The Cytoplasmic N,n'-Diacetylchitobiase Gene From Vibrio Parahaemolyticus: Sequence Analysis, Protein Secretion, and Secretion System Development.

Michael Hongli Wu
Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_disstheses

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_disstheses/6171

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700  800/521-0600
THE CYTOPLASMIC N,N'-DIACETYLCHITOBIASE GENE FROM VIBRIO PARAHAEOMOLYTICUS: SEQUENCE ANALYSIS, PROTEIN SECRETION, AND SECRETION SYSTEM DEVELOPMENT

A Dissertation

Submitted to the Graduate Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in

The Department of Biochemistry

by

Michael Hongli Wu
B.S., Beijing Forestry University, P. R. China, 1982
M.S., University of California, Riverside, CA, 1989
May 1996
ACKNOWLEDGMENTS

I am extremely grateful to my major professor, Dr. Roger A. Laine, for his understanding, patience, encouragement, and general support during this study and the preparation of manuscripts and this dissertation. He never seems to run out of new ideas that sometimes, somewhere, will shine some light on someone's project in the lab.

I am also indebted to other members of my advisory committee, Drs. William H. Daly (Chemistry), Donal F. Day (Audubon Sugar Institute), Ding S. Shih (Biochemistry), Patrick J. DiMario (Biochemistry), Jan L. VanSteenhouse (Veterinary Pathology, School of Veterinary medicine) for their valuable suggestions, corrections, and other efforts during reviewing this dissertation. In addition, I also highly appreciate Drs. Simon H. Chang (Biochemistry) and Jesse M. Jaynes (Formerly Biochemistry) for their general support and technical assistance to my research program while on my committee. My gratitude is extended to Dr. Betty C. R. Zhu for her assistance and valuable suggestions that are always available.

Last, but certainly not the least, I would like to express my gratitude to my wife, Kathy G. Wu. Without her understanding, patience and support, this study would not have been completed. I'm also in debt to my two sons, Peter and Clifford, with whom I should have played more as I always promised.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS .......................................................................................................................... iii

LIST OF FIGURES .................................................................................................................................... v

ABSTRACT ................................................................................................................................................ vi

CHAPTER 1 REVIEW OF THE LITERATURE .......................................................................................... 1
  1.1 INTRODUCTION .............................................................................................................................. 1
  1.2 CHITIN .............................................................................................................................................. 2
  1.3 CHITIN UTILIZATION BY MARINE BACTERIA ............................................................................ 3
  1.4 PROTEIN SECRETION .................................................................................................................... 8
  1.5 OVEREXPRESSION OF CLONED GENES ................................................................................... 29
  1.6 PCR AND RECOMBINANT DNA .................................................................................................... 36

CHAPTER 2 SEQUENCING AND SEQUENCE ANALYSIS ..................................................................... 48
  2.1 INTRODUCTION ............................................................................................................................ 48
  2.2 MATERIALS AND METHODS ......................................................................................................... 48
  2.3 RESULTS ....................................................................................................................................... 50
  2.4 DISCUSSION .................................................................................................................................. 65
  2.5 SUMMARY .................................................................................................................................... 69

CHAPTER 3 PROTEIN SECRETION ................................................................................................. 71
  3.1 INTRODUCTION ............................................................................................................................ 71
  3.2 MATERIALS ..................................................................................................................................... 72
  3.3 EXPERIMENTAL PROCEDURES AND RESULTS ....................................................................... 73
  3.4 DISCUSSION .................................................................................................................................. 91
  3.5 SUMMARY .................................................................................................................................... 95

CHAPTER 4 CONCLUSIONS ............................................................................................................. 98

REFERENCES ......................................................................................................................................... 101

APPENDIXES .......................................................................................................................................... 127
  A.1 SEQUENCING AND PCR PRIMERS ......................................................................................... 127
  A.2 SIGNAL SEQUENCES USED IN THIS STUDY ........................................................................... 128
  A.3 SECRETION VECTOR SEQUENCES ......................................................................................... 129
  A.4 GENOTYPES OF SOME HOST STRAINS .................................................................................... 131
  A.5 STRUCTURES OF RELEVANT COMPOUNDS ........................................................................... 132
  A.6 LIST OF ABBREVIATIONS ........................................................................................................... 137

VITA ......................................................................................................................................................... 138
LIST OF FIGURES

Fig. 1 Chitin degradation pathways in vibrios .......................................................... 6
Fig. 2 Construction of the tac promoter ................................................................. 30
Fig. 3 Vectorette and its application ....................................................................... 40
Fig. 4 Construction of a possible universal vectorette ............................................ 41
Fig. 5 Restriction map and nucleotide sequencing strategy .................................... 52
Fig. 6 Nucleotide sequence of the chitobiase gene .............................................. 53
Fig. 7 Amino acid alignment of three vibrio chitobiases ....................................... 56
Fig. 8 Amino acid sequence pileup ......................................................................... 60
Fig. 9 Hydrophobicity analysis ................................................................................ 67
Fig. 10 In-frame connection by recombinant PCR ................................................ 75
Fig. 11 Two clones for the secretion of cytoplasmic chitobiase ............................. 77
Fig. 12 Secretion and protein leakage ..................................................................... 79
Fig. 13 Silent mutagenesis ...................................................................................... 84
Fig. 14 Construction of pVerSec and pVerSec-SE .................................................. 85
Fig. 15 Application of secretion vectors .................................................................. 87
Fig. 16 Effects of EDTA on protein protection and host growth .......................... 88
Fig. 17 Chitobiase activity and EDTA concentration ........................................... 89
ABSTRACT

The nucleotide sequence of the gene encoding the cytoplasmic $N,N'$-diacetylchitobiase from *Vibrio parahaemolyticus* has been determined. The deduced peptide sequence surprisingly has minimum evolutionary relationship to two other reported $N,N'$-diacetylchitobiases from vibrios, except for highly conserved regions which are also homologous with lysosomal beta-hexosaminidases from eukaryotes including humans. In contrast, the other two sequenced chitobiases from vibrios are much more closely related to each other. This 85 kDa cytoplasmic protein, as revealed from the sequence, appears to be a unique protein, lacking a signal sequence and genetically distant from other known enzymes of similar function. This is consistent with its limited substrate specificity to small $N$-acetylglucosamine terminated oligosaccharides.

The signal peptide of an extracellular endochitinase from *V. parahaemolyticus* causes the mature chitinase to be efficiently secreted through the double membranes of both Gram negative *V. parahaemolyticus*, and of *Escherichia coli* JM101 when the gene is cloned therein. By using recombinant PCR, this signal sequence was fused in frame to the chitobiase coding sequence, and active chitobiase was found in the *E. coli* culture medium.

Two secretion vectors were developed during this study for the secretion of wild type proteins. PCR fragment of a structural gene can be inserted in frame with the signal sequence, which is under the strong *trc* promoter. The secretion system has been tested using the cytoplasmic chitobiase gene as a model system. Restriction sites and complete DNA sequence are not required for the structural gene to be cloned. The
chitobiase and the original chitinase as cloned for secretion were stabilized by EDTA added to the medium. This observation may prove generally useful for protecting cloned and secreted proteins in *E. coli*. 
CHAPTER 1 REVIEW OF THE LITERATURE

1.1 INTRODUCTION

1.1.1 Historical Overview

Chitin is a polymer of beta-D-N-acetyl-glucosamine (GlcNAc). Its chemical structure, natural abundance and biological function resemble those of cellulose found in plants. The primary source of this insoluble particulate is the marine environment, where many organisms use chitin as their major skeletal or cell wall components. Recycling of chitin for carbon and nitrogen sources is a complex biochemical process, involving many proteins. The Gram negative bacteria, vibrios, play a major role in the degradation of chitin.

Two parallel pathways in vibrios have been postulated for chitin catabolism (273). Chemotactic proteins act as receptors for trehalose and GlcNAc released from dead crustaceans; lectins help the bacteria adhere to the chitin (77, 274), and extracellular chitinase (123) and periplasmic chitodextrinase sequentially degrade the substrate into N,N'-diacetylchitobiose (15). The glycosidase/PTS system cleaves N,N'-diacetyl chitobiose to GlcNAc in the periplasm via a membrane bound chitobiase (15, 220, 265) after which GlcNAc is transported and phosphorylated by the bacterial phosphoenolpyruvate:glycosephosphotransferase system (PTS) (15). In the alternative system, an N,N'-diacetylchitobiase permease translocates this substrate to the cytoplasm, where it is cleaved by a cytoplasmic chitobiase (278) and phosphorylated by an ATP-dependent N-acetyl-D-glucosamine kinase (6, 12, 15). The cytoplasmic system works independently of the PTS (15). The two pathways both produce Fru-6-P from chitin, which enters glycolysis for further metabolism.
1.1.2 The Objectives of This Study

The cloning and characterization of the extracellular chitinase (123) and cytoplasmic chitobiase (278) from \textit{V. parahaemolyticus} have shed light on the chitinoclastic pathway involved in this organism. The objectives of this project include:

1. Nucleotide sequence determination of the gene encoding the cytoplasmic chitobiase from \textit{V. parahaemolyticus} to reveal the cytoplasmic features of the enzyme and its relationship to other chitobiases and beta-hexosaminidases.

2. The secretion of chitobiase by construction of a gene for a fusion protein using the signal peptide of chitinase gene to demonstrate that a 21 amino acid leader peptide is responsible for secretion of mature protein across membranes, and that the gene product in the supernatant of \textit{E. coli} culture is not simply the consequence of leakage or cell lysis.


4. Solve the problem of instability of the secreted chitinase.

1.2 CHITIN

Chitin is a homopolysaccharide composed of beta-(1,4)-linked 2-acetamido-2-deoxy-\textit{D}-pyranosyl residues. Considering the similarities in structure and function between chitin and cellulose, chitin can be considered as a derivative of cellulose in which the C2 hydroxyl groups of the building blocks have been replaced by acetamido residues (67). After cellulose, it is one of the most abundant organic compounds in nature. Chitin sediments produced as copepod exoskeletons alone have an estimated productivity of several billion tons annually (166).
The earliest description of this polysaccharide (29, 67) was in 1811 by Braconnot (26) who extracted this alkali resistant polysaccharide from several species of higher fungi (Agaricus volvaceus, A. acris, Hydnum repandum, H. hybridum, Merulius cantharellus, A. cantharellus and Boletus viscidus) and named it ‘fungine’. The name chitin (Greek, tunic or covering) was first proposed by Odier in 1923 to describe the substance that he isolated from the May beetles elytra (67). It is the major exoskeleton component of insect, crustacean and arachnoids. The molecular weight of this polysaccharide has been determined to be 1.0 x 10^6 Da for chitin purified from crabs, and higher than 1.8 x 10^6 Da from other crustaceans (30, 210). Because of the insoluble nature and its association with other macromolecules, isolation and purification of chitin frequently require drastic methods that partially degrade chitin (67). The purified chitin used in determining its molecular weight may be a degradation product of natural chitin.

1.3 CHITIN UTILIZATION BY MARINE BACTERIA

1.3.1 Vibrio Parahaemolyticus

V. parahaemolyticus is a member of the Gram negative, rod shaped bacterial genus, Vibrio. At least 11 vibrio species have been reported pathogenic to man. Three of them, namely V. cholera, V. parahaemolyticus and V. vulnificus, are of medical concern. The symptoms associated with V. parahaemolyticus infection include mild diarrhea, abdominal cramps, nausea, vomiting and fever (167). This was believed to be related to the high incidence of gastroenteritis and extraintestinal infections in coastal areas during warmer month of the year when the organism is prevalent.
1.3.2 The Bacterial PTS and Sugar Uptake

The bacterial phosphoenolpyruvate:glycose phosphotransferase system (PTS) is a complex system with diverse functions in bacterial cells. The PTS is widely distributed among most species of bacteria (149). At least 15 species of vibrios, including V. parahaemolyticus and V. vulnificus, contain a PTS (150). Several proteins in the PTS have been characterized. The general proteins of the pathway consist of Enzyme I (EI) (40, 118, 199, 253, 255) and HPr (20, 40, 253, 254). The sugar specific proteins in the PTS require specific sugar substrate. Known examples of sugar specific proteins characterized from S. typhimurium include: IIB^Man for mannone (119), III^Glc for Glucose (151), and II^Nag for GlcNAc and methyl-GlcNAc (83, 252). Some sugar specific proteins have more than one sugar substrate. On the other hand, some sugars (e.g., glucose) are substrates of several different proteins.

The system works by translocation and phosphorylation of a particular sugar through a series of membrane bound and cytosolic proteins. The phosphoryl group is transferred sequentially from phosphoenolpyruvate (PEP) to EI, to HPr, and then to a membrane bound sugar specific protein, depending on the sugar substrate. In the case of glucose (Glc), for example, the sugar specific protein would be either III^Glc or IIB^Man. The phosphorylated sugar specific protein(s) then transfer the phosphoryl group to the sugar concomitant with translocation of the sugar across the membrane (15, 149).

The significance of this system is best understood from its essential functions. In addition to phosphorylation and concomitant translocation of its sugar substrates (PTS sugars), it has regulatory functions on non-PTS
sugar transport. PTS-mediated repression of non-PTS sugar utilization operons is at the metabolic level. The product of crr gene, which encodes the III^{Glc} enzyme (200), is the key to understanding the underlying mechanism: a crr mutation simultaneously relieves four non-PTS operons from PTS-mediated regulation. It has been shown that III^{Glc} inhibits adenylase cyclase activity and thereby reduces the level of cAMP and cAMP-CAP level in the cell. III^{Glc} or phospho-III^{Glc} was also found to interact with permeases of lactose, glycerol, melibiose and maltose (149). The chemotaxis of bacterial cells towards sugar is known to be promoted by the PTS (140). There is evidence that the PTS regulates the transcription of operons required for the uptake and catabolism of non-PTS sugars, such as the E. coli lac operon for lactose utilization.

1.3.3 Chitinoclastic Pathways in Vibrios

Two parallel pathways have been postulated in marine vibrios for catabolism of chitin (Fig. 1), each possibly containing as many as 6-10 enzymes and a number of chemotactic proteins (273). In the common part of both pathways, chitin-binding proteins adhere to the substrate (77, 274), and extracellular chitinase (123) and periplasmic chitodextrinase work together to produce N,N'-diacetylchitobiose (15). The glycosidase/PTS system cleaves N,N'-diacetyl chitobiose to GlcNAc in the periplasmic space using a membrane bound chitobiase (15, 220, 265) after which GlcNAc is transported and phosphorylated (15). The second, parallel permease/glycosidase system resembles the E. coli lac permease/beta-galactosidase system (275, 276), utilizing an N,N'-diacetylchitobiose permease for transport of this substrate to the cytoplasm. The transported (GlcNAc)$_2$ is cleaved by the cytoplasmic chitobiase reported here (278)
**Fig. 1 Chitin degradation pathways in vibrios.** 1. Production of GlcNAc-6-P from chitin. Chitin is represented [GlcNAc]_{12}, where the number indicates chain length in GlcNAc. Inner membrane proteins for the translocation of GlcNAc and [GlcNAc]_{2}, respectively, are labeled as PTS and P, for permease. 2. GlcNAc metabolism. Chn, chitinase; Chd, chitodextranase; Chb, chitobiase. Adapted from Bassler et al. (15).
and phosphorylated by a ATP-dependent N-acetyl-\(\text{D}\)-glucosamine kinase (6, 12, 15). This cytoplasmic system works independently of the PTS (15).

The cytoplasmic \(N,N'\)-diacetylchitobiase (EC 3.2.1.14) from \textit{V. para-haemolyticus} (ATCC #27969) has been characterized and the gene cloned into \textit{Escherichia coli} (278). A novel 85 kDa cytoplasmic glycosyl hydrolase of restricted specificity participates in the utilization of chitin-derived 2-deoxy-2-acetamido-\(\text{D}\)-glucose (GlcNAc) as one of two parallel pathways for metabolism of \(N,N'\)-diacetylchitobiose (15). The chitobiose is then cleaved in the cytoplasm to yield GlcNAc which is then phosphorylated by a constitutively expressed, ATP-dependent \(N\)-acetyl-\(\text{D}\)-glucosamine kinase (6, 12, 15).

The two pathways, glycosidase/PTS and permease/glycosidase, yield a common product from chitin degradation, GlcNAc–6–P, which is further deacetylated by \(N\)-acetylglucosamine deacetylase (112, 196), followed by deamination of the intermediate by glucosamine-6-P deaminase, resulting in Fru-6-P, acetate, and ammonia (49, 233). As a substrate of phosphofructokinase (PFK), Fru-6-P enters the glycolysis pathway to provide energy (ATP) and reducing power (NADH) for the organism. The phosphoryl group of GlcNAc-6-P from the two parallel pathways come from different sources, phosphoenolpyruvate or ATP, depending on the pathway where it was produced.

Chitin-degrading enzymes have also been reported in organisms as diverse as bacteria, fungi, plants, insects and vertebrates (25, 80, 279), and the pathways involved in the degradation of chitin are as diverse as the distribution. It is clear that, at least in \textit{Vibrio}, secretion of the first catalytic enzyme to the environment is necessary to initiate the process,
since the outer membrane of the cell is not permeable to this particulate polymer. Subsequent enzymes in the pathway are chitodextrinases and/or chitobiases. The substrates of these enzymes are chitinase products, which gain entry into either the periplasm or the cytoplasm, depending on the organism, for further degradation. Other enzymes than the chitinases are either secreted into the periplasm or stay in the cytoplasm. Exceptions to this scheme do exist: chitobiase of the fungus, *Trichoderma harzianum*, is reported to be extracellular (235).

1.4 PROTEIN SECRETION

1.4.1 The Signal Hypothesis

The initial clue that some proteins require an amino terminal (N-terminal) peptide sequence for secretion through the membrane was obtained during the *in vitro* synthesis of an immunoglobulin chain by cell extracts. It was found that synthesis in the absence (but not the presence) of membrane yields a product with a ca. 3,000 Da extra segment at the N-terminus (158). This highly hydrophobic segment, the signal peptide, is intrinsic to the nascent secretory polypeptide possessing the information necessary and sufficient for targeting. It is the membrane-bound signal peptidase in the membrane fraction, but not the membrane *per se*, that plays the critical role in this process. The *signal hypothesis* was proposed by Blobel and Dobberstein (24): (i) mRNAs encoding proteins for secretion through the ER membrane contain *signal codons*, unique sequences of codons immediately following the initiation codon; (ii) Translation of the mRNAs is initiated by ribosomes that are *not* attached to the membrane; (iii) A signal sequence receptor(s) on the membrane binds the nascent polypeptide, and facilitates the initial penetration into the endoplasmic
reticulum (ER) membrane; (iv) There exists in the membrane a proteinaceous channel, formed in association with the signal sequence, that provides an aqueous environment for the transfer of hydrophilic peptide. It is this short sequence that signals the attachment of nascent polypeptide chain, thus the ribosome, to the membrane to produce the appearance of 'rough ER.' It is clear that the hypothesis assumed a cotranslational translocation mechanism.

Wickner (257, 258, 260) proposed an alternative hypothesis for membrane protein assembly known as the membrane trigger hypothesis. It was suggested in the hypothesis that some membrane proteins synthesized on soluble ribosomes can assume two conformations, one more stable in aqueous solutions and the other triggered by contact with the hydrophobic environment of the membrane. The soluble precursor diffuses to the membrane and undergoes a conformational change as it self-inserts into the bilayer. The precursors with triggered conformation are thought to be stabilized by cleavage of the signal peptide at the N-terminus. This hypothesis did not assume membrane attachment of the ribosome synthesizing the precursor. Membrane protein insertion is a self-assembly process.

