

CLONING AND EXPRESSION OF
TaqI RESTRICTION ENDONUCLEASE GENE
USING pMAL-c2 VECTOR SYSTEM
IN DIFFERENT STRAINS OF *Escherichia coli*

by

Elif ÖZKIRIMLI

B.S. in Ch.E., Boğaziçi University, 1996

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science
in
Chemical Engineering

Boğaziçi University

1998

Bogazici University Library



39001100122269

14

ACKNOWLEDGMENTS

The research presented in this thesis was conducted between September 1996 and June 1998 under the supervision of Prof. Dr. Betül Kırdar and Prof. Dr. Amable Hortaçsu.

I offer my deepest thanks to Prof. Dr. Betül Kırdar for her perpetual support and counsel throughout this research and for her patience in evaluating the numerous drafts of this thesis. I wish to express once again how fortunate I feel to work under her supervision and to be her student. Her kindness has made research not only an invaluable education and but also a cherished pleasure .

I would like to express my sincere gratitude to Prof. Dr. Amable Hortaçsu for her valuable suggestions throughout this research and for her friendly encouragement to pursue my ideals.

I would like to thank Prof. Dr. Zeynep İlşen Önsan for devoting her precious time to reading my thesis and for setting a model for me of the academician that I wish to be.

It is a pleasure for me to thank Assoc. Prof. Kutlu Ülgen for reading and evaluating my thesis and for her friendly and sincere attitude during my undergraduate and graduate studies.

I would like to thank Prof. Dr. Dilek Çalgan for kindly evaluating and commenting on my thesis.

I am indebted to the Biotechnology group for the fun we had and for the hard times we shared in our small realm.

Finally, I would like to thank my mother for being my best friend, my father for his encouragement to be who I want to be, my brother for his compassion, sweetness and for making me laugh when I cry, my grandmother for her knees, and Alptuğ for his love and for what we are. I dedicate this thesis to them.

ABSTRACT

The thermostable *TaqI* restriction endonuclease from *Thermus aquaticus* was cloned into *Escherichia coli* for overproduction and single step affinity chromatography purification. The gene was cloned into pMAL-c2 vector, resulting in the cytoplasmic expression of a MBP-*TaqI* restriction endonuclease fusion protein.

In the first part of this study, *TaqI* restriction endonuclease gene was amplified using *Taq* DNA polymerase and cloned into *E. coli*. Although the expected size fusion protein was produced, no endonuclease activity was obtained. Growth condition manipulations and expression in different strains did not result in the production of an active enzyme. Subsequent sequencing of the gene indicated a mutation at position 410 that might abolish activity.

In the second part, cloning was repeated after amplification of the gene with *Pfu* DNA polymerase which has proofreading activity. A colony coding for a protein with the expected size and activity was identified. Upon optimization of induction time, higher titer was obtained when the culture was induced at an absorbance of 0.9. Expression in three different strains of *E. coli*, indicated total recovery was higher in TB1 cells at 300,000 Units/liter of culture volume as compared to 100,000 Units/liter in XL1-Blue cells and 150,000 Units/liter in ER2508 cells.

ÖZET

Thermus aquaticus'un ısılkararlı *TaqI* restriksiyon endonükleazı yüksek miktarda üretim ve tek aşamalı afinite kromatografisi ile saflaştırma için *Escherichia coli*'ye klonlandı. Gen, pMAL-c2 vektörüne klonlanmış ve MBP-*TaqI* restriksiyon endonükleaz füzyon proteinin sitoplazmik ekspresyonu sağlandı.

Bu çalışmanın ilk aşamasında *Thermus aquaticus* DNA'sındaki *TaqI* restriksiyon endonükleaz geni *Taq* DNA polimeraz kullanılarak PCR metodu ile çoğaltıldı ve *Escherichia coli*'ye klonlandı. Doğru boyda füzyon proteininin yüksek miktarda üretilmesine karşın, endonükleaz aktivitesi elde edilemedi. Ortam koşullarının değiştirilmesi ve değişik *Escherichia coli* suşlarında ekspresyon sonucunda da aktif restriksiyon endonükleazı elde üretilmedi. Bunun üzerine, genin dizi analizi yapıldı ve 410uncu baz çiftinde aktiviteye engel olabilecek bir mutasyon saptandı.

Çalışmanın ikinci aşamasında *TaqI* restriksiyon endonükleaz geni *Pfu* DNA polimeraz kullanılarak çoğaltıldı. Doğru boy ve aktivitede füzyon proteini üreten bir koloni seçildi. İndükleme zamanının optimizasyonu sonucunda kültürün abzorbansı 0.9'a ulaşınca yapılan indüksiyon sonucunda daha yüksek restriksiyon endonükleaz aktivitesi elde edildiği belirlendi. TB1, XL1-Blue ve ER2508 adlarındaki üç değişik *Escherichia coli* suşunda ekspresyon sonucunda 1 litre TB1 hücre kültüründen 300,000 Ünite enzim aktivitesi elde edilirken. 1 litre XL1-Blue hücre kültüründen 100,000 Ünite enzim aktivitesi ve 1 litre ER2508 hücre kültüründen 150,000 Ünite enzim aktivitesi elde edildi.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	page iii
ABSTRACT	iv
ÖZET	v
LIST OF FIGURES	ix
LIST OF TABLES	x
LIST OF ABBREVIATIONS	xi
1. INTRODUCTION	1
2. OPTIMIZATION OF RECOMBINANT PROTEIN EXPRESSION IN <i>Escherichia coli</i>	3
2.1. Strain Selection	4
2.2. Transcriptional Regulation	6
2.3. Plasmid Stability	10
2.4. Translational Regulation	11
2.5. Codon Usage	12
2.6. Cytoplasmic Expression	13
2.6.1 Proteolysis	14
2.6.2 Inclusion Body Formation	15
2.7. Periplasmic Expression	18
2.8. Export of the Recombinant Protein to the Culture Medium	18
2.9. Fermentation Conditions	19
2.10. Fusion Protein	24
2.10.1 Maltose Binding Protein	31
2.10.2 pMAL-c2 Fusion and Purification System	38
3. <i>TaqI</i> RESTRICTION ENDONUCLEASE	40
4. MATERIALS	44
4.1. Bacterial Strains and Plasmids	44
4.2. Chemicals	44
4.3. Buffers and Solutions	45
4.3.1 Bacterial Cell Media	45
4.3.2 Buffers and Solutions for Plasmid DNA Isolation	46

4.3.3	Buffers used in PCR	47
4.3.4	Agarose Gel Electrophoresis Buffers and Solutions	48
4.3.5	Buffers and Solutions for the Isolation of DNA from Low-Melting Point Agarose	48
4.3.6	Restriction Enzyme Digestion Buffers	49
4.3.7	Buffers Used for Detection of Expression and Affinity Column Chromatography	50
4.3.8	Polyacrylamide Gel Electrophoresis Buffers and Solutions	50
4.3.9	Sequencing Buffers and Solutions	52
4.3.10	DNA and Protein Size Markers Used in Electrophoresis	52
4.3.11	Restriction Enzymes and Their Recognition Sites	53
4.4.	Primers for Amplification of the <i>TaqI</i> restriction endonuclease gene	53
4.5.	Laboratory Equipment	55
5.	METHODS	57
5.1.	Preparation of Preculture	57
5.2.	Isolation of Plasmid DNA	58
5.2.1	Large-Scale Plasmid DNA Isolation	58
5.2.2	Mini-Prep Method for Small-Scale Plasmid DNA Isolation	59
5.3.	Determination of Quantity and Quality of DNA	60
5.3.1	Spectrophotometric Measurement	60
5.3.2	Agarose Gel Electrophoresis	61
5.4.	Cloning of <i>TaqI</i> Restriction Endonuclease Gene into pMAL-c2 Vector	61
5.4.1	PCR Amplification of <i>TaqI</i> Restriction Endonuclease Gene	61
5.4.2	Isolation of PCR Products from Low-Melting Temperature Agarose Gels	62
5.4.3	Digestion of the Insert with Restriction Endonucleases	63
5.4.4	Digestion of the pMAL-c2 Vector with Restriction Endonucleases	63
5.4.5	Ligation Reaction	64
5.5.	Transformation of <i>Escherichia coli</i> Cells	64
5.5.1	Preparation of Competent Cells	64
5.5.2	Electroporation	65
5.6.	Screening System	66

5.6.1	α -Complementation	66
5.6.2	Plasmid Size	66
5.6.3	Restriction Enzyme Analysis	67
5.6.4	PCR Analysis	67
5.7.	Detection of the Expression of the MBP- <i>TaqI</i> Gene Fusions	67
5.7.1	Detection of MBP- <i>TaqI</i> Fusion Proteins in the Cytoplasm	67
5.7.2	Purification of <i>TaqI</i> from the Medium	68
5.8.	Chromatographic Purification of <i>TaqI</i> Restriction Endonuclease	69
5.8.1	Purification of the Fusion Protein on Cross-Linked Amylose	69
5.8.2	Cleavage with Factor Xa	69
5.9.	Determination of Protein Concentration	70
5.10.	Measurement of <i>TaqI</i> Restriction Endonuclease Activity	70
5.11.	SDS-Polyacrylamide Gel Electrophoresis of Proteins	71
5.12.	DNA Sequencing	72
5.13.	Determination of Growth Characteristics	74
6.	RESULTS AND DISCUSSION	75
6.1.	Amplification of <i>TaqI</i> Restriction Endonuclease Gene by <i>Taq</i> DNA Polymerase	75
6.1.1	Cloning of the Gene into pMAL-c2 Expression Vector	75
6.1.2	Expression of the Fusion Protein in <i>Escherichia coli</i> TB1 Cells	79
6.1.3	Sequencing of the Insert	83
6.2.	Amplification of <i>TaqI</i> Restriction Endonuclease Gene by PCR with <i>Pfu</i> DNA Polymerase	84
6.2.1	Optimization of PCR Conditions	84
6.2.2	Cloning of the <i>TaqI</i> Restriction Endonuclease Gene into pMAL-c2 Expression Vector	85
6.2.3	Expression of the MBP- <i>TaqI</i> Fusion Protein in <i>Escherichia coli</i> TB1 cells	89
6.2.4	Optimization of Induction Conditions for Recombinant Colony c22	93
6.2.5	Expression of the Fusion Protein in <i>Escherichia coli</i> XL1-Blue and ER2508 strains	95
7.	CONCLUSION AND RECOMMENDATIONS	98
8.	REFERENCES	103

LIST OF FIGURES

	page
FIGURE 2.2. pMAL-c2 expression vector	39
FIGURE 3.1. The Nucleotide Sequence of the <i>TaqI</i> Restriction Endonuclease Gene	43
FIGURE 6.1. Electrophoretic analysis of the PCR product amplified by <i>Taq</i> DNA polymerase on 1.2 per cent agarose gel	76
FIGURE 6.2. Strategy for cloning of <i>TaqI</i> restriction endonuclease gene into pMAL-c2 expression vector	77
FIGURE 6.3. Rapid analysis of plasmid size by cracking method	78
FIGURE 6.4. Electrophoretic analysis of the PCR products in suspected colonies	79
FIGURE 6.5. SDS-PAGE analysis of the proteins extracted from the suspected colonies and subjected to amylose affinity chromatography	80
FIGURE 6.6. SDS-PAGE analysis of the column fractions obtained by amylose affinity chromatography	81
FIGURE 6.7. SDS-PAGE analysis of cytoplasmic proteins obtained from culture induced with decreasing concentrations of IPTG	82
FIGURE 6.8. Electrophoretic analysis of the PCR product amplified by <i>Pfu</i> DNA polymerase	85
FIGURE 6.9. Electrophoretic analysis of the ligation reaction	86
FIGURE 6.10. Rapid analysis of the plasmid size by cracking method	87

FIGURE 6.11.	Restriction Map of the Recombinant pMAL-c2 Vector	88
FIGURE 6.12.	Restriction analysis of the recombinant colony harboring the plasmid pC22	89
FIGURE 6.13.	Analysis of <i>TaqI</i> restriction endonuclease activity in the cytoplasmic extracts of the suspected colonies.	90
FIGURE 6.14.	SDS-PAGE analysis of the crude and heat treated extracts from colony c22 induced for a period of 4 hours at midexponential phase	91
FIGURE 6.15.	SDS-PAGE analysis of the fusion proteins obtained by expression in late exponential phase	93
FIGURE 6.16.	Growth curve of the recombinant colony 22 with and without induction	94

LIST OF TABLES

	page
TABLE 2.1. Strain genotypes and their uses employed in recombinant protein overproduction	5
TABLE 2.2. A few promoters used for the high level expression of genes in <i>Escherichia coli</i> with their advantages and disadvantages	8
TABLE 2.3. General factors to be considered in optimization of heterologous protein overproduction in <i>Escherichia coli</i>	22
TABLE 2.4. Protein specific factors to be considered in optimization of heterologous protein overproduction in <i>Escherichia coli</i>	23
TABLE 2.3. Tail Types	26
TABLE 2.4. Cleavage Methods to Cleave Fusion Proteins	28
TABLE 4.1. Specifications of the primers	54
TABLE 6.1. <i>TaqI</i> restriction endonuclease activity of the crude extracts of the recombinant colony induced at OD ₆₀₀ = 0.9 for five hours	92
TABLE 6.2. <i>TaqI</i> restriction endonuclease activity of the crude extracts of the recombinant colony induced at different phases of growth	95
TABLE 6.3. <i>TaqI</i> restriction endonuclease activity of the crude extracts of three different strains	96

LIST OF ABBREVIATIONS

APS	Ammonium persulfate
Da	Daltons
ddNTP	dideoxyribonucleotide triphosphate
dNTP	deoxyribonucleotide triphosphate
EDTA	ethylenediaminetetraacetic acid
KDa	Kilodaltons
MBP	Maltose Binding Protein
OD	Optical Density
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
SDS	Sodium Dodecyl Sulfate
TEMED	N,N,N',N'-tetramethylethylenediamine
TQEcoRI	reverse primer
TQXmnI	forward primer
U	Unit
UV	Ultraviolet

1. INTRODUCTION

TaqI restriction endonuclease is a type II endonuclease which cleaves the DNA at the palindromic sequence TCGA. It is a thermostable enzyme, exhibiting maximum activity at 65 °C, isolated from the thermophilic microorganism *Thermus aquaticus*. The recognition site for *TaqI* restriction endonuclease exists at most of the polymorphic regions of the human genome. Therefore, *TaqI* restriction endonuclease is widely used in molecular diagnosis. Moreover, it is a model enzyme to study DNA-protein interactions. Therefore, obtaining large quantities of the enzyme is important. However, the isolation of *TaqI* restriction endonuclease from its native host *Thermus aquaticus* requires high fermentation costs and results in very low yields upon purification. Cloning and expression of the *TaqI* restriction endonuclease gene in *Escherichia coli* is considered to be beneficial for the high yield recovery of the enzyme.

Expression of genes with a fusion partner allows rapid purification of the target protein if the fusion partner can serve as an affinity tag. pMAL vectors express the target protein as a fusion protein with Maltose Binding Protein (MBP) fusion partner. Affinity of MBP to crosslinked amylose is employed to purify the protein. The MBP fusion protein is targeted to the periplasm with the pMAL-p2 vector, while the deletion of the signal sequence of MBP in the pMAL-c2 vector results in expression and accumulation of the target protein in the cytoplasm.

TaqI restriction endonuclease has previously been cloned into pMAL-p2 and pMAL-c2 expression vectors. As a result of periplasmic expression with pMAL-p2 vector, the MBP-*TaqI* restriction endonuclease fusion protein was purified by amylose column chromatography from the periplasmic and cytoplasmic protein pool and by salting out with ammonium sulfate from the medium protein pool. The activity of the fusion protein isolated from the medium was found to be 2 U/μl. (Özdinler, 1996) On the other hand, the

cytoplasmic expression of MBP-*TaqI* restriction endonuclease fusion protein resulted in production of the fusion protein with the expected size, but without endonuclease activity (Enünlü, 1996). It was suggested in this study that the recombinant construct might have the wrong gene sequence resulting in abolishment of activity.

Consequently, in the present study, the *TaqI* restriction endonuclease was cloned into pMAL-c2 vector. In the first part of this study, the *TaqI* restriction endonuclease gene was amplified by PCR with *Taq* DNA polymerase and cloned into pMAL-c2 expression vector. Although the fusion protein of the expected size was produced with a high yield, no endonuclease activity could be obtained. The production of an active restriction endonuclease could not be achieved by manipulation of growth conditions or expression in different strains. Subsequent sequencing analysis of the insert within the plasmid indicated a substitution within the gene that might abolish activity and hence cloning was repeated with higher fidelity of amplification.

In the second part of this study, *TaqI* restriction endonuclease gene was amplified using *Pfu* DNA polymerase which has proofreading activity. The cloned *Escherichia coli* cells produced the correct size fusion protein with endonuclease activity. Further experiments were conducted to optimize IPTG induction time and to compare the level of expression in three different strains of *Escherichia coli*. The fusion protein was purified using amylose column chromatography upon expression.

2. OPTIMIZATION OF RECOMBINANT PROTEIN EXPRESSION IN *Escherichia coli*

One of the practical aims of genetic engineering is to force a host cell to produce an optimal level of one or more desirable proteins. Genetically engineered cells are used for the production of large amounts of a recombinant protein or they are used for the production of intermediate enzymes that facilitate the synthesis of an end product in the host cell. In the latter case, genetic engineering techniques are employed, for example, to produce an enzyme that speeds up the limiting reaction in the production of a desired product (Georgiou, 1988).

In a commercial process, the overall objective is to maximize the yield of a desired product and to increase the scale of production if possible. However, the overproduction of a recombinant protein usually disrupts the coordination of metabolic processes and may ultimately result in cell death, and is, hence, unfavorable. Therefore, the maximization problem of the chemical engineer becomes an optimization problem in which the physiology of the genetically engineered cells must be taken into account (Georgiou, 1988).

Recombinant proteins can be produced in high levels in various expression systems, and the choice depends on (a) cell growth characteristics; (b) expression levels; (c) intracellular and extracellular expression; (d) posttranslational modifications; (e) biological activity of the protein when expressed with that specific expression system; (f) degree of authenticity required and regulatory issues in the production of therapeutic proteins (Makrides, 1996).

The choice of the host for the expression of a recombinant protein is the most important step in the selection of the expression system. For large-scale productions, the

host must propagate fast with simple nutritional requirements. It must not be pathogenic and must not produce any toxins that may contaminate the product. The cells should be able to grow to high cell densities to maximize protein production. *Escherichia coli* is by far the most widely used host for the high level production of recombinant proteins because it grows fast on simple nutrients, and there is extensive knowledge on the genetics and microbiology of *Escherichia coli* which makes it easier to model, predict and use in large scale productions (Georgiou, 1988). *Escherichia coli* is also versatile because there are many high expression vectors available for it. However, there are some drawbacks to this host, too. It is unable to perform posttranslational modifications required for eukaryotic proteins, it cannot facilitate disulfide bond formation necessary for the three dimensional conformation of some proteins and some proteins cannot assume their correct three dimensional structure that is essential for biological activity and instead precipitate inside the cell forming inactive protein aggregates called inclusion bodies. It cannot secrete the protein into the culture medium, which would be beneficial from a biochemical engineering point for the simple purification of the target protein. The stability and translational efficiency of mRNA, the ease of protein folding, degradation of the protein by host cell proteases, differences of codon usage between the foreign gene and *Escherichia coli* and the toxicity of the protein to the host determine the efficiency of expression of a protein in *Escherichia coli* (Gold, 1990; Marino *et al.*, 1989).

2.1. Strain Selection

The choice of strain for expression of recombinant protein depends on the target protein, the expression vector, the fermentation conditions and the particular aim of overproduction. For example, for proteins that are highly susceptible to proteolysis, it may be necessary to use protease deficient strains. When it is aimed to secrete the target protein to the culture medium, periplasmic “leaky” strains may be employed (Miller *et al.*, 1998). The genotype of the host is a useful guide in selection of the best strain for expression of a specific gene (Table 2.1).

TABLE 2.1. Strain genotypes and their uses employed in recombinant protein overproduction (Maniatis *et al.*, 1989)

Genotype	Use
<i>endA</i>	Activity of nonspecific Endonuclease I abolished. DNA preparations are thought to be of higher quality when prepared from <i>endA</i> strains.
F	A low-copy number of self-transmissible plasmid. F' factors carry portions of the <i>Escherichia coli</i> chromosome, most notably the <i>lac</i> operon and <i>proAB</i> on F' <i>lac-proA⁺B⁺</i> .
<i>lacI^q</i>	Overproduces the <i>lac</i> repressor, turning off expression from <i>lac</i> promoter
<i>lacZ</i>	β -galactosidase activity abolished.
<i>lacY</i>	Lactose permease activity abolished. $\Delta(lacZ)$ M15 expresses a fragment that complements the <i>lac</i> α - fragment encoded by many vectors. These vectors will yield blue color on X-Gal only if the host carries Δ M15.
<i>lon</i>	Activity of a protease responsible for degrading aberrant proteins abolished. Some eukaryotic proteins are stabilized in <i>lon</i> strains.
<i>malB</i>	The <i>malB</i> region encompasses the genes <i>malE</i> <i>FB</i> and <i>malK</i> <i>lamB</i> <i>malB</i> . $\Delta(malB)$ deletes most or all of this region and eliminates expression of Maltose Binding Protein (MalE).
<i>mcrA</i> , <i>mcrBC</i>	These affect methylcytosine-specific restriction systems. DNA containing methylcytosine in some sequences is restricted by <i>Mcr⁺</i> . <i>Dcm⁻</i> modified DNA is not restricted <i>Mcr⁺</i> . $\Delta(mcrC\ mrr)$ deletes six genes: <i>mcrC mcrB⁻ hsdS hsdM hsdR mrr</i> , <i>mcrA</i> is lost with <i>e14</i> .
<i>mrr</i>	A restriction system requiring cytosine or adenine methylation abolished; however, <i>dam⁻</i> , <i>dcm⁻</i> , or <i>EcoK</i> -modified DNA is not restricted by <i>Mrr⁺</i> . The methylcytosine-dependent activity is also known as <i>McrF</i> (3).
<i>recA</i>	Homologous recombination abolished; structural stability of the plasmid increases
<i>rpoH</i>	(also known as <i>htrR</i>) Lack of this heat-shock transcription factor abolishes expression of some stress-induced protease activities in addition to <i>lon</i> . Some cloned proteins are more stable in <i>rpoHam supCts</i> strains at high temperature.
<i>supE</i>	Strains carry a glutamine-inserting amber (UAG) suppressor tRNA required for growth of some phage vectors.
<i>traD</i>	Reduces the self-transmissibility of the F factor.
(Φ 80)	The cell that carries the lambdoid prophage Φ 80. A defective Φ 80 prophage carrying the <i>lac</i> M15 deletion is present in some strains.

TB1 has been reported to be the most widely used strain with pMAL expression vector. It is an α -complementing host, allowing selection of recombinant plasmids harboring an insert.

XL1-Blue is often used for high quality plasmid DNA purifications. It carries the gene endA1, preventing nonspecific activity of endonuclease 1 on DNA. It also carries the recA gene and hence homologous recombination is abolished increasing plasmid stability. This is particularly desirable when working with sequences containing direct repeats larger than 50 bp. Moreover, it is an α -complementing host. (Maniatis *et al.*, 1989)

ER2508 is a protease deficient strain. The *lon* protease is not expressed and hence this strain is most frequently used when proteolysis is a problem. The *malB* gene deletion includes the *malE* gene, so this strain does not produce any chromosomal MBP. This is beneficial when MBP is expressed by the expression vector as a fusion partner to a target protein. All the MBP produced is part of the fusion partner, increasing the efficiency of amylose column chromatographic purification. The proteins secreted to the periplasm can be released into the culture medium when expressed in a periplasmic leaky *Escherichia coli* host such as E609L.

JM109 is a frequently used strain. It carries the *lacI*^q gene which strongly represses the *lac* promoter and derivatives such as the *tac* promoter until induction. This is useful in preventing leaky expression of toxic recombinant proteins. However, this strain has been shown to be inferior to BL21 strain with regard to low acetate concentration maintenance (Shiloach *et al.*, 1996).

2.2. Transcriptional regulation

The gene coding for a desired protein acts as a template for the production of that protein. The gene may be inserted into the host chromosomal DNA for expression, however, the amount of mRNA determines the rate of expression of a protein. Therefore, if the number of copies of the target gene in the host can be increased, the rate of product

formation will be increased. To achieve this aim, the gene is inserted into high copy number plasmids that are propagated from one generation to the next and exist in several copies per cell. Plasmids containing a desired foreign gene and designed specifically to effect a high level of product synthesis are called expression vectors (Georgiou, 1988). An ideal plasmid should have a low molecular weight for easier handling and higher copy number, and should have single sites for a large number of restriction endonucleases. The expression vector should also harbor a gene that confers a phenotypic trait to simplify selection and propagation of the recombinant colonies. This is usually an antibiotic resistance gene. Hence, only those cells harboring the plasmid, i.e. recombinant cells, can grow in the presence of antibiotic supplemented to the medium (Old and Primrose, 1989). The essential components of an expression vector are the promoter, the ribosome binding site, the transcription terminator and the origin of replication.

The gene of interest must be placed under the control of an *Escherichia coli* promoter. Two types of promoters exist that initiate transcription; constitutive and regulated. Constitutive promoters are always turned on and the rate of mRNA synthesis is not modulated. On the other hand, a regulated promoter controls the level of mRNA synthesis and hence production of the protein under its control. Repressible or inducible promoters may be employed to control gene expression. The inducible promoters are usually preferred because they are induced by the addition of a substrate or by temperature shifts; however, repressible promoters require the dilution of the repressor from the growth medium by changing the culture medium, which is not practical (Georgiou, 1988).

