a high throughput tool was needed to study the gene expression profiles in a genomic scale. Previously, our group successfully used oligonucleotide probes to monitor the expression profiles of genes involved in the arginine deiminase pathway in *L. lactis*. However, this protocol was not sufficient for the genomic level studies due to limitations in sensitivity, specificity, and cost. Chapter 3 describes a novel filter-based oligonucleotide DNA macroarray protocol that was sensitive, highly specific, inexpensive, and easy to use. This meets the criteria for Objective 1.

Design and analysis of DNA arrays were computationally demanding tasks. There were several key computational problems that existed in DNA array design and data analysis, such as large-scale oligonucleotide probe design, spot digitalization, expression data normalization and statistical analysis. Flowcharts of the software components implemented in this project and their algorithms are available in Appendix A, B, C, and D, and in my degree report in computer sciences. Together these meet the criteria for Objective 2.

In Chapter 3, the utility of this protocol is demonstrated by profiling the stress response of intermediary metabolic genes in *L. lactis* spp. *lactis* IL1403 during heat, acid, and osmotic stress treatment. Since stress responses have been intensively studied in

lactic acid bacteria, a wealth of information was available to allow us to verify the performance of the DNA macroarray. In almost all cases, the results obtained with the macroarray agreed well with previously reported observations. In addition to confirming previous findings, we also gained new insights into the stress responses of *L. lactis* ssp. *lactis* IL1403. This meets the criteria for Objective 3a.

In Chapter 4, we profiled gene expression changes of *L. lactis* spp. *lactis* IL1403 growing in a peptide-limited medium, in a casitone-based peptide-rich medium, and in a casein hydrolysate produced by *B. linens* BL2 proteolytic enzymes. We found that *L. lactis* ssp. *lactis* IL1403 experienced nitrogen starvation even in the casitone-based medium where abundant peptide resources were available. This was because peptide transporter genes were not expressed. The repression of peptide transporter genes was due to transported intracellular branched chain amino acids (BCAA) via a BCAA-specific permease (*brnQ*). Conversely, a peptide pool generated by *B. linens* BL2 proteolytic activities sustained the growth of *L. lactis* ssp. *lactis* IL1403 because the repression of peptide transporter genes and other proteolytic related genes was relieved in this medium. Finally, we conducted a biochemical study on the extracellular protease of *B. linens* BL2 in Chapter 5. This enzyme was secreted as an inactive zymogen and required a maturation process to become active. Regulation of the total protease activity was controlled during maturation and proteolysis. These studies meet the criteria for Objective 3b.

References

1. **Adler-Nissen, J.** 1987. Bitterness intensity of protein hydrolyzates - chemical and organoleptic characterization. Proceedings of the 5th International Flavor Conference. Elsevier Science Publishers, Porto Karras, Chalkidiki, Greece.

2. **Aston, J. W., and J. R. Dully.** 1982. Cheddar cheese flavor. The Australian J. Dairy Tech.**42:**59-64.

3. **Beer, C.** 1998. Ph.D. dissertation. Utah State University, Logan, Utah.

4. **Bernal, A., U. Ear, and N. Kyrpides.** 2001. Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. Nucleic Acids Res. **29:**126-7.

5. **Blattner, F. R., G. Plunkett, 3rd, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:**1453-74.

6. **Bolotin, A., S. Mauger, K. Malarme, S. D. Ehrlich, and A. Sorokin.** 1999. Low-redundancy sequencing of the entire *Lactococcus lactis* IL1403 genome. Antonie Van Leeuwenhoek **76:**27-76.

7. **Bolotin, A., P. Wincker, S. Mauger, O. Jaillon, K. Malarme, J. Weissenbach, S. D. Ehrlich, and A. Sorokin.** 2001. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. Genome Res. **11:**731-53.

8. **Broadbent, J. R., M. Barnes, C. Brennand, M. Strickland, K. Houck, M. E. Johnson, and J. L. Steele.** 2002. Contribution of *Lactococcus lactis* cell envelope proteinase specificity to peptide accumulation and bitterness in reduced-fat Cheddar cheese. Appl. Environ. Microbiol. **68:**1778-85.

9. **Broadbent, J. R., M. Strickland, B. Weimer, M. E. Johnson, and J. L. Steele.** 1998. Peptide accumulation and bitterness in Cheddar cheese made using single-strain *Lactococcus lactis* starters with distinct proteinase specificities. J. Dairy Sci. **81:**327-337.

10. **Brown, M. P., W. N. Grundy, D. Lin, N. Cristianini, C. W. Sugnet, T. S. Furey, M. Ares, Jr., and D. Haussler.** 2000. Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc. Natl. Acad. Sci. USA **97:**262-7.

11. **Chen, J. J., R. Wu, P. C. Yang, J. Y. Huang, Y. P. Sher, M. H. Han, W. C. Kao, P. J. Lee, T. F. Chiu, F. Chang, Y. W. Chu, C. W. Wu, and K. Peck.** 1998. Profiling expression patterns and isolating differentially expressed genes by cDNA microarray system with colorimetry detection. Genomics **51:**313-24.

12. **Chou, L.** 2002. Ph.D. dissertation. Utah State University, Logan, Utah.

13. **Cox, J. M.** 2001. Applications of nylon membrane arrays to gene expression analysis. J. Immunol. Methods **250:**3-13.

14. **Davies, E. R.** 1997. Machine vision. 2nd ed. Academic Press, San Diego, Ca.

15. **de Vos, W. M., G. S. P. Vos, and S. David.** 1989. Gene organization and expression in mesophilic lactic acid bacteria. J. Dairy Sci. **72:**3398-3405.

16. **Decraene, C., I. Reguigne-Arnould, C. Auffray, and G. Pietu.** 1999. Reverse transcription in the presence of dideoxynucleotides to increase the sensitivity of expression monitoring with cDNA arrays. Biotechniques **27:**962-6.

17. **DeRisi, J. L., V. R. Iyer, and P. O. Brown.** 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science **278:**680-6.

18. **Detmers, F. J., E. R. Kunji, F. C. Lanfermeijer, B. Poolman, and W. N. Konings.** 1998. Kinetics and specificity of peptide uptake by the oligopeptide transport system of *Lactococcus lactis*. Biochemistry **37:**16671-9.

19. **Dorris, D., R. Ramakrishnan, T. Sendera, and S. Magnuson.** 2002. Oligonucleotide array technologies for gene expression profiling, p. 25-38. *In* E. Grigorenko (ed.), DNA arrays: Technologies and experimental strategies. CRC Press, New York, N. Y.

20. **Dudoit, S., Y. H. Yang, M. J. Callow, and T. P. Speed.** 2002. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Statist. Sinica. **12:**111-139.

21. **Duun, H. C., and R. C. Lindsay.** 1985. Evaluation of the role of microbial Streker-derived aroma compounds in unclean-type flavors of Cheddar cheese. J. Dairy Sci. **68:**2859-2870.

22. **Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein.** 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA **95:**14863-8.

23. **Even, S., N. D. Lindley, and M. Cocaign-Bousquet.** 2001. Molecular physiology of sugar catabolism in *Lactococcus lactis* IL1403. J. Bacteriol. **183:**3817-24.

24. **Exterkate, F. A.** 1990. Differences in short peptide-substrate cleavage by two cell-envelope- located serine proteinases of *Lactococcus lactis* subsp. *cremoris* are related to secondary binding specificity. Appl. Microbiol. Biotechnol. **33:**401-6.

25. **Exterkate, F. A., A. C. Alting, and P. G. Bruinenberg.** 1993. Diversity of cell envelope proteinase specificity among strains of *Lactococcus lactis* and its relationship to charge characteristics of the substrate-binding region. Appl. Environ. Microbiol. **59:**3640-7.

26. **Farkye, N. Y.** 1995. Contribution of milk-clotting enzymes and plasmin to cheese ripening. Adv. Exp. Med. Biol. **367:**195-207.

27. **Farkye, N. Y., and P. F. Fox.** 1992. Contribution of plasmin to Cheddar cheese ripening: effect of added plasmin. J. Dairy Res. **59:**209.

28. **Fedrick, I., J. Aston, S. Nottingham, and J. Dulley.** 1986. The effect of neutral fungal protease on Cheddar cheese ripening. N. Z. J. of Dairy Sci.. **21:**9-19.

29. **Fischetti, V. A., V. Pancholi, and O. Schneewind.** 1990. Conservation of a hexapeptide sequence in the anchor region of surface protein from Gram-positive bacteria. Mol. Microbiol. **4:**1603-1605.

30. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J. Gocayne, J. Scott, R. Shirley, L. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S.M. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, M. C. Fraser, H. O. Smith, and C. J. Venter.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269:**496-512.

31. **Foissy, H.** 1978. Aminopeptidase from *Brevibacterium linens*: activation and inhibition. Z. Lebensm. Unters Forsch. **166:**164-6.

32. **Fox, P. F.** 1989. Proteolysis during cheese manufacture and ripening. J. Dairy Sci. **72:**1379-1390.

33. **Fox, P. F., J. M. Wallace, S. Morgan, C. M. Lynch, E. J. Niland, and J. Tobin.** 1996. Acceleration of cheese ripening. Antonie van Leeuvenhoek **70:**271-297.

34. **Friedman, M. E., W. O. Nelson, and W. A. Wood.** 1953. Proteolytic enzymes from *Bacterium linens*. J. Dairy Sci. **36:**1124-1133.

35. **Golub, T. R., D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander.** 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science **286:**531-7.

36. **Guedon, E., P. Renault, S. D. Ehrlich, and C. Delorme.** 2001. Transcriptional pattern of genes coding for the proteolytic system of *Lactococcus lactis* and evidence for coordinated regulation of key enzymes by peptide supply. J. Bacteriol. **183:**3614-22.

37. **Gummalla, S., and J. R. Broadbent.** 2001. Tyrosine and phenylalanine catabolism by *Lactobacillus* cheese flavor adjuncts. J. Dairy Sci. **84:**1011-9.

38. **Haandrikman, A. J., J. Kok, and G. Venema.** 1991. Lactococcal proteinase maturation protein PrtM is a lipoprotein. J. Bacteriol. **173:**4517-25.

39. **Haandrikman, A. J., R. Meesters, H. Laan, W. N. Konings, J. Kok, and G. Venema.** 1991. Processing of the lactococcal extracellular serine proteinase. Appl. Environ. Microbiol. **57:**1899-904.

40. **Hansen, M. C., A. K. Nielsen, S. Molin, K. Hammer, and M. Kilstrup.** 2001. Changes in rRNA levels during stress invalidates results from mRNA blotting: fluorescence in situ rRNA hybridization permits renormalization for estimation of cellular mRNA levels. J. Bacteriol. **183:**4747-51.

41. **Hayashi, K., A. J. Cliffe, and B. A. Law.** 1990. Purification and preliminary characterization of five serine proteinases by *Brevibacterium linens*. International J. Food Sci. and Tech. **25:**180-187.

42. **Hayashi, K., D. F. Revell, and B. A. Law.** 1990. Accelerated ripening of Cheddar cheese with the aminopeptidase of *Brevibacterium linens* and a commercial neutral proteinase. J. Dairy Res. **57:**571-577.

43. **Hayashi, K., D. F. Revell, and B. A. Law.** 1990. Effect of partially purified extracelllular serine proteinases produced by *Brevibacterium linens* on the accelerated ripening of cheeddar cheese. J. Dairy Sci. **73:**579-583.

44. **Hoevel, T., H. Holz, and M. Kubbies.** 1999. Cisplatin-digoxigenin mRNA labeling for nonradioactive detection of mRNA hybridized onto nucleic acid cDNA arrays. Biotechniques **27:**1064-7.

45. **Juhasz, O., and B. Skarka.** 1990. Purification and characterization of an extracellular proteinase from *Brevibacterium linens*. Canadian J. Microbiol. **36:**510-512.

46. **Juhasz, O., and B. Skarka.** 1989. Purification and some properties of an extracellular proteinase from *Brevibacterium linens*. Biologia **44:**1063-1068.

47. **Kelly, M., P. F. Fox, and P. L. H. McSweeney.** 1996. Effect of salt-in-moisture on proteolysis in Cheedar-type cheese. Milchwissenschaft **51:**498-501.

48. **Kok, J.** 1990. Genetics of the proteolytic system of lactic acid bacteria. FEMS Microbiol Rev. **7:**15-42.

49. **Kunji, E. R., G. Fang, C. M. Jeronimus-Stratingh, A. P. Bruins, B. Poolman, and W. N. Konings.** 1998. Reconstruction of the proteolytic pathway for use of beta-casein by *Lactococcus lactis*. Mol. Microbiol. **27:**1107-18.

50. **Kunji, E. R., I. Mierau, A. Hagting, B. Poolman, and W. N. Konings.** 1996. The proteolytic systems of lactic acid bacteria. Antonie Van Leeuwenhoek. **70:**187-221.

51. **Larsen, L. B., C. Benfeldt, L. K. Rasmussen, and T. E. Petersen.** 1996. Bovine milk procathepsin D and cathepsin D: coagulation and milk protein degradation. J. Dairy Res. **63:**119-30.

52. **Larsen, L. B., and T. E. Petersen.** 1995. Identification of five molecular forms of cathepsin D in bovine milk. Adv. Exp. Med. Biol. **362:**279-83.

53. **Law, B.** 1980. Accelerated ripening of cheese. Dairy Industry International **45:**15-22.

54. **Law, B.** 1984. The accelerated ripening of cheese, p. 209-228. *In* F. L. Davies, and B. Law (ed.), Advances in the microbiology and biochemistry of cheese and fermented milk. Elsevier Applied Science Publishers, London.

55. **Law, B., and A. Wigmore.** 1983. Accelerated ripening of Cheddar cheese with a commercial proteinase and intracellular enzymes from starter streptococci. J. Dairy Res. **50:**519-525.

56. **Le Bourgeois, P., M. Lautier, M. Mata, and P. Ritzenthaler.** 1992. Physical and genetic map of the chromosome of *Lactococcus lactis* subsp. *lactis* IL1403. J. Bacteriol. **174:**6752-62.

57. **Li, F., and G. D. Stormo.** 2001. Selection of optimal DNA oligos for gene expression arrays. Bioinformatics **17:**1067-76.

58. **Lipshutz, R. J., S. P. Fodor, T. R. Gingeras, and D. J. Lockhart.** 1999. High density synthetic oligonucleotide arrays. Nat. Genet. **21:**20-4.

59. **Lockhart, D. J., H. Dong, M. C. Byrne, M. T. Follettie, M. V. Gallo, M. S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Horton, and E. L. Brown.** 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. **14:**1675-80.

60. **Manning, D. J., and J. C. Price.** 1977. Cheddar cheese aroma-the effect of selectively removing specific classes of compounds from cheese headspace. J. Dairy Res. **44:**357-368.

61. **McGarry, A., J. Law, A. Coffey, C. Daly, P. F. Fox, and G. F. Fitzgerald.** 1994. Effect of genetically modifying the lactococcal proteolytic system on ripening and flavor development in Cheddar cheese. Appl. Environ. Microbiol. **60:**4226-4233.

62. **Ney, K. H.** 1979. Bitterness of peptides: amino acid composition and chain length, p. 149-173. *In*: Bondreau, J. C. (ed.), Food taste chemistry, American Chemical Society, Washington, D.C.

63. **Quackenbush, J.** 2001. Computational analysis of microarray data. Nat Rev Genet. **2:**418-27.

64. **Rajeevan, M. S., I. M. Dimulescu, E. R. Unger, and S. D. Vernon.** 1999. Chemiluminescent analysis of gene expression on high-density filter arrays. J. Histochem. Cytochem. **47:**337-42.

65. **Ramdas, L., K. R. Coombes, K. Baggerly, L. Abruzzo, W. E. Highsmith, T. Krogmann, S. R. Hamilton, and W. Zhang.** 2001. Sources of nonlinearity in cDNA microarray expression measurements. Genome Biol. **2:**research0047.1–0047.7.

66. **Rattray, F. P., W. Bockelmann, and P. F. Fox.** 1995. Purification and characterizaiton of an extracellular proteinase from *Brevibacterium linens* ATCC 9174. Appl. Environ. Microbiol. **61:**3454-3456.

67. **Rattray, F. P., P. F. Fox, and A. Healy.** 1996. Specificity of an extracellular proteinase from *Brevibacterium linens* ATCC 9174 on bovine α_{s1}-casein. Appl. Environ. Microbiol. **62:**501-6.

68. **Rattray, F. P., P. F. Fox, and A. Healy.** 1997. Specificity of an extracellular proteinase from *Brevibacterium linens* ATCC 9174 on bovine β-casein. Appl. Environ. Microbiol. **63:**2468-71.

69. **Raychaudhuri, S., J. M. Stuart, and R. B. Altman.** 2000. Principal components analysis to summarize microarray experiments: application to sporulation time series. Pac. Symp. Biocomput. **5:**455-66.

70. **Rosenow, C., R. M. Saxena, M. Durst, and T. R. Gingeras.** 2001. Prokaryotic RNA preparation methods useful for high density array analysis: comparison of two approaches. Nucleic Acids Res. **29:**E112.

71. **Rouillard, J. M., C. J. Herbert, and M. Zuker.** 2002. OligoArray: genome-scale oligonucleotide design for microarrays. Bioinformatics **18:**486-7.

72. **Sabatti, C., L. Rohlin, M. K. Oh, and J. C. Liao.** 2002. Co-expression pattern from DNA microarray experiments as a tool for operon prediction. Nucleic Acids Res. **30:**2886-93.

73. **Salama, N., K. Guillemin, T. K. McDaniel, G. Sherlock, L. Tompkins, and S. Falkow.** 2000. A whole-genome microarray reveals genetic diversity among Helicobacter pylori strains. Proc. Natl. Acad. Sci. USA **97:**14668-73.

74. **Schadt, E. E., C. Li, B. Ellis, and W. H. Wong.** 2001. Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. J. Cell Biochem. **Suppl:**120-5.

75. **Schechter, I.** 1968. On the active site of proteases. 3. Mapping the active site of papain; specific peptide inhibitors of papain. Biochem. Biophys. Res. Commun. **32:**898-902.

76. **Schena, M., D. Shalon, R. W. Davis, and P. O. Brown.** 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270:**467-70.

77. **Schena, M., D. Shalon, R. Heller, A. Chai, P. O. Brown, and R. W. Davis.** 1996. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc. Natl. Acad. Sci. USA **93:**10614-9.

78. **Siezen, R. J.** 1999. Multi-domain, cell-envelope proteinases of lactic acid bacteria. Antonie Van Leeuwenhoek **76:**139-55.

79. **Siezen, R. J., and J. A. Leunissen.** 1997. Subtilases: the superfamily of subtilisin-like serine proteases. Protein Sci. **6:**501-23.

80. **Simard, R. E.** 1991. Evaluation of low fat cheese problems. p. 37-41. *In* Center for Dairy Research (ed.). Cheese Research and Technology Conference, Madison, Wi.

81. **Smyth, G. K., Y. H. Yang, and T. P. Speed.** 2002. Statistical Issues in cDNA Microarray Data Analysis. Technical Report, Department of Statistics, University of California, Berkeley. [Online.] http://stat-www.berkeley.edu/users/terry/zarray/TechReport/mareview.pdf. Accessed 10 February 2003.

82. **Tamayo, P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E. S. Lander, and T. R. Golub.** 1999. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc. Natl. Acad. Sci. USA **96:**2907-12.

83. **Tokita, F., and A. Hosono.** 1972. Studies on the extracellular protease produced by *Brevibacterium linens*. I. production and some properties of the extracellular protease. Jap. J. Zootech. Sci. **43:**39-48.

84. **Tusher, V. G., R. Tibshirani, and G. Chu.** 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA **98:**5116-21.

85. **Ummadi, M., and B. C. Weimer.** 2001. Tryptophan catabolism in *Brevibacterium linens* as a potential cheese flavor adjunct. J. Dairy Sci. **84:**1773-82.

86. **Urbach, G.** 1995. Contribution of lactic acid bacteria to flavor compound formation in dairy products. International Dairy J. **5:**877-903.

87. **van der Vossen, J. M., J. Kodde, A. J. Haandrikman, G. Venema, and J. Kok.** 1992. Characterization of transcription initiation and termination signals of the proteinase genes of *Lactococcus lactis* Wg2 and enhancement of proteolysis in *L. lactis*. Appl. Environ. Microbiol. **58:**3142-9.

88. **Velculescu, V. E., L. Zhang, B. Vogelstein, and K. W. Kinzler.** 1995. Serial analysis of gene expression. Science **270:**484-7.

89. **Visser, S.** 1993. Proteolytic enzymes and their relation to cheese ripening and flavor: an overview. J. Dairy Sci. **76:**329-350.

90. **Wei, Y., J. M. Lee, C. Richmond, F. R. Blattner, J. A. Rafalski, and R. A. LaRossa.** 2001. High-density microarray-mediated gene expression profiling of *Escherichia coli*. J. Bacteriol. **183:**545-56.

91. **Weimer, B.** 1990. Ph.D. dissertation. Utah State University, Logan, Utah.

92. **Weimer, B., B. Dias, U. Madhavi, J. Broadbent, C. Brennand, J. Jaegi, M. Johnson, F. Milani, J. Steele, and D. V. Sission.** 1997. Influence of NaCl and pH on intracellular enzymes that influence Cheddar cheese ripening. Le Lait. **77:**383.

93. **Weimer, B., and D. Mills.** 2002. Enhancing foods with functional genomics. Food Technology. **56**: 184-9.

94. **Weimer, B., K. Seefeldt, and B. Dias.** 1999. Sulfur metabolism in bacteria associated with cheese. Antonie Van Leeuwenhoek **76:**247-61.

95. **Weinstein, M. J., and R. F. Doolittle.** 1972. Differential specificities of thrombin, plasmin and trypsin with regard to synthetic and natural substrates and inhibitors. Biochimia. Biophsia. Acta. **258:**577-580.

96. **Wendisch, V. F., D. P. Zimmer, A. Khodursky, B. Peter, N. Cozzarelli, and S. Kustu.** 2001. Isolation of *Escherichia coli* mRNA and comparison of expression using mRNA and total RNA on DNA microarrays. Anal. Biochem. **290:**205-13.

97. **Xie, Y.** 2002. M.S. report. Utah State University, Logan, Utah.

98. **Yang, Y. H., M. J. Buckley, S. Dudoit, and T. P. Speed.** 2002. Comparison of methods for image analysis on cDNA microarray data. J. Computational and Graphical Statistics **11:**1-29.

99. **Yang, Y. H., S. Dudoit, P. Luu, and T. P. Speed.** 2001. Normalization for cDNA microarray data, p. 24-34. *In* M. L. Bittner, Y. Chen, A. N. Dorsel, and E. R. Dougherty (eds.), Microarrays: Optical technologies and informatics, Proceedings of SPIE, vol. 4266. The International Society for Optical Engineering. Bellingham, Washington.

100. **Ye, R. W., T. Wang, L. Bedzyk, and K. M. Croker.** 2001. Applications of DNA microarrays in microbial systems. J. Microbiol. Methods **47:**257-72.

101. **Yu, J., M. I. Othman, R. Farjo, S. Zareparsi, S. P. MacNee, S. Yoshida, and A. Swaroop.** 2002. Evaluation and optimization of procedures for target labeling and hybridization of cDNA microarrays. Mol. Vis. **8:**130-7.

Table 2-1. Specificity of proteinases in various *L. lactis* strains (3, 9, 25, 50).

		Lactococcal proteinase classes							
		a	*b*	*c*	*d*	*e*	*f*	*g*	*h*
Domains	Amino acid residue[a]	SK11 (P_{III})	AM2 (P_I)	E8 (P_{III}/P_I)	NCDO 763 (P_I)	Wg2 (P_I)	FD27 (P_I/P_{III})	HP (P_I)	S3 (?)
PRotease domain	131	Ser	Thr	Thr	Thr	Thr	Thr	Thr	Thr
	138	Lys	Thr	Thr	Thr	Thr	Thr	Thr	Thr
	142	Ala	Asp	Ala	Ala	Ser	Ala	Asp	Ala
	144	Val	Leu	Leu	Leu	Leu	Leu	Leu	Leu
	166	Asn	Asp	Asp	Asp	Asp	Asp	Asp	Gly
	177	Leu	Ile	Ile	Leu	Leu	Leu	Ile	Ile
A-domain	747	Arg	Leu	Arg	Leu	Leu	Leu	Leu	Leu
	748	Lys	Thr	Lys	Thr	Thr	Thr	Thr	Thr
	763	Asn	His	Asn	His	Asn	His	His	His
Fragments generated from a_{s1}-casein (f1-23)		1-16[1], 1-17[1], 1-21[2], 17-23[2], 18-23[2], 22-23[3]	1-16[2], 1-17[1], 1-21[2], 17-23[2], 18-23[2], 22-23[3]	1-16[3], 1-17[1], 1-21[3], 17-23[3], 18-23[3], 22-23[3]	1-8[3], 1-9[3], 1-16[3], 1-17[3], 1-19[3], 1-21[1], 22-23[3]	1-8[1], 1-9[1], 1-13[2], 1-16[2], 1-21[2], 9-23[2], 10-23[2], 17-23[3], 22-23[3]	1-8[2], 1-9[2], 1-13[1], 1-16[2], 1-17[2], 1-21[3], 9-23[3], 10-23[2], 17-23[3], 18-23[3]	1-8[1], 1-9[1], 1-13[2], 9-23[2], 10-23[2], 14-23[2]	1-9[2], 1-13[2], 1-16[1], 17-23[2], 14-23[2]

[a] Amino acid numbering is identical to *L. lactis* spp. *cremoris* SK11 proteinase amino acid sequence.
[1,2,3] Numbering indicate the cleavage rates to produce various fragments. The bigger the index number, the higher the cleavage rates.

Table 2-2. Lactococcal amino acid and peptide transport systems (3, 18).

Transport system	Mechanism[1]	Specificity
Branched-chain amino acids	PMF	Leu, Ile, Val
Neutral amino acids I	PMF	Ala, Gly
Neutral amino acids II	PMF	Ser, Thr
Basic amino acids I	PMF	Lys, Orn
Basic amino acids II	PMF	His[2]
Aromatic amino acids	PMF	Phe, Tyr, Trp[3]
Glutamate	ATP	Glu, Gln
Asparagine	ATP	Asn
Aspartate	exchange	Asp, Glu[4]
Arg-Orn	antiport	Arg, Orn, Lys
Proline	diffusion	Pro
Di- and Tripeptides	PMF/ATP	Di- and tripeptides[5]
Oligopeptides	ATP	oligopeptides (4-18 residues)[6]

[1] PMF = proton-motive-force, ATP = ATP-driven
[2] Substrate specificity has not been studied
[3] Substrate specificity has not been studied in detail
[4] System with low affinity for acidic amino acids
[5] Arginine containing peptides are not transported
[6] Neutral oligopeptides not containing proline residues

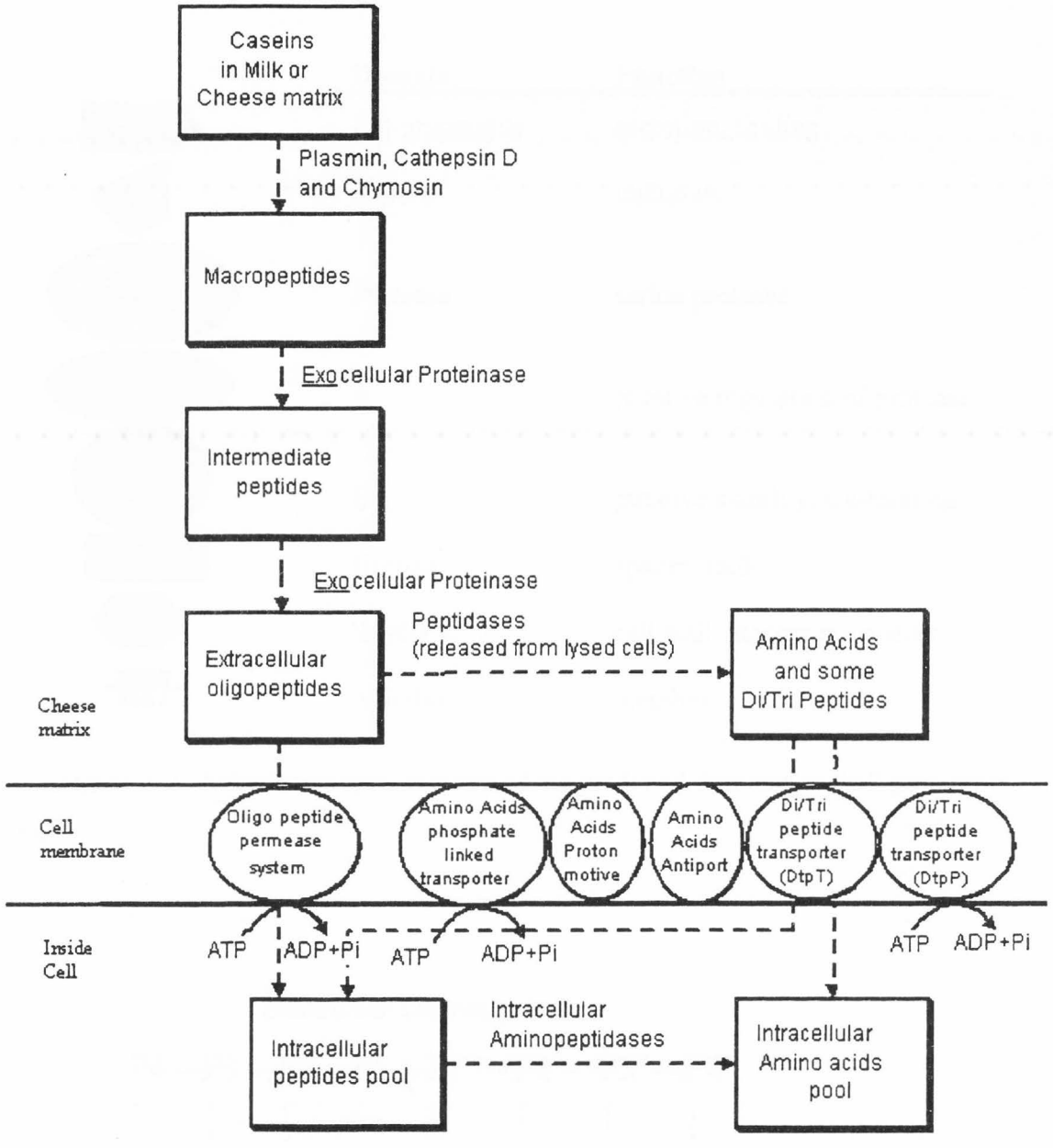
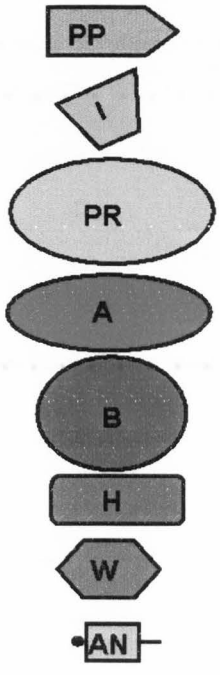


Figure 2-1. Proteolytic network in the cheese matrix (32, 50). See text for a complete description of all proteins.



Domain	Function
Pre-propeptide	secretion, folding
I(nsert)	unknown
Protease	serine protease
A	putative regulation of protease
B	putative stability, Ca-binding
H(elix)	spacer, stalk
W(all)	cell wall attachment, spacer
AN(chor)	membrane anchor

Figure 2-2. Schematic representation of the predicted domains in PrtP proteinase (78).

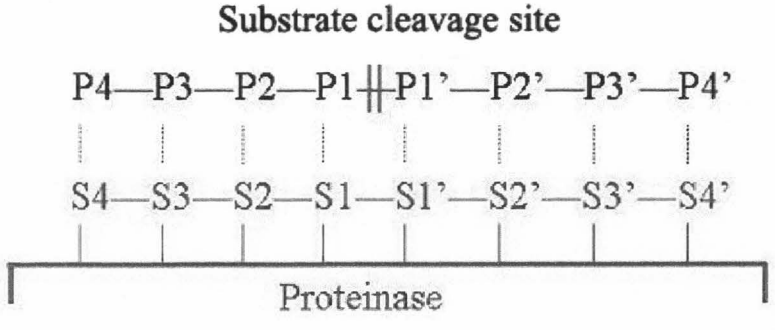


Figure 2-3. Number system to describe the specificity of a proteinase (75).

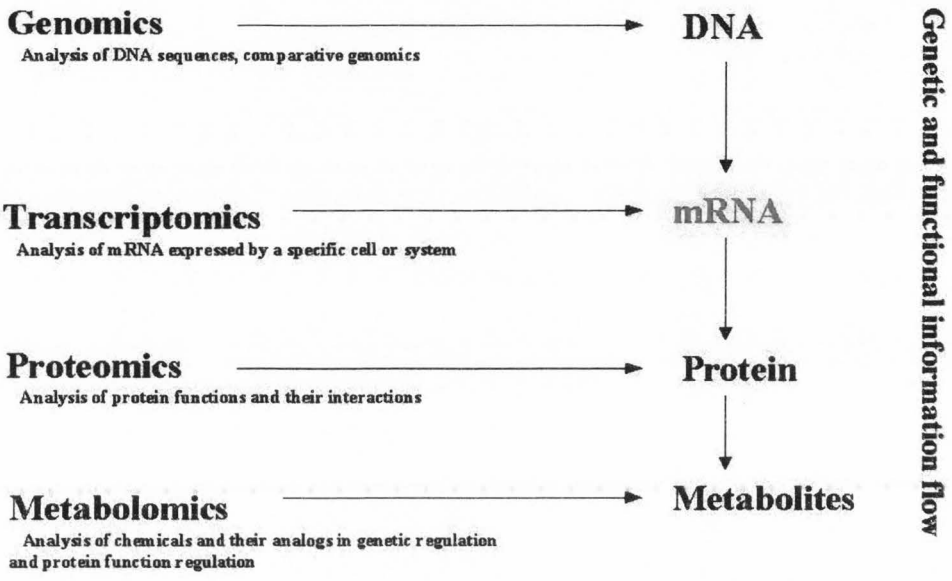


Figure 2-4. Genetic information flow and functional genomics.