The membrane trigger hypothesis could explain the secretion and integration of integral membrane proteins. Membrane protein secretion is not a favorite subject of the signal hypothesis proponents since a majority of these proteins are made without a cleaved leader sequence (39, 60, 188, 204, 263, 272). The membrane trigger hypothesis, however, is not without problems: there have been too few reported examples that fit into this model, probably due to the lack of practical testing systems. One
extensively studied protein that has been shown to follow this model is the M13 coat protein. This lipid soluble, 50 residue peptide, is synthesized as a precursor with a signal peptide 23 residues long. It becomes integrated into the membrane during the infection stage of the M13 bacteriophage life cycle (108). Insertion into the membrane also requires sequence at the carboxy terminus (C-terminus) of the mature protein. *In vitro* studies (261) demonstrated post-translational processing of this protein. Signal peptidase, membrane electrochemical potential and phospholipid are the only components required to mediate binding, processing and insertion.

The loop model proposed by Inouye (104) was based on the functional and structural relationship of the signal peptides (105). The positively charged N-terminal region and the extremely hydrophobic core region of these short leader peptides were believed to form a loop. The basic region interacts with the acidic surface of cytoplasmic membrane and the hydrophobic core inserts into the lipid bilayer. The model system used was the signal peptide of *E. coli* major outer membrane lipoprotein.

1.4.2 Cell Compartmentation and Protein Targeting

The chromosome of a bacterium occupies a certain area in the cell but is not membrane bounded. However, the bacterial cell is also compartmented and protein targeting is essential for the organism to survive. The periplasm of a Gram negative bacterium, sometimes called the periplasmic space, is surrounded by the inner (cytoplasmic) and outer (lipopolysaccharide and protein) membranes. In *Salmonella typhimurium* and *E. coli*, the volume of this compartment comprises 20 to 40% of the total cell volume and is isoosmotic to the cytoplasm (222). This
compartments contain at least 4% of total cell proteins (175) with up to 50 distinct polypeptide species (50). The outer membrane is a coarse molecular sieve that allows simple diffusion of hydrophilic and hydrophobic molecules up to a molecular weight of about 800 Da for *E. coli* and even higher for members of *Pseudomonas* and *Neisseria* (7) whereas the cytoplasmic membrane excludes almost all hydrophilic substances. The cell wall of a Gram negative bacterium includes both the outer membrane and a thin layer of murein or mucoprotein (peptidoglycan) as osmotic pressure barrier between the two membranes. Cell envelope is another term that includes everything outward from the cytoplasmic membrane.

Proteins synthesized in the cytoplasm of the cell have different destinations. In addition to integral membrane proteins that have hydrophobic regions to span the lipid bilayer, there are proteins that are targeted and functional in compartments other than the cytoplasm. It has been estimated that 25% of proteins synthesized in the bacterial cytoplasm are targeted either to the inner membrane, the periplasm, or to the outer membrane. The periplasm contains proteases, phosphatases and sugar binding proteins for active transport. Some toxins are secreted from the cell into the medium (8). Some enzymes in polymer utilization systems are secreted out to the external milieu. These include amylases from the Gram positive *Bacillus subtilis* (270), *B. amyloliquefaciens* (180, 225), *B. stearothermophilus* (223), and the Gram negative *Aeromonas hydrophila* (42). Chitinases from *V. harveyi* (219, 220) and *V. parahaemolyticus* (123), and pullulanases from *Klebsiella oxytoca* (51, 191) are known to be secreted across the double-membranes of these Gram-negative bacteria to
digest chitin and branched-starch, respectively. Other examples include the plant pathogens *Erwinia chrysanthemi* (134) and *Erwinia carotovora* (194) that secrete cell-wall degrading enzymes such as pectate lyase, polygalacturonase and cellulase.

1.4.3 **Bacterial Signal Sequences**

Secretory proteins of the Gram negative bacteria are those targeted to the periplasm, the external milieu and the inner and outer membranes. Not all secreted proteins require a signal sequence. For those that need one, it does not have to be at the N-terminus. For those that do have an N-terminal signal sequence, it is the underlying basis of protein targeting. Several lines of evidence from genetics and recombinant DNA demonstrated that it is this short piece of amino acid sequence at the N-terminus of the precursor that dictates the transport of the mature protein to the location where it is functional.

The bacterial signal peptides, like eukaryotic ones, have three distinct regions, a positively charged N-terminus (n-region), a hydrophobic core (h-region), and a hydrophilic C-terminus (c-region) (245). The n-region varies in length and has an overall positive charge. The contribution of this region to the secretion process is not well understood although it has been speculated that it may be responsible for binding to the negatively charged phospholipid head groups of the cytoplasmic surface of the membrane. The h-region is rich in hydrophobic residues (Phe, Leu, Val, Ile, Met, Trp). Often found in this region are sequences with potentials for forming alpha-helices, but the presence of some helix-disrupting amino acids is not unusual (242, 245). What is critical for translocation is its length (minimum of 7-8 residues) and overall hydrophobicity as a
continuous region. A consensus sequence to known signal sequences does not seem to exist even within the same organism and the analyses of structure and function are largely based on the classes of amino acids present in each of the three regions that contribute to the secondary structures.

The amino acid sequence of the c-region, generally around six residues in length, demonstrates a higher degree of stringency at the cleavage site of signal peptidase. Based on the signal sequences available, a '-1, -3' rule was proposed (243, 244) which states that the sequence at the cleavage site conforms to the formula A-X-B, where B contributes the carboxy group to the peptide bond that is cleaved and is Ala, Gly or Ser. The ‘A’ in the above formula at position -3 is any of B or Leu, Val or Ile. The -2 position is somewhat less restrictive where larger residues and charged residues are generally found. The signal peptidase prefers certain amino acids for their steric configuration that probably dictates the position for proteolytic cleavage. It is worth noting that these preferential residues at their perspective positions are for peptidase cleavage only. Mutations of these residues to bulky or charged residues prevent cleavage but not translocation. Extensive site-directed mutagenesis and genetics studies in these regions of signal peptides from OmpA, MPB, LamB, Staphlokinase, and beta-Lactamase agreed with the structural analysis above (71). A net positive charge at the N-terminus of the signal peptide is required for optimal processing and translocation. It is the overall hydrophobicity rather than any specific amino acid sequence that is essential for export function. It is also noteworthy from these studies that, for cleavage by signal peptidase, the N-terminal residue at +1 can be virtually any residue as
long as it does not disturb secondary structure, suggesting that this residue is not recognized by the enzyme. The specificity of the enzyme seems to rely on the signal peptide, which is not part of the functional exported protein. Restrictions on the N-terminal residue may limit the number of functional proteins translocated.

Attempts to elucidate the mechanistic contributions of the n-region and c-region during translocation using 2D-NMR (two dimentional nuclear magnetic resonance), CD (Circular dichroism) and infra-red measurements suggested that the peptide may be in beta-sheet structure when interacting electrostatically on the aqueous surface of the membrane, and in alpha-helix when inserted (28). Care should be taken for interpretations of such data since the principal assumption is that the translocated proteins interact with lipid bilayer. That secretory proteins get across the membrane does not signify interaction between the polypeptides and the lipid during translocation. Integral membrane proteins interact with the membrane and many of them are secretion products. However, there has been no conclusive evidence that integration is a direct consequence of translocation. In contrast, it has been demonstrated that the *E. coli* MalF protein, which spans the cytoplasmic membrane eight times, can be properly inserted into the membrane in the absence of its N-terminal export signal that also serves as the first of the eight transmembrane segments (61). It would be difficult to envision that these proteins simply stalk on the membrane in the middle of translocation and 'become' membrane inserted, without jamming the secretion pathway, especially for those that span the membrane many times, such as SecY, SecE (41, 56) as well as MalF (61).
Regardless of the underlying mechanism, the importance of signal sequences in protein secretion is obvious based on the fact that many signal peptides secrete foreign proteins, making this complex process visible and practical for cloning, expression and commercial production of proteins.

It has been demonstrated that, in E. coli, the signal sequences are interchangeable between the outer membrane porin PhoE and a similar outer membrane protein OmpF, and both functional proteins are localized in the outer membrane (231, 232). Similar to this is the substitution of periplasmic TEM-beta-lactamase signal sequence for that of the PhoA, also periplasmic, and the mature beta-lactamase is secreted into the periplasm (98, 142). The signal sequence from the alpha-amylase of the Gram positive Bacillus amyloliquefaciens was found capable of secreting periplasmic TEM-beta-lactamase of Gram negative bacteria into the growth medium of Bacillus subtilis (181, 236). This location can be considered as equivalent to the beta-lactamase destination in E. coli since Bacillus does not have an outer membrane and thus lacks a periplasm.

Prokaryotic signal sequences can be used to secrete proteins from eukaryotes. Examples include the expression and secretion of human growth hormone into the E. coli periplasm using signal sequences from E. coli alkaline phosphatase gene (177) and E. coli heat-labile enterotoxin gene (74, 161). This signal sequence was also used in the secretion of ribonuclease T1 from Aspergillus oryzae (69). Prokaryotic secretion machinery can also recognize, process and translocate eukaryotic precursors, as demonstrated in the case of rat preproinsulin gene that was
expressed, the precursor processed, and the proinsulin localized in *E. coli* periplasm (226, 227).

Other examples in light of signal sequence compatibility could be seen in numerous eukaryotic cellular systems. These include, but are not limited to: (i) *Saccharomyces cerevisiae* that secretes human epidermal growth factor using a synthetic leader sequence (45), human lysozyme using chicken lysozyme signal peptide (234), leech (*Hirudo medicinalis*) hirudin using the yeast invertase signal sequence (97), human lysozyme using synthetic signal sequences (267, 268), and the secretion of human lipocortin-1 by *S. diastaticus* using STA1 signal sequence from this yeast species (170); (ii) Insect cells (*Spodoptera frugiperda*, host cell line of baculovirus expression system) that secrete papain (a plant cystein protease, EC 3.4.22.2) using honeybee mellitin signal peptide (228), and human HIV-1 gp120 using signal peptide of honeybee mellitin or murine interleukin 3 (133). (iii) Mammalian cells (Chinese hamster ovary cells, CHO) that secrete endoglucanase of *Clostridium thermocellum* using signal sequences either from the pre-endoglucanase or from the human growth hormone gene (86), human granulocyte-colony stimulating factor (hG-CSF) using the conserved portion between hG-CSF and C-GSF (of the mouse) signal sequences (COS cells) (117), and neutral endopeptidase using pro-opiomelanocortin signal peptide (COS cells) (131); (iv) Transgenic tobacco cells that secrete the bacterial ChiA protein from *Serratia marcescens* using either the original signal sequence of pre-ChiA or that from the PR1b protein (138).

The above are some examples with respect to protein secretion in different systems. The enhanced secretion efficiency of papain and HIV-1
when their respective signal sequences were substituted for the insect-derived signal peptide immediately implies that the *Spodoptera frugiperda* translocation machinery is more compatible with signal sequences of insect origin. However, this has been shown to be not true, as the secretion of human tissue plasminogen activator (t-PA) was not increased by any of the signal sequences tested, most of which are insect-derived, including honeybee prepromellitin (111) that greatly enhanced the secretion of papain (228).

### 1.4.4 Secretion of beta-Galactosidase

The establishment of the signal hypothesis (22, 23, 24) and subsequent work on secretory systems led to partial understandings of the expression and secretion of several prokaryotic and eukaryotic proteins across the *E. coli* cytoplasmic membrane (18). One interesting protein that played a pivotal role in the early elucidation of the sec-dependent secretion pathway is beta-galactosidase. The monomer of this homotetrameric enzyme is encoded by the tricistronic mRNA of the *E. coli lac* operon (275, 276). Attractive features of this well-known enzyme as a secretion reporter include its ease of bioassay and its cytoplasmic location. Unfortunately, efforts in secreting this cytoplasmic protein have not been successful. The signal sequences used include those from the *malE* (periplasmic) (13, 14), *phoA* (periplasmic) (156), *phoE* (outer membrane) (230), *ompF* (outer membrane) (217), and *lamB* (outer membrane) (63). These fusion proteins could not be found in their expected locations. Instead, they were incorporated into the cytoplasmic membrane where they appear to block the secretory potential, causing accumulation of pre-cursors of most other would-be-exported proteins in the cell. The problem
had been noted earlier by von Heijne (241) who proposed that incompetence of beta-galactosidase in passing through the membrane is due to the presence of amino acid sequences within it. Genetic studies further specified that the nonexportability of the enzyme is due to folding of portion of the polypeptide in the cytoplasm. This model was further supported by the finding that overexpression of the GroEL chaperone protein and DnaK facilitate transfer of a LamB-beta-galactosidase hybrid protein (186, 205, 277). Another interesting feature of these systems is that these chimeric genes are lethal to the cell when over expressed. Careful investigation of malE-lacZ and lamB-lacZ demonstrated that the inhibition effect is proportional to the amount of chimeric proteins produced (107). Both of these chimeras can be induced by maltose since LamB and MalE proteins are involved in maltose metabolism. As the concentration of the fusion protein in the cytoplasmic membrane increases by maltose induction, there is an increase inhibition of export of cell envelope proteins. The precursors of these proteins accumulate in the cell but the mature proteins are barely found in their proper locations. Accumulation of the hybrid itself in the membrane severely affects the cells. A large fraction of the cells eventually lysed. The hybrids thus render the cell sensitive to maltose. A similar phenomenon was also observed in other fusions. The phoA-lacZ on a multicopy plasmid is lethal to a partially constitutive phoR E. coli strain, whereas the same gene on a single copy plasmid is not (156), another line of evidence implying the sensitivity of the cell to the amount of hybrid protein produced.
The inhibition effects may be due to the direct disruption of membrane organization by the hybrid protein, or indirectly through the blockage of secretion channels for naturally secreted proteins that have essential functions to the cell. The maltose induction assay system used by these investigators is rather ingenious since the beta-galactosidase activity of the hybrid protein is not measurable even after full induction with maltose (lamB-lacZ) (87) or considerably lower than expected without induction but is close to normal level upon induction (malB-lacZ) (13, 14). The inhibition effect itself is thus a better quantitative assessment to the hybrid than the enzyme activity.

1.4.5 The Bacterial Secretion Machinery

The pleiotropic detrimental effects of the LacZ fused with the N-terminal portions of secretory proteins led to the elucidation of bacterial protein secretion machinery. Mutations in *E. coli* carrying these chimeras mapped in loci that are components of the secretory machinery. These include secA(prlD), secB, secE (prlG), secY (prlA), secD, secF, and the two genes for leader peptidases, lepB (coding for leader peptidase) and lspA (coding for lipoprotein signal peptidase). Translocation and processing of precursors that require at least some of these proteins (*i.e.*, SecA and leader peptidase) are said to be sec-dependent, those that do not require any of these proteins are sec-independent. In both cases, translocation and processing which occur before chain elongation is complete are said to be cotranslational, and those that occur after elongation is complete are said to be post-translational. In *E. coli*, a known example of sec-independent protein transport is the 107 kDa haemolysin (HylA) that is secreted independently of an N-terminal signal sequence and the secA gene product
There exist extensive reviews on the genetics and enzymology of these genes and gene products (18, 205, 259) and this section only briefly the current understanding of what is known about the system in *E. coli*. On the basis of functional and structural interactions, these sec proteins are now believed to form a multisubunit enzyme, *translocase*, on the cytoplasmic membrane (256). The integral membrane domain of the enzyme is a complex of secY and secE, which can be co-purified under nondenaturing conditions and cross-immunoprecipitated (1, 37, 38). Both SecY (49 kDa) and SecE (13.6 kDa) proteins are highly hydrophobic with 10 and 3 membrane spanning regions, respectively (41, 56). The cytoplasmic SecA is associated with the membrane through its affinity for SecY/E complex (90). The ATPase activity of SecA is believed to catalyze cycles of precursor binding and release (206) and successive cycles of insertion into the membrane translocator (178). Both SecA and its ATPase activity are essential for sec-dependent translocation of proteins. SecB is a 16 kDa soluble protein that binds to both the leader and the mature portion of nascent chains of secretory proteins such as MBP, OmpA and PhoE (128, 193). It has been demonstrated *in vitro* that SecB binds only to denatured molecules and the binding prevents refolding (47, 128, 136). The homo-tetrameric SecB is required for translocation but not translation of precursors. *E. coli* supernatant depleted of SecB through an anti-SecB affinity chromatography column retains full translation activity but is unable to support translocation of preMBP into inverted vesicles (251). The products of *secD* and *secF* are homologous in their membrane spanning regions. These two proteins are required for the proton electrochemical gradient stimulation of preprotein translocation (5) and
are not considered as the subunits of the multisubunit translocase enzyme.

Proteins need to be unfolded prior to translocation and fully folded precursors are not secretion competent (192). The co-translational translocation process requires nothing but the translocase enzyme. This is consistent with the vectorial transfer as proposed in the signal hypothesis (24). In post-translational translocation, molecular chaperones play roles in maintaining precursor polypeptide ready for translocation. In addition to chaperones that are components of the translocase (SecB and SecA), GroEL and GroES are also known to have functions that maintain secretory precursors in translocation-competent conformations (144). These chaperonins, belonging to a subclass of chaperones, are the major 'heat shock' proteins but are essential for growth of E. coli at all temperatures. Mutations in both groEL and groES are defective in the export of pre-beta-lactamase but not pro-OmpA and pre-MBP.

The folding property of the mature portion or the folding status of the precursor affects the secretion of maltose binding protein (MPB), and rapid folding of the precursor in the cytoplasm of secB* cells results in a conformation incompatible with translocation across the membrane (192). Amino acid changes in the mature portion that suppress defects in the signal sequence decrease the rate of MPB folding (135). Mutations in secB slow the rate of MPB transport and also increase the amount of folded, translocation-incompetent preMPB in the cytoplasm. On the other hand, the defects in MPB export in secB* cells were suppressed by amino acid changes that slowed the folding rate of mature MPB (47), or by mutations in the signal sequence that increase the rate of export (48). Based on
these studies, Schatz and Beckwith (205) suggested that secB binds to preMBP during or soon after its synthesis and stabilizes the conformation to enable its association with later components of the secretion machinery. The signal sequence per se also plays a role, independently of chaperons, in slowing down the rate of precursors folding (173). The effect of precursor folding on protein secretion has been applied to practice. In one example, the production of active secreted protein is increased by 16 fold through the control of protein folding (224).

The mature protein found in the targeting location is the product of proteolytic cleavage by peptidases. There are two peptidases associated with protein secretion, leader peptidase and lipoprotein signal peptidase, each showing different substrate specificites. The lipoprotein signal peptidase is an endopeptidase that recognizes glycyl glyceride-cystein or alanyl glyceride-cystein as cleavage site [L-A(S)-G(A)-C-]. It cleaves lipid modified pro-lipoproteins into lipoproteins, which are further modified by N-acylation to form mature lipoproteins that attach to the outer membrane of Gram negative bacterial cells (191). The protein, also known as signal peptidase II, is the gene product of lspA (106). The lepB gene product, leader peptidase (also known as signal peptidase or leader peptidase I), cleaves the signal sequences from precursors other than that of pre-lipoproteins (52). Leader peptidase is a cytoplasmic membrane protein with its active site facing the periplasm. This 323 residue polypeptide (36 kDa) has three hydrophobic segments near the N-terminus with the first two spanning the cytoplasmic membrane and the third residing in the periplasmic space, as demonstrated by protease mapping studies (130, 160, 263). Studies on a B. subtilis signal sequence fused to the mature
portion of the TEM-beta-lactamase (238) showed that a 25-fold overproduction of this protease results in enhanced processing rate and improved release of mature beta-lactamase into the periplasm of E. coli. Since these effects were not detected on the wild-type pre-beta-lactamase and other secretory proteins, the author argued that the availability of this signal peptidase may be a limiting factor in protein export in E. coli, in particular with those showing low processing efficiencies.

