CONSTRUCTION OF VECTORS FOR THE OVEREXPRESSION OF RECOMBINANT HUMAN GROWTH HORMONE IN Bacillus megaterium

NORHASHIMA ABD. RASHID

UNIVERSITI SAINS MALAYSIA

2009

CONSTRUCTION OF VECTORS FOR THE OVEREXPRESSION OF RECOMBINANT HUMAN GROWTH HORMONE IN *Bacillus megaterium*

by

NORHASHIMA ABD. RASHID

Thesis submitted in fulfilment of the requirements for the degree of Master of Science

January 2009

PEMBENTUKAN VEKTOR UNTUK PENGEKSPRESSAN LAMPAU HORMON PERTUMBUHAN MANUSIA REKOMBINAN DI DALAM *Bacillus megaterium*

oleh

NORHASHIMA ABD. RASHID

Tesis yang diserahkan untuk memenuhi keperluan bagi Ijazah

Sarjana Sains

Januari 2009

This manuscript is dedicated to :

my parents
Abd. Rashid bin Ahmad
Zainab bt Abdul

my husband
Muhamad Bustani bin Abdul Latiff
&
our beloved son
Muhammad Afif Harith bin Muhamad Bustani

ACKNOWLEDGEMENTS

I would like to thank my supervisors Associate Professor Dr. Mohd. Razip Samian for his invaluable guidance, not excluding the meticulous proofreading of this thesis. I really appreciate him.

My sincerely thanks to Prof. Nazalan Najimudin, Assoc. Prof. Tengku Sifzizul Tengku Muhammad for their kindness of letting me use their lab equipment.

My deep gratitude is extended to all my friends from Lab 414, Lab 413 and Lab 403 for their co-operation and for the use of laboratory facilities and were always there to provide constructive ideas and discussion, and made my study in Universiti Sains Malaysia such a wonderful and enjoyable one.

I wish to extend my sincere thanks to all the staff of the School of Biological Sciences office for their kind help of scientific items purchasing in and delivery problems.

The financial support under the IRPA grant provided by the Ministry of Science, Technology and Innovation (MOSTI) is gratefully acknowledge.

Last but not least, a heartfelt of appreciation and special word of thanks goes to my beloved husband, Muhamad Bustani Abdul Latiff, my parents, Abd. Rashid Ahmad and Zainab Abdul for their warmth, love, unfailing support, blessing and patience.

NORHASHIMA ABD. RASHID

January 2009

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	**ii**
TABLE OF CONTENTS	**iii**
LIST OF TABLES	**x**
LIST OF FIGURES	**xi**
LIST OF GRAPHS	**xiv**
ABBREVIATIONS	**xv**
ABSTRAK	**xviii**
ABSTRACT	**xix**
1.0 INTRODUCTION	1
1.1 Research objectives	2
2.0 LITERATURE RIVIEW	
2.1 Human Growth Hormone	
2.1.1 Historical Background	3
2.1.2 Structure of the hGH genes	4
2.1.3 hGH synthetis	10
2.1.4 Biochemical and Physiological Characteristics of hGH	12
2.1.4.1 Carbohydrate metabolism	13
2.1.4.2 Lipid metabolism	13
2.1.4.3 Mineral metabolism	13
2.1.4.4 Prolactin-like effects	14
2.1.5 Recombinant hGH	14
2.2 Techniques of synthetic gene construction	
2.2.1 Background	15
2.2.2 Gene assembly techniques	16
2.2.3 In-frame cloning methods	18

2.2.4 Solid phase gene assembly 18

2.2.5 Chemical assembly of gene 19

2.2.6 PCR-mediated gene assembly 19

2.2.7 *Fok*I method of gene synthesis 21

2.3 *Bacillus megaterium* expression system

2.3.1 General features of *Bacillus megaterium* 21

2.3.2 Taxonomic position of *B. megaterium* 23

2.3.3 The cell structure of *B. megaterium* 23

2.3.4 Physiology and commercial products 25

2.3.5 *B. megaterium* as an expression host 26

2.3.6 *xyl* promoter and xylose utilization 29

2.3.7 Expression strain 30

2.3.8 Expression vector 31

2.3.9 Intracellular expression system 34

2.3.10 Protein secretion in *Bacillus* 36

2.3.11 Translocation and processing of *Bacillus* exoenzymes during secretion 38

2.3.12 Amino-terminal signal peptides 39

3.0 MATERIALS AND METHODS

3.1 General Equipment 43

3.2 Culture media, stock solutions and buffers 43

3.2.1 Media for transformation in *E. coli*

3.2.1.1 Luria Bertani Broth / Agar 46

3.2.1.2 M9 minimal medium 46

3.2.1.3 Agar plates 46

3.2.2 Media for protoplasting and transformation in *B. megaterium*

3.2.2.1 2X SMM 46

3.2.2.2 4X Bacto penassay broth (PAB) 47

3.2.2.3 SMMP 47

3.2.2.4 Polyethyleneglycol (PEG) 47

3.2.3 Media for plating / regeneration 47

3.2.3.1 DM3 regeneration medium 47

3.2.3.2 Agar plates 49

3.2.4 Stock solutions 49

3.2.5 Antibiotic 49

3.2.6 Sonication buffer 49

3.3 Host strains and vector

3.3.1 Bacterial strains 54

3.3.2 Cloning vector 54

3.3.3 Expression vector 54

3.3.4 Growth and maintenance of *B. megaterium* strain 54

3.4 Methods

3.4.1 Genomic DNA extraction for *B. subtilis* 57

3.4.2 Synthesis and modification of hGH gene for overexpression in *B. megaterium*

3.4.2.1 Construct 1: Construction of regulatory region of M7hGH gene 58

3.4.2.2 Construct 2: Construction of R2L4 signal peptide for hGH protein secretion in *B. megaterium* 60

3.4.2.2.1 Single-step assembly PCR 61

3.4.2.3 Construct 3: Construction of *Bacillus subtilis* CU1065 signal peptide and creation of fusion protein encoding *xylA* sequence and hGH sequence. 66

3.4.3 Agarose gel electrophoresis 69

3.4.4 Cloning of PCR products

3.4.4.1 Extraction of DNA from agarose gel 69

3.4.4.2 Ligation of the PCR product into pGEM®-T Easy vector 71

3.4.4.3 Preparation of competent cells 71

3.4.4.4 Transformation of competent cells 71

3.4.5 Identification of bacterial colonies that contain recombinant plasmid

3.4.5.1 Plasmid extraction for *E. coli* 72

3.4.5.2 Extraction of recombinant plasmid for *E. coli* 73

3.4.5.3 Conformation of recombinant plasmid with endonuclease digestion 74

3.4.6 Determination of nucleotide (DNA) sequence of recombinant plasmid 75

3.4.7 Cloning of the constructed synthetic hGH gene into B. megaterium

3.4.7.1 Restriction endonuclease digestion of recombinant pGEM®-T Easy vector to obtain the constructed hGH gene 75

3.4.7.2 Construction of pWH1520 expression vector 75

3.4.7.3 Preparation of *B. megaterium* protoplasts 76

3.4.7.4 Transformation of *B. megaterium* protoplasts 78

3.4.8 Identification of bacterial colonies that contain recombinant plasmid 78

3.4.8.1 Plasmid extraction for *B. megaterium* 79

3.4.8.2 Restriction analysis of small-scale preparation of plasmid DNA 80

3.4.9 Expression of recombinant *B. megaterium* WH320 80

3.4.9.1 Intracellular expression 80

3.4.9.2 Extracellular expression 81

3.4.10 Preparation of cell lysates for protein analysis 81

3.4.11 SDS-PAGE protein analysis 81

3.4.11.1 Preparation of SDS-polyacrylamide gel 82

3.4.11.2 SDS-PAGE 82

3.4.11.3 Coomassie Blue staining 84

3.4.11.4 Silver staining 84

3.4.12 Western blot analysis

3.4.12.1 Transfer of protein to nitrocellulose membrane 86

3.4.12.2 Immunodetection of recombinant hGH (rhGH) protein 86

3.4.13 Viable count

3.4.13.1 Growth of *B. megaterium* 88

3.4.13.2 Viable count of *B. megaterium* to determine the plasmid stability 89

4.0 RESULTS AND DISCUSSION

4.1.1 Construct 1: Construction of expression plasmid M7hGH for intracellular production of hGH gene

4.1.1.1 Plasmid DNA purification 91

4.1.1.2 Construction of regulatory region of M7hGH gene for intracellular overexpression 94

4.1.1.3 PCR to construct regulatory region of hGH gene for intracellular overexpression 96

4.1.1.4 Cloning of the PCR product to verify the nucleotide sequence 96

4.1.1.5 Sequencing of the amplified M7hGH 98

4.1.1.6 Expression of M7hGH in *B. megaterium*

4.1.1.6.1 Cloning of M7hGH into pWH1520 vector 100

4.1.1.6.2 Transformation 104

4.1.1.6.3 Screening of recombinant pWH1520 in *B. megaterium* 104

4.1.1.6.4 SDS-PAGE and Western Blot analysis 104

4.1.2 Constuct 2: Construction of R2L4 signal peptide for extracellular expression of hGH gene 110

4.1.2.1 The design of oligonucleotides 110

4.1.2.2 Secondary structure of mRNAs 113

4.1.2.3 Joining R2L4 signal peptide with hGH gene via single-step assembly PCR technique 115

4.1.2.4 Cloning of the PCR product in the pGEM®-T Easy Vector 119

4.1.2.5 Sequencing of the amplified R2L4sphGH 121

4.1.2.6 Expression of R2L4sphGH in *B. megaterium*

4.1.2.6.1 Cloning of R2L4sphGH into pWH1520 vector 121

4.1.2.6.2 Transformation 123

4.1.2.6.3 Screening of recombinant pWH1520 in *B. megaterium* 125

4.1.2.6.4 SDS-PAGE and Western Blot analysis 125

4.1.3 Construct 3: Construction of the *B. subtilis* CU1065 signal peptide and hGH gene in the *xylA* reading frame to create a translational or protein fusion 130