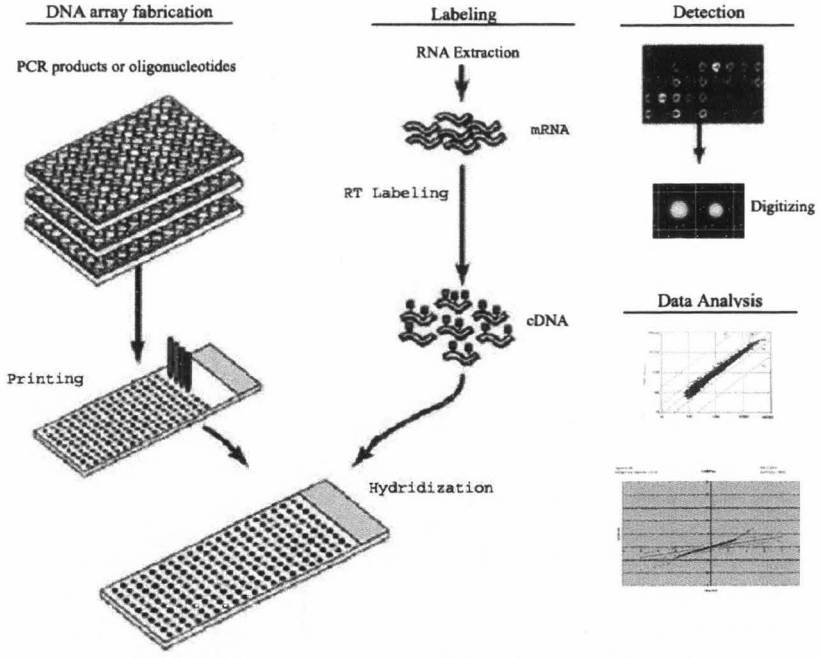


Figure 2-5. Scheme drawing of DNA array fabrication and use (76).

CHAPTER 3

EXPRESSION PROFILING OF *LACTOCOCCUS LACTIS* SSP. *LACTIS* IL1403 UNDER STRESSES WITH DNA MACROARRAY[1]

Abstract

DNA microarray technology is one of the primary tools used in functional genomics research. This paper describes an oligonucleotide-based filter DNA array protocol. The DNA macroarray was composed of short oligonucleotide probes immobilized onto a nylon membrane via polyinosine tails as spacers and presumably non-specific hybrid stabilizers. An indirect high-density labeling method was established to effectively incorporate biotin into the nucleic acids. The hybridization signals were detected with a chemiluminescence-based method and digitized with a desktop scanner. The utility of this protocol was demonstrated by profiling the expression of 375 metabolically related genes in *Lactococcus lactis* ssp. *lactis* IL1403 during heat, acid and osmotic stress. The macroarray accurately detected known stress responses in lactococci, indicating the array procedure and results were valid. Based on this observations, we made additional observations on the stress responses in *L. lactis* ssp. *lactis* IL1403. For example, the methionine biosynthesis pathway was induced during heat shock, but repressed during acid stress. Although peptide transporter genes (*opt* and *dtpT*) were similarly repressed in all three stresses, the amino acid transport genes were strongly induced after heat stress, but remained relatively unchanged in other stress treatments. In

[1] Coauthored by Yi Xie, Lan-szu Chou, Adele Cutler, and Bart Weimer.

addition to the peptide transport genes, several other genes were also repressed in all of the stress treatments, such as the β-glucoside specific PTS system and the arginine deiminase pathway genes. In conclusion, we developed a unique oligonucleotide array that was low cost, easy to use and highly specific.

Introduction

Bacteria were used as starter cultures in cheese manufacture long before the existence of microbiology. Cheese flavor compounds are derived from the metabolic activities of the microbial community living inside the cheese matrix, especially the starter culture *Lactococcus lactis* (42). Because of its economic importance, the physiology of this microbe is studied intensively. Recently, publication of the genome sequence of *L. lactis* ssp. *lactis* IL1403 opened many new research opportunities to further explore the metabolic capacities of this bacterium (2).

During the milk fermentation for cheese production, lactococci experience a variety of stresses, including acid, temperature, and osmotic shock. Gene regulation induced by these stresses has a direct influence on the fermentation process and flavor compound generation. Before the existence of DNA microarray technology, two-dimensional gel electrophoresis was the primary tool available to study global effects of stress responses in microorganisms (15, 16, 21). However, this method is labor-intensive, difficult to reproduce, and is limited to cytosolic proteins. While the regulation of stress responses is well established in *L. lactis* (32), the impact on expression of genes involved in intermediate metabolism has not been systematically explored.

DNA array technology can monitor the expression profiles of thousands of mRNA transcripts in parallel in a single experiment (37). The primary barriers preventing the implementation of array technology in many laboratories are the high startup cost and technical complexity associated with the current generation of DNA microarrays. In this work, we developed a low-cost oligonucleotide-based filter array protocol in the format of a DNA macroarray. The main difference between DNA microarrays and macroarrays is the distance between spots (pitch size). The pitch sizes in DNA macroarrays are around 1 mm, instead of 1 μm in DNA microarrays. The large pitch size makes the DNA macroarray an ideal choice for arrays with low or medium probe density, for example, bacterial arrays. In addition, we chose oligonucleotide as the probe in this protocol instead of cDNA. The oligonucleotide-based probes are more specific to the target molecules than the cDNA probes commonly used in DNA microarrays, and are able to differentiate the expression of genes with >90% identity (33). The whole protocol is similar to the reverse dot-blot procedure (20), and can be used by anyone familiar with Southern or Northern blotting techniques.

The utility of this novel DNA array protocol was demonstrated by profiling the expression patterns of 375 metabolic genes in *L. lactis* ssp. *lactis* IL1403 under heat, acid and osmotic stresses. These genes, designated as metabolic genes, were selected based on their annotated roles in carbohydrate, energy, fatty acid, nucleic acid, and amino acid metabolism, as well as protein degradation and transport pathways (2, 19). We found that the expression patterns of these genes were significantly altered by a variety of stresses. Approximately 13-18% of the investigated genes were differentially expressed under each of the environmental stress treatment. Many of our observations from the

DNA macroarray were in accord with the results from previous studies, which provided verification to our array protocol. In addition, we made some novel discoveries on the stress responses of *L. lactis* ssp. *lactis* IL1403.

Materials and Methods

Bacterial Strains

Lactococcus lactis ssp. *lactis* IL1403 (courtesy of Dr. Larry McKay, University of Minnesota) is a plasmid-cured strain that lacks plasmid-encoded features, such as lactose transporter, cell envelop protease, citrate transporter, and restriction/modification systems. IL1403 was grown at 30°C overnight without shaking in M17 broth (Difco, Sparks, Maryland) supplemented with 0.5% glucose (M17G). *Lactococcus lactis* ssp. *cremoris* ML3 was obtained from our laboratory collection. The working cultures of ML3 were grown at 30°C overnight without shaking in Ellikers broth (Difco, Sparks, Maryland).

Stress Treatment Conditions

Overnight cultures of *L. lactis* ssp. *lactis* IL1403 grown in M17G broth were harvested by centrifugation at 5000 X g for 10 minutes at 4°C, washed in 0.85% NaCl and then refreshed with fresh M17G broth for 10 minutes at 30°C. The cells were recollected by centrifugation at 5000 x g for 10 minutes at 4°C and suspended in fresh broth or stress conditioning media, as specified below. The cell densities in the treatment media were adjusted to optical density at 600 nm (OD_{600}) of 1.5.

The medium for the control (no stress treatment) and the heat shock treatment was sterile M17G broth. The acid stress medium was sterile M17G broth with the pH adjusted to 5.5 with lactic acid, and the osmotic shock medium was sterile M17G broth supplemented with NaCl to a final concentration of 4%. All treatments except heat shock were incubated at 30°C. Heat shock samples were incubated at 42°C. Samples were collected after 30 minutes of treatment, and the total RNA was extracted as described below. The OD_{600} and pH were also determined at the time of RNA isolation. The experiments, including the control, were replicated twice.

Total RNA Isolation

Cells from 1.5 ml cultures were collected by centrifugation at 16,000 X *g* for 2 minutes at room temperature, resuspended in lysozyme (50 mg/ml in 50 mM EDTA at pH 8.0), and incubated for 10 minutes at 25°C. After centrifugation at 5,000 x *g* for 1 minute, the pellet was re-suspended in 300 µl of the Lysis/Binding solution supplied in the RNAqueous kit (Ambion Inc., Austin, Texas). The total RNA was subsequently extracted according to the manufacturer's recommendations. Contaminating genomic DNA and protein were removed with RNeasy kit (Qiagen, Valencia, California) according to the manufacturer's recommendations, followed by LiCl precipitation. The mRNA transcripts (1 ng each) of RCP1 and XCP2 genes from *Arabidopsis thaliana* were added before the total RNA isolation and used as the positive controls for RNA extraction (Stratagene, La Jolla, California).

Total RNA Labeling with Biotin

Biotin was incorporated into the cDNA from the RNA transcripts with an indirect labeling procedure. Briefly, 100 μg of total RNA was reverse-transcribed into DNA and random hexmers with Superscriptase II (Life Technologies, Rockville, Maryland). The reverse transcription procedure described by the manufacturer was used, except the dNTP was replaced with aminoallyl-dNTP mixture (aa-dNTP). The 50X aa-dNTP stock mixture was composed of 10 mM dATP, dGTP, dCTP, 4 mM dTTP (Life Technologies, Rockville, Maryland), 6 mM aminoallyl-dUTP (Sigma, St. Louis, Missouri), and 1 mM ddATP (Sigma). Once reverse transcription was complete, the enzyme was heat-inactivated and the RNA templates were degraded with RNaseH (Epicentre, Madison, Wisconsin). The reaction mixture was cleaned with Qiaquick-PCR purification kit (Qiagen, Valencia, California), replacing the Tris-containing wash buffer with 75% ethanol. The purified ssDNA was eluted from the columns twice with total 100 μl PBS buffer (pH 7.2). Freshly prepared Sulfo-NHS-LC-biotin (10 μg/μl) (Pierce, Rockford, Illinois) was added (20 μl) to the purified amino-labeled single-strained cDNA (ssDNA) solution, and the mixture was incubated at 25˚C for 1 hour. The reaction was terminated by addition of 24 μmole hydroxylamine. After incubation at 25˚C for 15 minutes, the labeled DNA was purified using a Qiaquick-PCR purification kit (Qiagen). The elutant DNA mixture was heated at 100°C for 5 minutes and snap-cooled on ice for 5 minutes. All of the labeled target DNA was added to 10 ml of freshly prepared Pre/Hybridization buffer (see below) and hybridized with a printed DNA macroarray. The mRNA transcripts (1 ng each) of genes LTP4 and LTP6 from *Arabidopsis thaliana* were spiked

into the purified total RNA mixture before the RNA labeling reaction and used as the positive controls for labeling reaction (Stratagene).

Design and Fabrication of Oligonucleotide-Based DNA Macroarrays

A single probe per gene DNA macroarray was designed based on the published *L. lactis* ssp. *lactis* IL1403 genome sequence (2). A Perl script that drove a modified version of Primer3 (36) was implemented to design the oligonucleotide probes (46) (Appendix A) that were synthesized by Sigma-Genosys (Austin, Texas). The criteria used for oligonucleotide candidate selection were length (22-24mer), melting temperature (63-65°C), GC percentage (40-60%), the absence of significant secondary structure, and at least four mismatches with any sequence in the genome. The sequences and some characteristics of the 384 oligonucleotide probes used in this study are available in the supplemented web site (http://labgenome.usu.edu/macroarray/). The 375 probes were selected because their correspondent gene targets were annotated to have roles in protein degradation pathways and carbohydrate, energy, fatty acid, nucleic acid, and amino acid metabolisms (2). An additional 9 probes consisted of 4 positive control probes designed from the spiked *Arabidopsis thaliana* genes and 5 negative control probes (an empty spot, a random 15mer probe, a probe designed from reverse complementary strand of *dnaK*, and probes from the plasmid-encoded genes *prtP* and *prtM*).

A polyinosine (polyI) tail was added to each oligonucleotide using a terminal transferase. Each tailing reaction mixture was composed of 100 pmol oligonucleotide, 100 nmol dITP (Roche Applied Science, Indianapolis, Indiana), 20 U of terminal transferase (New England BioLabs, Beverly, Massachusetts), 5 mM $CoCl_2$, 0.2 M

potassium cacodylate, 0.25 mg/ml bovine serum albumin, and 25 mM Tris-HCl (pH 6.6). The tailing reactions were conducted at 37°C for 2 hours. Tailed oligonucleotides were spotted in duplicate onto a positively charged nylon membrane (Millipore, Bedford, Massachusetts) with a manual arrayer (VP Scientific, San Diego, California, see Figure E-1) and fixed by baking the membrane at 80°C for 30 minutes. The printed arrays were stored at room temperature until used.

Hybridization and Detection

DNA macroarrays were pre-hybridized for at least 1 hour at 52°C in a 15 ml Pre/Hybridization solution composed of 5X SSC, 0.1% N-laurosylsarcosine, 0.02% SDS, and 1X Blocking reagent (VectorLab, San Diego, California). After pre-hybridization, 10 ml of fresh hybridization solution was mixed with labeled DNA targets and added to the hybridization tube to replace the pre-hybridization solution. After hybridization at 52°C overnight, the membrane was washed twice at 50°C for 5 minutes with a 2X washing solution, containing 2X SSC and 0.01% SDS. The stringent wash was conducted twice with a 0.5X washing solution (0.5X SSC and 0.01% SDS) at 50°C for 15 minutes. The biotin-based chemiluminescent detection was conducted with a North2South Hybridization and Detection kit (Pierce, Rockford, Illinois) according to the kit instructions, except that blocking time was extended from 15 minutes to 1 hour. Each membrane array was exposed to chemiluminescence film (Roche Applied Science, Indianapolis, Indiana) for 20 and 60 minutes at room temperature.

Data Analysis of Gene Expression Data

The exposed films were digitized into 16-bit TIFF images with a desktop scanner at 800 dpi (Expression 1600, Espon, Long Beach, California). The images were processed with an ImageJ plugin using a manual gridding procedure (Appendix B) and a threshold-based spot segmentation algorithm (46, 47). The background-corrected intensity data from the images exposed for 20 and 60 minutes were combined using a linearity correction procedure (46, 47). Data from different experiments were calibrated using an adapted *lowess*-based procedure (48) with a reference data set generated by averaging the expression intensity of each gene over control and treatment conditions. The statistical significance of differential gene expression was calculated with a statistical package called SAM (41).

Results

Oligonucleotide-Based DNA Macroarray Protocol

Design of Oligonucleotide Probes for DNA Array

Proper design of oligonucleotide probes was critical, especially for the probe specificity to the targeted gene. Because of the small genome size of *L. lactis* ssp. *lactis* IL1403 (2.4 MB), very stringent oligonucleotide probe design criteria were applied. A genome-wise search for homologs was conducted for each oligonucleotide candidate to guarantee its specificity for the targeted gene. The specificity of the oligonucleotide probe was verified in the array hybridization results using the *arc*C and *arcD* genes.

These genes have three and two homologs in the IL1403 genome, respectively (2). The oligonucleotide probes designed in this study easily differentiated the expression of each gene from other paralogues (Figure 3-1), indicating that the probes were sufficiently specific to minimize false positives during hybridization.

Effect of Polyinosine Tails on Oligonucleotide Probe Hybridization

Addition of polyinosine (polyI) tails to the oligonucleotides significantly increased the hybridization efficiency of oligonucleotide probes. Oligonucleotide probes with a polyI tail hybridized the target and yielded a detectable hybridization signal, while the same probes without a polyI tail did not (Figure 3-2). In addition, we also demonstrated that with a polyI tail added, 23-mer oligonucleotide probes performed as well as a 40-mer oligonucleotide probe (Figure 3-2). The shortest PolyI-labeled probe successfully tested in this study was an 18-mer (data not shown). It generated a similar hybridization signal intensity as the 40-mer oligonucleotide probe when polyI tails were added to both probes.

Indirect Labeling of the cDNA Targets with Biotin

Initially, biotin was incorporated into cDNA using reverse transcription with biotin-16-dUTP. However, this method incorporated relatively small amounts of biotin in each cDNA molecule (Figure E-2), making it difficult to detect the low abundance transcripts, even with the addition of polyI tails to the oligonucleotide probes. To increase the signal strength, an indirect high-density biotin labeling method was devised by adapting two published protocols (7, 50). First, a small amount of ddATP was

incorporated into the dNTP pool (7). Second, dTTP in the reaction mixture was partially replaced with aminoallyl-dUTP. Subsequently, biotin was chemically coupled onto the incorporated amino groups on the cDNA molecules via sulfo-NHS esters (18, 50). With this high-density biotin labeling method, we achieved detection of low abundance transcripts, such as basal expression of arginine deiminase pathway genes in *L. lactis* ssp. *lactis* IL1403, which were undetectable previously (6) (Figure 3-1).

L. lactis ssp. *lactis* IL1403 under Environmental Stresses

In order to test the performance and demonstrate the utility of the newly devised expression array protocol, we profiled the gene expression of *L. lactis* ssp. *lactis* IL1403 under three well-studied environmental stresses: heat, acid and osmotic stresses. The DNA macroarray contained 375 oligonucleotide probes corresponding to the primary metabolic genes of *L. lactis* ssp. *lactis* IL1403. The investigated genes were selected based on their annotated roles in carbohydrate, energy, fatty acid, nucleic acid, and amino acid metabolism and protein degradation pathways (2, 19).

During the 30-minute incubation used for the stress conditions, the cell density of non-stressed control culture increased about 26%, while cell densities under heat, acid, and salt stresses increased 27%, 20%, and 10%, respectively. The pH of the media dropped from 7.2 to 6.2 in the control culture, to 6.4 during salt stress, and to 6.1 during heat stress. The pH of acid stress condition started at pH 5.5 and decreased to 5.1 during the 30-minute incubation period.

Representative expression array images during control, heat, acid, and osmotic shock conditions are shown in Figure 3-3. Additional array images and the calibrated

intensity data under treatments are available at a supplementary web site (http://labgenome.usu.edu/macroarray/). Table 3-1 showed the list of genes that were differentially expressed in each stress treatment relative to the control condition. Because of the statistical procedure used, this table may contain some errors (false positives), but it is impossible to know which spot is a false result. On average, the number of false positive genes reported by SAM were expected to be 0.47, 0.45, and 0.49 for heat, acid, and osmotic stress treatment, respectively.

Stress Responses During Heat Shock

During the 30-minute 42°C heat treatment, 64 of the 375 metabolic genes investigated were differentially expressed relative to the control treatment (Table 3-1). Thirty-four of those genes were positively regulated and the other 30 genes were negatively regulated. Expression of a putative amino acid ABC transporter (*yjgC-yjgD-yjgE*) increased nearly 100-fold, while aminopeptidase C (*pepC*) and the peptide transport system (*dtpT*, *optB*, *optC*, *optD*, and *optF*) were repressed approximately 20-fold. While the expression profiles of the proteolytic related genes were most actively regulated, genes involved in nucleotide salvage pathway, methionine biosynthesis, glutamine uptake and biosynthesis, arginine catabolism, and betaine uptake were also systematically regulated during heat shock stress. As expected the heat shock gene *dnaK*, which was used as an internal positive control, was induced 8-fold in this treatment.

Stress Responses During Acid Stress

Lactic acid stress altered the expression profiles of 50 genes compared to the control after 30-minute treatment (Table 3-1). Twenty-four genes were induced and

other 26 genes were repressed during the stress. As in heat stress response, the genes involved in proteolytic system were systematically regulated. However, in contrast to heat stress treatment, lactic acid stress led to repression of the amino acid transporter genes and other proteolytic genes. The expression of the β-glucoside-specific PTS system (*yedE-yedF* and *ptsH*) was repressed more than 100-fold. This was the most dramatically regulated system detected in this study. Among the genes induced during acid stress, both the choline transporter genes (*choQ-choS*) and tryptophan biosynthesis genes were induced about 20-fold. In addition, the citrate and malate fermentation genes were induced 2- to 15- fold.

Stress Responses During Osmotic Stress

During the period of 30-minute osmotic stress with 4% salt, 68 genes were differentially expressed relative to control cells (Table 3-1). Half of those genes (34 genes) were positively regulated while the others were negatively regulated. As expected, the genes involved in the compatible solute glycine-betaine transport (*busAA-busAB*) were induced 60-fold (28). The other genes induced during osmotic stress were involved in the nucleotide salvage pathway, glutamate biosynthesis, lysine biosynthesis, and peptidoglycan biosynthesis. The heat shock gene *dnaK* was also induced after osmotic stress treatment, as expected (21). The genes involved in fatty acid biosynthesis, citrate and malate fermentation, and arginine deiminase pathway were repressed. As in the other two stress responses, the gene expression of the proteolytic related proteins and the β-glucoside-specific PTS system were also repressed during osmotic stress.

Discussion

DNA array technology is one of the leading tools in functional genomic research. This technique provides a global picture of gene expression inside living cells (37). Previously, our group successfully used oligonucleotide probes (40mers) to monitor the expression profiles of genes involved in the arginine deiminase pathway in *L. lactis* with a protocol similar to a reverse northern blot (6). In this study, we further improved the sensitivity and specificity of the protocol and established a versatile, low cost oligonucleotide-based DNA array platform accessible to any laboratory.

Oligonucleotide based DNA arrays have many advantages over amplicon-based arrays (35). Amplicon arrays have problems with cross-hybridization between genes that are very similar, such as paralogues. For example, the expression levels of the alcohol dehydrogenase genes 1 and 2 (*adh1* and *adh2*) in yeast were differentiated using an oligonucleotide-based array (9), but were indistinguishable in a cDNA array because of their high level of sequence homology (88%) (8). To minimize the possibility of cross-hybridization, we adopted a very conservative criterion for probe selection, 4 mismatches in 22-24mer oligonucleotide probes. Three mismatch in a 30mer oligonucleotide probe is sufficient to differentiate the expression of genes with >90% identity (33). The short length of oligonucleotides (22-24mer) was also helpful to enhance the probe specificity. The shorter the oligonucleotide probes, the more specific they are because longer oligonucleotides are more likely to tolerate mismatches (35). In addition, use of an oligonucleotide as a probe made it possible to normalize probe characteristics, so that more stringent hybridization conditions were used to reduce possible false positive

signals without the risk of uneven hybridization among probes with different thermodynamic properties in the same array.

The major difficulty associated with oligonucleotide probes is the relative weak signal intensities. This in turn limits the sensitivity of oligonucleotide-based DNA arrays. Common practice to overcome this limitation is to use longer oligonucleotide (MWG's Pan® Arrays and Operon's Array-Ready Oligo Sets™). It is believed that a 70-mer is the minimal length required to generate sufficient signal intensities with oligonucleotide probes (3). In contrast, the protocol developed in this study was not dependent on long oligonucleotides (Figure 3-2) if a PolyI tail was added. The success of this oligonucleotide-based DNA array relied on PolyI tailing and indirect high-density biotin labeling.

Steric hindrance of probes is a primary source of low hybridization efficiency for surface-bound oligonucleotide hybridization. It is common to add a chemically inert spacer between the probe and the immobilization surface, so that the probes are extended into the solution and available for hybridization (44). Presumably, the polyI tails served a similar role in this study. Oligonucleotide probes with polynucleotide tail, such as polythymidine were first adopted for use in reverse dot-blot procedures to increase hybridization efficiency between the immobilized oligonucleotide probes and PCR products (4, 20). Oligonucleotides with a polyI tail was also used previously to screen bacterial clones printed on nylon membrane (40), a reverse application of our protocol.

PolyI tails reduced the strong effect of oligonucleotide length on signal intensity (Figure 3-2), commonly observed in other studies (3, 18, 39). This indicated that there was other roles played by polyI tails, in addition to as a spacer. We hypothesized that the

polyI tail in our protocol also functioned as a stabilizer for the DNA duplex formed between oligonucleotide probe and target cDNA molecules. In our model (Figure 3-4), the probe is divided into a short, specific head (oligonucleotide probe) and a long non-specific PolyI tail. While the head interacts with the target molecules in a sequence-dependent manner, the immobilized tail binds the rest of target sequence-independently. That is, the polyI tails in our protocol act as a nonspecific DNA hybrid stabilizer, similar to the tail portion of a long oligonucleotide as a specific hybrid stabilizer. Although inosine is a universal base that is compatible to any of bases, the hybrid formed is weaker than true Watson-Crick base pairing (24) and can be easily eliminated by stringent washes if there are no additional hybrid formed at specific head. This also explained the low background in our results (Figure 3-3). In principle, the engineered universal bases, such as 3-nitropyrrole 2'-deoxynucloside and 5-nitroindole 2'-deoxynucleoside, should have a similar or even better performance than inosine (24).

The second aspect that allowed success of this procedure was related to use of modified dNTP mixture containing ddATP and aminoallyl-dUTP. The reverse transcription with random primers has a stronger tendency for hot-spot priming and bias for longer RNA transcripts (7). Addition of dideoxynucleotide suppressed this bias because of the elongation termination effect, and increased the efficiency and reproducibility of reverse transcription (7). To overcome the problems associated with the bulkiness of biotin, an adapted indirect labeling procedure was used to achieve the high density labeling. The indirect labeling procedure was originally developed to efficiently incorporate fluorescent dye into cDNA (34). This method has many advantages over traditional direct reverse transcription labeling procedures, such as high

labeling efficiency, low cost and no dye bias (49). In this work, we adapted this method to incorporate biotin into cDNA.

Compared to the current generation of oligonucleotide or amplicon arrays, this oligonucleotide-based macroarray was easy to set up, convenient, highly specific, and extremely cost-effective. The whole protocol is similar to the reverse dot-blot procedures (20) and can be used by anyone familiar with Southern or Northern blotting techniques. It requires no specialized instruments that other DNA microarray techniques require. The results from chemiluminescence detection are acquired using a standard desktop scanner instead of a specific laser-directed fluorescence slide reader or confocal microscope. Unlike fluorescence-based detection, with proper spot layout an unlimited linear dynamic range can be achieved using the chemiluminescence detection (46, 47).

The rationale for the use of stress instead of other experimental conditions to validate the array procedure was that stress responses in *L. lactis* were intensively studied in the last decade (32). The literature is especially rich in heat, acid and salt stresses since these are the stresses that starter culture would regularly encounter in fermentations, especailly cheese making (11, 15, 16, 31). The rich source of literature provided us an opportunity to confirm our results with previous published genetic and physiological data, and also provided direct evidence to validate our array protocol.

Because of the massive parallel comparisons, pairwise comparison of expression array data may lead to large numbers of false positive results due to the "multiple testing" problem (10). For example, in an array with 1,000 spots, suppose there are 50 corresponding genes expressed that are truly different in control and treatment. With a Type I error rate at 0.05 per comparison, we would expect to detect $50+0.05\times(1,000-50)$

= ~98 differentially expressed genes. That is, almost 50% of the detected genes would be false positives. A software package called SAM was used to control the rate of false positives (41). This software is more sensitive on significance assignment than other statistical methods, but will introduce false positives into the analysis result. In order to reduce false positives, we adopted stringent criteria to assign significance, such as low false discovery rate 1% and a minimal 2-fold change in expression levels. In general, we expected that each result from expression profile comparison between stress and control contained an average of <1 false positive or at most 3-5 false positives.

In order to estimate the validity and performance of the array, we compiled the data on heat, acid and salt stress responses of *L. lactis* from the literature and compared them with the results from DNA macroarray (Table 3-2). The majority of the results (92%) from our expression profiling experiments were consistent with previous studies. For example, the internal control gene (*dnaK*) was induced 8-fold after heat treatment and induced ~2-fold by acid treatment. Previous studies based on Northern blots or two-dimensional gels showed that *dna*K was induced around 8- to 10- fold after 30-minute heat shock (14, 21) and only 2.1-fold by acid treatment (15).

The only case where there was inconsistency between our result and literature was the regulation of *deoB* gene during acid stress (Table 3-2). The *deoB* gene encoded phosphopentomutase, a key enzyme in purine salvage pathway. Insertion mutant *deoB* of MG1363 exhibited enhanced survival rates on both heat and acid stress treatments (31). However, in IL1403 we observed that *deoB* was induced after acid stress, instead of repressed as it was during heat stress. This inconsistency could be attributed to the strain variation between MG1363 and IL1403. Previous studies showed that lactic acid

resistance in strain MG1363 is chloramphenicol-sensitive (32), while it is not in strain IL1403 (15).

The close agreement between our results and literature gave us a confidence in the performance of the new DNA macroarray protocol. Based on these observations, we concluded that the protocol was functioning as expected and provided valid results for the stress response study in IL1403. In the next section, we focus on the discussion of the various stress responses of *L. lactis* ssp. *lactis* IL1403 that are allowed because of the global nature of the array information. Roughly in the order of carbohydrate metabolism, fatty acid metabolism, nucleotide metabolism, protein degradation and amino acids metabolism, and others.

The majority of genes in sugar metabolism and the glycolysis pathway were insensitive to stress treatments, although the β-glucoside specific PTS system genes were significantly repressed in all stress conditions. The secondary fermentation pathway had a major role in acid resistance in IL1403. The genes involved in converting citrate and malate to pyruvate (*citE*, *citF*, and *mleS*) were significantly induced during acid stress, as was lactate dehydrogenase gene (*ldh*), which ferments the pyruvate to lactate. The physiological importance of the induction of these two pathways during acid stress was attributed to the exchange of the divalent substrate citrate or malate in the medium with the monovalent fermentation product lactate (25, 30). This exchange generates a membrane potential and maintains a pH gradient across the membrane. It is noteworthy that although the plasmid-encoded *citP* gene was missing in IL1403, yet the regulation circuit fir citrate fermentation was still functioning, suggesting the lack of feedback regulation for this pathway or other unknown citrate transporters was used.

Osmotic stress resulted in significant repression ($p<0.0001$) of the genes involved in fatty acid synthesis. Previously, it was reported that fatty acid composition in the cell membrane of *L. lactis* changes during osmotic stress (13). In the IL1403 genome, fatty acid synthesis genes are organized as a multi-gene cluster with the order of *fabH*, *acpA*, *fabD*, *fabG1*, *fabF*, *accB*, *fabZ2*, *accC*, *accD*, and *accA* (2). Seven of these ten genes were repressed during salt stress, suggesting that they are likely organized as a single transcriptional unit. The activity of key enzyme of this pathway, acetyl coenzyme A carboxylase complex (*accA*, *accB*, *accC*, and *accD*), is tightly coupled with the cell growth control in *E. coli* and *B. subtilis* (23, 26). We suspect that suppression of fatty acid biosynthesis might be the result of growth inhibition by 4% NaCl, because the growth of the salt stressed IL1403 culture was significantly ($p<0.05$) slower than the control culture and cultures under other environmental stresses. Interestingly, the peptidoglycan biosynthesis genes were induced during osmotic stress. It is possible that the cells are attempting to change the cell wall structure, so that it is less susceptible to autolysis during osmotic stress (43).

Although the expression of the majority of genes involved in *de novo* synthesis of purine and pyrimidine ribonucleotides were insensitive to the environmental stresses, the expression of nucleotide salvage pathway genes were influenced by the environmental stresses. The transcriptional regulation of purine metabolism and the upstream pentose phosphate pathway genes have important implications in stress resistance in *L. lactis* ssp. *cremoris* MG1363 (11, 31). With transposon mutagenesis, insertion mutants of these genes in MG1363 (*hpt*, *relA*, *guaA*, *deoB*, and *tkt*) were shown to be resistant heat and/or acid stress treatments (11, 31). It is believed that the mutation of these genes altered the

intracellular (p)ppGpp pool and induced the stringent response, which conferred lactococci with general protection against multiple stresses (11, 31). With the exception of acid stress, IL1403 seemed to comply with this regulation scheme, where *deoB* was repressed with heat stress, and *tktA* was repressed with salt stress. In addition, ribose-phosphate pyrophosphokinase (*prsA-prsB* operon) was induced during salt stress. Since these enzymes share the substrate ribose-5-phosphate with phosphopentomutase (*deoB*), induction of these two genes would have a similar effect as repressing the expression of *deoB*. The genes in the pyrimidine salvage pathway were induced in all of the stress conditions.

Proteolysis is essential for supplying *L. lactis* with amino acids during growth in the complex media (22). The proteolysis system in lactococci includes a) a cell envelope proteinase (*prtP*, plasmid-encoded), b) transport systems for amino acids and peptides, and c) intracellular peptidases (22). In addition to lacking the plasmid-encoded *prtP*, IL1403 is also different from other lactococci in its peptide uptake system. In this study, IL1403 used the Opt system as the primary peptide uptake system, since the promoter of *opp* operon [encoding the primary oligopeptide transport system in other lactococci (22)] only had a limited activity in IL1403 (Chapter 4). While peptide transporter genes (*opt* and *dtpT*) were similarly repressed in all three stresses, the amino acid transporter genes behaved differently in each stress condition. With the exception of *glnP-glnQ* and *arcD1-arcD2* (see below), the amino acid transporter genes were strongly induced during heat stress, but repressed or unchanged in other stress treatments.

The expression of glutamine biosynthesis and transport system genes (*glnA*, and *glnP-glnQ* operon) were repressed during heat and acid stresses. The intracellular

glutamine pool is the central control point for Gram-positive bacteria to control nitrogen flow in metabolism (38). Disruption of glutamine transporter genes (*glnP-glnQ*) conferred a strong lactic acid resistance in the lactococci (31). After the osmotic stress treatment, expression of glutamate synthase genes (the *gltB-gltD* operon) were induced. In *Bacillus cereus*, glutamate functions as a compatible solute in balancing of external osmotic strength in the growth medium (38). It is logical that simultaneous suppression of glutamine biosynthesis and uptake would lead to glutamine deprivation and activate the global nitrogen starvation protection in lactococci, but additional physiological evidence is needed to support this speculaton.

One of the most unexpected findings in this study was the transcriptional response of sulfur amino acid metabolic genes to the environmental stresses. In heat shock, the *metA-metB1* and *metB2-cysK* operons in methionine and cysteine biosynthesis were induced 25- and 5-fold, respectively. In contrast, the methionine synthase (*metE*) was repressed in acid stress, while methionine catabolism gene *metK* was induced. The net result was the reduction of the internal methionine pool during acid stress. The concentration of intracellular methioine is an important parameter in cheese flavor production, because methionine is the precursor methanethiol, one of the primary flavor compounds found in Cheddar cheese (45).

L-asparginase (*ansB*) was induced more than 16-fold during heat stress, while the glutamine-dependent asparagine synthetase (*asnH*) was induced nearly 10-fold during acid stress. The asparagine biosynthesis consumes intracellular aspartate, while asparaginase produces aspartate and ammonia from asparagine. The physiological significance of this regulation in stress resistance mechanisms is unclear. However, these

observations were consistent with the regulation of methionine metabolism during heat and acid stress treatments, since asparate is a precursor for methionine biosynthsis.

The arginine deiminase pathway genes were repressed in all the stress treatments. Repression of the arginine degradation pathway during heat shock was observed in *Bacillus subtilis* as well (17). They suggested that the heat-liable transcriptional regulator AhrC was responsible for this repressing effect. The AhrC protein is the key regulator to coordinate arginine synthesis and catabolism pathways in lactococci as well (32). It is noteworthy that in IL1403 arginine deiminase pathway genes were not responsive to arginine induction in acid stress as is *L. lactis* ssp. *cremoris* ML3 (6).

Growth of *L. lactis* in the high salt environment relies on the accumulation of compatible solutes, such as glycine betaine (29). Lactococci have a limited capacity for *de novo* synthesis of these compounds and must acquire them from the environment (27). It is not surprising that both betaine and choline transport systems were induced during osmotic stress treatment. Interestingly, these compatible solute uptake genes were also strongly induced during heat and acid stresses. Previous studies demonstrated that the betaine transport activity in *L. cremoris* NCDO763 increased with high growth temperature as well as response to osmotic stress (13). This could be attributed to the thermoprotectant function of betaine. Both glycine betaine and choline protect enzymes from thermo-denaturation and help proteins to refold correctly after denaturation (5). We hypothesize that choline exerted a similar function in protecting enzymes and resisting the denaturation and misfolding introduced by low intracellular pH during lactic acid stress.