For proteins that export from the cytoplasm of Gram negative bacterial cell to the external milieu, an important but poorly addressed question is how the translocated protein is inserted into or exported across the outer membrane. Three mechanisms have been proposed for this process (9): through intermediate vesicles, direct translocation through the periplasm, and direct crossing from the inner to the outer membrane via Bayers Patches where the two membranes are in close vicinity. None of these were able to explain what dictates the final destination of these proteins.

Recent studies on the genetics of extracellular protein export have shed light on this problem. It has been shown that this step alone, from the periplasm to the external milieu of the organism, involves more than 13 genes, collectively called the out genes (in Erwinia) (134, 194) or pul genes (in Klebsiella) (51, 191). Molecular cloning and characterization of these genes from several Gram negative bacteria (51, 134, 191, 194) showed that they are highly conserved and are specific in function for the translocation of proteins targeting to the outside of the cell. Mutations in these genes do not affect periplasmic protease translocation (194). The extracellular starch-degrading enzyme pullulanase from Klebsiella oxytoca is
one of the best-studied examples, in which a *pulS* gene and a 13-gene operon (*pulC* through *pulO*) were shown to be necessary and sufficient for the secretion of the pullulanase enzyme across the outer membrane (51, 191). More recently, the *out* genes have been cloned and sequenced in two independent studies from the plant pathogens *Erwinia chrysanthemi* (134) and *Erwinia carotovora* (194) that secrete plant cell-wall degrading enzymes including pectate lyase, polygalacturonase and cellulase. Homologs of the *pul* or *out* genes have also been cloned from other Gram negative bacteria including *Pseudomonas aeruginosa* (10, 11, 55, 65), *Xanthomonas campestris* (57, 102), and *Aeromonas hydrophila* (113). This suggests that these systems are widely distributed in Gram negative bacteria. The gene arrangement and amino acid content of their ORFs are similar (194). Further investigation is needed to understand how the system works. Little is known, for example, about the signals in the extracellular proteins that target them to the secretion pathway, and how they interact with the components of the pathway in the periplasm. Although these *out* or *pul* genes are highly conserved in their gene arrangement, nucleotide and amino acid sequences and their biological functions, evidence from these works seemed to indicate that these gene clusters are species specific (134). Known exceptions are the *pul* and *out* clusters of *K. oxytoca* and *E. chrysanthemi*, respectively, that permit *E. coli* to secrete homologous proteins by a Sec-dependent pathway (93, 134, 190). In addition, the *P. aeruginosa xcp* gene cluster, another homolog of the *pul* gene system, complemented the protein secretion capacity of a *X. campestris* secretion mutant while the gene cluster from the latter
organism restored the secretion capacity of \textit{P. aeruginosa xcp} mutant only to a low degree (55).

Together, these findings are significant in that they are the starting point for understanding and for further probing the mechanism that dictates the final destination of proteins, since for almost two decades after the establishment of the signal hypothesis, little was known about how proteins transport beyond the cytoplasmic membrane in Gram negative bacteria.

1.4.6 Protein Secretion by Eukaryotes

Some examples are briefly mentioned in this section to show that protein targeting is a general phenomenon of life and that the bacterial system is only a part of it. Proteins encoded by nuclear genes in eukaryotic cells are synthesized in the cytosol and are targeted to specific locations where they are functional. Depending on the types of cells, the destinations include the nuclei, mitochondria, chloroplasts, lysosomes, as well as the lumen of endoplasmic reticulum (ER). The signals contained in the precursors play important roles in these processes.

Targeting of a nascent secretory protein to the ER lumen is initiated by the high affinity binding of newly emerged signal sequence to the signal recognition particle (SRP) (198). SRP is a cytosolic ribonucleoprotein complex consisting of six polypeptide chains and a 7S RNA (248). Its high affinity for signal sequence, ribosome, and SRP receptor makes it a critical factor in the association between the precursor and the ER. The SRP receptor, also known as docking protein (DP) (78, 94), was originally isolated from the canine pancreatic rough microsome, and was found to be required for translocation across the ER membranes in cell free assays (153, 154). The binding of SRP to the signal sequence halts chain
elongation until the complex interacts with the DP on the ER surface (155, 246, 247, 249). Precursor-synthesizing ribosomes bound to the ER gives rise to the 'rough ER' appearance. Two further events are mediated by the binding of SRP to DP: stable interaction between the ribosome and the ER membrane, and release of nascent polypeptide chain from the SRP. This process requires that DP be in the GTP-bound form. The bound GTP is hydrolyzed during the dissociation of SRP and DP. It is postulated that GTP hydrolysis provides a molecular proofreading mechanism to assure correct association of the nascent polypeptide with the ER membrane (78, 94). As with the bacterial secretion system, insertion of polypeptide into the ER membrane remains unclear, although several mechanisms have been postulated (260). One common model is that there exists a translocation complex on the ER membrane comprised of a 43 kDa signal binding unit and several other proteins. The roles of the complex include binding to the signal sequences and forming an aqueous transmembrane tunnel (8, 214). The signal sequence is then cleaved by the signal peptidase and the mature protein released into the ER lumen for further processing and sorting. The resistance of microsome peptidase activity to alkali treatment and trypsin digestion indicated that signal peptidase is an integral membrane complex with its active site positioned on the luminal side of the ER membrane (137, 250). There are extensive reviews on the import of proteins into organelles (mitochondria (91, 185), chloroplasts (208), and the nuclei (213)). Unlike that of ER, translocation of proteins into these organelles is posttranslational. Although similarities do exist among these and the ER/bacterial
systems, protein traffic into these organelles is complicated by further compartmentation of the organelles. Proteins that are nuclear-encoded can reside in one of several suborganellar compartments: four for a mitochondrion (Outer and inner membranes, intermembrane space, and matrix), and six for a chloroplast (Outer and inner membranes, intermembrane space, stroma, thylakoid membrane, and thylakoid lumen).

The nuclear envelop is constructed of two lipid bilayers with a perinuclear space in between. Accordingly, proteins targeted to their destinations require specific signals, in addition to unique translocation complex on each compartment.

Presequences of the mitochondria and chloroplasts (also called transit sequences) seem to accommodate this specific targeting. Hurt et al. (103) and Horwich et al. (101) were the first to demonstrate that the mouse cytosolic enzyme dihydrofolate reductase (DHFR) can be directed into the mitochondrial matrix by the 25 N-terminal presequence of cytochrome oxidase IV, an inner membrane protein. The first 12 residues of the presequence were found to be necessary and sufficient for the transport and subsequent processing in the mitochondria. On the other hand, even when 28 more N-terminal residues of the mature cytochrome oxidase IV were included, the fusion protein was still not found in the inner membrane, suggesting that information for intramitochondrial traffic must reside in the remaining amino acids of the mature protein. Another example is the presence of isozymes encoded by the same nuclear gene. Two isozymes of the yeast histidine tRNA synthetase are the products of a gene with dual transcription start sites (171). The longer mRNA is translated into a mitochondrial precursor, whereas the shorter one
produces a cytosolic isozyme. Mutation at the first start site leads to the lack of mitochondrial isozyme, resulting in respiratory deficiency, with no effect on the cytosolic isozyme.

The small subunit of Rubisco (Ribulose-1,5-bisphosphate carboxylase) is a nuclear gene product and the 40-residue transit sequence is cleaved in the stroma (207). The bacterial protein neomycin phosphotransferase was imported into chloroplast when fused to this transit sequence, showing that the transit sequence contains sufficient information to target mature proteins into the chloroplast (237). For proteins targeting to the nucleus, basic residue stretches called nuclear localization sequences (NLS), serve as targeting signals. Mutations in the transit sequence of nuclear proteins (e.g., SV40 T-antigen) cause cytoplasmic localization and exclusion of the protein from the nucleus (124, 125). Fusions between SV40 T-antigen and cytoplasmic proteins such as pyruvate kinase and beta-galactosidase localize the chimeras to the nuclei (114). Non-nuclear proteins coupled chemically to synthetic nuclear targeting peptides and microinjected into cells are also shown to be imported into the nucleus (126). Recently, a 60,000 Da protein called 'importin' that is essential for the first step of nuclear protein import has been isolated from Xenopus eggs (81). The current understanding is that importin binds NLS-containing protein and directs it to the nuclear pore proteins, followed by an energy dependent translocation process facilitated by the Ran/TC4 protein (189).

Like bacterial signal sequences, very little similarities have been found among presequences for ER lumen and organelles. Particularly noteworthy are nuclear localization sequences. Although most of them are short
(5-7 residues) and are rich in basic amino acids, a consensus sequence is also lacking.

1.5 OVEREXPRESSION OF CLONED GENES

1.5.1 Strong Promoters

The tac and trc promoters employed in the pKK expression vectors (e.g., pKK233-2/trc and pKK223-3/tac) are strong promoters derived from the E. coli lac (76, 109) and trp (187) operons. The E. coli bacteriophage T5 early promoters are naturally strong. Although this project does not involve these promoters, the rrnBT1T2 transcriptional terminators were first introduced down stream of the T5 early gene promoters (73) and subsequently of the tac and trc promoters in the pKK expression vectors. A brief introduction to T5 promoters will thus follow.

The tac and the trc promoters are hybrids composed of -10 region of the lacUV5 promoter and the -35 region of the trp promoter (see below). These promoters have been successfully used to drive the overproduction of several prokaryotic and eukaryotic genes including the dehydroquinate synthase from E. coli (68), the lipocortin I from rat (211), and the alpha-tubulin from human (266). The dehydroquinate synthase yield from E. coli K12 (MM294), for example, is 1000 fold higher when driven by the tac promoter than in the wild type E. coli K12, and 50 times more enzyme was produced than that from a clone in which the gene is driven by its own promoter (68).

During the production of human growth hormone (HGH) in E. coli, De Boer et al. (53, 54) used both trp, lacUV5 and a hybrid tandem trp-lac promoter they constructed (tac promoter). It was demonstrated that the promoter is 5 to 10 times as efficient as the lacUV5 promoter and is
The consensus sequences:

-10 Region: 5'-TAtAaT-3'
-35 Region: 5'-TTGACa-3'

**Fig. 2 Construction of the tac promoter.** 1. The lac, trp and tac promoter sequences (53, 54). The hgh gene was served as a reporter in the promoter strength assay and was labeled as the gene product, HGH. The distances between the -35 and -10 of the three promoters are 18 bp, 17 bp, 16 bp for lac, trp and tac, respectively. 2. Consensus sequences of the two hexamers of prokaryotic promoters (92, 132, 197).
controlled by the lac repressor. The lacUV5 is a UV-induced up mutant of the wild-type lac promoter. In the absence of cAMP-CAP complex, the wild type lac promoter is only 2% of full level (17). The GT in the -10 region of lac promoter is changed to an AA sequence, resulting in lacUV5 that restores 50% of full promotion (75). The trp -35 and lacUV5 -10 regions are identical to those of the consensus sequences, respectively (92, 132, 197). By bringing these two sequences together, these investigators were able to construct a hybrid promoter (Fig. 2) that is stronger than either of the two promoters, as demonstrated by the expression levels of HGH under the control of these promoters.

Expression vectors bearing the tac promoter are now commercially available and are widely used in revised forms that facilitate cloning simplicity and the capability of expressing genes from both prokaryotes and eukaryotes (2, 3, 4). The expression vector pKK223-2, constructed by Amamm and Brousius (2), carries a slightly modified form of tac. The promoter, known as trc, has the consensus distance, or consensus spacer length of 17 bp between the -10 and -35 regions instead of 16 bp in the original tac. It had been shown previously in vitro that trc is the strongest promoter characterized so far and confirmed that the consensus sequence, in terms of the two hexamers and spacer length in between, is the "best" (163). The expression levels of genes directed by tac and trc, however, are similar in vivo (2, 3, 33). The major improvement is the introduction of lacZ ribosome binding site (RBS) followed by an ATG translation initiation codon at an appropriate distance from the RBS. The ATG codon is exposed for insertion of incoming structural genes upon digestion with Nco I. The tac bearing plasmid pKK223-3 (34) contains a variety
of restriction sites for the insertion of genes to be expressed at high level between the regulatable tac and the rrnB transcriptional terminators (see below).

In addition to those mentioned above, it is also the driving force of cloned genes on vectors such as GST gene fusion vectors (pGEX series) (215), pTrc99A (pTrc series) (4) and pKK232-8, a promoter selection vector (31). Both pKK223-3 and pKK233-2 plasmids were used in this study. The trc promoter carried on pKK233-2 is the one that direct the expression of secreted chitobiase. The secretion systems developed during this study (pVerSec and pVerSec-SE) also use trc for high level expression and secretion of chitobiase and proteins from other sources.

1.5.5 The T5 Early Gene Promoters

The E. coli bacteriophage T5 and related phages have been extensively reviewed (146, 147). This coliphage is unique compared with other bacteriophages. It injects 7.9% of its 121.3 kbp chromosomal DNA into host E. coli upon attachment to the cell surface, with the remaining 92% DNA injected 3-4 minutes later. Promoters of its early genes are naturally strong.

Using filter binding, von Gabain and Bujard (239) measured both the relative rate of formation and stability of RNA polymerase-promoter complexes and demonstrated that the transcriptional activities of promoters, in vitro and in vivo, correlate with the rate of polymerase-promoter complex formation, but not with the stability of the complex. It has also been shown (240) that promoters of T5 early genes are very strong, stronger than those of pre-early and late promoters. Furthermore, T5 early gene promoters are stronger than other promoters tested, including those of
bacteriophage \( \lambda \), T7, fd, and of plasmid pML21 and pSC101. By setting up a promoter probing system (a plasmid used to select promoters by function), Gentz and Bujard \( ^{72} \) were able to select 11 highly efficient promoters from T5.

It should be pointed out that the structural aspects of promoters have been, and may only be, used to predict, but not to foresee, their functions. The ultimate biological activities of this important control element of a gene may not correlate well enough with these parameters. A synthetic prokaryotic promoter with consensus -10, -35 and optimal spacer length, for example, is only a moderate promoter \( ^{216} \) instead of a strong one as expected, although biologically functional.

1.5.6 Strong Transcription Termination Signals

Transcription of a prokaryotic gene by RNA polymerase begins by binding to the promoter. Once initiated, the newly synthesized transcript is elongated until the enzyme reaches a transcription termination signal, or a terminator. In general, the DNA from the promoter through the terminator defines a transcription unit. Strictly speaking, the transcription unit is what is meant by the word 'gene.' The open reading frame (ORF) is often called a structural gene to signify that it can not be naturally active as a gene because it lacks control elements such as the promoter, RBS and, at least for some genes, a proper terminator. Some genes have other elements such as enhancers and/or operators. In these cases, the transcription unit itself is active, but the biological and biochemical behaviors of these genes are not necessarily retained when mobilized through gene cloning.
The terminator is a base-paired hairpin on the newly synthesized RNA molecule. It is formed by GC rich self-complimentary sequences followed by a string of U residues. Nascent RNA spontaneously dissociates from RNA polymerase at the terminator. This prevents read-through of distal region on the DNA template by the RNA polymerase. Some terminators require an ancillary proteinaceous factor called \textit{rho} and are defined as \textit{rho}-dependent terminators. Those without the need of \textit{rho} for termination are \textit{rho}-independent terminators. Failure to terminate properly is obviously a disadvantage to the organism since it would be a waste of energy. It may have other negative effects if critical elements (origin of replication, promoters, etc.) that happened to be located nearby.

The strong T5 early gene promoters, as justified by \textit{in vitro} assay (239, 240), are attractive. Early attempts in cloning these promoters were unsuccessful when the promoter is carried on short inserts. These promoters were successfully cloned, however, when longer fragments that carry both the promoter and a strong transcription termination signal. Based on these and other observations, Gentz \textit{et al} (73) constructed plasmids bearing two reporter genes (\textit{lacZ'} and \textit{tet}) with a strong terminator (the \textit{fd} terminator) in between. Several T5 strong promoters were cloned and screened based on read-through of the terminator region and thus the expression of distal reporter gene \textit{tet}.

1.5.7 A Strong Tandem Terminator

A frequently seen termination signal used is a tandem terminator of the ribosomal RNA (rRNA) operon B of \textit{E. coli}, or \textit{rrnBT1T2}, occasionally labeled as T1T2 on plasmids. Expression vectors bearing this terminator include the above mentioned pKK series, pGEX series and the pTrc
series. On all of these systems, the promoter and terminator are arranged such that the inserted structural gene directed under the promoter is proximal to the terminator.

The *rrnB* operon is one of the seven separate *E. coli* rRNA operons, named *rrnA* through *rrnG*. These operons are not linked on the *E. coli* chromosome, although the encoded 16S, 22S and 5S rRNAs are homologous. In addition to rRNAs, these operons also encode transfer RNA (tRNAs) located between these rRNAs. All of these rRNAs and tRNAs are intracistronic, driven by the same tandem promoters P1 and P2 (separated by 80 bps) and terminated at a dual distal tandem terminator.

Processing of the 30S primary transcript by RNase III yields the rRNAs and tRNAs (132).

The *rrnB* operon was characterized and completely sequenced in the early 1980's (32, 35). The organization of the operon is similar to the other 6 operons, with gene products located on the operon in the following order: 16S rRNA, tRNA\(^{\text{Glu}}\), 23S rRNA, and 5S rRNA followed by *rrnBT1T2*. This dual terminator is located distal to the rRNA genes. T1 is adjacent to the 3' terminus of the 5S rRNA gene, and T2 is 175 nt further downstream. Both T1 and T2 are strongly homologous to known rho-independent terminators. Features of the T1T2 include a potential hairpin structure that ends with stretch of T residues (32). Understanding the operon organization is helpful in practice when some of the elements are involved in a system to be designed. One clear example is the 5S rRNA gene that always precedes T1T2 in expression systems. Since the 5S rRNA region has no function in terms of transcription termination, it can be deleted, partial or completely, without detrimental effect on the
expression of genes upstream, as long as the T1T2 themselves are retained. This idea has been applied to the secretion system construction as part of this project (chapter 4) and proven to be applicable.