There are basically four desirable features of a regulated promoter. First, it should be strong to increase the yield of the target protein to as high as 30 per cent of the total cellular protein (Goldstein and Doi, 1995). It must have a low basal expression level, meaning it must be tightly regulated and not “leaky”. Leaky protein expression before induction of the cells may result in the accumulation of the usually toxic recombinant protein, causing cell death due to excess toxicity or metabolic burden. Plasmids also lose stability as a result of expression of the foreign protein, and hence a decrease in growth rate occurs and recombinant protein production may drop significantly (Bentley *et al.*, 1990). The promoter must be easily transferable to other *Escherichia coli* strains to study expression in different strains. Finally, its induction must be simple and cost effective. For

laboratory research, strong promoters such as *tac* or *trc* are usually employed that may be induced by the addition of IPTG. However, IPTG is both toxic and very expensive, making it unsuitable for production of human therapeutic proteins and for large-scale use (Figge *et al.*, 1988). For large-scale production, promoters induced by thermal induction or chemical inducers are used (Table 2.2).

TABLE 2.2. A few promoters used for the high level expression of genes in *Escherichia coli* with their advantages and disadvantages (Weickert *et al.*, 1996)

Promoter	Induction	Characteristics
<i>lac</i>	IPTG, Thermal	Low level (about 5%), leaky, expression
<i>trp</i>	Trp starvation, indole acrylic acid	Leaky expression
<i>phoA</i>	Phosphate starvation	Limited media options
<i>recA</i>	Nalidixic acid	
<i>tac</i>	IPTG, Thermal	High level, leaky expression, induction results in cell death,
<i>trc</i>	IPTG	
T7	Thermal	Very high level, leaky expression
<i>cad</i>	acid pH	

Improvements have been made in promoter control that increase efficiency of transcription and provide tighter control at a lower cost some of which are listed in Table 2.2. A temperature sensitive mutant *lacI* gene that encodes a thermosensitive *lac* repressor is a recent development with which *lac* based promoters can be induced thermally with a tight regulation (Ward *et al.*, 1995). However, although thermal induction is cost effective and simple, it may result in the increased production of proteases, which degrade the product. The development of feed strategies for tight low-cost control of *trp* promoter in large scale fermentation and the development of thermally regulated, runaway replication plasmids that allow more efficient repression of the *trp* promoter are noteworthy improvements in promoter control. Some expression vectors use the moderately strong arabinose inducible *araB* promoter and the *araC* repressor. The *araB* promoter can both be

tightly repressed and modulated over a wide range of inducer conditions which is advantageous for the control of gene expression. The *araB* expression systems could be used with other promoters to independently regulate expression of two or more recombinant proteins in *Escherichia coli*. The regulatory region of the *cadA* gene was used to construct a pH inducible expression system. This is advantageous since bulk chemical inducers such as acids and bases can be employed. Another advantage of pH induction over temperature induction is that high temperatures induce formation of inclusion bodies and heat shock proteins including *lon* protease, which result in degradation (Weickert *et al.*, 1996).

The transcription terminator is located downstream of the target gene and is a signal to terminate transcription. Its presence prevents the transcription and translation of unnecessarily long gene segments which otherwise would increase the metabolic burden to the host. Moreover, undesirable secondary structures may form in the transcript due to overlong transcripts, which would reduce efficient translation. Finally, essential genes such as the ampicillin resistance gene or the origin of replication may be read through if termination is not present or is weak and hence they may not be transcribed and translated efficiently. If, for example, the origin of replication is read through, plasmid copy number decreases remarkably (Old and Primrose, 1988). Similarly, a transcription terminator placed upstream of the promoter will inhibit transcription through the promoter unless the promoter is induced and thus minimize background transcription. It also serves as a protective element composed of stem loop structures, protecting the mRNA from degradation, enhancing its stability and subsequently increasing the level of protein production. The use of a reversible transcription termination system derived from phage λ to control leaky expression of T7 RNA polymerase and use of tandem transcription terminators that prevent read through transcription into vector sequences are some of the improvements in the transcription termination (Nishihara *et al.*, 1994).

2.3. Plasmid Stability

The vector contains the origin of replication, which determines the average number of plasmid molecules per cell, which is also called the plasmid copy number (Georgiou, 1988). However, the exact number of plasmids in individual cells may differ. Moreover, the fraction of colonies harboring the plasmid may change throughout the fermentation. This is called plasmid instability. Plasmid stability is the ability of plasmid harboring cells to maintain the structure of the plasmid with a high copy number. Batch culture studies show that the recombinant plasmid suffers from segregational and structural instabilities. Structural instability is the result of insertions or rearrangements within the plasmid or the loss of some plasmid material by deletions (Old and Primrose, 1989). Segregational instability is due to the defective partitioning of plasmids resulting in the loss of the entire plasmid. Plasmid free cells have a higher growth rate and can outgrow the plasmid containing cells. Plasmid free cells do not produce the desired protein and hence their domination results in a rapid decline in the efficiency of the culture. Plasmids with small DNA inserts are more stable. The production of a protein toxic to the host cell results in the death of the plasmid containing cells. Hence the plasmid free cells dominate the culture much faster. Dissolved oxygen concentration, growth temperature, presence of essential amino acids, dilution rates are some of the environmental factors affecting plasmid stability. The most common method of preventing plasmid free cells from dominating is using ampicillin selection. The plasmid harbors the gene for ampicillin resistance and when the cells are grown in ampicillin containing media, the loss of the plasmid results in cell death (Georgiou, 1988). However, in large-scale cultivations, the antibiotic in the medium can be inactivated by a small portion of the cells and hence the plasmids free cells can propagate. Moreover, the cost and waste disposal considerations make antibiotic selection undesirable in large scale productions. The stability of the plasmid can also be increased when a plasmid-encoded protein is essential for cellular metabolism. The plasmid stability problem can be solved definitely by the insertion of the target gene into the bacterial chromosome at the expense of having only one copy of the gene per cell which would limit the mRNA synthesis and hence protein production (Pierce and Gutteridge, 1985). Plasmids can also be maintained if they harbor the partitioning function,

par, which ensures that the plasmid is efficiently segregated to the daughter cells at cell division (Primrose *et al.*, 1983).

2.4. Translational Regulation

The translational initiation region of most genes in *Escherichia coli* contains the initiation codon AUG coding for methionine. In addition to the AUG initiation codon, other conserved structures have also been recognized. Ribosome binding site is located downstream of the promoter (Alberts *et al.*, 1994). It consists of the Shine-Dalgarno sequence followed by an adenine and thymine rich translational spacer. The Shine-Dalgarno sequence interacts with the small ribosomal subunit. The ribosomal binding site is important because it determines the efficiency of protein synthesis per mRNA molecule (Gold *et al.*, 1981). The enrichment of the ribosome binding site with adenine and tyrosine residues increases efficiency of mRNA translation by reducing the potential for secondary structure formation at the 5' end of the transcript (Nishi and Itoh, 1986). The secondary structure at the translation initiation region of mRNA also is important in efficient gene expression. For good expression, the initiation and Shine-Dalgarno sites must be accessible. Thus the mRNA must be already folded in a specific secondary structure and then interact with the ribosome to initiate translation. The sequences immediately downstream of the start codon especially the triplet immediately preceding it are important as well. It is possible to increase protein expression by 20 fold by altering some nucleotides at the beginning of a gene (Hui *et al.*, 1984). However, the rate-limiting reagent of protein synthesis is the mRNA because it is highly unstable due to endonuclease and 3' exonuclease attack. The rate of synthesis and the rate of degradation determine the level of mRNA. The average half-life of an mRNA molecule is only one or two minutes. The rate of synthesis may be enhanced using strong promoters, but the same effect may be achieved by preventing or reducing degradation (Old and Primrose, 1989).

5' untranslated regions and the stem loop structures from the 3' untranslated regions and intercistronic regions stabilize the mRNA and protect it against degradation.

For example, 5' untranslated region of the *Escherichia coli* ompA transcript prolongs the half life of various heterologous mRNAs suggesting that addition of the ompA 5' stabilizer to heterologous genes might enhance gene expression. The addition of a protective hairpin structure has also been shown to enhance mRNA stability (Makrides, 1996).

2.5. Codon Usage

The genetic code is universal meaning that the same codons code for the same amino acids in all organisms. Moreover, it is degenerate, offering more than one codon for most of the amino acids. However, a bias exists in codon usage of organisms especially determined by the tRNA availability (Old and Primrose, 1989). In high level expression of heterologous proteins in *Escherichia coli*, errors in translation occur especially when the amino acid or charged tRNA is limited. Eukaryotic proteins usually contain codons infrequently used in *Escherichia coli*. Insulin, for example, has a cysteine content much higher than that of total *Escherichia coli* protein (Kurland and Gallant, 1996). Hence the low level of tRNA for these codons in *Escherichia coli* may hinder the rate and accuracy of translation of such proteins resulting in incomplete polypeptides or mistranslations. For example, it was observed that substitution of lysine for arginine occurred at the rare AGA codon (McPherson, 1988). Gene expression can be improved by replacement of the rare codons with more common ones, without modifying the protein product (Robinson *et al.*, 1984). The other approach that may be employed to circumvent codon usage problems is to coexpress the cognate tRNA genes along with the heterologous protein to ensure correct codon usage (Kurland and Gallant, 1996).

After overproduction, the protein may accumulate in the cytoplasm, may be secreted through the inner membrane to the periplasm or may be excreted to the medium.

2.6. Cytoplasmic Expression

The cytoplasmic expression usually yields high amounts of the desired protein. However, the drawbacks of the cytoplasm usually obliterate this advantage. First of all, the cytoplasm is crowded with all the other proteins of the cell, making the purification of a recombinant protein from this pool a very tedious task. Fusion protein strategies discussed in 2.10 that use fusion partners or tags with affinity to certain substances may be used to purify the protein using affinity chromatography. Applying heat to the total cell protein pool and then precipitating the denatured proteins may easily purify thermostable proteins.

Additionally, recombinant proteins usually contain the amino acid methionine as their initiator codon, which cannot be removed in the cytoplasm unless the enzyme methionine amino peptidase is coexpressed. This is a drawback especially in the expression of recombinant proteins used in the pharmaceutical industry because the recombinant protein with the extra amino acid might provoke an unexpected immune response in the patient. Furthermore, the extra amino acid might alter the conformation of the protein reducing or inhibiting its activity (Kavanaugh *et al.*, 1992).

The cytoplasm is a reducing environment and does not facilitate disulfide bond formation; this usually hinders the expression of eukaryotic proteins in *Escherichia coli* cytoplasm because they cannot assume their correct form. Bacterial cytoplasmic proteins contain few cysteine residues and most proteins that contain disulfide bonds are exported from the cytoplasm. Therefore, eukaryotic proteins that depend on the formation of disulfide bonds to be able to assume their correct three-dimensional form may not be formed in their correct conformation. There are two pathways in *Escherichia coli* that contribute to the reduction of disulfide bonds. One is the thioredoxin system composed of thioredoxin reductase and thioredoxin. Thioredoxin-reductase deficient *Escherichia coli* strains may be employed to facilitate correct disulfide bond formation. The other system is the glutaredoxin system, which consists of glutathione reductase, glutathione and three glutaredoxins (Weickert *et al.*, 1996, Schneider *et al.*, 1997).

2.6.1 Proteolysis

Among the proteins in the cytoplasm, there are proteases that degrade the protein of interest and seriously lower the amount of product obtained even if synthesis is very high. The cell recognizes some proteins as foreign or abnormal and destroys them so that the half-life of some proteins in the cytoplasm may be as low as two minutes (Talmadge and Gilbert, 1982). Proteolysis occurs due to incomplete polypeptides, mutations, excessive synthesis of subunits from multimeric complexes and overproduction of recombinant proteins. The stability of the protein depends on its amino acid sequence, especially in the amino terminal. Regions enriched in proline, glutamine, serine and threonine also destabilize proteins (Rogers *et al.*, 1986). Flexibility and accessibility of the termini, the presence of chemically blocking amino terminal groups such as the acetyl group determine stability. Exposure of protease cleavage sites on the surface of the protein make it more susceptible to proteolytic degradation (Old and Primrose, 1989).

Several experimental approaches exist to prevent degradation by proteolysis. It has been shown that the overproduction of foreign proteins triggers the production of the *lon* protease, which in turn degrades the protein. *Lon* protease is a DNA binding protein with ATP dependent proteolytic activity. Strains deficient in the *lon* protease may be used to reduce degradation. The *degP* mutant has been shown to stabilize fusion proteins in the periplasm and *ompT* mutants have proven useful in preventing cleavage between exposed basic residues (e.g. Arg-Arg) during preparation of the crude extracts. The observation that during bacteriophage T4 infection degradation of abnormal proteins decreases has been used to stabilize labile eukaryotic proteins. The gene responsible for this effect was cloned and called *pin* for protease inhibition. It was found that degradation of foreign genes was hampered in cells in which the T4 gene was expressed (Simon *et al.*, 1983). Starvation of the culture for carbon source or amino acids and oxygen limitations also increases degradation. This fact should especially be kept in mind for large-scale productions since glucose and oxygen limitations arise frequently. Growth conditions can be optimized to decrease the rate of degradation. Growth at suboptimal temperatures such as 30°C reduces degradation. Another approach is to use fusion partners. The protein may be fused to itself or to another protein to stabilize the target protein. Chaperones may be used to prevent

degradation. Replacement of specific amino acid sequences to eliminate the protease cleavage site is also possible. The hydrophobicity of the protein may also be modified. The formation of inclusion bodies also protects the protein from proteolytic degradation however the disadvantages of inclusion body formation outlined below must be considered (Georgiou, 1988). The protein may also be targeted to the periplasm or the culture medium where the concentration of proteases is low. It has been reported that rat pro-insulin that has a half life of two minutes in the cytoplasm has a half life over 10 times greater when located in the periplasm (Talmadge and Gilbert, 1982)

2.6.2 Inclusion Body Formation

The formation of inclusion bodies also is an important barrier of cytoplasmic gene expression. The charge average, turn forming residue fraction, cysteine fraction, proline fraction, hydrophobicity and the total number of residues affect the formation of inclusion bodies in the cytoplasm of *Escherichia coli* (Wilkinson *et al.*, 1991). A protein forms inclusion bodies especially when it is produced at an abnormally high rate (Sharma, 1986). In order to obtain the active protein in a soluble form, the purification conditions or the fermentation conditions can be manipulated. Precipitation of a protein into inclusion bodies sometimes can be an advantage, because inclusion bodies are insoluble and dense, and can be purified easily by centrifugation. Additionally, proteins that are susceptible to degradation are stabilized when expressed in the form of insoluble aggregates.

The insoluble aggregates can be solubilized after expression by refolding of the inclusion bodies. Addition of high concentrations of denaturants such as urea or guanidine HCl in the presence of reducing agents results in the denaturation of the protein. When the denaturants are removed by dilution or dialysis, the protein relaxes into its correct three-dimensional structure. However, the recovery of the product by refolding of inclusion bodies is unpredictable and moreover costly. Therefore, direct expression of the active soluble protein is a better choice (Georgiou, 1988).

Amino acid substitutions that suppress aggregation and favor the productive pathway leading to the native state may be employed. However, it is not possible to predict the result of amino acid substitutions and the effort required to characterize the effect of a mutation may not be commensurate to the benefit of obtaining higher soluble protein production (Georgiou and Valax, 1996)

The fermentation conditions can be altered to directly express the protein as a soluble active protein in the cytoplasm of *Escherichia coli*. The rate of protein synthesis may be reduced so that intracellular concentration of aggregation prone intermediates is minimized. The bacterial cells can be grown at lower temperatures to slow down the metabolic processes of the host and to allow sufficient time for the protein to fold correctly. When cultures of human interferon were grown at temperatures lower than 30°C, most of the protein was active and soluble while protein from cultures grown at 37°C was insoluble (Schein, 1989). Different *Escherichia coli* strains such as strains deficient in thioredoxin reductase that maintain a favorable redox potential can be selected that minimize inclusion body formation. Chaperones that interact with the nascent polypeptide chain preventing self-association may be coexpressed. It is also possible to coexpress foldases that catalyze kinetic barriers associated with the folding reaction (Georgiou and Valax, 1996). Thioredoxin may be used as a fusion partner or coproduced with the protein of interest (Proba *et al.*, 1995). Other highly soluble fusion partners may also aid formation of the native three-dimensional structure. Growth and induction of the cells under osmotic stress in the presence of sorbitol or glycyl betain, addition of nonmetabolizable sugars to the medium may also prove successful (Blackwell and Horgan, 1991).

Molecular chaperones facilitate attainment of correct conformation of the protein folding by posttranslational folding, assembly of polypeptides into oligomeric structures and localization of the protein. Elevated temperatures induce misfolded protein formation and their production boasts chaperone synthesis. Hence, chaperones are produced in dramatically high amounts at high temperatures, so they are also called heat shock proteins. The molecular chaperone binds to exposed hydrophobic regions and changes the structure of those regions that are likely to have misfolded from the molten globule state giving the protein another chance to fold (Alberts *et al.*, 1994). Three multicomponent chaperone systems have been identified in the cytoplasm. The 60 kDa heat shock protein

(Hsp60) is called GroEL and is associated with the small accessory protein GroES with which it helps other protein to fold. The GroEL chaperone has a broad specificity and is essential for viability. The Hsp70 comprises DnaK and the small proteins DnaJ and GrpE. DnaK is preferentially released in response to environmental stress. Nascent polypeptides are recognized and bound by DnaK in conjunction with DnaJ, released and then bound by GrpE (Georgiou and Valax, 1996). The third complex is the Clp system (ClpA and ClpX) which recognizes unfolded polypeptides that are to be targeted for degradation (Wickner *et al.*, 1994).

Molecular chaperones have been shown to improve correct protein folding in many cases, however, the success is unpredictable and protein specific. The correct protein-chaperone match has to be found by trial and error. Proteins fold into the state that is thermodynamically most stable. However, truncation of polypeptides, production of single domains from multisubunit protein complexes, lack of formation of disulfide bonds necessary for the correct conformation or the absence of posttranslational modifications such as glycosylation may make it possible to attain stability. Proteins that are so destabilized will or may fold incorrectly even in the presence of chaperones (Makrides, 1996). Therefore, strategies have been developed that use chaperone systems along with other optimization techniques. For example, the observation that different types of chaperones act in concert is used to overproduce chaperones in conjunction. The overproduction of DnaK alone resulted in plasmid instability, which was alleviated by coproduction of DnaJ. It may sometimes be necessary to coexpress chaperones cloned from the same source as the target protein (Cloney *et al.*, 1993). Growth temperature may have to be optimized. For example, GroES-GroEL increased the production of β -galactosidase at 30°C but not at 37°C or 42°C. However, DnaK and DnaJ conferred a more substantial increase in expression of soluble protein at all temperatures tested (Georgiou and Valax, 1996)

2.7. Periplasmic expression

Periplasmic expression is preferred when the high yield does not compensate for the disadvantages of cytoplasmic expression. The periplasm contains only about four per cent of the total cell protein (Nossal and Heppel, 1966) and hence purification of the target protein from the periplasmic pool is much simpler. Moreover, proteolytic degradation is prevented to a great extent, increasing the yield (Talmadge and Gilbert, 1982). The oxidizing environment of the periplasm facilitates the correct folding of proteins. Moreover, the signal peptide in the recombinant protein is removed during transportation to the periplasm, making the expression of the authentic native protein possible. The transport of a protein to the periplasm requires a signal sequence. The signal peptides used successfully for transportation to the periplasmic space include the *Escherichia coli* *phoA* signal, *ompA*, β -lactamase, protein A from *Staphylococcus aureus* and the human growth hormone signal (Makrides, 1996). Reduction in protein expression levels prevent the overloading of the translocation machinery and coproduction proteins that participate in membrane transport processes that improve the yield and efficiency of periplasmic expression. Chaperone system does not exist in the periplasm of *Escherichia coli* and therefore protein folding may sometimes present problems. Protein disulfide isomerases that catalyze the oxidation of disulfide bonds and peptidyl-prolyl-cis-trans-isomerases that catalyze the isomerization of X-Pro bonds may be over expressed with the target protein to facilitate protein folding in the periplasm. Coexpression of DsbA and heat shock factor σ^{32} at low temperature increased the yield of correctly folded T-cell receptor fragments (Hannig and Makrides, 1998)

2.8. Export of the recombinant protein to the culture medium

Excretion of the product to the culture medium is the best alternative from a biochemical-engineering point of view. The culture medium contains a very small amount of contaminating proteins facilitating the purification and recovery of the product.

Moreover, the absence of proteases in the medium stabilizes the protein (Hannig and Makrides, 1998). The most advantageous feature of excretion into the medium is the possibility of employment of continuous processes in recombinant product synthesis. However, if the cells grow to a high cell density, they will need vigorous agitation and aeration. This will result in air liquid interfaces that can denature the protein (Old and Primrose, 1989).

Excretion into the medium is possible by two basic methods. Either the protein is selectively secreted to the medium using pathways for truly selected proteins or the total cell proteins “leak”, selectively or not, into the medium by using signal sequences, fusion partners, permeabilizing proteins or nutrients. The first method truly offers the advantages of extra cellular excretion listed above (Makrides *et al.*, 1996) . For example vectors have been constructed in which the staphylococcal gene for the IgG-binding protein A was secreted into the medium upon induction. (Abrahmsen *et al.*, 1986). It was also reported that fusion of the human growth hormone to the protein A gene resulted in the export of the recombinant protein to the growth medium.

2.9. Fermentation conditions

Protein production can be increased significantly through the use of high-cell density culture systems in batch, fed batch and continuous systems. The composition of the growth medium, oxygen availability, pH, temperature and induction conditions are the parameters that influence recombinant protein overexpression. (Hannig and Makrides, 1998)

Acetate accumulation is a challenge in the production of proteins since acetate is a lipophilic agent that is detrimental to cell growth (Makrides, 1996). Uptake of glucose, when present in excess amounts normally exceeds the need for proper cell functions and subsequently leads to acetic acid formation. Acetate excretion can be reduced by monitoring glucose uptake rate. Methyl α -glucoside can be used as a nontoxic competitive

inhibitor to modulate cellular glucose uptake rate which delays the onset of acetate excretion. Initial concentration of glucose is important in acetate accumulation (Chou *et al.*, 1994). It has been reported that when the glucose concentration is higher than 5 g/L, protein production and cellular growth decreases (Narciandi *et al.*, 1996). Acetate accumulation is shown to be strain and growth medium specific. In BL21, high glucose concentration results in a reduction in growth rate after induction while low glucose concentration leads to exponential growth throughout induction. On the other hand, with JM109, high glucose concentration leads to linear growth throughout induction. It was reported that BL21 is a superior strain with regard to acetate accumulation suggesting that it probably possesses an acetate self-control mechanism (Shiloach *et al.*, 1996).

The supply of yeast extract to the culture medium has been reported to decrease the rate of proteolysis of the recombinant protein. It was proposed that the reason for this is that the proteases favor the smaller peptides supplied in the feed (Harrison *et al.*, 1997).

The oxygen supply to the feed is also important for protein production. Better oxygen supply to the culture provided by smaller working volumes and higher shaking speeds increases total and extracellular protein production. Moreover, glycerol increases secretion to the medium (Mikch, *et al.*, 1997).

The induction conditions affect recombinant protein production. The time and period of induction and amount of inducer are the important parameters that affect the recombinant protein expression.

Highest expression of genes cloned under the control of the *tac* promoter was obtained when IPTG was added during the exponential growth phase. It was demonstrated that the time of induction rather than the cell concentration is associated with recombinant protein activity. It was verified that IPTG induction in early or mid-exponential phase of growth retards growth of cells (Yarzabal *et al.*, 1997).

It was shown that higher growth rates prior to induction makes the cells better equipped to adapt to stressful conditions. When the pre-induction growth rate is higher,

post-induction growth rate, cell viability and cell mass concentration increase (Ryan *et al.*, 1996)

The amount of inducer can also be manipulated. The maximum activity of induced protein was found with one mM IPTG however, the activity was almost constant between IPTG concentrations 0.4 and one mM (Yarzabal *et al.*, 1997). It was also stated that different IPTG concentrations resulted in no difference in yield of protein and final growth. In such cases, the IPTG concentration can be reduced to the minimum amount possible to economize (Narciandi *et al.*, 1996)

On the other hand, it has also been reported in many cases that change in IPTG concentration also affects recombinant protein yields. It was demonstrated that an increase from 0.01 mM IPTG concentration to one mM resulted in 102 fold increase in activity while an increase from one mM to seven mM resulted in 23 per cent increase only. It has also been shown that an increase from 0.01 mM IPTG to 7.5 mM IPTG resulted in 42 fold increase in transcription while only 3.6 fold increase in mRNA level. These results prove that the mRNA stability is important and rate of degradation of cloned gene transcript increases as synthesis rate increases. They have also found that rRNA synthesis is highest when the culture is not induced. As amount of IPTG increases, load increases and there is less priority to rRNA synthesis toward mRNA synthesis (Wood and Peretti, 1991).

It was reported that the activity increase is not proportional with mRNA synthesis increases as IPTG concentration increases. They have also shown that as inducer level increases, plasmid retention decreases (Huang *et al.*, 1994).

Presence of glucose increases transcription rate and overall protein synthesis at exponential phase but specific recombinant protein level stays unaffected. This supports the hypothesis that translation rate and mRNA stability are significant factors in recombinant protein production (Li and Taylor, 1994).

The various strategies for optimization of recombinant protein production in *Escherichia coli* described herein are summarized in Table 2.3 and Table 2.4.