While the majority of the transcriptional responses were specific stress dependent, several groups of expression profile changes were shared among different stress responses (Figure 3-5). A number of stress responses were common in all three stress treatment, such as repression of several transporter genes (*yed* operon, *opt* operon and *arcD2*) and induction of two nucleotide kinase genes (*dukA* and *yfiG*). The stress gene *dnaK*, is known to be induced in all stress treatments investigated previously (1, 15, 21). These lines of evidence: 1) *pnpA* insertion mutant in MG1363 had high survival rate during heat stress (11); 2) *dtpT* was co-regulated with *opp* in MG1363 (12); 3) *arcA* was located in the same transcription unit with *arcD2* (2); 4) *yrcA* encoded an enzyme to hydrolyze the β-glucoside translocated inside cells by the PTS system encoded by *yedE-yedF*, and 5) *thyA* and *yfiG* encoded two enzymes in the pathway to interconvert dUMP↔dTMP↔thymidine suggest that these genes may belong to a general stress resistance regulon in lactococci.

In conclusion, we successfully developed a filter-based oligonucleotide DNA macroarray protocol. The success of this oligonucleotide-based DNA array was dependent on careful oligonucleotide probe design, namely PolyI tailing and indirect high-density biotin labeling. The macroarray was used to profile the transcriptional responses of *L. lactis* ssp. *lactis* IL1403 intermediary metabolism pathways during heat, acid and salt stress. Roughly 13-18% of the investigated genes were differentially expressed under each of the environmental stress treatments. In the overwhelming majority of cases (92%), the results obtained with the macroarray agreed with previously reported observations. Of the 375 metabolic genes only seven were regulated by all three

stress conditions. Use of the array allowed additional insights into the stress respons of *L. lactis* ssp. *lactis* IL1403 were found.

References

1. **Arnau, J., K. I. Sorensen, K. F. Appel, F. K. Vogensen, and K. Hammer.** 1996. Analysis of heat shock gene expression in *Lactococcus lactis* MG1363. Microbiology **142:**1685-91.

2. **Bolotin, A., P. Wincker, S. Mauger, O. Jaillon, K. Malarme, J. Weissenbach, S. D. Ehrlich, and A. Sorokin.** 2001. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. Genome Res. **11:**731-53.

3. **Bosch, J. T., C. Seidel, S. Batra, H. Lam, N. Tuason, S. Saljoughi, and R. Saul.** 2000. Poster: Validation of sequence-optimized 70 base oligonucleotides for use on DNA microarrays. The TIGR Genome Sequencing and Analysis Conference, Maimi Beach, Florida.

4. **Brown, T. J., and R. M. Anthony.** 2000. The addition of low numbers of 3' thymine bases can be used to improve the hybridization signal of oligonucleotides for use within arrays on nylon supports. J. Microbiol Methods **42:**203-7.

5. **Caldas, T., N. Demont-Caulet, A. Ghazi, and G. Richarme.** 1999. Thermoprotection by glycine betaine and choline. Microbiology **145:**2543-8.

6. **Chou, L.** 2002. Ph.D. dissertation. Utah State University, Logan, Utah.

7. **Decraene, C., I. Reguigne-Arnould, C. Auffray, and G. Pietu.** 1999. Reverse transcription in the presence of dideoxynucleotides to increase the sensitivity of expression monitoring with cDNA arrays. Biotechniques **27:**962-6.

8. **DeRisi, J. L., V. R. Iyer, and P. O. Brown.** 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science **278:**680-6.

9. **Dorris, D., R. Ramakrishnan, T. Sendera, and S. Magnuson.** 2002. Oligonucleotide array technologies for gene expression profiling. p. 25-38. *In* E. Grigorenko (ed.), DNA arrays: Technologies and experimental strategies. CRC Press, New York, N.Y.

10. **Dudoit, S., Y. H. Yang, M. J. Callow, and T. P. Speed.** 2002. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Statist. Sinica. **12:**111-139.

11. **Duwat, P., S. D. Ehrlich, and A. Gruss.** 1999. Effects of metabolic flux on stress response pathways in *Lactococcus lactis*. Mol. Microbiol. **31:**845-58.

12. **Guedon, E., P. Renault, S. D. Ehrlich, and C. Delorme.** 2001. Transcriptional pattern of genes coding for the proteolytic system of *Lactococcus lactis* and evidence for coordinated regulation of key enzymes by peptide supply. J. Bacteriol. **183:**3614-22.

13. **Guillot, A., D. Obis, and M. Y. Mistou.** 2000. Fatty acid membrane composition and activation of glycine-betaine transport in *Lactococcus lactis* subjected to osmotic stress. Int. J. Food Microbiol. **55:**47-51.

14. **Hansen, M. C., A. K. Nielsen, S. Molin, K. Hammer, and M. Kilstrup.** 2001. Changes in rRNA levels during stress invalidates results from mRNA blotting: fluorescence in situ rRNA hybridization permits renormalization for estimation of cellular mRNA levels. J. Bacteriol. **183:**4747-51.

15. **Hartke, A., Bouch, eacute, J. C. Giard, A. Benachour, P. Boutibonnes, and Y. Auffray.** 1996. The lactic acid stress response of *Lactococcus lactis* subsp. *lactis*. Curr. Microbiol. **33:**194-9.

16. **Hartke, A., J. Frere, P. Boutibonnes, and Y. Auffray.** 1997. Differential induction of the chaperonin GroEL and the Co-chaperonin GroES by heat, acid, and UV-irradiation in *Lactococcus lactis* subsp. *lactis*. Curr. Microbiol. **34:**23-6.

17. **Helmann, J. D., M. F. Wu, P. A. Kobel, F. J. Gamo, M. Wilson, M. M. Morshedi, M. Navre, and C. Paddon.** 2001. Global transcriptional response of *Bacillus subtilis* to heat shock. J. Bacteriol. **183:**7318-28.

18. **Hughes, T. R., M. Mao, A. R. Jones, J. Burchard, M. J. Marton, K. W. Shannon, S. M. Lefkowitz, M. Ziman, J. M. Schelter, M. R. Meyer, S. Kobayashi, C. Davis, H. Dai, Y. D. He, S. B. Stephaniants, G. Cavet, W. L. Walker, A. West, E. Coffey, D. D. Shoemaker, R. Stoughton, A. P. Blanchard, S. H. Friend, and P. S. Linsley.** 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat. Biotechnol. **19:**342-7.

19. **Kanehisa, M., S. Goto, S. Kawashima, and A. Nakaya.** 2002. The KEGG databases at GenomeNet. Nucleic Acids Res. **30:**42-6.

20. **Kawasaki, E., R. Saiki, and H. Erlich.** 1993. Genetic analysis using polymerase chain reaction-amplified DNA and immobilized oligonucleotide probes: reverse dot-blot typing. Methods Enzymol. **218:**369-81.

21. **Kilstrup, M., S. Jacobsen, K. Hammer, and F. K. Vogensen.** 1997. Induction of heat shock proteins DnaK, GroEL, and GroES by salt stress in *Lactococcus lactis*. Appl. Environ. Microbiol. **63:**1826-37.

22. **Kunji, E. R., I. Mierau, A. Hagting, B. Poolman, and W. N. Konings.** 1996. The proteolytic systems of lactic acid bacteria. Antonie Van Leeuwenhoek **70:**187-221.

23. **Li, S. J., and J. E. Cronan, Jr.** 1993. Growth rate regulation of *Escherichia coli* acetyl coenzyme A carboxylase, which catalyzes the first committed step of lipid biosynthesis. J. Bacteriol. **175:**332-40.

24. **Loakes, D.** 2001. Survey and summary: The applications of universal DNA base analogues. Nucleic Acids Res. **29:**2437-47.

25. **Magni, C., D. de Mendoza, W. N. Konings, and J. S. Lolkema.** 1999. Mechanism of citrate metabolism in *Lactococcus lactis*: resistance against lactate toxicity at low pH. J. Bacteriol. **181:**1451-7.

26. **Marini, P. E., C. A. Perez, and D. de Mendoza.** 2001. Growth-rate regulation of the *Bacillus subtilis* accBC operon encoding subunits of acetyl-CoA carboxylase, the first enzyme of fatty acid synthesis. Arch. Microbiol. **175:**234-7.

27. **Molenaar, D., A. Hagting, H. Alkema, A. J. Driessen, and W. N. Konings.** 1993. Characteristics and osmoregulatory roles of uptake systems for proline and glycine betaine in *Lactococcus lactis*. J. Bacteriol. **175:**5438-44.

28. **Obis, D., A. Guillot, J. C. Gripon, P. Renault, A. Bolotin, and M. Y. Mistou.** 1999. Genetic and biochemical characterization of a high-affinity betaine uptake system (BusA) in *Lactococcus lactis* reveals a new functional organization within bacterial ABC transporters. J. Bacteriol. **181:**6238-46.

29. **Obis, D., A. Guillot, and M. Y. Mistou.** 2001. Tolerance to high osmolality of *Lactococcus lactis* subsp. *lactis* and *cremoris* is related to the activity of a betaine transport system. FEMS Microbiol. Lett. **202:**39-44.

30. **Poolman, B., D. Molenaar, E. J. Smid, T. Ubbink, T. Abee, P. P. Renault, and W. N. Konings.** 1991. Malolactic fermentation: electrogenic malate uptake and malate/lactate antiport generate metabolic energy. J. Bacteriol. **173:**6030-7.

31. **Rallu, F., A. Gruss, S. D. Ehrlich, and E. Maguin.** 2000. Acid- and multistress-resistant mutants of *Lactococcus lactis* : identification of intracellular stress signals. Mol. Microbiol. **35:**517-28.

32. **Rallu, F., A. Gruss, and E. Maguin.** 1996. *Lactococcus lactis* and stress. Antonie Van Leeuwenhoek **70:**243-51.

33. **Ramakrishnan, R., D. Dorris, A. Lublinsky, A. Nguyen, M. Domanus, A. Prokhorova, L. Gieser, E. Touma, R. Lockner, M. Tata, X. Zhu, M. Patterson, R. Shippy, T. J. Sendera, and A. Mazumder.** 2002. An assessment of Motorola CodeLink microarray performance for gene expression profiling applications. Nucleic Acids Res. **30:**e30.

34. **Randolph, J. B., and A. S. Waggoner.** 1997. Stability, specificity and fluorescence brightness of multiply-labeled fluorescent DNA probes. Nucleic Acids Res. **25:**2923-9.

35. **Relogio, A., C. Schwager, A. Richter, W. Ansorge, and J. Valcarcel.** 2002. Optimization of oligonucleotide-based DNA microarrays. Nucleic Acids Res. **30:**e51.

36. **Rozen, S., and H. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. **132:**365-86.

37. **Schena, M., D. Shalon, R. W. Davis, and P. O. Brown.** 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270:**467-70.

38. **Schreier, H. J.** 1993. Biosynthesis of glutamine and glutamate and the assimilation of ammonia, p. 281-298. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (eds), *Bacillus subtilis* and other Gram-positive bacteria. biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.

39. **Stillman, B. A., and J. L. Tonkinson.** 2001. Expression microarray hybridization kinetics depend on length of the immobilized DNA but are independent of immobilization substrate. Anal. Biochem. **295:**149-57.

40. **Sugiyama, T., S. Ishii, K. Saito, J. Yamamoto, T. Isogai, and T. Ota.** 2000. Preparation of sensitive and specific oligonucleotide probes tailed using terminal transferase and dITP. Biotechniques **28:**486-90.

41. **Tusher, V. G., R. Tibshirani, and G. Chu.** 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA **98:**5116-21.

42. **Urbach, G.** 1995. Contribution of lactic acid bacteria to flavor compound formation in dairy products. International Dairy J. **5:**877-903.

43. **Vijaranakul, U., M. J. Nadakavukaren, B. L. de Jonge, B. J. Wilkinson, and R. K. Jayaswal.** 1995. Increased cell size and shortened peptidoglycan interpeptide bridge of NaCl-stressed *Staphylococcus aureus* and their reversal by glycine betaine. J. Bacteriol. **177:**5116-21.

44. **Walsh, M. K., X. Wang, and B. C. Weimer.** 2001. Optimizing the immobilization of single-stranded DNA onto glass beads. J. Biochem. Biophys. Methods **47:**221-31.

45. **Weimer, B., K. Seefeldt, and B. Dias.** 1999. Sulfur metabolism in bacteria associated with cheese. Antonie Van Leeuwenhoek. **76:**247-61.

46. **Xie, Y.** 2002. M.S. report. Utah State University, Logan, Utah.

47. **Xie, Y., A. Cutler, and B. Weimer.** 2002. Poster: Linearization of DNA macroarray data. The 10th International Conference on Intelligent System for Molecular Biology, Edmonton, Canada.

48. **Yang, Y. H., S. Dudoit, P. Luu, and T. P. Speed.** 2001. Normalization for cDNA microarray data, p. 24-34. *In* M. L. Bittner, Y. Chen, A. N. Dorsel, and E. R. Dougherty (eds.), Microarrays: Optical technologies and informatics, Proceedings of SPIE, vol. 4266. The International Society for Optical Engineering. Bellingham, Washington.

49. **Ye, R. W., T. Wang, L. Bedzyk, and K. M. Croker.** 2001. Applications of DNA microarrays in microbial systems. J. Microbiol Methods **47:**257-72.

50. **Yu, J., M. I. Othman, R. Farjo, S. Zareparsi, S. P. MacNee, S. Yoshida, and A. Swaroop.** 2002. Evaluation and optimization of procedures for target labeling and hybridization of cDNA microarrays. Mol. Vis. **8:**130-7.

Table 3-1. Responses of *L. lactis* spp. *lactis* IL1403 to stress treatments[a].

Pathway	Gene	Fold of change			Annotated function (2, 19)
		Heat[b,e]	Acid[c]	Salt[d]	
Phosphotransferase system (PTS)	*ptsH*		0.35		phosphocarrier protein Hpr
	ptsI	0.46			phosphotransferase system, enzyme I [EC:2.7.3.9]
	yedE		0.003	0.07	PTS system, beta-glucosides-specific IIA component, putative [EC:2.7.1.69]
	yedF	0.41	0.01	0.15	beta-glucoside-specific PTS system IIBC component [EC:2.7.1.69]
Sugar metabolism	*pmi*			0.18	mannose-6-phosphate isomerase [EC:5.3.1.8]
	uxuB		14.89		fructuronate reductase [EC:1.1.1.57]
	yrcA	0.22	0.46		phospho-beta-glucosidase [EC:3.2.1.86]
Glycolysis / Gluconeogenesis	*enoA*			0.45	enolase [EC:4.2.1.11]
	enoB			3.29	2-phosphoglycerate dehydratase [EC:4.2.1.11]
	fbp		0.07	0.05	fructose-1,6-bisphosphatase [EC:3.1.3.11]
	gapB	0.30		0.26	glyceraldehyde 3-phosphate dehydrogenase [EC:1.2.1.12]
Pentose phosphate pathway	*gnd*	0.19		0.28	6-phosphogluconate dehydrogenase, decarboxylating [EC:1.1.1.44]
	tkt			0.38	transketolase [EC:2.2.1.1]
	zwf			3.96	glucose-6-phosphate 1-dehydrogenase [EC:1.1.1.49]
Pyruvate metabolism	*citE*		2.83	0.07	citrate lyase beta chain [EC:4.1.3.6]
	citF		14.31		citrate lyase alpha chain [EC:4.1.3.6 2.8.3.10]
	ldh	0.49	2.31		L-lactate dehydrogenase [EC:1.1.1.27]
	mae			0.10	malate oxidoreductase [EC:1.1.1.38]
	mleS	0.03	2.91	0.36	malolactic enzyme [EC:1.1.1.38]
Fatty acid biosynthesis	*accA*		0.32	0.03	acetyl-CoA carboxylase carboxyl transferase subunit alpha [EC:6.4.1.2]
	accB			0.21	acetyl-CoA carboxylase biotin carboxyl carrier protein
	accC			0.02	biotin carboxylase [EC:6.3.4.14]

	fabD			0.43	malonyl CoA-acyl carrier protein transacylase [EC:2.3.1.39]
	fabG1			0.27	3-oxoacyl-[acyl-carrier protein] reductase [EC:1.1.1.100]
	fabH			0.17	3-oxoacyl-[acyl-carrier-protein] synthase III [EC:2.3.1.41]
	fabZ2			0.17	(3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase [EC:4.2.1.-]
Fatty acid catabolism	*fadA*	4.16	24.67		acetyl coenzyme A acetyltransferase [EC:2.3.1.9]
	lplL	0.05			lipoate-protein ligase [EC:6.-.-.-]
	thiL			3.80	acetyl coenzyme A acetyltransferase [EC:2.3.1.9]
Deoxyribonucleotide and ribonucleotide interconversions	*nrdE*	0.11		4.12	ribonucleoside-diphosphate reductase alpha chain [EC:1.17.4.1]
	nrdF	0.13			ribonucleoside-diphosphate reductase beta chain [EC:1.17.4.1]
Deoxyribonucleotide biosynthesis	*dukA*	21.92	4.27	6.19	deoxypurine kinase [EC:2.7.1.113]
	dut			10.29	deoxyuridine 5'-triphosphate nucleotidhydrolase [EC:3.6.1.23]
	thyA	2.75		5.72	thymidylate synthase [EC:2.1.1.45]
	yeaB			14.65	dTMP kinase [EC:2.7.4.9]
Purine ribonucleotide biosynthesis	*purA*		0.04		adenylosuccinate synthase [EC:6.3.4.4]
	purL	0.15		0.04	phosphoribosylformylglycinamidine synthase II [EC:6.3.5.3]
	purN	8.77			phosphoribosylglycinamide formyltransferase [EC:2.1.2.2]
Pyrimidine ribonucleotide biosynthesis	*pydA*	4.54			dihydroorotate dehydrogenase A [EC:1.3.3.1]
Salvage of nucleosides and nucleotides	*add*	3.26	0.08		adenosine deaminase [EC:3.5.4.4]
	deoB	0.43	3.45		phosphopentomutase [EC:5.4.2.7]
	prsA			2.35	ribose-phosphate pyrophosphokinase [EC:2.7.6.1]
	prsB			8.21	ribose-phosphate pyrophosphokinase [EC:2.7.6.1]
	udp	22.69			uridine phosphorylase [EC:2.4.2.3]
	upp	4.91			uracil phosphoribosyltransferase [EC:2.4.2.9]
	yfiG	9.39	6.39	6.21	thymidine kinase [EC:2.7.1.21]
Other ribonucleotide metabolism	*pnpA*		0.13	0.25	polyribonucleotide nucleotidyltransferase [EC:2.7.7.8]

	relA			2.11	GTP pyrophosphokinase [EC:2.7.6.5]
	ytfB	9.92			ADP-ribose pyrophosphatase [EC:3.6.1.13]
Aromatic amino acid biosynthesis	*aroA*			0.16	3-phosphoshikimate 1-carboxyvinyltransferase [EC:2.5.1.19]
	aroD		3.23		3-dehydroquinate dehydratase [EC:4.2.1.10]
	aroF			3.31	phospho-2-dehydro-3-deoxyheptonate aldolase [EC:4.1.2.15]
	trpF		19.53		phosphorybosyl-anthranilate isomerase [EC:5.3.1.24]
	trpG		6.12		anthranilate synthase component II [EC:4.1.3.27]
Aspartate family amino acid biosynthesis	*ansB*	16.58			L-asparaginase [EC:3.5.1.1]
	asnH		9.39		asparagine synthetase [EC:6.3.5.4]
	dapA			4.07	dihydrodipicolinate synthase [EC:4.2.1.52]
	dapB		2.62		dihydrodipicolinate reductase [EC:1.3.1.26]
	hom		0.28		homoserine dehydrogenase [EC:1.1.1.3]
	lysA			2.35	diaminopimelate decarboxylase [EC:4.1.1.20]
	metA	25.38			homoserine O-succinyltransferase [EC:2.3.1.46]
	metB1	24.04			cystathionine gamma-synthase [EC:4.2.99.9]
	metB2	5.74			cystathionine beta-lyase [EC:4.4.1.8]
	metE		0.26		5-methionine synthase [EC:2.1.1.14]
	thrC		0.32		threonine synthase [EC:4.2.99.2]
	ychH		0.12		2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase [EC:2.3.1.117]
Branch chain amino acid biosynthesis	*ilvC*		6.61		ketol-acid reductoisomerase [EC:1.1.1.86]
Glutamate family amino acid metabolism	*glnA*	0.05	0.27		glutamine synthetase [EC:6.3.1.2]
	gltB			3.97	glutamate synthase (NADPH) large chain [EC:1.4.1.13]
	gltD	0.36		2.66	glutamate synthase (NADPH) small chain [EC:1.4.1.13]
Histidine biosynthesis	*hisH*			0.16	amidotransferase [EC:2.4.2.-]
Serine family amino acid biosynthesis	*cysE*			5.50	serine acetyltransferase [EC:2.3.1.30]
	cysK	5.09			cysteine synthase [EC:4.2.99.8]

Amino acids catabolism	*araT*	0.38			aromatic amino acid specific aminotransferase [EC:2.6.1.-]
	arcA	0.15		0.02	arginine deiminase [EC:3.5.3.6]
	arcD1	0.04			arginine/ornitine antiporter
	arcD2	0.32	0.42	0.34	arginine/ornitine antiporter
	glmS			10.27	glucosamine--fructose-6-phosphate aminotransferase (isomerizing) [EC:2.6.1.16]
	metK		2.29		S-adenosylmethionine synthetase [EC:2.5.1.6]
	sdaB		6.03		L-serine dehydratase beta subunit [EC:4.2.1.13]
	yeiG		3.72		putative aminotransferase [EC:2.6.1.-]
	ytjE	33.36			aminotransferase [EC:2.6.1.-]
Peptidoglycan and polysaccharide biosynthesis	*dltA*	2.50			D-alanine activating enzyme [EC:6.3.2.-]
	murF			3.25	UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate--D- alanyl- D-alanyl ligase [EC:6.3.2.15]
	murG			7.47	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase [EC:2.4.1.-]
	murI	3.70			glutamate racemase [EC:5.1.1.3]
	racD	0.43			aspartate racemase [EC:5.1.1.13]
Menaquinone and ubiquinone biosynthesis	*hmcM*	0.09	0.30		hydroxymethylglutaryl-CoA synthase [EC:4.1.3.5]
	menD		4.69		2-oxoglutarate decarboxylase / 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase [EC:4.1.1.71 4.1.3.-]
	yhdB			17.24	O-succinylbenzoate-CoA synthase [EC:4.2.1.-]
	yhfE			44.23	hypothetical methlytransferase [EC:2.1.1.-]
Glutathione and thioredoxin	*gshR*		4.77		glutathione reductase [EC:1.6.4.2]
	trxB1	3.73			thioredoxin reductase [EC:1.6.4.5]
One carbon pool by folate	*folD*			2.68	methylenetetrahydrofolate dehydrogenase (NADP+) / methenyltetrahydrofolate cyclohydrolase [EC:1.5.1.5 3.5.4.9]
Folic acid biosynthesis	*pabA*			0.07	para-aminobenzoate synthetase component II [EC:4.1.3.-]
	pabB			0.12	para-aminobenzoate synthase component I [EC:4.1.3.-]

RNA polymerase	*rpoE*	2.90			DNA-directed RNA polymerase delta chain [EC:2.7.7.6]
RNA reaction	*truA*			19.08	tRNA pseudouridine synthase A [EC:4.2.1.70]
	truB			2.69	tRNA pseudouridine synthase B [EC:4.2.1.70]
	yljE	4.46			putative RNA methyltransferase [EC:2.1.1.-]
Aminoacyl-tRNA biosynthesis	*alaS*		5.67		alanyl-tRNA synthetase [EC:6.1.1.7]
	argS			3.01	arginyl-tRNA synthetase [EC:6.1.1.19]
	fmt			5.26	methyonyl-tRNA formyltransferase [EC:2.1.2.9]
	gatA	2.94			Glu-tRNA amidotransferase subunit A [EC:6.3.5.-]
	gltX		3.53		glutamyl-tRNA synthetase [EC:6.1.1.17]
	glyT		0.36	0.12	glycyl-tRNA synthetase beta chain [EC:6.1.1.14]
	ileS	0.29		0.10	isoleucyl-tRNA synthetase [EC:6.1.1.5]
	leuS			6.45	leucyl-tRNA synthetase [EC:6.1.1.4]
ABC transporters, prokaryotic	*busAA*	26.87		60.02	betaine ABC transporter ATP binding protein
	busAB	13.83		49.87	betaine ABC transporter permease and substrate binding protein
	choQ		7.77		choline ABC transporter ATP binding protein
	choS		24.65	2.51	choline ABC transporter permease and substrate binding protein
	potA			0.39	spermidine/putrescine ABC transporter ATP-binding protein
	potB			0.12	spermidine/putrescine ABC transporter permease protein
Amino acids transporter	*brnQ*		0.04		branch chain amino acid transporter
	glnP	0.07	0.04		glutamine ABC transporter permease and substrate binding protein
	glnQ	0.15	0.06		glutamine ABC transporter ATP-binding protein
	lysQ			0.01	lysine specific permease
	ydcC	7.04			amino acid ABC trasporter permease protein
	ydgC	3.06			amino acid permease
	yjgC	25.23			amino acid ABC transporter substrate binding protein
	yjgD	4.55	0.17		amino acid ABC transporter permease protein
	yjgE	86.80			amino acid ABC transporter ATP binding protein
	yvdF	2.45			amino acid ABC transporter substrate binding protein

Peptide transporter	*dtpT*	0.04	0.09		di-/tripeptide transporter
	optB	0.05	0.30	0.29	oligopeptide ABC trasporter permease protein
	optC	0.06			oligopeptide ABC trasporter permease protein
	optD	0.04	0.18		oligopeptide ABC trasporter ATP binding protein
	optF	0.07	0.25	0.17	oligopeptide ABC trasporter ATP binding protein
Peptidase	*gcp*	4.39			O-sialoglycoprotein endopeptidase [EC:3.4.24.57]
	pepC	0.03			aminopeptidase C [EC:3.4.22.40]
	pepDA			3.78	dipeptidase [EC:3.4.-.-]
	pepP			0.13	aminopeptidase P [EC:3.4.11.9]
	pepXP		0.22		X-prolyl dipeptidyl aminopeptidase [EC:3.4.14.11]
Protease	*yueE*	3.51			protease
	yuhB	4.71			protease
Heat shock protein	*dnaK*	8.33		10.34	DnaK protein

[a] The threshold for the software package SAM was minimal 2-fold change in expression ratio and maximal false discovery rate 1%.

[b] In heat shock responses, the median number falsely called significant genes was 0.46 with median false discovery rate 0.72%. That is, on average we expected less than one false positive in this column.

[c] In acid shock responses, the median number falsely called significant genes was 0.42 with median false discovery rate 0.85%. That is, on average we expected less than one false positive in this column.

[d] In osmotic shock responses, the median number falsely called significant genes was 0.45 with median false discovery rate 0.66%. That is, on average we expected less than one false positive in this column.

[e] One positive control significantly decreased were not shown.

Table 3-2. The literature consistency of stress responses of *L. lactis* spp. *lactis* IL1403 detected with DNA macroarray.

	heat		acid		salt			
Gene	Response[a]	Match literature[b]	Response[a]	Match literature[b]	Response[a]	Match literature[b]	Comments	References
citE			+	√	-	NA	Physiological data, but no *citP* in IL1403	(25)
citF			+	√			Physiological data, but no *citP* in IL1403	(25)
mleS	-		+	√	-	NA	Physiological data	(30)
deoB	-	√	+	X			High stress resistance from insertion mutant in MG1363	(31)
busAA	+	√			+	√	Physiological and genetic data	(13, 28, 29)
busAB	+	√			+	√	Physiological and genetic data	(13, 28, 29)
glnP	-	NA	-	√			High stress resistance from insertion mutant in MG1363	(31)
glnQ	-	NA	-	√			High stress resistance from insertion mutant in MG1363	(31)
dnaK	+	√			+	√	Northern blot, 2D Gel	(1, 21)

[a] In the response column, + symbol indicated that in this study a positive regulation on the gene expression was detected under the specified stress, while – symbol indicated a negative regulation on the gene expression.

[b] In the "match literature" column, √ symbol indicated that the result from this study was consistent with the literature, while X symbol indicated the inconsistency between the expression profiling results and the literature. NA indicated that there was no literature available for the result verification.

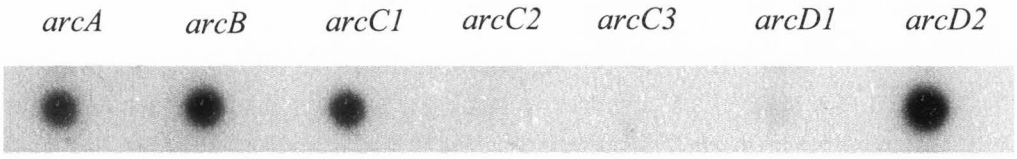


Figure 3-1. Detection of transcripts from the arginine deiminase pathway in *L. lactis* spp. *lactis* IL1403. *L. lactis* spp. *lactis* IL1403 after 30 minutes of growth in M17G at pH7.0 (i.e. control condition as in Materials and Methods). The *arcA*, *arcB*, *arcC1*, *arcC2*, *arcD1*, and *arcD2* genes encoded proteins involved in arginine deiminase pathway and belonged to *arc* gene clusters containing multiple promoter regions (6). The oligonucleotide probes for *arcA*, *arcB*, *arcC1*, *arcC2*, *arcC3*, *arcD1*, and *arcD2* genes were from probes designed for DNA macroarray (http://labgenome.usu.edu/macroarray).

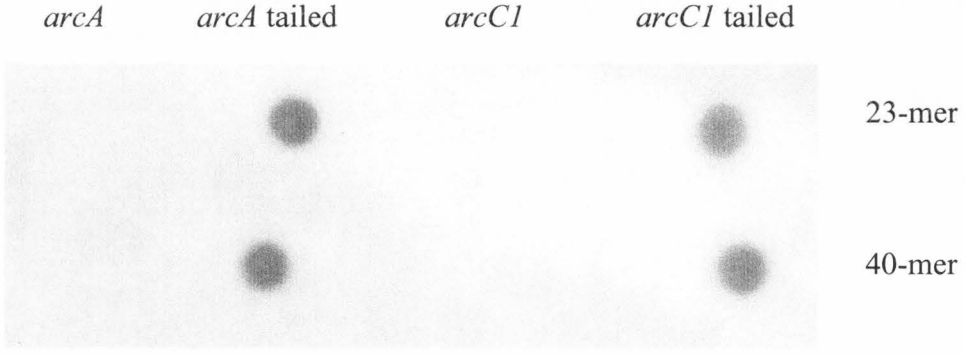


Figure 3-2. The influence of polyinosine tailing on hybridization signal. *L. lactis* ssp. *cremoris* ML3 was incubated in MOPS buffer (pH 5.0) supplemented with 2% arginine for 1 hour. Under this condition, *arcA* and *arcC1* genes were strongly induced (6). The 23-mer oligonucleotide probes of *arcA* and *arcC1* were from probes designed for DNA macroarray (http://labgenome.usu.edu/macroarray), and the 40-mer oligonucleotide probes were designed from sequences of *arc* gene clusters in ML3 (6).

A. Control

B. Heat shock

C. Acid shock

D. Osmotic shock

Figure 3-3. Examples of expression profiles of stress responses in *L. lactis* spp. *lactis* IL1403 with DNA macroarray. (A) Expression profiles in control condition; (B) Expression profiles after heat shock; (C) Expression profiles after acid shock; (D) Expression profiles after osmotic shock.

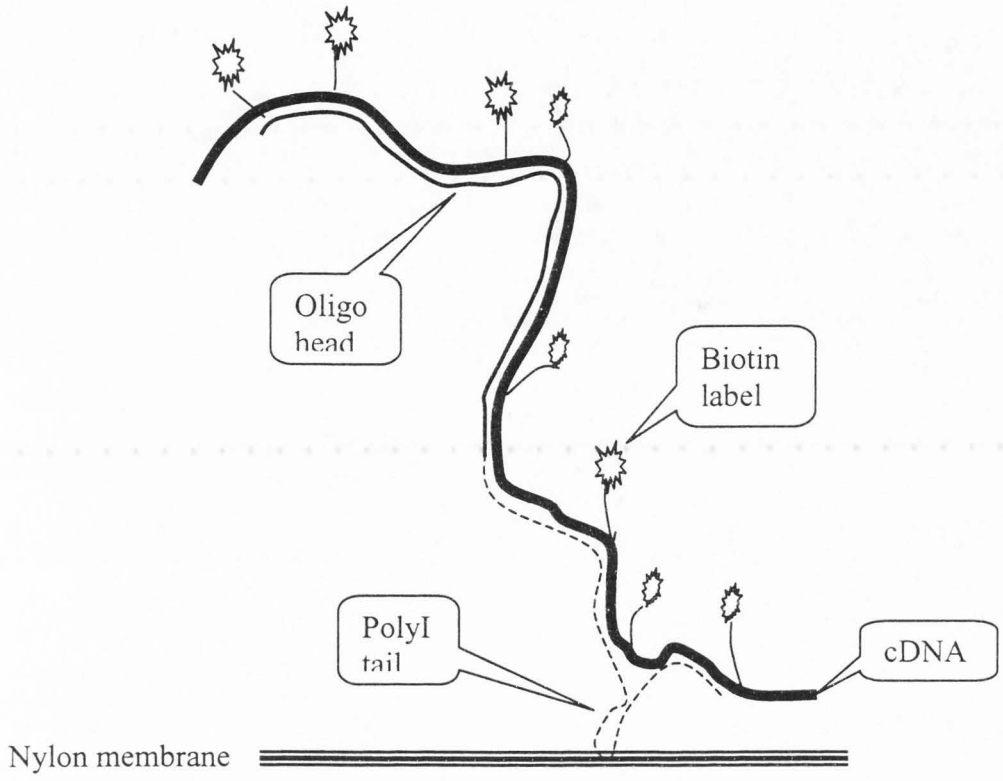


Figure 3-4. Proposed model for the effect of polyinosine tail on the high stability of hybrids. The thin curve line represents the specific oligonucleotide probe; and thin dash curve line represents the polyinosine tail that stabilized DNA duplex formed between probe and target DNA. The thick curve line is the target DNA molecules with biotin dense-labeled, which are represented by star symbols.