1.6 PCR AND RECOMBINANT DNA

The concept of the polymerase chain reaction (PCR) was first formulated and reduced to practice in the mid-1980s by Kary Mullis and co-workers (164, 165) using the Klenow fragment of E. coli DNA polymerase I and was applied to the amplification of human genomic DNA by a group in the Human Genetics Department at Cetus Laboratory (62, 201, 202). This ingenious invention was soon evolved into a powerful, yet simple, technique along with the discovery of the thermostable Taq DNA polymerase (70) and the introduction of thermocycling equipment (179). The operation of PCR since then became simple enough for end-users equipped with minimal background of molecular biology. The power of this revolutionary breakthrough to the recombinant DNA technology could be envisioned from several aspects of the technique. In addition to its simplicity, the reaction is extremely sensitive. One single copy of a target sequence buried in a mass of chromosomal DNA could be amplified. Another aspect is that the sequences to be amplified do not need to be known. Sequences flanking the target sequence is all that is required. As is seen below, with slight modifications of the procedure, PCR can also be used to amplify sequences flanking a known region. Since polymerase chain reaction tolerates mismatches in the primer binding regions away from the 3' ends of primers, mutations could be introduced to essentially anywhere in a given gene.
1.6.1 PCR and Beyond

The powerful, yet easy to use, polymerase chain reaction has revolutionized the world of molecular biology as did the discovery of restriction endonucleases. The amplification of specific DNA sequences is only the basis, but not the final solution, to solve real problems. Many modifications to the original invention have been made that greatly increased the efficiency of the reaction. Simple tagging of primers with restriction sites and control elements has been described in the original publication (164), and is thus considered in this writing as 'common knowledge.' This section only concerns major modifications thereafter.

**Inverse PCR** Originally presented by Ochman and co-workers (176), inverse PCR was the first method designed to amplify unknown DNA sequences flanking a known region. This is a rather theoretical complementation to the PCR technique, which was designed to amplify sequence between known sequences. The method is straightforward. Large DNA molecules such as chromosomal DNA is digested with a restriction enzyme followed by ligation of the digestion mixture. Circular molecules containing the known sequence are the template for the polymerase chain reaction. Primers are designed based on the known region such that they are extended away from each other, and thus the name 'Inverse PCR.'

**Vectorette PCR** Vectorette PCR can be practically considered as a further improvement to the Inverse PCR method to increase the specificity and efficiency of the reaction. In amplifying unknown DNA segments adjacent to regions with known sequences, an obvious idea would be to ligate digested chromosomal DNA with a synthetic linker. Two primer
binding sites, one in the known region of interest and the other in the linker region, can be used to amplify the unknown region. Since the linkers will ligate to both ends of any fragment of the digestion mixture, the resulting PCR products would be a smear because all the chromosomal fragments have the potential to ligate to the linker, and the linker can ligate to both ends of each fragment. Therefore, a primer sequence must be placed on the ends of DNA molecule in such a way that only those fragments containing the target sequence can be amplified. The design of 'vectorette' as a linker (58, 195) in this process solved the problem (Fig. 3). Two oligoes are synthesized such that they base-pair on both ends but not in the central region about 30 nt in length. One of the oligoes is longer than the other such that, upon base-pairing, an over-hang or a sticky end is formed, which is compatible with that produced with a known restriction enzyme that cut the chromosomal DNA (Fig. 3-1). This looped structure, termed a 'vectorette,' serves as a linker that ligates to the digested chromosomal DNA mixture. The ligation mixture, referred to as a vectorette library, is to be used as the template for subsequent PCR. Primers used are based on the target sequence (primer 1) and the central region of the vectorette (primer 2). Primer 2 sequence is the same as one of the two strands and it will not bind the other strand since the two strands are not complementary to each other. The first cycle of PCR using the ligation mixture as templates could only extend from the 3' end of primer 1 that binds to the target sequence (Fig. 3-2). Primer 2 will then bind to the PCR product from the first cycle, and a normal PCR is assumed (Fig. 3-3). Ligated molecules without the target sequence will not be amplified since neither of the two primers will be able to bind to them.
This technique has been applied to several projects, including the ordering of overlapping cosmid clones containing the YAC (yeast artificial chromosome) insert sequences (46, 195), and the analysis of chromosomal breakpoints (157). Vectorette PCR is very easy to use and is powerful in increasing reaction specificity and efficiency by preventing amplification of nonspecific sequences. One possible reason that it is not widely used today is that for each restriction site to be used for ligation, a separate vectorette needs to be synthesized, which contains two oligonucleotides each longer than 60 bases. In practice, several restriction enzymes need to be used to find a proper one.

A potential solution is to amplify or digest a multicloning sequence (MCS) from a commonly used cloning vector (i.e., pBS or pUC18) and ligate it to the synthesized vectorette (Fig. 4). This modified vectorette carries may unique restriction sites and is universal in that digestion of this vectorette with one of these restriction enzymes compatible with the one used for the chromosomal DNA has the same output as synthesizing another vectorette. In addition to reduced cost on oligonucleotide synthesis, a major advantage of this vectorette is that several vectorette libraries can be constructed simultaneously and used as templates in PCR’s that can be carried out at the same time.

For projects with more demanding usage of this structure, and for commercial purposes, two pairs of perfectly matched oligonucleotides can be synthesized such that, upon annealing, one strand from one pair will base pair with one strand from the other pair to form a vectorette. The two oligonucleotides within each pair are complementary and anneal to form a double stranded form. Appropriate unidirectional ligations of
Fig. 3 Vectorette and its application. 1. Ligation of vectorette and chromosomal DNA fragments. 2. The vectorette library as templates for the first cycle PCR. 3. Subsequent PCR cycles in the same reaction [Adapted from Eales and Stamps (58) and Riley et al. (195)].
Fig. 4 Construction of a possible universal vectorette. The MCS in this example is a Kpn I & Sac I fragment of pBlueScript KS. The Kpn I overhang on the vectorette is to designed as part of the two oligonucleotides for the vectorette structure.
these two fragments into pBS SK+ and pBS SK−, respectively, would allow production of a specific strand from each recombinant phagemid from *E. coli* upon rescue with a helper phage. Annealing of the two ss-DNA results in large amount of universal vectorette. Since PCR is not sensitive to sequences upstream to the 5' ends of primers, the length of the vectorette should not be of concern with regards to efficiency of the method.

**PCR Using Degenerate Primer(s)** A cloning project very often starts from DNA sequences deduced either from the N-terminal sequence of a protein or from a conserved region of another gene, presumably in the same gene family. In the first case, DNA sequence thus obtained is degenerate since the genetic code is redundant for all amino acids except methionine and tryptophan. In the second case, identical DNA sequence can not be expected since proteins with same function does not signify identity in their DNA (or protein) sequence. Degenerate primer is a mixture of oligoes that represents all possible sequences. The degeneracy can be reduced to some extent by observing codon usage of other genes in the same organism and further by substituting ambiguous bases with inosine (115, 182). The proportion of 'true' primer is still very low in the mixture, and addition of large amount, up to 10 µmol in place of the usual 20-100 pmol, has been used (79). However, since PCR tolerates mismatches in primer binding, some of the sequences in the mixture, although not necessarily perfect, can be used as primers to amplify prospective genes, directly or indirectly. This idea has seen some successful applications even though variability and uncertainty always remain. A DNA fragment generated this way was successfully used as a specific hybridization probe for porcin urate oxidase (129). Their degenerate primers were
deduced from within a 32 amino acid at the N-terminus of the protein. The short PCR product was then confirmed by sequencing and used successfully to isolate a full-length cDNA sequence from a porcine liver cDNA library.

Careful PCR parameter determination (annealing temperature and primer sequences in particular) are of critical importance for this kind of amplification since they are based on target sequences that are not available. Considerable time investment (trial and error) can be alleviated by using a gradient temperature thermocycler.

1.6.2 Recombinant PCR

Strictly speaking, recombinant PCR is a derivative of the original polymerase chain reaction (164) that results in chimeric DNA molecules. The purpose of an experimental design is often beyond what the ordinary PCR provides, although the underlying principle is the same. The resultant product of the reaction depends on particular problems but in almost all cases this is achieved through careful design of primers.

The possibility of combining two DNA fragments through overlapping regions using PCR was pointed out by Mullis et al. (165). The idea was to connect synthetic oligonucleotides to build up longer and longer synthetic DNA molecules than direct chemical synthesis. The same principle was applied in the development of phage displayed single chain variable region (scFv) technology in which the heavy and light chains for a functional antigen binding fusion protein is displayed on the surface of recombinant bacteriophage (100, 162, 262). The sources of the mRNAs for the cDNA templates were either from a hybridoma (43) or from unimmunized human lymphocytes (43, 143). Sequences of Joint-genes and leader
regions were used to prime both heavy and light chain variable-genes (143). A small linker sequence encoding (Gly$_4$Ser)$_3$ (148) has sequences complimentary to relevant ends of the PCR products and was used in subsequent PCRs to assemble the heavy and light chains.

The recombinant PCR protocols were introduced with several application potentials by Higuchi (95, 96). Base substitution is a matter of amplifying the fragment separately into two PCR products with overlapping sequences. This is achieved by using two overlapping primers: one as the lower primer in the first PCR, the other as the upper primer of the second PCR. These overlapping primers are often called the 'inside primers,' or 'internal primers' as they are positioned between the two primers at the ends of the initial fragment. The other two primers are similarly called the 'outside primers.' The two purified PCR products can be extended, when mixed together, in the absence of any primer because each fragment serves as the primer on the other one that is now the template. These inside primers are where manipulations can be introduced, thanks to the fact that primer binding tolerates mismatches away from the 3' end.

In addition to point mutations, desired length of DNA sequence can be deleted using the above protocol with minor modifications. With the 5' ends of inside primers located at the starting points of the sequences to be retained, homologous sequences to the 5' portion of the lower primer of the other PCR can be added. When the amplified products are self-assembled, the homologous region is where connection occurs and the sequence between the two inside primers is deleted.
Several bases can be added to the middle of the inside primers, and these are base insertions to the original DNA fragment. Since the added bases are not complementary to the template, they will loop out during initial binding. This does not happen during subsequent cycles since newly synthesized DNA molecules are the templates and are complementary to the primers, including the added region. The efficiency of this kind of PCR is generally not a concern.

The two PCR products to be connected do not need to be from the same longer fragment. A homologous region added to the 5' ends of the inside primers, sometimes called "add-on" (95), is all that is needed to combine the two PCR fragments. For example, one fragment could be a promoterless gene from one organism while the other could be the promoter region from another organism. It should also be clear that the power of recombinant PCR is to rearrange DNA sequences. The ultimate expression of a molecule is dependent on the genetic problem itself.

One obvious advantage of recombinant PCR is that there is no prerequisite, such as restriction site availability, to the manipulations discussed above. The mutations, in a broader sense, can be introduced into virtually anywhere on any DNA molecule, and the connections can be precisely controlled by careful primer design. The disadvantage is intrinsic to the Taq DNA polymerase itself: the fidelity of PCR is considerably lower than that of \textit{E. coli} holoenzyme (the \textit{E. coli} DNA-dependent DNA polymerase). The \textit{in vitro} PCR has higher mutation rate than the \textit{in vivo} amplification of a DNA fragment inserted into a replicon that is replicated by the \textit{E. coli} holoenzymes. The problem can be alleviated by using polymerases that have proof-reading capabilities, such as the \textit{Vent} DNA polymerase.
Functional screening, however, is still necessary unless the entire recombinant PCR product is to be confirmed by sequencing. There is also a limit of add-on's in length. Although add-on's as long as 45 bases have been reported (209), chemical synthesis of longer primer results in, besides higher cost, progressively smaller portion of full length products. Forty-five bases add-on plus the 3' portion (which is the only sequence responsible for primer binding) will add up to more than 60 bases, approaching the upper limit of chemical DNA synthesis.

1.6.3 PCR Cloning and the Choice of Polymerases

With the polymerase chain reaction in mind, essentially any fragment of DNA can be amplified and cloned into anywhere in a vector without concerning the availability of restriction sites. This has made many difficult experimental designs easier and is becoming indispensable to researchers familiar with this technology. One major issue deserves careful consideration if a PCR product is to be cloned: the fidelity of the polymerase. Taq polymerase is the first one that introduced the PCR revolution to average molecular biology laboratories but it is the very one that lacks 3' to 5' exonuclease (proof-reading) activity (70). Primers can be extended more reliably by other thermostable polymerases with 3'-5' proof-reading capability, such as the Vent DNA polymerase from *Thermococcus litoralis* (184), and the Pfu DNA polymerase from *Pyrococcus furiosus* (139), both available commercially. The Vent DNA polymerase was purified from the archaea *Thermococcus litoralis*, isolated from a submarine thermal vent (19) and capable of growth at up to 98°C (184).
By measuring opal codon reversion frequency, it has been shown that the error rate of DNA synthesis by Taq polymerase is 3/10,000 bases, 10 fold higher than observed for Vent polymerase that is in the range of 3-5/100,000 bases (59, 120, 145). With the average gene size of 2,000 bp, Taq-amplified PCR product does not seem to be a choice for cloning purposes. Even with Vent-amplified DNA fragments, functional screening is necessary after cloning.
CHAPTER 2 SEQUENCING AND SEQUENCE ANALYSIS

2.1 INTRODUCTION

Two parallel pathways have been postulated in marine vibrios for catabolism of chitin, possibly containing as many as 6-10 enzymes and a number of chemotactic proteins (273). In the common part of the pathway, chitin-binding proteins adhere to the substrate (77, 274), and extracellular chitinase (123) and periplasmic chitodextrinase work together to produce \( N,N' \)-diacetylchitobiose (15). The glycosidase/PTS system cleaves \( N,N' \)-diacetyl chitobiose to GlcNAc in the periplasmic space via a membrane bound chitobiase (15, 220, 265) after which GlcNAc is transported and phosphorylated by the PTS (15). The second, parallel permease/glycosidase system, resembling the \( E. coli \) lac permease/beta-galactosidase system (275, 276), utilizes an \( N,N' \)-diacetylchitobiose permease for transport of this substrate to the cytoplasm. The transported \( (\text{GlcNAc})_2 \) is cleaved by the cytoplasmic chitobiase reported here (278) and phosphorylated by a ATP-dependent \( N \)-acetyl-\( D \)-glucosamine kinase (6, 12, 15). The cytoplasmic system works independently of the PTS (15). This report gives the nucleotide sequence and the deduced polypeptide sequence of the gene encoding the cytoplasmic chitobiase (EC 3.2.1.14) from \( V. parahaemolyticus \) and shows an ancient evolutionary divergence for this unique beta-hexosaminidase when compared with other chitobiases.

2.2 MATERIALS AND METHODS

2.2.1 Host Bacterial Strains, Vectors and Phages

\( E. coli \) strains DH5\( \alpha \) and DH5\( \alpha F' \) were purchased from Gibco Bethesda Research Laboratories (Gaithersberg, MD). \( E. coli \) strain JM101,
phagemid vectors pBS II KS\(^+\) and pBS II SK\(^+\) and the interference resistant helper phages VCSM13 (\(kan^R\)) and R408 were from Stratagene (La Jolla, CA). The plasmid harboring the \textit{V. parahaemolyticus} chitobiase gene, PC120, was constructed in this laboratory and has been described previously (278).

2.2.1 Enzymes, Chemicals and Antibiotics

Restriction endonucleases, T4 DNA ligase, T4 DNA polymerase were purchased from Gibco Bethesda Research Laboratories (Gaithesberg, MD), New England Biolabs, Inc. (Beverly, MA) or United States Biochemicals Corp. (Cleveland, OH). T7 DNA polymerase DNA sequencing kit (Sequenase 2.0) was from United States Biochemicals Corp. (Cleveland, OH). 35S-deoxyadenosine 5'-[alpha-thio]-triphosphate was from Du Pont Co.-NEN (Boston, MA). Ampicillin was from Sigma Chemical, Co. (St. Louis, MO). Oligonucleotides used as sequencing primers were synthesized on an automated DNA/RNA synthesizer (Applied Biosystems, Model \textit{abi} 394) in GeneLab, School of Veterinary Medicine, Louisiana State University (Baton Rouge, LA).

2.2.2 Construction of Subclones for Sequencing

The chitobiase gene harboring plasmid PC120 (278) was digested with the restriction enzymes \textit{Pst} I, \textit{Sac} I and \textit{Hind} III, respectively. The 1.6 kbp, 2.1 kbp and 3.5 kbp fragments were gel purified and cloned into appropriate sites of pBS II SK\(^+\) vector for single strand DNA production. Insert orientations of the clones were identified by digestion with appropriate restriction enzymes that cut the inserts asymmetrically (\textit{Hind} III for the 2.1 kbp fragment and \textit{EcoR} V for the 3.5 kbp fragment), and the clones thus obtained were named as SKS162, SKP21-1, SKP21-2, SKH35-1 and
SKH35-2 (Fig. 5). The protocol for rescuing recombinant phagemid using VCSM13 was from Stratagene (La Jolla, CA) (except that *E. coli* DH5αF' instead of XL1-Blue, Luria Bertanic medium (LB medium) instead of Super-Broth, were used).

### 2.2.3 Sequence Analysis

Sequencing gel data were assembled and analyzed using Staden’s algorithm (221), an integrated part of the GCG software, Unix® version 8.0 (Genetic Computing Group, Madison, WI). The identified chitobiase open reading frame (ORF) and the predicted polypeptide sequence were then used as primary query sequences to search available nucleic acid and protein depository databases.

### 2.3 RESULTS

#### 2.3.1 Nucleotide Sequence Determination

The sequencing clones and the orientations of the inserts are shown in Fig. 5-1. The two phagemids, SK21-1 and SK21-2, contain the 2.1 kbp *Pst* I fragments but in opposite orientation. Both SKS162 and SKH351 overlap with the above two clones to facilitate assembly of sequencing gel data. The restriction map of PC120 as determined previously (278) and from the sequence data is depicted in Fig. 5-2. Using the Staden algorithm (221), a single ORF was identified from sequencing data from the clones shown in Fig. 5. The area sequenced and the coding region of *V. para-haemolyticus* chitobiase are also shown. The 47 amino acid residues at the N-terminus of the polypeptide are located between the *Pst* I and *Sac* I restriction sites, indicating that expression of the chitobiase gene on

---

1 The DNA sequence presented here has been deposited with the GenBank/EMBL (21) sequence data bank and is available under accession number U24658 (VPCHB).
PC120 is driven by the *V. parahaemolyticus* chitobiase gene promoter instead of the *lac* promoter on pUC18, the parental cloning vector. The nucleotide sequence and the deduced amino acid sequence of the gene are shown in Fig. 6. The N-terminal sequence of the polypeptide from Edman degradation (278) perfectly matched that deduced from the DNA sequence, suggesting the lack of a signal peptide at the N-terminus. This agrees with the cytoplasmic localization of the enzyme.

By attaching a signal peptide upstream of the first initiator AUG methionine codon, this enzyme has been shown to secrete across the two membranes of the Gram negative *E. coli* (Chapter 3), further indicating that the wild type chitobiase of *V. parahaemolyticus* is cytoplasmic. The predicted molecular weight of the deduced 741 amino acid polypeptide is 85 kDa, agreeing with that determined by SDS-PAGE (278).