TABLE 2.3. General factors to be considered in optimization of heterologous protein overproduction in *Escherichia coli* (Weickert *et al.*, 1996)

Technique	Comments
Host Strain	Choice of host strain impacts expression.
Plasmid copy number	Gene dosage, as manipulated through plasmid copy number, affects expression.
Antibiotic selection	Choice of antibiotic resistance on the expression plasmid can influence heterologous protein expression.
Promoter	Strong/weak, inducible/constitutive; regulation is a major influence on protein expression affected by relative orientation and strength of promoters on the plasmid.
Transcription termination	Effectiveness and spacing of transcription terminators affect expression.
mRNA stability	The stability of the mRNA impacts yield. Secondary structure, especially at the 5' end of the message, often plays a critical role.
Translation signal	The ribosome binding site affects the level of ribosome loading and clearance, and hence expression. Secondary structure at the 5' end of the message can affect the accessibility of the ribosome binding site.
Codon usage	Utilizing the optimal <i>E. coli</i> codons often improves yield.
Temperature	Temperature affects on protein folding and stability. Lower temperatures frequently improve soluble yields.
Growth conditions/media	Growth conditions, oxygen levels, growth rate, carbon source and fermentor configuration affect yield.
Fusion proteins	Fusing heterologous proteins to peptides that are typically highly soluble in <i>E. coli</i> often improves soluble yield.
Tagged Proteins	Fusion of short peptide tags simplifies recovery of heterologous proteins.

TABLE 2.4. Protein specific factors to be considered in optimization of heterologous protein overproduction in *Escherichia coli* (Weickert *et al.*, 1996)

Technique	Comments
Secretion	In the periplasm, disulfide bonds can form and certain deleterious proteins, like RNase A, can accumulate without threatening the cell.
Coexpression of chaperones	Co-induction of higher levels of bacterial chaperones, typically GroESL, improves soluble yields of some heterologous proteins.
Protease mutations	Protease-deficient strains can improve soluble yields of sensitive proteins.
Disulfides	Manipulation of genes associated with disulfide bond formation and/or redox state can improve soluble expression.
Cofactors	Provision of the necessary cofactors, either by exogenous feeding or supplementation of the natural capacity <i>E. coli</i> to produce cofactors, can improve soluble yields.
Coexpression of subunits	Multisubunit proteins can often be successfully assembled <i>in vivo</i> with concomitant expression of the subunits.
Expression as a polymer	Shorter peptides, or proteins that naturally polymerize, can be expressed in multimers to improve soluble yield.
Expression as a fragment	Extremely large, complex, heterologous proteins are seldom expressed successfully in <i>Escherichia coli</i> ; however, successful expression of domains can be achieved more readily.
Specific mutations	Mutagenesis can sometimes identify changes that specifically reduce or eliminate aggregation in a protein-specific manner.

2.10. Fusion Protein

The use of gene fusions has greatly facilitated the expression and subsequent purification of heterologous proteins in *Escherichia coli*. Specific fusion peptides may confer advantages to the target protein during expression, such as increased solubility, protection from proteolysis, improved folding, increased yield and secretion. The presence of the additional amino acid sequence stabilizes the hybrid protein against degradation. Fusion of short peptide tags can be used to target the protein to different compartments of the cell or even to the culture medium, allowing accumulation of the protein in the most suitable environment. Furthermore, fusion protein tags can be removed by *in vitro* cleavage of the fusion protein, which makes the formation of the native amino terminus of the protein possible. The highly soluble fusion partner facilitate folding of the target protein. A problem of gene fusions is that the fusion proteins are sometimes cleaved *in vivo* at the junction between the fusion partner and the target protein, which creates problems and decreases yield of target protein if the fusion partner serves as an aid in purification.

The protein of interest may be fused to itself to increase stability of the protein against proteolytic degradation. With this method, inclusion body formation may increase. The advantages and disadvantages of inclusion bodies have been described previously and must be taken into account (Ford *et al.*, 1991).

The gene may be spliced directly after a signal sequence. With this strategy, if it is possible to remove the tag, a recombinant protein with a native amino terminus may be obtained. Moreover, the signal sequence can target the gene to different compartments of the cell; hence it may serve as a secretion tag (Makrides, 1996).

C-terminal fusions are commonly used constructs in which the gene of interest is fused to the 3' end of a gene, which codes for a highly expressed protein. This configuration is used frequently because the signals which direct transcription and translation are located at the 5' end of a gene and will function efficiently in the presence of a variety of different fusions to the 3' end of the gene. Thus the promoter and the translation initiation signals are not changed by different target proteins and the expression

level is predictable. Moreover, many native proteins tolerate alterations in their N-terminal region (Carter, 1990).

N-terminal fusions can also be used to express genes. With N-terminal fusions, it is simple to perform N-terminal sequencing of the gene fusion and it is possible to design various expression assays. However, N-terminal fusions are disadvantageous because the transcription and translation signals must be engineered specific to the target gene. Additionally, the chemical cleavage of the fusion partner may result in nonnative N-terminal protein, which is undesirable especially in pharmaceutical recombinant proteins. (Ford *et al.*, 1991)

Fusion tails that can be used exhibit different biospecific and biochemical interactions are listed in Table 2.5. These include (a) entire enzymes with affinity for substrates or inhibitors, such as the β -galactosidase; (b) polypeptide binding proteins such as the staphylococcal protein A; (c) carbohydrate binding proteins or domains such as maltose binding protein, 4. biotin binding domains; (d) antigenic epitopes with affinity to immobilized monoclonal antibodies; (e) charged amino acids for use in charge based recovery methods such as the poly (Arg); (f) poly (His) residues for binding to metal chelates and recovery by immobilized metal affinity chromatography; (g) other poly(amino acid)s with various binding specificities (Ford *et al.*, 1991).

The selection of the fusion tail depends on the desired final destination and authenticity of the protein, the stability of the protein and the purification method suitable for the recovery of the protein. The advantages and disadvantages of cytoplasmic and periplasmic expression and secretion to the culture medium have been discussed. The end use of the protein determines the desired final authenticity of the protein. The affinity handles provide a simple purification for the proteins that are hard to purify using their own properties. (Makrides, 1996)

TABLE 2.5. Tail Types (Makrides, 1996; Ford *et al.*, 1991)

β Gal, β -galactosidase; CAT, chloramphenicol acetyltransferase; CBD, cellulose-binding domain; GSH, reduced glutathione; GST, glutathione-S-transferase; IA, immobilized antibody; IEC, ion-exchange chromatography; I-IgG, immobilized immunoglobulin G; IMAC, immobilized metal affinity chromatography; IS, immobilized substrate; MBP, maltose-binding protein; SPA, staphylococcal protein A; SPG, streptococcal protein G.

<u>FUSION PARTNER</u>	<u>C- OR N- TERMINUS</u>	<u>LIGAND /MATRIX</u>	<u>SEPARATION METHOD</u>
1) Enzymes			
β Gal	N, C	TPEG-Sepharose	IS, IA
GST	N	Glutathione Sepharose	IS (GSH)
CAT	N	Chloramphenicol-Sepharose	IS
2) Polypeptide-binding protein			
SPA	N	Immunoglobulin G-Sepharose	I-IgG
SPG	C	Albumin	I- albumin
3) Carbohydrate-binding domains			
MBP	N	Amylose resin	IS
CBD	C	Cellulose	IS
4) Biotin- Binding Domain	C	Biotin	I-avidin
5) Antigenic Epitopes			
RecA	C		IA
Flag	N	Antiflag monoclonal antibodies	IA
6) Charged Amino Acids			
Poly (Arg)	C	S-Sepharose	IEC
Glutamate	N		IEC
7) Poly (His) Tails	N, C	Ni ²⁺ -nitriloacetic acid	IMAC
8) Other Poly (aa) Tails			
Poly (Phe)	N	Phenyl-Superose	Phenyl-Sepharose
Poly (Csc)	N	Thiopropyl-Sepharose	Tripropyl-Sepharose

Secretion-affinity fusions are used both for targeting the protein to the culture medium and for simplifying purification of the protein by affinity chromatography. Secretion to the medium makes it possible to purify the product in a continuous system, an advantage from a chemical engineering point of view. Dual affinity fusions are especially used to express proteins that are highly susceptible to proteolysis. In this system, the gene of interest is fused to two fusion partners at each terminus, with specific affinity for two different ligands. This system requires two subsequent purification steps for complete recovery of the target protein. The solubility, stability, size, binding constants and subunit structure of both of the fusion partners must be considered in the construction of dual affinity fusions. It is also possible to fuse a secretion signal with a sequence that inserts the fusion product into the cell wall or one of the cellular membranes. This method is used to expose receptors or antigens on the outer surface of the bacteria or to assemble the protein into virus-like particles (Ford *et al.*, 1991).

The fusion partner may or may not be removed upon expression of the gene depending on the final desired authenticity of the protein and the biological interference of the fusion partner with the activity of the target protein. If the fusion partner interferes with the biological activity of the target protein, it is necessary to cleave the fusion protein and obtain the target in homogeneous form even in laboratory scale. If the target protein is to be used for therapeutic purposes, it is strictly necessary to obtain the target protein homogeneously in its authentic form not to cause immune response in the patient (Ford *et al.*, 1991; Uhlen and Moks, 1990).

It is possible to cleave the fusion protein enzymatically or chemically. The choice of method depends on the composition, sequence and characteristics of the particular protein. Different cleavage methods are summarized in Table 2.6. Chemical cleavage is efficient and the reagents are widely and cheaply available, making it possible to apply it on a large scale. However, it is nonspecific and harsh. Therefore, cleavage of the recombinant protein at additional cleavage sites occurs frequently. Moreover, the protein of interest may not always be able to withstand the harsh conditions that must be applied for chemical cleavage. Thirdly, a limited extent of amino acid chain modification may occur, leading to heterogeneity of the target protein. For example, oxidation of cysteine residues by Cyanogen bromide has been reported. Additionally, the chemical used may be

highly toxic, such as Cyanogen bromide, limiting its use in preparative scale applications. (Carter, 1990)

Enzymatic cleavage is desirable for relatively mild reaction conditions and most importantly, for the high degree of specificity exhibited. Exopeptidases and endopeptidases may be used for cleavage of the fusion protein. However, enzymatic cleavage is affected by steric factors, therefore the cleavage site must be carefully engineered to be accessible by the enzyme. Among the useful enzymes are factor Xa, thrombin, enterokinase, renin and collagenase. All of these enzymes have extended substrate recognition sequences (upto seven amino acids in the case of renin), which reduces the likelihood of unwanted cleavages elsewhere in the protein. Factor Xa and enterokinase are the most advantageous because they cleave on the carboxy-terminal of their recognition sequences, allowing the release of the target protein with its authentic N-terminal (Ausubel *et al.*, 1995).

TABLE 2.6. Cleavage Methods to Cleave Fusion Proteins (Uhlen and Moks, 1990)

<u>Cleavage Method</u>	<u>Recognition Sequence</u>
1) Chemical	
Cyanogen Bromide	-Met-/-
Formic Acid	-Asp-/-Pro-
Hydroxylamine	-Asn-/-Gly-
2) Enzymatic	
Collagenase	-Pro-Val-/-Gly-Pro-
Enterokinase	-Asp-As-Asp-Lys-/-
Factor Xa	-Ile-Glu-Gly-Arg-/-
Thrombin	-Gly-Pro-Arg-/-
Trypsin	-Arg-/-
Clostripain	-Arg-/-
Ala ⁶⁴ -subtilisin	-Gly-Ala-His-Arg-/-

There are increasing number of studies reporting the use of fusion tails in the production and recovery of recombinant proteins.

A bacterial protein encoded by the MalF gene with known gene sequence but which could not be identified by conventional methods was identified by labeling by fusion to β -galactosidase. The product is an essential component of the maltose-maltodextrin transport system. MalF gene has been fused to the lacZ gene with MalF in the amino terminal and lacZ in the carboxyl terminal. The fusion protein is expressed in *Escherichia coli* using the ribosome binding site and the translational start of lacZ gene. This facilitates the purification of the recombinant protein after it is expressed. The recombinant protein is purified using anti- β -galactosidase affinity columns. A variety of strains have been used in this study all of which contain the lacY^{mal} gene. Thus, they have the ability to grow on maltose even though they lack the normal maltose transport system (Shuman *et al.*, 1980).

β -galactosidase fusion system was improved by facilitating easier separation of the target gene from the marker after a simple purification step of the fusion protein. The target cistron DNA is fused to the marker cistron via a piece of DNA that codes for a linker peptide. The target cistron DNA in this case is the replication initiator cistron of the plasmid R6K, the marker cistron is β -galactosidase and the linker peptide is collagen. The β -galactosidase provides easier purification as well as stabilization of a highly labile protein. The fusion protein is readily purified by β -galactosidase specific affinity chromatography. Then, by controlled collagenolysis with collagenase, the replication initiator of R6K is separated from the β -galactosidase moiety. This method also makes it possible to isolate mutant proteins or domains of proteins (Germino and Bastia, 1984).

Protein A of *Staphylococcus aureus* has been widely employed as a fusion partner. The effect of overproduction of EcoRI restriction endonuclease on the plasmid stability and product concentration has been studied (Wu *et al.*, 1993). EcoRI restriction endonuclease is toxic to *Escherichia coli* cells, therefore, a protection plasmid encoding the DNA methylase of *Escherichia coli* was coexpressed. Additionally, the effect of expression of a plasmid carrying the gene for the repressor protein is also investigated. Strain JM109 has been employed in this study. The fusion protein expression is induced by

shift in temperature. The comparison of the expression of EcoRI alone and as a fusion with SpA has shown higher stability of the repressor plasmid, higher enzyme activity and easier purification. Expression without the protection plasmid results in loss of the expression and repressor plasmids during induction and the activity is low. When the protection plasmid is coexpressed, the copy numbers of the plasmids remain constant throughout the experiment.

Histidine tag has been employed to overproduce and purify biologically active human growth hormone in *Escherichia coli*. Especially for small proteins, large fusion partners may be problematic since the major portion of the fusion is the fusion partner and could be immunogenic. With the His₆ tag, this problem is circumvented since the tag is small and does not interfere with the biological function of the human growth hormone. The His₆ tagged protein binds to Ni²⁺-nitriloacetic acid column and can be easily eluted with imidazole. The induction is performed with IPTG and the protein is expressed as inclusion bodies with high levels. The expression as inclusion bodies is advantageous reported to be advantageous for easier separation of the protein and prevents degradation of the protein in this study (Mukhija *et al.*, 1995).

The expression of ubiquitin fusion proteins has been studied. It has been shown that ubiquitin provides stability to heat allowing expression of the fusion protein at higher temperatures and hence providing *in vivo* heat treatment. It has also been shown to stabilize nascent peptide chains by its resistance to proteolytic degradation. Moreover, it acts as chaperonin and allows correct folding. The DH5αF'IQ cells are induced with IPTG for two or three hours after which cell lysis occurs and the specific yield decreases. As a result, 60 per cent or more of the total cell protein is ubiquitin fusion protein. In addition, the protein content of ubiquitin fusion strain has been reported to increase by induction, which is unusual, since the protein decreases due to metabolic burden to the cell. The reason for this has been proposed to be resistance of the ubiquitin fusion protein to protein turnover (Pilon *et al.*, 1996).

2.10.1 Maltose Binding Protein

Maltose Binding Protein (MBP) is the product of the *MalE* gene of *Escherichia coli* K12. *MalE* gene encodes for the pre-protein of 396 amino acid residues, 26 of which constitute the signal peptide. The expression of MBP is controlled by *malEp*, a strong promoter that is activated by protein *MalT* in the presence of inducing maltose or maltodextrins and repressed by glucose (Bedouille and Duplay, 1988). It is a binding protein specific for maltose and maltodextrins. The binding site recognizes the α 1-4 glycosidic bond linking the glucose moieties of maltose. MBP is essential for the energy dependent translocation of maltose and maltodextrins through the cytoplasmic membrane (Duplay *et al.*, 1984).

MBP is exported across the cytoplasmic membrane and localized in the periplasm. It has been proposed that the signal sequence is responsible for initiating export through the membrane. Mutations in the signal sequence prevent MBP export and provoke its accumulation in the cytoplasm.

The crystal structure of the MBP has been identified and it has been proposed that the protein is made up of two discontinuous globular domains that surround a cleft that forms the binding site for maltose and maltodextrins (Spurlino *et al.*, 1991). This multidomain structure was exploited to determine the permissive sites at which short peptide insertions and/or deletions can be functionally accommodated (Duplay *et al.*, 1987). Using this information, β -lactamase (Bla) was inserted into three of these permissive sites and fused to the C-terminus of *MalE*. It was observed that the placement of Bla towards the C-terminus of *MalE* results in a higher level of soluble proteins in the periplasmic fractions. The expressed hybrid proteins were purified by amylose column chromatography. The presence of the degradation products on SDS-PAGE suggested that the hybrid protein still binds to amylose and therefore degradation occurs within the Bla moiety. Proteolytic breakdown is higher in end-to-end fusion protein than with inserted Bla (Betton *et al.*, 1997).

The use of MBP as a fusion partner in expression and purification of foreign proteins in *Escherichia coli* has many advantages. MBP and its precursor can be purified by affinity chromatography on crosslinked amylose and then the proteins can be eluted by competition with maltose. Hence, fusion proteins can be purified from cell extracts in one step with a high yield and high degree of purity. The materials used are easy to prepare and inexpensive and crosslinked amylose is reusable, making large-scale purification simple and feasible. Moreover, the purification takes place under very mild, physiological conditions, enabling the protein to maintain its full activity (Maina *et al.*, 1988). In contrast, purification by fusion to protein A necessitates the pH to be decreased to 3.0, with the risk of inactivation (Bedoulle and Duplay, 1988). Finally, MBP does not contain any cysteine residues that could interfere with the disulfide formation within the target peptide (Maina *et al.*, 1988).

The Staphylococcal aureus nuclease A and the Klenow fragment of the cytoplasmic *Escherichia coli* DNA polymerase I have been fused to MalE, the periplasmic maltose-binding protein of *Escherichia coli*, under the control of the promoter malEp to test the potential of MalE as a vector. The strain PD28 was used in the expression of the hybrid proteins. The deletion Δ malE44 of gene malE in this strain avoided the expression of wild type protein MalE from the bacterial chromosome and its copurification with the MalE hybrids. Gene expression was repressed with glucose until induction with a wash of the cells. The hybrid proteins were purified on crosslinked amylose with 85 per cent efficiency. The nuclease was exported to the periplasm while the polymerase was expressed and accumulated in the cytoplasm due to a mutation in the signal sequence of the malE gene. Standard, nonoptimal growth conditions were employed because toxicity problems were expected. Moreover, it was observed that upon shift of growth temperature from 30°C to 42 °C, toxicity increased and syntheses decreased. This may be due to the increase in the abnormally folded proteins that may be more toxic to the cell or the increase in the expression of the *lon* protease that may degrade the products (Bedoulle and Duplay, 1988).

PstI endonuclease and β -galactosidase have been expressed as fusion proteins with the maltose binding protein under the control of the *lac* promoter. *Escherichia coli* strains RR1, CG864, SF1362, MZ20, JM83 and 71-18 have been used to express the fusion

proteins. Cytoplasmic and periplasmic expression of the PstI restriction-modification system were studied. The cells were induced with 1 mM IPTG for a period of 2 hours and the crude extract was purified by amylose column chromatography. A major proportion of the restriction endonuclease activity passed through the column and the remaining five-10 per cent was eluted by maltose. The yield of restriction endonuclease activity upon purification by amylose column chromatography was 100,000 units/liter culture. On the other hand, 760 000 units of β -galactosidase was obtained per liter culture. The presence of β -galactosidase activity in the flow through fraction of the amylose column chromatography suggested the possibility of cleavage of the hybrid protein to give an active β -galactosidase fragment lacking maltose binding activity. The two drawbacks of this system are that the stability of the hybrid protein is hard to predict as seen from the degradation products on SDS-PAGE analysis and that the hybrid protein carries MBP, which may not be acceptable in some cases such as the production of pharmaceuticals (di Guan *et al.*, 1988).

The fusion of target proteins to MBP has been improved in a system that allows the separation of the target domain by a site-specific proteolytic cleavage after purification. The fusion protein in this system contains the recognition sequence for the blood coagulation factor Xa. Factor Xa recognizes and cleaves the protein at the tetrapeptide Ile-Glu-Gly-Arg. The presence of the recognition site before the target domain in a fusion would allow the cleavage at that site, releasing the target protein without any N-terminal residues. This system has been employed for the expression of the prokaryotic β -galactosidase and the eukaryotic paramyosin in the strain 71-18. The $lacI^q$ coded repressor strongly represses the *lac* promoter on plasmids they have used and allows cloning of genes that may be lethal. Upon induction by IPTG and purification of the crude extract by amylose column chromatography, the purified protein was analyzed on SDS-PAGE and multiple bands with the expected size of the fusion protein and smaller proteins were observed. This may be due to premature termination of the fusion protein or the degradation of the expressed protein by *Escherichia coli* proteases (Maina *et al.*, 1988).

The cytoplasmic and dimeric Gene 5 protein (G5P) of phage M13 was expressed as a fusion protein with MBP. The strain PD28 was used in the periplasmic expression of the hybrids and it was observed that 3.5 mg of protein was obtained from the periplasmic

extract with a yield of 50000 hybrid protein molecules per *Escherichia coli* cell. The expression of the G5P gene as a fusion with MalE makes the purification and study of the mutant G5P having an impaired DNA-binding activity possible, since with this system, purification is dependent on the MBP moiety of the hybrid protein (Blondel and Bedoulle, 1990).

Genetic fusion with MBP was used for the purification and characterization of the PufQ gene of *Rhodobacter capsulatus*. Amylose column chromatography could be used to purify the MBP-PufQ fusion protein which was, cleaved by Factor Xa and further purified by reversed-phase HPLC (Fidai *et al.*, 1993).

The function of proliferating-cell nuclear antigen (PCNA) was determined by expression of the rice PCNA in the form of a fusion protein with MBP. pMAL-c2 vector system was used in the expression studies with TB1 as the host strain. The fusion protein from the crude extracts was purified by amylose column chromatography. However, complete digestion with Factor Xa was found to be unsuccessful. The fusion protein was observed to be less effective in stimulating human polymerase δ activity than human PCNA either due to the interference by the large MBP or due to inefficient heterologous combination of rice PCNA and human polymerase δ (Matsumoto *et al.*, 1994).

Basic fibroblast growth factor (FGF2) is purified by heparin affinity chromatography. However, traces of heparin remain in the purified products and interfere with the biological activity of FGF2. To solve this problem, the FGF2 gene was inserted into the pMAL-c vector system to produce the fusion protein with MBP via recombinant fusion protein with MBP fusion partner. The purification of the fusion protein by amylose affinity chromatography yields pure fusion protein with no heparin. The purified fusion protein was then subjected to cleavage with Factor Xa and the activity of the intact fusion protein and the cleaved FGF2 were determined and compared. The results showed that the cleaved FGF2 was more active than MBP-FGF2. In addition to the ease and high efficiency of purification, this system provides a slow releasing device for FGF2. Culture dishes, membranes or biomaterials may be coated with MBP-FGF2 and release of FGF2 may be possible under the control of Factor Xa (Lemaitre *et al.*, 1995).

Full length Rep78 of Adeno-Associated Virus (AAV) has been cloned, expressed and purified as a fusion protein with MBP. In order to investigate the various functional aspects of AAV Rep78, its expression with the vector pMAL-c2 was studied. Obtaining high yields and readily purifying them with this protein expression-purification system solved the problem of generating sufficient amounts of the protein to study its functions. Maximum expression was obtained upon 0.3 mM IPTG induction for 45 minutes. five per cent of the total cellular proteins was observed to be the fusion protein. The lower molecular contaminants seen on the SDS-PAGE analysis of the amylose column eluates were either due to immature chain termination of the fusion protein or due to its degradation by host proteases. It was determined that the fusion protein retained all the biochemical activities of wild type Rep78 (Batchu *et al.*, 1995).

In order to demonstrate two additional functions of the Rep proteins, the Rep68 protein of AAV has been cloned into the pMAL-c2 vector. The expression and purification of the protein yielded large amounts of protein sufficient for characterization studies. DH5 α was used for the expression of the fusion protein and the cellular extracts prepared were purified by one step amylose affinity chromatography (Wonderling *et al.*, 1995).

Streptolysin O (SLO) belongs to the family of bacterial cytolysins and forms oligomeric pores on cell membranes and lyses the cells. Its expression in *Escherichia coli* is a possible yet time consuming and inefficient task because the SLO is rapidly inactivated. The regulated expression of SLO as translational fusion with MBP gene has made the structure-function studies possible. The protein was expressed in the periplasm and the cytoplasm of the strain TG1 using the vectors pMAL-p2 and pMAL-c2 respectively. The cells were induced with 0.4 mM IPTG for 1.5 hours. 87.5 mg of purified protein was obtained from three liter culture. The periplasmic expression resulted in degradation of the fusion protein; however, high yields of the fusion protein were obtained with cytoplasmic expression. The fusion protein was purified by a single affinity chromatography step over an immobilized starch gel column. The purified proteins were then subjected to cleavage with Factor Xa, thrombin, trypsin and plasmin. It was observed that the digestion products of cleavage with Factor Xa and thrombin were of the expected size of the SLO. On the other hand, cleavage with trypsin and plasmin resulted in active

but smaller, truncated forms of SLO. SLO fused to MBP was determined to exhibit full hemolytic and cytotoxic activity (Weller *et al.* 1996).