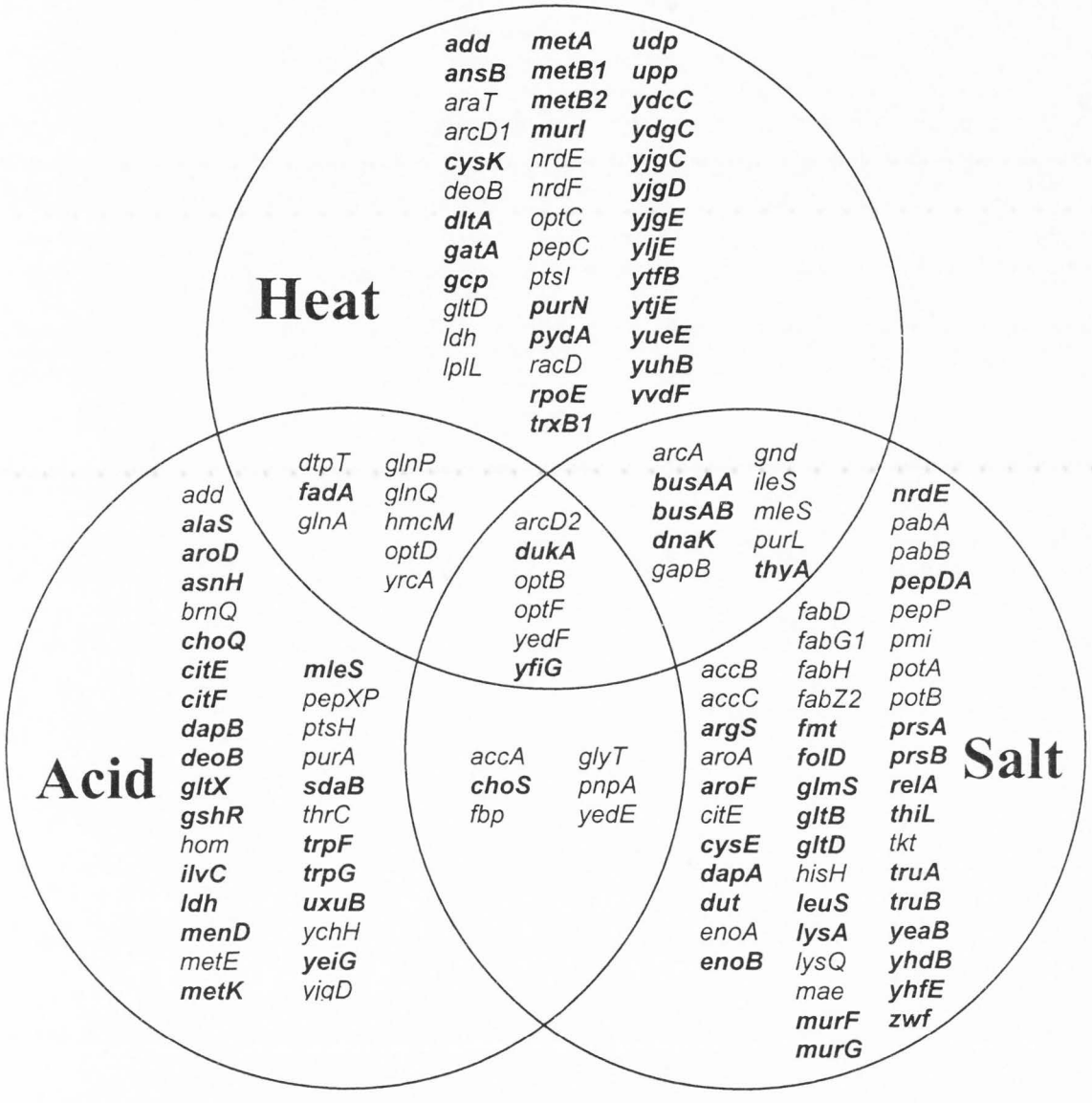


Figure 3-5. Venn diagram of shared stress responses of *L. lactis* ssp. *lactis* IL1403. The gene names were grouped according to its response to each stress treatment and present in each circle. The common stress responses were represented in the intersection area among the circles. The gene induced in the specific stress treatment was labeled with bold fonts, while repressed gene was labeled using normal fonts.

CHAPTER 4

EFFECT OF PEPTIDE SOURCES ON THE EXPRESSION PROFILE OF *LACTOCOCCUS LACTIS* SSP. *LACTIS* IL1403[1]

Abstract

Extracellular proteolytic enzymes in cheese matrix influence metabolic processes of the starter culture and therefore cheese flavor production. In this report, we studied the impact of peptide pools generated from different proteolytic processes on metabolism in lactococci. Using a DNA macroarray, we profiled the gene expression changes of *Lactococcus lactis* spp. *lactis* IL1403 growing in a peptide-limited medium, in a casitone-based peptide rich medium, and in a casein hydrolyte made from the proteolytic action of *Brevibacterium linens* BL2. When IL1403 was grown in a peptide-limited medium and casitone medium, the cells did not utilize peptides, and experienced nitrogen starvation even with abundant peptide resources available. To uptake the essential branched chain amino acids (BCAA), cells utilized a BCAA-specific permease (*brnQ*), instead of relying on peptide transporters, which were not actively expressed in these media. Conversely, a peptide pool generated by *B. linens* BL2 proteolytic activities sustained the growth of *L. lactis* ssp. *lactis* IL1403. The repression of peptide transporter genes and other proteolytic related genes was relieved in this medium. In addition, the Opt system, a di-tripeptide transporter, was used as the primary peptide transport system in IL1403, as the *opp* operon was not actively expressed.

---

[1]Coauthored by Yi Xie, Lan-szu Chou, Paul Joseph, Adele Cutler and Bart Weimer

Introduction

Lactococcus lactis was used as starter culture in cheese manufacture long before the existence of microbiology. Cheese flavor compounds are derived from the metabolic activities of the microbial community living inside the cheese matrix, especially the starter culture *L. lactis* (22). Since the catabolism of amino acids results in the formation of most volatile flavor compounds, a proper intracellular pool of amino acids must be established in order to promote a desirable flavor production in cheese. Generation of this pool of amino acids requires complex interactions among the community of bacteria in cheese, proteolytic enzymes and casein and its derivatives. In the cheese matrix, the casein is hydrolyzed by the proteolytic enzymes that come from milk, coagulant, starter bacteria, non-starter lactic acid bacteria, and adjunct microorganisms (12). Subsequently, the starter culture transports the resulting peptides and amino acids via peptide and amino acid transport systems (12). Inside the starter culture cell, peptides are further hydrolyzed into amino acids via a complex peptidases network (12).

During cheese ripening, starter cultures coexist in the cheese matrix with a variety of other microbes, including non-starter lactic acid bacteria and other flavor adjunct bacteria. *Brevibacterium linens*, as an adjunct bacterium, helps to produce a more flavorful and softer-bodied low fat Cheddar cheese (24). This indicates that there are some interactions between starter culture and brevibacteria that modulate the cheese flavor and texture development. In this study, we focused on studying the influence of extracellular proteases from *B. linens* BL2 on the proteinase deficient *L. lactis* IL1403. Using growth media containing peptides derived from various proteolytic enzymes

including BL2 protease, we compared the influence of these peptide pools on the growth and expression profiles of IL1403.

DNA array technology is used to monitor the profiles of thousands of mRNA transcripts in parallel (17). Recently, we developed a low-cost oligonucleotide-based filter DNA macroarray (Chapter 3). With this tool and reverse phase HPLC, we demonstrated that the peptide pool modulated the expression of *L. lactis* ssp. *lactis* IL1403 proteolytic genes. In one peptide rich medium containing casitone, the cells experienced nitrogen starvation even with abundant peptide resources available because they were unable to express peptide transporter genes. In contrast, the cells incubated in casein hydrolyte produced by the *B. linens* protease did expressed peptide transporter genes (except *opp*) and were able to utilize the extracellular peptides.

Materials and Methods

Bacterial Strains and Media

Lactococcus lactis ssp. *lactis* IL1403 (courtesy of Dr. Larry McKay, University of Minnesota) is a plasmid-cured strain that lacks plasmid-encoded features, such as lactose transporter, cell envelop protease, citrate transporter, and restriction/modification systems. The working cultures of IL1403 were grown at 30°C overnight without shaking in M17 broth (Difco, Sparks, Maryland) supplemented with 0.5% glucose (M17G). *Brevibacterium linens* BL2 was obtained from our culture collection. The stock culture was prepared by mixing 1 ml culture with an equal volume of sterile Tryptic Soy Broth (TSB broth, Difco), containing 30% glycerol, and stored in -70°C. Before each use, a

frozen stock culture was thawed, inoculated at 2% into TSB broth, and grown at 25°C for 48 hours with aeration.

The 220 medium was designed to optimize the production of extracellular protease in *B. linens* BL2 (23). The base 220 medium was composed of 1% yeast extract, 1% NaCl, 1% $K_2HP_4 \cdot 3H_2O$, and 0.1% glucose and was used as the peptide-limited medium in this study. The peptide rich casitone medium was composed of base 220 supplemented with 1% casitone (Difco). Another peptide rich medium, BL2 spent medium, was prepared by growing *B. linens* BL2 in base 220 medium supplemented with 1.5% glucose and 3% casein (Research Organics, Inc., Cleveland, Ohio) for 72 hours at 25°C with aeration at 250 rpm. Cells were removed from the culture medium by centrifugation (7,000 X *g* for 10 minutes at 4°C twice) and the supernatant was filter-sterilized with 0.22 µm filter system (Cornings Inc., Acton, Massachusetts). The resulted mixture was further supplemented with sterile 1% casein, and incubated at 30°C for 48 hours to allow casein hydrolysis by the mixture of proteolytic enzymes from *B. linens* BL2. Before the inoculation, 0.1% glucose was supplemented the medium to sustain the initial growth of IL1403. This medium was referred to as the BL2-peptide medium.

Growth of *L. lactis* ssp. *lactis* IL1403 in Treatment Media

Overnight cultures of *L. lactis* ssp. *lactis* IL1403 grown in M17G broth were harvested by centrifugation at 5000 x g for 10 minutes at 4°C, washed in 0.85% NaCl and then refreshed with fresh M17G broth for 10 minutes at 30°C. The cells were recollected by centrifugation at 5000 X *g* for 10 minutes at 4°C and suspended in treatment media.

The initial cell densities in the treatment media were adjusted to optical density at 600 nm (OD_{600}) of 2.0. Samples were collected after 2 hours of growth at 30°C. Peptide profiles and the total RNA was analyzed as described below. The OD_{600} and pH were also determined at the time of RNA isolation. Each experiment was replicated twice.

RP-HPLC Analysis of Peptide Profiles

The medium samples (10 ml) were prepared for HPLC analysis by centrifugation at 5,000 x *g* for 15 minutes at 4°C. The supernatant was collected and filtered with G/FA filter paper (Whatman International Ltd., England) and concentrated with a 10 K MWCO Centriprep™ (Pall Filtron, East Hill, New York). The filtrate was further concentrated to 3 ml using a speed-vacuum drier (Speedvac AES 1010; Savant Instruments Inc., Farmingdale, New York). A portion of the concentrate (1 ml) was fractionated with a 3 K MWCO Centricon ™ (Pall Filtron) using centrifugation (3000 x *g* for 3 hours at 4°C), and the permeate was analyzed by RP-HPLC. All samples were stored at -20° until use.

RP-HPLC analysis of peptide profiles was performed using a C8 column (Brownlee Aquapore RP-300, Perkin-Elmer/Applied Biosystems, Norwalk, Connecticut) on a Beckman gradient HPLC system (Beckman Instruments, Fullerton, California) as previously described (2). Peptides were eluted from the column with a linear gradient of 0.1% trifluoroacetic acid (TFA) and 0.085% TFA in 80% acetonitrile. During each run, the linear gradient of TFA/aceonitrile in TFA was increased from 0 to 50% over a 65-minute period at a flow rate of 0.2 ml/minute. The detection of peptides was done at 214 and 280 nm.

Design and Fabrication of Oligonucleotide-Based DNA Macroarray

A single probe per gene DNA macroarray was designed based on the published *L. lactis* ssp. *lactis* IL1403 genome sequence (1). A Perl script that drove a modified version of Primer3 (15) was used to design the oligonucleotide probes. In total, 384 oligonucleotides including nine controls were synthesized (Sigma-Genosys, Austin, Texas). These probes were selected because their correspondent gene targets were annotated to have roles in protein degradation pathways and carbohydrate, energy, fatty acid, nucleic acid, and amino acid metabolisms (1). The nine controls included four positive control probes designed for spiked *Arabidopsis thaliana* genes and five negative control probes, such as an empty spot, a random 15mer probe, a probe designed from a reverse complementary strain of *dnaK*, and probes from the plasmid-encoded genes *prtP* and *prtM*. Each gene-specific probe was tailed using polyinosine with terminal transferase (Roche Applied Science, Indianapolis, Indiana), and printed in duplicate on a nylon membrane (Millipore, Bedford, Massachusetts) as outlined in Figure 4-1.

Total RNA Extraction, Labeling, and Hybridization

The general procedure described in Chapter 3 was followed (Figure 4-1). Briefly, after collection from the medium IL1403 cells were treated with lysozyme and total RNA was extracted with RNaqeous kit (Ambion Inc., Austin, Texas). To further remove contaminated DNA and protein, RNA was purified with RNeasy kit (Qiagen, Valencia, California) and LiCl precipitation. Total 50 μg RNA from each sample was labeled with a high-density indirect labeling method. This labeling method relied on a non-biased

reverse transcription and an indirect labeling scheme to efficiently incorporate biotin molecules into the target cDNAs (Chapter 3). Biotin-labeled target DNA molecules were hybridized with the DNA macroarray at 52°C overnight, and washed. The hybridization was detected with a North2South Hybridization and Detection Kit (Pierce, Rockford, Illinois). The mRNA transcripts of genes from *Arabidopsis thaliana* (Stratagene, La Jolla, California) were spiked as previously described (Chapter 3).

Data Analysis

The DNA macroarray data were analyzed after the exposed films were digitized into 16-bit TIFF images with a desktop scanner (Expression 1600, Espon, Long Beach, California). The array images were processed with a custom-written ImageJ plugin using semi-automatic gridding and threshold-based spot segmentation algorithms (25, 26). After background correction, linearization, and data normalization (25, 26), the data were further analyzed for different patterns of gene expression. For array data derived from *L. lactis* ssp. *lactis* IL1403 grown on peptide-limited and peptide rich casitone media, the list of genes with intensity higher than background spots were identified using the *mt.minP* function in R package *multtest* (4). The expression profiles of *L. lactis* ssp. *lactis* IL1403 grown on BL2-peptide medium were analyzed using a statistical package called SAM (20), with the gene expression profiles of IL1403 on the peptide-limited and casitone media as references. The total peak areas in peptide utilization profiles from RP-HPLC were compared using a two-sample t test.

Results

Growth and Peptide Utilization Profiles of *L. lactis* spp. *lactis* IL1403

Over 2 hours of growth at 30°C, the cell density of *L. lactis* ssp. *lactis* IL1403 increased 247% in BL2-peptide medium, while the cell density increased 32% in peptide-limited and peptide-rich casitone media. The media pH dropped from 7.2 to 5.35 when *L. lactis* ssp. *lactis* IL1403 were grown in BL2-peptide media, but only decreased to 6.65 and 6.70 in peptide-limited or casitone medium, respectively. This clearly indicated that growth was limited in the peptide-limited or peptide rich casitone medium.

In order to understand the reason for differential growth and acid production of IL1403, we examined peptide utilization in these media (Figure 4-2). Most of the peptide peaks decreased during growth of IL1403 in BL2-produced peptide medium. In this medium, the total peptide peak area decreased significantly ($p < 0.05$) from 889 before incubation to 696 after the 2 hours growth of IL1403, indicating that peptides from casein hydrolysis by BL2 proteolytic enzymes were transported by IL1403 (Figure 4-2). During incubation of IL1403 in the peptide-limited medium, the total peptide pool did not decrease (total peptide peak area was 442 at 0 minutes and 423 at 120 minutes). Incubation of IL1403 in the casitone-supplemented casitone medium resulted in a slight, but insignificant ($p > 0.05$) increase in the total peptide pool (1223 at 0 minutes to 1324 at 120 minutes).

Expression Profiles of *L. lactis* spp. *lactis* IL1403 in the Peptide-Limited or Peptide-Rich Casitone Medium

Initial inspection of DNA macroarray images revealed that the expression profiles of IL1403 in peptide-limited and peptide rich casitone media were nearly identical (Figure E-3). After data extraction and normalization, the expression intensities of IL1403 genes in these two media were compared using the statistical package SAM (20). Only 4 of 375 spotted genes on the arrays were differentially expressed with a median false discovery rate of 25% (i.e., on average it was expected that one of these four genes was a false positive). This result confirmed that the expression profiles of *L. lactis* ssp. *lactis* IL1403 were similar to each other when grown in peptide-limited and casitone media.

Consequently, we asked which genes in IL1403 were still actively expressed in these two media. Since the majority of genes were expressed levels were identical, the expression data from both conditions were pooled for further statistical analysis. The pooled array data were compared with the spotted background controls, including an empty spot, a random 15mer probe, a probe designed from reverse complementary strand of *dnaK*, and probes from the plasmid-encoded genes *prtP* and *prtM*. Among the investigated genes, 33 genes (~9%) were expressed at significantly (adjusted p-value < 0.05) higher levels than the background (Table 4-1).

The majority of these actively expressed genes were classified into four groups according to their annotated functions (1) (Table 4-1). The first group was composed of genes involved in β-glucoside uptake and hydrolysis, including the β-glucoside PTS Enzyme II component (*yedE-yedF*, and *ptcB*) and 6-phospho-β-glucosidase (*bglA*). The

second group of genes was primarily involved in pyruvate metabolism, including enolase (*enoB*), pyruvate kinase (*pyk*), pyruvate dehyrogenase complex (*pdhA*, *pdhB*, and *pdhC*), acetolactate synthase (*als*), and α-acetolactate decarboxylase (*aldB*). The genes in the third group were associated with ribose and purine metabolism, including ribokinase (*rbsK*), transketolase (*tkt*), phosphopentomutase (*deoB*), purine-nucleoside phosphorylase (*deoD*), and ribose-phosphate pyrophosphokinase (*prsA*). The last group of genes was involved in proteolytic pathways. This group included the housekeeping protease *htrA*, aminopeptidases (*pepF*, *pepM*, *pepO*, *pepP*, and *pepT*), and two amino acid transporters (*brnQ* and *glnP*). Notably, expression of the peptide transporter genes (*opp* and *opt*) was missing, despite their presence in the genome (1). With the exception of *brnQ* and *glnP*, amino acid transporter genes were largely missing from the actively expressed gene list. These observations agreed well with the absence of peptide utilization in these peptide-limited and peptide rich casitone media.

Expression Profiles of *L. lactis* spp. *lactis* IL1403 in the BL2-peptide Medium

Contrary to growth in the other two media, fifty-eight genes were differentially expressed during growth in the BL2-peptide medium (Table 4-2). The majority of these differentially expressed genes were up regulated, as suggested by the high growth rate and acid production when IL1403 was grown in BL2-peptide medium. Three genes were repressed significantly, including *yedE-yedF* operon (β-glucosides-specific PTS system, Enzyme II components) and *enoB* (2-phosphoglycerate dehydratase).

Peptide utilization (Figure 4-2) indicated that active peptide transporters were present in IL1403 during its growth in BL2-peptide medium. In lactococci, there were

several peptide transporter systems (12). Consistent with the peptide utilization, transporter gene *optA*, *optB*, and *optD* in the *optSABCDF* operon were induced significantly in BL2-peptide medium. The induction of the primary peptide transporter *opp* was not observed. In fact, we have never observed the expression of *opp* operon in IL1403 even in optimal growth media like M17G (Figure 4-3). Two amino acid transporters with unknown specificities (*ydcB-ydcC* operon and *ydgB-ydgC*) were also significantly induced in BL2-peptide medium. In addition to transporter genes, the genes that encoded intracellular proteolytic enzymes were also induced, including aminopeptidase C (*pepC*), aminopeptidase T (*pepT*) and putative protease (*yueF*).

Glycolytic genes were significantly induced in the BL2-peptide medium, including several key genes that were not expressed previously, such as *fbaA*, *gapB*, *pgiA*, *tpiA*, and two genes of the *lac* operon (*pyk* and *ldh*). These observations were consistent with the strong acid production in this medium. As expected due to acid production, several genes involved in acid resistance were differentially expressed in this BL2-peptide medium, including repression of the *yedE-yedF* operon and induction of *citE*, *citF*, *ldh*, and *choQ*. Regulation of these genes are part of the acid stress responses in IL1403 (Chapter 3).

We observed that the *opp* operon was not expressed in IL1403, irrespective of the stress condition (Figure 4-3). Subsequent sequence analysis revealed that IL1403 contained a potential promoter P*oppD* with an imperfect -35 box. Although the sequence homology between IL1403 and *L. lactis* SSL135 was poor in this region (21) (Figure 4-4), the downstream structural genes were identical or nearly identical in the amino acid

sequence comparison with products of *oppDFBCA* operon, , 98.2%, 100%, 99.4%, 100%, and 87.8%, respectively.

Discussion

Lactococci are fastidious organisms with multiple amino acid auxotrophies. As a result, their growth is dependent on an efficient system of the protein degradation and transport of amino acids and small peptides (12). In *L. lactis* ssp. *lactis* IL1403, the plasmid-encoded cell envelope proteinase (PrtP) is missing, making it possible to study the influence of exogenous proteolytic enzymes on growth and metabolism of lactococci. In this study, we examined the effect of different extracellular peptide pools on the growth and gene expression profiles of *L. lactis* ssp. *lactis* IL1403. The results allowed us to identify the genes associated with the protein metabolic network in lactococci.

Unexpectedly, the peptide pool from casitone did not sustain the growth of IL1403 as it does with other lactococci (10). The growth of IL1403 in the casitone-based peptide rich medium was indistinguishable from that of a peptide-limited medium, and much slower than in the BL2-peptide medium. The RP-HPLC peptide profiling indicated that this lack of growth was due to the absence of peptide utilization in these media. In addition, the results from DNA macroarrays also revealed that IL1403 had nearly identical expression profiles during growth in these two media due to the lack of peptide transport gene expression (Table 4-1). These two lines of evidence lead us to conclude that the lack of growth of *L. lactis* ssp. *lactis* IL1403 in the casitone-based peptide rich medium was the result of the lack of peptide transport and intracellular peptide deficiencies that lead to nitrogen starvation in IL1403. It is possible that repression of

glycolysis and energy metabolism also had some influence on this process, but their role was unclear.

L. lactis ssp. *lactis* IL1403 is auxotrophic for branched chain amino acids (BCAA) (6). Although several genes in BCAA biosynthesis pathway are cryptic in IL1403, the transcription of *ilv* and *leu* operons are induced during BCAA starvation (6). Lack of active expression of BCAA biosynthesis genes in peptide-limited or casitone medium indicated that IL1403 was not starved for BCAA in these conditions. In addition to synthesizing BCAA, lactococci acquire these essential amino acids primarily via oligopeptide transport systems and intracellular peptidolytic enzymes during normal growth (12). In the peptide-limited and casitone-based peptide rich medium, peptide transport systems were not actively expressed, but instead *brnQ* (branch-chain amino acids permease) was one of the most actively expressed genes. We hypothesize that this transporter was responsible for the uptake of BCAA to support cellular activities in IL1403 during these conditions. Interestingly, internalized BCAA are also a signal for the transcriptional regulator CodY that represses the expression of the oligopeptide transport system, exocellular protease and peptidases in *L. lactis* ssp. *cremoris* MG1363 (11). This may explain why IL1403 did not actively express the peptide transporter genes, even in the casitone medium.

Repression of the proteolytic-related genes was relieved in BL2-peptide medium, as indicated by the induction of the peptide transport system (*optA*, *optB*, and *optD*) and peptidase genes (*pepC* and *pepT*). The peptide transport system Opt was likely responsible for the reduction of peptides from the BL2-peptide medium (Figure 4-2) since the expression of another peptide transporter genes (*opp* operon) was not observed.

The *opt* genes had a high sequence homology with recently characterized di-tripeptide ABC transporter *dpp* in *L. lactis* ssp. *cremoris* MG1363 (16). The amino acid sequence similarities between *optSABCDF* in IL1403 and *dppAdppPBCDF* in MG1363 were 95.1%, 98.5%, 95.4%, 97.1%, 99.1%, and 98.4%, respectively. This result indicated that the Opt system was likely a peptide transporter with a substrate length preference of 2-3 amino acid residues, instead of oligopeptide transporter as annotated (1). This explained the rapid growth and peptide utilization of IL1403 in the BL2-peptide medium. *Brevibacterium linens* is rich in extracellular proteolytic and peptidolytic enzymes (14). With the combined activities of these enzymes, the casein molecules were hydrolyzed into short peptides (≤ 3 amino acid residues) that were transportable by the Opt transporter of lactococci. These peptides subsequently were utilized to sustain the growth of IL1403 in this medium.

The Opp system is the primary oligopeptide transporter in lactococci (12, 21), and transports oligopeptides with lengths of 4-18 amino acids residues (3). However, we noticed that the *opp* operon was not actively expressed in IL1403 in all the conditions (Figure 4-3). Sequence analysis revealed that IL1403 had a potential promoter P*oppD* with an imperfect -35 box, implying that an activator might be required for transcription initiation. Although the sequence homology in this region between IL1403 and *L. lactis* SSL135 was poor (21) (Figure 4-4), the downstream structural genes in IL1403 and SSL135 shared a very high amino acid sequence identity, 98.2%, 100%, 99.4%, 100%, and 87.8%, with products of *oppDFBCA* operon. The *opp* operon is located close to a genome inversion site in lactococci and is structurally similar to a transposon (13). It is tempting to speculate that the P*oppD* region in IL1403 was the footprint of complex

recombination events. The recombination corrupted essential elements such as the operator sequence for the transcriptional activator, thus prevents the expression of this operon. An additional imperfect extended -10 box was located upstream of *oppA* in both IL1403 and SSL135. Comparing with the vegetative P*oppA* in *L. lactis* ssp. *cremoris* MG1363, which carries a perfect extended -10 box, this promoter probably also has limited activity in IL1403.

There are rich sources of extracellular aminopeptidases in *B. linens* (5), which could lead to a pool of extracellular amino acids available to IL1403 in BL2-peptide medium. Not surprisingly, the amino acid transporter genes (*ydcB-ydcC* and *ydgB-ydgC* operons) were strongly induced in this medium. In the meantime, the abundance levels of *brnQ* mRNA decreased in the BL2-peptide medium, although they were not statistically significant. The synchronized regulation of *brnQ* and *opt* operon with the proteolytic regulon suggests that they belong to the proteolytic regulon modulated by the pleiotrophic regulator CodY (11). Indeed, it was shown that in *Bacillus subtilis* purified CodY was able to bind the *dpp* promoter *in vitro* (18).

In addition to nitrogen starvation, glucose was likely depleted during the growth of IL1403 in peptide-limited or casitone medium because low glucose concentration (0.1%) and high starting cell density (OD_{600} = 2.0). It is possible that after depletion of glucose, the cells shut down the majority of glycolysis gene expressions while still maintaining the active expression of *eno* and *pyk*. These genes were essential for lactococci to survive under carbohydrate starvation because the cells relied on enolase (*enoB*) and pyruvate kinase (*pyk*) to obtain energy from phosphoenolpyruvate and 2-glycerate phosphate (19). The derived pyruvate then was used as the precursor to synthesize both acetyl-CoA and

α-acetolactate pools via the pyruvate dehydragenase complex (*pdhABCD*) or acetolactate synthase (*als*), respectively. α-Acetolactate is a central metabolite in lactococci and is involved in both anabolism (biosynthesis of BCAA) and catabolism (production of acetoin) (7). The fate of α-acetolactate is modulated by α-acetolactate decarboxylase (*aldB*), a gene whose activity is tightly regulated at the transcriptional, translational, and enzymatic levels in response to intracellular BCAA concentrations (8, 9). The absence of BCAA biosynthesis genes *ilv* and *leu* and presence of actively expressed branch amino acid permease gene (*brnQ*) suggests that BCAA are present intracellularly, which positively regulated the production and activity of α-acetolactate decarboxylase. Hence, it is likely that most of the α-acetolactate was converted into acetoin.

In conclusion, we found that *L. lactis* ssp. *lactis* IL1403 experienced nitrogen starvation even in the peptide rich casitone medium. This was due to lack of expression of peptide transporter genes. Instead of using peptide transporters, the cells transported the essential BCAA with a BCAA-specific permease (*brnQ*). The internalized BCAA in turn served as a signal to repress the peptide transporter genes (11). In contrast, growth in peptide-limited and casitone medium, peptides generated by *B. linens* BL2 proteolytic activities sustained the growth of *L. lactis* ssp. *lactis* IL1403. Repression of peptide transporters and other proteolysis genes was relieved in this growth medium. Unexpectedly, the Opp system, the primary peptide transporter in lactococci, was not actively expressed in IL1403. Instead, the Opt system, a di-tripeptide transporter, was used as the primary peptide transport system in IL1403. The *opp* operon was not actively expressed probably because a deficient promoter of the *opp* operon.

References

1. **Bolotin, A., P. Wincker, S. Mauger, O. Jaillon, K. Malarme, J. Weissenbach, S. D. Ehrlich, and A. Sorokin.** 2001. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. Genome Res. **11:**731-53.

2. **Broadbent, J. R., M. Strickland, B. Weimer, M. E. Johnson, and J. L. Steele.** 1998. Peptide accumulation and bitterness in Cheddar cheese made using single-strain *Lactococcus lactis* starters with distinct proteinase specificities. J. Dairy Sci. **81:**327-337.

3. **Detmers, F. J., F. C. Lanfermeijer, R. Abele, R. W. Jack, R. Tampe, W. N. Konings, and B. Poolman.** 2000. Combinatorial peptide libraries reveal the ligand-binding mechanism of the oligopeptide receptor OppA of *Lactococcus lactis*. Proc. Natl. Acad. Sci. USA **97:**12487-92.

4. **Dudoit, S., J. P. Shaffer, and J. C. Boldrick.** 2002. Multiple hypothesis testing in microarray experiments. Technical report 110. Division of Biostatistics, University of California, Berkeley. [Online.] http://www.bepress.com/cgi/viewcontent.cgi?article=1014&context=ucbbiostat. Accessed 10 February 2003.

5. **Foissy, H.** 1978. Aminopeptidase from *Brevibacterium linens*: activation and inhibition. Z. Lebensm. Unters Forsch. **166:**164-6.

6. **Godon, J. J., C. Delorme, J. Bardowski, M. C. Chopin, S. D. Ehrlich, and P. Renault.** 1993. Gene inactivation in *Lactococcus lactis*: branched-chain amino acid biosynthesis. J. Bacteriol. **175:**4383-90.

7. **Goupil, N., G. Corthier, S. D. Ehrlich, and P. Renault.** 1996. Imbalance of leucine flux in *Lactococcus lactis* and its use for the isolation of diacetyl-overproducing strains. Appl. Environ. Microbiol. **62:**2636-40.

8. **Goupil-Feuillerat, N., M. Cocaign-Bousquet, J. J. Godon, S. D. Ehrlich, and P. Renault.** 1997. Dual role of alpha-acetolactate decarboxylase in *Lactococcus lactis* subsp. *lactis*. J. Bacteriol. **179:**6285-93.

9. **Goupil-Feuillerat, N., G. Corthier, J. J. Godon, S. D. Ehrlich, and P. Renault.** 2000. Transcriptional and translational regulation of alpha-acetolactate decarboxylase of *Lactococcus lactis* subsp. *lactis*. J. Bacteriol. **182:**5399-408.

10. **Guedon, E., P. Renault, S. D. Ehrlich, and C. Delorme.** 2001. Transcriptional pattern of genes coding for the proteolytic system of *Lactococcus lactis* and

evidence for coordinated regulation of key enzymes by peptide supply. J. Bacteriol. **183:**3614-22.

11. **Guedon, E., P. Serror, S. D. Ehrlich, P. Renault, and C. Delorme.** 2001. Pleiotropic transcriptional repressor CodY senses the intracellular pool of branched-chain amino acids in *Lactococcus lactis*. Mol. Microbiol. **40:**1227-39.

12. **Kunji, E. R., I. Mierau, A. Hagting, B. Poolman, and W. N. Konings.** 1996. The proteolytic systems of lactic acid bacteria. Antonie Van Leeuwenhoek **70:**187-221.

13. **Le Bourgeois, P., M. L. Daveran-Mingot, and P. Ritzenthaler.** 2000. Genome plasticity among related *Lactococcus* strains: identification of genetic events associated with macrorestriction polymorphisms. J. Bacteriol. **182:**2481-91.

14. **Rattray, F. P., and P. F. Fox.** 1999. Aspects of enzymology and biochemical properties of *Brevibacterium linens* relevant to cheese ripening: a review. J. Dairy Sci. **82:**891-909.

15. **Rozen, S., and H. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. **132:**365-86.

16. **Sanz, Y., F. C. Lanfermeijer, P. Renault, A. Bolotin, W. N. Konings, and B. Poolman.** 2001. Genetic and functional characterization of dpp genes encoding a dipeptide transport system in *Lactococcus lactis*. Arch. Microbiol. **175:**334-43.

17. **Schena, M., D. Shalon, R. W. Davis, and P. O. Brown.** 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270:**467-70.

18. **Serror, P., and A. L. Sonenshein.** 1996. Interaction of CodY, a novel *Bacillus subtilis* DNA-binding protein, with the dpp promoter region. Mol. Microbiol. **20:**843-52.

19. **Thompson, J., and T. D. Thomas.** 1977. Phosphoenolpyruvate and 2-phosphoglycerate: endogenous energy source(s) for sugar accumulation by starved cells of *Streptococcus lactis*. J. Bacteriol. **130:**583-95.

20. **Tusher, V. G., R. Tibshirani, and G. Chu.** 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA **98:**5116-21.

21. **Tynkkynen, S., G. Buist, E. Kunji, J. Kok, B. Poolman, G. Venema, and A. Haandrikman.** 1993. Genetic and biochemical characterization of the oligopeptide transport system of *Lactococcus lactis*. J. Bacteriol. **175:**7523-32.

22. **Urbach, G.** 1995. Contribution of lactic acid bacteria to flavor compound formation in dairy products. International Dairy J. **5**:877-903.

23. **Weimer, B.** 1990. Ph.D. dissertation. Utah State University, Logan, Utah.

24. **Weimer, B., B. Dias, U. Madhavi, J. Broadbent, C. Brennand, J. Jaegi, M. Johnson, F. Milani, J. Steele, and D. V. Sission.** 1997. Influence of NaCl and pH on intracellular enzymes that influence Cheddar cheese ripening. Le Lait. **77**:383.

25. **Xie, Y., A. Cutler, and B. Weimer.** 2002. Poster: Linearization of DNA macroarray data. The 10th International Conference on Intelligent Systems for Molecular Biology, Edmonton, Alberta, Canada.

26. **Xie, Y.** 2002. M.S. report. Utah State University, Logan, Utah.

Table 4-1. Expression profiles of metabolic related gene in *L. lactis* ssp. *lactis* IL1403 under starvation condition.