4.1.3.1 Chromosomal DNA purification 130

4.1.3.2 Joining the *B. subtilis* CU1065 signal peptide to the synthetic hGH gene 132

4.1.3.3 Cloning of the PCR product in pGEM®-T Easy Vector System 135

4.1.3.4 Sequencing of the amplified spBsubhGH 137

4.1.3.5 Expression of spBsubhGH in *B. megaterium* 137

4.1.3.5.1 Cloning of spBsubhGH into pWH1520 vector in *E. coli* JM109 137

4.1.3.5.2 Transformation of Construct 3 (spBsubhGH) in *B. megaterium* WH320 140

4.1.3.5.3 Screening of recombinant pWH1520 in *B. megaterium* 141

4.1.3.5.4 SDS-PAGE and Western Blot analysis 141

4.1.4 Viable Count

4.1.4.1 Determination the doubling time of Construct 1 (M7hGH), Construct 2 (R2L4sphGH) and Construct 3 (spBsubhGH) 145

4.1.4.2 Determination of plasmid stability in Construct 1 (M7hGH), Construct 2 (R2L4sphGH) and Construct 3 (spBsubhGH) 148

4.1.4.2.1 Segregational stability 151

4.1.4.2.2 Structural instability 151

5.0 CONCLUSION 153

REFERENCES 155

LIST OF PUBLICATIONS 174

LIST OF TABLES

			Page
Table	2.1	Effects of hGH	8
Table	2.2	Examples of industrial uses of *B. megaterium*	27
Table	2.3	Automated sequence analysis using a Beckman model 890 C sequenator	35
Table	3.1	Materials and reagents used and their suppliers	44
Table	3.2	Stock solution for the preparation of DM3 regeneration media	48
Table	3.3	Solutions for bacterial (*E. coli*) transformation and selection	50
Table	3.4	Solution for electrophoresis DNA	50
Table	3.5	Stock solutions for *E. coli* plasmid extraction	51
Table	3.6	Stock solutions for *B. megaterium* plasmid extraction	52
Table	3.7	Stock solutions for *B. subtilis* genomic extraction	53
Table	3.8	Genotype of bacterial strains used	55
Table	3.9	Oligonucleotides for R2L4 signal peptide-hGH (R2L4sphGH) amplification	63
Table	3.10	Oligonucleotides used for PR amplification of the *B. subtilis* signal peptide	67
Table	3.11	SDS-PAGE buffers and solutions	83
Table	3.12	Coomassie Blue staining solutions	85
Table	3.13	Silver staining solutions	85
Table	3.14	Western Blot buffers and solutions	87

LIST OF FIGURES

		Page
Figure 2.1	Covalent structure of hGH	5
Figure 2.2	Map of human chromosome 17	7
Figure 2.3	Theoretical representations of the hGH gene and the mRNA	9
Figure 2.4	The mRNA sequence and the amino acids sequence of hGH	11
Figure 2.5	Phylogenetic trees of some members of the genus *Bacillus* based on 16S rRNA sequence analysis	24
Figure 2.6	Map of pWH1520, a shuttle vector for *E. coli*/*B. megaterium*	32
Figure 2.7	Nucleotide sequence of (incomplete) *xylA* gene from *B. megaterium* including multiple cloning sites (MCS)	33
Figure 2.8	Model for signal peptide insertion into the cytoplasmic membrane and cleavage by SPase I.	40
Figure 2.9	Predicted signal peptides.	42
Figure 3.1	The circular map of the pGEM®-T Easy Vector	56
Figure 3.2	P1a and P2b primer	59
Figure 3.3a	R2L4sphGH RNA secondary structure before modification	62
Figure 3.3b	R2L4sphGH RNA secondary structure after modification	62
Figure 3.4	Overview of single-step assembly PCR method to generate R2L4sphGH	64
Figure 3.5	Overview of amplification of *B. subtilis* signal peptide from genomic sequence	70
Figure 3.6a	*B. megaterium* cells after 15 minutes added with 1 mg/ml lysozyme	77
Figure 3.6b	*B. megaterium* protoplasts after 45 minutes incubation with lysozyme	77
Figure 3.7	Overview of *B. megaterium* plasmid stability determination	90
Figure 4.1	The overview workflow of Construct 1 (M7hGH)	92
Figure 4.2	The purified pPic3.5K plasmid obtained from Loh (2005)	93
Figure 4.3	Amplified hGH gene using primer P1a and P2b using pPic3.5K as template	97

Figure 4.4	Gel purified product of PCR amplification using P1a and P2b primers	97
Figure 4.5	The purified pGEMM7hGHplasmids for sequencing with T7 and SP6 universal primer to determine the order of nucleotides of hGH gene	99
Figure 4.6	The restriction endonuclease digestion of plasmid M7hGH	99
Figure 4.7	M7hGH fragment sequence using T7 forward primer with length 624 bp	101
Figure 4.8	Constructed regulatory region of M7hGH gene for intracellular overexpression comprised of stop codon, RBS, Added codons (taaca) and start codon	102
Figure 4.9	The double-digestion product of pWH1520 with *Spe* I and *Bam*H I restriction enzymes confirming succeed of subcloning of the M7hGH gene in pWH1520 plasmid	103
Figure 4.10	Restriction analysis of plasmid DNA from recombinant *B. megaterium* cell to determine the presence of ~624 bp insert and the ~8 kb pWH1520 shuttle vector	105
Figure 4.11	Analysis of protein expressed in control strain pWH1520 (without insert) on 12.5% (v/v) SDS-polyacrylamide gel	107
Figure 4.12a	Analysis of rhGH protein expressed in recombinant clone of Construct 1 (M7hGH) on 12.5% (v/v) SDS-polyacrylamide gel	107
Figure 4.12b	Immunodetection (Western Blot) of rhGH	107
Figure 4.13	The synthetic hGH gene containing AGA codon were inserted at the first 8, 16, 19, 64, 77, 94, 127, 134, 167 and183 codons of hGH gene	109
Figure 4.14	Overview workflow of Construct 2 (R2L4sphGH)	111
Figure 4.15	Six overlapping oligonucleotides used to construct 147 bp DNA fragments comprised of synthetic signal peptide, RBS sequence and cloning sites	114
Figure 4.16	The single-step assembly PCR which was allowed to assemble within the range of 55 ^{o}C to 65 ^{o}C	117
Figure 4.17	The amplification of 147 bp size via gradient PCR which indicates the 6 oligonucleotides were assembled accordingly	117
Figure 4.18	The amplification of 702 bp synthetic regulatory region and hGH gene	118
Figure 4.19	The purified pGEMR2L4sphGH plasmids	120

Figure 4.20	The restriction endonuclease of digested pGEMR2L4sphGH to confirm the insertion of the gene	120
Figure 4.21	R2L4sphGH fragment sequence using T7 Forward primer with length 709 bp	122
Figure 4.22	The purified recombinant pWHR2L4sphGH plasmids in *E. coli* JM109	124
Figure 4.23	The restriction endonuclease digestion of R2L4sphGH (in *E. coli* JM109) to determine the presence of the recombinant product	124
Figure 4.24	Restriction analysis of isolated DNA from recombinant *B. megaterium* WH320 cell to determine the presence of ~709 bp insert and ~8 kb pWH1520 shuttle vector	126
Figure 4.25	Analysis of protein expressed in control strain pWH1520 (without insert) on 12.5% (v/v) SDS-polyacrylamide gel	127
Figure 4.26a	Analysis of intracellular rhGH protein expressed in recombinant clone of Construct 2 (R2L4sphGH) on 12.5% (v/v) SDS-polyacrylamide gel	127
Figure 4.26b	Immunodetection (Western Blot) of rhGH	127
Figure 4.27	The overview workflow of Construct 3 (spBsubhGH)	131
Figure 4.28	Construction of the *B. subtilis* CU1065 signal peptide and hGH gene in the *xylA* reading frame to create a translational or protein fusion	133
Figure 4.29	The amplified *B. subtilis* signal peptide (spBsub) (~146 bp)	134
Figure 4.30	The amplified fragment of the spBsubhGH	136
Figure 4.31	Screening of recombinant spBsubhGH in pGEM®-T Easy Vector	138
Figure 4.32	The restriction endonuclease digestion product	138
Figure 4.33	spBsubhGH fragment sequence using T7 Forward primer with length 719 bp	139
Figure 4.34	Restriction analysis of isolated DNA from recombinant *B. megaterium* cell to determine the presence of ~719 bp insert and ~8 kb pWH1520 shuttle vector	142
Figure 4.35	Analysis of protein expressed in control strain pWH1520 (without insert) on 12.5% (v/v) SDS-polyacrylamide gel	143
Figure 4.36a	Analysis of intracellular rhGH protein expressed in recombinant clone of Construct 3 (spBsubhGH) on 12.5% (v/v) SDS-polyacrylamide gel	143
Figure 4.36b	Immunodetection (Western Blot) of rhGH	143