### 2.3.2 Homology with Other Chitobiases

The deduced polypeptide sequence of *V. parahaemolyticus* was used as a query sequence to search the GenBank/EMBL (21) and Swiss-Prot genetic data bases using GCG program (FastA) based on the algorithm of Pearson and Lipman (183). Limited homologies, aligning in a 60 amino acid area of a composite map from residues 341-400, were found between this enzyme and those of *V. harveyi*, an outer membrane protein (110, 220) and *V. vulnificus* (218). Except for lysosomal chitobiases from human and rabbit which cleave (GlcNAc)$_2$ from the reducing end and are related to each other (66), a remarkable homology was found among 31 amino acids out of 930 to all sequenced beta-hexosaminidases in the composite map locations 296-325, 341-359, 380-400 and 358-465 (Fig. 7). Particularly interesting in this region is position 359, at which an
Fig. 5 Restriction map and nucleotide sequencing strategy of the sequenced region containing the $N,N'$-diacetylcitobiase gene from *V. parahaemolyticus*. 1. Inserts in subclones used to generate single stranded DNA for sequencing. Arrows represent insert orientation relative to the chromosomal DNA fragment in PC120 (278). 2. Restriction sites used to generate the inserts shown above and the sequencing strategy. Arrows below the restriction map are individual sequencing gel data used in sequence assembly. Sequence presented in this paper is boxed. The solid portion indicates the chitobiase ORF. Only restriction sites used for subcloning are shown. $P,$ *Pst I*; $S,$ *Sal I*; *Sc,* *Sac I*; $H,$ *Hind III.*
Fig. 6 Nucleotide sequence of the chitobiase gene from V. parahaemolyticus. Presented in the figure is the coding strand (non-template strand) of the DNA. Putative regulatory elements and the N-terminal sequence of the predicted polypeptide as determined by Edman Degradation (278) are underlined. Numbering of nucleotides is based on the putative transcription start site (+1). -35 and -10: -35 and -10 regions of the promoter, respectively; SD: Shine-Dalgarno site or ribosome binding site (RBS). Note the three consecutive translation Opal stop codons (TGA) at the end of the ORF.

(Fig. con’d)
GCA GAT CAA CCA GCC TCA ATG GAA GTG GTC TGG TCT TCA ATC AAA GAC CGA CCA CGT TTT 822
Gly Asp Glu Pro Ala Ser Met Glu Val Val Cys Cys Ser Ile Lys Asp Arg Pro Arg Phe 260

GCT TAC CGC GGT ARG ATG CTA GAT TGT GCT CC GAT TCT TAC CCC TGG GAA GTC AAA 882
Arg Tyr Arg Gly Met Met Leu Arg Cys His His Ser Val Glu Glu Val Lys

GCT TCG ATC AAC CAG TTG GCT CAC TAC AAC TTT TTC ATC AAC TTT ACC GAT 942
Arg Leu Ile Asp Glu Leu Ala His Tyr Lys Phe Asn Thr Phe His Thr Thr Arg 300

GAT GAA GTG TG GAA ATG AGT AAA TCA TTG CCT CAA CTA ACC TAT GAT GCC GCA TGG 1002
Asp Glu Gly Trp Arg Ile Glu Ile Lys Ser Leu Pro Glu Leu Thr Asp Ile Gly Ala Trp 320

TGT ATC TAC TCC GAA ATC GAC GTC TGG CCT GAT CCC GCA GCA ACC TAC AAC GAA 1182
Arg Leu Val Glu Ala Glu Asp Thr Thr Glu Tyr Arg Ser Ile Gln His Tyr Asp Asn 400

AAA GAA CG CG GAA AG TGC TGG TGC ATT AAC CCA GCT TCT GCG GGC ATT GAG 1302
Val Ile Asp Pro Ala Leu Pro Gly Ser Tyr Glu Phe Asp Lys Val Leu Glu Glu Val 420

TTT ATC TAC TCC GCT CCC CTP TTT CTA TCG CGG AGT AAC ACT AAC GAC GCA TCG 1362
Ser Arg Cys Ser Cys Ala Leu Met Phe Ala Ser Val Arg Thr Asn Gly Val Trp 440

TCA AAA AGC CCT GCA TGG CAA GCA CTA ATG GAA CAA CTA GAT TAC AGC AAC TAC AAA GAG 1422
Ser Lys Ser Pro Arg Cys Ala Glu Ala Leu Met Glu Glu Leu Gly Tyr Ser Asp Tyr Lys Ser 460

TTA CAA GGG CAC TTC TTG GTG CAT GCC GAC AAA CGC AAA CCT TGC GAA CGG ATG 1482
Leu Gln Gly His Phe Leu Arg His Ala Glu Asp Lys Leu Arg Lys Gly Lys Arg Met 480

CTG GTT TCG GAA GAA GCA CAG CAT GGC GAC AAA GTC AGC AAA GAC ACA GTG ATC TAT TCG 1542
Leu Gly Trp Glu Ala Glu His Gly Asp Lys Val Ser Lys Asp Thr Val Ile Tyr Ser 500

TTG TTA GAC GAA GGC GGG TCG AAC TGC CCC CCC CAA GTT TCC GAT GCG TGG CTA CAA 1602
Trp Leu Ser Glu Glu Ala Asn Cys Ala Arg Glu Gly Phe Asp Val Leu Glu 520

CCT GCG GAC ACC TAC TTA GAT ATG ACC CAA GAT TAC GCA CCA GAA GCA CGC GCC GTG 1662
Pro Ala Gln Thr Tyr Leu Asp Met Thr Glu Asn Tyr Ala Pro Glu Glu Pro Gly Val 540

GAT TGG CTT AAC CCA TGG CCG CTA GAA AAA GCT TAC AAC TAT GAA CCA CTC GCT GAA GTC 1722
Asp Trp Ala Asp Pro Leu Pro Leu Gly Glu Asp Tyr Asn Tyr Glu Pro Leu Ala Glu Val 560

CCA GCC GAT CAC CAA ATA GGG AAA GCC ATT TGG GGC ATT CAA ACA GCA TTG TGG TCG GAA 1782
Pro Ala Asp Pro Asp Arg Asp Ala Ile Trp Gly Ile Gln Thr Ala Leu Trp Cys Glu 580

ATC ATC AAC AAC CAG TCT CTT GTG GAC TAC ATG GTG TTC CCC CCG TTA ACC GCA ATG GCA 1842
Ile Ile Asn Asn Glu Ser Met Asp Tyr Met Val Phe Pro Arg Leu Thr Ala Met Ala 600

GAA GCA TGT TGG ACA GAC AGG CAA CAC CGA GAC TGG ACC GAC TAT TTA TCA CGT TGG AAA 1902
Glu Ala Cys Trp Thr Asp Lys Gin His Arg Asp Trp Thr Asp Tyr Leu Ser Arg Leu Lys 620

GCA CAC CTA CGT CTT GAT CAG GQA GTG TAT AAC TAC CTC CAC GCA GAT AGT ACA 1962
Gly His Leu Pro Leu Leu Asp Leu Gin Leu Val Asn Tyr Asn Arg Arg Gly Ser Asn Thr 640

GAG CAT TGT ATG AGA AGC ATC AGC CTT GAA GAG TTT TTA ATT TTT GCC TGC AGC GAA 2022
Glu His Cys Ser Arg Ser Ile Thr Leu Glu Glu Phe Leu Asn Phe Gly Cys Arg Ser 660

TTT CGA AAA AGG AAT ACA CAA ATG AAA TAC GCC CAT TGT TCC GAT AAC GAG ATG CTT GAA TAC 2082
Phe Val Lys Arg Asn Thr Glu Met Lys Tyr Gly Tyr Phe Asn Glu Asn Arg Glu Arg 680

(Fig. con'd)
GTC ATT ACT GGC CCT GAT GTA CCT GCT GCT TGG ACC AAC TAC CTA GCT ACG GAA AAA TTC
Val Ile Thr Arg Pro Asp Val Pro Ala Pro Thr Asn Tyr Leu Gly Thr Glu Lys Phe

TGT ACC GTT ATC TCG CAT AAC GCA GGT GGC TAT TCG TTC TAC AAC TCT CCA GAA TAC AAC
Cys Thr Val Ile Ser His Asn Ala Gly Gly Tyr Ser Phe Tyr Asn Ser Pro Glu Tyr Asn

CGT GTT ACT AAG TTC GCT CTA AAT GCG ACA TTT CGA TCG CCC AGG ACA CTG TTA CCT
Arg Val Thr Lys Phe Arg Pro Asn Ala Thr Phe Arg Ser Pro Arg Thr Leu Arg Leu Pro

ACG TGA TGA TGA GACGGGAGATTACGGTCAATCTCTTGGCAACCAGTTGCAAAAGCCTAGACGAAGCGAACTACG
Thr TER TER TER

AAGTTCGTCAGGTTTGGCTACTCTAAATTCTAAGTGTTAACACGGGATATGTCAGGCTACACG
2337

AAAAGGCGAAGATGCAAACATAGCTTTGATGTGTTAGCTCAAGCCTGTGGCCACTGACG
2416

TTTGTGATGCTCTGGGCAGCCACACTTCAAGTACAACCAACTTCTTGTGTCTCAGC
2495

ACAAGCCGGCTATGTAACCTGTACTACAAACACTAAGCTGCTGACGCTGACG
2574

ATCGAGATTCAATGCGAACACGGCTCAAGGATTCTCTACACATCTACACTACACTAAACG
2653

CAGGTAAGGTTACACGCAACTATCGCTGTGTGCTCTGCTCTACTGCTGTATT
2732

2786
Fig. 7 Amino acid alignment of three vibrio chitobiases. Shown are regions displayed by the FastA program. Numbers above or beneath sequences indicate positions of amino acids in the original polypeptides, starting from the amino termini. 1. vpchb vs. vhchb, 28.6% identity in 336 amino acid overlap; 2. vpchb vs. vvhchb, 31.7% identity in 284 amino acid overlap; 3. vhchb vs. vvhchb, 34.4% identity in 785 amino acid overlap. vhchb: V. harveyi chitobiase (220); vvhchb: V. vulnificus beta-hexosaminidase (218); vpchb: V. parahaemolyticus chitobiase, this study. (Fig. con'd)
(Fig. con'd)
(Fig. con’d)
Fig. 8 Amino acid sequence pileup of chitobiases and beta-hexosaminidases from 10 organisms. 1. Dendrogram of aligned amino acid sequences. Distance along the horizontal axis is proportional to the difference between sequences. 2. Amino acid sequence alignment. Numbers above the sequences represent positions in the alignment, not for individual polypeptide. Highlighted residues are identical to those in vpchb. The boxed arginine residues at position 359 include R^{178} of hshxa and R^{211} of hshxb, respectively (36). The arginine position in the chitobiase polypeptide is 271. hschb: *H. sapiens* chitobiase (66); rnchb: *Rattus norvegicus* chitobiase (66); hshxa and hshxb: alpha- and beta-polypeptides, respectively, of *H. sapiens* beta-hexosaminidase (116); mmhxa and mmhxb: alpha- and beta-polypeptide, respectively, of *Mus musculus* beta-hexosaminidase (16, 269); ddhxa: *Dictyostelium discoideum* beta-hexosaminidase (82); vhchb: *V. harveyi* chitobiase (220); vvchb: *V. vulnificus* beta-hexosaminidase (218); vpchb: *V. parahaemolyticus* chitobiase, this study.

(Fig. con'd)
(Fig. con'd)
(Fig. con'd)
arginine residue is conserved for all the enzymes listed except for lysosomal chitobiases from human and rabbit. It has been shown (36) that Arg$^{178}$ and Arg$^{211}$ (aligned at position 359 in Fig. 8) in the alpha- and beta-subunits of human beta-hexosaminidase, respectively, are "active" residues. They are part of the catalytic sites, but they do not participate in substrate binding. Fig. 6 also shows that periplasmic chitobiases from V. harveyi and V. vulnificus have extensive homology with each other while homology of either of these with the cytoplasmic chitobiase from V. para-haemolyticus is much lower. This implies that cytoplasmic chitobiases from vibrios took a very different line of evolution than periplasmic chitobiases and these signal-sequence-containing enzymes are more closely related to beta-hexosaminidases from higher organisms.

The structural gene and the deduced amino acid sequence of the V. para-haemolyticus chitobiase were progressively piled up (64, 172) to those of chitobiases and beta-hexosaminidases from other organisms including other vibrios and higher organisms. The results are shown as a dendrogram (Fig. 8-1) as well as sequence alignment (Fig. 8-2). Identical relationships among these organisms were obtained using either the DNA or the amino acid sequences for comparison (only the amino acid data are shown). The clustering relationships show the uniqueness of the cytoplasmic chitobiase from V. para-haemolyticus among the chitobiases and beta-hexosaminidase from all three vibrios.

2.4 DISCUSSION

Roseman et al. showed that V. furnissii possessed a separate N,N'-di-acetylchitobiase with cytoplasmic localization which depended on a (GlcNAc)$_2$ permease (15, 273, 274). We isolated such an N,N'-diacetyl
chitobiase of *V. parahaemolyticus* from the cytoplasm (278). The lack of signal peptide at the N-terminus in the gene confirms this assignment. It appears that a hydrophobic patch is absent throughout the entire length of the polypeptide sequence, as shown in Fig. 9. In addition, a naturally secreted endo-chitinase has been characterized and the gene cloned from *V. parahaemolyticus* (123). The major product of chitin cleavage by this chitinase is \(N,N'\)-diacetylchitobiose, the substrate of the periplasmic and cytoplasmic \(N,N'\)-diacetylchitobiases. A chitin binding protein to facilitate *V. parahaemolyticus* to adhere to chitin has also been reported (77). The presence of a permease-like protein to translocate chitobiose across the cytoplasmic membrane of *V. parahaemolyticus* is thus obligatory. The involvement of a permease is supported by *in vitro* data from *V. furnisii* (15). Complementation analysis (219) showed that a lac permease defective *E. coli* strain (LE392, lac\(Y^-\) mutant) carrying plasmids with chromosomal DNA fragment from *V. harveyi* were able to hydrolyze o-nitrophenyl-beta-D-galactoside (ONPG), implying that the proposed putative chitopermease is equivalent in function to the lac\(Y\) gene product and is able to translocate lactose across the cytoplasmic membrane of *E. coli*. Identification and characterization of the permease protein in *V. parahaemolyticus* will establish the complete chitinoclastic pathway in this organism.

Although it has been shown that the PTS exists in *V. parahaemolyticus* (150) and all vibrios tested (149, 150), the establishment of an alternative pathway for the utilization of chitin by *V. parahaemolyticus* (150) requires further investigation, particularly the presence of a second \(N,N'\)-diacetyl-chitobiase that is periplasmic.
Fig. 9 Hydrophobicity analysis of chitobiases from two vibrios according to Kyte and Doolittle (122). 1. *V. parahaemolyticus* (this study); 2. *V. harveyi* (220). Parameters used for the plots: average window size of 11 amino acids, compression factor of 4. The highly hydrophobic N-terminal region (the signal peptide) in *V. harveyi* prechitobias is indicated by arrow.
It is worth noting that sequences in the -10 and -35 region of the *V. parahaemolyticus* chitobiase gene promoter are close to the prokaryotic consensus promoters sequences (5'-TATAAT-3' for -10 and 5'-TTGACA-3' for -35) (92, 197), especially in the -35 region where the only difference is that the *V. parahaemolyticus* chitobiase gene has a G instead of an A as in the consensus sequence, implying high expression of the gene driven by its own promoter. This agrees with the high yield of the enzyme using PC120, the clone with the chromosomal insert from *V. parahaemolyticus*. On the other hand, the chitobiase level in *V. parahaemolyticus* itself is at least 30 fold lower (Zhu, B. C., personal communication). One possibility is that the expression of chitobiase gene is regulated in the original organism but is not when cloned into *E. coli*, although other alternatives such as higher gene dosage in *E. coli* carrying PC120 may also play a role. In contrast to the chitin degradation systems of *V. harveyi* and *V. vulnificus* in which catalytic and non-catalytic proteins are organized as simple *chi*-operons (219, 265), the chitinase and chitobiase genes for the two major catalytic enzymes in *V. parahaemolyticus* do not seem to be in the same cistron, since both have their own promoters and ribosome binding sites; nor are they in close vicinity, since sequences 1000 nt upstream and 500 nt downstream of the chitobiase ORF do not overlap with the chitinase gene (data not shown). At this point, we do not yet have enough data to conclude the presence or the absence of a *chi*-operon in this organism. The regulatory effect involved in the chitobiase gene of *V. parahaemolyticus* may be in trans.

Two highly conserved regions were revealed from Fig. 8-2 among all chitobiases and hexosaminidases except for human and rabbit lysosomal
chitobiases which are exoglycosidases that split the GlcNAc-beta-D-(1-4)GlcNAc chitobiose core of asparagine-linked glycoproteins from the reducing ends of the dimeric sugar (66, 121). It is highly likely that amino acid residues in one or both of these regions participate in the catalysis of their respective substrate, especially the region from position 341-400 that includes the catalytic arginine residues in the alpha- and beta-subunits of human beta-hexosaminidase (36).

This report shows the uniqueness of the cytoplasmic chitobiase gene of *V. parahaemolyticus* when compared to other chitobiase genes cloned from vibrios or other organisms. A cytoplasmic chitobiase activity was proposed by Roseman's group (15) and cloned and isolated by Zhu et al. (278). The sequence and homology comparisons in this report establish an evolutionary relationship among similar enzymes and further show the extent of genetic investment in chitin degradation by vibrios.

### 2.5 SUMMARY

The cytoplasmic *N,N*-diacetylchitobiase (EC 3.2.1.14) from *Vibrio parahaemolyticus* has been characterized and the gene cloned into *E. coli* [Zhu, B. C. et al. (1992) *J. Biochemistry (Tokyo)* 112, 163-167]. The nucleotide sequence of the gene encoding this unusual beta-hexosaminidase has now been determined, and the deduced peptide sequence surprisingly has minimum evolutionary relationship to two other reported *N,N*-diacetylchitobiases from vibrios, except for highly conserved regions which are also homologous with lysosomal beta-hexosaminidases from eukaryotes including humans. In contrast, the other two beta-hexosaminidases with reported sequences from vibrios are much more closely related to each other. This novel 85 kDa cytoplasmic
glycosyl hydrolase of restricted specificity participates in the high level utilization of chitin-derived 2-deoxy-2-acetamido-D-glucose (GlcNAc) by vibrios as one of two parallel pathways for metabolism of N,N'-diacetylchitobiose [Bassler, B. L., Yu, C., Lee, Y. C. & Roseman, S. (1991) J. Biol. Chem. 266, 24276-24286]. This complex chitin utilization system contains as many as 6-10 enzymes and a number of chemotactic proteins [Yu, C., Bassler, B. & Roseman, S. (1993) J. Biol. Chem. 268, 9405-9409]. These pathways use chitin-binding proteins for the adherence of the bacterial chitinase to the substrate [Yu, C., Lee, A. M., Bassler, B. L. & Roseman, S. (1991) J. Biol. Chem. 266, 24260-24267; Gildemeister, O., Zhu, B. & Laine, R. (1994) Glycoconjugate J. 11, 518-526], and extracellular chitinase and periplasmic chitodextrinase to produce N,N'-diacetylchitobiose. At this point the pathways branch. On one hand, a periplasmic, membrane anchored N,N'-diacetyl-chitobiase cleaves (GlcNAc)\textsubscript{2} and the liberated GlcNAc is simultaneously transported to the cytoplasm and phosphorylated by the bacterial phosphoenolpyruvate: glycose phosphotransferase system (the PTS). The alternate pathway utilizing the cytoplasmic N,N'-diacetyl chitobiase requires a permease for N,N'-diacetyl-chitobiose, and is mechanistically equivalent to the lactose permease of E. coli. The V. parahaemolyticus cytoplasmic N,N'-diacetyl-chitobiase appears to be an unique protein, lacking a signal sequence and genetically distant from other known chitinoclastic beta-N-diacetyl-hexosaminidases. This is consistent with its limited substrate specificity to small GlcNAc terminated oligosaccharides.
CHAPTER 3 PROTEIN SECRETION

3.1 INTRODUCTION

To facilitate collection and purification of recombinant proteins, several signal sequences have been fused to proteins from various sources and cloned into \textit{E. coli} (69, 74, 85, 141, 161, 177, 224), yeast (45, 89, 97, 170, 174, 234, 267, 268), insect cells (\textit{Sf} cell line, a baculovirus host) (133, 228), transgenic tobacco plant cells (138), and mammalian cells (CHO and COS cells) (86, 117, 131). In \textit{E. coli}, the secreted protein often accumulates in the periplasmic space (18, 85, 141, 224). Signal sequences have been used to construct vectors for protein secretion from cells of yeast (27, 159, 174), insect (228), and \textit{E. coli} (84, 141, 224). In the \textit{E. coli} secretion systems, some of which are commercially available (84, 141, 224), collection of gene products requires disruption of the cell wall, because the target location of the recombinant proteins is the periplasm. Gram negative bacteria do not secrete many proteins. To our knowledge, no secretion vector has been available for the harvest of wild type cytoplasmic proteins in \textit{E. coli} culture medium.

