การระบุเพปไทด์ที่จำเพาะต่อการจับเดกซ์ทรานด้วยเทคนิคกลูแคนบายดิงแอสเสย์

นางสาว สุริษา สุวรรณรังษี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต

สาขาวิชาวิทยาศาสตร์ชีวภาพ

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2550

IDENTIFICATION OF PEPTIDES SPECIFIC FOR DEXTRAN BINDING BY GLUCAN BINDING ASSAY TECHNIQUE

Miss Surisa Suwannarangsee

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Biological Sciences
Faculty of Science
Chulalongkorn University
Academic year 2007

Thesis Title	IDENTIFICATION OF PEPTIDES SPECIFIC FOR DEXTRAN BINDING BY GLUCAN BINDING ASSAY TECHNIQUE
By	Miss Surisa Suwannarangsee
Field of Study	Biological Sciences
Thesis Advisor	Associate Professor Warawut Chulalaksananukul, Ph.D.
Thesis Co-advisor	Professor Magali Remaud-Simeon, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

..... Dean of the Faculty of Science
(Professor Supot Hannongbua,Ph.D.)

THESIS COMMITTEE

..... Chairman
(Associate Professor Preeda Boon-long, Ph.D.)

..... Thesis Advisor
(Associate Professor Warawut Chulalaksananukul, Ph.D.)

..... Thesis Co-advisor
(Professor Magali Remaud-Simeon, Ph.D.)

..... External Member
(Associate Professor Kosum Chansiri, Ph.D.)

..... Member
(Assistant Professor Alisa Vangnai, Ph.D.)

..... Member
(Teerada Wangsomboondee, Ph.D.)

สุริษา สุวรรณรังษี : การระบุเพปไทด์ที่จำเพาะต่อการจับเดกซ์ทรานด้วยเทคนิคกลูแคนบายดิงแอสเสย์. (IDENTIFICATION OF PEPTIDES SPECIFIC FOR DEXTRAN BINDING BY GLUCAN BINDING ASSAY TECHNIQUE) อ. ที่ปรึกษา: รศ.ดร. วรวุฒิ จุฬาลักษณานุกูล, อ. ที่ปรึกษาร่วม: Prof. Dr. Magali Remaud-Simeon, 180 หน้า.

ขั้นตอนหนึ่งที่สำคัญในการศึกษาลักษณะและหน้าที่ของโปรตีนคือ กระบวนการแยกโปรตีนให้บริสุทธิ์ โดยการนำโปรตีนติดตามมารวมกับโปรตีนที่สนใจสามารถแยกโปรตีนที่ต้องการออกจากโปรตีนและสิ่งปนเปื้อนอื่นๆ ในสารละลายโดยผ่านกระบวนการอัฟฟินิตีโครมาโทกราฟี เมื่อพิจารณาโดเมนด้านปลาย C ของเอนไซม์ dextransucrase จากเชื้อ *Leuconostoc mesenteroides* NRRL B-512F (DSR-S) มักถูกเรียกว่า glucan-binding domain (GBD) อย่างไรก็ตาม ในปัจจุบันยังไม่สามารถระบุลำดับกรดอะมิโนที่มีบทบาทในการจับกลูแคน วัตถุประสงค์ของงานวิจัยนี้จึงมุ่งศึกษาหน้าที่ของโดเมนด้านปลาย C ของเอนไซม์ DSR-S และนำชิ้นส่วนเพปไทด์เล็กที่สุดที่สามารถจับกลูแคนมาใช้เป็นโปรตีนติดตามสำหรับการทำโปรตีนให้บริสุทธิ์โดยวิธีอัฟฟินิตีโครมาโทกราฟีและใช้ตัวตรึงที่มีเดกซ์ทรานเป็นส่วนประกอบ การศึกษาครั้งนี้ ได้ทำการออกแบบและทดสอบความสามารถในการจับกลูแคนของชิ้นส่วนเพปไทด์ขนาดต่างๆ กันจากโดเมนด้านปลาย C รวม 14 ชิ้น พบว่า เพปไทด์ขนาดเล็กที่สุดคือ GBD-7 ซึ่งมีขนาด 14 กิโลดาลตันสามารถจับกลูแคนได้ โดยอยู่บริเวณด้านปลาย C ของGBD และประกอบด้วย YG repeat 6 หน่วย เพปไทด์นี้มีค่าคงที่ในการแยกกันของโปรตีน (Kd) เท่ากับ 2.8x10 M และสามารถจับกับ isomaltohexaose และเดกซ์ทรานสายยาวๆ ได้ แสดงให้เห็นว่าเพปไทด์นี้มีความสามารถในการจับเดกซ์ทรานสูง

การทดสอบการนำ GBD-7 ไปใช้เป็นโปรตีนติดตามสำหรับการทำโปรตีนให้บริสุทธิ์ร่วมกับตัวตรึงที่มีราคาถูก เช่น Sephacryl® S300HR พบว่า elution yield เท่ากับ 58% โปรตีนที่ได้มีความบริสุทธิ์มากกว่า 95% หลังจากทำการ elute โดยเดกซ์ทรานขนาด 1.5 กิโลดาลตัน นอกจากนี้ rare codon ที่มีอยู่ในลำดับเบสของ GBD-7 ถูกแทนที่โดย codon ที่ถูกใช้บ่อยใน *E. coli* โดยวิธี site-directed mutagenesis วิธีดังกล่าวนี้สามารถเพิ่มการแสดงออกของโปรตีนถึง 4.8 เท่าจากปกติ ซีรีส์ของชิ้นส่วนสั้นๆ ของ GBD-7 ถูกสร้างขึ้นโดยเทคนิค random deletion เพื่อเพิ่ม solubility และลดขนาดของโปรตีน GBD-7 พร้อมกัน ผลการทดลองแสดงให้เห็นว่าสามารถลดขนาดของโปรตีน GBD-7 ได้อีก 12 กรดอะมิโนที่ปลายด้าน N โดยยังคงความสามารถในการจับเดกซ์ทราน รวมถึงการตัดปลาย C บริเวณ V5 epitope ของโปรตีน *Thioredoxin*-GBD-7-*His* และ GBD-7-*His* สามารถเพิ่มการแสดงออกของโปรตีน GBD-7 ได้

สาขาวิชา วิทยาศาสตร์ชีวภาพ ลายมือชื่อนิสิต.....
ปีการศึกษา 2007 ลายมือชื่ออาจารย์ที่ปรึกษา.....
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

4573870223 : MAJOR BIOLOGICAL SCIENCES
KEY WORD: GLUCAN-BINDING DOMAIN / DEXTRANSUCRASE / AFFINITY TAG / PROTEIN PURIFICATION / DIRECTED EVOLUTION

SURISA SUWANNARANGSEE: IDENTIFICATION OF PEPTIDES SPECIFIC FOR DEXTRAN BINDING BY GLUCAN BINDING ASSAY TECHNIQUE. THESIS ADVISOR: ASSOC. PROF. WARAWUT CHULALAKSANANUKUL, Ph.D., THESIS CO-ADVISOR: PROF. MAGALI REMAUD-SIMEON, Ph.D., 180 pp.

One of the key steps in protein characterization and functional analysis is protein purification. Fusion of an affinity tag to protein of interest is enabling the protein separation from other contaminants in the extract through an affinity chromatography step. Considering a C-terminal domain of dextransucrase from *Leuconostoc mesenteroides* NRRL B-512F (DSR-S), it often refers as a glucan-binding domain (GBD), however, amino acid sequences responsible for the glucan binding have not been totally elucidated yet. The objectives of our study are to investigate the precise role of GBD in glucan binding and to generate shortest fragment that could be employed as a tag for affinity purification onto available cross-linked dextran supports. Fourteen truncated forms of DSR-S C-terminal domain were tested for their ability to bind dextrans. With a 14 kDa molecular weight, the shortest fragment (GBD-7) exhibiting a strong interaction with dextran was found to be localized at C-terminal end of the GBD and consists of six YG repeats. With a dissociation constant K_d of 2.8 10^{-9} M, this motif shows a very high affinity for isomaltohexaose and longer dextrans, confirming a very strong dextran binding property.

Potential application of GBD-7 to be used as an affinity tag for rapid purification onto cheap resins like Sephacryl® S300HR was demonstrated with the elution yield of 58% and purity higher than 95% after competitive elution using 1.5 kDa dextran. Moreover, rare codons representing in the GBD-7 sequences were replaced by frequently-used codons in *E. coli* by site-directed mutagenesis. This approach could enhance the fusion protein expression up to 4.8 folds of a wild-type. In order to increase its solubility and simultaneously reduce its size, nested truncated forms of GBD-7 proteins were generated. From the GBD-7 sequences, it can be deleted only 12 amino acids at N-terminus with retained binding activity. The truncations at C-terminal V5 epitope part of the GBD-7 fusion protein could enhance its expression in with or without *Thioredoxin* tag.

Field of study Biological Sciences Student's signature Surisa Suwannarangsee
Academic year 2007 Advisor's signature Warawut Chulalaksananukul
Co-advisor's signature

ACKNOWLEDGEMENTS

I wish to express my most sincere appreciation to my advisor, Associate Professor Dr. Warawut Chulalaksananukul, and my co-advisor, Professor Dr. Magali Remaud-Simeon for their kind supervision, instruction, and excellent encouragement during this study.

Also I am very grateful to my thesis committee, Associate Professor Dr. Preeda Boon-long, Assistant Professor Dr. Alisa Vangnai, Dr. Teerada Wangsomboondee, and Associate Professor Dr. Kosum Chansiri for giving thoughtful suggestions.