LIST OF GRAPHS

		Page
Graph 4.1	The growth curve of Construct 1 (M7hGH) in LB medium	146
Graph 4.2	The growth curve of Construct 2 (R2L4sphGH) in LB medium	146
Graph 4.3	The growth curve of Construct 3 (spBsubhGH) in LB medium	146
Graph 4.4	The stability of Construct 1 (M7hGH) plasmid	149
Graph 4.5	The stability of Construct 2 (R2L4sphGH) plasmid	149
Graph 4.6	The stability of Construct 3 (spBsubhGH) plasmid	149

ABBREVIATIONS

A	adenine
bp	base pair
C	cytosine
$CaCl_2$	calcium chloride
cDNA	complementary deoxyribonucleic acid
CNBr	Cyanogen Bromide
Da	dalton
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleoside triphosphates
ECL	electrochemiluminescence
EDTA	ethylene diaminetetraacetic acid
EtBr	ethidium bromide
FDA	Food and Drug Administration
G	guanine
g	gram
g	relative centrifugal force (centrifugation)
GH	growth hormone
h	hour
hGH	human growth hormone
H_3PO_4	phosphoric acid
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	kilo base
kDa	kilo Dalton
K_2HPO_4	di-potassium hydrogen phosphate
KH_2PO_4	potassium dihydrogen phosphate
KOH	potassium hydroxide

kPa	kilo Pascal
LB	Luria-Bertani
mg	milligram
$MgCl_2$	magnesium chloride
$MgSO_4$	magnesium sulphate
min	minute
mL	millilitre
mRNA	messenger RNA
NaCl	sodium chloride
NEB	New England Biolabs
ng	nanogram
$(NH_4)_2SO_4$	ammonium sulphate
nm	nanometer (wavelength)
nt	nucleotides
OD	optical density
OD_{600}	optical density at 600 nm
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
Pfu	Pyrococcus furiosus
psi	pounds per square inch
RBS	ribosome binding site
rhGH	recombinant hGH
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis

sec	second
somatotropin	growth hormone
SMM	Spizizen's minimal medium
T	thymine
T_a	annealing temperature
TAE	tris-acetate-EDTA
Taq	*Thermus aquaticus*
TBS	tris buffered saline
TEMED	N, N, N', N'-tetramethylethylenediamine
T_m	melting temperature
tRNA	transfer RNA
Tween® 20	polyoxyethylene-sorbitan mono-laurate
U	unit of enzyme activity
UV	ultraviolet
v/v	volume/ volume
w/v	weight/ volume
X-Gal	5-bromo-4-chloro-3 indolyl-β-D-galactopyranoside
xyl	xylose utilization

PEMBENTUKAN VEKTOR UNTUK PENGEKSPRESSAN LAMPAU HORMON PERTUMBUHAN MANUSIA REKOMBINAN DI DALAM *Bacillus megaterium*

ABSTRAK

Bagi mengekspreskan gen hGH di dalam *Bacillus megaterium*, 3 konstrak telah direka. Konstrak-konstrak ini dinamakan sebagai Construct 1 (M7hGH), Construct 2 (R2L4sphGH) dan Construct 3 (spBsubhGH). Construct 1 direka dengan memasukkan gen pengawalatur (kodon penamat, tapak perlekatan ribosom, jujukan TAACA dan kodon pemula) yang akan diekspreskan di dalam sel. Manakala Construct 2 telah dicipta untuk diekpreskan di luar sel yang mana mengandungi isyarat peptida daripada *Bacillus brevis* dan gen. Bagi Konstrak 3, isyarat peptida daripada *Bacillus subtilis* CU1065 digunakan untuk pengekspresan di luar sel, tetapi gen pengawalatur (kodon penamat, tapak perlekatan ribosom dan kodon pemula antara protein *xylA* dengan hGH) telah disingkirkan untuk menghasilkan protein gabungan antara *xylA* dan gen hGH. Walaubagaimanapun, tiada pengekspresan diperolehi daripada Construct 1 dan Construct 2, mungkin berikutan dengan ketidakhadiran penggunaan kodon yang sesuai. Gen hGH sintetik mengandungi 10 kodon arginina (AGA) (4.8%) yang frekuensinya lebih tinggi daripada 1.7% AGA dan 0.9% AGG yang dianggarkan dalam *E. coli*. Contruct 3 menghasilkan ~20 kDa protein yang tidak dapat dikenalpasti maka, analisis lanjutan ke atas protein tersebut diperlukan. Untuk menentukan kestabilan plasmid, purata peratusan sel yang membawa plasmid telah dihitung ke atas ketiga-tiga Konstrak. Bagi Konstrak 1 (M7hGH), kadar peratusan pengurangan sel yang membawa plasmid adalah 10% bagi setiap generasi, Konstrak 2 (R2L4sphGH) adalah 5% bagi setiap generasi, dan 4.8% bagi Konstrak 3 (spBsubhGH).

CONSTRUCTION OF VECTORS FOR THE OVEREXPRESSION OF RECOMBINANT HUMAN GROWTH HORMONE IN *Bacillus megaterium*

ABSTRACT

In order to express the hGH gene in *Bacillus megaterium*, 3 constructs were designed. The constructs were named as Construct 1 (M7hGH), Construct 2 (R2L4sphGH) and Construct 3 (spBsubhGH). Construct 1 was designed with regulatory features [stop codon, ribosome binding site (RBS), TAACA sequence and start codon] meant for intracellular expression. Construct 2, was created to express human growth hormone (hGH) extracellularly by using the signal peptide from *Bacillus brevis*. For Construct 3, the signal peptide from *B. subtilis* CU1065 was used to express hGH extracellularly but the regulatory gene [stop codon, ribosome binding site (RBS) and start codon between *xylA* protein and hGH] was excluded to create a fusion protein combining *xylA* and the hGH gene. However, no expressions were obtained from Construct 1 and 2 probably due to the presence of unfavourable codons. The synthetic hGH gene contains 10 arginine codons (AGA) (4.8%) in which the frequency is much higher than the 1.7% AGA and 0.9% AGG estimated for *E. coli*. Construct 3 yielded an unidentified ~20 kDa protein and further analysis is required to identify the protein. The average percentage of cells carrying plasmid was calculated among these 3 constructs in order to determine the plasmid stability For Construct 1 (M7hGH), the average rate of reduction of cells that carry plasmid is 10% for each generation, Construct 2 (R2L4sphGH) is 5% for each generation and Construct 3 (spBsubhGH) is 4.8% for each generation.

INTRODUCTION

1.0 Introduction

Growth hormone (GH) plays an important role in metabolism, protein synthesis and cell proliferation (Kostyo and Isaksson, 1977). Human growth hormone (hGH), somatotropin, is a protein consisting of 191 amino acids protein with a molecular weight of ~22,000 Da and contains two disulfide bonds. This hormone is essential for linear growth and its application in the treatment of hypopituitary dwarfism is well established (Raben, 1958; Goodman *et al*., 1968). Growth hormone may also be effective in the treatment of other disorders including bone fractures, burns and bleeding ulcers (Singh *et al*., 1998). A number of strategies have been employed for its expression (Goeddel *et al*., 1979a; Hsiung *et al*., 1989). hGH is a non-glycosylated protein and hence a prokaryotic expression system is preferred. hGH expression in culture media (Hsiung *et al*., 1989), inclusion bodies (Schoner *et al*., 1985) and periplasmic space (Chang *et al*., 1987; Ghorpade and Garg, 1993) has been documented.

Among Gram-positive bacteria, *Bacillus subtilis* is genetically the most thoroughly studied organism. However, other *Bacillus* species are of great biotechnological interest, since they produce a variety of industrially important enzymes, such as proteases and amylases and also vitamins (Priest, 1977). For the production of enzymes on an industrial scale, on the other hand, different species of the genus *Bacillus* are more important such as *B. amyloliquefaciens* as a source of α-amylase, *B. licheniformis* as a producer of alkaline protease and *B. megaterium* for glucose dehydrogenase (Heilmann *et al*., 1988), and mutarotase (Rygus and Hillen, 1991).

Recently, the construction of a vector for constitutive expression of the homologous glucose dehydrogenase and the heterologous chloramphenicol acetyltransferase in *B. megaterium* has been reported (Meinhardt *et al*., 1989). Besides that, there was no indication of proteolytic instability of the products in *B.*

megaterium (Rygus and Hillen, 1991). The apparent stability of the products expressed and the plasmids encoding these genes motivated us to construct a heterologous regulated expression for *B. megaterium*.

This thesis describes the construction of plasmid encoding the regulated expression of the heterologous human growth hormone (hGH) gene by using *Pichia pastoris* codon preference at high levels in *B. megaterium*. The regulatory elements used for the construction have been derived from the *B. megaterium*-encoded operon for xylose utilization (Rygus *et al*., 1991).

1.1 Research objectives

Due to the importance of rhGH to humanity and potential demand in biopharmaceutical market, the application of genetic engineering was widely used to produce this protein in genetically selected and modified microorganisms. In recent years, *Bacilli* have become very important experimental models in both microbiological and genetic research.