Chitinase is a glycohydrolase whose target polymer must be degraded outside the periplasm of vibrios. Resultant oligosaccharides are processed by periplasmic chitodextrinases and chitobiases or transported into the cytoplasm for degradation (15). Cloning of the extracellular chitinase into \textit{E. coli} resulted in high level secretion of the enzyme (123), which apparently has the correct architecture to pass the protein through the two membranes of Gram negative bacteria. To test the universality of this chitinase signal sequence (MIRFLCAAGVALALSGAAVA), we fused the DNA fragment encoding this peptide to the
cytoplasmic chitobiase structural gene, using recombinant polymerase chain reaction. This report shows highly efficient secretion of the active chitobiase into *E. coli* culture medium and the construction of two secretion vectors that could be potentially useful for the secretion of other proteins.

### 3.2 MATERIALS

#### 3.2.1 Host Strains and Plasmids

*Escherichia coli* strains DH5α and DH5αF" were purchased from Gibco Bethesda Research Laboratories (Gaithersberg, MD). *E. coli* strain JM101 was from New England Biolabs, Inc. (Beverly, MA). The phagemid vectors pBS SK+ was from Stratagene (La Jolla, CA). The expression vectors pKK233-2*(trc)* and pKK223-3*(tac)* were from Pharmacia Biotech Inc. (Alameda, CA). The plasmid harboring the *Vibrio parahaemolyticus* chitobiase gene, PC120, has been described previously (278). The plasmid pKKA1 carrying the *V. parahaemolyticus* chitinase gene and its signal sequence was from the *V. parahaemolyticus* chitinase gene was constructed previously (123). The chitobiase structural gene was amplified from the plasmid PC120 (278), based on the nucleotide sequence reported previously (Wu and Laine, submitted), from this laboratory.

#### 3.2.2 Enzymes, Chemicals and Antibiotics

Restriction endonucleases, T4 DNA ligase, T4 DNA polymerase, T4 polynucleotide kinase and calf intestine phosphatase were purchased from Gibco Bethesda Research Laboratories (Gaitherberg, MD), New England Biolabs, Inc. (Beverly, MA) or United States Biochemicals Corp. (Cleveland, OH). T7 DNA polymerase DNA sequencing kit (Sequenase 2.0) was from United States Biochemical Corp. (Cleveland, OH). $^{35}$S-ATP
was from Du Pont Co.-NEN (Boston, MA). Vent DNA polymerase, deoxyribonucleotides and MgSO₄ solution were from New England Biolabs, Inc. (Beverly, MA). The polymerase chain reactions were carried out as previously described (164, 165). Ampicillin and p-nitrophenyl derivatives were purchased from Sigma Chemical, Co. (St. Louis, MO). Oligonucleotides used as sequencing and PCR primers were synthesized on an automated DNA/RNA synthesizer (Applied Biosystems, Model abi 394) in GeneLab, School of Veterinary Medicine, Louisiana State University (Baton Rouge, LA).

3.3 EXPERIMENTAL PROCEDURES AND RESULTS

3.3.1 In-frame Connection of DNA Fragments

To fuse the *V. parahaemolyticus* chitinase signal peptide to the N-terminus of chitobiase, two PCR products were connected, using the recombinant PCR method modified from that by Higuchi (95, 96). Briefly, the two pairs of primers were used to amplify the regions (PCR1 and PCR2) encompassing the signal sequence and the chitobiase structural gene, using pKKA1 and PC120 as templates, respectively (Fig. 10). The 10-base long sequence at the 5' end of the lower primer for PCR1 was from the 5' end of the chitobiase structural gene. Similarly, the 10 base stretch at the 5' end of the upper primer for PCR2 was from the 3' end of the chitinase signal sequence, as indicated by the raised portion of the two arrows in Fig. 10-2. These two oligonucleotides thus have a perfect match region of 20 bases in length and are designated as "inside primers." The other two primers are similarly designated as "outside primers."

The two inside primers were designed in order to fuse the N-terminal methionine to the carboxy terminal of the first 6 amino acids of the
mature chitinase, immediately following the signal peptide (Fig. 10-1). The two PCR products, PCR1 (annealing: 56°C, 1 min.; polymerization: 72°C, 25 sec.; denaturation: 94°C, 1 min.; Vent DNA polymerase: 2 unit; Mg²⁺: 16 mM; 25 cycles) and PCR2 (polymerization: 2.3 min.), were then gel-purified and used as PCR primers and templates to each other, and the reaction allowed for 10-12 cycles (polymerization: 2 min.) in a 100 μl volume. Two units of Vent DNA polymerase and 2 μl of 20 μM outside primer mixture were then added and the reaction proceeded for 25 cycles under the same conditions (Fig. 10-3). The products of the 3 PCR reaction were monitored by agarose gel electrophoresis (Fig. 10-4).

In addition to the signal sequence and chitobiase structural gene, PCR3 also contains the strong trc promoter (33, 163) and the E. coli lacZ Shine-Dalgarno ribosome binding site (S.D.). The 3' end of PCR3 encompasses the 3 consecutive translational stop codons (TGA TGA TGA) of the chitobiase gene. PCR3 is therefore an intact transcriptional unit that, when carried on a plasmid, is expected to express and secrete the recombinant chitobiase.

### 3.3.2 Construction of Chitobiase Secretion Clones

For convenience of cloning, the upper primer for PCR1 was designed such that the Bam HI site was at the 5' end of PCR1 (and PCR3). PCR3 was blunt-ended with T4 DNA polymerase and inserted into the EcoRV site of pBS SK⁺, and the ligation mixture was used to transform E. coli JM101. White colonies were incubated at 37°C for 20 hrs. The supernatant and cell pellet suspension were used to test the activity of chitobiase, using p-NP-GlcNAc as substrate. One of the colonies was found to express and secrete chitobiase according to the assay, and was
Fig. 10 In-frame connection by recombinant PCR. 1. Internal primer sequences and their arrangement during self-priming. Also shown in the top panel is the fusion of proteins between the serine from chitinase gene and the initiator Met of chitobiase. 2. Amplification of signal sequence from pKKAl and chitobiase structural gene from PC120. Primers used are indicated by arrows on the plasmids and PCR products (PCR1 and PCR2) are shown with their respective sizes. 3. Connection of the fragments. Shown are the two PCR products that serve as both templates and primers during self-priming for the extension of the two fragments. The resulting product, PCR3, is reamplified by adding outside primers. 4. Agarose gel stained with ethidium bromide showing the 3 PCR products. Five µl was used from each of 100 µl PCR reaction. M: 1 kb DNA ladder; P1, P2, P3: PCR1, PCR2 and PCR3, respectively.

(Fig. con'd)
2

\begin{align*}
\text{P}_{trc} & \quad \text{chb} \\
\text{PCR 1} & \quad \text{PCR 2} \\
\text{B} & \quad \text{chb} \\
\text{PCR 3} & \quad \text{(2647 bp)}
\end{align*}

3

\begin{align*}
\text{PCR2 Upper Primer} & \quad \ldots \text{GCT CCG ACC GCA CCA AGT ATG GAA TAT CGT GTT GAT CT} \ldots \\
\text{PCR1 Lower Primer} & \quad \ldots \text{A P T A P S M E Y R V D L} \ldots \\
\text{PCR1 Product} & \quad \ldots \text{G A G C G T G G C T G G T T C A T A C T T A T A G C A C A A C T A G A} \ldots \\
\text{PCR2 Product} & \quad \ldots \text{G C G T G T C A T A C T T A T A G C A C A A C T A G A} \ldots
\end{align*}

4

\begin{align*}
\text{M} & \quad \text{P1} \quad \text{P2} \quad \text{P3} \quad \text{M} \\
3054 \text{ bp} & \quad \text{3054 bp} \\
2036 \text{ bp} & \quad \text{2036 bp} \\
1636 \text{ bp} & \quad \text{1636 bp} \\
1018 \text{ bp} & \quad \text{1018 bp} \\
506 \text{ bp} & \quad \text{506 bp}
\end{align*}
Fig. 11 Two clones for the secretion of cytoplasmic chitobias. 1. The PCR3 fragment is cloned into the EcoRV site of pBS II SK+. The BamHI fragment from the resultant plasmid pSKPCR3 is ligated with the larger BamHI fragment of pKK223-3, resulting in pK3PCR3. 2. Nucleotide and amino acid sequences at the juncture between signal sequences and the chitobias structural gene sequences. Arrow indicates signal peptidase cleavage.

(Fig. con’d)
ATG ATT CGA TTT AAC CTA TGT GCA GCT GGG GTT GCT TTA
Met Ile Arg Phe Asn Leu Cys Ala Ala Gly Val Ala Leu

GCG CTA TCA GGT GCT GCA GTC GCA GCT CCG ACC GCA CCA
Ala Leu Ser Gly Ala Ala Val Ala Ala Pro Thr Ala Pro

AGT ATG GAA TAT CGT GTT GAT CTC GTC CTA ...
Ser Met Glu Tyr Arg Val Asp Leu Val Val ...

**Diagram:**

- **pK3PCR3**
  - 6983 bp
  - **ori**
  - **chb**
  - **Amp**
  - **T**
  - **5S**
  - **BamHI**
  - **ScaI**

- **pSKPCR3**
  - 5608 bp
  - **ori**
  - **chb**
  - **Amp**
  - **BamHI**
  - **ScaI**
  - **NaeI**
  - **trc**
  - **fl**

- **BamHI** is marked in both circles.
Fig. 12 Secretion and protein leakage. *E. coli* JM101 cells transformed with pK3PCR3 and PC120 were incubated in LB medium at 37°C for 18 hours. Cultures were spun for 3 minutes and 100 μl of the supernatants was added to a cuvette containing 880 μl TE buffer, pH 8.0 and 20 μl p-NP-GlcNAc. Reactions were allowed for 20 minutes in a spectrophotometer (RT) and the release of p-nitrophenol (A420) was measured at time intervals indicated. The spectrophotometer was calibrated at 420 nm with pK3PCR3 supernatant reaction mixture without p-NP-GlcNAc. Inocula used were overnight colonies.
TABLE 1  Specificity of recombinant chitobiase. The recombinant chitobiase from the supernatant fraction of an overnight culture harboring pK3PCR3 was precipitated with 50% (NH₄)₂SO₄ (w/v) and incubated with the compounds listed. Activity towards each of the p-nitrophenyl derivatives was measured as the release of p-nitrophenol (A₄₂₀). Reaction conditions (in 0.5 ml): 50 mM phosphate buffer, pH 7.0, 50 μM p-NP sugar substrates, 37°C, 30 min. Reactions were promptly stopped by adding 2 mM Na₂CO₃. The data for wild type chitobiase purified from V. parahaemolyticus were previously determined (278). nd: not determined.

<table>
<thead>
<tr>
<th>Substrates</th>
<th>Chitobiase Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Secreted</td>
</tr>
<tr>
<td>p-NP-β-N-GlcNAc</td>
<td>100</td>
</tr>
<tr>
<td>p-NP-α-N-GlcNAc</td>
<td>0</td>
</tr>
<tr>
<td>p-NP-β-N-GalNAc</td>
<td>29</td>
</tr>
<tr>
<td>p-NP-α-N-GalNAc</td>
<td>0</td>
</tr>
<tr>
<td>p-NP-β-N,N'-diacetylchitobiose</td>
<td>60</td>
</tr>
<tr>
<td>p-NP-β-Glu</td>
<td>0</td>
</tr>
<tr>
<td>p-NP-α-Glu</td>
<td>0</td>
</tr>
<tr>
<td>p-NP-β-Man</td>
<td>0</td>
</tr>
<tr>
<td>p-NP-α-Man</td>
<td>0</td>
</tr>
<tr>
<td>p-NP-β-Gal</td>
<td>0</td>
</tr>
<tr>
<td>p-NP-α-Gal</td>
<td>0</td>
</tr>
<tr>
<td>p-NP-β-N-acetyl-1-thio-GlcN</td>
<td>14</td>
</tr>
</tbody>
</table>

designated pSKPCR3. The orientation of the insert in pSKPCR3 was determined by KpnI/SacI digestion followed by agarose gel electrophoresis, and is indicated by arrow in Fig. 11.

Digestion of pSKPCR3 with BamHI results in the release of PCR3 and the fragment is used to replace the small BamHI fragment of pKK223-3. This results in pK3PCR3. The chitobiase activity assay and the insert orientation determination used were the same as those for the pSKPCR3. Although pKK223-3 was used as a vector, the promoter used in pK3PCR3
for the chitobiase gene is *trc* instead of *tac*. Both *tac* and *trc* have similar promoter strength (2, 3, 33, 163).

For both pSKPCR3 and pK3PCR3, the DNA and amino acid sequence at the juncture between signal sequence and the chitobiase structural gene are the same, as depicted in Fig. 11-2. The six amino acids (APTAPS) at the N-terminus do not seem to have any effect on the enzyme, since the secreted recombinant chitobiase specificity towards several *p*-nitrophenyl derivatives (Table I) is the same compared with the wild type chitobiase purified from *V. parahaemolyticus* (278).

### 3.3.3 Secretion and Leakage

To rule out the possibility that the chitobiase in the supernatant is due to cell lysis instead of secretion, the plasmid PC120 was used to transform JM101 and the resultant clone was incubated in parallel with pK3PCR3 for 18 hr. at 37°C. Chitobiase activities from the supernatants were assayed using *p*-NP-GlcNAc. Absorbance at 420 nm (A420) was measured and the results are shown in Fig. 12. No chitobiase activity was detected in 20 minutes from the supernatant of PC120 clone. This was expected since no signal sequence is present at the N-terminus. It is clear that cell lysis, which is possible and will result in protein leakage, does not contribute to the detection of chitobiase in the supernatant, and the appearance of chitobiase in the supernatant of pK3PCR3 is due to the function of signal peptide. This is also the evidence that the chitinase gene product in the supernatant is a result of secretion.
3.3.4 Construction of Secretion Vectors

Although the secreted chitobiase using either pSKPCR3 and pK3PCR3 does not seem to be different in function from that purified from V. para-haemolyticus, it is still desirable to collect the wild type enzyme from the supernatant. The six amino acids at the N-terminus of the chitobiase can be removed using PCR. It is also possible to construct expression vectors that could secrete wild type chitobiase, which have the added advantage of secrete proteins from other sources. The signal peptidase cleavage site between the two Ala residues (Fig. 13-1) can be mutated to create a Nae I site in the DNA without changing the amino acid sequence. Digestion of the plasmid thus exposes the 3' blunt end of the signal sequence that can be used to fuse in frame the incoming structural gene PCR product, provided that the Nae I site thus created is unique on the vector. Another possibility is the existence of a potential Sfi I site near the 3' end of the signal sequence. Silent mutations can be introduced such that the ORF is maintained. DNA sequence from the Sfi I site to the 3' end of the signal sequence, inclusive, could be added to the 5' end of the upper primer so that the PCR product of the structural gene to be cloned could be tagged with Sfi I site, which could be used to fuse to the signal peptide through ligation (Fig. 13-2).

To eliminate other Nae I sites on the plasmid, pKK223-3 was digested with Nae I and Pvu II, both blunt end cutters, and the large fragment is self-ligated, resulting in a modified version of the vector, pK3-NP (Fig. 14-1, a). The BamHI fragment from pK3PCR3 was then used to replace the

---

2 Materials in sections 3.3.4 and 3.3.5 have been submitted to the Office of Technology Transfer, Louisiana State University and A&M College for disclosure to the university patent attorneys.
small BamHI fragment of pK3-NP (Fig. 14-1, b). The resulting plasmid is the same as pK3PCR3 except that sequence between Nae I and Pvu II in pKK223-3 has been removed. The lower primer used for pVerSec was tagged with Nae I. Self-ligation of the PCR product results in pVerSec with a unique Nae I site which, upon digestion, will expose the 3' end of the signal sequence for in frame fusion to the structural gene of interest Fig. 14-1, c). Since the structural gene is promoterless, only half of the ligation products could be functional. By using a different lower primer tagged with Sfi I site, the same was done using pK3NPCR3 as template to produce pVerSec-SE. The upper primer for pVerSec-SE was also tagged with an Eco RV site (Fig. 14, d), so that the ligation of the structural gene into the vector is unidirectional. Detailed DNA and the deduced amino acid sequences around the structural gene insertion sites are shown in Fig. 14-2.

Both vectors have been used to express and secrete the wild type chitobiase. The lower primer used was that used in the amplification of PCR2. For cloning into pVerSec, the 5' end of the upper primer starts from the translation initiator methionine codon, ATG. Ligation of the PCR product results in pVS-CHB (Fig. 15-1). For cloning into pVerSec-SE, a 10 base sequence was attached to the 5' end of the upper primer. The PCR product was trimmed with T4 DNA polymerase followed by digestion with Sfi I. Ligation of this processed PCR product gives rise to pVSSE-CHB (15-2). The juncture amino acid sequences between the signal peptide and the structural gene are the same using either of the two vectors.
Figure 13 Silent mutagenesis. 1. pK3PCR3 with juncture sequence in which mutagenesis is to be made. The 4 Nae I sites are indicated by stars (*). 2. Wild type and mutated forms of signal sequences. Arrows indicate lower primers for secretion vector constructions. Tagged restriction sites within the primers are also indicated and their recognition sequences underlined. Lowercase bases were mutated.
Fig. 14 Construction of pVerSec and pVerSec-SE. 1. The four Nae I sites were removed from pKK223-3 by cutting with Nae I and Pvu II and self-ligation of the large fragment (a). The small Bam HI fragment of pK3-NP is replaced by PCR3 (from pSKPCR3) (b). The fragment encompassing ori were amplified, and the products self-ligated, resulting in pVerSec (c). 2. pVerSec-SE construction by PCR using primers tagged with Sfi I and EcoRV. 3. Junction sequences in pVerSec and pVerSec-SE. lowercase indicate base substitution.

(Fig. con'd)
2

**pK3NPCR3**
5318 bp

Amp

BamHI

SacI

BamHI

SfiI

EcoRV

3

**pVerSec**
2975 bp

Amp

BamHI

NdeI

Trc

ΔS

ATG ATT CGA TTT AAC CTA
Met Ile Arg Phe Asn Leu

TGT GCA GCT GGG GTT GCT
Cys Ala Ala Gly Val Ala

TTA GCG CTA TCA GGT GCT
Leu Ala Leu Ser Gly Ala

GAT AAATCAGAAC ...