I would like to express my gratitude to Dr. Claire Moulis, Dr. Gabrielle Potocki-Veronese, Professor Dr. Pierre Monsan, and all members of Equipe Ingénierie Enzymatique Moléculaire, Institut National des Sciences Appliquées de Toulouse, France for cordial assistances, laboratory suggestions, and close cooperations.

I am greatly appreciative to all members in the Biofuel Production by Biocatalyst Research Unit, Department of Botany, Faculty of Science, Chulalongkorn University for their helpfulness and kindly supports.

I am most delighted to thank the Royal Golden Jubilee Ph.D. program and The French Embassy in Thailand to support me for all thesis periods and for giving me the chance to obtain wonderful experiences in France.

Thank you to Department of Botany, Faculty of Science, Chulalongkorn University and Laboratoire Biotechnologie Bioprocédés, Institut National des Sciences Appliquées de Toulouse, France for all facilities performed in this study.

Eventually, I wish to convey my deep appreciation to my family and my friends for their love, supports, and encouragements.

CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

LIST OF ABBREVIATIONS

Abs	absorbance
bp	base pair
BSA	bovine serum albumin
BCIP	5-bromo-4-chloro-3-indolyl phosphate
°C	degree Celsius
CCD	charge-coupled device
DMSO	Dimethyl sulfoxide
DNA	deoxyribonucleic acid
DSR-S	dextransucrase from *Leuconostoc mesenteroides* NRRL B512F
DTT	dithiothreitol
EDTA	ethylenediamine tetraacetic acid
et al.	Et. Alii (latin), and others
g	gram
GBD	glucan-binding domain
GTF	glucosyltransferase
h	hour
kb	kilobase
kd	dissociation constant
kDa	kilodalton
l	litre
µg	microgram
µl	microlitre
M	molar
ml	milliliter
mM	millimolar
MW	molecular weight
NBT	Nitroblue Tetrazolium
nm	nanometer

NMR	nuclear magnetic resonance
NTA	nitrilo triacetic acid
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffer saline
PCR	polymerase chain reaction
PMSF	phenyl methylsulfonyl fluoride
rpm	revolution per minute
SDS	sodium dodecyl sulfate
TEMED	N,N,N',N'-Tetramethylethylenediamine
v	volt

CHAPTER I
INTRODUCTION

In proteomic researches, diverse kinds of proteins are investigated for their function and structure. A process to separate an individual protein from several thousands of other proteins in the extract is a major important step to achieve that goal. However, proteins are different in chemical and structural properties and not easily purified in the same way. Fusion of affinity tag with the various recombinant proteins can be purified on a specific ligand under a standard condition. This process is usually called affinity tag purification and appropriate for purification of numerous protein targets. Nowadays, many affinity tag systems are commercially available e.g. His (Janknecht *et al.*, 1991), glutathione S-transferase (GST) (Smith and Johnson, 1988), maltose-binding protein (MBP) (Kellerman and Ferenci, 1982). Most of them were formerly genetically engineered into a vector which is ready to generate a fusion protein. Aside from protein purification, the tag is also expected to facilitate detection, soluble expression, and not to interfere with the biological activity of the protein partner. Until now, no affinity tag matches for all purposes even the most widely used, His tag. A significant drawback of the His tag is that proteins purified by these systems are usually contaminated. This may not be sufficiently pure for some purposes (Keefe *et al.*, 2001). Therefore, it is great interest to increase a variety of affinity tag systems.

Glucansucrases are enzymes mainly produced by lactic acid bacteria such as *Leuconostoc mesenteroides* (soil bacteria), *Streptococcus* species (oral flora), or Lactobacilli. They catalyse the synthesis of glucan polymers and the concomitant release of fructose from sucrose (Monchois *et al.*, 1999). Various types of glucosidic linkages are formed by glucansucrases (α-1,2; α-1,3; α-1,4 or α-1,6 linkages) depending on the enzyme specificity (Monsan *et al.*, 2001). In the presence of acceptors, glucansucrase may transfer the glucosyl residue from sucrose to the acceptor to the detriment polymer synthesis. This generally results in the formation of oligosaccharides when the acceptor is well recognized. These enzymes are thus very useful for the production of glucan polymers or oligosaccharides. For example,

dextran from *L. mesenteroides* B-512F was one of the first biopolymers to be produced on an industrial scale in 1948. Dextran can be used in several applications in the medical field, as blood plasma substitute, and in the field of separation media (Sephadex®) (Monsan *et al.*, 2001). Oligosaccharides with α-1,2/α-1,6 linkages produced by *L. mesenteroides* NRRL B-1299 are produced at the industrial scale to be employed as prebiotic compounds (Remaud-Simeon *et al.*, 2000).

Seventy glucansucrase encoding genes (named *gtf* for *Streptococcus* sp. enzymes and *dsr* for *L. mesenteroides* enzymes) have been cloned and sequenced by recombinant DNA techniques (http://www.cazy.org). The calculated molecular weights of these enzymes are range from 160 to 300 kDa (Shah *et al.*, 2004). The availability of glucansucrase primary sequences allowed sequence comparison of these enzymes which were classified in the family 70 of glycoside hydrolases from sequence similarities (Henrissat and Romeu, 1996). All of them possessed the same organization which comprises 4 regions (Figure 1). The first one corresponded to the N-terminal signal peptide allowing protein secretion. The second one was the variable region which was not very much conserved among glucansucrases. The other domain was the highly conserved N-terminal catalytic domain, which contains the catalytic amino acid residues. The last interesting domain covering from 300 to 400 amino acids was a domain often referred as C-terminal glucan binding domain (Monchois *et al.*, 1998; Monsan *et al.*, 2000 cited in Russell 1990; Wilke-Douglas *et al.*, 1989)

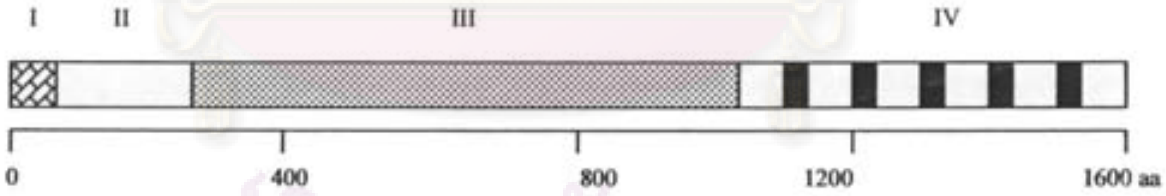


Figure 1. Schematic primary structure of glucansucrases. I, signal peptide; II, variable region; III, N-terminal catalytic domain; IV, C-terminal glucan binding domain.

This domain is generally composed of a series of YG repeats. The consensus sequence of the YG repeat is NDGYYFxxxGxxH°x(G/N)xH°H°H° (x, non-conserved

amino acid residue; H^{ϕ}, hydrophobic amino acid residue) (Giffard and Jacques, 1994). The number and organization of these repeats are specific to each source of enzymes (Monchois *et al.*, 1999). However, the function of C-terminal domain has not been totally elucidated yet. Identification of a minimal motif responsible for glucan binding should be performed. Moreover, the dextran-binding domain (DBD) of GTF from *S. sobrinus* has been successfully applied as an affinity tag for recombinant protein purification (Kaseda *et al.*, 2001). Nevertheless, they used the full-length DBD (30 kDa) that might harm the folding of fusion partner and cause high metabolic burden of host cell. Elution with high concentration 50 mg/ml of 18.1 kDa dextran might not desired for some applications like protein structural determination. Therefore, using the GBD as the affinity tag should be improved for developing an efficient and cost effective purification system.

Objectives

The objectives of this research were:

1. To screen the minimal GBD sequences of a dextransucrase from *L. mesenteroides* NRRL B-512F which interact specifically to dextran
2. To investigate the potential application of small GBD fragments as an affinity tag onto cross-linked dextran supports
3. To reduce size and improve soluble expression of small GBD fragment by directed evolution

Expected outcome

The amino acid data of GBD responsible for binding from this study will give more information about structure/function relationship of GBD for better understanding the enzyme mechanism. Moreover, the small fragment of GBD could be employed as a chromatographic affinity tag for recombinant protein purification process.

Methodology

1. Screening of GBD sequence specific to dextran binding
    1.1 Generation of the truncated forms of GBD encoding gene

1.2 Cloning and expression of these proteins in *E. coli* TOP10 cells

1.3 Estimation affinity of the proteins to dextran by glucan binding assay

1.4 Confirmation of the glucan binding property by affinity electrophoresis against an insoluble dextran

1.5 Determination of dissociation constant to dextran among the short fragments

2. Evaluation of the shortest GBD motif for affinity purification onto cross-linked dextran supports

2.1 Test the purification onto different cross-linked dextran supports

2.2 Optimization of the purification process in term of binding time, dextran size and concentration.

2.3 Comparison of the purification by GBD to His tag

3. Generation of nested deletion libraries of the shortest GBD to reduce its size and improve the solubility

3.1 Optimization of phenotypic high-throughput screening for GBD

3.2 Creating of the libraries and screening for size and soluble expression

3.3 Comparison of the GBD variant expression to that of the wild-type

Contents of the thesis

The organization of this manuscript is started with an introduction of the research including rational and problem, objectives, methodology. Then, the chapter II presents the literature reviews about current situation of affinity chromatography, the role of C-terminal domain of glucansucrases and integration of directed evolution technique for protein improvement. Materials and methods used in this study are also mentioned in the following chapter. The chapter IV, results and discussion, is divided into 3 parts corresponding to: (i) rational engineering of the DSR-S glucan-binding domain, (ii) recombinant protein purification by GBD-dextran support affinity chromatography, (iii) evolving GBD-7 towards a smaller size and a better expression. After that, all results are summarized in the chapter V and future perspectives are also described. Appendices indicated other information that did not included in the previous chapters.