Therefore, in this research, the cloning and expression of rhGH in *B. megaterium* expression system was carried out. The objectives of this research are:

(1) To express the recombinant hGH protein in *B. megaterium* ;

(2) To express synthetic hGH gene extracellularly by adding the signal peptide sequence.

CHAPTER TWO

LITERATURE REVIEW

2.0 Literature Review

2.1 Human Growth Hormone

2.1.1 Historical Background

Growth hormone (GH) has been in the forefront of endocrine research for almost one hundred years. Early research studied by Bell (1917), Evans and Long (1921) and Smith (1927 & 1930) found out how hypophysectomy affected the body size of the rat, the frog, and the dog. They also found that in the young animal, hypophysectomy stunted growth, and in the mature animal it caused weight loss and visceral atrophy. Exogenous administration of pituitary extracts reversed these effects. From these observations emerged the concept of a pituitary growth hormone or somatotropin.

Li and co-workers isolated GH from bovine pituitary glands in 1945 (Li *et al*., 1945). Subsequent work by Li, as well as by Wilhelmi and others in 1948 (Wilhelmi *et al*., 1948) established procedures for isolating human GH (hGH) in what was considered to be homogenous from by the standards of the 1950s, e.g. ion-exchange resin and cellulose column chromatography, paper electrophoresis, and sedimentation analysis (Smith *et al*., 1949). It was recognized that the hormone not only stimulated growth of the skeletal and soft tissues, but also influenced protein, carbohydrate, and fat metabolism.

For the past 25 years, human growth hormone (hGH) has been used for treatment of certain disease states associated with short stature, especially growth hormone deficiency (Rabens, 1958). Unlike other peptide hormones such as insulin, the homologous hormone from other species is inactive in man (Bergenstal and Lipset, 1960). Therefore the only source of growth hormone has been human pituitary gland obtained at necropsy. In the United States the collection of pituitary glands, purification of growth hormone, and distribution of hGH to research-workers

for use in patients has been centralized by the National Pituitary agency under the auspices of the National Institutes of Health. This program has yielded much useful research on growth hormone therapy and hundreds of patients have benefited.

Techniques in molecular biology have led to the ability to isolate the hGH mRNA and to make partly synthesized hGH gene. In 1979, hGH was directly expressed in *Escherichia coli* using pBR322 as a cloning vector (Goeddel *et al*., 1979a). This biosynthetic hGH has exactly the same amino acid sequence as pituitary hGH, with the additional of a methionyl residue on the N-terminal end. This development provides an essentially unlimited supply of hGH for research and treatment. Previous work has shown that this synthetic methionyl human growth hormone (met-hGH) is biologically active in animals without detectable adverse side-effects (Olson *et al*., 1981).

Human growth hormone was first used to treat stunted growth in children. Clinically, deficiencies in growth hormone or receptor defects manifest as growth retardation or dwarfism. The effect of excessive secretion of growth hormone is also very dependent on the age of onset and is seen as two distinctive disorders such as giantism that begin in young children and acromegaly that occur in adults. Later, it was used in people with HIV disease to treatment the gauntness of AIDS-related wasting and, more recently, the fat accumulations associated with lipodystrophy. HGH may also play a role in immune reconstitution. Lately, the application of hGH was approved by FDA for the treatment of AIDS-related wasting syndrome (Roehr, 2003).

2.1.2 Structure of the hGH Genes

Human growth hormone (hGH) is a single chain polypeptide of 191 amino acids with two loops formed by disulfide bridges (Figure 2.1). The molecule is predominantly α-helical in secondary structure (Chawla *et al*., 1983).

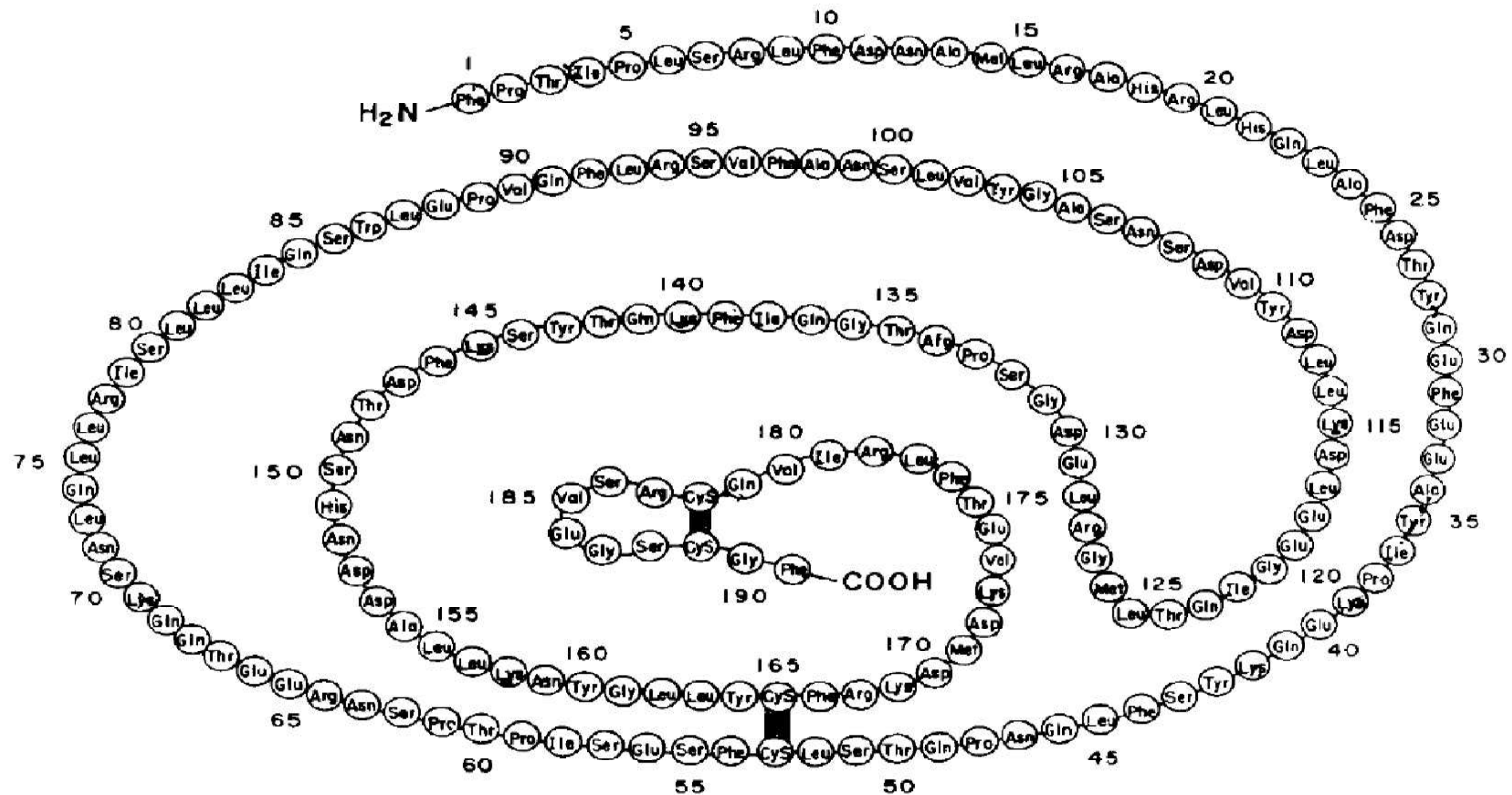

Figure 2.1 Covalent structure of hGH. hGH has two disulphide bonds between residues 53-165 and 182-189. (Picture adapted from Chawla *et al*., 1983)

In humans, the long arm of chromosome 17 (q22-24) contains cluster of five GH-related genes. A report has been made by Martial *et al*. and by Roskam & Rougeon in 1979 where they constructed recombinant DNA copies of pre-hGH mRNA from human pituitary glands. The sequence was found to be 94% homologous with the sequence of human chorionic somatomammotropin (hCS). Fiddest *et al*. (1979) found that there were two nonallelic copies of hGH genes and at least three copies of hCS genes in each hGH and hCS gene cluster. These include the normal pituitary hGH or hGH-N gene (DeNoto *et al*., 1981; Seeburg, 1982) in the somatotroph cell (Wood, 2001), the nonallelic variant hGH or hGH-V gene (Seeburg, 1982) produced by placenta during pregnancy (Wood, 2001), an hCS gene whose sequence differs slightly from that of the major component of hCS mRNA (Seeburg, 1982) and another hCS gene (Chawla *et al*., 1983). However, the exact number and linear array of hGH and hCS genes are not known (Chawla *et al*., 1983). This family of gene is shown in Figure 2.2 while Table 2.1 lists the molecular forms of pituitary human growth hormone.