Pathway	Gene	Intensity Average	Adjusted p-value	Gene Function (1)
Sugar catabolism	*bglA*	2032976	0.0468	6-phospho-beta-glucosidase
	galM	229492	0.0101	aldose 1-epimerase
	ycbD	64196	0.0330	UDP-glucose 4-epimerase
Phosphotransferase system	*ptcB*	90173	0.0002	cellobiose-specific PTS system IIB component
(PTS)	*yedE*	27634759	0.0002	PTS system, beta-glucosides-specific IIA component, putative
	yedF	92432279	0.0002	beta-glucoside-specific PTS system IIBC component
Glycolysis	*enoB*	262720105	0.0002	2-phosphoglycerate dehydratase
Pyruvate metabolism	*pdhA*	9978653	0.0002	pyruvate dehydrogenase E1 component, alpha subunit
	pdhB	65289	0.0018	pyruvate dehydrogenase E1 component, beta subunit
	pdhC	15494148	0.0002	dihydrolipoamide acetyltransferase component of PDH complex
	pyk	4692875	0.0002	pyruvate kinase
Butanoate metabolism	*aldB*	5450045	0.0002	alpha-acetolactate decarboxylase
	als	296021653	0.0002	acetolactate synthase large subunit
Pentose phosphate pathway	*rbsK*	2419	0.0487	ribokinase
	tkt	1483994	0.0002	transketolase
Purine metabolism	*deoB*	106971	0.0002	phosphopentomutase
	deoD	1935206	0.0002	purine-nucleoside phosphorylase
	prsA	5391454	0.0002	ribose-phosphate pyrophosphokinase
Arginine metabolism	*arcC1*	6860029	0.0002	carbamate kinase
	argG	505801	0.0006	argininosuccinate synthase
Cysteine metabolism	*cysM*	2846951	0.0030	cysteine synthase
Lysine biosynthesis	*asd*	68688	0.0468	aspartate-semialdehyde dehydrogenase
Aminoacyl-tRNA biosynthesis	*aspS*	654238	0.0002	aspartyl-tRNA synthetase
Proteolytic enzymes	*brnQ*	2597232	0.0101	branch-chain amino acids permease
	glnP	43301	0.0202	glutamine ABC transporter permease and substrate binding protein

	htrA	19993	0.0173	exported serine protease
	pepF	654036	0.0002	oligoendopeptidase F
	pepM	48924	0.0407	methionine aminopeptidase
	pepO	25689	0.0444	neutral endopeptidase
	pepP	109539	0.0216	aminopeptidase P
	pepT	5797	0.0478	peptidase T
Non-enzymes	*prmA*	50824	0.0096	ribosomal protein L11 methyltransferase
	dnaK	3868374	0.0002	DnaK protein

[a]Genes with adjusted p-value smaller than 0.05 were shown.

[b]Four positive controls were not shown. All of four positive controls were significant above the background with adjusted p-value 0.0002.

Table 4-2. Influences of exogenous proteolytic enzymes from *B. linens* BL2 on expression of *L. lactis* ssp. *lactis* IL1403.

Pathway	Gene	Fold Change	Gene Function (1)
Glycolysis	*enoB*	0.30	2-phosphoglycerate dehydratase
	fbaA	NA	fructose-bisphosphate aldolase
	gapB	NA	glyceraldehyde 3-phosphate dehydrogenase
	pgiA	7.64	glucose-6-phosphate isomerase
	tpiA	14.80	triosephosphate isomerase
Sugar catabolism	*pmi*	NA	mannose-6-phosphate isomerase
Pentose phosphate pathway	*gnd*	167.32	6-phosphogluconate dehydrogenase, decarboxylating
	tkt	2.45	transketolase
	zwf	NA	glucose-6-phosphate 1-dehydrogenase
Pyruvate metabolism	*aldB*	4.31	alpha-acetolactate decarboxylase
	als	1.26	acetolactate synthase large subunit
	citE	151.52	citrate lyase beta chain
	citF	NA	citrate lyase alpha chain
	ldh	15.59	L-lactate dehydrogenase
	mae	49.10	malate oxidoreductase
	pdhA	2.54	pyruvate dehydrogenase E1 component, alpha subunit
	pyk	7.51	pyruvate kinase
Fatty acid biosynthesis (path 1)	*acpA*	NA	acyl carrier protein
Purine and pyrimidine metabolism	*apt*	NA	adenine phosphoribosyltransferase
	deoD	1.65	purine-nucleoside phosphorylase
	guaC	NA	GMP reductase
	prsB	NA	ribose-phosphate pyrophosphokinase
	purL	288.03	phosphoribosylformylglycinamidine synthase II
	purM	NA	phosphoribosyl-aminoimidazole synthetase
	udk	12.20	uridine kinase
Aromatic amino acids biosynthesis	*araT*	19.12	aromatic amino acid specific aminotransferase
	trpB	107.25	tryptophan synthase beta chain

Brach chain amino acids biosynthesis	*ilvC*	NA	ketol-acid reductoisomerase
Proline biosynthesis	*proA*	NA	glutamate-5-semialdehyde dehydrogenase
Methionine metabolism	*metB2*	90.04	cystathionine beta-lyase
	metK	10.22	S-adenosylmethionine synthetase
Peptidoglycan biosynthesis	*dltA*	100.10	D-alanine activating enzyme
	glmS	53.89	glucosamine--fructose-6-phosphate aminotransferase (isomerizing)
	murE	NA	UDP-MurNac-tripeptide synthetase
Ubiquinone biosynthesis	*hmcM*	NA	hydroxymethylglutaryl-CoA synthase
	menD	NA	2-oxoglutarate decarboxylase / 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase
RNA polymerase	*rpoA*	6.10	DNA-directed RNA polymerase alpha chain
Aminoacyl-tRNA biosynthesis	*argS*	NA	arginyl-tRNA synthetase
	gltX	2.67	glutamyl-tRNA synthetase
	metS	4.89	methionyl-tRNA synthetase
	trpS	NA	tryptophanyl-tRNA synthetase
ABC transporters, prokaryotic	*busAA*	NA	betaine ABC transporter ATP binding protein
	busAB	127.87	betaine ABC transporter permease and substrate binding protein
	choQ	9.72	choline ABC transporter ATP binding protein
Phosphotransferase system (PTS)	*ptsI*	13.15	phosphotransferase system, enzyme I
	yedE	0.01	PTS system, beta-glucosides-specific IIA component, putative
	yedF	0.05	beta-glucoside-specific PTS system IIBC component
Proteolytic related proteins and enzymes	*optA*	NA	oligopeptide ABC transporter substrate binding protein
	optB	NA	oligopeptide ABC trasporter permease protein
	optD	78.70	oligopeptide ABC trasporter ATP binding protein
	pepC	11.57	aminopeptidase C
	pepT	7.48	peptidase T
	ydcB	103.32	amino acid ABC transporter ATP binding protein
	ydcC	NA	amino acid ABC transporter permease protein
	ydgB	7.30	amino acid permease
	ydgC	NA	amino acid permease
	yueF	NA	protease

Enzyme	*bar*	NA	acyltransferase

[a]The threshold setting for SAM leads to median number falsely called significant genes was 0.91 with median false discovery rate 1.51% (90 percentile of falsely called significant genes was 9.06 with false discovery rate 15.1%).

[b]NA means that the expression values were negative in starvation conditions because of background correction procedures, and therefore the fold of changes cannot be calculated.

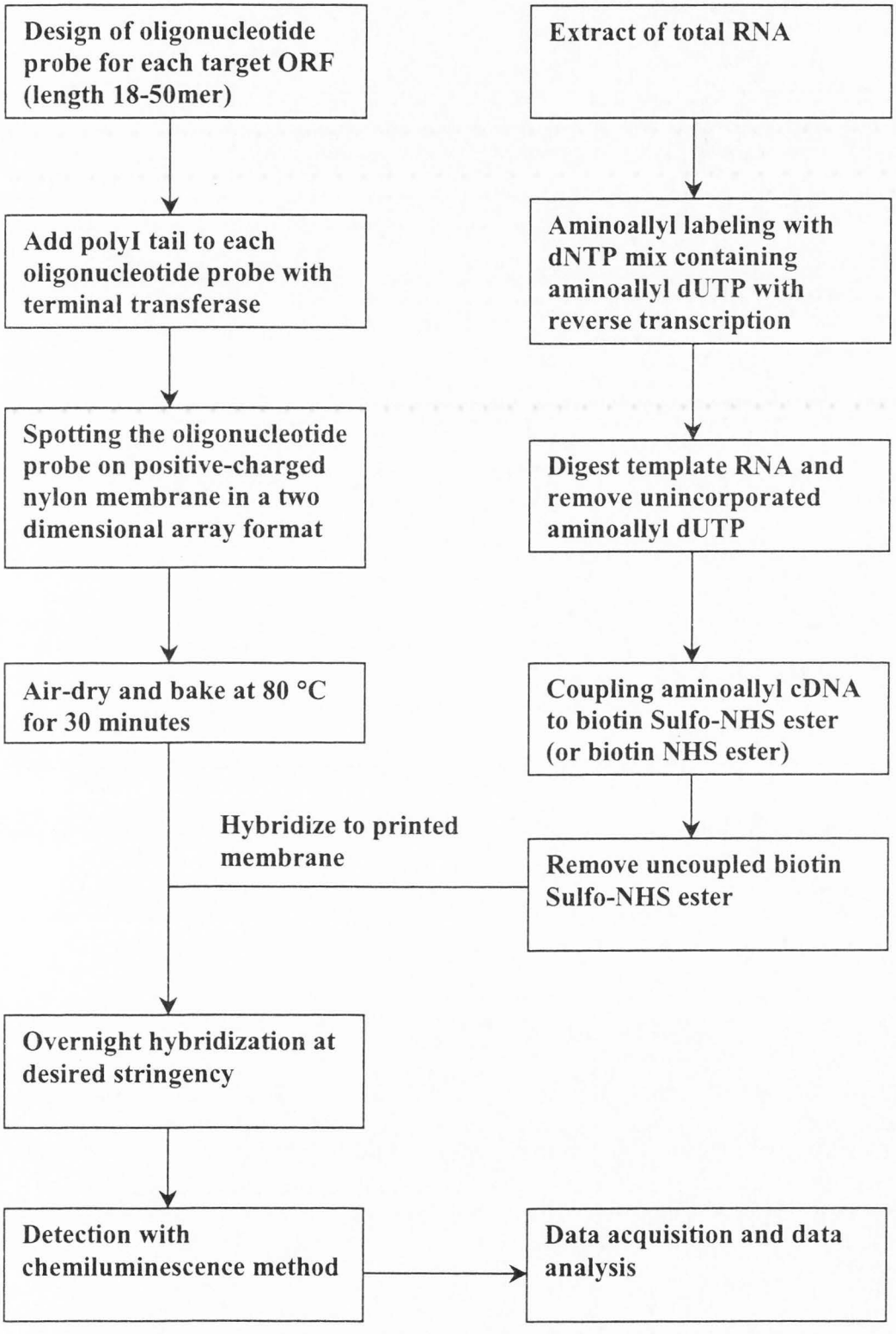


Figure 4-1. Flow chart of DNA macroarray protocol for gene expression.

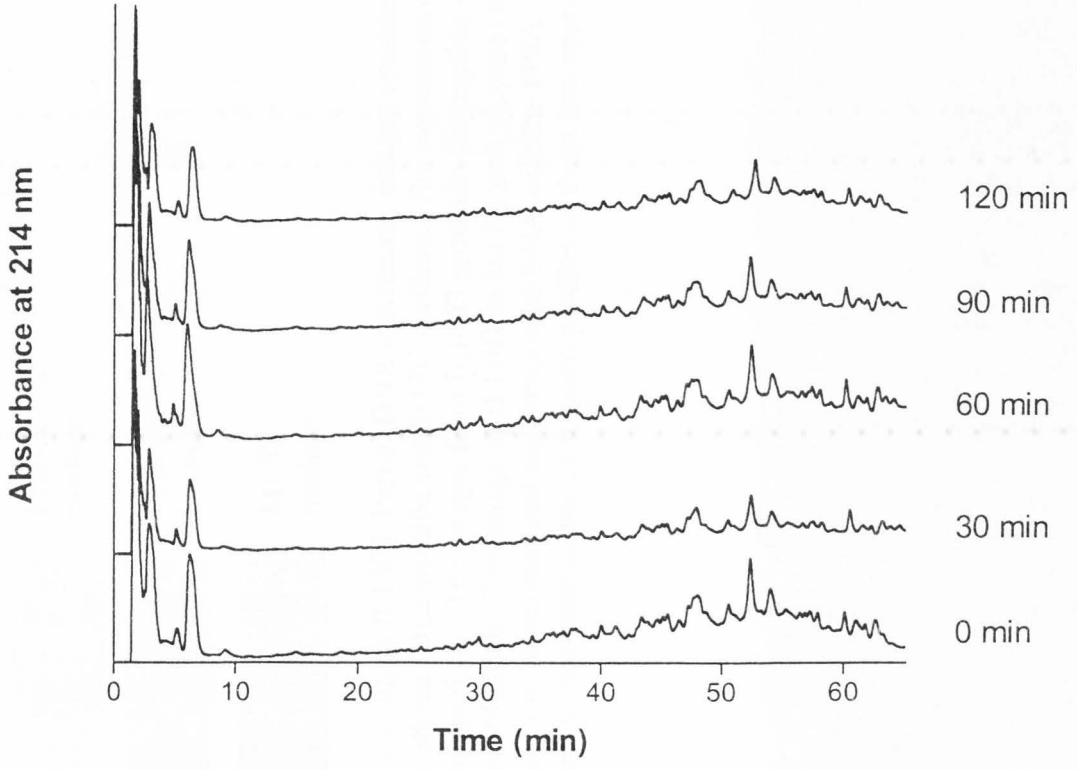


Figure 4-2. Peptide utilization profiles of *L. lactis* ssp. *lactis* IL1403 in BL2-peptide medium. The samples were taken from the BL2-peptide medium after *L. lactis* ssp. *lactis* was grown in it for 0, 30, 60, 90, and 120 minutes, respectively. The RP-HPLC chromatograms were labeled with the corresponding growth time of IL1403.

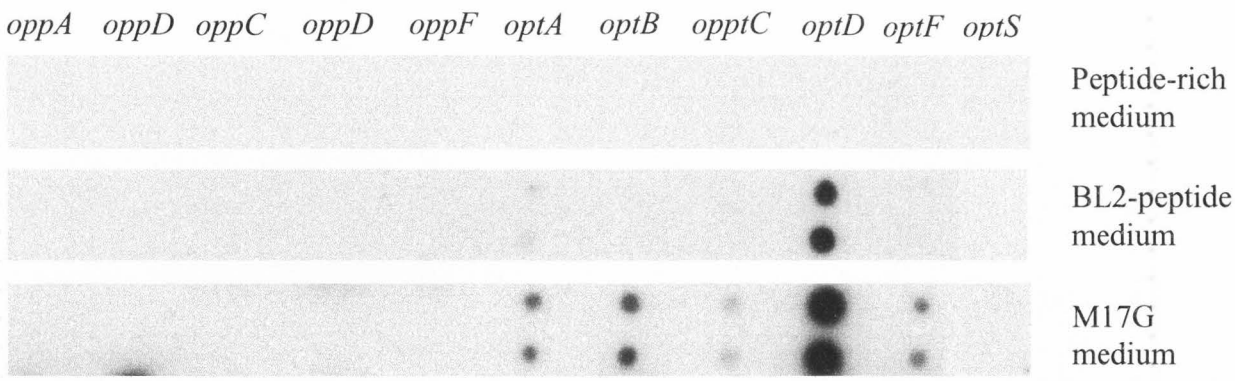


Figure 4-3. Expression of *opp* and *opt* operons in *L. lactis* ssp. *lactis* IL1403. Partial DNA macroarray images contained the expression profiles of the *opp* operon of IL1403 grown in the casitone, BL2-peptide and M17G medium. The expression of the *opt* operon in these media was also shown for the purpose of comparison. The images from IL1403 grown in the peptide-rich and BL2-peptide media were prepared as described in Materials and Methods. The image of IL1403 in M17G was prepared by incubating IL1403 in M17G (M17 supplemented with 0.5% glucose) for 30 minutes and analyzing total RNA with the DNA macroarray protocol. This image was used here as a positive control to demonstrate the expression profiles of these two operons when grown in the optimal medium.

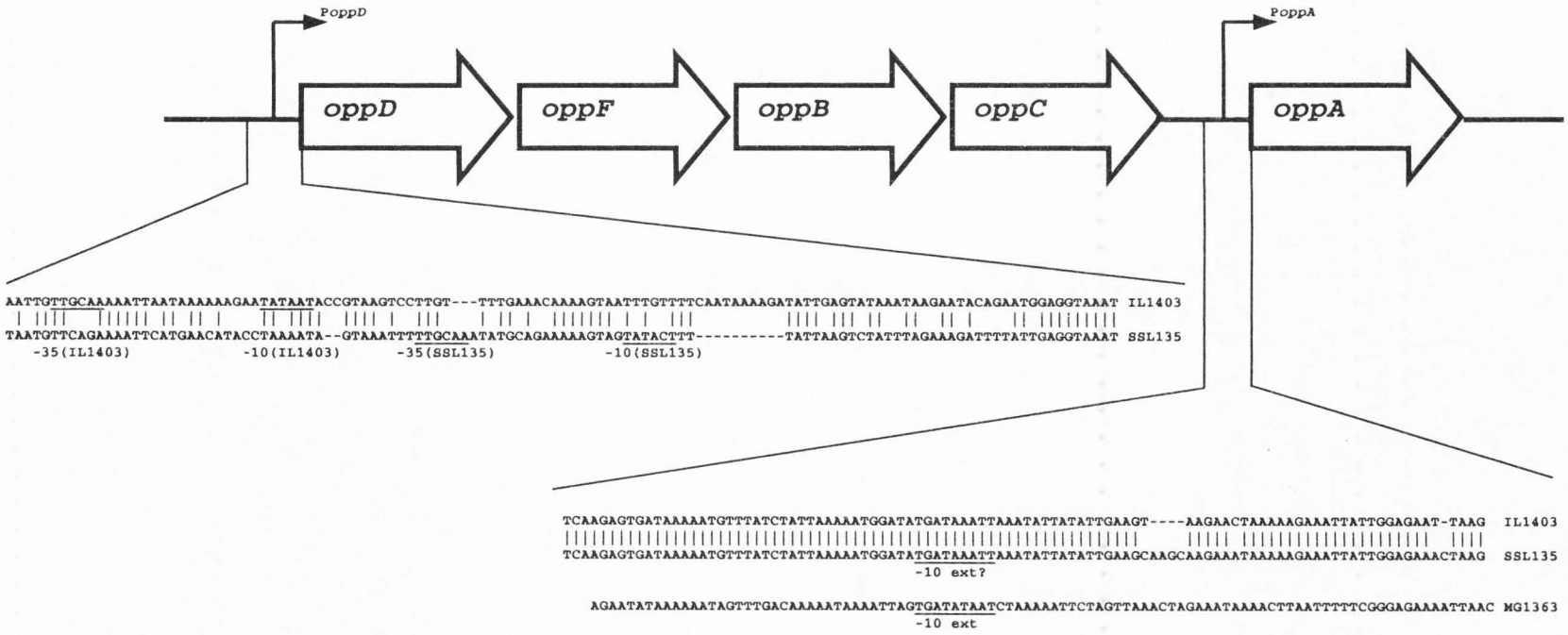


Figure 4-4. Sequence comparison of promoter regions of *opp* operon in *L. lactis* ssp. *lactis* IL1403 and SSL135. The promoter sequences were from *L. lactis* ssp. *lactis* IL1403 and SSL135 (1, 21). The -35 and -10 boxes of each *oppD* promoter from IL1403 and SSL135 were labeled by underline with the strain name shown parentheses. The sequence diversity in this region was shown by a pairwise alignment result (Martinez-NW method of MegaAlign package in DNAStar, Madison, Wisconsin). For *oppA* promoter, the putative extended -10 boxes for promoters in both IL1403 and SSL135 were underlined. A functional P*oppA* promoter with a perfect extended –10 box in MG1363 was shown for the comparison (Genbank access number AF245305).

CHAPTER 5

THE PROTEOLYSIS AND MATURATION OF *BREVIBACTERIUM LINENS* PROTEASE[1]

Abstract

Brevibacterium linens is a key microorganism in the production of Limburger cheese. It generates extracellular peptides via the activities of proteases and produces flavor compounds from downstream amino acid catabolism. Although the extracellular protease has been studied extensively, there is a little knowledge on the maturation process of this protease. We observed a delay in protease activity followed by protease secretion and hypothesized that this was due to zymogen maturation. Results from this study indicated that the extracellular protease from *B. linens* BL2 was secreted as a non-active zymogen with a molecular weight of 76 Kdal. This zymogen matured to an activated protease of 13 Kdal. Conversion of the zymogen to the fully activated protease was an autocatalytic process and involved several intermediate states. Both exogenous proteases and protease inhibitors influenced the maturation process. A detailed investigation of the kinetics of proteolysis and maturation process was presented. The proteolytic process was regulated by substrate inhibition, while the maturation process was also repressed once a certain amount of active protease was generated. Therefore, the total extracellular protease activity was tightly controlled in *B. linens* BL2 even after it was secreted outside the cell.

---

[1]Coauthored by Yi Xie, Rodney Brown, and Bart Weimer

Introduction

Brevibacterium linens is a coryneform bacterium used in the production of surface-ripened cheeses. This bacterium is also used as a flavor adjunct to produce a flavorful low fat Cheddar cheese (15). The extracellular protease from *B. linens* plays an important role for the bacterium used as an adjunct or as a ripening accelerating reagent (4, 15). Although the extracellular protease of *B. linens* is known to be a serine protease, there are conflicting reports about other properties of this enzyme. For example, pH and temperature optima were reported with ranges of 5.0-11.0 and 25-55 °C (3, 5, 7), respectively. It is also possible that the brevibacterial protease has more than one active form (3), although only a 56 Kdal protein has been purified to homogeneity (7). Interestingly, the extracellular proteolytic activity of *B. linens* was uniquely cycled during the long-term growth (11, 14), suggesting that the production of this enzyme is highly regulated.

Most bacterial extracellular and exocellular serine proteases are synthesized in the pre-pro form (13). After secretion, a signal peptidase hydrolyzes the leading peptide from the N-terminal of the pro-protein. The pro-protease undergoes additional maturation processes to become fully active, relying on a helper protein or via a self-mediated maturation process. Despite the industrial importance of this organism and its extracellular protease, there is a little information available on the maturation and its kinetics of the brevibacterial protease (8).

The aim of this study was to define the maturation process of the brevibacterial zymogen and determine the proteolytic kinetics of the active enzyme. In this study, we

demonstrated that *B. linens* BL2 secreted the extracellular protease as a zymogen that required further maturation to form an active enzyme. Both exogenous proteases and inhibitors, such as Mn^{2+} and serine protease inhibitors, influenced the maturation process. A detailed investigation of the kinetics of both the proteolytic and maturation processes was also investigated.

Materials and Methods

Bacterial Strain, Media, and Chemicals

Brevibacterium linens BL2 was obtained from our laboratory culture collection. The stock culture was prepared by transferring 1 ml culture into 1 ml sterile Tryptic Soy Broth (TSB broth, Difco, Sparks, Maryland), containing 30% glycerol, and stored in –70°C until used. Before each use, a frozen stock culture was thawed, inoculated (2% v/v) into TSB broth and grown at 25°C for 48 hours with aeration by shaking at 250 rpm.

Protease inhibitors, bovine plasmin, and cathepsin D were obtained from Sigma Chemical Co (St. Louis, Missouri). Chymax® chymosin was obtained from Chr Hansen Co. (Milwaukee, Wisconsin). The subtilisin Carsberg (a protease from *Bacillus subtilis* Carsberg) was obtained from Fluka Chemical Co. (Milwaukee, Wisconsin).

Assay of Protease Activity

Protease activity was measured by hydrolysis of fluorescently-labeled whole casein (EnzCheck, Molecular Probes, Eugene, Oregon). Unless specified, the protease was diluted 10-fold in 10 mM Tris-HCl buffer (pH 7.8), containing 20 mM sodium azide. The substrate working solution was composed of 10 μg/ml fluorescently-labeled casein

dissolved in 10 mM Tris-HCl buffer (pH 7.8). There were two types of assays. In single point assays, 0.5 ml of the enzyme working solution was mixed with an equal volume of substrate working solution and incubated at 30°C for 24 hours. The increase in fluorescence intensity was measured in a spectrofluorophotometer (model RF-1501, Shimadzu Co., Japan) with excitation/emission wavelengths of 485 nm/530 nm, respectively. Negative controls were made by mixing Tris-HCl buffer with substrate working solution and used to adjust activity value.

For multiple time point assays, the assay mixture was prepared in a 96-well plate (FluoroNunc, Nalge Nunc International, Denmark) in a total volume of 200 µl per well, and incubated at 30°C in a fluorescent plate-reader (model HTS 7000, Pelkin Elmer Co., Norwalk, Connecticut). The fluorescence intensities were monitored at 10-minute intervals for 24 hours at excitation/emission wavelengths of 485 nm/535 nm, respectively. For experiments involved addition of proteases or protease inhibitors, these reagents were diluted in 10 mM Tris-HCl (pH 7.0). These solutions were used to prepare the reaction mixtures. All assays were conducted in replicate.

Purification of Proteases from *B. linens* BL2

Preparation of Cell-free Concentrated Supernatant

Working cultures were prepared from frozen culture stocks as described above. To produce the protease for isolation, cells were inoculated (2%) into TSB broth, and incubated for 48 hours at 25°C with aeration at 250 rpm, subcultured (2%) into sterilized TSB broth, and further incubated for 60 hours at 25°C with aeration at 250 rpm. Cells

were removed from the culture medium by centrifugation twice (5,000 X *g* for 30 minutes at 4°C). The supernatant volume was reduced 50-fold in a stirred-cell ultrafiltration unit with a 50 kDa membrane (YM50 membrane; Amicon, Inc., Beverly, Massachusetts). The retentate was collected, filter sterilized using a 0.22-µm sterile syringe filter (Amicon, Inc.), and stored in a sterile container at -20°C for further use.

Purification of Zymogen

The filter-sterilized retentate was incubated with 100 mM sodium phosphate (pH 7.6) at 37°C for 30 minutes. The treated solution was applied to a Superose-12 column on the FPLC system (Amersham Pharmacia Biotech, Piscataway, New Jersey) pre-equilibrated with 1 *M* Tris-HCl buffer (pH 7.0) and eluted using 10 mM phosphate buffer at a linear flow rate of 0.5 ml/minute. One-ml fractions were collected and assayed for protease activity. The fractions containing protease activity were then pooled and further concentrated approximately10-fold with Centricon® 10 (Amicon, Inc.) at 5,000 x *g* for 2 hours at 4°C. The retentate contained the protease zymogen and used for further investigation.

Purification of Active Proteases

The filter-sterilized retentate (500 µl) was applied to an FPLC anion exchange column (Mono Q column HR10/10; Amersham Pharmacia Biotech). A linear gradient of 0.8 M NaCl in 0.01 M sodium phosphate buffer (pH 7.0) was used to elute the proteins at a linear flow rate of 2 ml/minute. Two-ml fractions were collected and assayed for protease activity using EnzCheck (Molecular Probes). The fractions containing protease activity were pooled and further purified with size exclusion chromatography using a

Superose-12 column with 0.01 M sodium phosphate (pH 7.0) at a linear flow rate of 0.5 ml/minute. Fractions were collected at 0.5-ml intervals and assayed for protease activity. Two active fractions were pooled and their protein content was determined with a SDS-PAGE gel.

Electrophoresic Analysis

The PhastGel SDS-PAGE gel system was used according to manufacturer's instructions and protein bands were detected with silver staining (Amersham Pharmacia Biotech). Pre-cast SDS-PAGE gels with an 8-25% polyacrymide gradient or homogeneous 12.5% polyacrymide were used. A 10 Kdal protein ladder was used as the protein standard (Life Technologies, Rockville, Maryland).

Data Analysis

The non-linear regression curve fitting function in Prism 3.0 (GraphPad Software, Sorento, California) was used to derive multiple parameters in the kinetic models (Eq 1 and Eq 3) described in Results section. The approximation of the data to the kinetic model was measured with the corrected r-square reported by the software package.

Results

Initial observations provided evidence of unusual casein hydrolysis patterns for crude and purified protease from *Brevibacterium linens* BL2. Hydrolysis of the fluorescently labeled casein substrate yielded a sigmoidal curve (Figure 5-1). The first derivative of this curve revealed that there were two distinct stages during the casein hydrolysis process, an initial slow linear stage followed by a second phase where the rate

of casein hydrolysis increased exponentially. The time required for the protease to switch to the second stage was enzyme concentration-dependent (data not shown). This two-stage hydrolysis pattern was different from the behavior of a subtilisin from *Bacillus subtilis* Carlsberg (subtilisin Carlsberg). The hydrolysis of casein by subtilisin Carlsberg yielded a typical enzymatic reaction curve (Figure 5-1). These observations lead us to hypothesize that this enzyme was synthesized as a zymogen and required further processing before it became fully active. To prove this hypothesis, we purified both the zymogen and activated proteases from *B. linens* BL2.

Purification of Proteases from *B. linens* BL2

In order to study the protease maturation process, the zymogen from *B. linens* BL2 was purified to homogeneity (Figure 5-2). We explored several buffers for their ability to stabilize the zymogen during purification process and yet allow it to maturate into the active form after the purification. We found that the pre-incubation of the crude mixture in phosphate buffer stabilized the protease zymogen, while EDTA and other strong metal chelators irreversibly inhibited the maturation process. The isolation of the *B. linens* protease zymogen was achieved based on a unique property of this protein, an affinity to Tris-treated Superose resin. Although most resins used in size-exclusion chromatography are chemically inert, we observed that washing the column with 1 M Tris buffer (pH 7.0) activated the Superose resin so it specifically bound the *B. linens* BL2 protease zymogen. Subsequently, phosphate buffer (10 mM at pH 7.0) was used to elute the zymogen from the column. Those fractions with the characteristic kinetics of the zymogen contained proteins of 76 Kdal (P76) (Figure 5-2).

Two activated forms (56 Kdal and 13 Kdal) of the enzyme were also purified to homogeneity (Figure 5-2). The smaller active form of protease had a molecular weight 13 Kdal (P13) and required no further maturation for proteolytic activity. However, this protease was unstable in isolation. We suspected that this protein was the end product of the zymogen maturation, although the size of this protease was relatively small in comparison to other serine proteases.

Maturation Process of *B. linens* BL2 Protease

Maturation Process Monitoring

To monitor the maturation process, the zymogen (P76) was pre-incubated for various times at 30°C and the samples were divided into two parts. One part was used for an activity assay (Figure 5-3A), and the another was run on an SDS-PAGE gel (Figure 5-3B). The delay in appearance of activity was shortened with increasing pre-incubation time at 30°C. Concurrent with the disappearance of zymogen P76, several additional bands appeared and disappeared during the process, indicating that the maturation involved multiple stages and protein intermediates. The maturation process produced the active protease P13. The maximal protease activity was obtained after 14 hour of incubation at 30°C. Once maximally activated, the protease activity was stable for at least additional 35 hours (Figure 5-3B). Activity optima for temperature, pH, and compounds that modulate totally activity are in Appendix F.

Influence of Exogenous Proteases or Chemical Compounds on the Maturation Process

We also evaluated the influences of exogenous factors on the activation of the *B. linens* BL2 protease zymogens. Multiple proteolytic enzymes, including the *B. linens* BL2 zymogen itself, were added to compare their efficiency in accelerating the maturation of zymogen. Cathepsin D showed the highest activation efficiency, followed by *B. linens* BL2 protease (Figure 5-4A). Chymosin was exceptional among the exogeneous enzymes tested because it extended the lag time for zymogen activation. It is noteworthy that all of the exogeneous proteases activated the *B. linens* BL2 zymogen. In contrast to all other tested proteases, the BL2 protease zymogen led to a decrease in activity after initial activation.

Inhibitors also influenced the activation of the *B. linens* protease zymogen. Treatment with several inhibitors, including Mn^{2+} and the serine protease inhibitors of p-hydroxy mercuribenzoate (PCMB) and iodoacetic acid, eliminated the activation phase in the enzyme kinetics while maintained the initial slow casein hydrolysis phase (Figure 5-4B).

Enzyme Kinetics of *B. linens* BL2 Protease

Kinetics of the Mature Protease on Casein Hydrolysis

To avoid interference of the maturation process, we measured the proteolysis kinetic data using *B. linens* proteases pre-incubated at 30°C for 16 hours. The plot of initial velocity versus substrate concentration showed that with casein concentrations

over 30 μg/ml, proteolytic activity was controlled by substrate inhibition (Figure 5-5A). A poor fit with the Michaelis-Menten kinetic model ($r^2 = 0.50$) confirmed that the activity of this enzyme did not follow typical enzymatic reaction patterns. Therefore, the Haldane equation (Eq. 1), which considers substrate inhibition effect (1), was used to fit these data. The data approximated well with the prediction of this model ($r^2 = 0.86$) and confirmed the observation that the high concentration of substrate inhibited the *B. linens* protease activity. Using the Haldane equation, the derived kinetic parameters were V_{max} = 3.6X10^5 RFU/hour, K_m = 2.3X10^4 μg casein/ml, and K_i = 0.67 μg casein/ml.

$$v = \frac{V_{max}}{1 + \frac{K_m}{[S]} + \frac{[S]}{K_i}} \qquad \text{Eq. 1}$$

Kinetics of the Self-Mediated Maturation Process

In parallel to proteolysis kinetic studies, we examined the maturation kinetics of the *B. linens* BL2 protease zymogen. The maturation process was assayed as the fluorescent product increase due to the newly formed active protease. In the self-mediated maturation process, the protease zymogen functioned as both enzyme and substrate, therefore influenced two parameters in Michaelis-Menten equation, V_{max} and substrate concentration *[S]*. To rule out this effect, we had to adapt this equation. In this analysis, we assumed that at the initial stage of maturation, the ratio of active protease to total protease is a constant. Based on this assumption, we deduced the Eq. 2 as:

$$V_{max} = k_1 [E_{active}] = k_1 * k_2 [E_{total}] = k_0 [E_{total}] \qquad \text{Eq. 2}$$

Replacing V_{max} in Michaelis-Menton equation with Eq. 2, a new form of the kinetic equation was established as:

$$v = \frac{k_0[E]^2}{K_m + [E]} \qquad \text{Eq. 3}$$

The kinetic data obtained by varying the zymogen concentration followed by analysis with Eq. 3 were k_0 = 2.0 RFU/(hr*μg casein/ml) and K_m = 2.5X10^4 μg casein/ml with a correlation coefficient of r^2 = 0.89 between observations and predictions of the model. The initial velocities versus zymogen concentration plot revealed their non-linear exponential relationship (Figure 5-5B). This confirmed our initial observations that the higher the zymogen concentration, the shorter the interval to enzyme activated. This is different from enzymatic reactions that follow Michaelis-Menton kinetics, where initial velocities of enzyme reaction reach a plateau value with increasing substrate concentrations.