**pVerSec-SE**
2990 bp

Amp

BamHI

NdeI

Trc

ΔS

ATG ATT CGA TTT AAC CTA
Met Ile Arg Phe Asn Leu

TGT GCA GCT GGG GTT GCT
Cys Ala Ala Gly Val Ala

TTA GCG CTA TCA GGT GCT
Leu Ala Leu Ser Gly Ala

SfiI

GCG CCT GCT CCG ATG
Val Ala Ala Pro Met

EcoRV

AGA GAA GAT aTc CAG ...
Arg Glu Asp Ile Gln ...
Fig. 15 Application of secretion vectors. 1. Cloning into pVerSec. Insertion of amplified chitobiase structural gene into the Nae I site results in pVS-CHB. 2. Cloning into pVerSec-SE. The amplified chitobiase structural gene was digested with Sfi I and ligated into pVerSec-SE.
Fig. 16 Effects of EDTA on protein protection and host growth. A. Activity of chitobiase from the supernatant. B. Activity of chitobiase from the bacterial cell pellet washed with, and then suspended in, LB medium prior to assay. Overnight cultures supplemented with known concentrations of EDTA at inoculation were centrifuged in an Eppendorf tube for 5 min., and 20 μl of 5 mM p-NP-beta-GlcNAc solution was added to 200 μl each of the supernatant fractions. Reactions were proceeded at 25°C for 5 min., after which the cell pellets were suspended in 200 μl LB media and treated the same as the supernatant fractions (except that incubations were increased to 10 min.). Residual chitobiase in all reactions was inactivated by heating at 65°C for 10 min. C. Growth of E. coli carrying the recombinant chitobiase gene in the presence of EDTA. 200 μl of overnight cultures from corresponding tubes were diluted in 800 μl LB medium and turbidity was measured as A600. Optical densities of E. coli culture were adjusted based on Toennies and Gallant (229).
Fig. 17 Chitobiase activity and EDTA concentration. Five microliters of chitobiase concentrated from the supernatant with (NH$_4$)$_2$SO$_4$ was added to 995 µl of water containing 100 µM p-NP-beta-GlcNAc and 0.0 mM to 1.0 mM EDTA. Reactions were stopped at 5 min. by heating at 65°C for 5 min.
3.3.5 Protection of Secreted Proteins

Colonies of JM101 harboring pK3PCR3 grown overnight (18 hours) in LB medium showed chitobiase activity in the supernatant fraction (Fig. 12) but subsequent cultures using previous overnight culture as inoculum lead to dramatic declines in active chitobiase in the supernatant. For each of the colonies tested, no chitobiase activity could be detected after three sequential overnight cultures using immediate previous culture as inoculum. Incubation longer than overnight resulted in a decrease in chitobiase activity regardless of inoculum used (data not shown). Secreted chitobiase was probably being degraded during stationary growth of the culture since a plateau of chitobiase activity in the supernatant fraction would be expected otherwise, even when the expression and secretion terminate.

When an appropriate concentration of EDTA was added into the medium, overnight cultures inoculated with the same culture showed similar chitobiase activity as that inoculated with a colony, suggesting a protective effect. LB medium was supplemented with EDTA from 0.0 mM to 1.0 mM and inoculated with culture that no longer showed detectable chitobiase activity. Cells from the overnight cultures were pelleted and 200 µl supernatant fraction from each assayed for p-NP-beta-GlcNAc hydrolysis at 100 µM. The reaction mixtures were incubated for 5 minutes and 800 µl of water were added after reactions were stopped by incubation for 10 minutes at 65°C. As shown in Figure 16, optimal EDTA concentration was 0.6 mM. Less than 0.3 mM EDTA did not protect, and high concentrations of EDTA inhibited cell growth, and thus the production and secretion of the enzyme. No chitobiase activity was detected at 1.0
mM EDTA. Unexpectedly, chitobiase from the cell pellets also showed a similar response to EDTA under the same assay condition (except that the incubation time was 10 minutes instead of 5 minutes) (Fig. 16), implying that chitobiase degradation is not limited to the results of host cell lysis.

The decline of chitobiase activity above 0.6 mM of EDTA is not from inhibition of the enzyme which tolerates EDTA up to 1.0 mM (Fig. 17). Higher levels of EDTA affect the level of secreted enzyme by inhibiting growth of the host bacteria (Fig. 16). EDTA levels of 0.4 mM to 0.8 mM inhibited bacterial growth by ca. 25% (Fig. 16). Above 0.8 mM of EDTA dramatically inhibited bacterial growth, with complete inhibition at 1.0 mM.

EDTA has a slight inhibitory effect on *E. coli* growth at moderate concentrations (0.4 mM–0.8 mM). At 0.6 mM of EDTA during *E. coli* growth, secreted chitobiase was best protected probably due to balance between cell growth and proteolysis protection. The observation that stability of the enzyme retained in the cells has a similar response to EDTA in the culture medium implies that degradation of chitobiase is not host cell lysis dependent. The culprit is probably a metalloprotease released by *E. coli*. This observation may prove generally useful for protection of cloned secreted proteins in the supernatant fraction of *E. coli* and other bacterial cultures.

3.4 DISCUSSION

A stretch of six amino acids (-Ala-Pro-Thr-Ala-Pro-Ser-), located at the extreme N-terminus of mature chitinase, is noteworthy. The peculiar arrangement of two potential turn-making residues in this short peptide suggested a possible function for the signal peptidase. The first Ala
residue is where proteolysis occurs. It is for this reason that some clones for the secretion of cytoplasmic chitobiase retained these residues, and the secretion product is thus a fusion protein containing them. Clones with or without these residues, however, did not show any differences in protein secretion. In addition, these residues attached to the chitobiase do not have detectable effects on the function of the protein. These amino acid residues are therefore ordinary components of the mature chitinase and have no contribution to the nature of the signal peptide immediately preceding it.

Several protein secretion vectors have been reported to secrete fusion proteins from organisms such as *E. coli* yeast, and higher eukaryotic cells. In *E. coli*, as all Gram negative bacteria, the existence of two membranes (outer and inner membranes) seems to have complicated the secretion process. Collection of gene products often requires disruption of the cell wall, which makes subsequent purification difficult. Genes to be cloned need to have unique restriction sites at certain locations, which is not always available. In addition, proteins obtained this way are often recombinants, with additional amino acids attached to the final products, which are obviously not desirable. For example, gene products directed by the *malE* signal peptide (on the pMAL-p2 vector) target to the periplasm of *E. coli* (84, 141, 224). Periplasmic extract prepared by osmotic shock is used to purify fusion proteins. Attached peptides to the proteins of interest are then cleaved by Factor Xa at a specific site (127, 168, 169), leaving some residues at the amino termini of the resultant protein, depending on restriction sites used during cloning. In addition, the quality of proteins also depend on other factors, including the efficiency of Factor Xa
cleavage and the presence of additional Factor Xa cleavage sites in the mature proteins. The secretion vectors reported here do not require restriction site nor nucleotide sequence information except sequences flanking the structural gene for primer binding. These vectors also have the added advantage of producing wild type proteins independent of the restriction sites on the gene to be cloned.

As reviewed in Chapter 1, the placement of a strong transcription termination signal distal to a strong promoter is believed to be necessary. The evidence against this hypothesis is provided in this study. The PCR fragment, PCR3, is an intact transcription unit with a signal sequence placed in frame with the chitobiase ORF. It contains the trc promoter and the SD sequence of *E. coli*. A terminator-like stretch of sequence down stream of the chitobiase gene translation stop codons was not present (Chapter 2). Further more, the lower primer for the chitobiase structural gene amplification is in the stop codon region, ruling out the possibility that a unidentified termination signal could be carried over with the PCR product.

Insertion of PCR3 into pBS SK+ results in pSKPCR3. The fragment was then excised from pSKPCR3 with *Bam*H I, and ligated into pKK223-3, resulting in pK3PCR3. The level of expression and secretion of chitobiase from the two clones, pSKPCR3 and pK3PCR3, are similar. It is unlikely that sequences on pBS SK+ could serves as a transcription terminator of the chitobiase gene. The orientation of the chitobiase gene transcription unit is opposite to the *lacZ'* gene. The stop codons of chitobiase gene is very close (<500 bp) to the origin of replication of the plasmid (ori). No terminator-like sequence was evident in this region. The *E. coli* strains used
may play a role but is unlikely in this system since DH5αF' does not support neither clones whereas JM101 supports both.

A by-product of this project is the observation that plasmids containing the signal sequence followed by a ORF can not survive in DH5αF' but replicate and express faithfully in JM101. Less than 1.0 ng of miniprepped plasmid DNA pSKPCR3 or pK3PCR3 were routinely used to successfully transform JM101 competent cells whereas no colonies of DH5αF' were observed with up to 1.0 µg of the same DNA samples of either plasmid. When the chitobiase ORF is removed, the plasmid can survive in both DH5αF' and JM101. Examples of such plasmids from this study include pVerSec and pVerSec-SE and several others. As a matter of fact, only DH5αF' was used for large preparation of plasmid DNA to eliminate contamination by plasmid with the chitobiase ORF. Another reason that DH5αF' is preferred to JM101 is that it grows slower than JM101, and thus the plasmid DNA is relatively cleaner. All single-strand DNAs for sequencing were prepared from DH5αF' except for secretion clones that can only transform JM101.

Both strains of \( E. \ coli \) (DH5αF' and JM101) are commercial products and their genotypes are known from various sources (see Appendices). However, these laboratory \( E. \ coli \) strains, as well as others, are the results of extensive mutagenesis. Different strains may carry different, so far undiscovered, mutations that have unknown effects to a particular system. The secretion plasmids in this study seem to have unfortunately fallen into this situation. The known genotypes (mutations) of these strains include, among others, plasmid protection (\( e14(McrA^-, McrBC^-) \), \( mrr, hsdRS, recABCFJ, etc. \)), colony identification (\( \Delta lacZ, \phi 80(lacM15,\)
etc.), strain maintenance (proAB, etc.) and bacteriophage infection for ss-DNA production (F', carrying the pili genes). These genotypes are obviously useful for general cloning purposes but less so for analysis in gene expression and secretion.

The observation that the secretion plasmids can survive in JM101 but not in DH5αF' has been very advantageous in this project, regardless of the underlying mechanism. During pVerSec and pVersec-SE developments, there are several steps that require amplification of vector portion followed by self-ligation of the PCR product. High fidelity amplification demands lower cycle number whereas high yield PCR product requires larger amount of template. Since the secretion vectors are the results of stepwise chitobiase structural gene removal and insertion, prevention of template-carry-over is critical if sequencing of each intermediate is to be avoided. One way to alleviate the problem is to gel-purify PCR products before ligations. This did not work well because the PCR products co-migrate with the supercoiled form of the template plasmid DNA. Although restriction enzyme digestion before gel electrophoresis would solve the problem, PCR products need to be purified before digestion. The alternative solution, which has been used in this study, is to transform DH5αF' with the ligation mixture. Since the template DNA can not survive in this strain, PCR product can be ligated and directly used for transformation. DH5αF' is also used to propagate these vectors to eliminate contamination. Considerable time and effort were saved by using this E. coli strain.

3.5 SUMMARY
The signal peptide of an extracellular endochitinase from the marine bacterium Vibrio parahaemolyticus causes the mature chitinase to be
efficiently secreted through both membranes of its Gram negative bacterial host, and of *Escherichia coli* JM101 when cloned. By using recombinant PCR, this signal sequence, including the N-terminal six amino acids of the mature chitinase, was fused in frame to the gene encoding a cytoplasmic chitobiase [Zhu, *et al.*, 1992, *J. Biochem. (Tokyo)*, 112(1):163-167]. The construct caused the cytoplasmic chitobiase secreted efficiently into the *E. coli* culture medium, suggesting that this signal sequence could be useful for extracellular production of other proteins.

To facilitate ease of cloning, two secretion vectors were developed. PCR fragment of a structural gene can be inserted into the vectors such that it is in frame with the signal sequence. The secretion system has been tested using the cytoplasmic chitobiase gene as a model system for the secretion of wild type proteins. Secreted chitobiase was assayed both on the transformation plate and in *E. coli* culture supernatant. The signal sequence on the two vectors, together with the incoming structural genes when fused, is driven by the strong *trc* promoter. PCR fragment of the structural gene of interest can be used directly, or following simple restriction enzyme digestion, for ligation into the vectors for secretion. Restriction sites and complete DNA sequence are not required for the structural gene to be cloned. EDTA was shown to have a protective effect on cloned proteins secreted into the medium of host cell *E. coli* JM101. The protective concentration was rather narrow at 0.5-0.7 mM EDTA due to inefficiency at lower concentrations and cell growth inhibition at higher levels. EDTA was also shown effective in protecting proteins inside *E. coli* cells, implying that cytoplasmic proteins cloned for secretion have limited stability in the cytoplasm. This observation may
prove generally useful for protecting proteins cloned for secretion in *E. coli* and other bacterial cultures.
CHAPTER 4 CONCLUSIONS

The 85 kDa cytoplasmic N,N'-diacetylchitobiase of restricted specificity participates in the high level utilization of chitin-derived 2-deoxy-2-ace-tamido-\(\text{D}\)-glucose (GlcNAc) by vibrios as one of two parallel pathways for metabolism of N,N'-diacetylchitobiose. This chitin utilization system contains many enzymes and chemotactic proteins. The system is obviously complex in general, and is probably so in *Vibrio parahaemolyticus*. While this study has answered some questions, many remain for further studies.

The nucleotide sequence and the predicted amino acid data revealed that the cytoplasmic N,N'-diacetylchitobiase appears to be a unique protein, lacking a signal sequence and genetically distant from other known chitinoclastic beta-N-diacyltel-hexosaminidases, consistent with its limited substrate specificity to small GlcNAc terminated oligosaccharides. Sequence analysis and comparison with other published sequences also implied that Arg271 (Fig. 8-2) of the enzyme may be involved in catalysis (36).

Of pivotal importance is the cytoplasmic nature of chitobiase as one of the major catalytic proteins in the chitinoclastic pathway. In particular, the involvement of a chitoperm ease is based on the properties of cytoplasmic chitobiase. These two proteins together determine which one of the two pathways is involved in chitin degradation by a particular organism. The fusion of chitinase signal peptide to the chitobiase coding sequence provided several lines of evidence. (i) Consistent with the general belief, it is now clear that the chitinase signal sequence is required and sufficient to target the mature chitinase to the external milieu of *E. coli*, a Gram
negative bacterium. The mature polypeptide does not seem to have any function in this regard. (ii) Although the DNA sequence and the physical location of chitinase strongly argue that the enzyme is secreted, it was not possible to rule out the possibility that the presence of the enzyme in the culture supernatant is a consequence of cell lysis. The secretion of the cytoplasmic chitobiase by this signal sequence provided evidence that cell lysis does not contribute to the presence of enzymes in the supernatant, since chitobiase activity was not detected without a signal peptide. (iii) The secretion of chitobiase by fusion to the signal peptide, together with the lack of a signal peptide as revealed from the sequence data of chitobiase, also proved that the chitobiase is cytoplasmic. The hydrophobicity analysis demonstrated that the polypeptide does not interact with the membrane. All of these points to the notion that the enzyme is not involved in any reaction in the periplasm. This is important because the presence of a chitopermase is necessary to translocate the substrate of chitobiase to the cytoplasm across the cytoplasmic membrane.

The artificial secretion of cytoplasmic chitobiase also suggests that the chitinase signal sequence can be used to produce other proteins extracellularly. This idea has been translated into the construction of secretion vectors during this study. While there is no way to predict the secretability of a particular protein, the same is true for other systems. The advantages of this system are: (i) Easy to use for cloning; (ii) Facile collection of gene products; (iii) Does not require restriction site and sequence information except for the partial sequences at both ends of the coding sequence of interest; (iv) The secreted gene products are of wild type instead of fusion proteins. Of course, the unfortunate lack of glycosyltransferase
systems in *E. coli* puts a barrier to the secretion of functional glycoproteins.

EDTA was shown to have a protective effect on cloned proteins secreted into the medium of host cell *E. coli* JM101. The protective concentration was rather narrow at 0.5-0.7 mM EDTA due to inefficiency at lower concentrations and cell growth inhibition at higher levels. This observation may prove generally useful for protecting proteins cloned for secretion in *E. coli* and other bacterial cultures.
REFERENCES


mapping of recombinant plasmid containing the \textit{rrnB} ribosomal RNA operon of \textit{E. coli}. \textit{Plasmid}. \textbf{6}, 112-118.


amino acid sequence: cloning of urate oxidase. Science. 239, 1288-1291.


peptide from the small subunit of ribulose 1,5-bisphosphate carboxylase. *Nature.* 313, 358-363.