Figure 2.3 shows the theoretical representation of the hGH gene and the mRNA from it. There are five exons or coding sequence designated I-V, which are separated by four introns or interverning sequences designated A-D. The positions and lengths of the four introns are identical in the hGH-N, hGH-V and hCS genes (Chawla *et al*., 1983). Exon I contains some 5' untranslated nucleotides, trinucleotide codons -26 to -24 of the pre-hGH signal peptide, and the first nucleotide of codon -23. The second exon codes for the remainder of the signal peptide and amino acids 1 to 31 of hGH. Exons III, IV, and V code for amino acids 32 to 71, 71 to 126 and 127 to 191, respectively. Introns A, B, C and D contain 256, 209, 93 and 253 base pairs. Each intron begin with a GH dinucleotide and ends with an AG. Coding sequences are spliced together at these junctions during the post-transcriptional

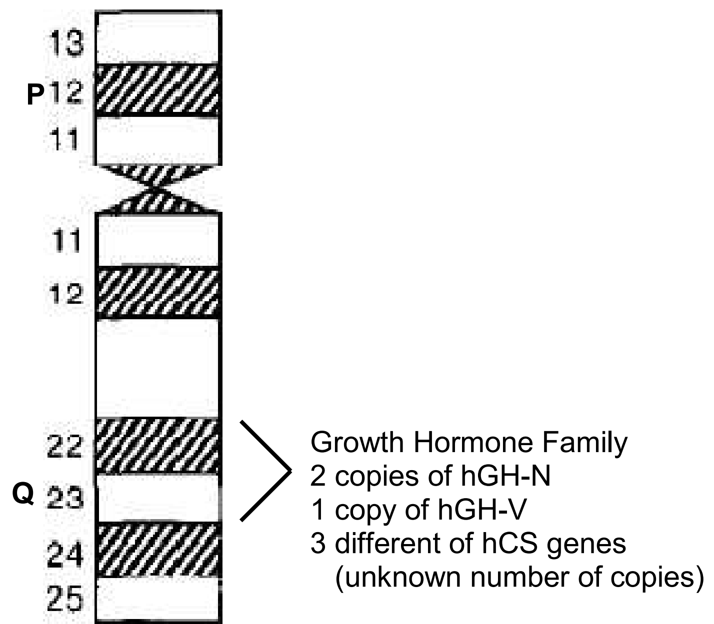

Figure 2.2 Map of human chromosome 17. The dark and the light bands are given numbers outward from the centromere to facilitate identification. The GH gene family is located on the long arm, designated Q, somewhere within the region of bands 22-24. There are two identical copies of hGH-N gene, one copy of the hGH-V gene, and an unknown number of copies of three different hCS genes. (Picture adapted from Chawla *et al*., 1983)

Table 2.1 Effects of hGH (Table was adapted from Chawla *et al*., 1983)

Metabolic	Stimulates amino acid transport
	Stimulates protein synthesis in most cell types
	Stimulates DNA/RNA synthesis in most cell types
	Stimulates polyamine synthesis
	Inhibits insulin action on glucose metabolism
Physiologic	Increases renal blood flow, glomerular filtration rate, and tubular reabsorption of PO_4
	Increases basal metabolic rate
	Stimulates new bone formation
	Stimulates erythropoiesis
	Expands extracellular fluid space
Anatomic	Accelerates linear growth
	Reduces adipose mass and enlarges lean body mass (muscle, liver, kidney, heart, pancreas, skeleton, connective tissue)

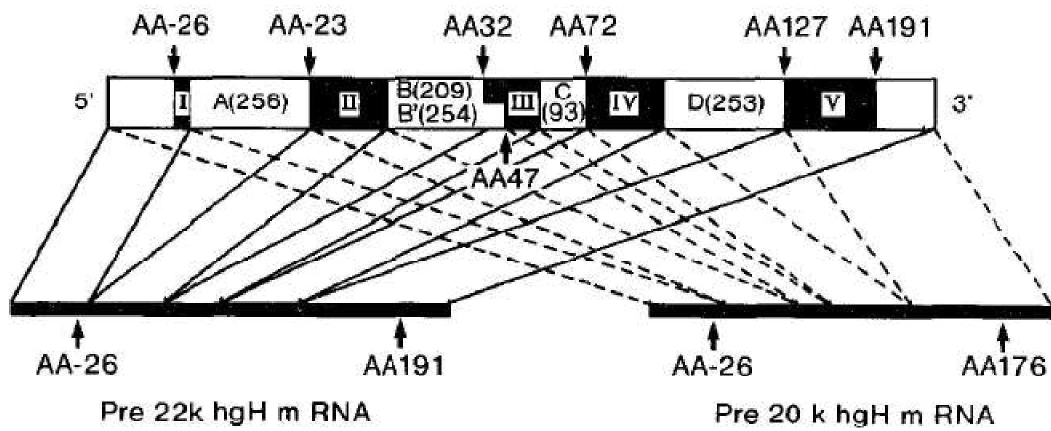

Figure 2.3 Theoretical representations of the hGH gene and the mRNA. The hGH gene contains three introns: intron A between amino acids (AA) 24, intron B between AA 31 and 32, and intron C between AA 71 and 72. These are excised out when the gene is transcribed into the mRNA for the hormone (Picture adapted from Chawla *et al.*,1983)

processing of mRNA (Chawla *et al*., 1983). The hGH-N gene contains an alternate splice point [designated B' in Figure 2.3 preceding the codon for amino acids 47 (DeNoto *et al*., 1981; Seeburg, 1982)]. 32 to 46 of 22K hGH (Lewis *et al*., 1980). As suggested by Wallis (1980), the 20K hGH, which lacks amino acids 32 to 46 of 22K hGH (Lewis *et al*., 1980), may arise from splicing of the pre-hGH messenger precursor RNA at this point.

Human growth hormone binds to specific receptors on cultured human lymphocytes; on the plasma membrane fraction of leporine or rodent hepatocytes and of leporine mammary glands; and on human, canine, or rodent adipocytes (Lesniak *et al*., 1973; Moore and Jin, 1978; Fagin *et al*., 1980).

Transcription and translation give rise to two forms of hGH in the pituitary. The two forms are a 22-kDa form, which accounts for approximately 90% of GH produced, and a 20-kDa form resulting from an alternative splice site on exon 3, which accounts for 5-10% of hGH production. The 20-kDa form of hGH lacks a sequence of 15 amino acids present at position 32-46 in the 22-kDa molecule and has a total of 176 residues compared with the 191 residues in the 22-kDa peptide. The bioactivity of 20-kDa hGH is less than that of the 22-kDa form in most of the assay systems tested *in vitro* but the *in vivo* bioactivity in humans is unknown (Wood, 2001). Figure 2.4 shows the mRNA sequence and the amino acids sequence of hGH (Roskam and Rougeon, 1979).

2.1.3 hGH synthesis

Human growth hormone (hGH) is synthesized in somatotropes, a subclass of the pituitary acidophilic cells where somatotropes are the most abundant cells in the gland. The concentration of hGH in the pituitary is 5-15mg/g, which is much higher than the microgram per gram quantities of other pituitary hormones (Murray *et al*., 1993). The expression of hGH is restricted to the pituitary and regulated by pituitary transcription factor-1 (GHF-1) which binds to the hGH promoter acting in concert with