Discussion

Most extracellular and exocellular proteases in bacteria are synthesized as pre-pro-enzymes (13). During secretion, the protein molecule is processed into a pro-enzyme that requires further maturation into a properly folded active protease. In this study, we observed a sigmoidal-shaped curve during the course of casein hydrolysis by the *B. linens* protease (Figure 5-1). This observation is consistent with other bacterial zymogens, such as subtilisin E and a streptococcal cysteine protease (2, 16). In addition, the protease from *B. linens* is also known for its characteristic activity cycles (11, 14), and existing in

multiple active forms (3). Based on the evidence, we hypothesized that the *B. linens* BL2 protease is synthesized as a zymogen and requires processing for activity.

The initial evidence in support of the maturation hypothesis was our purification of three forms of protease with sizes of 76 Kdal (P76), 56 Kdal (P56), and 13 Kdal (P13) (Figure 5-2). The other research groups also observed multiple forms of serine proteases from *B. linens* (3, 11), although there had been no in depth studies reported on the cause of this phenomenon or the activity associated with any specific intermediate. It appeared that the P76 protein was the zymogen, while P13 was the fully processed and active form of protease (Figure 5-3). It was unclear whether the 56 Kdal protein (P56) was an intermediate of the maturation process or a protease isozyme. However, it had a similar size with a previously purified protease from *B. linens* ATCC9174 (7).

There are two distinct models for the protease maturation process dependent on the source of the helper protein or peptide. The first model is an intermolecular model where helper protein is heterogeneous as exemplified by the maturation of PrtP protease in lactococci. PrtP maturation processing depends on a protein called PrtM, whose gene is located upstream and in the opposite orientation of protease gene (12). The second model of maturation process is an intramolecular model. This model was illuminated with subtilisin E where the maturation is mediated by the intramolecular chaperone, the pro-peptide (9). The mechanism of action of this intramolecular chaperone in subtilisin E was studied intensively (10). Maturation of the zymogen in *B. linens* BL2 was an autocatalytic event, that is, the maturation was occurred without addition of helper protein (Figure 5-3). Therefore, the mechanism of the *B. linens* BL2 protease activation is more likely to follow the intramolecular model as subtilisin E. In addition,

concentration-independent behavior of BL2 protease maturation processes also indicated a unimolecular (i.e. intramolecular) mechanism (6). During the maturation process, zymogen band was disappeared and in the meantime multiple additional intermediate protease bands were formed in the SDS-PAGE as the total proteolytic activity of the mixture reaches its maximum (Figure 5-3). This suggested that mechanism of the *B. linens* BL2 protease maturation is more complex than subtilisin E model.

Although the maturation of *B. linens* zymogen was an autocatalytic process, the exogenous factors such as exogenous proteases and protease inhibitors also influenced maturation process of *B. linens* zymogen. Addition of exogenous proteases altered maturation of the *B. linens* BL2 protease zymogen (Figure 5-4A). Interestingly, the BL2 protease P76 showed a unique activation pattern where a decrease in protease activity followed the initial increases. This suggests that the maturation process of *B. linens* BL2 was repressed after a certain extent of maturation process. Noticeably, this regulation mechanism was absent in the maturation process catalyzed by other proteases. We also surveyed the influences of various protease inhibitors on pre-activated *B. linens* protease. The control had a sigmodial-shaped curve with a shorter delay time compared to non-activated protein, while addition of Mn^{2+} and serine protease inhibitors resulted in a linear enzymatic reaction curve. One possible explanation was that the maturation process was more sensitive to these inhibitors than the proteolytic activity of this protease.

The extracellular proteolytic activity cycled during the long-term growth of *B. linens* (11, 14). This implied that the activity of protease in *B. linens* was highly regulated. Our evidence indicated that this regulation occurred by at lease two

mechanisms. A good fit between our data and Haldane equation ($r^2 = 0.86$) provided a direct evidence that the proteolytic activity of activated protease was controlled by substrate inhibition (Figure 5-5A). We also observe loss of protease activity (Figure 5-4A) and protease (Figure 3A) under some conditions, indicating that the protease activity can be regulated at maturation and post-maturation levels. There are two possible explanations for these observations. This could be caused by the inhibition of maturation process by the end product pro-peptide, as demonstrated in protease in *Brevibacterium brevis* (8), or by a self-degradation of active protease (Figure 5-3A).

References

1. **Cornish-Bowden, A.** 1995. Fundamentals of enzyme kinetics. Portland Press, London.

2. **Doran, J. D., M. Nomizu, S. Takebe, R. Menard, D. Griffith, and E. Ziomek.** 1999. Autocatalytic processing of the streptococcal cysteine protease zymogen: processing mechanism and characterization of the autoproteolytic cleavage sites. Eur. J. Biochem. **263:**145-51.

3. **Hayashi, K., A. J. Cliffe, and B. A. Law.** 1990. Purification and preliminary characterization of five serine proteinases by *Brevibacterium linens*. International J. Food Sci. and Tech. **25:**180-187.

4. **Hayashi, K., D. F. Revell, and B. A. Law.** 1990. Effect of partially purified extracelllular serine proteinases produced by *Brevibacterium linens* on the accelerated ripening of Cheddar cheese. J. Dairy Sci. **73:**579-583.

5. **Juhasz, O., and B. Skarka.** 1990. Purification and characterization of an extracellular proteinase from *Brevibacterium linens*. Canadian J. Microbiology **36:**510-512.

6. **Mach, L., J. S. Mort, and J. Glossl.** 1994. Maturation of human procathepsin B. Proenzyme activation and proteolytic processing of the precursor to the mature proteinase, in vitro, are primarily unimolecular processes. J. Biol. Chem. **269:**13030-5.

7. **Rattray, F. P., W. Bockelmann, and P. F. Fox.** 1995. Purification and characterizaiton of an extracellular proteinase from *Brevibacterium linens* ATCC 9174. Appl. Environ. Microbiol. **61:**3454-3456.

8. **Serkina, A. V., T. F. Gorozhankina, A. B. Shevelev, and G. G. Chestukhina.** 1999. Propeptide of the metalloprotease of *Brevibacillus brevis* 7882 is a strong inhibitor of the mature enzyme. FEBS Lett. **456:**215-9.

9. **Shinde, U., and M. Inouye.** 1995. Folding mediated by an intramolecular chaperone: autoprocessing pathway of the precursor resolved via a substrate assisted catalysis mechanism. J. Mol. Biol. **247:**390-5.

10. **Shinde, U., and M. Inouye.** 1995. Folding pathway mediated by an intramolecular chaperone: characterization of the structural changes in pro-subtilisin E coincident with autoprocessing. J. Mol. Biol. **252:**25-30.

11. **Tokita, F., and A. Hosono.** 1972. Studies on the extracellular protease produced by *Brevibacterium linens*. I. production and some properties of the extracellular protease. Jap. J. Zootech. Sci. **43:**39-48.

12. **van der Vossen, J. M., J. Kodde, A. J. Haandrikman, G. Venema, and J. Kok.** 1992. Characterization of transcription initiation and termination signals of the proteinase genes of *Lactococcus lactis* Wg2 and enhancement of proteolysis in *L. lactis*. Appl. Environ. Microbiol. **58:**3142-9.

13. **Wandersman, C.** 1990. Secretion, processing and activation of bacterial extracellular proteases. Mol. Micro. **3:**1825-1831.

14. **Weimer, B.** 1990. Ph.D. dissertation. Utah State University, Logan, Utah.

15. **Weimer, B., B. Dias, U. Madhavi, J. Broadbent, C. Brennand, J. Jaegi, M. Johnson, F. Milani, J. Steele, and D. V. Sission.** 1997. Influence of NaCl and pH on intracellular enzymes that influence Cheddar cheese ripening. Le Lait. **77:**383.

16. **Yabuta, Y., H. Takagi, M. Inouye, and U. Shinde.** 2001. Folding pathway mediated by an intramolecular chaperone: propeptide release modulates activation precision of pro-subtilisin. J. Biol. Chem. **276:**44427-34.

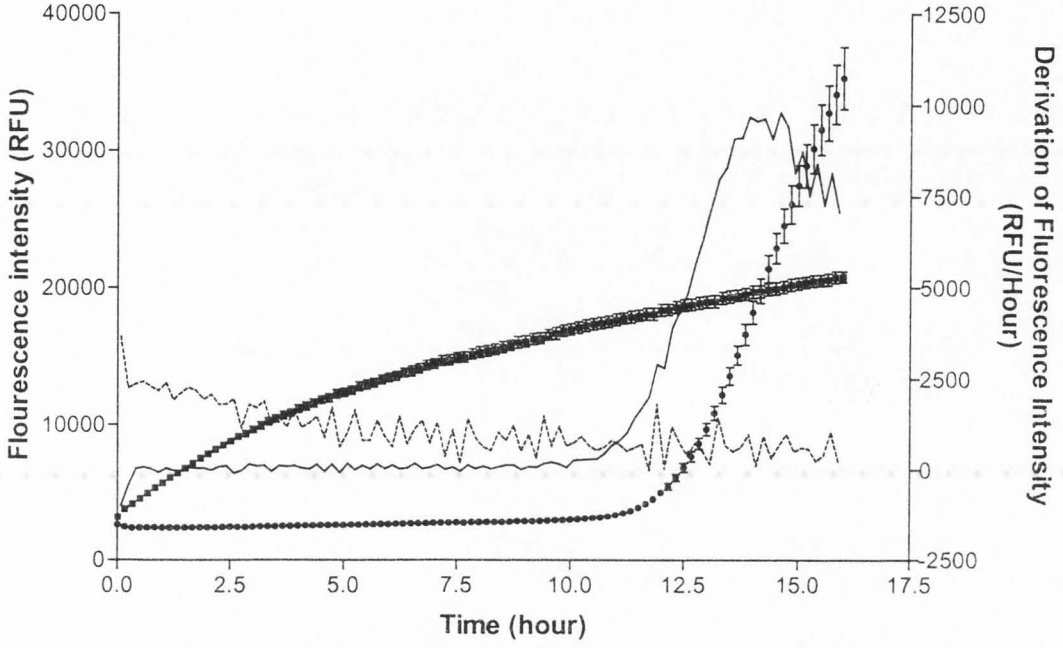


Figure 5-1. Time course of hydrolysis of EnzCheck substrate by *Brevibacterium linens*BL2 protease and subtilisin Carlsberg, ■ subtilisin Carlsberg (1mUnit/ml); ● *B. linens* BL2 protease (3.0 μg/ml); — first derivative curve of subtilisin Carlsberg (1mUnit/ml); --- first derivative curve of *B. linens* BL2 protease (3.0 μg/ml).

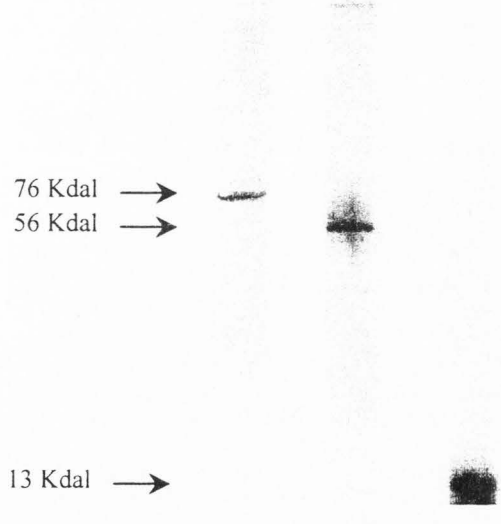


Figure 5-2. Purification of the protease from *B. linens* BL2. Lane 1 shows the 76 Kdal zymogen; lane 2 shows the 56 Kdal protease; and lane 3 shows the 13 Kdal fully activated protease.

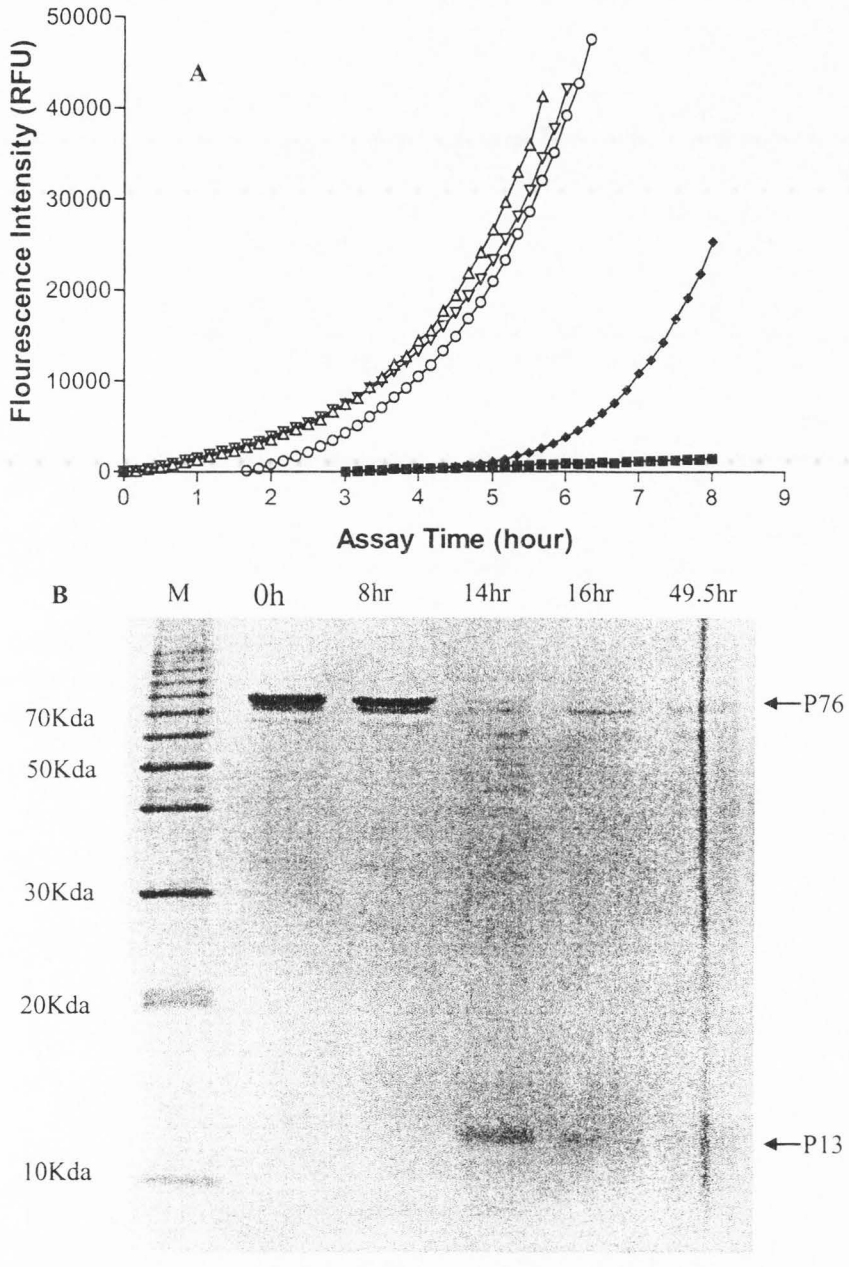


Figure 5-3. Maturation process of *B. linens* BL2 protease. Panel A: ■ *B. linens* BL2 protease with no pre-incubation; ◆ BL2 protease pre-incubated for 8 hours at 30°C; Δ BL2 protease pre-incubated for 14 hours at 30°C; ∇ BL2 protease pre-incubated for 16 hours at 30°C; ○ BL2 protease pre-incubated for 49.5 hour at 30°C; Panel B: Lane M: 10Kda protein ladder; Lane 0hr: *B. linens* BL2 protease with no pre-incubation; Lane 8hr: BL2 protease pre-incubated for 8 hours at 30°C; Lane 14hr: BL2 protease pre-incubated for 14 hours at 30°C; Lane 16hr: BL2 protease pre-incubated for 16 hours at 30°C; Lane 49.5hr: BL2 protease pre-incubated for 49.5 hours at 30°C.

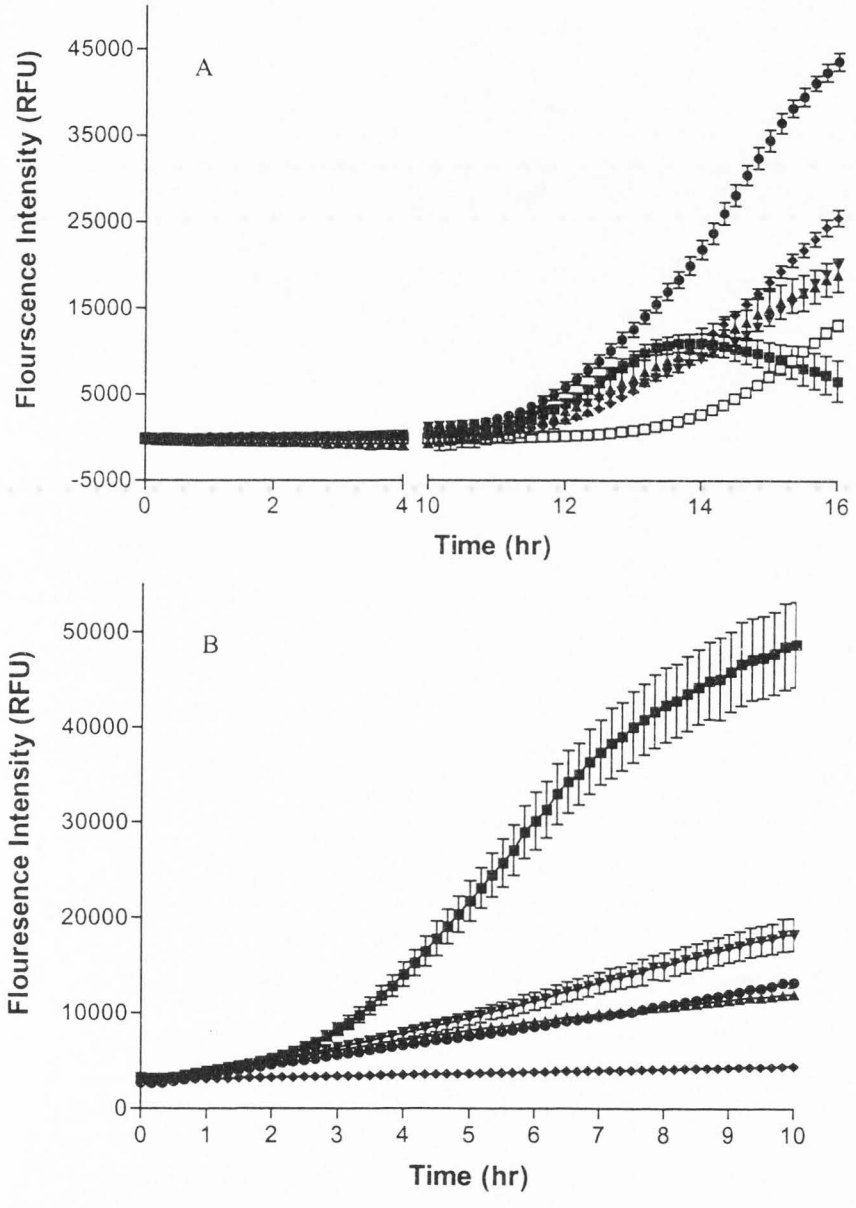


Figure 5-4. Effect of various activators and inhibitors on maturation of *B. linens* BL2 protease. Panel A: For heterogeneous protease activation experiments, 29.1 μg/ml *B. linens* BL2 protease is used. Proper heterogeneous protease activity controls are subtracted. In the case of *B. linens* BL2 protease, total 58.2 μg/ml BL2 protease is used and the protease activity above 29.1 μg/ml protease control is represented. ▲ subtilisin Carsberg (1 mU/ml); ■ subtilisin Carsberg (0.1 mU/ml); ◆ Bovine plasmin (1 mU/ml); ● Cathepsin D (10 mU/ml); □ Chymosin (1 Milk Clotting Unit/ml); ▪ *B. linens* BL2 protease (29.1 μg/ml); Panel B: For each of inhibitor experiment, 29.1 μg/ml *B. linens* BL2 protease is used. ◆ EDTA (1 mM); ▲ *p*-Hydroxy mercuribenzoate (PCMB) (1 mM); ▼ Iodoacetic acid (1 mM); ● Mn^{2+} (1 mM); ▪ *B. linens* BL2 protease control.

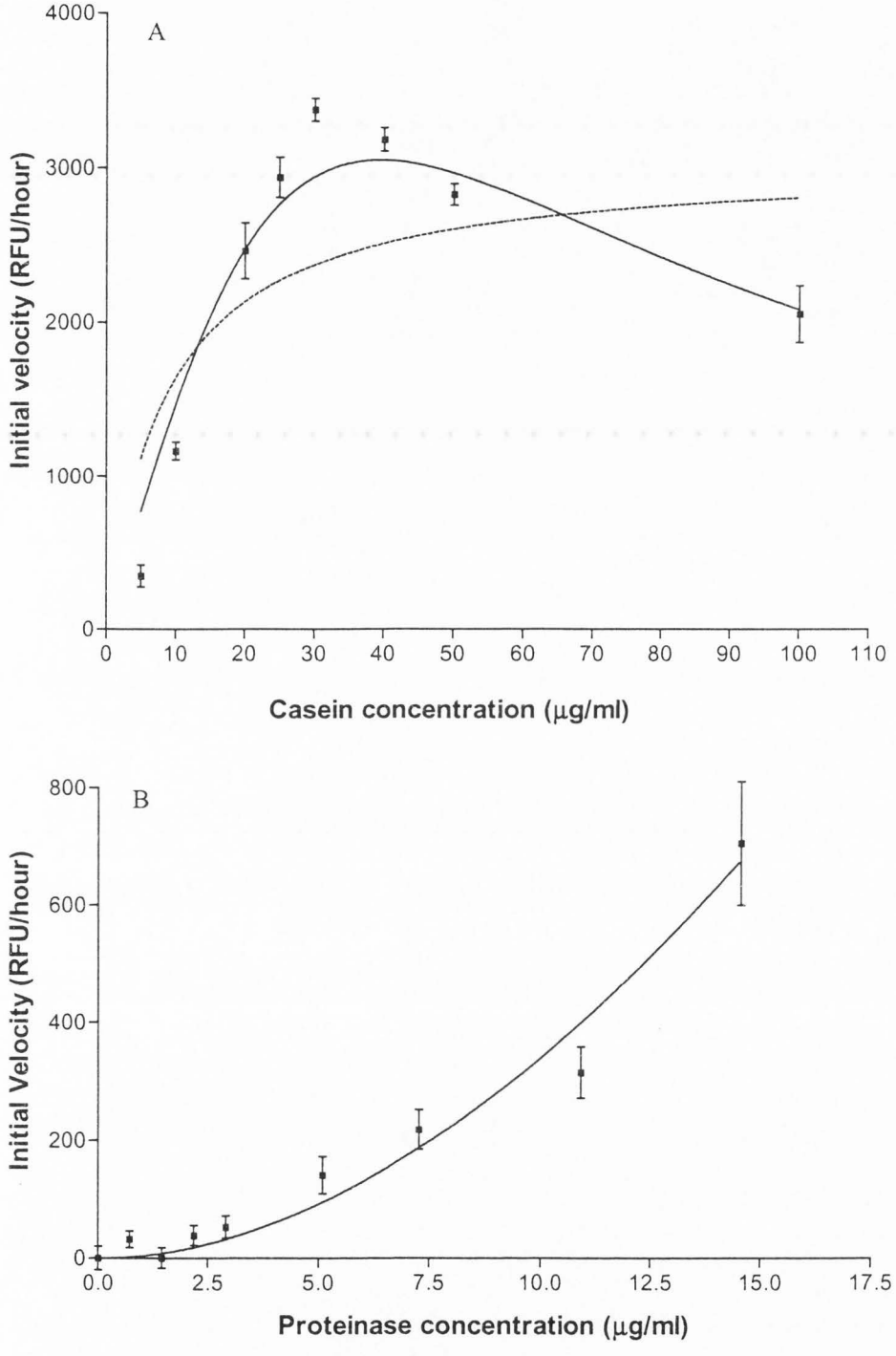


Figure 5-5. Kinetics of proteolysis and maturation of *B. linens* BL2 protease. Panel A. — represents the results of non-linear fitting based on Haldane equation. --- represents the results of non-linear fitting based on Michaelis-Menton equation. Panel B. — represents the results of non-linear fitting based on adapted Michaelis-Menton equation.

CHAPTER 6

SUMMARY AND CONCLUSIONS

Lactococcus lactis has been used as starter culture in cheese manufacture long before the existence of microbiology. Cheese flavor compounds are derived from the metabolic activities of the microbial community living inside the cheese matrix, especially by *L. lactis* (8). Since the catabolism of amino acids forms most volatile flavor compounds, a proper intracellular pool of amino acids must be established in the bacteria to promote desirable flavor compound production during cheese ripening. Generation of this pool of amino acids requires complex interactions among the community of bacteria, proteolytic enzymes, and casein and its derivatives.

During cheese ripening, lactococci coexist in the cheese matrix with a variety of other microbes, including non-starter lactic acid bacteria and flavor adjunct bacteria. Addition of *Brevibacterium linens* as an adjunct bacterium with the starter culture promotes more flavorful and softer-bodied low fat Cheddar cheese (9). This indicates that important interactions occur between the starter culture and the brevibacteria that modulate the cheese flavor and texture development. In this work, we focused on studying the influences of the proteolytic activity of *B. linens* BL2 extracellular protease BL2 on the expression of the metabolic genes in the proteinase-deficient starter culture *L. lactis* spp. *lactis* IL1403.

Hypothesis

Casein hydrolysis by the extracellular proteinases of *Brevibacterium linens* BL2 changes the expression profiles of proteolytic related metabolism genes in *Lactococcus lactis* spp. *lactis* IL1403.

Objectives

1. Develop an oligonucleotide-based DNA array technique suitable for monitoring gene expression in *L. lactis* spp. *lactis* IL1403.
2. Establish statistical and computational procedures that are suitable to analyze the data generated with the DNA arrays.
3. Demonstrate and utilize this technique in: a) profiling the stress responses of metabolic-related genes in *L. lactis* spp. *lactis* IL1403; b) determining the effect of the extracellular proteolytic activities of *B. linens* BL2 on the growth and gene expression profiles in *L. lactis* IL1403.

Development of a Filter-Based Oligonucleotide DNA Macroarray Protocol

To monitor global gene regulation in *L. lactis* ssp. *lactis* IL1403, a high throughput tool was needed to study the gene expression profiles at a genome scale. Previously, our group used oligonucleotide probes (40 mers) to monitor the expression profiles of genes involved in the arginine deiminase pathway in *L. lactis* (1). However, this protocol was not sufficient for genome level studies due to the limitations in sensitivity, specificity, and cost. Chapter 3 described a novel filter-based oligonucleotide

DNA macroarray protocol that was sensitive, highly specific, inexpensive, and easy to use.

The success of this oligonucleotide-based DNA array was dependent on technical innovations including polyinosine tailing, indirect high-density biotin labeling, careful probe design, and integrated computational data analysis. The polyinosine tails in this protocol served as steric spacers and non-specific DNA hybrid stabilizers. The indirect high-density labeling method was used to efficiently incorporate biotin into nucleic acids. The validity and utility of this protocol was demonstrated in Chapter 3 by profiling the expression of 375 metabolically related genes in *Lactococcus lactis* ssp. *lactis* IL1403 during heat, acid and osmotic stresses. This protocol also allowed us to access the impact of different peptide sources on the metabolism of *Lactococcus lactis* ssp. *lactis* IL1403 (Chapter 4). Some additional procedures and related results on DNA macroarray fabrication and usage were present in Appendix E.

Establish Computational Procedures for DNA Macroarray Data Analysis

Design and analysis of DNA arrays is very computationally demanding. Several key computational problems exists in DNA array design and data analysis (4, 6). These include large-scale oligonucleotide probe design, spot digitalization, expression data normalization and statistical analysis. The software components implemented in this project are described in Chapter 3, Appendixes A-D and in the report (11).

Design of the oligonucleotide probes was critical to this project. We utilized a modified version of Primer3 (5) to handle the main designing task of the oligonucleotide probe design, while a Perl script was implemented to handle the cross-hybridization

library assembly and the other input and output tasks. Both a command line user interface and a graphic user interface were available in this program. Oligonucleotide probes were selected based on length (22-24mer), melting temperature (63-65°C), GC percentage (40-60%), the absence of significant secondary structures, and at least four mismatches with any sequence outside the target in the genome.

Spot detection and quantification had an important influence on the accuracy of downstream data analysis. An ImageJ plugin, called SpotFinder, was implemented in Java to provide this spot detection function. Gridding was done with a semi-automatic procedure based on user controls of grid layout and start and end points. Subsequently, the spot in each grid cell was detected and quantified with a threshold-based spot segmentation algorithm. This program automatically detected the majority of spots with strong signals. A manual segmentation tool was also available for spots with weak or no signals and to correct any mis-called strong spots.

There were two data preprocessing steps used in the analysis of DNA macroarray data: intensity linearization and data calibration. Intensity linearization was required because of the dynamic range limitation the detection device (X-ray film) often underestimated the true signal intensities for high abundance transcripts due to film saturation. A linear regression based procedure was developed to merge the intensity dataset from two films with different lengths of exposure from a single DNA macroarray. With this approach, we not only recovered the linearity between signal intensities and mRNA concentrations for high abundance transcripts, but also maintained the sensitivity for low abundance transcripts. It was impossible to compare the raw data without data standardization between experiments. Intensity calibration (normalization) across

different samples was conducted with an adapted *lowess*-based procedure (10) with a reference data set generated by averaging the expression intensity of each gene over control and treatment conditions. Both of these software components were implemented in the statistical language R.

In order to calculate the statistical significance of differential gene expression, two software packages were used, SAM and *multtest*. SAM was developed at Stanford University and is available as an Excel add on (7). This software was used when the goal was to identify differentially expressed genes. The *multtest* package was developed at UC Berkeley (2) and is available as an R package in Bioconductor. This package was used when the aim was to assign biological functions to the genes. SAM relied on controlling the false discovery rate to minimize false positives, while *multtest* relied on controlling the family-wise error rate to prevent false positives.

Stress Responses of Metabolic-Related Genes in *L. lactis* spp. *lactis* IL1403

Roughly 13-18% of the investigated genes was differentially expressed under heat, acid and osmotic stress treatments. These data were described as part of Chapter 3. Since stress responses have been intensively studied in lactic acid bacteria, a wealth of information was available to allow us to verify the performance of the DNA macroarray. In almost all cases, the results obtained with the macroarray agreed well with previously reported observations. Based on this evidence, we concluded that the DNA macroarray protocol functioned as desired and provided valid results for the stress responses of IL1403.

In addition to confirming of previous findings, we also gained new insights into *L. lactis* ssp. *lactis* IL1403 stress responses. For example, the methionine biosynthetic pathway was induced during heat shock, but repressed during acid stress. Since aspartate is a precursor of methionine biosynthesis, genes in the metabolism of aspartate were also regulated differently between heat and acid stress. Although all three stresses repressed peptide transporter genes (*opt* and *dtpT*), amino acid transporter genes were strongly induced after heat stress, but remained relatively constant in other stress treatments. In addition to peptide transporter genes, several other genes were also repressed in all of the stress treatments. These genes included the β-glucoside specific PTS system (*yedE-yedF*) and the arginine deiminase pathway genes. It is likely that the regulations of these genes belonged to a general stress protection mechanism.

Influence of the Extracellular Proteinase of *B. linens* BL2 on the Growth and Gene Expression Profiles in *L. lactis* IL1403

With DNA macroarray analysis, we profiled gene expression changes of *Lactococcus lactis* spp. *lactis* IL1403 growing in a peptide-limited medium, in a casitone-based peptide-rich medium, and in a peptide-rich casein hydrolyte produced by the proteolytic enzymes of *Brevibacterium linens* BL2. These findings are described in Chapter 4. We found that *L. lactis* ssp. *lactis* IL1403 experienced nitrogen starvation, despite abundant peptide resources in casitone. This nitrogen starvation was caused by the lack of expression of peptide transporter genes in this medium. In the absence of peptide transporter gene expression, the cells transported essential branch chain amino

acids (BCAA) with a BCAA-specific permease (*brnQ*). The internalized BCAA probably then served as a signal to repress the peptide transporter genes (3).

Conversely, a peptide pool generated by the action of *B. linens* BL2 proteaseson caseins was sufficient to sustain the growth of *L. lactis* ssp. *lactis* IL1403. The repression of peptide transporter genes and other proteolytic related genes was relieved in this medium. In addition, the Opt system, a di-tripeptide transporter, was used as the primary peptide transport system in IL1403, as the *opp* operon was not actively expressed due to a functional deficiency in the promoter.

Proteolysis and Maturation of *Brevibacterium linens* BL2 Protease

The extracellular protease from *B. linens* BL2 was the enzyme responsible for the generation of peptides transportable by *L. lactis* IL1403. In this study, we further undertook a biochemical investigation of this enzyme. These results are described in Chapter 5 and Appendix F. We observed a characteristic delay in appearance of protease activity and hypothesized that this was due to zymogen maturation. Data from this study indicated that the extracellular protease from *B. linens* BL2 was secreted as a nonactive zymogen with a molecular weight of 76 Kdal. Maturation of this zymogen was an autocatalytic proteolysis and produced a 13 Kdal activated protease. Conversion of the zymogen to the fully activated protease involved several intermediate states.

Both exogenous proteases and inhibitors influenced the maturation process. Regulation of the total protease activity was controlled during maturation and proteolysis. A detailed investigation on the kinetics of both the proteolytic process and maturation process is also presented. In addition to studies presented in Chapter 5, we also

characterized the optimal pH and temperature of pre-activated enzyme on α-, β-, κ-, and whole casein substrates (Appendix F).

Conclusion

We hypothesized that the products of casein hydrolysis by the extracellular proteinases of *Brevibacterium linens* BL2 alters the expression profile of proteolytic related metabolism genes in *L. lactis* spp. *lactis* IL1403. We developed a novel oligonucleotide-based filter DNA array protocol and used a functional genomic approach to test this hypothesis. The utility and validity of the DNA macroarray protocol was demonstrated by profiling the expression of 375 metabolically related genes in *Lactococcus lactis* ssp. *lactis* IL1403 during heat, acid, and osmotic stresses. We found that the peptide pool generated from casein hydrolysis by the *B. linens* BL2 proteases sustained growth of *L. lactis* ssp. *lactis* IL1403. In contrast, casitone-based peptide-rich medium did not support growth. The extracellular protease from *B. linens* BL2 was secreted as a non-active zymogen, and its maturation and proteolytic activity were tightly regulated. Collectively, this work demonstrated that the proteases of *B. linens* BL2 were able to generate a pool of peptides transportable by *L. lactis* IL1403 and induced the changes in gene expression profiles in *L. lactis* IL1403. Consequently, this body of work validated the working hypothesis.

References

1. **Chou, L.** 2002. Ph.D. dissertation. Utah State University, Logan, Utah.

2. **Dudoit, S., J. P. Shaffer, and J. C. Boldrick.** 2002. Multiple hypothesis testing in microarray experiments. Technical report 110. Division of Biostatistics, University of California, Berkeley. [Online.] http://www.bepress.com/cgi/viewcontent.cgi?article=1014&context=ucbbiostat. Accessed 10 February 2003.