The *Rhodobacter capsulatus* NtrB and NtrC (RcNtrB and RcNtrC) system has also been characterized. RcNtrB has been expressed as a fusion with MBP because overexpression of RcNtrB alone in the T7 overexpression system gave very low yields which was not sufficient for characterization studies. RcNtrB was expressed in TB1 with the pMAL-c2 expression vector system. The cells were induced with one mM IPTG for three hours. The fusion protein was overexpressed to 5 per cent of total cell proteins. Upon purification by amylose column chromatography, the fusion protein was obtained with 95 per cent homogeneity. Several major forms of the MBP-RcNtrB smaller than the fusion protein were present in the eluates of the chromatography probably as a result of degradation (Cullen *et al.*, 1996).

The human phenylalanine hydroxylase (hPAH) has been expressed as a fusion protein in the pMAL system to study the phosphorylation of the protein. The hPAH was fused to the C terminal end of MBP through the target sequences of the restriction protease Factor Xa or enterokinase. The purified proteins were subjected to Factor Xa or enterokinase cleavage. Enterokinase cleavage resulted in 100 per cent full-length enzyme. Factor Xa cleavage produced enzyme that lacks 13 N-terminal residues probably due to a secondary cleavage site. It was also observed that when the enzyme was phosphorylated before exposure to Factor Xa, it was protected against secondary cleavage. This system is valuable because it allow the production of the mutant forms of the hPAH in patients with phenylketonuria and hyperphenylalaninanaemia since some mutant forms are unstable as isolated enzymes. When they are expressed as fusion proteins, their catalytic properties may be studied due to higher stability and yields (Doskeland *et al.*, 1996).

The fatty acid synthase and its subdomains were expressed in the vector systems pMAL-c2 and pET-32b as fusion proteins with MBP and thioredoxin-His-tags as fusion partners. TOP10F' was used in the expression of the fusion protein with the pMAL-c2 system. The MBP fusion proteins were purified by fractionation with ammonium sulfate to get rid of *Escherichia coli* enzymes, dialysis and then amylose column chromatography. This study showed that expression of a large (272 kDa) is possible with pMAL-c2 vector

system. However, the interesting feature of this work is that the subdomains of a large protein may be expressed separately making the determination of the function of each domain possible. This should facilitate mutation analyses of the structure-function relationship of complex enzymes (Jayakumar *et al.*, 1996).

The active chimeric pediocin AcH was fused in frame to the MBP. Pre-pediocin AcH can be secreted and processed in *Escherichia coli* with the aid of a specialized secretion machinery. Expression as a fusion to MBP provides a method to secrete active AcH to the periplasm. Moreover, they have succeeded in releasing the chimeric protein into the culture medium by expression in a periplasmic leaky *Escherichia coli* host, E609L. In this host, the gene encoding the outer membrane Braun's lipoprotein is disrupted. Only 10-50 per cent of the periplasmic proteins is released into the culture medium. It has been observed that minimal amounts of protein are released by cell lysis. The MBP fusion-periplasmic leaky system is useful for the production and screening of bioactive peptides such as bacteriocins (Miller *et al.*, 1998).

2.10.2 pMal-c2 Fusion And Purification System

The pMAL vectors contain the *malE* gene that encodes for MBP under the control of the strong inducible *tac* promoter and the translation initiation signals of MBP. The pMAL-c2 vector is 6646 base pairs long. It has an ampicillin resistance gene. There are approximately 20 copies of the vector per cell. The *lacI^q* gene on the vector encodes the *lac* repressor, which keeps expression low until the *tac* promoter is induced with IPTG. Unique restriction endonuclease sites between *malE* and the *lacZ α* gene allow the insertion of coding sequences of interest, to generate fusions with foreign polypeptide (Figure 2.1). Periplasmic signal sequence of *malE* gene is deleted in pMAL-c2 resulting in cytoplasmic expression of the fusion protein. β -galactosidase α -fragment activity of the *malE-lacZ α* gene fusion results in blue colony formation when α -complementing hosts such as TB1 are transformed with the pMAL-c2 vector. Insertion of a DNA fragment into the polycloning site interrupts the *malE-lacZ α* gene fusion, destroying α fragment activity. Hence recombinant colonies do not turn blue, allowing blue-to-white screening for inserts.

There is a spacer sequence coding for ten asparagine residues between the *malE* sequence and the polylinker. This spacer insulates MBP from the target gene increasing affinity binding to amylose resin. The vectors also have a sequence coding for the recognition site of the protease factor Xa. Factor Xa cleaves after the amino acids Ile-Glu-Gly-Arg allowing the protein of interest to be cleaved after purification. The polylinker includes an XmnI site superimposed on the sequence coding for the Factor Xa site. Digestion of the vector at the XmnI site yields a blunt end, and when this is used in cloning of a foreign gene, no vector-derived residues remain on the target protein upon cleavage with Factor Xa.

The M13 origin of DNA replication in the vector allows the production of single stranded DNA for sequencing and oligonucleotide directed mutagenesis.

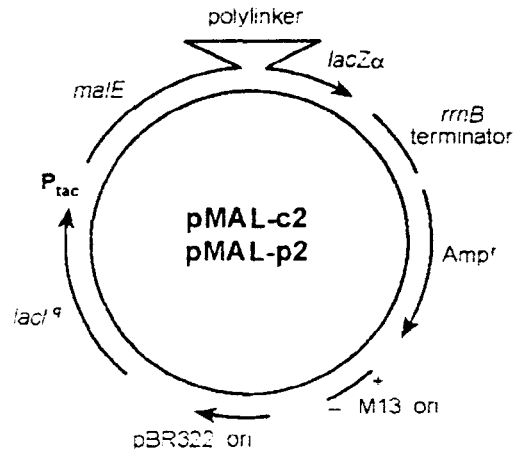


FIGURE 2.1. pMAL-c2 expression vector. pMAL-c2 (6646 base pairs) has an exact deletion in the *malE* signal sequence. Arrows indicate the direction of transcription.

The advantages of using pMAL-c2 vectors include the advantages of MBP fusion proteins as well as the advantages of C-terminal fusions. In this system, the MBP is fused to the N-terminal of the target gene, hence the transcriptional and translational signals that control expression are brought with the vector and are always identical, allowing predictability of expression. Moreover, transcription is under the control of the strong *tac* promoter which can be strongly repressed until induction with IPTG.

3. *TaqI* RESTRICTION ENDONUCLEASE

Restriction endonucleases cleave the DNA at specific sequences, which makes them the first line of defense of a bacterium against invading phages and valuable tools in various fields of molecular biology such as DNA mapping, sequencing or recombinant *in vitro* cloning. Therefore, it is desirable to obtain them in large amounts.

The isolation of a new thermophilic bacterium, *Thermus aquaticus* was described in 1969 (Brock, 1969). The optimum growth temperature of this microorganism is 70° C with generation time of 50 minutes. *TaqI* is a restriction endonuclease that has been purified from *Thermus aquaticus* YT-1 strain and is found to be very specific to the target sequence. *TaqI* restriction endonuclease recognizes 5' TCGA 3' sequence, binds specifically and cleaves between the thymidine and cytosine residues leaving a 5' staggered end upon digestion. Its specificity is several million fold relative to sites that differ by only one base pair, a property common to the type II restriction endonucleases (Mayer and Barany, 1995). The enzyme cleaves bacteriophage lambda DNA at 121 sites, pBR322 at seven sites, sV40 at one site and bacteriophage phiX174 RF DNA at ten sites (Sato S, *et al.*, 1977). It has a temperature optimum of 65° C and a molecular weight of 31612 Daltons. Kinetic analysis has shown that both high temperature and magnesium are required for specific binding and catalysis (Zebala *et al.*, 1992a).

In addition to its wide usage as a restriction endonuclease, it is also a model enzyme used to study DNA-protein interactions because of its high specificity. The *TaqI* restriction endonuclease contains each nucleotide only once in its recognition sequence, avoiding any ambiguity as to which modification might cause a given effect. Additionally, it contains both AT and GC pairs so that the recognition determinants in both type of base pairs might be probed (Bodnar *et al.*, 1983). Moreover, its requirement for the cofactor Mg₂ makes monitoring its activity simple (Zebala, *et al.*, 1992a)

TaqI restriction endonuclease is a thermostable enzyme, maintaining its activity upto temperatures as high as 70° C. Moreover, *TaqI* restriction endonuclease is stable under conditions that would denature most of the proteins, it can cleave its canonical sequence in 50 per cent formamide or 7 M urea at 65° C (Barany, 1988b). The advantage is the inactivation of nonspecific endo- and exonucleases at this temperature.

The *TaqI* restriction endonuclease exhibits star activity under conditions of low salt, high pH organic solvents such as Me₂SO, high enzyme concentrations, 20 per cent glycerol 10 per cent DMSO and substitution of Mn²⁺ for Mg²⁺. It was found that greater relaxation in specificity was found in 9LS buffer containing 20 per cent DMSO (9LS buffer: 10 mM Tris-HCl pH 9.0, 5 mM MgCl₂, 6 mM mercaptoethanol). Star activity is defined as the ability to cleave sites that differ by one base pair from the canonical site (Zebala *et al.*, 1992b, Barany, 1988a).

It is reported that insertions at amino acid positions 47, 99, 101 and 108 resulted in less than 0.1 per cent wild type activity, suggesting that amino acids in those regions interact directly with the substrate (Mayer and Barany, 1995).

There are seven cognate sequences in the *TaqI* restriction endonuclease gene for the enzyme, while there are none in the methylase gene. This is thought to be important in the sequential expression of the two genes. The restriction endonuclease gene cannot be expressed until the DNA has become fully modified. When the restriction endonuclease is expressed, it cleaves its own DNA, interrupting its further transcription and hence preventing synthesis (Slatko *et al.*, 1987).

The *TaqI* restriction endonuclease and methylase genes were cloned and sequenced in *Escherichia coli*. The genes were cloned by selecting *in vitro* for self-modified plasmids. It was found that the restriction endonuclease gene is 702 bp in length coding for the enzyme of a molecular weight of 27523 Daltons. *TaqI* methylase was found to methylate adenine residues in the cognate sequence and hence sequences were no longer susceptible to cleavage by the restriction endonuclease. The strain dependent viability of the recombinant cells was studied and it was observed that the restriction endonuclease is not lethal in the absence of modification methylase. This could be due to the fact that the cells

are grown at 37 °C and hence the restriction endonuclease can only nick the DNA at its recognition sequence, which is repairable by *Escherichia coli* ligases (Slatko *et al.*, 1987).

The *TaqI* restriction endonuclease gene was cloned under the control of the alkaline phosphatase (*phoA*) promoter to isolate and characterize *TaqI* restriction endonuclease mutants and to study specificity of protein DNA interactions. The aim of this study was to separate the methylase and the endonuclease, place the restriction endonuclease behind a strong promoter and to remove extraneous restriction sites thus allowing for subsequent two-codon insertion mutagenesis. When *TaqI* gene is introduced into a small pBR322 derivative, the recovery was 5000 U from five ml of culture. The *TaqI* restriction endonuclease gene was then cloned under the *phoA* promoter yielding 1000 U from five ml of culture. When the nucleotide sequences between *phoA* and *TaqI* gene were deleted the yield increased to 10000 U from five ml of culture. The endonuclease and *phoA* was then cloned under T7 promoter yielding 30000 Units from five ml culture. When the endonuclease gene was finally reunited with the methylase gene, the yield was 45000 units from five ml culture (Barany, 1987).

This study also revealed that insertion of two amino acids by two codon mutagenesis at position nine or ten had no effect on activity while in frame deletions between amino acid ten and 22 almost completely abolished activity, indicating the importance of the N terminus.

The *TaqI* endonuclease gene was fused to the first four codons of the alkaline phosphatase signal sequence and overproduced under the *phoA* promoter with a yield of five per cent of the total cellular proteins. The sequence of the *TaqI* restriction endonuclease gene sequence revealed a 14 base pair hairpin structure. Replacement of the hairpin region by degenerate codons increased the yield. When the methylase gene was added again, the increase in yield was remarkable, overproducing the restriction endonuclease to 30 per cent of the cellular protein. The strains RR1, HB101 and MM294 were used in this study. It was observed that RR1 strain tolerates *TaqI* methylase containing plasmids because it lacks an endonuclease that cleaves DNA containing N⁶-methyladenine (Barany, 1988b).

The sequences of both *TaqI* methylase and restriction endonuclease were corrected and the C terminus of the methylase gene was shown to overlap with the N terminus of the restriction endonuclease gene, giving a restriction endonuclease of 1263 base pairs and a methylase gene of 1263 base pairs (Figure 3.1). It was suggested that the overlap region allows for formation of a hairpin structure, which allows termination of transcription. Failure to translate a full-length methylase might result in premature transcription termination at the hairpin loop, preventing the synthesis of restriction endonuclease, which is lethal, since the methylase is not expressed. However, proper translation of the methylase would occlude formation of the hairpin, allowing transcription to continue. Thus, the overlap with a hairpin region serves as a protector against degradation of DNA and hence loss of viability (Barany *et al.*, 1992).

```

1 ATGGCTTCCA CACAAGCCCA GAAAGCGCTC GAAACTTTTG AGCGTTTTCT
51 CGCAAGCTTG GACCTCGAGT CCTACCAGCA AAAGTACCGC CCTATCAAAA
101 CGGTTGAACA AGACCTGCCT AGGGAGCTGA ACCCGCTTCC GGACCTGTAC
151 GAGCATTATT GGAAAGCGCT TGAGGATAAC CCTTCCTTCC TGGGCTTCGA
201 AGAGTTCTTT GACCACTGGT GGGAAAAGCG CCTACGGCCC TTGGACGAGT
251 TCATACGCAA ATACTTTTGG GGATGCTCCT ACGCGTTTGT TCGCTTGGGC
301 CTCGAGGCTA GGCTGTACCG AACAGCCGTT TCCATCTGGA CTCAGTTTCA
351 CTTCTGCTAC CGCTGGAACG CCTCCTGCGA GCTTCCTCTA GAAGCTGCCC
401 CAGAACTCGA CGCCCAAGGG ATAGACGCGC TGATTCATAC AAGCGGGTCC
451 TCAACAGGAA TCCAGATCAA AAAGGAAACT TACCGTTCCG AGGCCAAGAG
501 CGAGAACCGC TTTTAAAGGA AGCAAAGAGG CACCGCCCTC ATCGAGATTC
551 CCTACACCCT GCAGACACCA GAGGAGCTCG AAGAAAAAGC CAAACGGGCA
601 AGAGTGAACG GAGAAACCTA CCGTCTATGG GCCAAGGTTG CACACCATTT
651 GGACCGTCTA GAAAACGGAT TCGTCATTTT TCGGGAAAGT TATGTGAAAA
701 GCATTGAGCT TTTTCTCCAG AAAAACGCTC CTACCCTATC TGGGCTCATC
751 CGCTGGGACA GGGTGGCCCA GGAAGCCCTC ACCGCCCCGT GAGGTAGACA
801 CGAAGCACAA GCCCCACAGC AAAGAGCAGG CCCACGGCAC GG

```

FIGURE 3.1. The Nucleotide Sequence of the *TaqI* Restriction Endonuclease Gene (Barany *et al.*, 1992).

4. MATERIALS

4.1. Bacterial Strains and Plasmids

Escherichia coli strain TB1 [*ara* Δ (*lac-proAB*) *rpsL* Φ 80*dlacZ* Δ M15 *hsdR* ($r_k^- m_k^+$)], ER2508 [*lon::Tn10* Δ I6 Δ I7 Δ *malB* *zjb::Tn5* Δ *lac* U169 *hsdS*20 *ara*14 *galK*2 *rpsL*20 *xyl*5 *mtl*1 *supE*44 *leuB*6 *fhuA*] and pMALc-2 plasmid were purchased from New England Biolabs.

Escherichia coli strain XL1-Blue [*F'*::*Tn10* *proA*⁺*B*⁺ *lacI*^f Δ (*lacZ*)M15/*recA*1 *endA*1 *gyrA*96 (Nal^r) *thi* *hsdR*17 ($r_k^- m_k^+$) *supE*44 *relA*1 *lac*] was from our laboratory stock.

4.2. Chemicals

All chemicals and solutions used in this study were from MERCK (GERMANY) or SIGMA (USA) or DIFCO (USA) unless stated otherwise in the text.

The restriction endonucleases were purchased from PROMEGA (USA) and New England Biolabs (USA). The *Taq* DNA polymerase was purchased from PROMEGA (USA) and *Pfu* DNA polymerase was purchased from STRATAGENE (USA). DNA and protein size markers were purchased from PROMEGA (USA) and SIGMA (USA) respectively. Bradford Reagent was from SIGMA (USA). The OmniBaseTM DNA Cycle Sequencing System and the OmniBase Sequencing Enzyme Mix were purchased from PROMEGA (USA).

The instant films (667) were from POLAROID (USA).

4.3. Buffers and Solutions

4.3.1 Bacterial Cell Media

LB Medium

Tryptone	10g
Yeast Extract	5 g
NaCl	5 g
per liter of deionized and distilled water.	

LB-Agar Medium

Agar	15-20 g
per liter of LB Medium.	

100 µg/ml of ampicillin, 1mM IPTG and 10 µg/ml X-Gal were added into either the liquid medium or the agar medium after they cooled down to 50 °C following sterilization.

SOC Medium

Tryptone	0.4 g
Yeast Extract	0.1 g
NaCl	0.012 g
KCl	0.004 g
MgCl ₂	0.04 g
MgSO ₄	0.024 g
Glucose	0.08 g
in 20 ml of distilled water.	

4.3.2 Buffers and Solutions For Plasmid DNA Isolation

STE Buffer

10 mM Tris-HCl (pH8),
100 mM NaCl,
1 mM EDTA (pH8)

Solution I

25 mM Tris-HCl (pH8),
50 mM D-Glucose,
10 mM EDTA (pH 8.0)

Solution II

0.2 M NaOH,
1% SDS

Solution III

5 M Potassium Acetate	60 ml
Glacial Acetic Acid	11.5 ml
per 100 ml of dH ₂ O	

TE Buffer

10 mM Tris-HCl (pH 8.0),
1mM EDTA (pH 8.0)

Phenol

50 mM TE saturated phenol

Chloroform/Isoamylalcohol

24:1 chloroform and isoamylalcohol

NH₄acetate

3 M ammoniumacetate

Ethanol

100 %

Cracking 2X Buffer

5 M NaOH	2 ml
10 % SDS	2.5 ml
Sucrose	10 g
per 100 ml of dH ₂ O	

Dye Solution Used in Cracking

4 M KCl
0.4 % bromophenol blue

4.3.3 Buffers Used in PCR10X PCR Buffer (Mg²⁺ free) for amplification with *Taq* DNA polymerase

500 mM KCl,
100 mM Tris-HCl (pH 9.0)
1 % Triton X-100

10X PCR Buffer for amplification with *Pfu* DNA polymerase

100 mM KCl
60 mM (NH₄)₂SO₄
200 mM Tris-HCl (pH 8.0)
20 mM MgCl₂
1 % Triton X-100
100 µg/ml nuclease free BSA

4.3.4 Agarose Gel Electrophoresis Buffers and Solutions

10X Loading Buffer

2.5 mg/ml bromophenol blue,
1 % SDS in 1 ml glycerol.

5X TBE Buffer

445 mM Trizma base,
10 mM EDTA
445 mM Boric acid (pH 8.3)

Ethidium bromide

10 mg/ml

4.3.5 Buffers and Solutions for the Isolation of DNA from Low-Melting Point Agarose

TE Buffer

10 mM Tris-HCl (pH 8.0),
1mM EDTA (pH 8.0)

Phenol

50 mM TE saturated phenol

Chloroform/Isoamylalcohol:

24:1 chloroform and isoamylalcohol

NH₄acetate

3 M ammoniumacetate

Ethanol

100 %

4.3.6 Restriction Enzyme Digestion Buffers

10 X Buffer for EcoRI

50 mM NaCl,
100 mM Tris-HCl,
10 mM MgCl₂,
0.025% Triton X-100(pH 7.5)

10 X Buffer for HindIII

50 mM NaCl,
10 mM Tris-HCl,
10 mM MgCl₂,
1mM dithiothreitol (pH 7.9)

10 X Buffer for PstI

100 mM NaCl,
50 mM Tris-HCl,
10 mM MgCl₂,
1mM dithiothreitol (pH 7.9)

Multicore Buffer for *TaqI*

25 mM Tris-Acetate (pH 7.8 at 25°C)
100 mM Potassium Acetate
10 mM Magnesium Acetate
1 mM DTT

10 X Buffer for XmnI

50 mM NaCl,
10 mM Tris-HCl
10 mM MgCl₂,
1 mM DDT (pH 7.9)

4.3.7 Buffers Used for Detection of Expression and Affinity Column Chromatography

Amylose Column Wash Buffer

20 mM Tris-HCl (pH 7.9),
200 mM NaCl,
1 mM EDTA,
1 mM DTT

4.3.8 Polyacrylamide Gel Electrophoresis Buffers and Solutions

<u>APS</u>	10 % Ammonium persulphate
<u>TEMED</u>	N, N, N, N-tetramethylethylenediamine

Acrylamide-bisacrylamide mixture (30:0.8)

Acrylamide	29.2 g
N'N'-bis-methylene-acrylamide	0.8 g
per 100 ml of dH ₂ O	

Electrophoresis Separating Gel (12.5%)

1.5 M Tris-HCl pH 8.8	2.5 ml
Acrylamide/Bis (30:0.8)	4.2 ml
10% (w/v) SDS	100 µl
10 % Ammonium persulfate	50 µl
TEMED	5 µl
per 10. ml of dH ₂ O	

Electrophoresis Stacking Gel (5%)

0.5 M Tris-HCl (pH 6.8)	2.5 ml
Acrylamide/Bis (30:0.8)	1.7 ml
10% (w/v) SDS	100 µl
10 % Ammonium persulfate	100 µl
TEMED	10 µl
per 10. ml of distilled water	

5 X Running Buffer

Tris-base	15 g
Glycine	72 g
SDS	5 g
per 1 liter of distilled water	

Sample Buffer

0.5 M Tris-HCl (pH 6.8)	1 ml
Glycerol	0.8 ml
10% (w/v) SDS	1.6 ml
2-β-mercaptoethanol	0.4 ml
0.05%(w/v) Bromophenol blue	0.2 ml

Stain

0.1 % Coomassie blue R-250,
 40% Methanol
 10 % Acetic acid

Destain

25 % Isopropanol,
 10 % Acetic Acid

Fixing

7 % Acetic acid

4.3.9 Sequencing Buffers and Solutions

6 % Sequencing Gel

Acrylamide	5.7 g
Bisacrylamide	0.3 g
Urea	42 g
TBE (10 X Buffer)	10 ml
10 % APS	500 µl
TEMED	50 µl
in 100 ml deionized water	

Sequencing 5 X Buffer

250 mM Tris-HCl, pH 9.0
10 mM MgCl ₂

Sequencing Stop Solution

10 mM NaOH
95 % formamide
0.05 % bromophenol blue
0.05 % xylene cyanol

4.3.10 DNA and Protein Size Markers Used in Electrophoresis

<u>λ- DNA / Hind III DNA size marker</u>	23130, 9416, 6557, 4361, 2322, 2027 564, 125 bp
--	--

SDS-PAGE Protein Molecular Weight Markers

Myosin	205	kDa
β-Galactosidase	116	kDa
Phosphorylase b	97.4	kDa
Albumin, bovine	66	kDa
Albumin, egg	45	kDa
Carbonic Anhydrase	29	kDa

4.3.11 Restriction Enzymes and Their Recognition Sites

EcoRI	5'...G/AATTC...3'
Hind III	5'...A/AGCTT...3'
Pst I	5'...CTGCA/G...3'
<i>TaqI</i>	5'...T/CGA...3'
Xmn I	5'...GAANN/NNTTC...3'

4.4. Primers for amplification of the *TaqI* restriction endonuclease gene

TQXmnI:	5'-CCCGAATTGCTTCCACACAAGCCCAGAAAGCGCTCGAA-3'
TQEcoRI:	5'-TATGAATTCTCACGGGGCGGTGAGGGCTTCCTGGGCCAC-3'

TABLE 4.1. Specifications of the primers (Enünlü, 1996)

	TQXmnI		TQEcoRI	
Base	Number	Percent	Number	Percent
A	12	31.6	6	15.8
C	14	36.8	9	23.7
G	7	18.4	14	36.8
T	5	13.2	9	23.7
A + T	17	44.7	15	39.5
G + C	21	55.3	23	60.5

Method	T _m (°C) TQXmnI	T _m (°C) TQEcoRI
Nearest neighbor	93.7	95.5
Percent GC	86.4	88.6
2*(A+T)+4*(G+C)	118	122

4.5. Laboratory Equipment

Autoclaves	Model MAC-601, (Eyela, Japan)
Balances	80A-200M, (Precisa, Switzerland)
Camera	DS-34, Direct Screen Instant Camera (Polaroid, USA)
Centrifuges	SORVALL RC-5B Refrigerated Superspeed Centrifuge, (DuPont Inst., USA) SORVALL RC 285, (DuPont Inst., USA) Centrifuge 5415 C, (EPPENDORF, Germany)
Cold room	4 °C (VWR Scientific, USA)
Deepfreezers	-80°C, (HETO, Denmark) -20°C, (BOSCH, Germany) -20°C, (ARCELIK, Turkey)
Electroporation Equipment	Gene Pulser II, Pulse Controller II, Capacitance Extender II, (BIO-RAD, USA)
Electrophoretic Equipment	Horizon 58, Model 200, Horizontal Gel Electrophoresis Apparatus (BRL, USA) Miniprotean II, (BIO-RAD, USA)
Fraction Collector	Model 2110, (BIO-RAD, USA)
Freeze Drier	VR-1, (HETO, Denmark) Vacuum Dryer, (BIOMETRA, Germany) GD-1, (HETO, Denmark)
Ice Machine	AF-30, (Scotsman, USA)
Incubators	Lab-Line Ambi-High-Low Chamber, (Lab-Line Instruments, Inc., USA) EN500, (NÜVE, Turkey) FN 500, (NÜVE, Turkey)
Laminar Flow	HBB 2460 LaminAir, (Holten, Denmark)

Orbital Shakers	GFL 3032, (GFL, Germany) INNOVA 4340 Illuminated refrigerated Incubator Shaker (New Brunswick Scientific, USA)
Power Supply	Power/Pac 300, (BIO-RAD, USA)
Refrigerators	+4°C (ARCELIK, Turkey) +4°C (SIMTEL, Turkey)
Sonication Equipments	Labsonic 1510, (B. Brown, Germany) Branson Sonifier 450, (VWR Scientific, USA)
Spectrophotometers	Lambda 3 UV/VIS (PERKIN-ELMER CETUS, USA) DU 640, (Beckman, USA)
Thermocyclers	Thermal Reactor TR1, (Hybaid, UK) GeneAmp PCR System 9600 (PERKIN-ELMER CETUS, USA)
Transilluminators	Reprostar II, (CAMAG, Switzerland) Foto/uv 15, (Fotodyne, USA)
UV Monitor	Econo UV Monitor, (BIO-RAD, USA)
Water Baths	BM102, (NÜVE, Turkey) DT Hetotherm, (HETO, Denmark) CB 8-30e DT ₁ , (HETO, Denmark) CB 8-30e AT ₁₁₀ , (HETO, Denmark)
Water Purification Systems	MILLI-Q, UF Plus, (MILLIPORE, USA) MILLI-RO, Plus, (MILLIPORE, USA) Model 2004, (GFL, Germany)

5. METHODS

During the experiments, sterilized equipment was used. Pipetman tips, microfuge tubes, centrifuge tubes, solutions and culture media were all sterilized at 1.02 atm and 121°C for 20 minutes in an autoclave. Glassware was sterilized at 180°C for two hours in an oven.