APPENDIXES

A.1 SEQUENCING AND PCR PRIMERS

A.1.1 Sequencing Primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUC-FOR-P</td>
<td>5'-CGC CAG GGT TTT CCC AGT CAC GAC-3'</td>
</tr>
<tr>
<td>pUC-REV-P</td>
<td>5'-TCA CAC AGG AAA CAG CTA TGA C-3'</td>
</tr>
<tr>
<td>M13-FOR-40</td>
<td>5'-GTT TTC CCA GTC AGC AC-3'</td>
</tr>
<tr>
<td>KSP21-P1</td>
<td>5'-TGG CAG GTA AAC CTG AAC-3'</td>
</tr>
<tr>
<td>KSP21-P2</td>
<td>5'-TTC AAT GAA GCC TGT GCT-3'</td>
</tr>
<tr>
<td>KSP21-P3</td>
<td>5'-GCC AAT TAA GCT AAA GAG-3'</td>
</tr>
<tr>
<td>SKS161-P1</td>
<td>5'-CAG GGT GGT GGT GAT AAT-3'</td>
</tr>
<tr>
<td>SKS21-P1</td>
<td>5'-TCT GCC AGT CTC GGT GTT-3'</td>
</tr>
<tr>
<td>SKS21-P2</td>
<td>5'-CTT TTT CTA GCC GCA ATG-3'</td>
</tr>
<tr>
<td>SKS21-P3</td>
<td>5'-GGG AAG AAC GCC AGA CTG-3'</td>
</tr>
<tr>
<td>SKS21-P4</td>
<td>5'-GAC TCA ATT CCG CAA CCT-3'</td>
</tr>
<tr>
<td>SKH35-P1</td>
<td>5'-TGG ATT TGC AGG GAG TGA-3'</td>
</tr>
<tr>
<td>KSP-1</td>
<td>5'-GGG CAG GAT TAC GGT CAA-3'</td>
</tr>
<tr>
<td>KSP-2</td>
<td>5'-CTA TCC GCC AGA ATG TAG-3'</td>
</tr>
<tr>
<td>KSP-3</td>
<td>5'-ACT GCC GCT TGT CTA TTC ATG-3'</td>
</tr>
<tr>
<td>KSP-4</td>
<td>5'-GGC TGG TGC ATC GCC AAC-3'</td>
</tr>
<tr>
<td>KSP-10</td>
<td>5'-CCA AGG AGG AGG TAC ATC-3'</td>
</tr>
<tr>
<td>KSP-11</td>
<td>5'-GCT AAT GCC GTG TTA ATG-3'</td>
</tr>
<tr>
<td>KSP-12</td>
<td>5'-GAC CGG GAT TGC GAC TAA-3'</td>
</tr>
<tr>
<td>SFP-5</td>
<td>5'-TGC TCC CTA TTT TCA TTC CTC-3'</td>
</tr>
<tr>
<td>SFP-6</td>
<td>5'-GGT GCT TGC CTT TGC TTC TGC-3'</td>
</tr>
<tr>
<td>SFP-7</td>
<td>5'-GGA ACA ACT GGG TTA CAG-3'</td>
</tr>
<tr>
<td>SFP-8</td>
<td>5'-GGC CAA GGT TGC GAT GTG-3'</td>
</tr>
<tr>
<td>SFP-9</td>
<td>5'-AAG GCA TTT GGG GCA TTC-3'</td>
</tr>
<tr>
<td>CHNSIG-F</td>
<td>5'-CCT ATG TGC AGC TGG GGT TGC TTC ATG-3'</td>
</tr>
</tbody>
</table>

A.1.2 PCR Primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR-PU1</td>
<td>5'-GGA GAT CCA TAC TCT GCC GCA TTC-3'</td>
</tr>
<tr>
<td>PCR-PL1</td>
<td>5'-GAT ATT CCA TAC TTG GTG CGG TCG-3'</td>
</tr>
<tr>
<td>PCR2-PU1</td>
<td>5'-CGG ACC AAG TAT GGA ATA TCG TGT TGA TCT TCA-3'</td>
</tr>
<tr>
<td>PCR2-PL1</td>
<td>5'-GAT TGA CCG TAA TCT CCC GTC TCA TCA-3'</td>
</tr>
<tr>
<td>pVS-pU</td>
<td>5'-GGC TGA TAC AGA TTA AAT CAG CAA-3'</td>
</tr>
<tr>
<td>pVS-pl2</td>
<td>5'-GTC GGG CCG CCG ACT GCA GCA CTA CTA-3'</td>
</tr>
<tr>
<td>pVS-SE-U</td>
<td>5'-ATG AGA GAA CAT ATC CAG CTC GAT ACA TAC-3'</td>
</tr>
<tr>
<td>pVS-SE-L</td>
<td>5'-CGG AGA GCC CAC TGC GGC CCC TGA TAG CTC-3'</td>
</tr>
<tr>
<td>PCR4-pU</td>
<td>5'-ATG GAA TAT CTT GCT GAT CTC-3'</td>
</tr>
<tr>
<td>PCR5-pU</td>
<td>5'-GGG GCC GCA GTG GCC ATG GAA TAT CTT GGT GAT C-3'</td>
</tr>
<tr>
<td>PBGAL-U</td>
<td>5'-GGG GCC GCA GTG GCC ATG AGC GAA AAT TAC ATC GTC AC-3'</td>
</tr>
<tr>
<td>PBGAL-L</td>
<td>5'-CGA AAT ACG GGC AGA CAT GG-3'</td>
</tr>
</tbody>
</table>

3 Not all primers in the chitobiase gene region perfectly match the template since they were based on primary gel data. Some of the primers listed are commercial products.
A.2 SIGNAL SEQUENCES USED IN THIS STUDY

A.2.1 Wild Type Signal Sequence

ATG ATT CGA TTT AAC CTA TGT GCA GCT GGG GTT GCT TTA GCG CTA TCA GGT
M I R F N L C A A G V A L A L S G

GCT GCA GTC GCA GCT CCG ACC GCA CCA AGT ...
A A V A A P T A P S ...

A.2.2 Signal Sequence Used in Secretion Vectors

pVerSec:

ATG ATT CGA TTT AAC CTA TGT GCA GCT GGG GTT GCT TTA GCG CTA TCA GGT
M I R F N L C A A G V A L A L S G

GCT GCA GTC GCc
A A V A

pVerSec-SE:

ATG ATT CGA TTT AAC CTA TGT GCA GCT GGG GTT GCT TTA GCG CTA TCA GGG
M I R F N L C A A G V A L A L S G

GCc GCA GTC GCC
A A V A

---

4 Underlined is the N-terminal sequence of the mature chitinase (123). The last Ala not underlined donates the carboxy group of the signal peptide.

5 Underlined is the minimal add-on to the upper primer of structural genes to be amplified. Lowercase bases in A.22 and A.23 represent silent mutations.
A.3 SECRETION VECTOR SEQUENCES
pVerSec (2975 bp):

```
GGCCACGGATG CGTCGGCGGT AGAGGATCCG GAATTCTCAT GTTTGACAGC TTATCATCGA
CTGCACGGTG CACCAGCTTG TCTGGCGTCA GGCAGCCATC GGAAGCTGTG GATTGGCTGT
GCAGGTCGTA AATCACTGCA TAATTCGTGT CCGCTCAAGGC GCACTCCCGT TCTGGATAAT
GTTTTTTGCG CCGACATATA AACGGTTCTG GCAAATATTC TGAAATGAGC TGTTGACAAT
TAATCATCCG GCTCGTATAA TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA
CAGACCATCG ATCCGGCGGT CGCCGCGCCC ACGGCTGATA CAGATTAAAT
```

```
GAGACCATCG ATCCGGCGGT CGCCGCGCCC ACGGCTGATA CAGATTAAAT
```

```
GAGACCATCG ATCCGGCGGT CGCCGCGCCC ACGGCTGATA CAGATTAAAT
```

```
CCCAGAAACG CTGGTGAAAG TAAAAGATGC TGAAGATCAG TTGGGTGCAC GAGTGGGTTA
CATTGGTAAC TGTCAGACCA AGTTTACTCA TATATACTTT AGATTGATTT AAAACTTCAT
```

```
GCCGCAGGAG CGGAGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC
```

```
GCCGCAGGAG CGGAGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC
```

```
GCCGCAGGAG CGGAGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC
```

```
GCCGCAGGAG CGGAGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC
```

```
GCCGCAGGAG CGGAGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC
```

```
GCCGCAGGAG CGGAGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC
```

```
GCCGCAGGAG CGGAGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC
```

```
GCCGCAGGAG CGGAGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC
```

```
GCCGCAGGAG CGGAGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC
```

```
GCCGCAGGAG CGGAGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC
```

```
GCCGCAGGAG CGGAGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC
```

```
GCCGCAGGAG CGGAGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC
```

```
GCCGCAGGAG CGGAGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC
```

```
GCCGCAGGAG CGGAGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC
```

```
GCCGCAGGAG CGGAGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC
```

```
GCCGCAGGAG CGGAGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA CGCCAGCAAC
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
GCCAGTGGCG AAAAAACCC TACATACCTC GCTCTGCTAA TCCTGTTACC AGTGGCTGCT
```

```
AAGAACTCTG TAGACCTTCA GCACTGCGGA CGCTGTCGCT CGGGAGCTGC ATGTGTCAGA
GGTTTTCACC GTCATCACCG AAACGCGCGA GCCAAG
```

```
129
```
pVerSec-SE (2990 bp):

GGCCACGATG CTGCCCGGCCT AGAGGAATTG GATTTGACAGC TTATCATCGA 60
CTGCCAGCTG CACCAATGCT TCTGGCGTCA GGCAGCCATC GGAAGCTGTG GATTGGCTGT 120
GACGATGCTTA AATCCTGCTAT GTAACCGCTG CTGCAACGGC CGACCTCGCTG TCTGGAATAAT 180
GAAAAGCGTG CCGACTATAA AAGGTTCTGC GCAAAATTTC TGAGATGAGC TGGTAGCAAT 240
TAAATCTCCG GCTCAGTATAA AACTGTCCGT CTTATGGCTG TATTTACCATG TAAACCTGCA 300
AGAGCTACTG CTTAGAGATAA TATGAAATGG TATGAAATGG TATGAAATGG TATGAAATGG 360
GAGTTGCTTTA GGTGTACATCG AACTGGATCT CAACAGCGGT AAGATCCTTG AGAGTTTTCG 420
CTGCAGAATAA AATACCGCTG TTTTTTTGCG CCGACATATA AACGGTTCTG GCAAATATTC TGAAATGAGC TGTTGACAAT 480
GCAGGTCGTA AATCACTGCA TAATTCGTGT CGCTCAAGGC GCACTCCCGT TCTGGATAAT 540
GTTTTTTGCG CCGACATATA AACGGTTCTG GCAAATATTC TGAAATGAGC TGTTGACAAT 600
TAATCATCCG GCTCGTATAA TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA 660
CTGAGGNGAA CAAATTCCTGC GGCAGCCCATC GGAAGCTGTG GATTGGCTGT 720
TGGCGGCGAG GACGCCGGCC ATAAACTGCC AGGCATCAAA TTAAGCAGAA GGCCATCCTG 780
ACCGGATGCG GCTGCCCGCC ATAAACTGCC AGGCATCAAA TTAAGCAGAA GGCCATCCTG 840
ATTGTAATCG CTCATGAGAC AATAACCCTG ATAAATGCTT CAATAATATT GAAAAAGGAA 900
GATCGTACG ATTCACACTT CCTGCTGCGC CCTTATCCCG TTTTATTCGG CATTTTGCCT 960
TCCCTTTTTT TCTCAGAGAA AAGCCTGGTG GAAATGAAAA GATCTGCAAG ATCATGGTG 1020
TGACAGACGG GGTATGACCTG AACTGGATCT CAACAGCGGT AAGATCCTTG AGAGTTTTCG 1080
CCCGGAAAG TCTGTTTGAG CCGACACTAT GGTGGCGGCG TTTTATTCGG CATTTTGCCT 1140
ATCGGAGCGG GAGGAGCTGGG GGAGCCGAGG TTTTATTCGG CATTTTGCCT 1200
CTGGTGGTAG TACTCAGAGC TGCAAGAAA GCACTTACGG GATGGCCAGA CAGTACAGAA 1260
APGATATCGT GCTGCTGATGA AAACTTTTGG TTATATTGAT CGAGTCAGCC AAGGTGCTGG 1320
GATCGTACG ATTCACACTT CCTGCTGCGC CCTTATCCCG TTTTATTCGG CATTTTGCCT 1380
CCTTGATGCT GGGAGCAGCG ACTGTAAGTG ACGCTAACC AAGCGAGCG GTCGACACAC 1440
TGCTGATAG GAAATTGACG AAACTTTTGG TTATATTGAT CGAGTCAGCC AAGGTGCTGG 1500
GCTTCCGCCG TTTATTCGG CATTTTGCCT 1560
TCTGCTGATGA AAACTTTTGG TTATATTGAT CGAGTCAGCC AAGGTGCTGG 1620
TCTGCTGATGA AAACTTTTGG TTATATTGAT CGAGTCAGCC AAGGTGCTGG 1680
TACAGCAGCG GAGGAAAGCG TTTTATTCGG CATTTTGCCT 1740
GCCCTGCTGCTG TACAAGCTATG TACAGCTAGT CAACACTTCC AGTCTGCTGG 1800
GATTTAAAAAC TTTCTTTTTA ATTAAAAAGG ACTGTTGCTG AGATGCTTTT GATATATCTG 1860
ATGAAACAAAT TTTTATTCGG GAGGAAAGCG TTTTATTCGG CATTTTGCCT 1920
ATCAAGAAAT TTTTATTCGG GAGGAAAGCG TTTTATTCGG CATTTTGCCT 1980
AAACCGGCC TTTTATTCGG GAGGAAAGCG TTTTATTCGG CATTTTGCCT 2040
AAGGATACTG GCTCGCTGCGC AGGGAGGATA CAACATGATG CTTCTTTTGG TGATCGAGCA 2100
TGTGGCCACC TTTTATTCGG GAGGAAAGCG TTTTATTCGG CATTTTGCCT 2160
TTAACAGTTG GTXGGCAGAG CTTTTTATTC CATTTTGCCT 2220
TGAAGTCAGG TTTTATTCGG GAGGAAAGCG TTTTATTCGG CATTTTGCCT 2280
GCGAGGTCG TTTTATTCGG GAGGAAAGCG TTTTATTCGG CATTTTGCCT 2340
ACCGCTGCAGG GAAAGAAATG TTTTATTCGG GAGGAAAGCG TTTTATTCGG CATTTTGCCT 2400
GGAGGAGCTGG AAGGAAAGCG TTTTATTCGG CATTTTGCCT 2460
GCCCTCCTGCT CTTTTTATTCG GAGGAAAGCG TTTTATTCGG CATTTTGCCT 2520
AAAAAACCGT GCAACGCGAG GAGGAAAGCG TTTTATTCGG CATTTTGCCT 2580
AGGTGAGATC TTTTATTCGG GAGGAAAGCG TTTTATTCGG CATTTTGCCT 2640
GCTGTGATGCT TTTTATTCGG GAGGAAAGCG TTTTATTCGG CATTTTGCCT 2700
GAAGAGGAGT GTTTTATTCGG GAGGAAAGCG TTTTATTCGG CATTTTGCCT 2760
GAGGAGTCTG TTTTATTCGG GAGGAAAGCG TTTTATTCGG CATTTTGCCT 2820
TATGTCAGCTG TTTTATTCGG GAGGAAAGCG TTTTATTCGG CATTTTGCCT 2880
CCTGCGAGGGGT GAGGAAAGCG TTTTATTCGG GAGGAAAGCG TTTTATTCGG CATTTTGCCT 2940
GCTGCATGCTG TTTTATTCGG GAGGAAAGCG TTTTATTCGG CATTTTGCCT 3000
A.4 GENOTYPES OF SOME HOST STRAINS

JM100 Series

JM101: F' traD36 lacIΔ(lacZ)M15 proA+B+/supE thi Δ(lac-proAB) (271). Supports the growth of vectors carrying amber mutations (152).

JM103: F' traD36 lacIΔ(lacZ)M15 proA+B+/endA1 supE sbcBC thi-1 rpsL (Str') Δ(lac-pro) (P1) (r_k⁺ m_k⁺ m_p1⁺) (88, 271).

JM105: F' traD36 lacIΔ(lacZ)M15 proA+B+/endA1 sbcB15 sbcC? hsdR4 (r_k⁺ m_k⁺) Δ(lac-proAB) (271). Supports the growth of vectors carrying amber mutations (152) and will modify but not restrict transfected DNA (271).

JM107: F' traD36 lacIΔ(lacZ)M15 proA+B+/e14· (McrA') Δ(lac-proAB) thi gyrA96 (Na') endA1 rpsR17 (r_k⁺ m_k⁺) relA1 supE44 (271). Supports the growth of vectors carrying amber mutations (152) and will modify but not restrict transfected DNA (271).


Other Strains


DH5αF': F'/endA1 hsdR17 (r_k⁺ m_k⁺) supE44 thi-1 recA1 gyrA (Na') relA1 Δ(lacZYA-argF)U169 deoR (ϕ80 dlacΔ(lacZ)M15) (264).

KS1000: F'lacIΔlacI+/ara Δ(lac-pro) nalA argI(am) rif thi-1 Δ(tsp)::Kanr eda-51::Tn10 (tet'). E. coli tsp mutant. The tsp gene (tail specific protease) encodes a periplasmic protease that may degrade secreted or cytoplasmically overexpressed proteins after cell lysis (212).

6 Listed are known, but not all, genotypes. Since these strains are the results of extensive mutagenesis, phenotypes can not be predicted based on these genotypes. Additional notes on some of these strains can be found in several laboratory manuals (8, 189).
A.5 STRUCTURES OF RELEVANT COMPOUNDS

A.5.1 Monosaccharides

Glucose

Galactose

Mannose

Glucosamine

\( N\text{-Acetylglucosamine} \)

---

\(^7\) All sugars, monomeric or as components of other compounds, are beta anomers of \(D\)-sugars unless otherwise indicated.
A.5.2 Disaccharides

Lactose

1,6-Allolactose

Maltose

Cellobiose

Chitobiose
A.5.3 Polysaccharides

Cellulose

Chitin ($n=4500-8000$, $M_r=1.0-1.8\times10^6$ Da)
A.5.4 Sugar Analogs

**X-Gal** (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside)

**IPTG** (Isopropyl-beta-D-thiogalactopyranoside)

**p-NP-GlcNAc** (p-Nitrophenyl-beta-D-N-acetylgalactosamine)
α-³⁵S-dATP (³⁵S-Deoxyadenosine 5'-(alpha-thio)triphosphate); FW: 507.2; Half-life: 87.4 days; Catalog #: NEG-034H, from DuPont-NEN, Sequencing grade for use with Sequenase 2.0.
A.6 LIST OF ABBREVIATIONS

ATCC  American type culture collection
CAMP  3',5'-Cyclic adenosine monophosphate
bp  base-pair (of DNA)
CAP  Catabolite activator protein
CHO  Cell line derived from Chinese hamster ovary
C-terminal  Carboxy terminal
ds-DNA  Double stranded DNA
ER  Endoplasmic reticulum
EDTA  Ethylene diamine tetraacetic acid, disodium salt
Fru-6-P  Fructose-6-phosphate
Gal  Galactose
GalNAc  N-acetylgalactosamine
GCG  Genetics Computing Group, GCG software vendor
Glc  Glucosyl
GlcN  Glucosamine
GlcNAc  N-acetylglucosamine
GlcNAc-6-P  Glucose-6-phosphate
Glu  Glucose
IPTG  Isopropyl-beta-D-thiogalactopyranoside
kDa  Kilodalton
LB  Luria-Bertani medium
Man  Mannose
MPB  Maltose binding protein
N-terminal  Amino terminal
nt  (Deoxy)ribonucleotide (as a unit of RNA or DNA)
ORF  Open reading frame
p-NP  para-nitrophenyl
p-NP-GlcNAc  para-nitrophenyl-N-acetylglucosamine
PBS II  pBluescript II
PCR  Polymerase chain reaction
PTS  Bacterial phosphoenolpyruvate:glycose phosphotransferase system
recPCR  Recombinant polymerase chain reaction
rrnBT1T2  rRNA operon B transcriptional tandem terminator
S.D.  Shine-Dalgarno sequence (= RBS)
SDS-PAGE  Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
ss-DNA  single stranded DNA
Tris  Tris(hydroxymethyl)aminomethane
X-gal  5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

---

8 Not included are those fully defined whenever used. Names of genes (products), operons, mutants, amino acids and nucleotides follow standard notations.
VITA

Michael Hongli Wu was born on April 27, 1956 in Hebei province, People's Republic of China. He went to Beijing Forestry University in 1978, where he received his Bachelor of Science in Agriculture in 1982. After working for four years in the Forestry Institute of Shaanxi province in Northwest China, he came to the University of California at Riverside in 1986, where he worked on baculovirus genetics under professor Brian A. Federici and graduated with a Master of Science in Entomology in 1989. He started his doctoral program in the Department of Biochemistry, Louisiana State University, in the summer of 1991, under the direction of professor Roger A. Laine, and will receive the doctor of philosophy degree in the spring of 1996. His project is on DNA sequencing and protein secretion of the cytoplasmic chitobiase gene from the marine bacterium, *Vibrio parahaemolyticus*. In addition to his proposed project, he is also interested in the development of cloning vectors for the production of extracellular proteins cloned from other sources. The major techniques used during his study in professor Laine's lab include gene cloning, DNA sequencing, and recombinant polymerase chain reaction (recPCR).
Candidate: MICHAEL HONGLI WU

Major Field: BIOCHEMISTRY

Title of Dissertation: THE CYTOPLASMIC N,N'-DIACETYLCHITOBIASE GENE FROM VIBRIO PARAHAEOMLYTICUS: SEQUENCE ANALYSIS, PROTEIN SECRETION, AND SECRETION SYSTEM DEVELOPMENT

Approved:

Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

Date of Examination:

OCT. 31, 1995