CHAPTER II
LITERATURE REVIEW

1. Affinity chromatography

Protein purification is a powerful and versatile tool for separating an individual protein apart from crude extract. Nowadays, the protein purification has become one of the most commonly technique for studying biochemical properties, protein-protein interaction, and structural genomics. A number of chromatographic methods were developed for separating protein of interest from the mixture by charge, molecular weight, isoelectric point, hydrophobicity, and biological affinity (Janson and Rydén, 1998). However, size and biochemical characteristics of proteins are distinguished depend on their amino acid codes. So, the purification strategy of each protein must be different. The only chromatographic purification, which has general applicability, is affinity chromatography base on affinity-mediated isolation (Negrouk *et al.*, 2004). In post-genomic era, important of affinity chromatography is increased owing to high-throughput expression and large-scale purification for structural determination is required (Routzahn and Waugh, 2002). Moreover, fusion with a tag in the affinity chromatography has also been found to improve the yield and solubility of target proteins (Terpe, 2003; Hammarstrom *et al.*, 2002; Braun *et al.*, 2003; Waugh, 2005).

1.1 Principle of affinity chromatography

Affinity chromatography exploits the biological affinity for adsorption to a solid phase (Janson and Rydén, 1998). In this approach, genetic manipulation is required to produce a fusion of target protein and affinity tag. This affinity tag usually defines as exogenous amino acid sequences with high affinity for a specific biological or chemical ligand (Arnau *et al.*, 2006). Most of the times, protease cleavage site is included between the gene and the affinity tag sequences for further eliminating the tag. The presence of the affinity tag may affect important characteristics or functions of the target protein (Terpe, 2003). After the recombinant protein was constructed to an expression vector, the fusion protein is expressed in specific host cells. Protein extract derived from this construct is applied to a solid phase or support which is specifically bound to the carrier protein. Other proteins in crude extract can be

removed during washing steps. Protease is then added in order to separate the target protein from the fusion tag (Figure 2). In some case, the fusion protein is eluted altogether by adding competitors which can contend the binding of affinity tag to support.

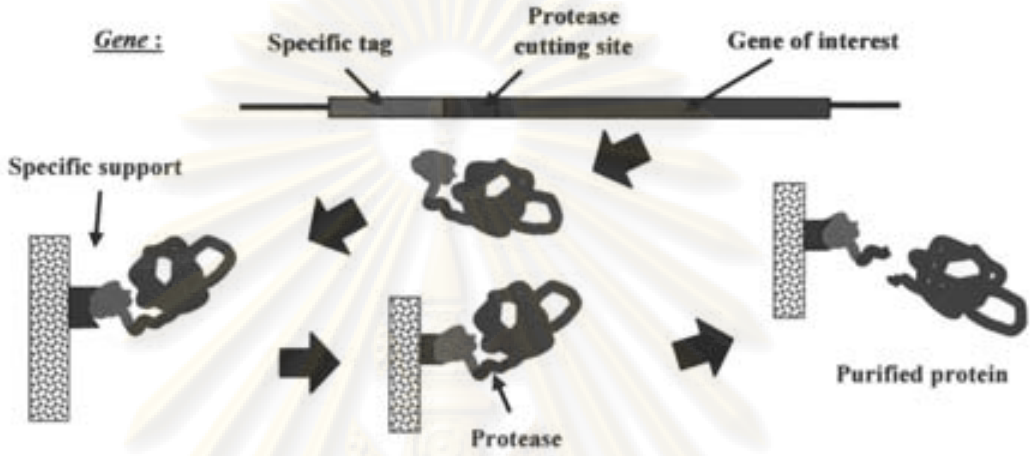


Figure 2. Affinity chromatography.

In general, a good affinity tag should possess these following characteristics (Lichty *et al*., 2005; Waugh, 2005; Terpe, 2003).

(a) Offer a high yield and purity

(b) Has no effect on biological activity and folding of partner protein

(c) Can be placed at any position (N- and C-terminus)

(d) Can be used to detect the fusion protein

(e) Reversibly bound to an inexpensive matrix

(f) Facilitate the recombinant protein expression in any heterologous host

(g) Can be eluted in mild condition without irreversibly affecting the protein to be isolated

(h) Enhance solubility and proper folding

(i) Can be easily removed to produce the native protein

However, no single affinity tag is optimal with respect to all of these parameters. Deleterious effects of fusion tagging have been investigated which is resulting in undesirable characteristics (Arnau *et al.*, 2006). For example, protein conformation changing (Chant *et al.*, 2005), low protein yield (Goel *et al.*, 2000), inhibition of enzyme activity (Kim *et al.*, 2001; Cadel *et al.*, 2004), alteration of enzyme activity (Fonda *et al.*, 2002), undesired flexibility in structural studies (Smyth *et al.*, 2003), and toxicity (de Vries *et al.*, 2003). Selection of the suitable system is relied on the target protein itself (e.g. stability, hydrophobicity), the expression system, and the application of the purified protein (Terpe, 2003).

1.2 Current available affinity tag purification system

Most of affinity tag systems have been continually developed during this past 20 years. They differ in matrix used, elution method, condition for purification, and nature of the tag itself. The last characteristic was used by Lichty *et al.* (2005) to categorize the tags into the following 3 classes:

Class I: Epitope affinity tags that use peptide of protein fusion to bind to small molecule ligands linked to a solid support e.g. hexahistidine (His) tag bind to immobilized metal ion (Hochuli *et al.*, 1987), glutathione S-transferase (GST) tag bind to glutathione attached to resin (Smith and Johnson, 1988).

Class II: Peptide tags bind to a protein-binding partner immobilized onto chromatography resin e.g. CBP tag (calmodulin-binding peptide) which is bind specifically to calmodulin resin (Stofko-Hahn *et al.*, 1992).

Class III: Epitope affinity tags which a binding support is an antibody recognizing a specific peptide epitope. This class may be considered as a subset of the second class. For example, FLAG peptide which can be used with one of several anti-FLAG antibody resins (Brizzard *et al.*, 1994).

Class I

1.2.1 Polyhistidine tag (His tag)

The polyhistidine tag is one of the most widely used tags which typically consists of five or six consecutive His residues (Hochuli *et al.*, 1988). The fusion could be performed in both N- and C-terminus of target proteins (Hochuli *et al.*, 1988). Optimal placement of the tag is protein specific (Terpe, 2003). Purification using the His tag is based on an immobilized metal ion affinity chromatography (IMAC) which was presented by Porath *et al.* (1975). The IMAC is based on the interaction between a transition metal ion (Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}) immobilized on a matrix and specific amino acid side chains. Histidine is the amino acid that exhibits the strongest interaction with immobilized metal ion matrices, as electron donor groups on the histidine imidazole ring readily form coordination bonds with the immobilized transition metal (Terpe, 2003). In 1987, Hochuli *et al.* firstly reported a nitrilotriacetic acid (NTA) as adsorbent which was able to strongly chelate Cu^{2+} and Ni^{2+}. One year later, they were successfully purified dihydrofolate reductase fused with the His tag onto Ni^{2+}-NTA resin (Hochuli *et al.*, 1988). The NTA molecule occupied four of six ligand binding site of the nickel ion (Figure 3), leaving two site free to bind with the His tag (Figure 4). Nowadays, Ni^{2+}-NTA matrices are commercially available in many scientific companies, such as Qiagen, Invitrogen, Sigma etc., with reasonable prices.

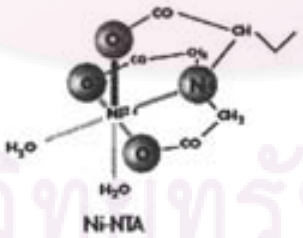


Figure 3. Interaction of NTA molecule to Ni ion (Qiagen).

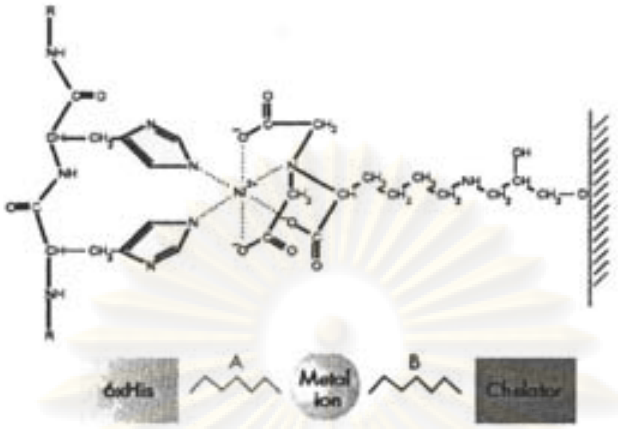


Figure 4. Capture of His tagged proteins by metal-chelated affinity matrix (Qiagen).

Purification using His tag is started by adsorption of His-tagged proteins to an adsorbent or resin at neutral or slightly alkaline pH in either native or denaturing condition (Nilsson *et al.*, 1997). Washing step is then performed using buffer containing trace amount of imidazole (10-20 mM) in order to prevent non-specific binding of host proteins. The imidazole ring is the part of the structure of histidine and can also bind to nickel ion affecting the binding of histidine residues in untagged proteins (Figure 5). The His-tagged proteins are finally recovered using high concentration of imidazole (100-250 mM) or low pH condition with purity of more than 90% in a single step using Ni^{2+}-NTA as a support (Hochuli *et al.*, 1988, Qiagen's manual).