The bacterial strains were stored as glycerol stocks at - 70°C. Frozen glycerol cultures were used by streaking on LB agar plates for the preparation of master plates. Master plates were kept at 4°C after being sealed with parafilm.

Preparation of plates, inoculation and all other experimental work that require sterile environment were carried out under sterile laminar flow.

5.1. Preparation of Preculture

10 ml of sterile liquid medium was inoculated with a single colony of bacteria. The inoculating loop was sterilized by heating in a flame and cooled on the agar plate to avoid death of the living cells to be transferred. The preculture was incubated in an orbital shaker at 37°C overnight.

5.2. Isolation of Plasmid DNA

5.2.1 Large-scale Plasmid DNA Isolation

Large scale plasmid DNA isolation was carried out by alkaline lysis method (Maniatis *et al.*, 1989). 50 ml culture of the bacterial strain carrying the plasmid of interest was grown to a high cell medium. 500 ml LB medium containing 100 µg/ml ampicillin was inoculated with 15 ml of the overnight culture and incubated for 16-20 hours at 37°C with vigorous shaking.

Bacterial cells were harvested by centrifugation at 9000 rpm, using GSA type of rotor, for 6 minutes at 4°C. The pellet was resuspended in 100 ml STE Buffer and collected by centrifugation at the previous conditions.

The washed bacterial pellet was resuspended in 18 ml Solution I containing 100 µg Lysozyme. The cells were lysed by the addition of 40 ml of Solution II with gentle swirling. 20 ml of ice-cold Solution III was then added to the bacterial lysate to neutralize the alkali and shaken vigorously until a flocculent white precipitate formed. The precipitate contains cell debris and chromosomal DNA. After 15 minutes of incubation on ice, the lysate was centrifuged at 6000 rpm for 15 minutes, using SS34 type of rotor, at 4°C and the supernatant was filtered into a sterile centrifuge tube.

The nucleic acids were precipitated by the addition of 0.6 volume of isopropanol was added to the supernatant, mixed and incubated 30 minutes at room temperature. Nucleic acids were recovered by centrifugation at 10000 rpm, using SS34 type of rotor, for 15 minutes at 4°C. The pellet was dissolved in sterile TE Buffer.

DNase free pancreatic RNase was added to this solution to a final concentration of 10 µg/ml and the mixture was incubated for one hour at room temperature. One volume of TE saturated phenol was added to the solution and was shaken vigorously to remove the proteins. After centrifugation at 6000 rpm, using SS34 type of rotor, for five minutes, the

upper phase was mixed with one volume of chloroform:isoamylalcohol (24:1). Plasmid DNA was precipitated by the addition of 0.3 M ammonium acetate and 2.5 volumes of ice-cold absolute ethanol.

The final mixture was kept at - 80°C for one hour or at - 20°C overnight. The plasmid DNA was recovered by centrifugation at 15000 rpm, using SS34 type of rotor, for 20 minutes at 4°C. The pellet was dissolved in sterile TE Buffer.

5.2.2 Mini-prep Method for Small-Scale Plasmid DNA Isolation

In order to analyze large numbers of transformants, plasmid DNA was isolated from overnight cultures of *Escherichia coli* picked from single colonies by the mini-prep method (Rodriguez and Tait, 1983, Maniatis *et al.*, 1989). Ten ml of LB medium containing 100 µg/ml ampicillin was inoculated with a single colony and incubated overnight at 37°C with vigorous shaking.

1.5 ml of the overnight culture was placed into a minicentrifuge tube and was centrifuged at 12000 rpm for two minutes. The medium was removed and the cell pellet was resuspended in 100 µl of ice-cold Solution I. After five minutes of incubation at room temperature, 200 µl of freshly prepared Solution II was added, mixed by inversion and incubated for five minutes on ice. 150 µl of Solution III was added to the viscous lysate formed, vortexed gently and incubated for five minutes on ice.

The mixture was centrifuged in a microcentrifuge, at 12000 rpm for five minutes and the supernatant was transferred to a new tube, avoiding the white precipitate. RNase was added to the supernatant to a final concentration of 20 µg/ml and the reaction mixture was incubated at 37°C for 20 minutes.

To remove cell proteins, the solution was extracted once with TE saturated phenol and once with chloroform:isoamylalcohol (24:1) by vortexing for one minute and centrifuging at 12000 rpm for three minutes at each step, using a minicentrifuge.

Plasmid DNA was recovered from the aqueous phase by the addition of 0.3 M ammonium acetate to get rid of dNTPs and shorter oligonucleotides and 2.5 volumes of cold absolute ethanol. After thirty minutes of incubation at -70°C, the plasmid DNA was precipitated by centrifugation at 14000 rpm for 20 minutes. The pellet was washed with 70 per cent ethanol, dried and dissolved in 30 µl sterile TE buffer.

5.3. Determination of Quantity and Quality of DNA

Spectrophotometric measurement and agarose gel electrophoresis were used to determine the quantity and quality DNA.

5.3.1 Spectrophotometric Measurement

DNA was diluted to a ratio of 1:500 with distilled water and the UV absorption of DNA was read at 260 nm using quartz cuvettes. Since 50 µg of double stranded DNA has an absorbance of 1.0 at 260 nm, the concentration of the DNA in the samples were calculated using the formula:

$$50 \mu\text{g/ml} \times \text{OD}_{260} \times \text{dilution factor} = \text{Concentration in } \mu\text{g/ml} \quad (\text{Eqn 5.1})$$

The $\text{OD}_{260}/\text{OD}_{280}$ ratio was also used to determine the purity of the DNA samples. A ratio of 1.8 was considered to be pure whereas higher and lower values suggested RNA and protein contamination respectively.

5.3.2 Agarose Gel Electrophoresis

Gels were prepared by boiling the appropriate mass of agarose in 0.5 x TBE, cooling to 50°C, adding ethidium bromide to a concentration of 5 µg/ml and setting this in the gel former. Different dilutions of DNA samples were loaded on the appropriate concentration of agarose gel, after mixing with 0.5 volume of DNA loading buffer and were run at 150 V with the known amounts of DNA. Ethidium bromide intercalates DNA, and in this state fluoresces when illuminated by UV light. DNA was visualized by illuminating the gel with short wave UV light on a transilluminator. The unknown amount was estimated by comparing the intensity of the fluorescence in each sample.

5.4. Cloning of *TaqI* Restriction Endonuclease Gene into pMAL-c2 Vector

5.4.1 PCR Amplification of *TaqI* Restriction Endonuclease Gene

Genomic DNA from *Thermus aquaticus* YT-1 strain was used as a template for the PCR amplification of the *TaqI* restriction endonuclease gene using primers TXmnI and TEcoRI designed by Enünlü, 1996. All PCR reactions were carried out in a reaction volume of 25 µl containing 12.5 pmole of each primer, 2.5 µl of 10X Promega Mg⁺⁺ free buffer, 1.5 mM MgCl₂, 0.25 mM dNTP, 1 unit of *Taq* DNA Polymerase enzyme and 0.2 µg of *Thermus aquaticus* genomic DNA.

Three-step amplification in the Perkin Elmer CETUS thermal cycler Because *TaqI* restriction endonuclease gene contains a high amount of GC residues, the denaturation step at the beginning was carried out at 96°C for 5 minutes followed by thirty five cycles with incubations of thirty seconds of denaturation at 93°C, thirty seconds of primer annealing at 66°C and 1.5 minutes of chain elongation at 72°C, ending with ten minutes of incubation at

72°C. After amplification, 2 µl aliquot from each tube was mixed with 1X loading buffer and run on 1.2 per cent agarose gel to check for amplification.

In the second part of the study *TaqI* restriction endonuclease gene was amplified using the high fidelity *Pfu* DNA polymerase. All PCR reactions were carried out in a reaction volume of 25 µl containing 0.5 µM of each primer, 300 µM dNTPs and 1.25 µunits of *Pfu* DNA polymerase with *TaqI* restriction endonuclease gene as the template DNA.

After initial denaturation step at 96°C for five minutes, twenty five cycles of PCR were carried out with incubations of 96°C for 30 seconds, 66°C for thirty seconds and 2 minutes of chain elongation at 72°C. After the cycles, the reaction mixture was kept at 72°C for ten minutes to complete chain elongation of the formed fragments. The PCR products were then identified by 1.2 per cent agarose gel electrophoresis.

5.4.2 Isolation of PCR Products From Low-Melting Temperature Agarose Gels

The PCR amplification product was purified from excess primers, dNTPs, DNA polymerase and unspecific products using low melting point agarose gel electrophoresis (Maniatis *et al.*, 1989). PCR products were collected together, mixed with 1X Loading Buffer and loaded onto one per cent low melting point agarose gel which was prepared in 1X TBE Buffer. After electrophoresis, the 808 bp fragment was cut out of the gel and was kept at 65°C for 15 minutes. When the gel melted, equal volume of TE buffer was added and the solution was kept at 65°C for an additional 20 minutes until it melted completely.

Then the mixture was extracted twice with hot phenol, and once with chloroform. The desired DNA fragment was recovered from the aqueous phase by the addition of 0.1 volume of 3M ammonium acetate and two volumes of ice-cold ethanol and centrifugation at 15000 rpm, using SS34 type of rotor, for 30 minutes. The pellet was dried and dissolved in sterile TE buffer.

5.4.3 Digestion of the Insert With Restriction Endonucleases

One μg of the purified PCR fragment was digested with 36 U XmnI in a total volume of 50 μl as described by the manufacturer and incubated at 37°C overnight. Then the digest was extracted once with equal volume of TE saturated phenol and once with chloroform:isoamylalcohol (24:1) and precipitated with ice-cold ethanol. The dried pellet was dissolved in 10 μl sterile distilled water, digested with 36 U EcoRI in a total volume of 50 μl as described by the manufacturer and incubated at 37 °C overnight.

The insert that was digested at both ends was purified by phenol:chloroform:isoamylalcohol extraction, and the DNA was recovered by ethanol precipitation. The pellet was then dissolved in 10 μl of distilled water and it was kept at -20°C.

5.4.4 Digestion of the pMAL-c2 Vector With Restriction Endonucleases

Ten μg of pMAL-c2 vector was digested with 60 U XmnI in a total reaction volume of 100 μl at 37°C overnight as described by the manufacturer. The digested vector was first checked for the completion of the digestion by loading a 2 μl aliquot of the digestion reaction on 0.8 per cent agarose gel. The linearized vector was purified by phenol:chloroform:isoamylalcohol extraction and ethanol precipitation.

The dried pellet was dissolved in 10 μl sterile distilled water, digested with 60 U EcoRI in a total volume of 100 μl as described by the manufacturer and incubated at 37°C overnight. The pMAL-c2 vector that was digested with XmnI and EcoRI was purified by phenol chloroform extraction and was precipitated with ethanol. The dried pellet was then dissolved in 10 μl of distilled water and was kept at -20°C.

5.4.5 Ligation Reaction

In a microfuge tube, 100 ng linear vector prepared as described in Methods 5.4.4 and 300 ng insert prepared as described in Methods 5.4.3 were pipetted and incubated at 65°C for five minutes and were then transferred on ice where 1 µl of 10 X Ligase Buffer and 800000 U T4 DNA Ligase were added to a final reaction volume of 10 µl. The reaction was carried out at 16°C overnight.

After the completion of the reaction, 2-3 µl of samples were loaded on a 0.8 per cent agarose gel to check the efficiency of the ligation reaction before transforming the product.

5.5. Transformation of *Escherichia coli* Cells

5.5.1 Preparation of competent cells

100 ml of LB was inoculated with one ml of overnight *Escherichia coli* cell culture and was kept with vigorous shaking at 37°C until the OD₆₀₀ reached 0.5. The cells were then chilled on ice for thirty minutes before being centrifuged at 4°C at 6000 rpm, using SS34 type of rotor, for six minutes.

The LB medium was discarded and the pelleted cells were washed twice with 100 ml of sterile distilled water, kept on ice for ten minutes before centrifugation at 6000 rpm for six minutes at 4°C using SS34 type of rotor,. The cells were then washed with two ml of ten per cent glycerol centrifuged and resuspended in 200 µl of ten per cent glycerol. The competent cells were then aliquoted in 40 µl volume of samples and were kept at -70°C for storage.

5.5.2 Electroporation

Electroporation was employed to transform the competent cells with the recombinant plasmid (Maniatis *et al.*, 1989). Three tubes were prepared which contained 40 μ l of competent cells. 2 μ l of the ligation mixture was added to the first tube, 1 μ g of undigested, untreated pMAL-c2 vector was added to the second tube and nothing was added to the last tube. Competent cells cannot survive in the presence of ampicillin unless they harbor a plasmid that confers antibiotic resistance. The third tube was thus used as a control to check if the competent cells were free of any contamination.

Competent cells were then placed into the electroporation cuvette on ice avoiding bubbling in the cuvette. The electroporation apparatus was adjusted to a voltage of 2.5 kV, resistance of 200 Ω , capacitance of 25 μ F before applying the electric shock to the competent cells.

2 ml of cold SOC medium was immediately added to the electroporation cuvette and the competent cells were then transferred to 37°C water bath for two hours of incubation with gentle shaking time to time.

Different dilutions of the transformed cells from were spread on to LB agar plates containing ampicillin. The plates were kept at room temperature for about ten minutes to allow the spread cells to diffuse completely and were then transferred to 37°C oven for 16-20 hours of incubation.

5.6. Screening System

5.6.1 α -Complementation

Using a sterile toothpick, each colony on the LB ampicillin plate was transferred to new LB agar plates containing ampicillin (100 $\mu\text{g/ml}$), X-Gal (10 $\mu\text{g/ml}$), and IPTG (1mM). After an overnight incubation at 37°C, the recombinant colonies were determined to be white due to their disturbed β -galactosidase activity, while the colonies that harbor the pMAL-c2 vector alone gave blue color on the selective plate. The white colonies were then subjected to further testing.

Ten ml precultures were prepared from these white colonies and their plasmids were isolated by the miniprep plasmid isolation method as described in Methods 5.2.2.

5.6.2 Plasmid Size

To determine if those white colonies carry an insert, the size of the vector that they carry is analyzed by the cracking method using the pMAL-c2 vector for comparison (Maniatis *et al.*, 1989). Single colonies from agar plates were picked by using a sterile tooth-pick or pipette tip under laminar flow, transferred to small microfuge tubes and were dissolved in 25 μl of 10 mM EDTA, pH 8.0. 25 μl of freshly prepared cracking 2X Buffer was added and the samples were incubated at 70°C for 15 minutes and were cooled down to room temperature.

1.5 μl of 4M KCl and 0.5 μl of 0.4 per cent bromophenolblue were added to the samples, vortexed and incubated for 10 minutes on ice. Then the samples were centrifuged at 12000 rpm in a microcentrifuge for three minutes and 20 μl of the supernatant was immediately loaded on agarose gel.

5.6.3 Restriction Enzyme Analysis

To determine if the insert in the white colonies are the desired *TaqI* restriction endonuclease gene, three different restriction enzymes giving different band patterns were used. 100 ng of plasmid DNA was digested with 20 units of *EcoRI*, *XmnI*, *HindIII* and *PstI* restriction endonucleases in a reaction volume of 20 μ l, as described by the manufacturer. Incubations were carried out at 37°C for three hours to overnight.

The digestions were checked on 1.2 per cent agarose gel and the plasmids giving the correct band pattern were identified.

5.6.4 PCR analysis

The plasmids of the suspected colonies were isolated by miniprep method as described in Methods 5.2.2. Using the plasmid as the template and the identical PCR conditions used in amplification of the insert, the insert contained within the recombinant plasmids were amplified. Thus, the presence and the size of the insert were confirmed.

5.7. Detection of the Expression of the MBP-*TaqI* Gene Fusions

5.7.1 Detection of MBP-*TaqI* Fusion Proteins in the Cytoplasm

200 ml of LB medium containing ampicillin was inoculated with 200 μ l of overnight culture, induced with 1 mM IPTG and harvested by centrifugation at 6000 rpm for 10 minutes using SS34 type of rotor. The harvested recombinant cells were resuspended ten ml of Column Wash Buffer and were frozen overnight at -20°C.

The cells were thawed in cold water slowly and lysed by three to four rounds of sonication for 30 seconds until a clear lysate was obtained. The sonication was carried out with constant cooling on ice. Insoluble proteins, and membranes were removed from the lysate by centrifugation at 9000 rpm, using SS34 type of rotor, for 20 minutes. The supernatant, which is the crude cellular extract, was analyzed by SDS-PAGE.

To quick-check for the total cytoplasmic proteins, 500 μ l of overnight culture was centrifuged and the pellet was dissolved in 50 μ l of sample buffer before denaturation at 90°C for eight minutes. The samples were then analyzed by SDS-PAGE.

5.7.2 Purification of *TaqI* from the Medium

200 ml of LB medium was inoculated with 2 ml of overnight culture of recombinant XL1 cells and was kept growing at 37°C with vigorous shaking until the OD₆₀₀ reached 0.9. The cells were then induced with IPTG, to a final concentration of 1 mM. The induced cells were kept growing for an additional twenty hours and they were centrifuged for twenty minutes at 6000 rpm, using GSA type of rotor, at 4°C.

The medium was then transferred into a sterile GSA and it was precipitated with 75 per cent ammonium sulfate at 4°C with continuous gentle shaking. The precipitated proteins were recovered by centrifugation at 12000 rpm for 20 minutes using GSA type of rotor. The recovered pellet was then washed with five ml of distilled water to remove the precipitated ammonium sulfate. The solution was then centrifuged at 15000 rpm, using SS34 type of rotor, for 30 minutes. Then the pellet was then dissolved in 500 μ l of *TaqI* buffer and was stored at -20°C. The supernatant was checked for *TaqI* endonuclease activity.

5.8. Chromatographic Purification of the *TaqI* Restriction Endonuclease

5.8.1 Purification of the Fusion Protein on Cross-linked Amylose

A sterile injection needle with a total volume of 15 ml was used during chromatographic purification experiments. The column was first stabilized by a piece of glass wool and filled with five ml of amylose resin. The column was washed with 3 column volumes of column wash buffer and the flow rate was adjusted to be one drop per second. Two ml of crude cell extract was loaded onto the column after dilution to total volume of 10 ml with column wash buffer and the flow-through fractions were collected into sterile tubes for further analysis. The column was then washed with three column volumes of column wash buffer to remove the unbound proteins.

The MBP-fusions were eluted with ten ml column wash buffer containing 10 mM Maltose. 500 µl eluate fractions were collected on ice and their protein concentration was measured by the Bradford method as described in Methods 5.9 and the activity of the fractions were determined.

5.8.2 Cleavage with Factor Xa

The eluates containing the fusion protein were collected and their concentration was adjusted to one mg/ml. The protease digestion was accomplished by adding 0.2-1 mg Factor Xa to 50 mg of the fusion protein in buffer containing 20 mM Tris-HCl, 100 mM NaCl and 2mM CaCl₂. Reaction mixtures were incubated at room temperature for 16 hours.

5.9. Determination of Protein Concentration

The protein concentration was determined by using a Sigma protein assay kit, based on the Bradford Dye-binding Method (Bradford, 1976). Bradford reagent was diluted at a ratio of 1:4. A series of tubes containing one to ten μg of Bovine Serum Albumin were prepared in a final volume of 1000 μl . The dilutions of BSA were added to 1 ml of Bradford reagent. After five minutes of incubation at room temperature, the absorbance of each sample were read at 595nm against a blank containing only the diluted Bradford reagent and the protein buffer. A standard curve showing protein amount in μg versus absorbance at 595 nm was obtained. The protein concentration in each sample was then calculated using this standard curve.

5.10. Measurement of *TaqI* Restriction Endonuclease Activity

A series of dilutions were prepared from the isolated *TaqI* restriction endonuclease and parallel digestions were carried out with the commercial *TaqI* endonuclease to compare the activities at different dilutions. The isolated *TaqI* restriction endonuclease was dissolved in column wash buffer and the digestion reaction of the purified enzyme was carried out using Multicore Buffer. All conditions for digestion were kept constant both for the isolated and the commercial *TaqI* restriction endonuclease.

Activity was analyzed by digesting 2 μg of plasmid pUC18 at 65°C waterbath for one hour by covering the top with mineral oil to prevent evaporation. The digestions were then run on 1.2 per cent agarose using pUC18 digested with commercial *TaqI* restriction endonuclease as a positive control.

5.11. SDS-polyacrylamide gel electrophoresis of proteins

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis is a low cost, reproducible and rapid method for quantifying, comparing and characterization of proteins. This method separates proteins based on their molecular weights (Laemmli, 1970). SDS binds to hydrophobic portions of a protein, disrupting its folded structure and allowing it to exist stably in solution in an extended conformation. As a result, the length of the SDS-protein complex is proportional to its length (Bollag *et al.*, 1996).

After the gel sandwich is assembled the separating gel is prepared with TEMED and 10 % APS added last. Then the solution is poured carefully into the gel sandwich about 2 cm from the top of the front plate. The top of the separating gel is covered with water to prevent contact with air and to keep the gel surface flat. The gel is allowed to polymerize and then the water is poured off. The stacking gel is prepared with TEMED and 10 % APS added last again and is poured on top of the separating gel. The comb is carefully inserted and the stacking gel is allowed to polymerize for about half an hour. After polymerization, the comb is removed and the gel is placed in the electrophoresis chamber. Electrophoresis buffer is added to the inner and outer reservoirs and any air bubbles trapped in the wells are removed by rinsing.

The protein sample is mixed with sample buffer in an eppendorf tube and the mixture was heated at 98° C for 3 minutes. The sample solution is introduced into the well and molecular weight standards are introduced to one or more of the wells. The electrode plugs are attached to the proper electrodes and the power supply is turned on to 200 V. when the dye front has migrated to the bottom of the gel, the power supply is turned off, the gel plates are pried apart and the gel is removed and placed into staining solution for half an hour. The stain is then poured out and destain is added to get rid of unbound stain. After half an hour, destain is poured out and fixing solution is added to fix the proteins. The gels were then dried onto chromatography paper under vacuum for 30 minutes at 80°C.

5.12. DNA sequencing

The dideoxy chain termination method of sequencing of the insert was performed using the Sanger method of dideoxy mediated chain termination. Synthesis is initiated at the site where a primer anneals to the template. Elongation of the 3' end of the annealed primer is catalyzed by DNA polymerase in the presence of 2'-deoxynucleoside-5'-triphosphates (dNTPs) and is terminated by the incorporation of a 2',3'-dideoxynucleoside-5'-triphosphate analog (ddNTP) that will not support continued DNA elongation, hence the name chain termination. Sequence determination is carried out as a set of four separate reactions, each of which contains all of four dNTPs supplemented with a limiting amount of one of the ddNTPs per reaction. The growing oligonucleotide is terminated selectively at G, A, T or C. The resulting fragments, each with a common origin but ending in a different oligonucleotide, are separated according to size.

The insert in the plasmid is first amplified using the malE primer as the forward primer supplied in the pMAL fusion and purification kit and the EcoRI primer as the reverse primer, the dNTPs and *Pfu* DNA polymerase. The PCR product is identified and determined to be pure using 1.2 % agarose gel electrophoresis.

The PCR products were purified using the modified glass beads protocol (Vogelstein and Gellespie, 1989). DNA binds specifically to prepared glass particles in the presence of sodium iodide. The resulting precipitate is washed to remove NaI and impurities and then DNA is eluted from the glass with water or TE buffer.

All the PCR reactions were collected together and three volumes of NaI solution was added to the reaction mixture. The resuspended silica powder suspension was added and the mixture was incubated at 55°C for 5 minutes. The silica powder/DNA complex was centrifuged briefly for 5 seconds. The supernatant was removed. The pellet was washed with ice cold wash buffer (200-600 µl) three times. The pellet was resuspended completely during each washing. The DNA was then eluted with water. The pellet was resuspended in sterile deionized buffer and incubated at 55°C for 5 minutes. The tube was centrifuged briefly and the supernatant was removed avoiding the pellet. Elution was performed once

more to purify as much of the DNA as possible. The eluted DNA was then checked on agarose gel.