3. **Guedon, E., P. Serror, S. D. Ehrlich, P. Renault, and C. Delorme.** 2001. Pleiotropic transcriptional repressor CodY senses the intracellular pool of branched-chain amino acids in *Lactococcus lactis*. Mol. Microbiol. **40:**1227-39.

4. **Quackenbush, J.** 2001. Computational analysis of microarray data. Nat. Rev. Genet. **2:**418-27.

5. **Rozen, S., and H. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. **132:**365-86.

6. **Smyth, G. K., Y. H. Yang, and T. P. Speed.** 2002. Statistical issues in cDNA microarray data analysis. Technical report, Department of Statistics, University of California, Berkeley. [Online.] http://stat-www.berkeley.edu/users/terry/zarray/TechReport/mareview.pdf. Accessed 10 February 2003.

7. **Tusher, V. G., R. Tibshirani, and G. Chu.** 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA **98:**5116-21.

8. **Urbach, G.** 1995. Contribution of lactic acid bacteria to flavor compound formation in dairy products. International Dairy J. **5:**877-903.

9. **Weimer, B., B. Dias, U. Madhavi, J. Broadbent, C. Brennand, J. Jaegi, M. Johnson, F. Milani, J. Steele, and D. V. Sission.** 1997. Influence of NaCl and pH on intracellular enzymes that influence Cheddar cheese ripening. Le Lait. **77:**383.

10. **Yang, Y. H., S. Dudoit, P. Luu, and T. P. Speed.** 2001. Normalization for cDNA microarray data, p. 24-34. *In* M. L. Bittner, Y. Chen, A. N. Dorsel, and E. R. Dougherty (eds.), Microarrays: Optical technologies and informatics, Proceedings of SPIE, vol. 4266. The International Society for Optical Engineering. Bellingham, Washington.

11. **Xie, Y.** 2002. M.S. report. Utah State University, Logan, Utah.

APPENDICES

Appendix A. OligoDesigner for Large Scale Design of Oligonuncleotide Probes for DNA Macroarray

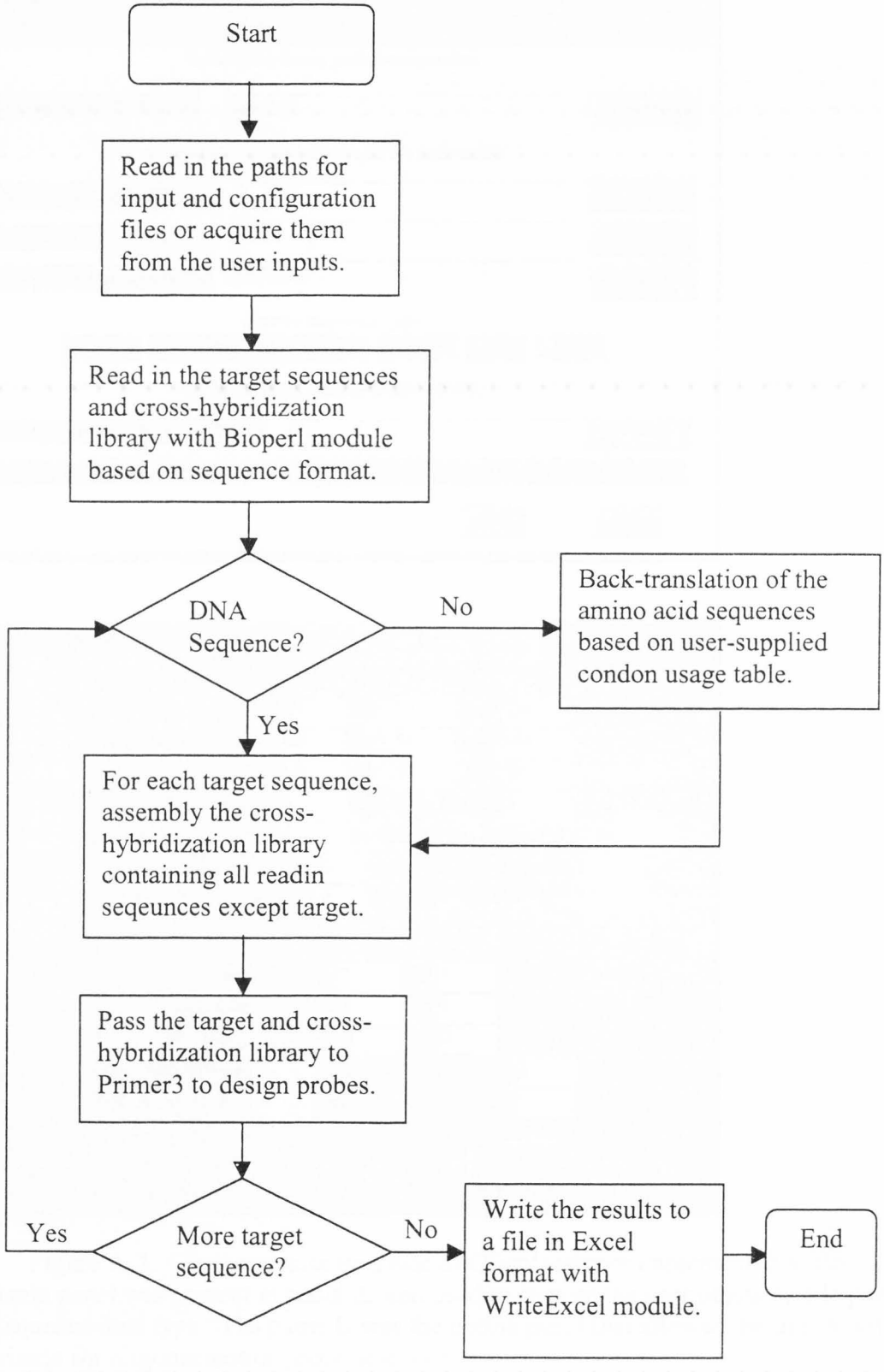
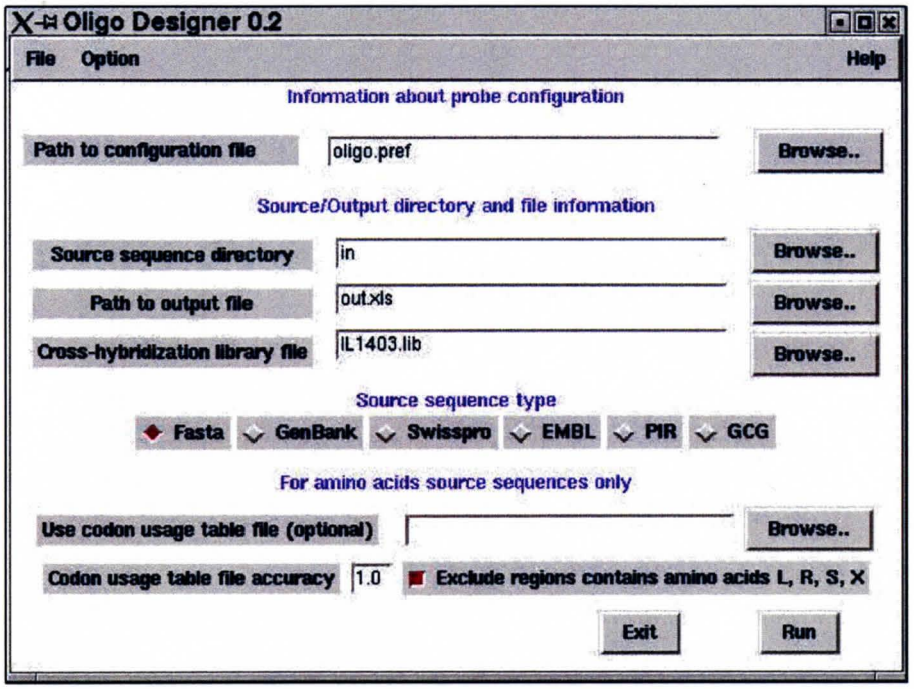


Figure A-1. The flowchart for the large-scale oligonucleotide designing program.

A



B

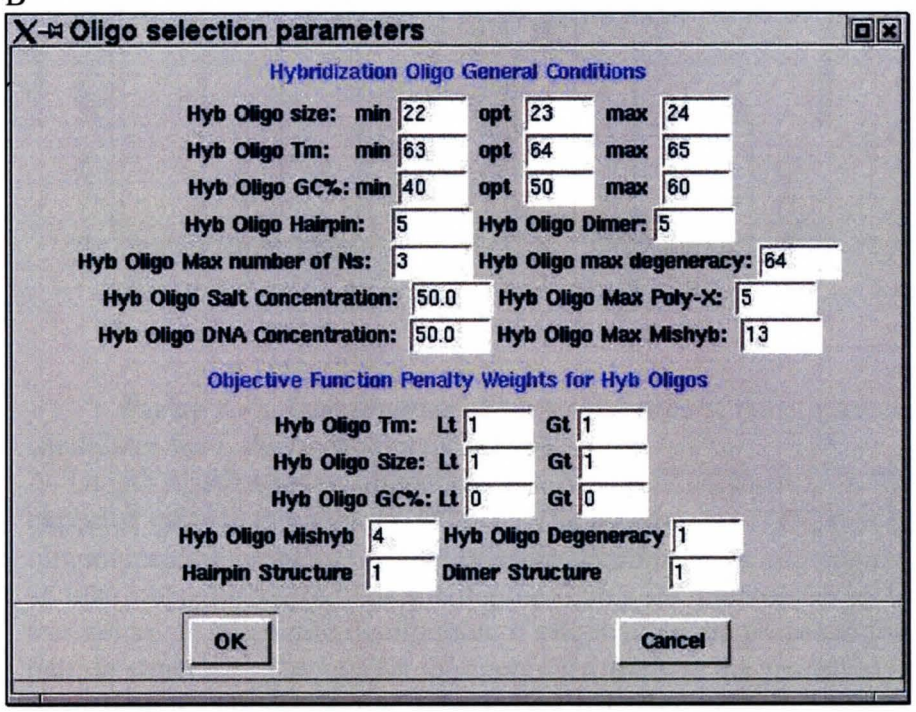


Figure A-2. Graphical user interface of OligoDesigner implemented in Perl. The main panel was present in panel A, and used to acquire the user inputs on file path and sequence data type. The panel B was the option panel that allowed the user to set up the criteria for oligonucleotide probe selection.

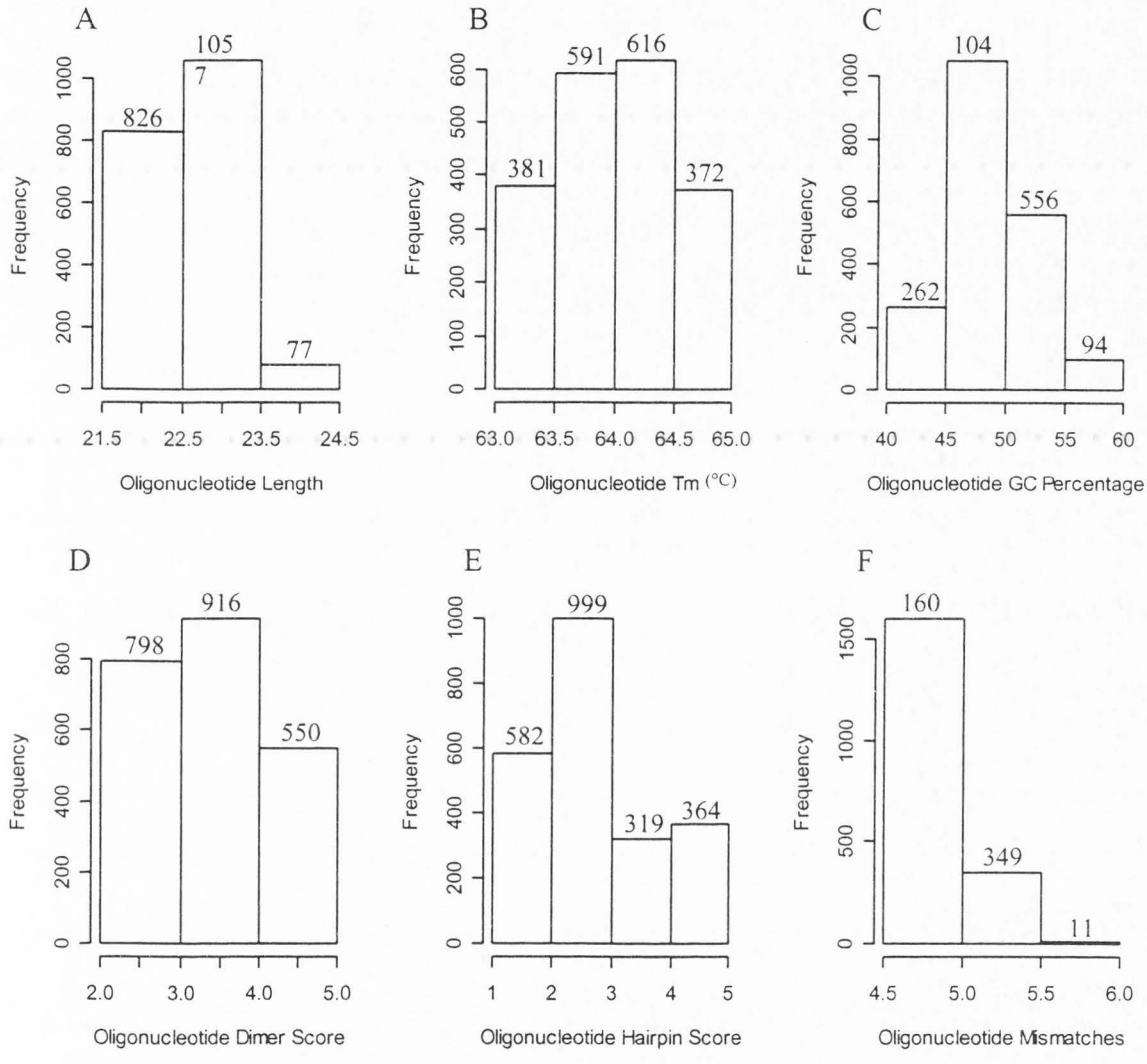


Figure A-3. Distributions of some attributes of the oligonucleotide probe candidates for *L. lactis* ssp. *lactis* IL1403.
A. Length distribution of oligonucleotide probes (22-24mers). B. The distribution of expected melting temperature (°C). C. The distribution of GC percentage of oligonucleotdie probes (%). D. The score distribution of oligonucleotide probes to form an intermolecule dimer. The smaller the scores are, the less the probability for dimer formation. E. The score distribution of oligonucleotide probes to form an intramolecule hairpin structure. The smaller the scores are, the less the probability for hairpin formation. F. The number of mismatches that oligonucleotide have with their closest homologues.

Appendix B. The Algorithm for Spot Detection and Quantification on DNA Macroarrays

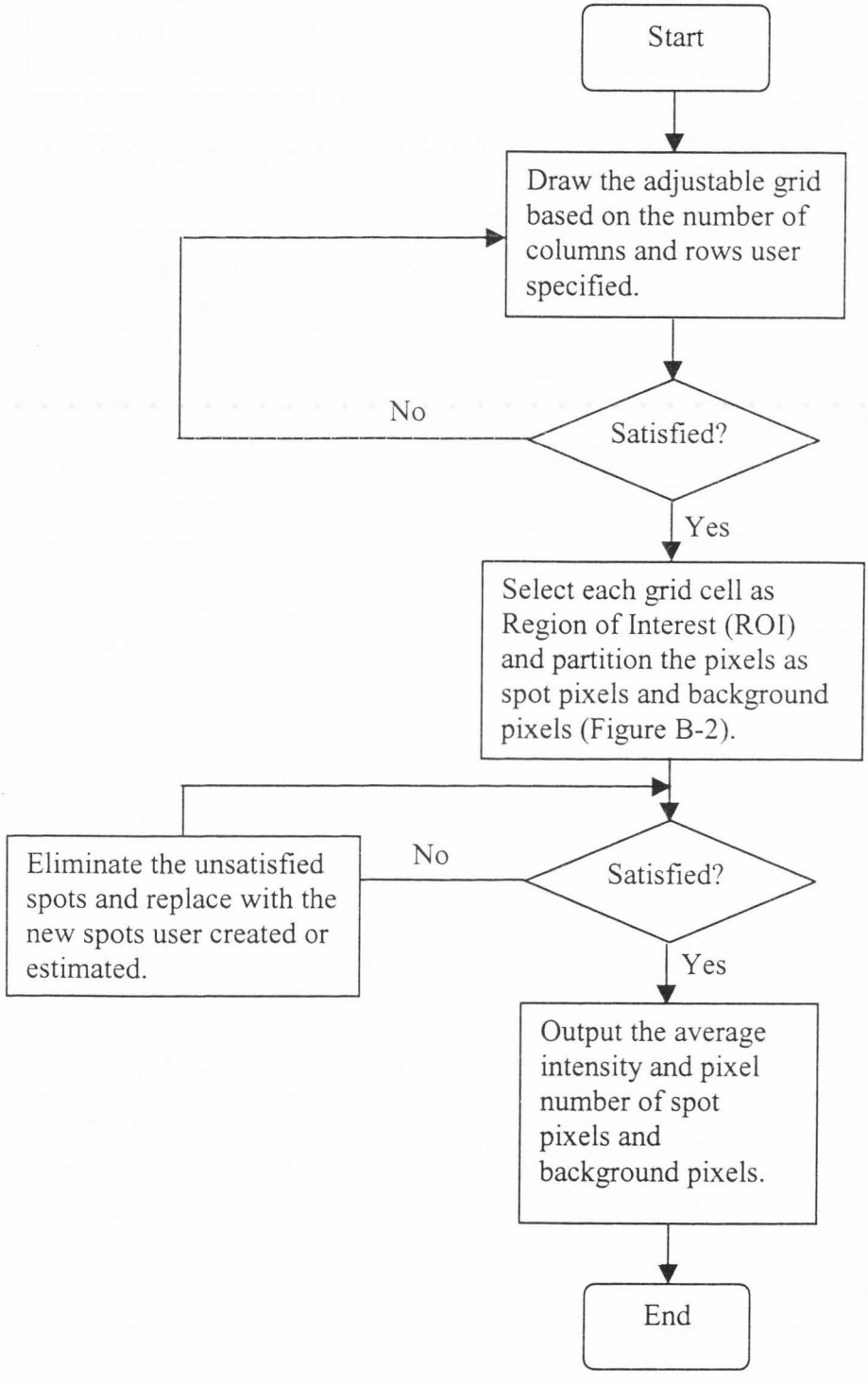


Figure B-1. General work flow of the spot detection and quantification program.

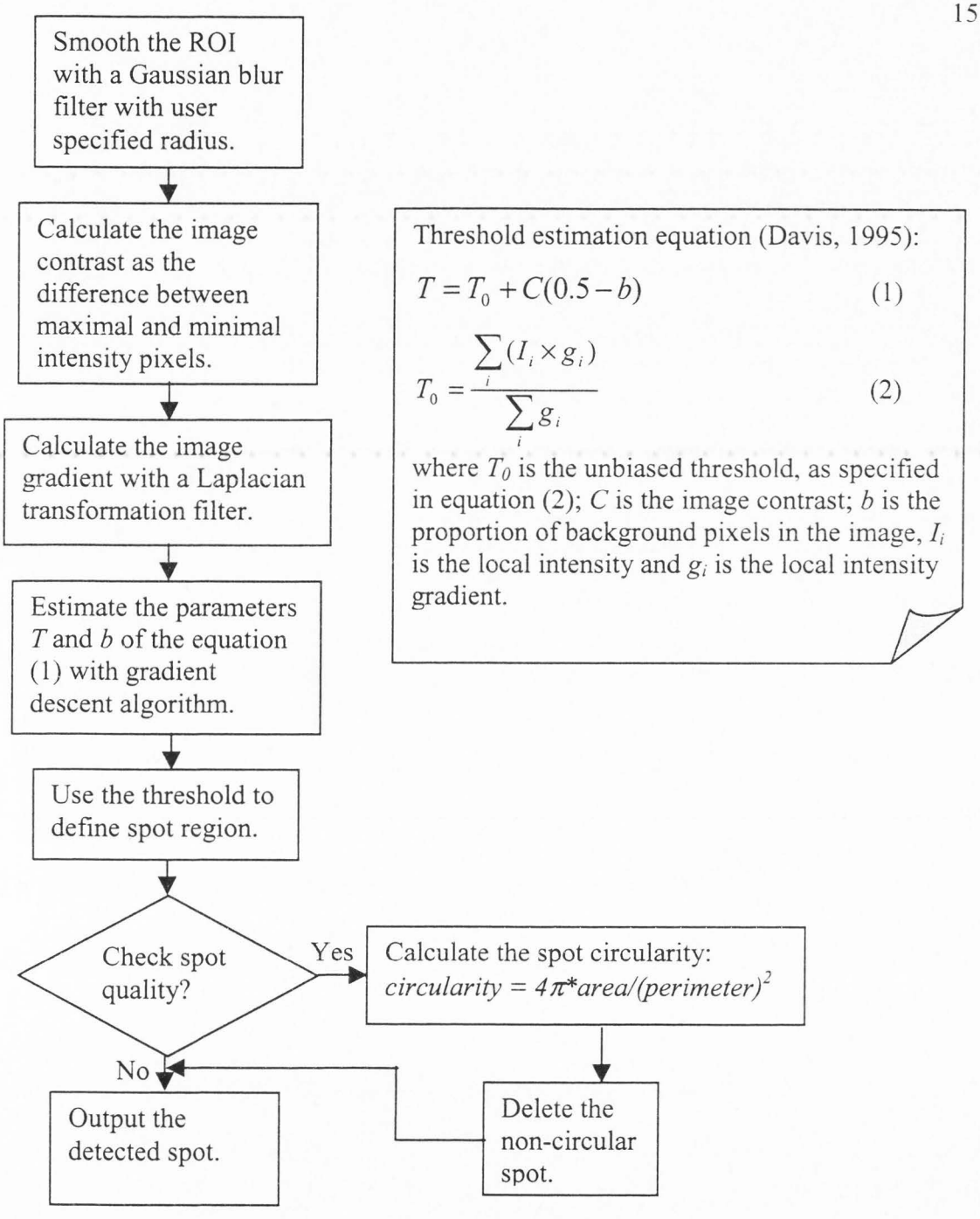
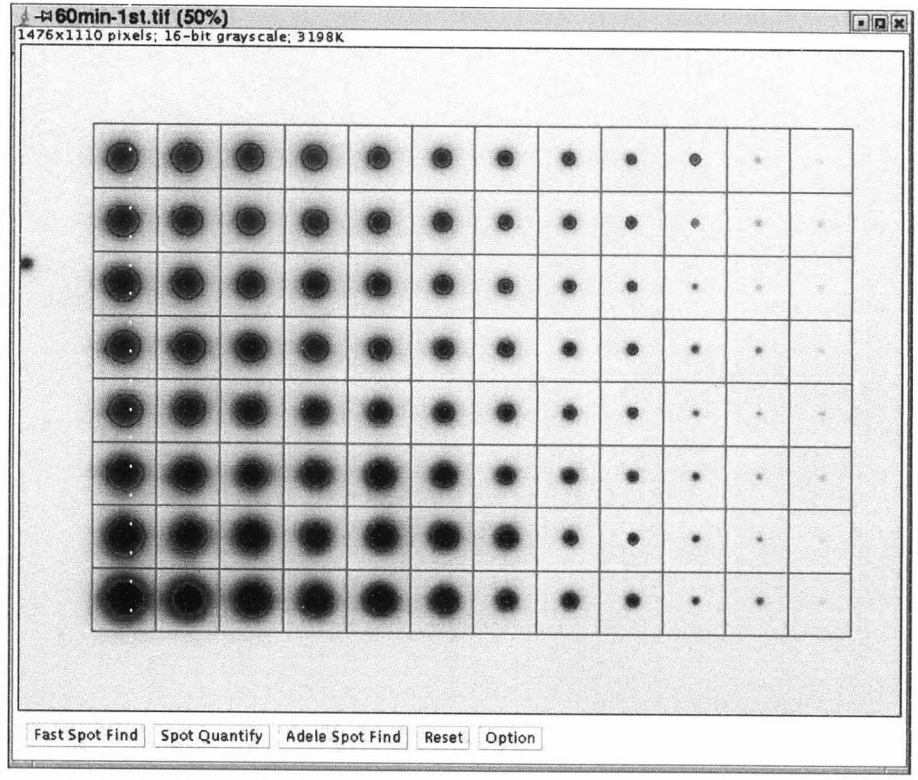


Figure B-2. Detailed flowchart for the spot detection algorithm.

A



B

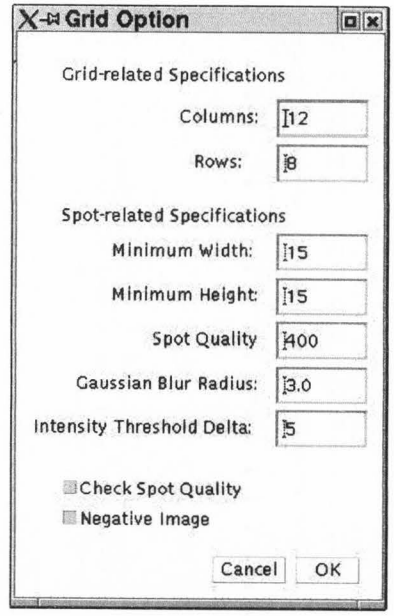


Figure B-3. Graphical user interface of SpotFinder, an ImageJ plugin implemented in Java. The main panel was present in panel A, displaying semi-automatic gridding and spot detection results. The panel B was the option panel that allowed the user to set up necessary parameters in the spot detection algorithm.

Appendix C. Preprocessing Procedures of DNA Macroarray Data

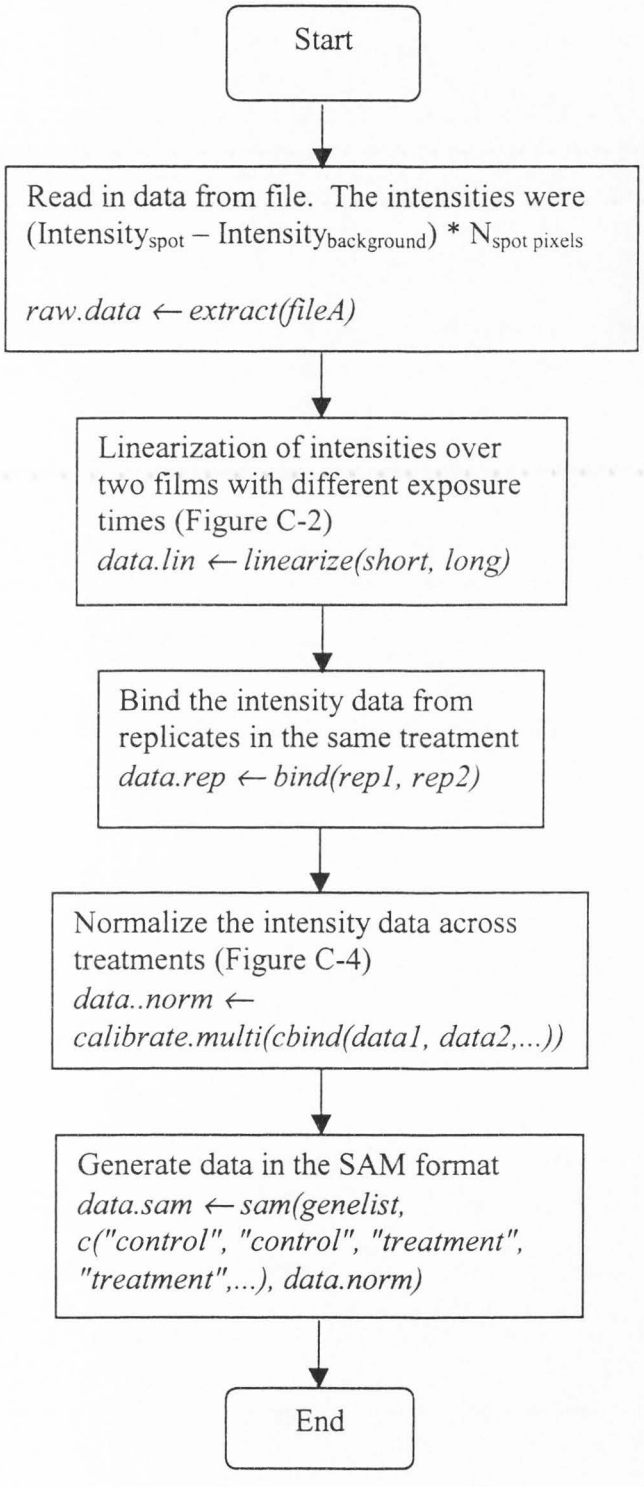


Figure C-1. General flowchart of preprocessing of DNA macroarray data. The functions written in R language, their parameters and return results were present as *italics*. The description of the purpose of each function was present with normal font.

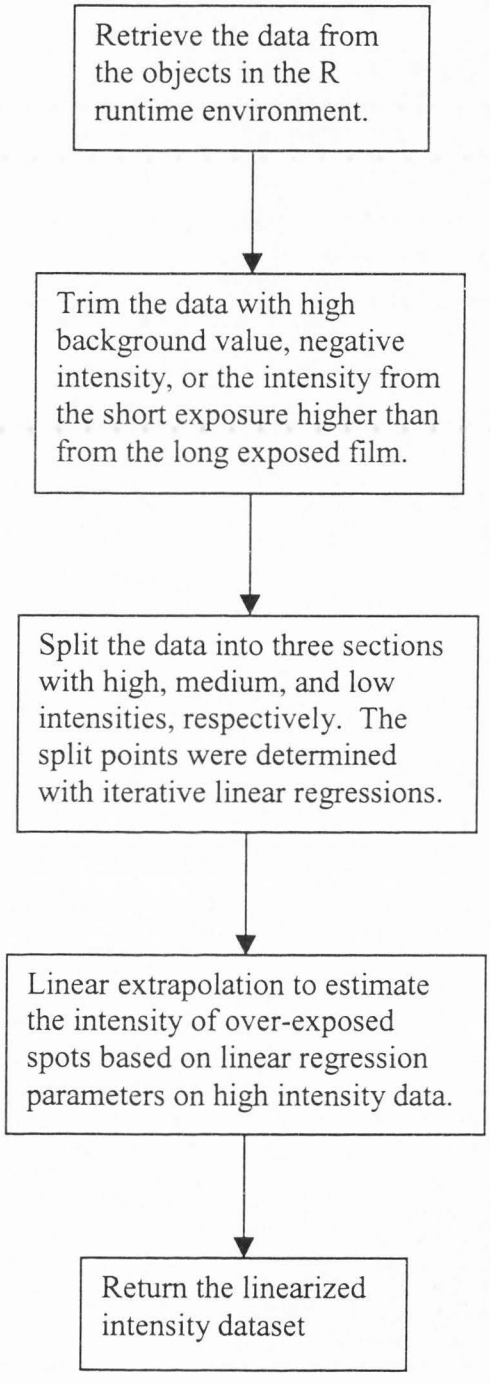


Figure C-2. Detailed flowchart of the algorithm in *linearize* function.

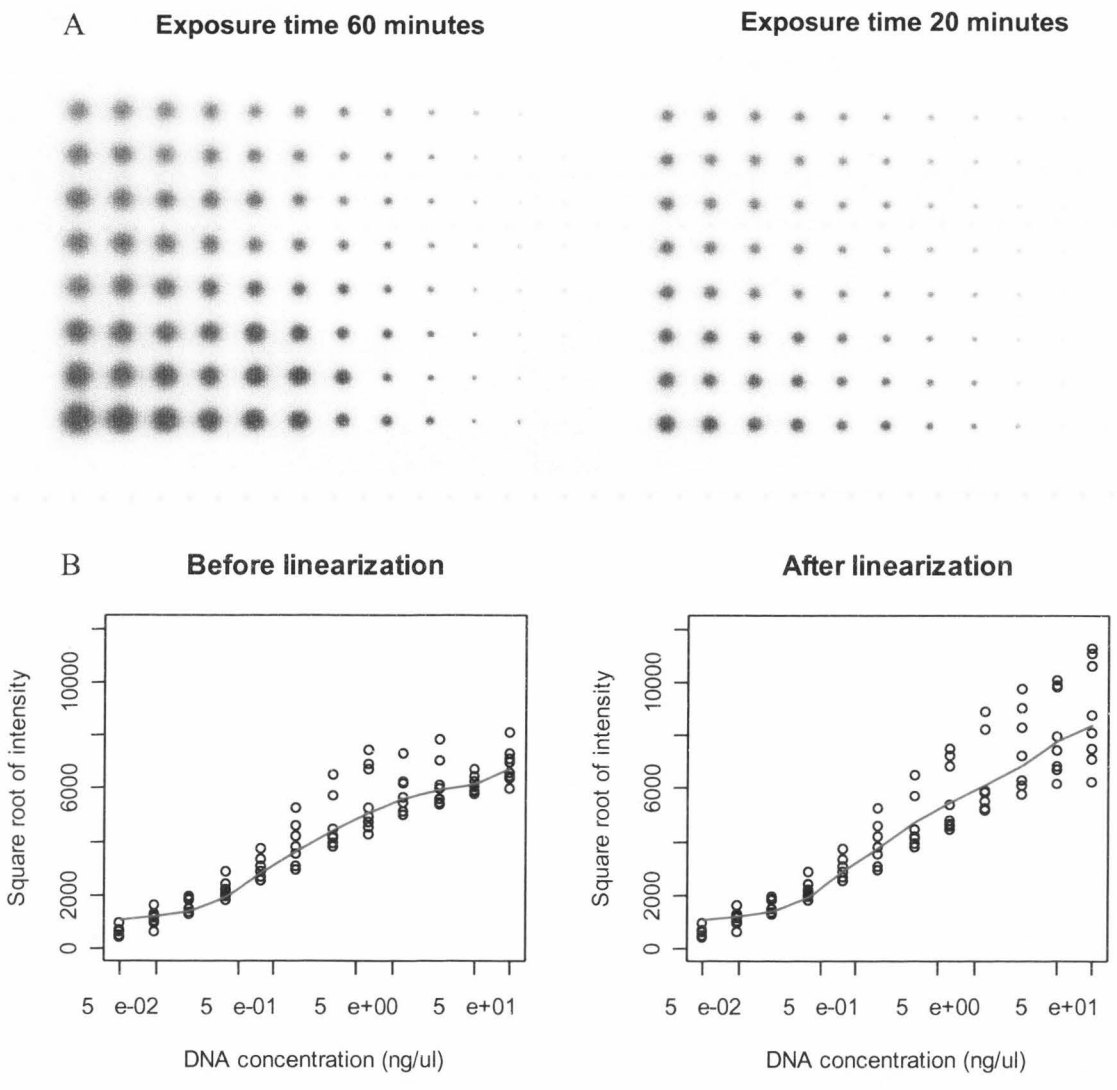


Figure C-3. Linearization of DNA macroarray data.
A. Differentially exposed films derived from the same DNA macroarray blot. The DNA macroarray blot was prepared with fragmented DNA labeled with BrightStar® Psoralen-Biotin Nonisotopic Labeling Kit (Ambion Inc., Austin, Texas). The biotinylated DNA was then diluted at the concentrations of 100, 50, 25, 12,5, 6.25, 3.125, 1.563, 0.781, 0.391, 0.195, 0.098, 0.049 µg/µl, and spotted on the nylon membrane from left to right with eight column-wise replicates. The biotin-based chemiluminescent detection was conducted with a North2South Hybridization and Detection kit (Pierce, Rockford, Illinois) according to the kit instructions, except that blocking time was extended from 15 minutes to 1 hour. The membrane blot was exposed to chemiluminescence films (Roche Applied Science, Indianapolis, Indiana) for 20 and 60 minutes at room temperature. B. Before and after linear correction, the relationship between square root of spot intensity from 60-minute exposure film and true labeled DNA concentration. Red line was the *lowess* curve to represent the trends in the data.