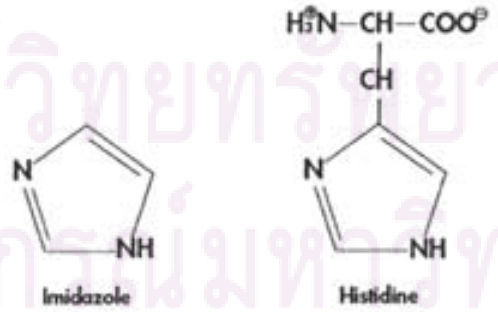


Figure 5. Chemical structures of imidazole and histidine.

Currently, another matrix for purifying His tagged proteins was developed by Clontech Company which is TALON™ metal affinity resin. The TALON resin consists of a Co^{2+}-carboxylmethylaspartate (Co^{2+}-CMA). The advantages of the TALON resin over the Ni-NTA are less contaminating proteins obtained and no metal leakage (Chaga *et al.*, 1999). However, the presence of imidazole can influence NMR experiments, competition studies, and crysllographic trials, and often result in protein aggregates (Terpe, 2003; Hefti *et al.*, 2001) Moreover, unless specific protocols are optimized for each protein, the His tag can not completely remove contaminants. The His tag is also not appropriated if the target protein activity is compromised by immobilized ion or chelating groups. Samples containing EDTA, EGTA, or DTT cannot be purified using Ni^{2+}-NTA without additional sample treatment (Keefe *et al.*, 2001). Purification under anaerobic condition is also not recommended because Ni^{2+}-NTA is reduced (Terpe, 2003).

1.2.2 Glutathione S-transferase (GST)

In 1988, recombinant protein purification using a 26 kDa GST enzyme of *Schistosoma japonicum* or GST tag was firstly demonstrated in *E. coli* (Smith and Johnson, 1988). Fusion protein could be efficiently purified from crude cell lysate by glutathione-affinity chromatography. Bound fusion proteins can be recovered by competition with 10 mM reduced glutathione under native condition. The vectors have been introduced site-specific protease sequences such as thrombin or blood coagulation factor X, so that the GST tag can be cleavage from the fusion proteins. The tag and non-cleaved fusion protein can be removed by absorption on glutathione-agarose resin. PreScission protease contained the human rhinovirus 3C protease including the GST tag was successfully engineered providing removal of this protease after proteolysis by affinity chromatography on gluthathion-agarose (Terpe, 2003). In 1994, it was able to produce both N- and C-terminal fusion to the GST tag in T7 expression vector engineered by Sharrocks (1994).

Until now, several evidences showed that the tag could be located at both N- and C- terminus and was successfully applied in *E. coli* (Smith and Johnson, 1988), yeast (Lu *et al.*, 1997), mammalian cells (Rudert *et al.*, 1996), and

baculovirus infected insect cells (Beekman *et al.*, 1994). Detection of the GST tag was available both enzyme assay and immunoassay (e.g. Novagen, Amersham biosciences). It was also found that the tag can prevent intracellular protease cleavage and stabilize the fusion protein (Terpe, 2003).

By contrast, the GST can form a homodimer which can complicate purification of fusion proteins and renders this affinity tag unsuitable for isolation of oligomeric proteins. Furthermore, four cysteine residues in each subunit of the GST dimer can lead to a significant degree of oxidative aggregation (Kaplan, *et al.*, 1997; Waugh, 2005). The GST tag is considered as large protein module and can result in interference to structural characterization of a fusion protein. Also, proteins purified solely on the basis of this tag are not sufficiently pure for some purposes (Keefe *et al.*, 2001).

1.2.3 Cellulose-binding domain (CBD)

On the basis of amino acid sequence similarities, more than 180 different CBDs have been identified and classified into 13 families (Tomme *et al.*, 1995). Most of the reported CBDs belong to families I, II, and III. In family I, CBDs are compact polypeptides of 32–36 residues, which are found only in fungi. The CBDs of families II and III are much larger and contain 90–100 and 130–172 residues, respectively (Carrard *et al.*, 2000). CBDs can vary in size from 4 to 20 kDa and can locate at different positions, N-terminal, C-terminal, and internal, within polypeptides (Terpe, 2003). Among them, some CBDs, family I, bind reversibly to cellulose and can be used as an affinity tag for protein purification. It was shown that the CBDs function normally when fused to heterologous polypeptides and can be used for purification and immobilization purposes (Greenwood *et al.*, 1989; Ong *et al.*, 1989).

Table 1. CBD tag sequences used in Novegen's vector (Novagen).

CBD tag	CBD Family	Molecular weight	Protein source
CBDclos	IIIa	17 kDa	Cellulose-binding Protein A; *Clostridium cellulovorans*
CBDcenA	IIa	11.7 kDa	Endo-1,4-b-glucanase A; *Cellulomonas fimi*
CBDcex	IIa	10.8 kDa	Exo-1,4-b-glucanase A; *Cellulomonas fimi*

CBDs can bind to cellulose at a pH range from 3.5 to 9.5 and can be placed at the N- or C- terminus of the target protein. The affinity of the tag is strong, thus an immobilized fusion protein can only be released with buffers containing urea or guanidine hydrochloride. Refolding step of proteins is required after this denaturing elution (Terpe, 2003). The advantage of cellulose is that it is inert, has low non-specific affinity, is available in many different forms, and has been approved for many pharmaceutical and human uses (Terpe, 2003). Currently, CBDs from *Clostridium cellulovorans* (CBDclos) and *Cellulomonas fimi* (CBDcenA and CBDcex) are commercially available by Novagen as shown in Table 1. These CBD domains are engineered into pET expression vectors and can efficiently bind both amorphous and crystalline celluloses (Novagen).

1.2.4 Maltose-binding protein (MBP)

MBP (40 kDa) is a binding protein specific for maltose and maltodextrins produced by the *mal*E gene of *E. coli* K12 (Duplay *et al.*, 1988). MBP can be employed as an affinity tag for recombinant protein purification on matrix containing cross-linked amylose. Competitive elution of bound proteins can be achieved using buffer containing 10 mM maltose (Kellerman and Ferenci, 1982; Terpe, 2003). In 1988, specific vectors for the MBP tag were firstly constructed allowing fusion proteins of foreign peptides and MBP (Guan *et al.*, 1988). The fusion protein can be directed to the periplasm by including the leader sequence from the *pho*A gene on the vector. The recognition sequence of blood coagulation factor Xa protease was inserted between the MBP and fusion partner (Maina *et al.*, 1988). After cleavage by factor Xa, the target protein can be separated from the MBP tag by

repeating the affinity chromatography step. Nowadays, MBP tag is usually included in commercial vectors such as pMAL with a protease cleavage site (Figure 6).

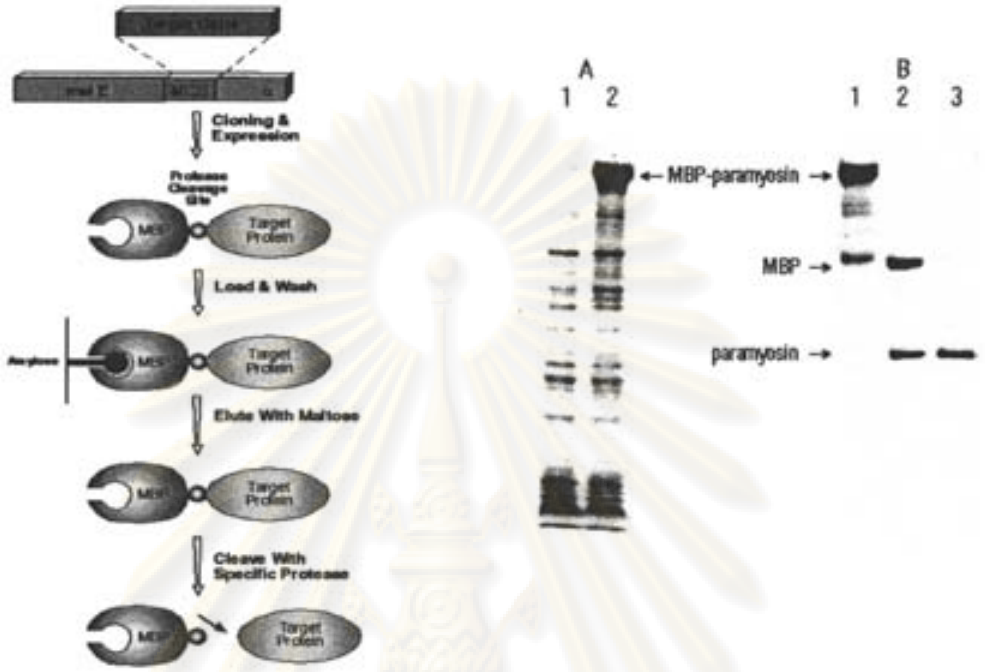


Figure 6. Left: Schematic of pMAL system, Right: SDS-PAGE of fractions from the purification of MBP-paramyosin-ΔSal. A: Lane 1: uninduced cells. Lane 2: induced cells. B: Lane 1: purified protein eluted from amylose column with maltose. Lane 2: purified protein after Factor Xa cleavage. Lane 3: paramyosin fragment eluted from second amylose column (New England Biolabs).