The OmniBase DNA Cycle Sequencing System of Promega was employed in the sequencing reactions. Four tubes were labeled (G, A, T, C). 2 μ l of the appropriate d/ddNTP mix was added to each tube. 500 fmol (7 μ l) of template DNA, 3.0 pmol of primer, 10 μ Ci of [α - ^{35} S]dATP, DNA sequencing 5X buffer and nuclease free water were mixed for each set of four sequencing reactions. 1.0 μ l of OmniBase Sequencing Enzyme Mix was added to this mix and mixed gently. Then 4 μ l of this mix was added to each tube containing d/ddNTP mix. The tubes were then placed in the thermal cycler preheated to 95 °C. The PCR reactions were carried out as 2 minutes of initial denaturation at 95 °C, 30 seconds of denaturation at 95°C, 30 seconds of annealing and extension at 70 °C for a total of 30 cycles. 3 μ l of stop solution was added to terminate the reactions at the end of the cycles.

The products of the sequencing reactions were run on 6 per cent denaturing polyacrylamide gels. Electrophoresis is carried out in 1XGTG buffer at constant 80 W. The gel is preheated to 55°C to keep the samples denatured during electrophoresis. The samples are denatured at 95°C for 2 minutes and chilled on ice immediately before loading onto gel. The long run of 5 hours allows determination of the sequences in the middle of the gene while the short run of 2 hours allows the determination of the sequences at the two termini of the gene.

After electrophoresis, the plates are separated and a 3 mm Whatman paper is placed onto the gel. When the Whatman paper is removed, the gel is stuck onto it. The gel was then dried on a gel drier for 1 hour at 80° C. The gel was exposed to an X-ray film with an intensifying screen in a cassette at -70° C overnight or at room temperature without the screen. The X-ray film was developed and fixed.

5.13. Determination of Growth Characteristics

Fifty ml of LB medium containing ampicillin was inoculated with a single recombinant colony and grown overnight at 37 °C. Two 2 liters of LB media were inoculated with twenty ml of this preculture and incubated at 37 °C, 180 rpm in an orbital shaker. One ml samples were taken from the cultures every thirty minutes or one hour and their absorbance values at 600 nm wavelength against sterile LB blank were recorded. 5 ml of sample was taken from the cultures, centrifuged for 10 minutes in a Nüve benchtop centrifuge. The supernatant was discarded and the suspension was transferred to a sterile preweighed eppendorf. The suspension was centrifuged again in an Eppendorf centrifuge and the supernatant was discarded. The pellet was dried at 65 °C for 6 hours and then the eppendorfs were weighed again to determine the dry mass. SDS samples were also taken from the cultures. 500 ml of sample was taken, centrifuged and the supernatant was discarded. The pellet was dissolved in SDS sample buffer and left at -20 °C until loading.

6. RESULTS AND DISCUSSION

The gene encoding *TaqI* restriction endonuclease was cloned into the pMAL-c2 vector to produce MBP-*TaqI* restriction endonuclease fusion protein in *Escherichia coli*. *TaqI* restriction endonuclease gene from *Thermus aquaticus* genomic DNA was amplified using both *Taq* DNA polymerase and *Pfu* DNA polymerase. *Pfu* DNA polymerase is known to provide higher fidelity in the PCR amplification of DNA. Expression of the cloned gene in three different strains of *Escherichia coli* was investigated.

6.1. Amplification of the *TaqI* Restriction Endonuclease gene by *Taq* DNA polymerase

6.1.1 Cloning of the gene into pMAL-c2 expression vector

In the first part of this study, *Taq* DNA polymerase was used to amplify the *TaqI* restriction endonuclease gene by PCR with *Thermus aquaticus* genomic DNA as template as described in Methods 5.4.1. 792 base pairs of *TaqI* restriction endonuclease gene and 16 additional nucleotides of the tailored ends of primers TQXmnI and TQEcoRI resulting in 808 base pairs of PCR product was identified by electrophoresis on 1.2 per cent agarose gels (Figure 6.1). The amplified product was then isolated from low melting point agarose gel as described in Methods 5.4.2.

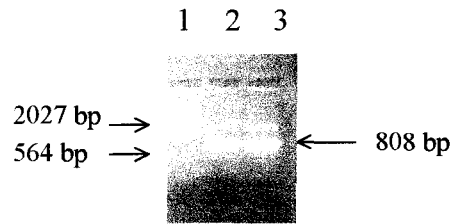


FIGURE 6.1. Electrophoretic analysis of the PCR product amplified by *Taq* DNA polymerase on 1.2 per cent agarose gel. Lane 1: λ-HindIII marker, Lanes 2 and 3: 1 µl of the PCR product from two different samples amplified under identical conditions

The insert purified from low melting point agarose gel and the vector isolated from *Escherichia coli* cells harboring pMAL-c2 were both digested with XmnI and EcoRI as described by the manufacturers and ligated using T4 DNA ligase as described in Methods 5.4.5. The resulting recombinant plasmid is 7432 bp long and harbors the *TaqI* restriction endonuclease gene downstream from the *malE* gene of *Escherichia coli*. The construction of the recombinant vector is described in Figure 6.2.

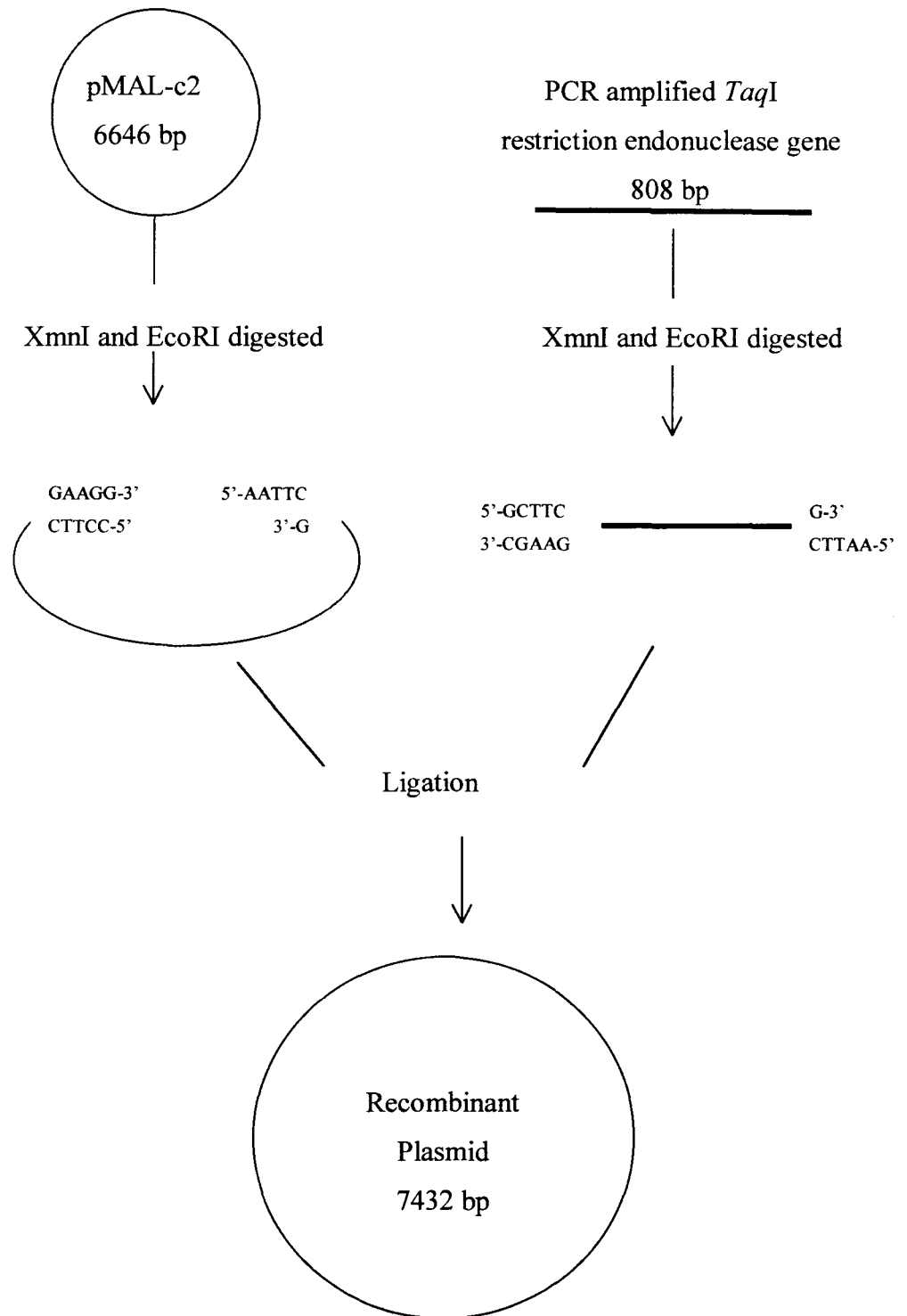


FIGURE 6.2. Strategy for cloning of *TaqI* restriction endonuclease gene into pMAL-c2 expression vector

Escherichia coli TB1 strain was transformed with this ligation mixture by electroporation as described in Methods 5.5.2. The cells harboring the plasmid pMAL-c2 or the recombinant plasmid confer resistance to ampicillin. Therefore, the transformed cells were first selected on LB-agar plates supplemented with ampicillin.

Colonies harboring plasmids larger than the pMAL-c2 vector were identified by cracking method as described in Methods 5.6.2 and 12 colonies were selected. Cracking method provided a rough yet rapid estimation of plasmid size useful for the selection of recombinant colonies. Figure 6.3 shows the results of cracking, a colony suspected to harbor a larger plasmid is indicated by an arrow.



FIGURE 6.3. Rapid analysis of plasmid size by cracking method Lane 1: pMAL-c2. The plasmid suspected to have a larger size is indicated by an arrow.

These 12 colonies harboring the plasmid larger than pMAL-c2 were grown and their plasmids were isolated by the miniprep method as described in Methods 5.2.2. Using these plasmids and the forward and reverse primers TQXmnI and TQEcoRI, the insert was amplified by PCR. The amplified products were inspected by agarose gel electrophoresis. The results indicated that 6 of these suspected colonies harbor an insert of approximately expected size (Figure 6.4).

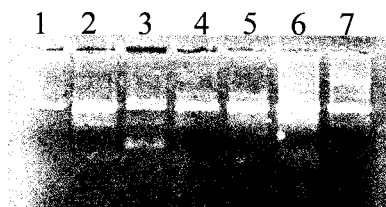


FIGURE 6.4. Electrophoretic analysis of the PCR products in suspected colonies. Lane 1: fragment obtained by PCR amplification of the *Thermus aquaticus* genomic DNA used in cloning as a control, Lane 2: colony 7, Lane 3: colony 5, Lane 4: colony 4, Lane 5: colony 6, Lane 6: colony 8, Lane 7: colony 10

6.1.2 Expression of the fusion protein in *Escherichia coli* TB1 cells

In the pMAL-c2 expression system, the target gene is expressed as a fusion protein with maltose binding protein. Maltose binding protein is a periplasmic protein of *Escherichia coli*, however, pMAL-c2 has an exact deletion of the *malE* sequence, resulting in the cytoplasmic expression and accumulation of the fusion protein. The target gene is inserted downstream from and in the same reading frame as the *malE* gene. The *malE* gene of the vector is under the control of the strong *tac* promoter (di Guan *et al.*, 1988), hence expression of the gene is induced by a gratuitous inducer; isopropyl-1-thio- β -D-galactopyranoside (IPTG) which is a lactose derivative.

These 6 recombinant colonies selected in the present study were grown and induced with 1 mM IPTG in the mid-exponential phase of growth at an optical density of 0.7. The cells were harvested after an induction period of 3 hours and the crude extract was prepared as described in Methods 5.7.1 and the endonuclease activity in the crude extracts was assayed, but no endonuclease activity could be obtained in any of them. The size of the proteins in the crude extracts was analyzed by SDS-PAGE; however, it was observed that the number of proteins was too high to detect the presence of the fusion protein. Therefore, we decided to purify the crude extract and isolate the fusion protein from the cytoplasmic protein pool by amylose column chromatography.

The molecular mass of the maltose binding protein expressed by the pMAL-c2 vector is 42700 Da, including the amino acid residues coded for by the polylinker upto the XmnI site. (Duplay *et al.*, 1984) The molecular mass of full length *TaqI* restriction endonuclease was reported to be 31600 Da (Barany *et al.*, 1992). Hence, the resulting fusion protein expressed by the pMAL-c2-*TaqI* system is expected to have a molecular mass of approximately 74300 Da. The cytoplasmic proteins were purified by amylose affinity chromatography as described in Methods 5.8.1. Two of these colonies (c6 and c7) were found to encode a protein with the expected size of 74 kDa by SDS-PAGE analysis on 10 per cent polyacrylamide gels, as seen in Figure 6.5. However, neither the crude extracts nor the purified fusion proteins from the colonies gave endonuclease activity. Therefore, c7 was chosen as the recombinant colony for further research since it encoded for an expected size protein with a high yield.

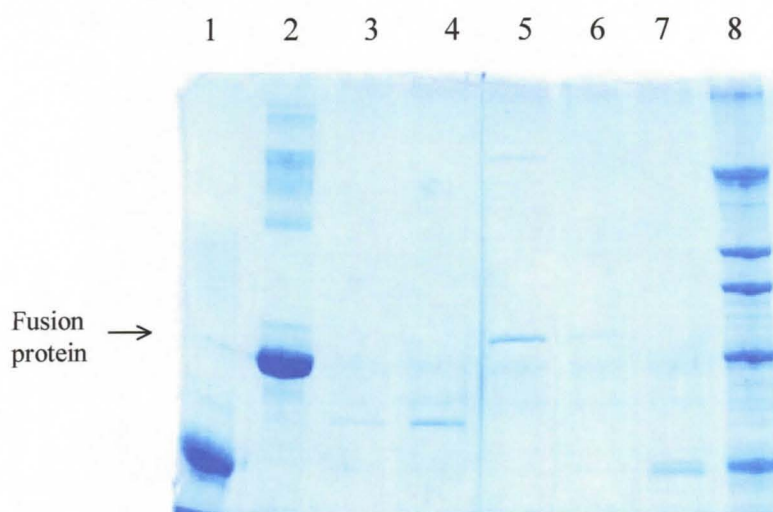


FIGURE 6.5. SDS-PAGE analysis of the proteins extracted from the suspected colonies and subjected to amylose affinity chromatography. Lane 1: 45 kDa protein marker, Lane 2: 66 kDa protein marker, Lanes 3-7: recombinant colonies pC11, pC8, pC7, pC6, pc5 respectively, Lane 8: 45, 66 ,97.4 ,116, 205 kDa protein markers

The recombinant colony c7 producing the expected size of protein was grown in LB medium containing ampicillin and induced for 4 hours with IPTG. The crude extract was prepared and was applied to an amylose column. The unbound proteins were washed out with column wash buffer. The MBP fusion proteins were then eluted with column wash buffer containing 10 mM maltose. The SDS-PAGE analysis showed that the eluted fractions from the column contain the fusion protein of expected size in very large amounts (Figure 6.6).

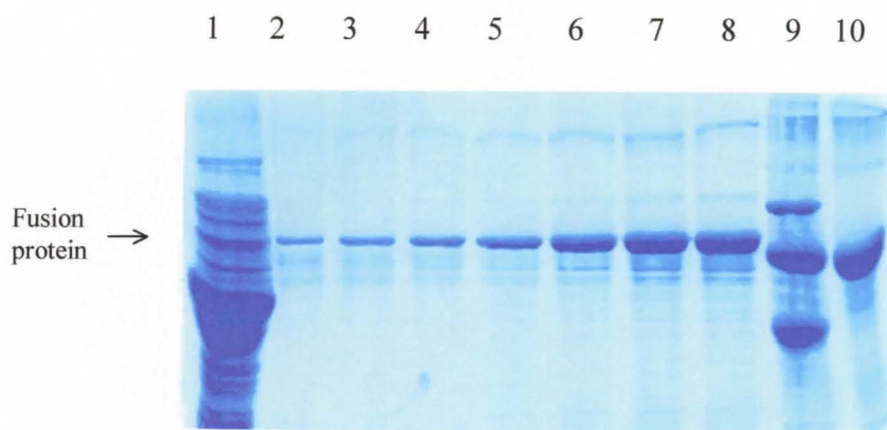


FIGURE 6.6. SDS-PAGE analysis of the column fractions obtained by amylose affinity chromatography Lane 1: cytoplasmic extract of the colony c7, Lanes 2-8: Fractions 1-7, Lane 9: 45, 66, 97 kDa protein markers, Lane 10: 66 kDa protein marker

Purified fusion protein was assayed for *TaqI* restriction endonuclease activity and found to display no activity. The cleavage of the MBP portion of the fusion protein with Factor Xa resulted in the production of a protein of 32 kDa corresponding to the expected size of *TaqI* restriction endonuclease. However, no restriction endonuclease activity could be observed in these digested samples either.

It has been reported that production of the protein in the reducing environment of the cytoplasm may prevent formation of disulfide bonds necessary for proper folding (Schein, 1989). Moreover, high level expression of the protein may result in the incorrect folding of the protein and may form insoluble aggregates or inclusion bodies (Georgiou *et*

al., 1996). Therefore, in order to prevent the possible formation of inclusion bodies due to overexpression in the cytoplasm, expression under different cultivation conditions in different strains was examined.

First, the effect of IPTG concentration on gene expression was investigated. The cells were induced with 0.5, 0.15, 0.05 and 0.025 mM of IPTG and the cell extracts were prepared. No endonuclease activity could be detected in any of these cell extracts. SDS-PAGE analysis of cytoplasmic proteins showed the presence of the fusion protein in all cell cultures (Figure 6.7). Even very low levels of the inducer was found to be sufficient for the synthesis of large amounts of the fusion protein.

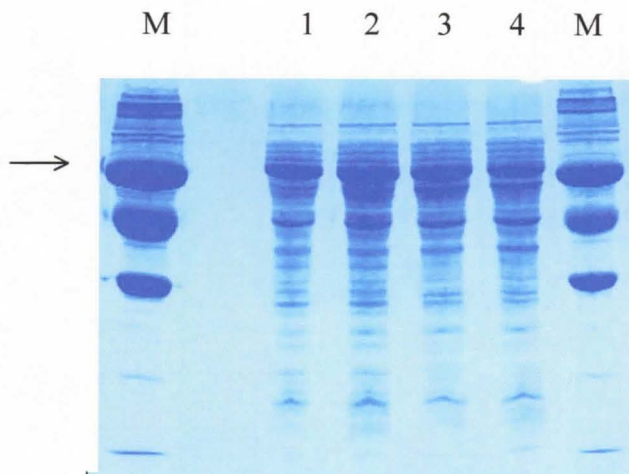


FIGURE 6.7. SDS-PAGE analysis of cytoplasmic proteins obtained from culture induced with decreasing concentrations of IPTG, M: 45, 66 97.4 kDa protein size markers. Lane 1: 0.5 mM, Lane 2: 0.15 mM, Lane 3: 0.05 mM, Lane 4: 0.025 mM, Arrow indicates the fusion protein

The reduction in growth temperature has been reported to favor formation of soluble proteins due to factors such as decrease in the driving force for protein self association, slower rate of protein synthesis that allows sufficient time for proper folding and changes in the folding kinetics of the polypeptide chain (Georgiou *et al.*, 1996). Consequently, the cells were grown at 37°C and the temperature was shifted to 30°C and 25°C after induction. The induction period was prolonged for 2 more hours for every

approximately 7°C of reduction in growth temperature. Hence, induction period was 6, 8 and 10 hours for the cells induced at 37°C, 30°C and 25°C respectively. The results indicated a decrease in the yield of fusion protein as induction temperature decreased, however, the growth of cells at lower temperatures did not result in the production of an active protein.

The cells were also grown in the presence of glycerol. Glycerol provides osmotic stress, decreases diffusivity and the partial molar volume of the proteins thereby reducing the rate of aggregate producing solute interactions (Schein, 1996). No improvement in the production of active *TaqI* restriction endonuclease could be achieved, although the expected size of protein was produced.

Expression of the *TaqI* restriction endonuclease fusion protein in different strains of *Escherichia coli*, namely XL1-Blue and ER2508 was also investigated. XL1-Blue is a *recA* strain, which provides higher stability of the recombinant plasmid, and ER2508 is a protease deficient strain. These strains were transformed with the recombinant plasmid (pC7) constructed. Cloning of the recombinant plasmid into XL1-Blue strain resulted in the expression of the gene with a very high yield as determined by SDS-PAGE analysis, but no endonuclease activity could be observed in either strain.

6.1.3 Sequencing of the insert

Since the growth and induction condition manipulations and expression in different strains proved unsuccessful in the expression of active protein, the sequence of the insert in the recombinant colony was determined as described in Methods 5.12. It was found that the adenine in position 410 within the *TaqI* restriction endonuclease gene was replaced by a thymine, resulting in the production of a valine, instead of aspartic acid (V137D). The substitution of valine to aspartic acid at codon 137 resulted in the complete loss of activity suggesting that codon 137 is a part of the active site. This substitution in the gene sequence may possibly be due to lack of fidelity of *Taq* DNA polymerase. Therefore, we decided to clone the gene using an amplification system with higher fidelity.

6.2. Amplification of the *TaqI* restriction endonuclease gene by PCR with *Pfu* DNA polymerase

Pfu DNA polymerase is a thermostable DNA polymerase from *Pyrococcus furiosus*, which has 5' to 3' DNA polymerase and 3' to 5' proofreading exonuclease activity that increases the fidelity of DNA synthesis. This DNA polymerase is used in applications where high fidelity of PCR is required, such as the amplification of a gene used for cloning. The accuracy of amplification is important since the amino acid sequence determines the folding and enzyme activity of a protein. Therefore, *Pfu* DNA polymerase was used in the second part of this study for the amplification and cloning of the *TaqI* restriction endonuclease gene.

6.2.1 Optimization of PCR conditions

The PCR conditions used in the first part of this study for the amplification of the *TaqI* restriction endonuclease gene with *Taq* DNA polymerase proved to be unsuccessful in PCR with *Pfu* DNA polymerase mainly because of its exonuclease activity. Therefore, the reaction conditions were optimized again. *Pfu* DNA polymerase has a chain elongation speed of 2.0 minutes/kb which is 5 fold slower than the rate of nucleotide incorporation of *Taq* DNA polymerase. Hence, *Pfu* DNA polymerase requires a longer extension time. Furthermore, *Pfu* DNA polymerase requires a high level of dNTPs to ensure that its exonuclease function will not start to degrade the DNA product when one of the nucleotides cannot be incorporated. Optimum conditions of the PCR was determined and PCR was carried out in the presence of 300 μ M dNTPs, 0.5 μ M of each primer and 1.25 μ Units of *Pfu* DNA polymerase with *Thermus aquaticus* genomic DNA as the template. The reactants were incubated at 96°C for 5 minutes as the initial denaturation step. Then, 25 cycles of PCR were carried out with 30 seconds of denaturation at 96°C, 30 seconds of primer annealing at 66°C and 2 minutes of chain elongation at 72°C. The number of cycles was kept low to make sure that the dNTPs were not exhausted. The final chain elongation

step was at 72°C for 10 minutes. Electrophoretic analysis of the PCR product indicated the presence of a fragment approximately 800 basepairs long as expected (Figure 6.8).

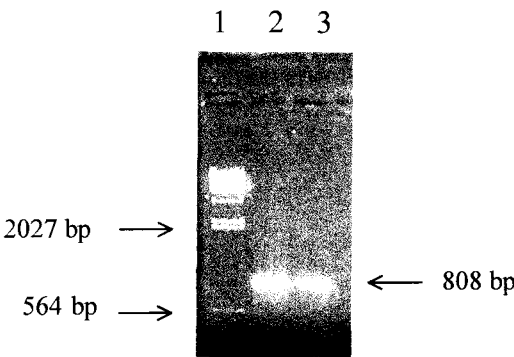


FIGURE 6.8. Electrophoretic analysis of the PCR product amplified by *Pfu* DNA polymerase Lane 1: λ -HindIII marker, Lanes 2 and 3: 1 μ l of the PCR product from two different samples amplified under identical conditions.

6.2.2 Cloning of the *TaqI* restriction endonuclease gene into pMAL-c2 expression vector

The PCR product and the pMAL-c2 vector were both digested first with XmnI and then with EcoRI. T4 DNA ligase was used to ligate the insert and the vector. Analysis of the ligation products by electrophoresis on 1.2 per cent agarose gel showed the presence of dimers and trimers of insert, the recircularized plasmid as well as the recombinant plasmid which has a slightly larger size than the vector (Figure 6.9).

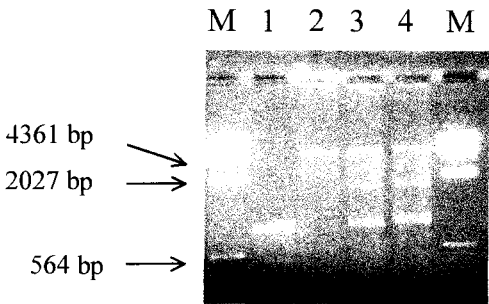


FIGURE 6.9. Electrophoretic analysis of the ligation reaction. M: λ -HindIII size marker, Lane 1: double digested *TaqI* restriction endonuclease gene, Lane 2: double digested vector, pMAL-c2, Lane 3: ligation mixture with 7 μ l of vector, 4 μ l of insert, Lane 4: ligation mixture with 7 μ l of vector and 7 μ l of insert

The competent cells were prepared as described in Methods 5.5.1 and transformed with the ligation mixture.