-26	-25	-24	-23	-22	-21	-20	-19	-18	-17	-16	-15	-14	-13	-12	-11	-10
AUG	**GCU**	**ACA**	**GGC**	**UCC**	**CGG**	**ACG**	**UCC**	**CUG**	**CUC**	**CUG**	**GCU**	**UUU**	**GGC**	**CUG**	**CUC**	**UGC**
Met	Ala	Thr	Gly	Ser	Arg	Thr	Ser	Leu	Leu	Leu	Ala	Phe	Gly	Leu	Leu	Cys
-9	-8	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8
CUG	**CCC**	**UGG**	**CUU**	**CAA**	**GAG**	**GGC**	**AGU**	**GCC**	**UUC**	**CCA**	**ACU**	**AUA**	**CCA**	**CUA**	**UCU**	**CGT**
Leu	Pro	Trp	Leu	Gln	Glu	Gly	Ser	Ala	Phe	Pro	Thr	Ile	Pro	Leu	Ser	Arg
9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
CUA	**UUC**	**GAU**	**AAC**	**GCU**	**AUG**	**CUU**	**CGU**	**GCU**	**CAU**	**CGU**	**CUU**	**CAU**	**CAG**	**CUG**	**GCC**	**UUU**
Leu	Phe	Asp	Asn	Ala	Met	Leu	Arg	Ala	His	Arg	Leu	His	Gln	Leu	Ala	Phe
26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
GAC	**ACC**	**UAC**	**CAG**	**GAG**	**UUU**	**GAA**	**GAA**	**GCC**	**UAU**	**AUC**	**CCA**	**AAG**	**GAA**	**CAG**	**AAG**	**UAU**
Asp	Thr	Tyr	Gln	Glu	Phe	Glu	Glu	Ala	Tyr	Ile	Pro	Lys	Glu	Gln	Lys	Tyr
43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59
UCA	**UUC**	**CUG**	**CAG**	**AAC**	**CCC**	**CAG**	**ACC**	**UCC**	**CUC**	**UGU**	**UUC**	**UCA**	**GAG**	**UCU**	**AUU**	**CCG**
Ser	Phe	Leu	Gln	Asn	Pro	Gln	Thr	Ser	Leu	Cys	Phe	Ser	Glu	Ser	Ile	Pro
60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76
ACA	**CCC**	**UCC**	**AAC**	**AGG**	**GAG**	**GAA**	**ACA**	**CAA**	**CAG**	**AAA**	**UCC**	**AAC**	**CUA**	**GAG**	**CUG**	**CUC**
Thr	Pro	Ser	Asn	Arg	Glu	Glu	Thr	Gln	Gln	Lys	Ser	Asn	Leu	Glu	Leu	Leu
77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93
CGC	**AUC**	**UCC**	**CUG**	**CUG**	**CUC**	**AUC**	**CAG**	**UCG**	**UGG**	**CUG**	**GAG**	**CCC**	**GUG**	**CAG**	**UUC**	**CUC**
Arg	Ile	Ser	Leu	Leu	Leu	Ile	Sln	Ser	Trp	Leu	Glu	Pro	Val	Gln	Phe	Leu
94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110
AGG	**AGU**	**GUC**	**UUC**	**GCC**	**AAC**	**AGC**	**CUA**	**GUG**	**UAC**	**GGC**	**GCC**	**UCU**	**GAC**	**AGC**	**AAC**	**GUC**
Arg	Ser	Val	Phe	Ala	Asn	Ser	Leu	Val	Tyr	Gly	Ala	Ser	Asp	Ser	Asn	Val
111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127
UAU	**GAC**	**CUC**	**CUA**	**AAG**	**GAC**	**CUA**	**GAG**	**GAA**	**GGC**	**AUC**	**CAA**	**ACG**	**CUG**	**AUG**	**GGG**	**AGG**
Tyr	Asp	Leu	Leu	Lys	Asp	Leu	Glu	Glu	Gly	Ile	Gln	Thr	Leu	Met	Gly	Arg
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144
CUG	**GAA**	**GAU**	**GGC**	**AGC**	**CCC**	**CGG**	**ACU**	**GGG**	**CAG**	**AUC**	**UUC**	**AAG**	**CAG**	**ACC**	**UAC**	**AGC**
Leu	Glu	Asp	Gly	Ser	Pro	Arg	Thr	Gly	Gln	Ile	Phe	Lys	Gln	Thr	Tyr	Ser
145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161
AAG	**UUC**	**GAC**	**ACA**	**AAC**	**UCA**	**CAC**	**AAC**	**GAU**	**GAC**	**GCA**	**CUA**	**CUC**	**AAG**	**AAC**	**UAC**	**GGG**
Lys	Phe	Asp	Thr	Asn	Ser	His	Asn	Asp	Asp	Ala	Leu	Leu	Lys	Asn	Tyr	Gly
162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178
CUG	**CUC**	**UAC**	**UGC**	**UUC**	**AGG**	**AAG**	**GAC**	**AUG**	**GAC**	**AAG**	**GUC**	**GAG**	**ACA**	**UUC**	**CUG**	**CGC**
Leu	Leu	Tyr	Cys	Phe	Arg	Lys	Asp	Met	Asp	Lys	Val	Glu	Thr	Phe	Leu	Arg
179	180	181	182	183	184	185	186	187	188	189	190	191				
AUC	**GUG**	**CAG**	**UGC**	**CGC**	**UCU**	**GUG**	**GAG**	**GGC**	**AGC**	**UGU**	**GGC**	**UUC**				
Ile	Val	Gln	Cys	Arg	Ser	Val	Glu	Gly	Ser	Cys	Gly	Phe				

Figure 2.4 The mRNA sequence and the amino acids sequence of hGH. The numbers above the mRNA sequence represent the location of the amino acids; 1-191 represent mature hGH and -26 to -1 represent signal peptide of hGH (Reproduced from Roskam and Rougeon, 1979).

several other more ubiquitous DNA binding proteins (Miyachi *et al*., 1993). The secretion of hGH is regulated by growth hormone releasing hormone (GHRH) and somatostatin. GHRH controls hGH synthesis by stimulating transcription of hGH mRNA while somatostatin determines the timing and amplitude of hGH pulses (Miyachi *et al*., 1993). The pulsatile GH secretion is influenced by a number of neurogenic, metabolic and hormonal factors.

2.1.4 Biochemical and Physiologic Characteristics of hGH

Human growth hormone is essential for postnatal growth and for normal carbohydrate, lipid, nitrogen and mineral metabolism. The growth related effects are primarily mediated by IGF-I (Insulin-Like Growth Factor-I), a member of the insulin-like gene family (Javier *et al*., 1996). Another closely related peptide found in human plasma, IGF-II (Insulin-like Growth Factor-II) has activity similar or identical to what is often referred to in the rat as multiplicational-stimulating activity (MSA) (Javier *et al*., 1996). IGF-I and IGF-II both bind to membrane receptors; however, they can be differentiated on the basis of specific radioimmunoassay. IGF-I has 70 amino acids and IGF-II has 67. Plasma level of IGF-II are twice those of IGF-I, but it is IGF-I that correlates most directly with growth hormone effects (Javier *et al*., 1996). Individuals who lack sufficient IGF-I but have IGF-II fail to grow normally (Franchimont and Burger, 1975).

Table 2.1 summarizes the biologic effects associated with hGH. The primary function of the hormone is to promote proportionate growth of both soft and skeletal tissues. These anabolic effects are at least in part mediated by hGH-dependent growth factors called somatomedin (Wilhelmi, 1982).

The growth-promoting characteristic of hGH is measured by either (a) the weight gain test, in which the increase in body weight of the young hypophysectomized rat is monitored during ten daily injections of the hormone, or (b) the tibia assay, in which the growth of the proximal epiphysis of the tibia of the young

hypophysectomized rat is measured after daily hGH injections. Other assays measure metabolic actions of the hormone on muscle, adipose tissue and mammary gland. The latter organ reveals the lactogenic property of hGH (Rudman, 1981).

2.1.4.1 Carbohydrate metabolism

Growth hormone generally antagonizes the effects of insulin. Hyperglycemia after growth hormone administration is the combined result of decreased peripheral utilization of glucose and increased hepatic production via gluconeogenesis (Franchimont and Burger, 1975). In the liver, growth hormone increases liver glycogen, probably from activation of gluconeogenesis from amino acids. Impairment of glycolysis may occur at several steps and the mobilization of fatty acids from triacylglycerol stores may also contribute to the inhibition of glycolysis in muscle (Franchimont and Burger, 1975). Prolonged administration of growth hormone may result in diabetes mellitus (Franchimont and Burger, 1975).

2.1.4.2 Lipid metabolism

Growth hormone promotes the release of free fatty acids and glycerol from adipose tissue, increases circulating free fatty acids and causes increased oxidation of free fatty acids in the liver. Under conditions of insulin deficiency as diabetes, increased ketogenesis may occur. These effects and those carbohydrate metabolisms probably are not mediated by IGF-I (Franchimont and Burger, 1975).

2.1.4.3 Mineral metabolism

Growth hormone promotes a positive calcium, magnesium and phosphate balance and causes the retention of Na^+, K^+ and Cl^-. The first effect probably relates to the action of growth hormone in bone, where it promotes growth of long bones at the epiphysial plates in growing children and appositional or acral growth in adults.

In children, growth hormone also increases formation of cartilage (Franchimont and Burger, 1975).

2.1.4.4 Prolactin-like effects

Growth hormone binds to lactogenic receptors and thus has many of the properties of prolactin, such as stimulation of the mammary glands, lactogenesis (Franchimont and Burger, 1975)

2.1.5 Recombinant hGH

Since the introduction of recombinant human growth hormone (rhGH), several clinical trials have examined the characteristics of growth hormone deficiency (GHD) and the effect of growth hormone (GH) replacement therapy in adults with GHD. Adults with GHD appear to have muscle weakness that can be improved with rhGH therapy (Cuneo *et al*., 1990; Rutherford *et al*., 1994; Beshyah *et al*., 1995; Johannsson *et al*., 1997; Sartorio *et al*., 1995). Due of the main cause of the decrease in muscle strength is a decrease in muscle mass, which is frequently reported in adults with GH deficiency. Normal quadriceps strength per thigh muscle mass, estimated by anthropometry, was reported in a small group (n=6) of patients with childhood-onset (Sartorio *et al*., 1995).

Since the successful synthesis of hGH by recombinant DNA technology (Goeddel *et al*., 1979b), hGH therapy in Turner's syndrome began systematically as clinical trials (Takano *et al*., 1986; Takano *et al*., 1989; Raiti *et al*., 1986; Rosenfeld *et al*., 1988).

Patients with thermal burns have increased protein catabolism, impaired wound healing, depressed immune response and high incidence of sepsis (Wolfe, 1986 and Batstone *et al*., 1982). Decreased bioavailability as well as low circulating levels of anabolic hormones (GH, insulin, IGF-1 and testosterone) has been reported in major burn injury that correlate with negative nitrogen balance (Jenkins and Ross,

1996). Recombinant human growth hormone (rhGH) therapy in burned patients has been reported to reduce protein catabolism, improve donor site healing and decrease mortality (Jenkins and Ross, 1996). Severe protein catabolism, delay in wound healing, sepsis and multiorgan dysfunction are major factors influencing morbidity and mortality in burns (Jenkins and Ross, 1996).

2.2 Techniques of synthetic gene construction

2.2.1 Background

Within recent years, a great deal of interest has been focused on the cloning of DNA sequences coding for biologically and medically important proteins. Recent advances in the chemical synthesis of oligodeoxyribonucleotides have now made it possible to chemically synthesize DNA sequences coding for proteins composed of well over 100 amino acids (Edge *et al*., 1981). The advantages of chemically synthesizing such sequences, reside in the greater potential for engineering certain desired features into these DNAs. These may include conveniently placed restriction endonuclease sites and transcriptional and translational regulatory signals (Edge *et al*., 1981; Itakura *et al*., 1977; Crea *et al*., 1978; Goeddel *et al*., 1979b) as well as codon usage designed to maximize the use of the most abundant species of tRNAs for a given organism (Edge *et al*., 1981 and Ikemura, 1981).