Fusion to MBP can be performed at N- or C-terminus when the fusion proteins were expressed in bacterial cells (Sachdev and Chirgwin, 2000). The MBP tag can act as a solubility enhancer to the fusion partner (Kapust and Waugh, 1999; Waugh, 2005). The tag is also suitable for large-scale protein purification. However, some proteins do not bind efficiently in the presence of 0.2% Triton X-100 or 0.25% Tween 20. Denaturing agent cannot be used (Terpe, 2003). Moreover, the size of MBP is relatively large leading high metabolic burden of host cells (Waugh, 2005).

1.2.5 Chitin-binding domain (ChBD)

C-terminal chitin-binding domain (ChBD) of chitinase A1 from *Bacillus circulans* WL-12 is a small peptide consisting of 45 amino acids (Chern and Chao, 2005). The ChBD was reversibly bound to insoluble chitin via hydrophobic interaction and can be function over a wide range of pH (Watanabe *et al.*, 1994; Hashimoto *et al.*, 2000). Application of the ChBD to be an affinity tag is attractive owing to its small size and specific binding to naturally abundant chitin as support (Wen *et al.*, 2002; Chern and Chao, 2005).

ChBD tag has been commonly available in combination with inteins (INTerventing protEINS) to serve as a self-cleaving affinity tag (Terpe, 2003; Sharma *et al.*, 2006). The tag can facilitate highly selective capture of the fusion protein on an inexpensive substrate-chitin (IMPACT® system, New England Biolabs) in both low cell density and high cell density cultures (Sharma *et al.*, 2006). Commercial expression vectors correspond to the IMPACT purification system have been constructed for both N- and C-terminal fusion (Figure 7).

Moreover, recent development is focused on utilization of ChBD as immobilizing tag. It was explored immobilization of the ChBD with d-hydantoinase on chitin given the enzyme could be reusable (Chern and Chao, 2005). In addition, immobilization of cells with surface-displayed ChBD was well described (Wang and Chao, 2006).

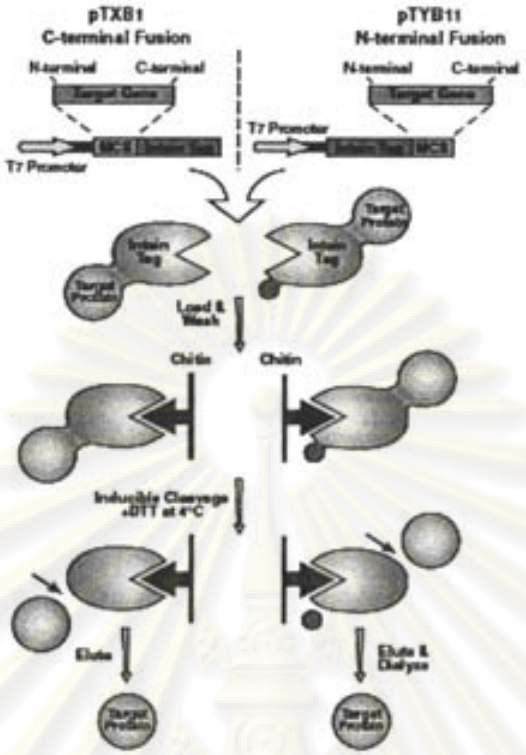


Figure 7. Schematic illustration of IMPACT purification system (New England Biolabs).

1.2.6 His-patch thioredoxin (Hp-Trx)

The expression system based on the use of *E. coli* thioredoxin (trxA, 11.7 kDa) as a gene fusion partner has been demonstrated by the production of a variety of mammalian cytokines and growth factors (LaVallie *et al.*, 1993). The fusion with thioredoxin can prevent the forming of an insoluble aggregates or inclusion bodies and increase solubilities of these proteins when expression in *E. coli*. A major disadvantage of the tag, however, is it cannot be purified with a specific matrix through affinity chromatography (Terpe, 2003). The engineering of a metal affinity site on the surface of the thioredoxin molecule can facilitate such problem (Lu *et al.*, 1996). They showed that a metal ion binding capacity can be conferred on thioredoxin by changing two or three surface-exposed residues to histidine and that these mutant forms of thioredoxin (termed "histidine patch" or His-patch thioredoxins) retain the wild type molecule's ability to act as an effective fusion partner (Lu *et al.*, 1996). Structure of Hp-thioredoxin mutant compare to wild-type is

illustrated in Figure 8. The Hp-trx fused proteins can be purified on Ni^{2+}-NTA column on the basis of IMAC and is now commercially available.

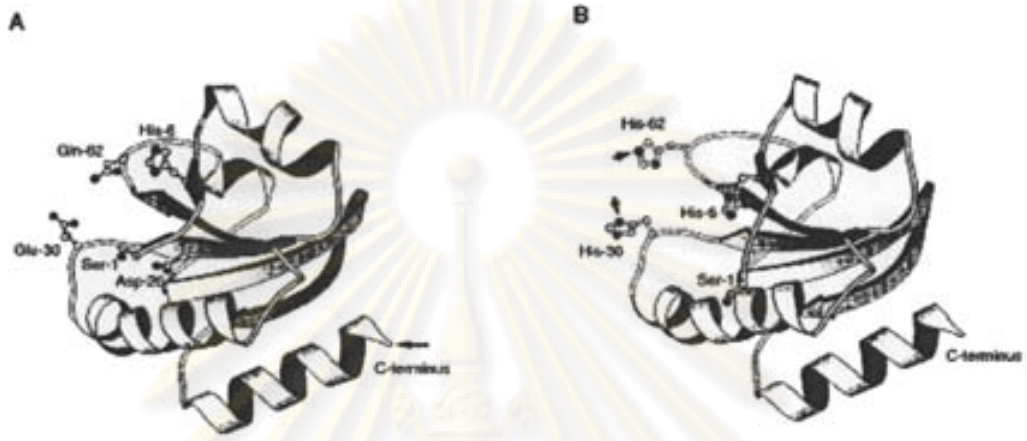


Figure 8. A: Structure of wild-type thioredoxin that was determined by x-ray crystallography. The arrow indicates the COOH-terminal carboxyl group where the fusion linker is attached. The side chains of five residues are shown in the figure: Ser-1, His-6, Asp-26, Glu-30, and Gln-62. In various thioredoxin mutants Ser-1, Glu-30, and Gln-62 were changed to histidine, His-6 was changed to asparagine, and Asp-26 was changed to alanine, either individually or in combination, resulting in the different thioredoxin mutants; B: A molecular model of hp2TrxA in which histidine side chains replaced the native glutamate and glutamine side chains at positions 30 and 62, respectively. The ND1 nitrogen of His-30 and the NE2 nitrogen of His-62, the proposed metal chelating groups, are indicated by arrowheads. The drawings were made with the program MOLSCRIPT (Lu *et al.*, 1996).

Class II

1.2.7 Calmodulin-binding peptide (CBP)

Purification by affinity chromatography on immobilized calmodulin is suitable for recombinant proteins expressed in bacteria because endogenous bacterial calmodulin-binding proteins are unknown. Importantly,

calmodulin is a relatively inexpensive protein (Pestov and Rydström, 2007). One step affinity chromatography using CBP tag has been established since 1992 (Stofko-Hahn *et al.*, 1992). They constructed the fusion protein of three functional units: a phosphorylation site for the cAMP dependent protein kinase, a recognition site for the coagulation factor Xa, and a peptide derived from the C-terminus of skeletal muscle myosin light-chain kinase that binds calmodulin in the presence of free calcium. These CBP-fused proteins could bind to calmodulin-agarose resin and could be eluted with calcium chelators such as EGTA or EDTA. After purification, the affinity tail can be cleaved by proteolysis with factor Xa.

The CBP consists of 26 amino acid residues and can be placed at both N- and C-terminus (Terpe, 2003). The translational efficiency, however, may decrease by the N-terminal fusion, while the high expression level can be achieved by the fusion at the C-terminus (Zheng *et al.*, 1997). Another protocol for purification of recombinant membrane proteins fused with CBD tag was described as illustrated in Figure 9.

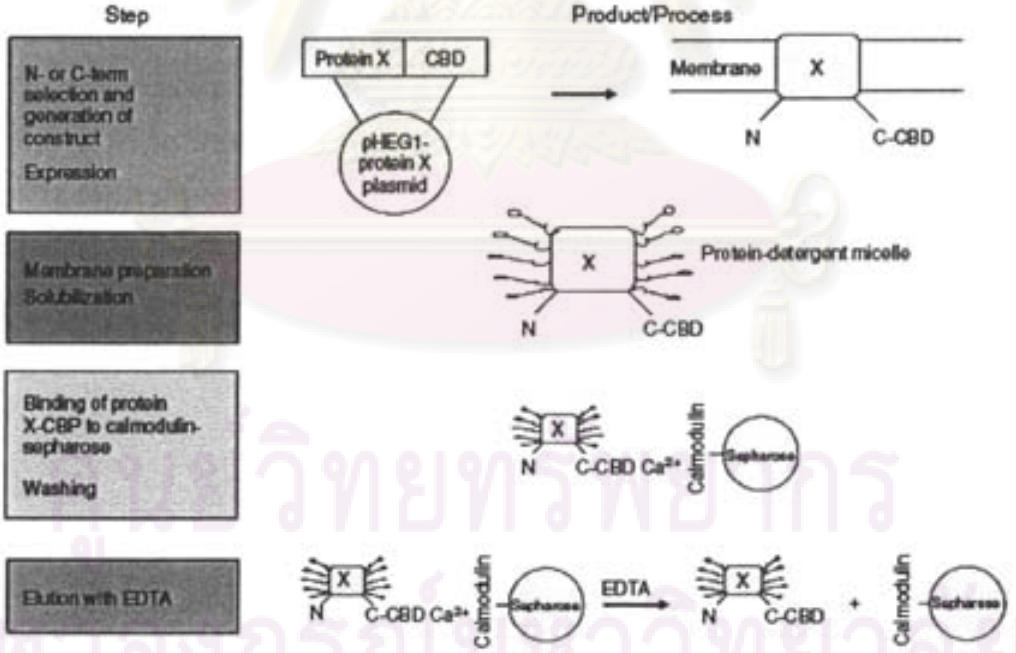


Figure 9. Summary of steps using CBD for purifying membrane proteins (Pestov and Rydström, 2007).