Ampicillin selection was used as the first step in the selection of recombinant colonies since the vector harbors an ampicillin resistance gene. 175 colonies were identified on LB agar plates supplemented with ampicillin indicating the presence of the vector with or without the insert.

α -complementation was used to select the colonies that carried an insert within their plasmids. The pMAL-c2 vector confers a lac^+ phenotype on strains containing the $\text{lacZ}\Delta\text{M15}$ deletion. The insertion of a DNA fragment into the polylinker abolishes the α -complementation. Thus, colonies containing the vector with an insert can easily be selected on LB agar plates containing 100 $\mu\text{g/ml}$ ampicillin, 10 $\mu\text{g/ml}$ X-Gal and 1 mM IPTG. Colonies containing the vector alone will turn blue while recombinant colonies with their lacZ gene disturbed will remain white. We have screened 175 colonies using this test. 20 colonies turned blue, while 155 colonies remained white in the present study.

Cracking method was then used as described in Methods 5.6.2 to determine the colonies harboring the plasmid larger than the vector. 155 colonies were screened by

cracking method and 14 colonies were shown to have plasmids larger than pMAL-c2 (Figure 6.10).

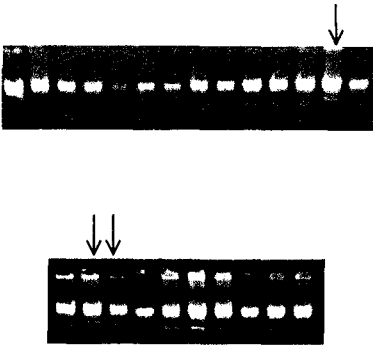


FIGURE 6.10. Rapid analysis of the plasmid size by cracking method lane 1: pMAL-c2
The plasmids suspected to have a larger size than the vector without an insert are indicated by arrows.

The colonies harboring a plasmid larger than pMAL-c2 were further analyzed by digestion with restriction enzymes. The plasmids from these 14 colonies were isolated by miniprep method as described in Methods 5.2.2. The restriction enzyme digestions were performed as described by the manufacturer and the size of the DNA fragments obtained upon digestion were compared with the size estimated from the restriction map of the recombinant plasmid.

The pMAL-c2 vector has a size of 6646 basepairs. It has been digested by XmnI and EcoRI resulting in the loss of 7 basepairs. The PCR product has 808 base pairs 15 of which are lost upon digestion with XmnI and EcoRI. Hence the recombinant vector should have a size of 7432 basepairs. *TaqI* restriction endonuclease gene has unique restriction sites for HindIII and PstI. Moreover, the PstI and HindIII sites in the polycloning site of pMAL-c2 have been preserved. As a result, HindIII digestion of the recombinant plasmid containing *TaqI* restriction endonuclease gene as an insert should result in the formation of fragments of 774 and 6658 basepairs, while PstI digestion should yield fragments of 261

and 7171 basepairs. The restriction map of the recombinant pMAL-c2 vector is presented in Figure 6.11.

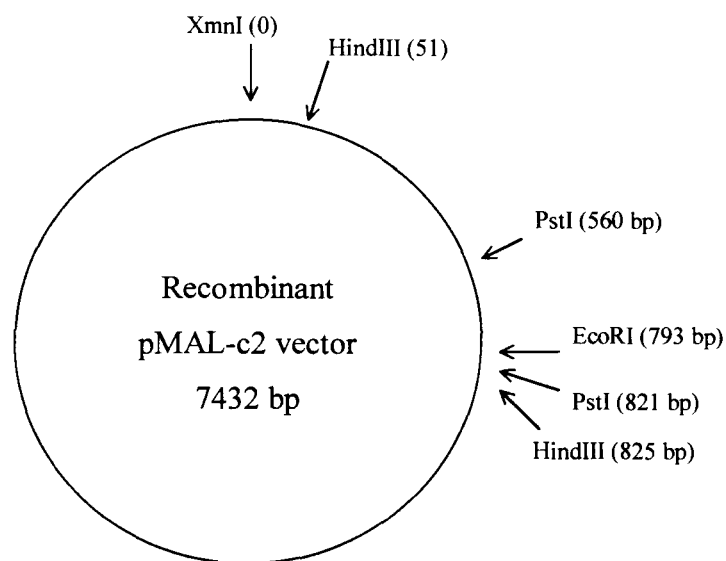


FIGURE 6.11. Restriction Map of the Recombinant pMAL-c2 Vector

The restriction enzyme analysis performed on the 14 colonies indicated the presence of a recombinant colony (c22) giving the expected pattern of restriction (Figure 6.12). As a result, one colony harboring the insert within the plasmid giving the correct restriction digestion pattern was identified.

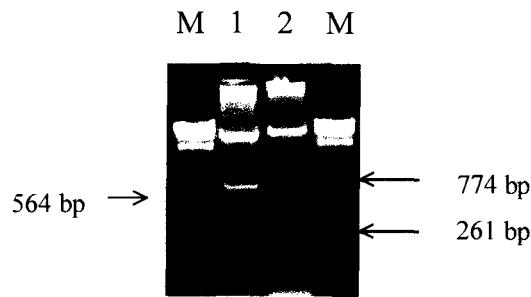


FIGURE 6.12. Restriction analysis of the recombinant colony harboring the plasmid pC22
M: λ -HindIII marker, Lane 2: pC22-HindIII digest, Lane 3: pC22-PstI digest

6.2.3 Expression of the MBP-*TaqI* fusion protein in *Escherichia coli* TB1 cells

Even though one colony (c22) was identified to give the expected restriction enzyme analysis results, expression in all 14 suspected colonies was studied as the final step in the selection of the correct recombinant construct.

The suspected recombinant TB1 colonies were grown in 20 ml LB medium containing ampicillin and induced with 1 mM IPTG in their mid-exponential phase. Cells were harvested after an induction period of 3 hours and the crude extract was prepared as described in Methods 5.7.1. Analysis of the crude extracts for endonuclease activity by digestion of λ -DNA at 65 °C revealed that partial endonuclease activity was present in the cytoplasmic extract of the colony c22 which was also selected by restriction enzyme analysis (Figure 6.13).

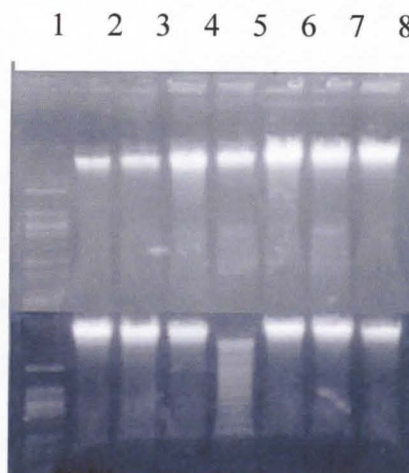


FIGURE 6.13. Analysis of *TaqI* restriction endonuclease activity in the cytoplasmic extracts of the suspected colonies. Upper panel: Lane 1: λ -DNA digested with commercial *TaqI* restriction endonuclease, Lanes 2-8: colonies 1, 32, 65a, 7, 67, 66, 24 Lower panel: Lane 1: λ -DNA digested with commercial *TaqI* restriction endonuclease, Lanes 2-8: colonies 40, 45, 11a, 22, 11b, 44, 65b

The sequence of the insert within the plasmid of this colony was determined as described in Methods 5.12. Sequencing of the insert confirmed that it had the DNA sequence of *TaqI* restriction endonuclease as expected. Thus, this colony harboring the correct recombinant construct with the correct insert and coding for an active MBP-*TaqI* restriction endonuclease fusion protein was identified for further studies.

A pilot experiment was performed to verify the production of a fusion protein with expected size and activity and to determine optimal induction time and period for the recombinant colony c22. The cells were grown in parallel batch cultures to $OD_{600}=0.6$ and $OD_{600}=0.9$ respectively and induced with 1 mM IPTG. Samples were taken 3, 4, 5 and 10 hours after induction and the crude extracts were prepared. The results indicated a slight increase in the recovery of *TaqI* restriction endonuclease activity when cells were induced at $OD_{600}=0.9$ for a period of 5 hours.

The crude extracts were then heat treated for 10 minutes at 65°C for the partial purification of this fusion protein containing thermostable *TaqI* restriction endonuclease. The recovery of active proteins was unreliable and unpredictable with this purification method, often resulting in total loss of activity.

The SDS-PAGE analysis of the crude extracts confirmed the presence of the fusion protein with the expected size (Figure 6.14)

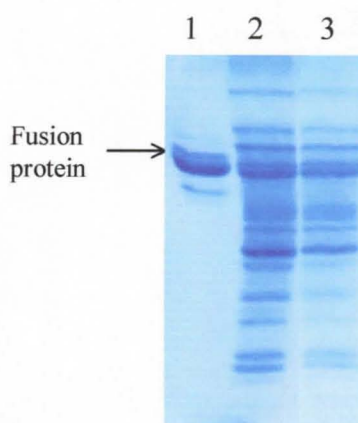


FIGURE 6.14. SDS-PAGE analysis of the crude and heat treated extracts from colony c22 induced for a period of 4 hours at midexponential phase. Lane 1: Fusion protein obtained by amylose column chromatography in the first part of this study as a size marker, Lane 2: crude extract, Lane 3: heat treated extract

The cells were grown and induced at $OD_{600} = 0.9$ for 5 hours using 1 mM IPTG. Cell extracts were prepared and enzyme activity was measured. 5 Units of *TaqI* restriction endonuclease activity was obtained per μl of crude extract. The specific activity of the enzyme was determined as 1041 U/mg. The recovery on the other hand was 245,000 Units of enzyme per liter of culture volume and 40500 Units of enzyme per gram of wet cell weight. The results of induction at $OD_{600} = 0.9$ are presented in Table 6.1.

TABLE 6.1. *TaqI* restriction endonuclease activity of the crude extracts of the recombinant colony induced at $OD_{600} = 0.9$ for five hours

Induction OD	Wet cell weight (g)	Protein conc. (mg/ml)	Total protein (mg)	Volumetric activity U/ μ l	Total units	Specific activity U/mg	U/liter of culture volume	U/g of wet cell weight
0.9	1.21	4.8040	48.04	5	50000	1041	250000	41322

Consequently, the active fusion proteins were purified from the total cytoplasmic protein pool by amylose column chromatography. 0.5 U/ μ l of activity was detected in the fraction with the highest activity. This fraction contained a protein concentration of 0.1782 g/ml. Hence the specific activity of the fraction with the highest activity is 2806 U/mg.

The SDS-PAGE analysis of the crude extract, heat treated samples and the column fractions reveal that the fusion protein is cleaved upon expression resulting in degradation products slightly smaller than the fusion protein observed as a multiple banding pattern (Figure 6.15). These cleavage products retain their MBP activity as confirmed by their purification with affinity column chromatography. The presence of a protein with the approximate size of MBP may be the result of chromosomal MBP expression that did not have the time to be transported to the periplasm, or it may be due to cleavage of the fusion protein. In either case, the presence of MBP and cleavage products might reduce specific activity of amylose column fractions, decreasing the efficiency of purification.

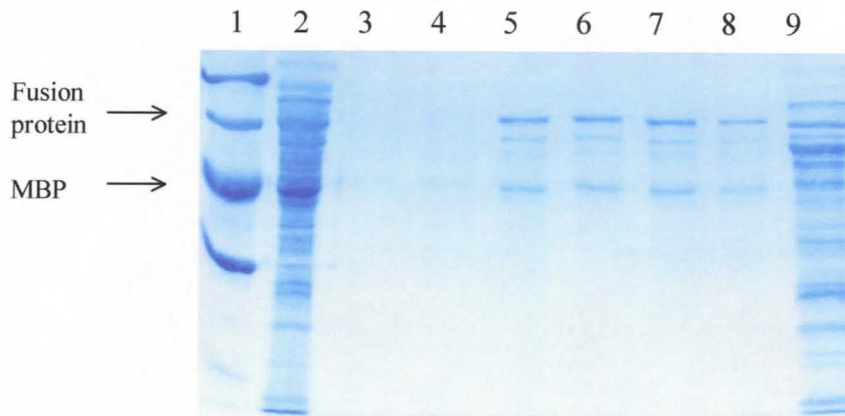


FIGURE 6.15. SDS-PAGE analysis of the fusion proteins obtained by expression in late exponential phase. Lane 1: 45, 66, 97, 116 kDa protein markers, Lane 2: Total cytoplasmic proteins, Lanes 3-8: Column fractions 2-7 obtained by amylose affinity chromatography, Lane 9: Heat treated sample

6.2.4 Optimization of induction conditions for recombinant colony c22

In order to improve the recovery of endonuclease activity obtained, growth and induction conditions need to be optimized. Therefore, growth characteristics of the recombinant colony were investigated. The cells were grown in 2 liters of LB medium in a total volume of 6 liters and induced with 1 mM IPTG in exponential phase when the OD_{600} was 0.6 . Samples were taken from the culture in one hour or half hour intervals to determine the absorbance of the culture and the dry weight of the cells as described in Methods 5.13. Growth characteristics of the recombinant colony were compared with those of the uninduced recombinant colony grown under the same conditions (Figure 6.16). It was observed that the induced cells reached stationary phase 4 hours after induction, however 6 hours were needed for the uninduced cells to reach the stationary phase.

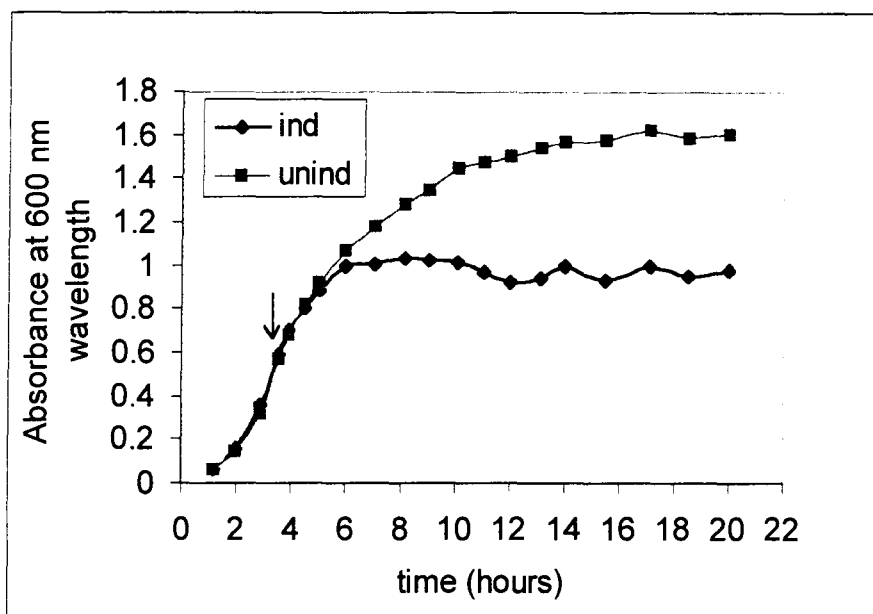


FIGURE 6.16. Growth curve of the recombinant colony 22 with and without induction. Time of induction is indicated by an arrow

It is an established fact that recombinant protein production retards cell growth. The results found in the present study for the recombinant colony expressing the MBP-*TaqI* restriction endonuclease fusion protein confirm that cell growth is retarded upon induction and as a result, entered the stationary phase earlier than uninduced recombinant colony.

To optimize time of induction of fusion protein expression, 3 batch fermentation experiments were performed in parallel in 200 ml of LB medium in a total volume of 1 liters. The cells were induced at $OD_{600} = 0.6$ at $OD_{600} = 0.9$ and at $OD_{600} = 1.2$ respectively. Cell extracts were prepared after 5 hours of induction. *TaqI* restriction endonuclease activity and protein concentration were measured in these cell extracts. 300000 Units of enzyme per liter of culture was obtained from 1.31 g of cells when the cells were induced at late exponential phase of the growth. On the other hand, 200000 Units of *TaqI* restriction endonuclease could be obtained from 1.18 and 1.44 g of cells when the cells were induced at $OD_{600} = 0.6$ and $OD_{600} = 1.2$ respectively.

The effect of the time of induction on *TaqI* endonuclease production in flask cultures of recombinant cells is summarized in the Table 6.2. The wet cell weights obtained after harvesting the cells confirm that later induction allows growth to higher cell densities. Total recovery of endonuclease activity obtained after induction in late exponential phase is the highest, at 300000 Units/liter of culture volume and 45802 U/gram of cell weight of enzyme could be obtained in the cell extracts. The specific activity was found to be slightly higher when induction was made in mid-exponential phase because of the lower amount of total cell protein.

TABLE 6.2. *TaqI* restriction endonuclease activity of the crude extracts of the recombinant colony induced at different phases of growth

Induction OD	Wet cell weight (g)	Protein conc. (mg/ml)	Total protein (mg)	Volumetric activity U/ μ l	Total units	Specific activity U/mg	U/liter of culture volume	U/g of wet cell weight
0.6	1.18	4.5378	45.378	4	40000	881	200000	33898
0.9	1.31	7.6401	76.401	6	60000	785	300000	45802
1.2	1.44	5.6312	56.312	4	40000	710	200000	27778

6.2.5 Expression of the fusion protein in *Escherichia coli* XL1-Blue and ER2508 strains

The level of fusion protein expression was determined to be variable between different strains of *Escherichia coli* in the overproduction of recombinant MBP-decorin core protein fusions (Hering *et al.*, 1996). Significantly higher yield of recombinant decorin fusion protein was reported in XL1-Blue as compared to TB1.

We have used three different strains of *Escherichia coli* in order to study the effect of strain selection on the expression of recombinant MBP-*TaqI* restriction endonuclease fusion protein. The recombinant plasmid from colony 22 that expresses the fusion protein

with the expected size and expected endonuclease activity was isolated from the recombinant TB1 *Escherichia coli* cells. *Escherichia coli* XL1-Blue and ER2508 strains were transformed with the recombinant plasmid by electroporation. Expression in these strains was studied and compared.

Three batch fermentation experiments were performed with recombinant TB1, XL1-Blue and ER2508 cells harboring the plasmid that expressed the *TaqI* endonuclease gene as a fusion protein with MBP. All three cultures were induced with 1 mM IPTG at optical density of 0.9 at 600 nm wavelength, for an induction period of 5 hours. The crude extracts were prepared and the *TaqI* restriction endonuclease activity and the protein concentration of these extracts were determined. The results are presented in Table 6.3. 300000 Units of enzyme per liter culture was obtained from 1.31 g of TB1 cells, 100000 Units of enzyme per liter culture was obtained from XL1-Blue cells and 150000 Units of active enzyme was obtained from ER2508 cells. On the other hand, the specific activity of 785 U/mg for TB1 cells is not significantly higher than the specific activity of 530 U/mg for XL1-Blue cells or 505 U/mg for ER2508 cells.

TABLE 6.3. *TaqI* restriction endonuclease activity of the crude extracts of three different strains

Strain	Wet cell weight (g)	Protein conc. (mg/ml)	Total protein (mg)	Volumetric activity U/ μ l	Total units	Specific activity U/mg	U/liter of culture volume	U/g of wet cell weight
TB1	1.31	7.6401	55.380	6	60000	785	300000	45802
XL1	1.71	3.7747	37.747	2	20000	530	100000	11696
ER2508	2.13	5.9410	59.410	3	30000	505	150000	14085

It was previously reported that prolonged IPTG induction of periplasmic expression of *TaqI* restriction endonuclease led to the secretion of the enzyme to the medium. (Özdinler H., 1996). Hence, to detect the presence of active endonuclease in the medium, recombinant cells were grown and induced for a period of 20 hours. The medium proteins were concentrated using ammonium sulfate precipitation as described in Methods 5.10 and

TaqI restriction endonuclease activity was measured. The absence of enzyme activity in the medium indicated that fusion protein is not excreted into the medium if it is not specifically targeted to the periplasm.

7. CONCLUSION AND RECOMMENDATIONS

TaqI restriction endonuclease is a thermostable enzyme with high specificity for its recognition site. It is a model enzyme in studying DNA-protein interactions as well as a valuable restriction endonuclease in molecular biology research. The low yield and high cost of isolation of the enzyme from its host, *Thermus aquaticus*, made it necessary to clone and express *TaqI* restriction endonuclease in *Escherichia coli* and to develop purification techniques.

pMAL-c2 expression vector which expresses the target protein as a fusion protein with Maltose Binding Protein (MBP) can be used for the high level expression of recombinant proteins. The affinity of MBP to amylose is exploited for the one-step chromatographic purification of the fusion protein. The protein is expressed under the control of the *tac* promoter, yielding a high recovery of the target protein.

In the framework of this thesis, *TaqI* restriction endonuclease was cloned into pMAL-c2 expression vector for overproduction in *Escherichia coli*. In the first part of this study, the *TaqI* restriction endonuclease gene was amplified by PCR using *Thermus aquaticus* genomic DNA as template and *Taq* DNA polymerase. The amplified gene was inserted into pMAL-c2 vector and the recombinant vector was then transferred to *Escherichia coli* TB1 cells.

Screening of the recombinants were performed by rapid estimation of plasmid size and by PCR analysis. It is very important to note that PCR analysis does not confirm that the insert in the recombinant vector is the target gene. PCR analysis gives proof only that the primer annealing sites are preserved and that the size of the insert is correct. The gene might have mutations that might hinder gene expression but that cannot be detected by PCR analysis. Seven colonies were identified as recombinant after the screening steps and

were further analyzed for endonuclease expression as the final step in selection of recombinant colonies. The crude extracts of the suspected colonies were subjected to amylose column chromatography and one colony was identified to code for a protein of the expected size. However, no detectable endonuclease activity could be observed. It has been reported that expression with MBP as fusion protein yields biologically active Staphylococcus nuclease A, Klenow fragment of the cytoplasmic *Escherichia coli* DNA polymerase I (Bedoulle *et al.*, 1988), PstI endonuclease, β -galactosidase (Guan *et al.*, 1988), paramyosin (Maina *et al.*, 1988), Gene 5 protein of phage M13 (Blondel *et al.*, 1990), Rep78 of Adeno-Associated Virus (Batchu *et al.*, 1995), Streptolysin O (Weller *et al.*, 1996), decorin core protein (Hering *et al.*, 1996), fatty acid synthase (Jayakumar *et al.*, 1996), and *TaqI* restriction endonuclease (Ozdinler, 1996). The cleavage of the MBP portion of the purified fusion protein resulted in the production of a protein with a molecular weight corresponding to that of *TaqI* restriction endonuclease. This protein did not display any enzymatic activity. These results supported by literature findings cited above indicated that MBP portion of the fusion protein is not the reason for absence of its activity. The presence of the fusion protein without endonuclease activity may be due to incorrect folding leading to the formation of insoluble inclusion bodies or due to the misincorporation of a nucleotide into the gene sequence resulting in a mutation in the active site of the protein.

Decreasing the concentration of the inducer or growth of the cells at lower temperatures to decrease the level of gene expression and to allow sufficient time for correct protein folding did not result in any improvement in the production of an active enzyme. The absence of an active protein in the cells grown in medium containing glycerol to provide osmotic stress and to reduce aggregate formation gave additional evidence that formation of inclusion bodies may not be the reason for the absence of enzymatic activity. Furthermore, expression of the gene in two other strains of *Escherichia coli*, namely XL1-Blue and ER2508 did not result in the production of an active fusion protein.

The sequencing analysis of the insert within the recombinant plasmid showed that the adenine in the position 410 was replaced by a thymine resulting in the production of a valine instead of aspartic acid at the residue 137 of the protein. It was previously reported that insertions at amino acid positions 47, 99, 101 and 108 resulted in loss of activity

(Mayer and Barany, 1995). Our results suggest that aspartic acid at codon 137 may also be important for the activity of the enzyme and may be a part of the active site. Further investigation is necessary to determine the contribution of codon 137 to the active site, for example by mutagenesis performed previously (Mayer *et al.*, 1985).

The use of *Pfu* DNA polymerase which has proofreading activity in the PCR amplification resulted in the identification of at least one colony that expresses the active MBP-*TaqI* restriction endonuclease fusion protein. The sequencing of the insert revealed no mutations in the target gene confirming the importance of the gene sequence in protein production. It has been reported that the error rate of *Taq* DNA polymerase is 2.0×10^{-5} errors/bp while the error rate of *Pfu* DNA polymerase is 1.6×10^{-6} (Lundberg *et al.*, 1991). The *Pfu* DNA polymerase exhibits 12 fold higher fidelity of amplification and therefore, it can be suggested that *Pfu* DNA polymerase has to be used for the amplification of target genes prior to cloning and the DNA sequence analysis needs to be incorporated into the screening step of cloning.

Purification of the fusion proteins from the total cytoplasmic extract and their analysis by SDS-PAGE revealed that degradation products slightly smaller than the fusion protein and a protein with the approximate size of MBP were also eluted with the fusion protein. The degradation products were probably cleaved within their *TaqI* restriction endonuclease portions since their MBP portion was still intact and active as confirmed by their affinity to amylose. The protein with the expected size of MBP may be endogenous MBP or a truncation product of the fusion protein. The elution of these proteins reduces efficiency of chromatographic purification. In order to prevent degradation of the fusion protein, shorter induction periods may be employed.