Modification of a gene in a controlled fashion is an important tool in the fields of molecular genetics and protein engineering. Small alteration can be obtained by either an oligonucleotide-directed "site-specific" mutagenesis or localized random mutagenesis of a gene already cloned in a plasmid vector (Gilliam *et al*., 1979; Miyada *et al*., 1982; Shortle *et al*., 1980; Botstein and Shortle, 1985). However, if extensive or multiple alterations are required, for example, in the preparation of homologous proteins of different species, the oligonucleotide-directed mutagenesis would appear to be too cumbersome (Sung *et al*., 1986). In this situation, separate

synthesis of individual genes would be a more appropriate approach, albeit a laborious one (Jay *et al*., 1984).

The assembly of DNA sequences from oligos finds applications in DNA synthesis, gene expression and in vitro mutagenesis. Several publications describe the assembly of DNA sequences from oligos by gene assembly techniques (Agarwal *et al*., 1970), in-frame cloning method, solid phase gene assembly, chemical assembly of gene PCR-mediated gene assembly and by the *Fok*I method of gene synthesis (Mandecki *et al*., 1988). Recently, DNA shuffling was introduced as a method for in vitro recombination by combinatorial assembly of genes from random fragments generated by partial DNaseI digestion (Stemmer, 1994*a*, b), or from a mixture of oligos and random fragments (Crameri and Stemmer, 1995).

2.2.2 Gene assembly techniques

There are several approaches for the assembly of synthetic genes. The principal method for assembling DNA duplexes from synthetic oligonucleotides was the joining of complementary overlapping complexes with the aid of T4 DNA ligase. This method was first elaborated in the late 1960s by Har Gobind Khorana (Agarwal *et al*., 1970). In this approach, a series of sequentially overlapping oligonucleotides were annealed under optimized condition to form a double stranded DNA fragment containing nicks on both strands, which were sealed with DNA ligase. The first successful case of this technique was reported by Sekiya *et al*. (1979) in the synthesis of tyrosine suppressor tRNA gene using polynucleotide kinase and polynucleotide ligase to join 10 oligonucleotide segments to form the 62-nucleotide-long DNA fragment. Later, Smith and coworkers (1982) also reported the synthesis of human β-urogastrone gene from 23 oligonucleotides using T4 DNA ligase to ligate the phosphorylated oligonucleotides sequentially. Subsequently, this strategy was applied broadly to the synthesis of many genes prior to expression in *E. coli*, for

instance, human immune interferon (Tanaka and Robey, 1983) and human interferon-α_2 gene (Edge *et al.*, 1981).

Rossi *et al*. (1982) developed an alternative strategy where two sets of two long oligonucleotides (about 40 bases length) were constructed of which the complementary 3'-ends (about 10 bases) of the oligonucleotides overlapped, thus, allowed to anneal. The two constructs were completed to a full-length double strand DNA by a subsequent filling-in reaction in the presence of DNA polymerase I (Klenow fragment) with dNTPs. After polymerization, overhanging ends were generated on the double stranded DNA fragment with restriction endonucleases *Eco*R I and *Pst* I, respectively, prior to cloning into *Eco*R I-*Pst* I-digested pXJ001 plasmid.

An alternative strategy to synthesize simultaneously two DNA duplexes was reported by Sung *et al*. (1986). This strategy was called 'hybrid gene synthesis', which produced both human and mouse epidermal growth factors (hEGF and mEGF) simultaneously. In this approach, four oligonucleotides encoding hEGF and three oligonucleotides encoding mEGF were synthesized. The positive strand encoded hEGF while the negative strand encoded mEGF as complementary sequence. Upon annealing, overlapped oligonucleotides containing specific regions of mismatched bases were ligated to linearized pUC8 plasmids, yielding heteroduplex plasmids. After transformation, each plasmid strand act as a template yielding two plasmid progenies bearing two related genes.

There was another 'in vitro' method of assembling a synthetic gene reported by Narang *et al*. (1986) whereby a mixture of linearized plasmid containing six synthetic complementary oligonucleotides was directly transformed into competent cells. They found that 1 out of 100 transformants was positive in colony hybridization using one of the synthetic fragment probe (Narang *et al*., 1986). This technique is simple and not suitable for a large gene synthesis. Furthermore, the assembly efficiency of the host cells is lows, thus, multiple screening test needed to be carried out for positive result.

2.2.3 In-frame cloning methods

Another assembly technique utilizing cloning approach was reported by Adams *et al*. (1988) whereby the HIV-1 *tat* synthetic gene was generated. This technique involved the synthesis of large oligonucleotides covering one strand of the entire desired gene (positive strand) and the utilization of short complementary bridging oligonucleotides (negative strand) to direct the assembly of the large DNA fragments. The partially single-stranded intermediate was ligated to the cloning vector and transformed directly into the recipient bacterial host where the recombinant plasmid was repaired. This approach exploited the nature of *E. coli* DNA repair mechanisms to 'fill-in' breaks or gaps in a partially constructed gene. Again, this technique is simple, however, multiple screening tests needed to be carried out for positive result.

2.2.4 Solid phase gene assembly

With the advent of automated gene synthesis, the use of magnetic beads to synthesize larger DNA fragment was established. This strategy enables the rapid and cost-effective preparation of long duplex DNA region (Beattie and Fowler, 1991). Hostomsky *et al*. (1987) reported the construction of cow colostrum trypsin inhibitor gene using this technique. Basically, this approach is comprised of three steps. In the first step, an anchor oligonucleotide was covalently bound to the CNBr-activated Sepahcryl S-500 support. Next, with proper washing after each step, assembly of the gene by stepwise hybridization of the phosphorylated oligonucleotides to the immobilized complementary sequence was carried out. In the last step, a linearized vector was ligated to the assembled gene. Finally, the complete gene construct was released from the solid support with a restriction enzyme, circularized and used for transformation.

Hostomsky and coworkers (1987) found a number of advantages in this approach. Most of the enzymes or enzyme systems, including T4 DNA ligase,

polynucleotide kinase, restriction endonuclease and splicing extracts, were active in Sephacryl S-500 with bound nucleic acids comparable with that of homogenous solutions. This system enabled rapid and efficient change of buffers, as well as removal of unbound reaction products via washing steps. The gene assembly process was straight forward and purification step was simple with no need of electrophoretic separations up to the stage of harvesting the recombinant clones.

2.2.5 Chemical assembly of gene

Extensive studies of chemical reactions in DNA duplexes have made it possible to develop a procedure alternative to the enzymatic method for the assembly of extended double-stranded DNAs from oligonucleotides using a chemical reagent. Shabarova *et al*. (1990) reported a total chemical assembly genes using condensing agent (cyanogen bromide) for the assembly of 35- to 53-membered oligonucleotides to generate a 183 bp gene. The reaction is complete within several minutes at $0^{o}C$ in buffer. Shabarova and coworkers (1991) found that this approach demonstrated a number of advantages over enzymatic method, which include a high reaction rate (1 or 2 min versus 12-14 h with DNA ligase) with the absence of by-product, higher possibility of introducing various modifications and low cost of the reagents as compared with the enzymes. Nevertheless, to generate a relatively long DNA fragment, the utilization of enzyme for joining oligonucleotides together was essential (Shabarova *et al*., 1991).

2.2.6 PCR-mediated gene assembly

Jayaraman and Shah (1987) introduced a PCR-mediated gene synthesis approach that involved a single-step ligation of overlapping oligonucleotides comprising the entire gene followed by PCR amplification of the crude ligation mixture, with two outer primers, to generate the full-length gene. With this method, they successfully synthesized several genes, including the horseradish peroxidase

gene (Jayaraman *et al*., 1991). Jayaraman and Puccini (1992) applied the PCR-mediated gene synthesis approach to the assembly of oligonucleotides that correspond to only one strand of the gene. In that study, a mixture containing 3 long olgonucleotides (>100 bp) corresponding to only one of the strands, short oligonucleotides (~20 bp) that were complementary at the junction of the long oligonucleotides and two outer primers (~20 bp) was subjected to PCR to generate a full-length double-stranded gene.

Dillon and Rosen (1990) also developed a rapid method for the construction of synthetic gene using the PCR strategy. Multiple overlapping oligonucleotides were used to generate synthetic DNA through several sequential rounds of Klenow-based PCR amplification. In their report, two-step PCR protocol using the thrmostable *Taq* polymerase was carried out to create a synthetic gene for the HIV-2 Rev protein. Upon annealing, 4 overlapping oligonucleotides (each 105 nucleotides in length with 20-25 complementary sites) served as 'template' as well as 'primers' in the primary PCR amplification. Then, the PCR product was added as template into a mixture containing 5'- and 3'-flanking primers for secondary PCR amplification to generate a full-length double-stranded gene.