1.2.8 Streptavidin-binding peptide (SBP; Strep-tag II)

A nine-amino-acid peptide (Ala-Trp-Arg-His-Pro-Gln-Phe-Gly-Gly) with affinity toward streptavidin was firstly developed as an affinity tag for recombinant protein purification (Schmidt and Skerra, 1993). Competitive elution can be performed with biotin or biotin derivative as it occupy the same surface pocket of Strep tag:streptavidin complex (Schmidt *et al.*, 1996). Subsequently, a modified Strep tag termed Strep tag II which is compose of eight amino acids (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys) was created to increase affinity binding of streptavidin under the same condition (Schmidt *et al.*, 1996). However, the interaction of the Strep-tag II to streptavidin was still fairly weak (72 μM). The affinity of this interaction was then improved by a streptavidin mutant or Strep-Tactin that bind the Strep-tag II with the affinity constant about 1 μM (Voss and Skerra, 1997). The Strep-tag II fusion proteins are generally eluted from the Strep-Tactin affinity column under physiological buffer conditions using a low concentration of a biotin derivative for competition (Figure 10; Schmidt and Skerra, 2007).

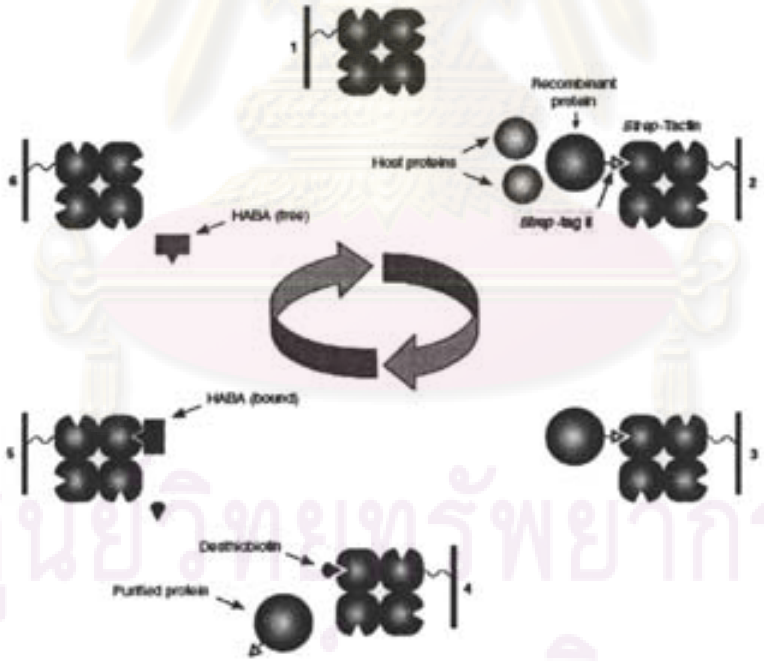


Figure 10. Schematic illustration of the Strep-tag purification cycle. The cell lysate containing the Strep-tag II fusion protein is applied to a column with

immobilized Strep-Tactin (Steps 1 and 2). After a short washing step with a physiological buffer, host proteins are removed (3) and the purified Strep-tag II fusion protein is specifically displaced via competition with a low concentration of D-desthiobiotin (4). D-Desthiobiotin removal is accelerated by application of a HABA solution and indicated via color change from yellow-orange to red (5). Finally, HABA is quickly removed by washing with a small volume of running buffer (6), thus making the column ready for the next purification run (1) (Schmidt and Skerra, 2007).

Recently, the new streptavidin-binding peptides possess nanomolar-affinity for streptavidin was presented in termed Nano-tag (Lamla and Erdmann, 2004). The first nano-tag_{15} is 15 amino acids long and binds to streptavidin with a dissociation constant of 4 nM. The Nano-tag_9 is 9 amino-acid peptides with a dissociation constant of 17 nM. Since the interactions between these two new tags and streptavidin are moderate when compare to the biotin-streptavidin system, they have been attracted for reversible immobilization on the available streptavidin-coated surface plasmon resonance (SPR) biosensor chip (Li *et al.*, 2006). The system is great benefit to various biosensors, biochips, and immunoassay applications based on the streptavidin capture surface (Li *et al.*, 2006).

Alternatively, a streptavidin-binding peptide named SBP-tag with a length of 38 amino acids was screened from peptide library and was used as affinity tag for protein purification. The dissociation constant of the tag is 2.5 nM and the binding capacity of immobilized streptavidin is 0.5 mg per ml of matrix (Wilson *et al.*, 2001; Keefe *et al.*, 2001).

Class III

1.2.9 FLAG tag

An affinity gene fusion system that has become popular in recent years is the so-called FLAG™ system, based on the fusion of an eight-amino-acid peptide to target protein for immunoaffinity chromatography on immobilized monoclonal antibodies (Nilsson *et al.*, 1997). This small tag is compost of Asp-Tyr-

Lys-Asp-Asp-Asp-Asp-Lys peptide sequence. The Tyr residue containing aromatic ring is the major factor in antigen-antibody interaction (Einhauer and Jungbauer, 2001). The peptide sequence is hydrophilic and contains an internal enterokinase cleavage recognition sequence (Asp-Asp-Asp-Asp-Lys) that enables the removal of the tag from target protein (Nilsson *et al.*, 1997; Einhauer and Jungbauer, 2001).

Until now, three monoclonal antibodies (M1, M2, M5) to the FLAG sequence were developed and usually be called anti-FLAG. The monoclonal antibody M1 can bind to the FLAG-tagged protein with calcium-dependent manner allowing protein elution by chelating agents (Hope *et al.*, 1996). One limitation of the anti-FLAG M1 is that it can only be located at extremely N-terminus of the fusion protein (Nilsson *et al.*, 1997; Einhauer and Jungbauer, 2001). Overcoming this drawback, anti-FLAG M2 which can bind to the fusion protein of FLAG tag at C-terminus was established (Brizzard *et al.*, 1994). The elution of the fusion protein from the anti-FLAG M2 is performed at low pH or competition with high concentration of synthetic FLAG peptides. A third monoclonal antibody (M5) was raised against the sequence Met-Asp-Tyr-Lys-Asp_4-Lys. The binding of the anti-FLAG M5 does not require calcium and exhibits a high affinity to for N-terminal FLAG fusion protein (Einhauer and Jungbauer, 2001).

A variety of host cells have been reported to use the FLAG tag system such as bacteria (Blanar and Rutter, 1992; Su *et al.*, 1992), yeast (Hopp *et al.*, 1988; Einhuauer *et al.*, 2002), and mammalian cell (Zhang *et al.*, 1991). However, the FLAG system cannot be applied in the presence of denaturants as they affect antibody or fusion stability. For this reason, expression of FLAG fusion proteins in inclusion bodies may not be the best choice (Einhauer and Jungbauer, 2001). Another disadvantage of the system is that the purification yield per volume of matrix is low because antibodies are used as the capture agents (Keefe *et al.*, 2001). In addition, immunoaffinity chromatography, applied to the production of therapeutic proteins has several drawbacks: ligand leakage, instability, need for validation of antibody production, and required removal of antigenic FLAG tag from therapeutic proteins (Einhauer and Jungbauer, 2001).

1.2.10 C-myc tag

The c-Myc tag consists of 10 amino acid residues (Glu-Gln-Lys-Leu-Ile-Ser-Glu-Glu-Asp-Leu). The tag was isolated from the transcription factor c-Myc which was an oncoprotein that plays a role in cell cycle regulation, metabolism, apoptosis, differentiation, cell adhesion, and tumorigenesis (Dang *et al.*, 1999; De Buck *et al.*, 2004). The epitope tag can be expressed in different protein contexts and still confer recognition by the 9E10 immunoglobulin developed by Evan *et al.* (Munro and Pelham, 1986; Evan *et al.*, 1985). The c-Myc tag can be used in many applications including Western blotting, immunoprecipitation, ELISA, and flow cytometry (Kipriyanov, 1996). Moreover, a single-step purification of the c-Myc fused proteins can be undertaken using the 9E10 antibody immobilized affinity column (Kramer *et al.*, 1997). However, it is a wildly used detection system but is rarely applied for purifications (Terpe, 2003).

While it was reported for expression in bacteria (Dreher *et al.*, 1991; Vaughan *et al.*, 1996), problem in expression of the c-Myc fused protein was occurred in *Legionella pneumophila* (De Buck *et al.*, 2004). It was observed that the amount of mRNA transcripts in the presence of the tag was lower than in absence. The authors assumed that there was either a degradation of the c-Myc tagged proteins or a decrease in protein production of the protein (De Buck *et al.*, 2004).