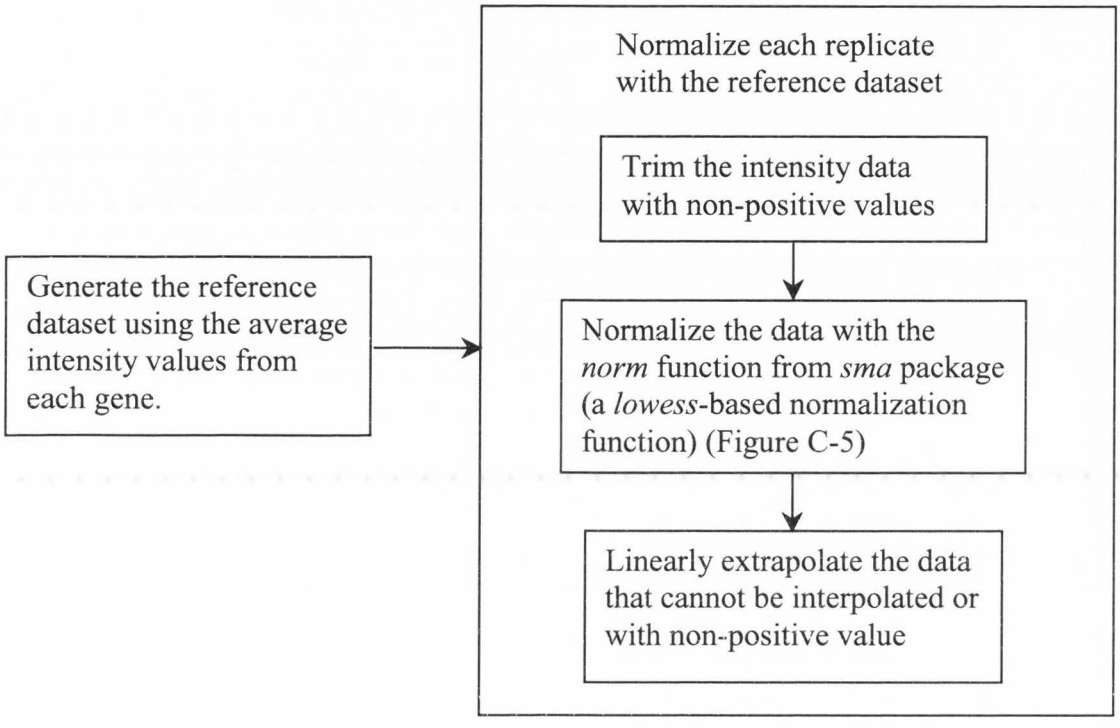


Figure C-4. Detailed flowchart of the algorithm in *calibrate.multi* function.

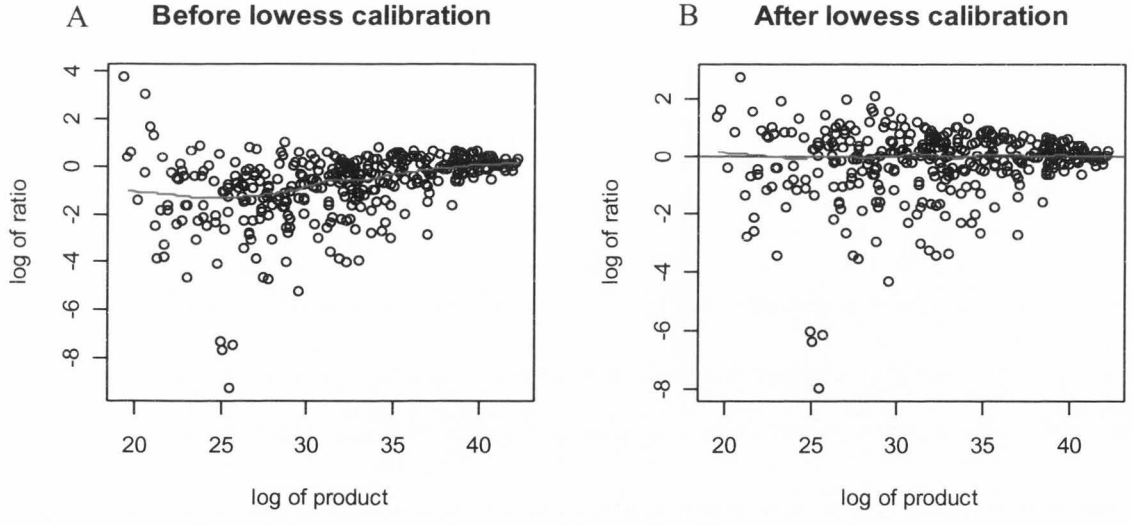


Figure C-5. MA plot of normalization on DNA macroarray data.
A. MA plot of the un-normalized data. B. After the *lowess* calibration, the data were normalized so that the ratio between sample and standard was around 1. Red line was the *lowess* curve and green line was the curve where the log ratio was 0.

Appendix D. Supplemental Data on Determine of Differentially Expressed Genes

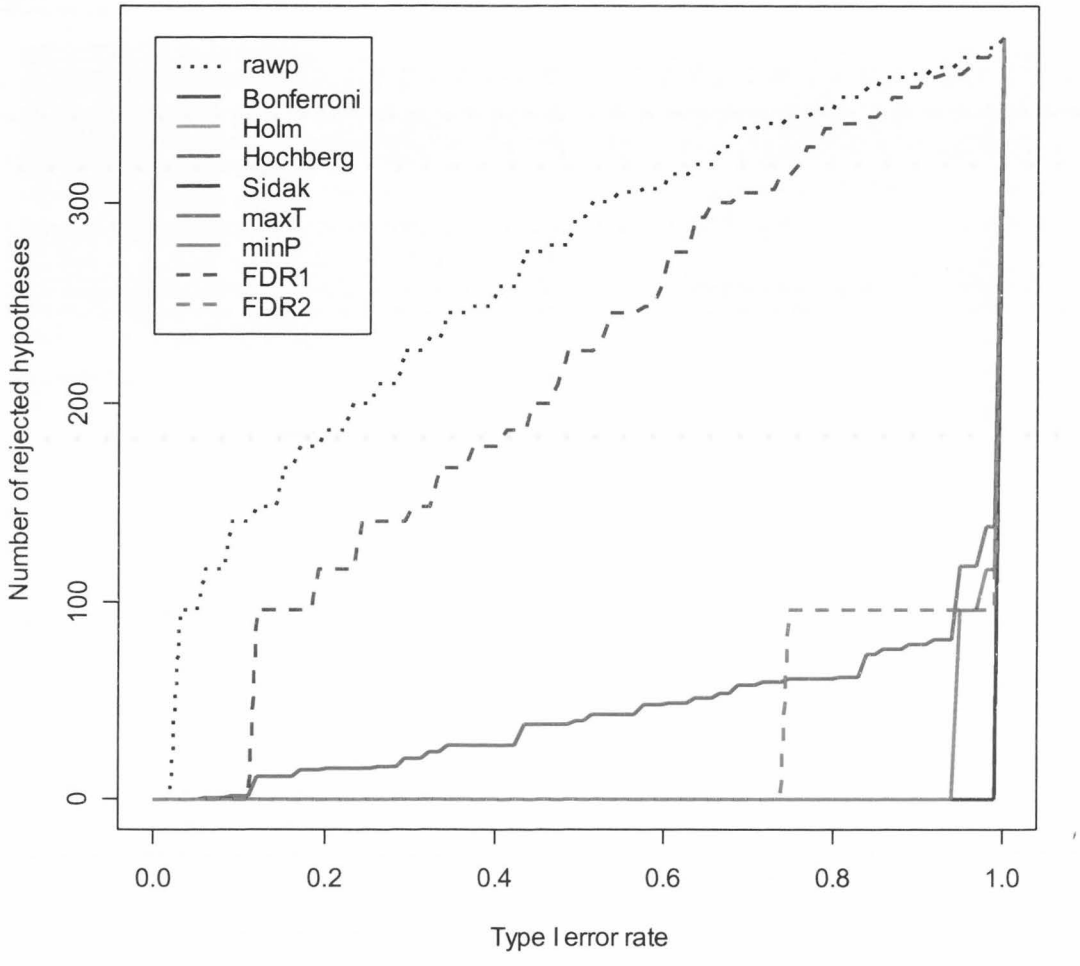
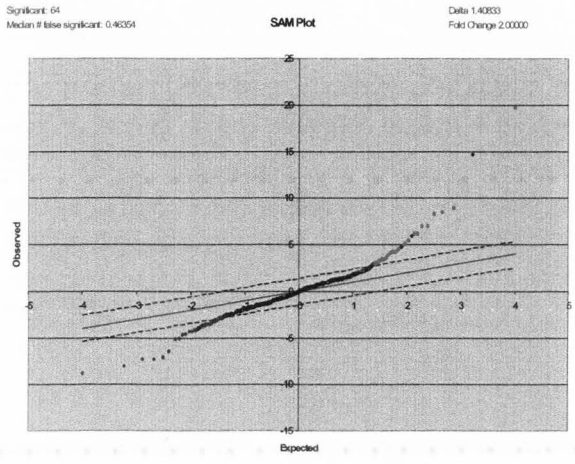
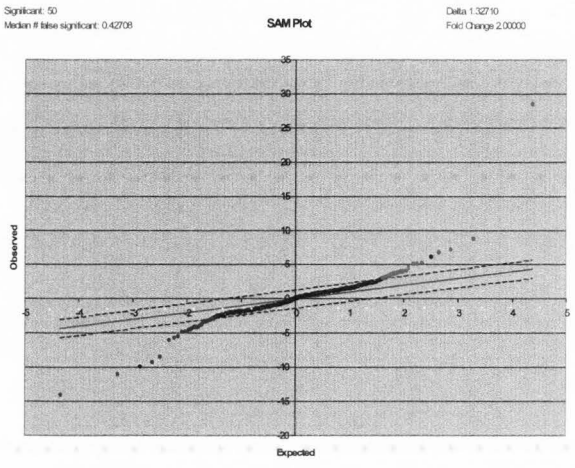


Figure D-1. Comparison of performance of different statistical procedures on detection of differentially expressed genes.
Bonferroni, Holm, Hochberg, Sidak, and Westfall&Yang (maxT and minP) methods were controlling FWER. Benjamini&Hochberg (FDR1) and Benjamini&Yekutieli (FDR2) methods were FDR-based methods. The dataset was derived from analysis of the expression profiles of *L. lactis* ssp. *lactis* IL1403 under normal and heat stress conditions.

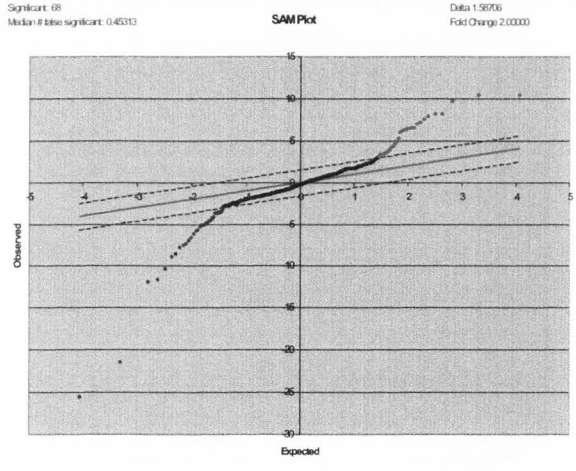
A. Heat vs. Control



B. Acid vs. Control



C. Salt vs. Control



D. Peptide starvation vs. BL2 peptide

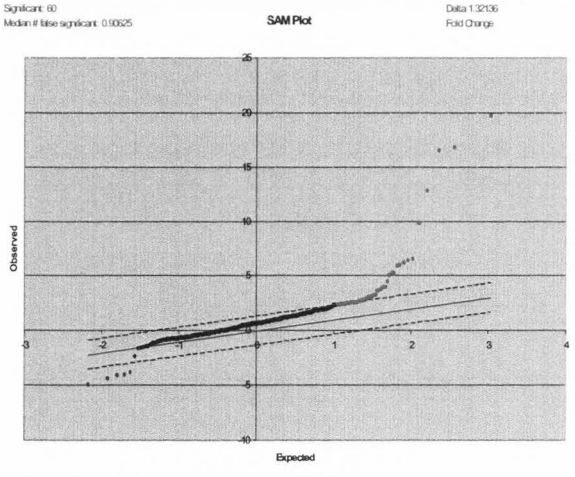
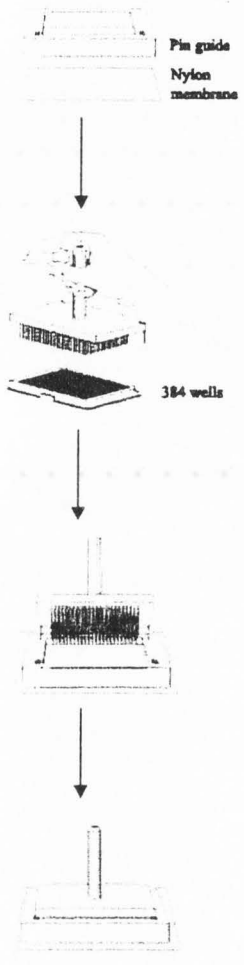


Figure D-2. Significance analysis for differentially expressed genes with SAM. The criteria used to generate these plots were low FDR (< 1%) and minimal 2-fold change in the expression profile ratio, excepting peptide utilization plot (D). A. Analysis of heat shock responses relative to control. B. Analysis of acid shock responses relative to control. C. Analysis of osmotic shock responses relative to control. D. Analysis of expression profiles in BL2-peptide medium relative to the ones in peptide-limited and casitone media. The red dots indicated the genes that were induced significantly, while green dots indicated the genes that were repressed significantly. The blue line was the diagonal line, where ideal null hypotheses were located. The value of delta determined the regions between two dashed black lines, which contained the genes whose expressions were not differentially expressed represented by black dots.

Appendix E. Supplemental Data on DNA Macroarray Fabrication and Usage



Assembly pin registration tool and nylon membrane

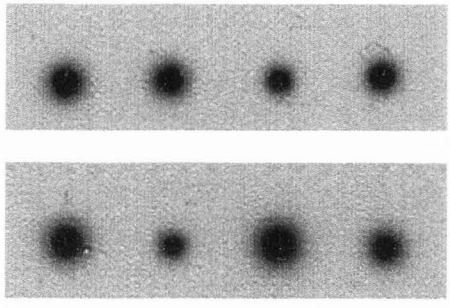
Dip the pins into 384 well microplate containing polyI tailed oligonucleotide probes

Align guiding pins into proper alignment holes

Deliver the liquid onto nylon membrane

Figure E-1. Flowchart of DNA macroarray fabrication.

A.



B.

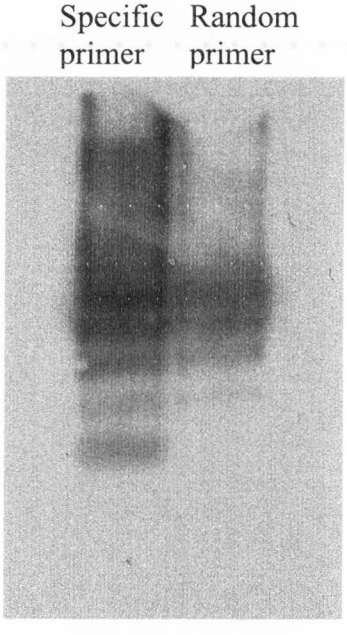
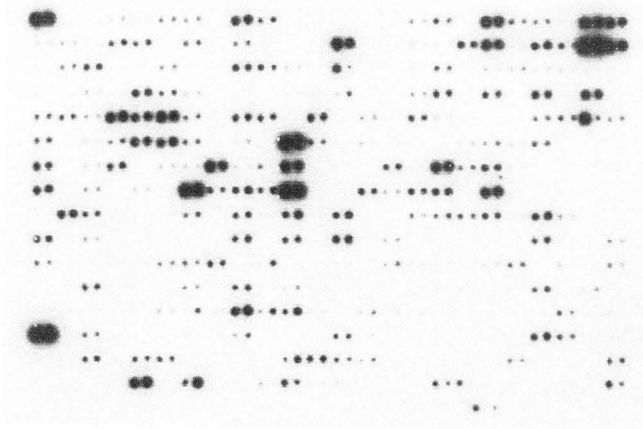


Figure E-2. Direct labeling of biotin via reverse transcription by random primers. A. Effect of increasing the ratio of biotin-16-dUTP in dNTP mixture. B. Effect of using specific primer versus using random primer.

A. Initial (M17G Referesh)



B. BL2-peptide medium (T=120min)

C. Peptide-limited medium (T=120min)

D. Peptide-rich casitone medium (T=120min)

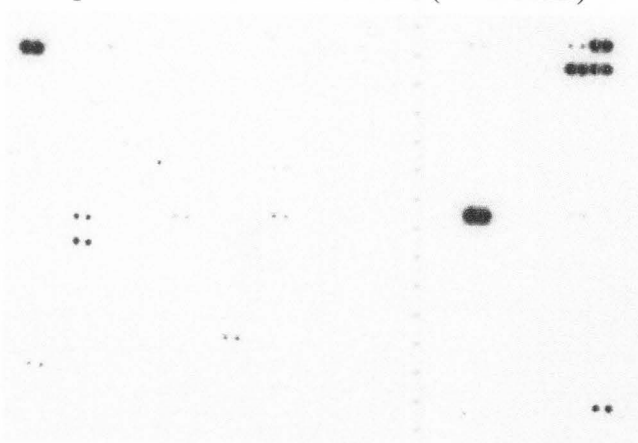


Figure E-3. Expression profiles of *L. lactis* spp. *lactis* IL1403 in media with different peptide pool.
(A) Expression profiles after 10 min fresh in M17G medium; (B) Expression profiles in aBL2-peptide medium; (C) Expression profiles in apeptide-limited medium; (D) Expression profiles in a peptide-rich casitone medium.

Appendix F. Supplemental Data on Characterization of Preactivated Protease from *Brevibacterium linens* BL2

Protease Activity Assaying

Protease activity was measured by hydrolysis of fluorescent-labeled casein (EnzCheck, Molecular Probes, Eugene, OR). The assay mixture was prepared on a 96-well plate (FluoroNunc, Nalge Nunc International, Denmark). The protease (pre-activated at 30°C for 16 hr) was diluted 10 fold in 10 mM Tris-HCl buffer (pH 7.8), containing 20 mM sodium azide. The enzyme working solution was then mixed with an equal volume of fluorescent-labeled casein substrate solution (10 μg/ml in the same Tris-HCl buffer) into total volume 200 μl, and incubated at 30°C on plate-reader (HTS 7000, Pelkin Elmer Co., Norwalk, CT). The fluorescence intensities were monitored at 10 minute time intervals for 2 hours at excitation/emission wavelengths of 485 nm/535 nm. For inhibitor and heterogeneous protease activation experiments, the specific inhibitor or protease was first diluted with 10mM Tris-HCl buffer (pH 7.0), and then this solution was used to prepare enzyme dilutes. All assays were conducted in replicates.

In addition to the whole casein substrate supplied in EnzCheck® kit, the individual casein component was also custom-conjugated with different fluorescence dyes (Molecular Probes, Eugene, OR), and used in optimal pH and temperature determination. The tetramethlyrhodamine was conjugated into α-casein at the density of 11.7 moles dye/mole, and detected excitation/emission wavelengths of 540 nm/565 nm. The Texas Red was conjugated into β-casein with at the density of 7.4 moles dye/mole, and detected with excitation/emission wavelengths of 580 nm/605 nm. The fluorescein was conjugated into κ-casein with at the density of 8.1 moles dye/mole, and detected with excitation/emission wavelengths of 480 nm/520 nm.

Optimal Temperatures of Preactivated *B. linens* BL2 Protease on Different Caseins

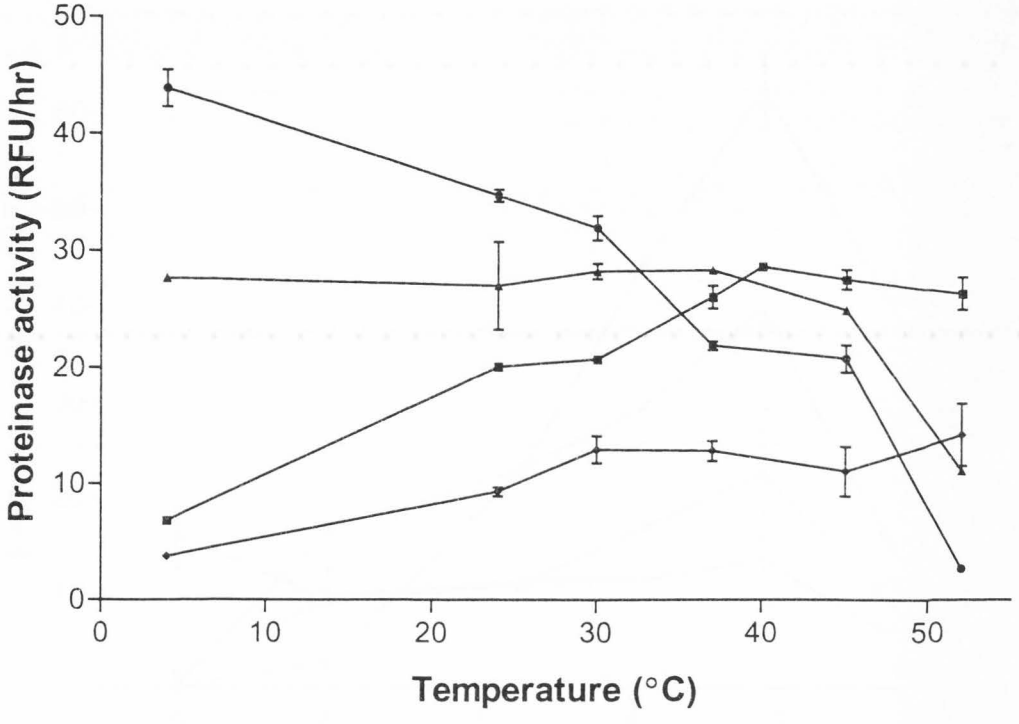


Figure F-1. Optimal temperature of preactivated *B. linens* BL2 protease. ■ protease activity of *B. linens* BL2 with whole casein as substrate; ▲ protease activity of *B. linens* BL2 with α-casein as substrate; ◗protease activity of *B. linens* BL2 with β-casein as substrate; ● protease activity of *B. linens* BL2 with κ-casein as substrate.

The optimal temperature profiles between whole casein and other casein components were fairly different. With whole casein as the hydrolysis substrate, the optimal temperature was 40°C, while with labeled β-casein as substrates, the optimal temperatures were around 30-37°C. When labeled κ-casein was used as the substrate, the active protease from *B. linens* BL2 was significantly active in 4°C. There was no obvious optimal temperature when labeled α-casein was the substrate.

Optimal pH of Preactivated *B. linens* BL2 Protease on Different Caseins

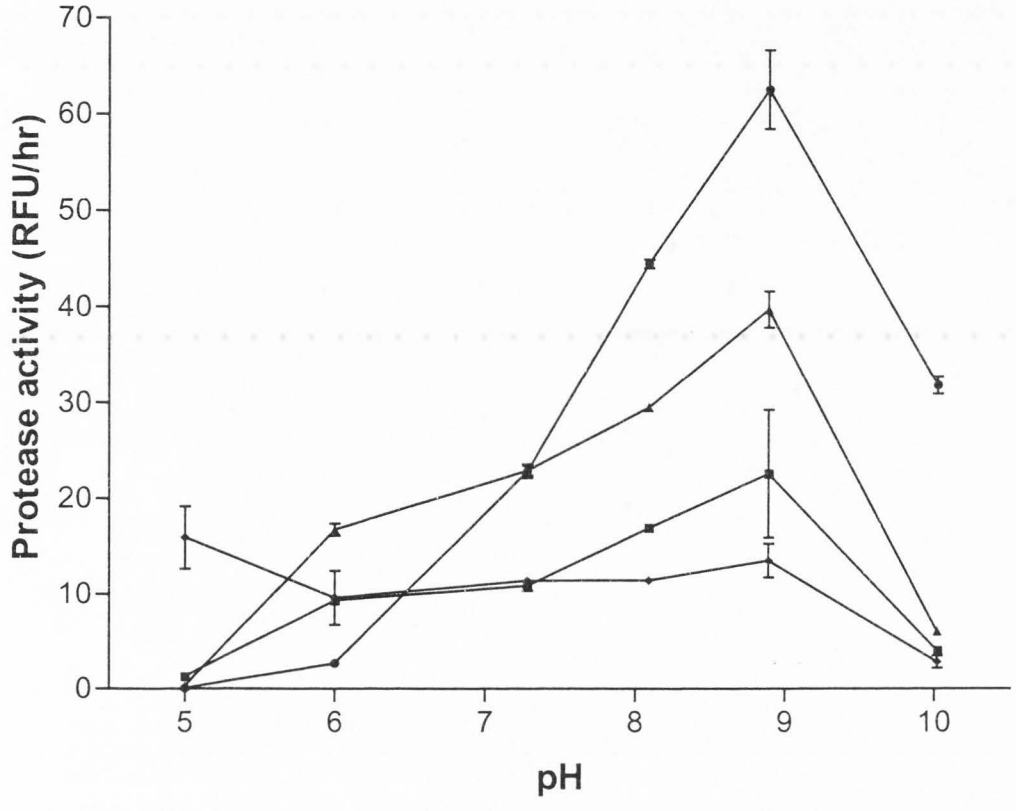


Figure F-2. Optimal pH of preactivated *B. linens* BL2 protease.
■ protease activity of *B. linens* BL2 with whole casein as substrate; ▲ protease activity of *B. linens* BL2 with α-casein as substrate; ◗protease activity of *B. linens* BL2 with β-casein as substrate; ● protease activity of *B. linens* BL2 with κ-casein as substrate.

The pre-activated protease from *B. linens* BL2 had an optimal pH 8.8 on whole casein, as did on α-casein, β-casein, and κ-casein. Proteolytic activity profiles over the range of pH 5-10 were identical across labeled whole casein, α-casein, and κ-casein, but slightly different on β-casein, where active protease from *B. linens* BL2 had a wide range of activities over pH 5-9.

Influence of Activators and Inhibitors of *B. linens* BL2 Protease

Table F-1. Activators and inhibitors of protease activity from *B. linens* BL2.

Effector	Residual activity (%)
Mg^{2+}	139.3*
Ca^{2+}	138.6*
Cysteine	113.8
β-mercaptoethanol	111.6
Iodoacetic acid	80.7*
Mn^{2+}	80.2**
PCMB	79.8**
Fe^{3+}	77.3**
PMSF	64.0*
Fe^{2+}	53.8**
Ni^{2+}	48.6**
Co^{2+}	46.9**
Cu^{2+}	28.7**
Zn^{2+}	11.4**
EDTA	9.3**
Cd^{2+}	6.2**
Hg^{2+}	NA[1]**

[1] The net activity with Hg^{2+} was negative, so residual activity cannot be estimated.

* The protease activity was different from the control activity with p-value smaller than 0.05.

** The protease activity was different from the control activity with p-value smaller than 0.001.

With the pre-activated protease from *B. linens* BL2, we also examined the influences of a variety of potential activators and inhibitors on the performance of this enzyme (Table A-1). This protease was essentially a serine protease, as it was sensitive to serine protease inhibitors, iodoacetic acid, PCMB, and PMSF. It was also very sensitive to metal chelator EDTA, and activated by Mg^{++} and Ca^{++} ions, suggesting that these two metal ions were likely to have structural roles as suggested in purification. In addition, Cd^{2+} and Hg^{2+} were extremely toxic to this enzyme and 1mM concentration of these heavy metal ions was sufficient to completely inhibit the proteolytic activity from

activated protease. Other ions also had different levels of inhibition effects on the proteolytic activity.

Appendix G. Permission Letters

January 30, 2003

Yi Xie
Department of Nutrition and Food Science
Utah State University
Logan, Utah 84322
E-mail: yixie@cc.usu.edu

Dear Dr. Chou:

I am in the process of preparing my dissertation in the Influence of the Extracellular Proteinase from *Brevibacterium linens* on the Metabolism of *Lactococcus lactis* spp. *lactis* at Utah State University. I hope to complete my degree in the spring of 2003.

Two articles, Expression Profiling of *Lactococcus lactis* ssp. *lactis* IL1403 under Stresses with DNA Macroarray, and Expression Profiling of *Lactococcus lactis* ssp. *lactis* IL1403 with Different Peptide Sources, of which you are one of co-authors, compose an essential part of my dissertation research. I would like your permission to reprint it as a chapter in my dissertation. Please note that USU sends dissertation to Bell & Howell Dissertation Services to be made available for reproduction.

I will include an acknowledgement to the article on the first page of the chapters, as shown below. Copyright and permission information will be included in a special appendix. If you would like a different acknowledgement, please so indicate.

Please indicate your approval of this request by signing in the space provided, and attach any other form necessary to confirm permission. If you have any questions, please send me an email message at the above address. Thank you for your assistance.

Yi Xie

Xie Yi

I hereby give permission to Yi Xie to reprint the requested material in his dissertation.

Signed _____

Date Jan 31, 2003

January 30, 2003

Yi Xie
Department of Nutrition and Food Science
Utah State University
Logan, Utah 84322
E-mail: yixie@cc.usu.edu

Dear Dr. Brown:

I am in the process of preparing my dissertation in the Influence of the Extracellular Proteinase from *Brevibacterium linens* on the Metabolism of *Lactococcus lactis* spp. *lactis* at Utah State University. I hope to complete my degree in the spring of 2003.

An article, the Proteolysis and Maturation of *Brevibacterium linens* Protease, of which you are one of co-authors, compose an essential part of my dissertation research. I would like your permission to reprint it as a chapter in my dissertation. Please note that USU sends dissertation to Bell & Howell Dissertation Services to be made available for reproduction.

I will include an acknowledgement to the article on the first page of the chapter, as shown below. Copyright and permission information will be included in a special appendix. If you would like a different acknowledgement, please so indicate.

Please indicate your approval of this request by signing in the space provided, and attach any other form necessary to confirm permission. If you have any questions, please send me an email message at the above address. Thank you for your assistance.

Yi Xie

Xie Yi

I hereby give permission to Yi Xie to reprint the requested material in his dissertation.

Signed /ax

Date January 31, 2003

CURRICULUM VITAE

Yi Xie

Education

Ph. D. in Food Microbiology, Utah State University, Logan, Utah. GPA: 3.9/4.0
M.S. in Computer Science, Utah State University, Logan, Utah. GPA: 3.9/4.0
M. S. in Microbiology, Fudan University,Shanghai, China. GPA: 3.6/4.0.
B. E. in Food Engineering, Wuxi Institute of Light Industry Wuxi, China. GPA: 3.2/4.0

Professional Experience

RESEARCH ASSISTANT
1995 - Present Utah State University Logan, Utah
Develop a low-cost oligonucleotide-based filter DNA macroarray protocol and integrated computational tools for experimental design and data analysis
Profiling gene expression changes of *Lactococcus lactis* IL1403 under stress conditions with DNA macroarray
Monitoring gene expression of starter culture during cheese ripening by using DNA macroarray
Purification and characterization of proteinase from *Brevibacterium linens*
Characterization and kinetic studies of activity and maturation proteinase from *B. linens*
Cloning, expression, isolation the MscL membrane protein from *E. coli*

RESEARCH ASSISTANT
1992 - 1995 Fudan University Shanghai, China
Studies of stabilizing mechanism for *Kluveromyces* lactis β-galactosidase
Immobilization of whole-cell *K. lactis* β-galactosidase
Leading the research project on enzymatic production of fructo-oligosaccharides

ASSISTANT MANAGER
1990 - 1992 Information Center of Light Industry, Government of Heilongjiang Province Harbin, China
Data collection from light industry related manufactures
Data processing and reporting to the Governor

Professional Training

2001/07 – 2001/07 Pittsburgh Supercomputing Center Pittsburgh, PA
Summer Workshop of Nucleic acids and Protein Sequence Analysis sponsored by National Institute of Health (NIH)

1998/07 – 1998/08 University of Georgia Athens, GA
Summer Workshop of Microbial Physiology sponsored by National Science Foundation (NSF)

Publications and Presentations

L. Chou, **Y. Xie**, and B. Weimer (2001)
Studies of arginine metabolism and its genetic regulations in *Lactococcus. lactis* spp. *lactis* ML3 (submitted)

Y. Xie, P. Jiang, and J. Guo (1999)
Studies on stability of *Kluveromyces lactis* β-galactosidase. Journal of Fudan University (Natural Science) Vol. 38 (5) p524-8

Y. Xie, A. Cutler and B. C. Weimer 2002
Linearization of DNA macroarray data. The 10th International Conference on Intelligent Systems for Molecular Biology, Edmonton, Canada

Weimer, B. C., Lan–Szu Chou, and **Yi Xie**. 2002.
Using expression arrays to determine gene expression during cheese ripening. LAB7, The Netherlands.

Weimer, B. C., P. Joseph, R. Koka, and **Y. Xie**. 2002.
Use of non-lactic acid bacterial enzymes to modify the peptide pool in Cheddar cheese. LAB7, The Netherlands.

Y. Xie and B. C. Weimer. 2002.
Use of oligonucleotide-based DNA macroarray to profile the metabolic responses of *Lactococcus lactis* spp. *lactis* IL1403 to an exogenous microbial protease. The American Society for Microbiology Annual Meeting, Salt Lake City.

Weimer, B. C., P. Joseph, R. Koka, and **Y. Xie**. 2002.
Use of non-lactic acid bacterial enzymes to degrade casein. Institute of Food Technologists annual meeting, Anaheim.

Weimer, B. C., **Y. Xie**, and Rod Brown. 2000.
Autocatalytic processing of the protease from *Brevibacterium linens* BL2: a kinetic analysis for the degradation of casein. International Dairy Federation Biennial Cheese Flavor Conference, Banff, Alberta.

Honors and Awards Received

- Gandhi Fellowship of Utah State University. 1998
- President Fellowship of Utah State University. 1995
- Numerous times on Utah State University President Honor Roll, 1995-2000
- 1st in the nation-wide graduate entrance exam at the College of Life Sciences, Fundan University. 1992
- Excellent Graduate of Wuxi Institute of Light Industry. 1990
- 1st in Higher Mathematics Competition at Wuxi Institute of Light Industry. 1987
- Winner of National High School Mathematics Competition in China. 1985