It has been reported that the recovery of active proteins differ in different strains. XL1-Blue strain of *Escherichia coli* has been shown to yield higher recovery of fusion protein production with pMAL-c2 expression vector (Hering *et al.*, 1996). In the XL1 strain, the structural plasmid stability is higher due to *recA* gene and the degradation of plasmid by nonspecific endonucleases is less due to *endA* gene. Moreover, the *lacI^q* gene expresses the lac repressor, preventing leaky expression of toxic recombinant proteins. The ER2508 strain does not make any MBP from its chromosome due to the *malB* deletion and

furthermore is lon protease deficient, preventing proteolysis of the protein. Investigation of expression of the recombinant plasmid in two different strains of *Escherichia coli*, XL1 and ER2508 showed that the total recovery of active endonuclease from TB1 cells was higher than the recovery from XL1 and ER2508 cells. The deficiency of the lon protease in ER2508 may have increased the production of toxic proteins resulting in cell death and hence a decrease in yield.

Heat treatment resulted in total loss of the activity. This may be due to the fact that MBP is not a thermostable protein and as a result the fusion protein is denatured as a whole. Milder heat treatment conditions may be necessary for partial purification of the fusion protein.

The expression of the *TaqI* restriction endonuclease was shown to inhibit cell growth. Therefore, cloning of the protective *TaqI* methyltransferase gene might reduce cell death and enhance *TaqI* restriction endonuclease production.

The biological activity of a restriction endonuclease is determined by its specificity to its cognate sequence and its homogeneity is determined by its freedom from contaminating non-specific endonucleases, nucleases and exonucleases. It has been shown in this study that the presence of the MBP fusion partner does not inhibit the biological activity of the *TaqI* restriction endonuclease. However, the equivalence of the affinity column purified MBP-*TaqI* restriction endonuclease fusion protein to the pure *TaqI* restriction endonuclease must be established before the fusion protein can be used as a substitute. Cut-ligate-recut test is necessary to show that the enzyme preparation is free of detectable exonucleases and phosphatases. Overdigestion test should be carried out to show that the purified fusion protein is free of detectable levels of DNases. To assay for nonspecific endonucleases, the fusion protein should be incubated with a supercoiled plasmid lacking the recognition sequence of *TaqI* restriction endonuclease.

The effect of temperature on the expression of enzyme can be investigated. Growth at high temperatures may provide *in vivo* heat purification of the thermostable *TaqI* restriction endonuclease. On the other hand, proteolytic degradation can be avoided by the use of low temperatures.

The induction period optimum for enzyme accumulation and the optimum concentration of IPTG for high level expression of MBP-*TaqI* restriction endonuclease fusion protein can be determined.

The effect of culture medium manipulations on fusion protein expression such as concentration of glucose, aeration, addition of glycerol can be investigated.

8. REFERENCES

- Abrahmsen, L., T. Moks, B. Nilsson and M. Uhlen, "Secretion of Heterologous Gene Products to the Culture Medium of *Escherichia coli*," *Nucleic Acids Research*, Vol. 14, pp. 7487-7500, 1986
- Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson, *The Cell*, Third Edition, Garland Publishing Inc., New York, 1994
- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (editors), *Short Protocols in Molecular Biology*, John Wiley and Sons Inc., USA, 1992
- Barany, F., "A Genetic System for Isolation and Characterization of *TaqI* Restriction Endonuclease Mutants," *Gene*, Vol. 56, pp. 13-27, 1987
- Barany, F., "The *TaqI* Star Reaction: Strand Preferences Reveal Hydrogen-Bond Donor and Acceptor Sites in Canonical Sequence Recognition," *Gene*, Vol. 65, pp. 149-165, 1988a
- Barany, F., "Overproduction, Purification and Crystallization of *TaqI* Restriction Endonuclease," *Gene*, Vol. 65, pp. 167-177, 1988b
- Barany, F., B. Slatko, M. Danzitz, D. Cowburn, I. Schildkraut, and G.G. Wilson, "The Corrected Nucleotide Sequences of The *TaqI* restriction and modification enzymes reveal a thirteen-codon overlap," *Gene*, Vol. 112, pp. 91-95, 1992
- Batchu, R. B., D. A. Miles, T. M. Rechlin, R. R. Drake, and P. L. Hermonat, "Cloning, Expression and Purification of Full Length Rep78 of Adeno-Associated Virus as a

- Fusion Protein with Maltose Binding Protein in *Escherichia coli*,” *Biochemical and Biophysical Research Communications*, Vol. 208, pp. 714-720, 1995
- Bedouelle, H., and P. Duplay, “Production in *Escherichia coli* and One-Step Purification of Bifunctional Hybrid Proteins Which Bind Maltose,” *European Journal of Biochemistry*, pp. 541-549, 1988
- Bentley, W. E., N. Mirjalili, D. C. Andersen, R. H. Davis, and D. S. Kompala, “Plasmid-encoded Protein: The Principal Factor in the “Metabolic Burden” Associated with Recombinant Bacteria,” *Biotechnology and Bioengineering*, Vol. 35, pp. 668-681, 1990
- Betton, J-M, J. Jacob, M. Hofnung, and J. Broome-Smith, “Creating a Bifunctional Protein by Insertion of β -lactamase into the Maltodextrin-Binding Protein,” *Nature Biotechnology*, Vol. 15, pp. 1276-1279, 1997
- Blackwell, J. R., and R. Horgan,”A Novel Strategy for Production of a Highly Expressed Recombinant Protein in Active Form,” *FEBS Letters*, Vol. 295, pp. 10-12
- Blondel, A., and H. Bedouelle, “Export and Purification of a Cytoplasmic Dimeric Protein by Fusion to the Maltose-Binding Protein of *Escherichia coli*,” *European Journal of Biochemistry*, Vol. 193, pp. 325-330, 1990
- Bodnar, J. W., W. Zempsky, D. Warder, C. Bergson, and D. C. Ward, “Effect of Nucleotide Analogs on the Cleavage of DNA by the Restriction Enzymes AluI, DdeI, HinfI, RsaI and TaqI,” *The Journal of Biological Chemistry*, Vol. 258, pp. 15206-15213, 1983
- Bollag, D. M., and M. D. Rozycki, S. J. Edelstein, *Protein Methods*, Second Edition, JohnWiley and Sons Inc., New York, 1996
- Bradford, M., “A Rapid and Sensitive Method for the Quantification of Microgram Quantities of Protein Utilizing the Principle of Protein Binding,” *Analytical Biochemistry*, Vol. 72, pp. 248-254, 1976

- Brock, T. D., "Life at High Temperatures," *Science*, Vol. 158, pp. 1012-1019, 1967
- Carter, P., "Site Specific Proteolysis of Fusion Proteins," in M. R. Ladisch, R. C. Willson, C.C. Painton, and S.E. Builder (Eds.), *Protein Purification: From Molecular Mechanisms to Large-Scale Processes*, pp. 181-193, American Chemical Society Symposium Series, 1990
- Chou, C., G. N. Bennett, and K. San, "Effect of Modulated Glucose Uptake on High-Level Recombinant Protein Production in a Dense *Escherichia coli* Culture," *Biotechnology Progress*, Vol. 10, pp. 644-647, 1994
- Cloney, L. P., D. R. Bekkaoui, and S. M. Hemmingsen, "Coexpression of Plastid Chaperonin Genes and a Synthetic Plant Rubisco Operon in *Escherichia coli*," *Plant Molecular Biology*, Vol. 23, pp. 1285-1290, 1993
- Cullen, P. J., W. C. Bowman, and R. G. Kranz, "*In vitro* Reconstitution and Characterization of the *Rhodobacter capsulatus* NtrB and NtrC Two-component System," *The Journal of Biological Chemistry*, Vol. 271, No. 11, pp. 6530-6536, 1996
- di Guan, C., P. Li, P. D. Riggs, and H. Inouye, "Vectors that Facilitate the Expression and Purification of Foreign Peptides in *Escherichia coli* by Fusion to Maltose-Binding Protein," *Gene*, Vol. 67, pp. 21-30, 1988
- Doskeland, A. P., A. Martinez, P. M. Knappskog, and T. Flatmark, "Phosphorylation of Recombinant Human Phenylalanine Hydroxylase: Effect on Catalytic Activity, Substrate Activation and Protection Against Non-Specific Cleavage of the Fusion Protein by Restriction Protease," *Biochemistry Journal*, Vol. 313, pp. 409-414, 1996
- Duplay, P., H. Bedouelle, A. Fowler, I. Zabin, W. Saurin, and M. Hofnung, "Sequences of the *malE* Gene and Its Product, the Maltose-Binding Protein of *Escherichia coli* K12," *The Journal of Biological Chemistry*, Vol. 259, No. 16, pp. 10606-10613, 1984

- Duplay, P., S. Szmecman, H. Bedouille, and M. Hofnung, "Silent and Functional Changes in the Periplasmic Maltose-binding Protein of *Escherichia coli* K-12," *Journal of Molecular Biology*, Vol. 194, pp. 663-673, 1987
- Enünlü, İ., "Cloning of *TaqI* Endonuclease Gene into *E. coli* and its Expression," MS. Thesis, Boğaziçi University, 1996
- Fidai, S., G. B. Kalmar, W. R. Richards, and T. J. Borgford, "Recombinant Expression of the *pufQ* Gene of *Rhodobacter capsulatus*," *Journal of Bacteriology*, Vol. 175, No. 15, pp. 4834-4842, 1993
- Ford, C. F., I. Suominen, and C. E. Glatz, "Fusion Tails for Recovery and Purification of Recombinant Proteins," *Protein Expression and Purification*, Vol. 2, pp. 95-107, 1991
- Georgiou, G., "Optimizing the Production of Recombinant Proteins in Microorganisms," *AIChE Journal*, Vol. 34, pp.1233-1248, 1988
- Georgiou, G., and P. Valax, "Expression of Correctly Folded Proteins in *Escherichia coli*," *Current Opinion in Biotechnology*, Vol. 7, pp. 190-197, 1996
- Germino, J., and D. Bastia, "Rapid Purification of a Cloned Gene Product by Genetic Fusion and Site Specific Proteolysis," *Proceedings of the National Academy of Sciences*, Vol. 81, pp. 4692-4696, 1984
- Germino, J., J. G. Gray, H. Charbonneau, T. Vanaman, and D. Bastia, "Use of Gene Fusions and Protein-Protein Interaction in the Isolation of a Biologically Active Regulatory Protein: The Replication Initiator Protein of Plasmid R6K," *Proceedings of the National Academy of Sciences USA*, Vol. 80, pp. 6848-6852, 1983
- Gold, L., "Expression of Heterologous Proteins in *Escherichia coli*," *Methods in Enzymology*, Vol. 185, pp. 11-14, 1990

- Gold, L., D. Pribnow, T. Schneider, S. Shinedling, B. S. Singer, and G. Stormo, "Translation Initiation in Prokaryotes," *Annual Review in Microbiology*, Vol. 35, pp. 365-369, 1981
- Goldstein, M. A., and R. H. Doi, "Prokaryotic Promoters in Biotechnology," *Biotechnology Annual Review*, Vol. 1, pp. 105-128, 1995
- Hannig, G., and S. C. Makrides, "Strategies for Optimizing Heterologous Protein Expression in *Escherichia coli*," *Trends in Biotechnology*, Vol. 16, pp. 54-60, 1998
- Harrison, J. S., E. K. Moore, P. Dunnill, M. J. Berry, A. Fellingner, and L. Frenken, "Factors Affecting the Fermentative Production of a Lysozyme-Binding Antibody Fragment in *Escherichia coli*," *Biotechnology and Bioengineering*, Vol. 53, No. 6, pp. 611-622, 1997
- Hering, T. M., J. Kollar, T. D. Huynh, and J. B. Varelas, "Purification and Characterization of Decorin Core Protein Expressed in *Escherichia coli* as a Maltose-Binding Protein Fusion," *Analytical Biochemistry*, Vol. 240, pp. 98-108, 1996
- Huang, C., S. Perretti, and J. Bryers, "Effects of Inducer Levels on a Recombinant Bacterial Biofilm Formation and Gene Expression," *Biotechnology Letters*, Vol. 16, pp. 903-908, 1994
- Ito, K., "Purification of the Precursor Form of Maltose-Binding Protein, a Periplasmic Protein of *Escherichia coli*," *The Journal of Biological Chemistry*, Vol. 257, No. 17, pp. 9895-9897, 1982
- Jayakumar, A., W. Huang, B. Raeyz, S. S. Chirala, and S. J. Wakil, "Cloning and Expression of the Multifunctional Human Fatty Acid Synthase and its Subdomains in *Escherichia coli*," *Proceedings of the National Academy of Sciences*, Vol. 93, pp. 14509-14514, 1996

- Kavanaugh, J. S., P. H. Rogers, and A. Arnone, "High Resolution X-Ray Study of Deoxy Recombinant Human Hemoglobins synthesized from β -globins Having Mutated Amino Termini," *Biochemistry*, Vol. 31, pp. 8640-8647, 1992
- Kurland, C., and J. Gallant, "Errors of Heterologous Protein Expression," *Current Opinion in Biotechnology*, Vol. 7, pp. 489-493, 1996
- Laemmli, U. K., "Cleavage of Structural Proteins During the Assembly of the Head of Bacteriophage T4," *Nature*, Vol. 227, pp. 680-685, 1970
- Lemaitre, G., K. Laaroubi, L. Soulet, D. Barritault, and M. Miskulin, "Production and Purification of Active FGF2 via Recombinant Fusion Protein," *Biochimie*, Vol. 77, pp. 162-166, 1995
- Li, X., and K. B. Taylor, "Effect of Glucose on the Expression Parameters of Recombinant Protein in *Escherichia coli* During Batch Growth in Complex Medium," *Biotechnology Progress*, Vol. 10, pp. 160-164, 1994
- Lundberg, K. S., D. D. Shoemaker, M. W. Adams, J. M. Short, J. A. Sorge and E. J. Mathur, "High-fidelity Amplification Using a Thermostable DNA Polymerase Isolated from *Pyrococcus furiosus*," *Gene*, Vol. 108, pp. 1-6, 1991
- Maina, C., P. Riggs, A. Grandea III, B. Slatko, L. Moran, J. Tagliamonte, L. McReynolds, and C. di Guan, "An *Escherichia coli* Vector to Express and Purify Foreign Proteins by Fusion to and Separation from Maltose-Binding Protein," *Gene*, Vol. 74, pp. 365-373, 1988
- Makrides, S.C., "Strategies for Achieving High-Level Expression of Genes in *Escherichia coli*," *Microbiological Reviews*, pp. 512-538, 1996
- Maniatis, T., E. Frisch, and J. Sambrook, *Molecular Cloning- A Laboratory Manual*, Cold Spring Harbor Laboratory Publications, New York, 1989

- Marino, M. H. "Expression Systems for Heterologous Protein Production" *BioPharm.* Vol. 2, pp. 18-33, 1989
- Matsumoto, T., S. Hata, I. Suzuka, and J. Hashimoto, "Expression of Functional Proliferating-cell Nuclear Antigen from Rice (*Oryza sativa*) in *Escherichia coli*," *European Journal of Biochemistry*, Vol. 223, pp. 179-187, 1994
- Mayer, A.M., and F. Barany, "Photoaffinity of Cross-Linking of *TaqI* Restriction Endonuclease Using Aryl Azide Linked to the Phosphate Backbone," *Gene*, Vol. 153, pp.1-8, 1995
- McPherson, D. T., "Codon Preference Reflects Mistranslational Constructs: A Proposal," *Nucleic Acids Research*, Vol. 16., pp. 4111-4120, 1988
- Miksch, G., R. Neitzel, E. Fiedler, K. Friehs, and E. Flaschel, "Extracellular Production of a Hybrid β -glucanase from *Bacillus* by *Escherichia coli* under Different Cultivation Conditions in Shaking Cultures and Bioreactors," *Applied Microbial Biotechnology*, Vol. 47, pp. 120-126, 1997
- Miller, K. W., R. Schamber, Y. Chen, and B. Ray, "Production of Active Chimeric Pediocin AcH in *Escherichia coli* in the Absence of Processing and Secretion Genes from the *Pediococcus pap* Operon," *Applied and Environmental Microbiology*, Vol. 64, No. 1, pp. 14-20, 1998
- Mukhija, R., P. Rupa, D. Pillai, and L. Garg, "High Level Production and One Step Purification of Biologically Active Human Growth Hormone in *Escherichia coli*," *Gene*, Vol. 165, pp. 303-306, 1995
- Narciandi, R. E., F. J. Morbe, and D. Riesenber, "Maximizing the Expression of Recombinant Kringle 1 (Streptokinase) Synthesized in *Escherichia coli*. Influence of Culture and Induction Conditions," *Biotechnology Letters*, Vol. 18, No.11, pp. 1261-1266, 1996

- Nishi, T., and S. Itoh, "Enhancement of Transcriptional Activity of the *Escherichia coli* trp Promoter by Upstream A + T- rich Regions," *Gene*, Vol. 44, pp. 29-36, 1986
- Nishihara, T., T. Iwabuchi, and T. Nohno, "A T7 Promoter Vector with a Transcriptional Terminator for Stringent Expression of Foreign Genes," *Gene*, Vol. 145, pp. 145-146, 1994
- Nossal, N. G., and L. A. Heppel, "The Release of Enzymes by Osmotic Shock from *Escherichia coli* in Exponential Phase," *Journal of Biological Chemistry*, Vol. 241, pp. 3055-3062, 1966
- Old, R. W., and S. B. Primrose, *Principles of Gene Manipulation*, Fourth Edition, Blackwell Scientific Publications, Oxford, 1989
- Özdinler, H., "Cloning of *TaqI* Endonuclease Gene into *E. coli* and Isolation of the Enzyme by Using pMAL-p2 Vector System," MS. Thesis, Boğaziçi University, 1996
- Pierce, J., and S. Gutteridge, "Large Scale Preparation of Ribulobisphosphate Carboxylase from a Recombinant System in *Escherichia coli* Characterized by Extreme Plasmid Instability," *Applied Environmental Microbiology*, Vol. 49, pp. 1094-1098, 1985
- Pilon, A. L., P. Yost, T. E. Chase, G. L. Lohnas, and W. Bentley, "High Level Expression and Efficient Recovery of Ubiquitin Fusion Proteins from *Escherichia coli*," *Biotechnology Progress*, Vol. 12, pp. 331-337, 1996
- Primrose, S. B., P. Derbyshire, I. M. Jones, M. E. Nugent, and W. C. A. Tacon, "Hereditary Instability of Recombinant DNA Molecules," in L. J. Nisbet and D. J. Winstanley (Eds.), *Bioactive Microbial Products 2: Development and Production*, pp. 63-77, Academic Press, London, 1983
- Proba, K. L. M. Ge, and A. Plückthun, " Functional Antibody Single Chain Fragments from the Cytoplasm of *Escherichia coli*: Influence of Thioredoxin Reductase (TrxB)," *Gene*, Vol. 15-, pp. 203-207, 1995

- Robinson, M., R. Lilley, S. Little, J. S. Emtage, G. Yarranton, P. Stephens, A. Millican, M. Eaton, and G. Humphreys, "Codon Usage Can Affect Efficiency of Translation of Genes in *Escherichia coli*," *Nucleic Acids Research*, Vol. 12, pp. 6663-6671, 1984
- Rodriguez, R., and R. C. Tait, *Recombinant DNA Techniques: An Introduction*, Addison-Wesley Publishing Company, London, 1983
- Rogers, S., R. Wells, and M. Rechsteiner, "Amino Acid Sequences Common to Rapidly Degraded Proteins: The PEST Hypothesis," *Science*, Vol. 234, pp. 364-368, 1986
- Ryan, W., P. Collier, L. Loredó, J. Pope, and R. Sachdev, "Growth Kinetics of *Escherichia coli* and Expression of a Recombinant Protein and Its Isoforms under Heat Shock Conditions," *Biotechnology Progress*, Vol. 12, pp. 596-601, 1996
- Sato, S., C. Hutchison III, and J. Harris, "A Thermostable Sequence-Specific Endonuclease from *Thermus aquaticus*," *Proceedings of the National Academy of Sciences*, Vol. 74, pp. 542-546, 1977
- Schein, C., "Production of Soluble Recombinant Proteins in Bacteria," *Bio/Technology*, Vol. 7, pp. 1141-1149, 1989
- Schneider, E. L., J. G. Thomas, J. A., Bassuk, E. H. Sage, and F. Baneyx, "Manipulating the aggregation and oxidation of human SPARC in the cytoplasm of *Escherichia coli*," *Nature Biotechnology*, Vol. 15, pp. 581-585, 1997
- Sharma, S. K., "On the Recovery of Genetically Engineered Proteins from *Escherichia coli*," *Separation Science Technology*, Vol. 21, pp. 701-705, 1986
- Shiloach, J., J. Kaufman, A. S. Guillard, and R. Fass, "Effect of Glucose Supply Strategy on Acetate Accumulation, Growth, and Recombinant Protein Production by *Escherichia coli* BL21 (λ DE3) and *Escherichia coli* JM109," *Biotechnology and Bioengineering*, Vol. 49, pp. 421-428, 1996

- Shuman, H. A., T. J. Silhavy, and J. R. Beckwith, "Labeling of Proteins with β -Galactosidase by Gene Fusion," *The Journal of Biological Chemistry*, Vol. 255, No. 1, pp. 168-174, 1980
- Simon, L. D., B. Randolph, N. Irwin and G. Binkowski, "Stabilization of proteins by a bacteriophage T4 gene cloned in *Escherichia coli*," *Proceedings of the National Academy of Sciences USA*, Vol. 80, pp. 2059-2062, 1983
- Slatko, B., J. Benner, T. Jager-Quinton, L. Moran, T. Simcox, E. Van Cott, and G. Wilson, *Nucleic Acids Research*, Vol. 15, pp. 9781-9796, 1987
- Spurlino, J.C., G. Lu, and F. A. Quiocho, "The 2.3-Resolution Structure of the Maltose- or Maltodextrin-binding Protein, a Primary Receptor of Bacterial Active Transport and Chemotaxis," *The Journal of Biological Chemistry*, Vol. 266, No. 8, pp. 5202-5219, 1991
- Talmadge, K., and W. Gilbert, "Cellular Location Affects Protein Stability in *Escherichia coli*," *Proceedings of the National Academy of Sciences*, Vol. 79, pp. 1830-1833, 1982
- Uhlen, M., and T. Moks, "Gene Fusion for Purpose of Expression," *Methods in Enzymology*, Vol. 185, pp. 129-143, 1990
- Vogelstein, B., and D. Gillespie, "Preparative and Analytical Purification of DNA from Agarose," *Proceedings of the National Academy of Sciences USA*, Vol. 76, pp. 615-619, 1979
- Ward, G. A., C. K. Stover, B. Moss, and T. R. Fuerst, "Stringent Chemical and Thermal Regulation of Recombinant Gene Expression by Vaccinia Virus Vectors in Mammalian Cells," *Proceedings of the National Academy of Sciences, USA*, Vol. 92, pp. 6773-6777, 1995
- Weickert, M., D. Doherty, E. Best, and P. Olins, "Optimization of Heterologous Protein Production in *Escherichia coli*," *Current Opinion in Biotechnology*, Vol. 7, pp. 494-499, 1996

- Weller, U., L. Müller, M. Messner, M. Palmer, A. Valeva, J. Trantum-Jensen, P. Agrawal, C. Biermann, A. Döbereiner, M. A. Kehoe, and S. Bhakdi, "Expression of Active Streptolysin O in *Escherichia coli* as a Maltose-Binding-Protein-Streptolysin-O Fusion Protein," *European Journal of Biochemistry*, Vol. 236, pp. 34-39, 1996
- Wickner, S., S. Gottesman, D. Skwyra, J. Hoskins, K. MacKenney, and M. Maurizi, "A Molecular Chaperone, ClpA, Functions Like DnaK and DnaJ," *Proceedings of the National Academy of Sciences USA*, Vol. 91, pp. 12218-12222, 1994
- Wilkinson, D. L., and R. G. Harrison, "Predicting the Solubility of Recombinant Proteins in *Escherichia coli*," *Bio/Technology*, Vol. 9, pp. 443-448, 1991
- Wonderling, R. S., S. R. M. Kyöstiö, and R. A. Owens, "A Maltose-Binding Protein/Adeno-Associated Virus Rep68 Fusion Protein has DNA-RNA Helicase and ATPase Activities," *Journal of Virology*, Vol. 69, No. 6, pp. 3542-3548, 1995
- Wood, T., and S. Perretti, "Effect of Chemically-Induced, Cloned Gene Expression on Protein Synthesis in *Escherichia coli*," *Biotechnology and Bioengineering*, Vol. 38, pp. 397-412, 1991
- Wu, X., L. Brandes, H. E. Maschke, and K. Schugerl, "Overproduction of EcoRI and SpA: EcoRI Fusion Protein by *Escherichia coli* Employing the Combinations of Three Different Types of Multicopy Plasmids: Comparison of Plasmid Stabilities and Product Concentrations," *Journal of Chemical Technology and Biotechnology*, Vol. 56, pp. 383-387, 1993
- Yarzabal, A., M. Bastidas, L. Avilan, J. Cruz, and J. Pig, "Induction Conditions for Maximizing Recombinant Staphylokinase Expression in *Escherichia coli*," *Biotechnology Letters*, Vol. 19, pp. 633-637, 1997
- Yuan, Z-H., U. Kumar, H. C. Thomas, Y-M. Wen, and J. Monjardino, "Expression, Purification, and Partial Characterization of HCV RNA Polymerase," *Biochemical and Biophysical Research Communications*, Vol. 232, pp. 231-235, 1997

Zebala, J., J. Choi, G. Trainor, and F. Barany, "DNA Recognition of Base Analogue and Chemically Modified Substrates by the *TaqI* Restriction Endonuclease," *The Journal of Biological Chemistry*, Vol. 267, pp. 8106-8116, 1992a

Zebala, J., J. Choi, and F. Barany, "Characterization of Steady State Single-Turnover, and Binding Kinetics of the *TaqI* Restriction Endonuclease," *The Journal of Biological Chemistry*, Vol. 267, pp. 8097-8105, 1992b