Other tag systems, which are not described in detail here, are also available e.g. poly-arginine (Sassenfeld and Brewer, 1984), S-tag (Connelly *et al.*, 1990), and NusA (Davis *et al.*, 1999) etc. Even though, new affinity tag systems are still continually developed in the recent year due to the widespread of using affinity tags throughout biotechnology. New peptide tag like AviD-tag is recently developed for purification on NeutrAvidin, a chemical modified avidin for reducing non-specific interactions (Pierce) (Gaj *et al.*, 2007). This tag can be used for protein purification and immobilization with higher yield than a Strep tag. In addition, a unique viral peptide sequence (R tag) and its specific monoclonal antibody are reported as an immunoaffinity system (Jones *et al.*, 2007). This system is suitable for application in

E. coli, *Arabidopsis thaliana*, or *Pichai pastoris* as it do not naturally presence in these host cells.

1.3 Comparison of advantage and disadvantage of affinity tag systems

As no single affinity tag is optimal with respect to all applications, each tag has its strengths and weaknesses (Waugh, 2005). Selection of an appropriate affinity tag depends on nature of target proteins, type of application of the proteins, stage of development of the proteins, costs of different chromatographic supports, and purification condition (Terpe, 2003; Arnau *et al.*, 2006). Some useful information about the common affinity tag systems were summarized in Table 2. Sizes of the tags are varying from 5 to 396 amino acid residues. The use of a small peptide tag should not harm the fused protein activity and may not need to remove, whereas the use of a large peptide or protein can increase solubility and promote proper folding of the target protein (Terpe, 2003). Another consideration is the cost of resin that they interact. The expensive resin with low binding capacity is less attractive for high-throughput applications. In addition, advantages and disadvantages of various tags were described in Table 3 with respect to their elution condition, solubility enhancement, and cost of resin (Waugh, 2005).

More recently, there was a systematic comparison of affinity tags in term of purification yield and purity (Lichty *et al.*, 2005). Different affinity tags were fused with dihydrofolate reductase (DHFR) which can be easily performed a quantitative spectrophotometric assay. The silver staining of SDS-PAGE gels revealed purity gained from each purification tag. Epitope tags such as the FLAG tag showed highest purity (Figure 11, lane 7), in contrast to the His tag that produced many contaminants after purification on Talon cobalt metal resin (Figure 11, lane 1-3). However, the His tag provided the highest elution yield of almost 70% for the tag/resin tested (Table 4).

Table 2. Comparison of common affinity tag systems (Nilsson *et al.*, 1997; Terpe, 2003; Lichty *et al.*, 2005).

Tag	Size (a.a.)	Ligand	Detection	Elution	Source	Capacity[1]	Cost /ml resin[1]
His	5-15	TALON	Anti-His	Imidazole, low pH	Clontech	5-14 mg/ml	$9
		Ni^{2+}-NTA			Qiagen		$10
GST	201	Glutathione	Anti-GST	Reduced glutathione	Amersham	5-10 mg/ml	$16
CBD	27-189	Cellulose (CBIND)	Anti-CBD	Family I: guanidine HCl or urea > 4M Family II,III: ethylene glycol	Novagen	N/A	N/A
MBP	396	Amylose	Anti-MBP	Maltose	NE Biolabs	3 mg/ml	$10
ChBD	52	Chitin	Anti-ChBD	Fused with intein:30-50 mM DTT, β-mercaptoethanol or cysteine	NE Biolabs	2 mg/ml	$2
Hp-Trx	109	Ni2+-NTA	Anti-Thio	Imidazole	Qiagen	N/A	$10
CBP	26	Calmodulin	Anti-calmodulin	EGTA	Strategene	2 mg/ml	$23
Strep II	8	Strep-tactin	Strep-tactin	Desthiobiotin	IBA	50-100 nmol/ml	$44
FLAG	8	Anti-FLAG mAb	Anti-FLAG M2	EDTA, low pH, FLAG peptide	Sigma	0.6 mg/ml	$63
c-Myc	10	Red Anti-c-Myc	Anti-c-Myc	Low pH, c-Myc peptide	Sigma	1-1.5 mg/ml	$510

[1] Capacity of resins for the fusion protein and cost information was obtained from manufacturer's catalogs and websites.

Table 3. Advantages and disadvantages of some commonly used fusion partners (Waugh, 2005).

Tag[1]	Advantage	Disadvantage
GST	Efficient translation initiation	High metabolic burden
	Inexpensive affinity resin	Homodimeric protein
	Mild elution conditions	Does not enhance solubility
MBP	Efficient translation initiation	High metabolic burden
	Inexpensive affinity resin	
	Mild elution conditions	
	Enhance solubility	
NusA	Efficient translation initiation	High metabolic burden
	Enhance solubility	
	Not an affinity tag	
Thioredoxin	Efficient translation initiation	Not an affinity tag[2]
	Enhance solubility	
Ubiquitin	Efficient translation initiation	Not an affinity tag
	Might enhance solubility	
FLAG	Low metabolic burden	Expensive affinity resin
	High specificity	Harsh elution conditions
His_6	Low metabolic burden	Specificity of IMAC is not as high as other affinity methods
	Inexpensive affinity resin	
	Tag works under both native and denaturing conditions	Does not enhance solubility
STREP	Low metabolic burden	Expensive affinity resin
	High specificity	Does not enhance solubility
SET	Enhance solubility	Not an affinity tag
CBP	Low metabolic burden	Expensive affinity resin
	High specificity	Does not enhance solubility
	Mild elution conditions	
S-tag	Low metabolic burden	Expensive affinity resin
	High specificity	Harsh elution conditions (or on-column cleavage)
		Does not enhance solubility

1 GST, glutathione S-transferase; MBP, maltose-binding protein; NusA, N-utilization substance A; FLAG, FLAG-tag peptide; His6, hexahistidine tag; STREP, streptavidin-binding peptide; SET, solubility-enhancing tag; CBP, calmodulin-binding peptide.

2 Derivatives of thioredoxin have been engineered to have affinity for immobilized metal ions (His-patch thioredoxin) or avidin/strepavidin (Smith *et al.*, 1998).

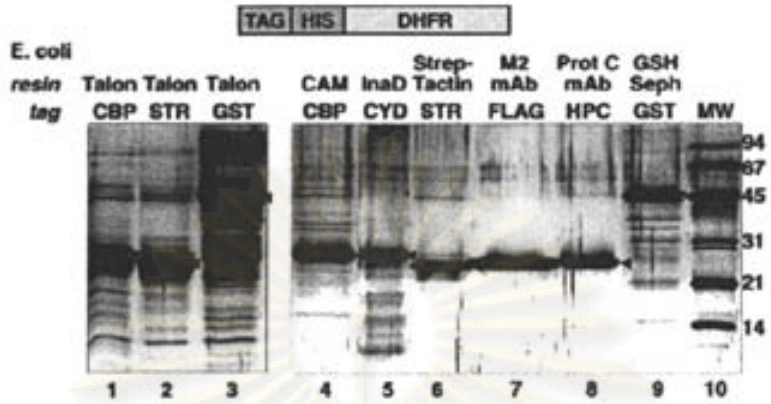


Figure 11. Comparison of affinity tags to purify tagged DHFR proteins expressed in *E. coli*. As shown in the schematic above the gels, each of the tagged recombinant proteins contain a specific N-terminal tag (identified by TAG label just above each lane), followed by the His tag and DHFR. The CYD tagged polypeptide was C-terminally tagged as CBP-HIS-DHFR-CYD. Eluted fractions were fractionated by SDS–PAGE and the gels were silver stained. Each tagged protein is highlighted by a red arrowhead. The identity of the tag used for each purification is shown immediately above each lane and the identity of the affnity resin just above that. CBP, calmodulin-binding peptide; STR, Strep II; GST, glutathione S-transferase; CYD, covalent yet dissociable NorpA peptide; and HPC, heavy chain of protein C. Abbreviations used for affnity resins: CAM, calmodulin; InaD, PDZ domain of InaD protein; M2 mAb, anti-FLAG M2 monoclonal antibody; Prot C mAb, anti-protein C (clone HPC4) monoclonal antibody; GSH Seph, GSH–Sepharose.

Table 4. Purification yields for active tagged DHFR proteins from *E. coli* extracts determined by NADPH assay (affinity tag used to purify the protein highlightd in bold) (Lichty *et al.*, 2005).

Tag	Resin	Yield (elution fraction) (%)	Recovery (flow through and fractions) (%)
CBP-HIS	Talon	67.7 ± 4.3	82.4 ± 2.4
CBP-HIS	Calmodulin	55.3 ± 7.6	88.6 ± 5.4
STR-HIS	Strep-Tactin	61.5 ± 6.1	87.6 ± 5.7
FLAG-HIS	M2 MAb	56.1 ± 2.3	84.5 ± 2.5
HPC-HIS	Prot C MAb	62.5 ± 1.2	80.3 ± 3.9
CYD (c-term)	InaD PDZ	57.6 ± 4.0	81.1 ±4.6
GST-HIS	GSH-Seph	40.8 ± 5.0	77.8 ±4.6

1.4 The tag removal by proteolytic enzymes

Either large or small affinity tags have the potential to interfere with the biological activity of a protein or certain applications such as crystallization or NMR experiments. Also the affinity tag can cause unwanted immune responses, alter the properties of the target protein. Besides the chemical cleavages which are performed at rather harsh condition, removal of the tag by proteolytic enzymes at mild condition is more preferable (Nilsson *et al.*, 1997). The removal of the tag can simply introduced the specific cleavage site of the protease between the tag and the target protein. The most commonly used proteases are enterokinase, tobacco etch virus (TEV), thrombin, and factor Xa (Table 5; Terpe, 2003). Finally, a downstream process for the protease removal has to be performed to obtain the real purified protein. Some of these proteases have been fused to an affinity tag allowing the protease removal on an affinity resin (Figure 12). For example, PreScisson protease™ can bound to a glutathione resin as it was previously fused with GST tag (Amersham Bioscience).