Another simple and rapid single-step assembly gene approach involving PCR by using a thermostable polymerase for the filling-in reactions of overlapping complementary oligonucleotides was reported by Stemmer *et al*. (1995). This strategy was derived from DNA shuffling (Stemmer, 1994b) independent of DNA ligase activity but instead relied on DNA polymerase for longer DNA construction. In this approach, a 1.1 kb DNA fragment containing TEM-1 β-lactamase-encoding gene (*bla*) was assembled in a single-step polymerase chain reaction from a total of 56 oligonucleotides, each 40 nucleotides in length. Norazmi and coworkers (1999) successfully applied this technique for the cloning of a malarial epitope into *Mycobacterium smegmatis*. Barnes and Frawley (2003) carried out a comparison study between their method and that of Stemmer *et al*. (1995) and reported that the

smearing phenomenon at the first step assembly PCR product could be reduced using less amounts of DNA (1 pmole each oligonucleotide in 100 µL reaction) and longer extention time (20 minutes) in less PCR cycles (25 cycles).

2.2.7 ***Fok*I method of gene synthesis**

Mandecki and Bolling (1988) developed a fast and simple method for synthesis of a gene, or any DNA fragment with a defined sequence. This method was applied to the synthesis of a gene fragment encoding the N-terminal 143 amino acid residues of human immunodeficiency virus transmembrane protein. This method is based on the observation that large (approximately 100 bp long) inserts can be cloned into plasmid using a technique of oligonucleotide-directed double-strand break repair. The procedure involves tramsformation of *E. coli* with a denatured mixture of an insert-carrying oligo and linearized plasmid DNA (Mandecki, 1986). The nucleotide sequences are inserted between two *Fok*I restriction nuclease sites in one of four pUC-derived plasmids. Since *Fok*I makes a staggered double-strand break at a DNA site 9 and 13 nucleotides away from its recognition site, upon cleavage of the plasmid DNA with *Fok*I, a restriction fragment is liberated that by design contains unique 4-nucleotide-long 5'-protruding ends. The uniqueness of ends permits efficient and directed simultaneous ligation of several restriction fragments to form a gene. This method offers flexibility due to the modular-type assembly and does not require any restriction sites within the constructed gene. The sequence error rate is low and just about one error per 4000 bp of DNA cloned (Mandecki and Bolling, 1988).

2.3 ***Bacillus megaterium* expression system**

2.3.1 **General features of *Bacillus megaterium***

Bacillus megaterium has fascinated microbiologist since it was first described over 100 years ago. It is interesting especially because of its physiology, unusual

and useful enzymes and products, also the wide range of ecological habitats. It is also capable of sporulation, a simple cell differentiation cycle that serves as a model system for understanding gene regulation during temporal and morphological development. Moreover, the source of the significant name "*megaterium*" was the large size of the vegetative cells and spores make it especially amenable to morphological analysis (Harwood *et al*., 1990; Priest, 1989).

B. megaterium is able to grow on a wide variety of carbon sources and has, thus, been found in many ecological niches, such as waste from meat industry or petrochemical effluents. Also documented has been the degradation of persistent insecticides by *B. megaterium* (Sexana *et al*., 1987; Selvanayagam and Vijaya, 1989) offering potential applications as detoxifying agent. One of the genetic regulatory elements for carbon utilization is the xylose operon. It has been described by Rygus and Hilen (1991).

Several *B. megaterium* proteins are of importance. A family of P-450 cytochrome monooxygenases is similar to eukaryotic P-450 playing a role in many diseases. Industrial applications of enzymes excreted by *B. megaterium* are diverse, starting from amylases used in the bread industry to penicillin amidase and steroid hydrolases, which is used for generation of new synthetic antibiotics (Vary, 1994). It is the major aerobic producer of vitamin B_{12} and is one of the organisms involved in fish spoilage.

During the 1980s, genetic techniques of transduction, plasmid transformation, protoplast fusion and transportation became developed enough in *B. megaterium* to apply them to the study of many of its metabolic and developmental functions. Moreover, it is increasingly used as a host to produce foreign genes since it has been found to express, secrete and process foreign proteins without degradation.

2.3.2 Taxonomic position of *B. megaterium*

In order to understand the species, it is necessary to be aware of the great diversity in the taxonomy of *Bacillus*, and within *B. megaterium* (Claus *et al*., 1989; Priest, 1993). As can be seen from Figure 2.5, it is within the *B. subtilis* group, but much more distantly related to *B. subtilis* than *B. licheniformis*, *B. cereus*, *B. anthracis* or *B. pumilus* by 16S rRNA sequence analysis (Ash *et al*., 1991; Priest, 1993).

Major research strains of *B. megaterium* include QM B1551, KM, 216, DSM 319, ATTC 10778 and ATTC 19213. Strains QM B1551, 216 and IWG3, as well as the plasmidless strains PV361, DSM 319 and VT1660 are well-known industrially (Vary, 1994).

2.3.3 The cell structure of *B. megaterium*

B. megaterium is one of a few *Bacillus* strains that have a cell width greater than 1 µm (Vary, 1994). This large size and its ability to take up diaminopimelic acid have been exploited in several morphological studies. It has been used effectively to study cell wall synthesis as well as membrane and spore structure as reviewed recently (Archibald *et al*., 1993).

Walls of Gram-positive bacteria are dynamically variable and flexible structure that enclose and protect the underlying cytoplasmic membranes. They are intimately involved in cell growth and morphogenesis, cell division and its environment, and movement of materials into and out of the cell. During growth, the wall has to enlarge and change shape to accommodate the exponentially increasing volume of the cell and must divide to allow the formation of two daughter cells. The forces that drive wall growth and the ways which growth is accomplished without compromising the wall's strength is the ability to protect the protoplast (Koch and Burdett, 1986).

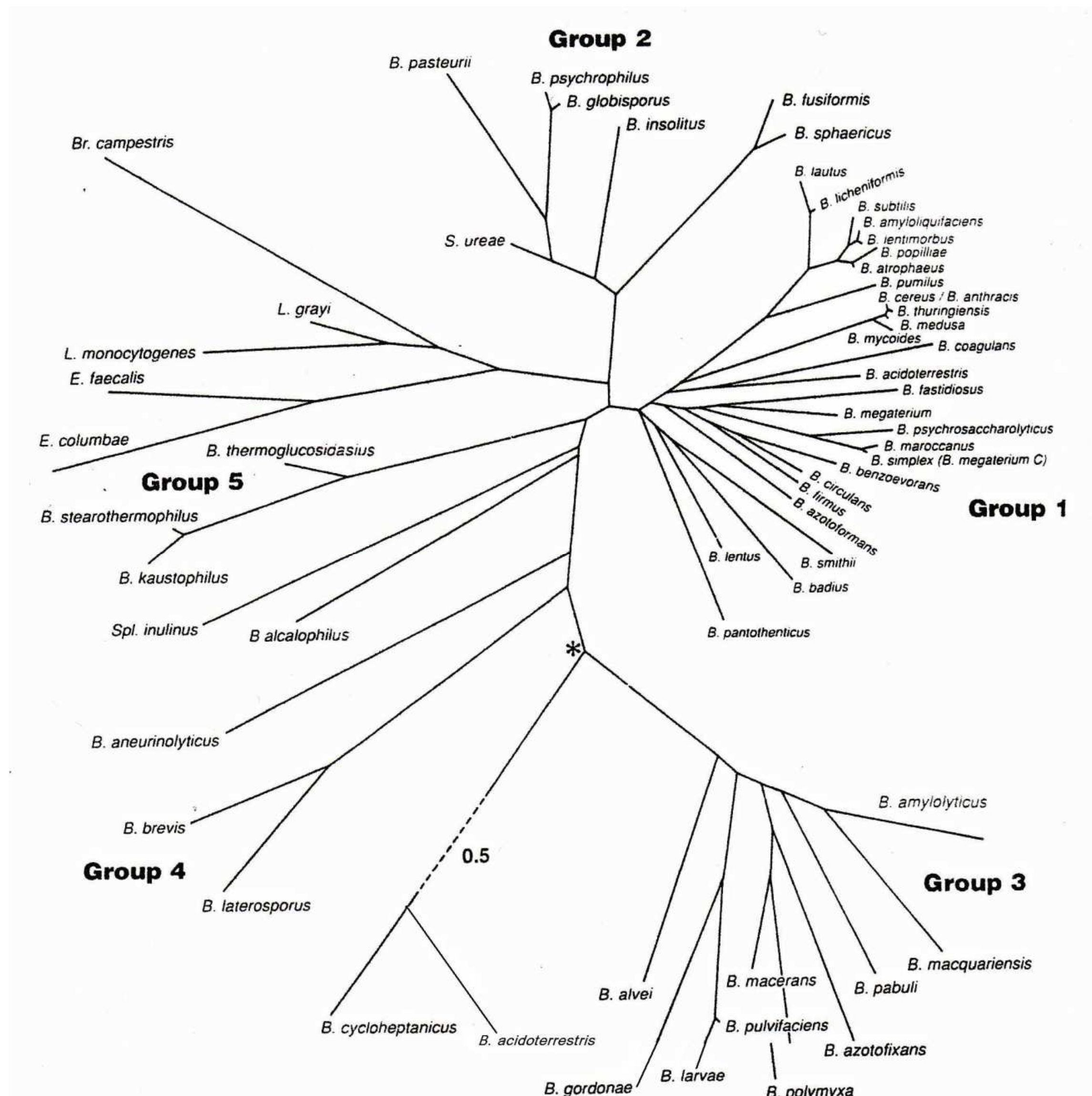

Figure 2.5 Phylogenetic trees of some members of the genus *Bacillus* based on 16S rRNA sequence analysis (Ash *et al*., 1991). Picture adapted from Vary (1994).