Table 5. Recognition sequences of proteolytic enzymes.

Protease	Recognition sequences	References
Enterokinase	Asp-Asp-Asp-Asp-Lys	Terpe, 2003
TEV	Glu-Asn-Leu-Tyr-Phe-Gln-Ser	Carrington and Dougherty, 1988
Thrombin	Leu-Val-Pro-Arg-Gly-Ser	Nilsson *et al.*, 1997
Factor Xa	Ile-Glu-Gly-Arg	Nagai and Thogerson, 1984
PreScisson protease	Leu-Glu-Val-Leu-Phe-Gln-Gly-Pro	Arnau *et al.*, 2006

In addition, separation of the tag from the target protein can accomplished using a self-splicing intein (Xu *et al.*, 2000). Inteins are naturally occurring protein sequences capable of post-translational self excision from a host-intein precursor protein through a process known as "protein splicing" (Banki and Wood, 2005). The first commercially available intein purification system (IMPACT™) was developed by the combination of a chitin-binding domain and modified *Saccharomyces cerevisiae* vacuolar ATPase subunit A intein (New England Biolabs). The mutation taken in this intein lead it enable the cleavage at the N-terminus in the presence of 30 mM DTT or β-mercaptoethanol over a wide pH range (5.5-9.0) (Banki and Wood, 2005). However, the intein system has not been tested for high-throughput applications. The large size of the catalytic machinery that must be incorporated into the fusion protein can increase the metabolic burden in the cells. Moreover, the slow rate of autoprocessing is also a drawback of the system (Waugh, 2005).

Figure 12. Concept of affinity-tagged protease removal in a single affinity chromatography step (Nilsson *et al.*, 1997).

1.5 Future directions of affinity tag applications

A variety of affinity tags and methods for tag removal have been developed. Most of these methods are suitable for laboratory scale applications (Arnau *et al.*, 2006). On the other hand, in a competitive pharmaceutical market where production costs are high, the development of more economically sound production and purification processes may be a trend to follow (Arnau *et al.*, 2006). Utilization of a combined affinity tag such as the tandem affinity purification (TAP) tag might be applicable. The TAP tag is a combination of IgG binding domain of *Straphylococcus aureus* protein A and a calmodulin binding peptide (Riguat *et al.*, 1999; Puig *et al.*, 2001). The TAP purification system enables purification of target protein and removal of the tag in two sequential affinity columns (Figure 13).

Moreover, large-scale structural genomics centers on the world have made great step in the development of automated tools for high-throughput protein expression and purification, although only a few different strategies for the protein production and purification have been reported (Waugh, 2005). For example, an automated method for high-throughput protein purification applied for a His tag and a GST tag has been reported (Scheich *et al.*, 2003). In addition, directed evolution of existing proteins will also benefit from a universal platform for production and purification (Arnau *et al.*, 2006). Another prospective is the use of affinity tag in peptide/protein chip design for determination of binding constant and drug delivery application (Terpe, 2003).

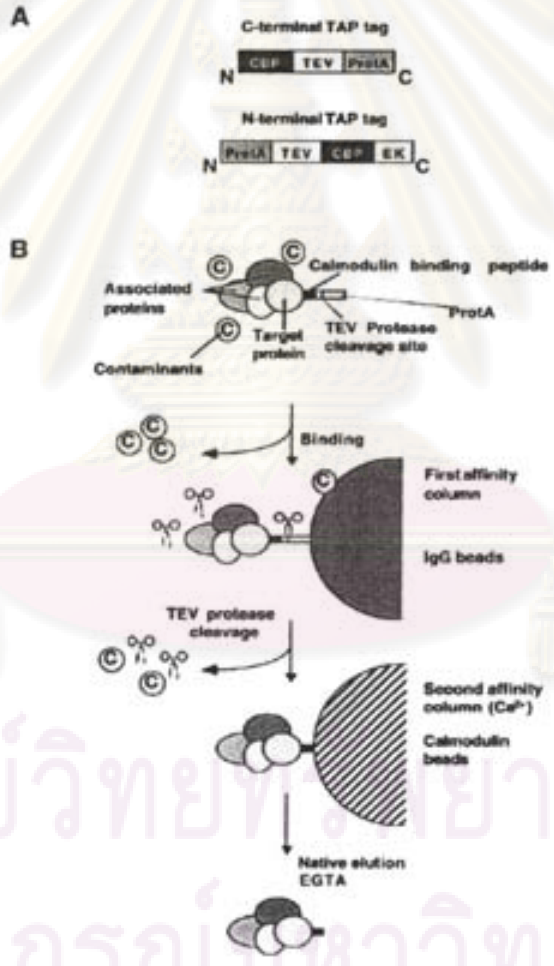


Figure 13. A: Schematic representation of the C- and N-terminal TAP tags. B: Overview of the TAP purification strategy (Puig *et al.*, 2001).

2. C-terminal domain of glucansucrases

2.1 Glucansucrases

Glucansucrases (E.C. 2.4.1.5) are extracellular enzymes mainly produced by the soil bacterium *Leuconostoc mesenteroides* (commonly termed dextransucrase, DSR), *Streptococcus* species from the oral flora (commonly termed glucosyltransferase, GTF), and by lactic bacteria *Lactococci*. They catalyse the synthesis of high molecular weight D-glucose polymer or glucan from sucrose (reaction 1). When efficient acceptors, like maltose or isomaltose, are added to the reaction medium, the glucansucrases catalyze the synthesis of low molecular weight oligosaccharides instead of high molecular weight glucan (Monchois *et al.*, 1999).

Reaction 1: n Sucrose $\longrightarrow$ $(Glucose)_n$ + n Fructose (Monsan and Paul, 1995)

As sucrose is a widely available and cheap substrate, they are very attractive for the synthesis of novel oligosaccharides and polymers (Remaud-Simeon *et al.*, 2003). Furthermore, the GSF were found to involve in the plaque formation causing dental caries (Hamada and Slade, 1980). Therefore, many researches were investigated their genetic, reaction mechanism, and structure-function relationship.

2.1.1 Enzymatic classification and characterization

In the enzyme nomenclature, glucansucrases were characterized to be E.C. 2.4.1.5 based on their reaction catalyzed and product specificity. They can also categorize to the glucoside hydrolase (GH) family 70 according to amino acid sequence similarities (Henrissat and Davies, 1997). Currently, 70 members of glucansucrases are belonging to the GH family 70 on carbohydrate active enzyme database available at http://www.cazy.org (Coutinho and Henrissat, 1999). The glucansucrases from different strains produce different kinds of glucans with different sizes and structures.

The products are compost of a main linear chain of D-glucose polymer, either principally linked through α(1-6) glucosidic bonds (dextran polymer),

or through α(1-3) glucosidic bonds (mutan polymer) as well as linked alternately through α(1-6) and α(1-3) glucosidic bonds (alternan polymer) (Monchois *et al.*, 1999). Moreover, glucans also differ from α(1-2), α(1-3), α(1-4), and α(1-6) glucosidic bonds, the degree of branching, the length of branch chains and their spatial arrangement (Monchois *et al.*, 1999). The main characteristics of some glucansucrases are presented in Table 6.

Table 6. Main characteristics of some glucansucrases. (Monchois *et al.*, 1999).

Strain	Gene	Glucan[1]	Size (a.a.)	M.W. (kDa)
S. mutans GS5	*gtf-B*	87% α(1-3)	1475	150
		13% α(1-6)		
	gtf-C	85% α(1-3)	1375	140
		15% α(1-6)		
	gtf-D	85% α(1-3)	1430	155
		13% α(1-6)		
S. downei Mfe28	*gtf-I*	88% α(1-3)	1556	160
		12% α(1-6)		
	gtf-S	10% α(1-3)	1328	147
		90% α(1-6)		
S. sobrinus OMZ176 (serotype d)	*gtf-T*	27% α(1-3)	1542	163
		73% α(1-6)		
	gtf-I	Nd[2]	1590	175
S. gordonii (*S. sanguis*)	*gtf-G*	40% α(1-3)	1578	170
		60% α(1-6)		
L. mesenteroides NRRL B-512F	*dsr-S*	5% α(1-3)	1290	170
		95% α(1-6)		
L. mesenteroides NRRL B-1299	*dsr-A*	15% α(1-3)	1290	146
		85% α(1-6)		
	dsr-B	5% α(1-3)	1508	167
		95% α(1-6)		

1 Structure of the produced glucan presented in percentage of total linkages

2 not determined

2.1.2 Reaction catalyzed by glucansucrases

Glucansucrases catalyze the cleavage of glycosidic bond of their substrate sucrose and couple a glucose unit to a growing glucan (polyglucose) chain (transglucosylation), water (hydrolysis), or other acceptor substrate (acceptor reaction) (Figure 14). The energy released by cleavage of the energy-rich glycosidic bond in sucrose is used for synthesis of new glucosidic bonds (van Hijum *et al.*, 2006). For the acceptor reaction, the acceptor can be a saccharide molecule like maltose, isomaltose, and O-α-methylglucoside causing the production of oligosaccharide instead of a glucan polymer (Monchois *et al.*, 1999).

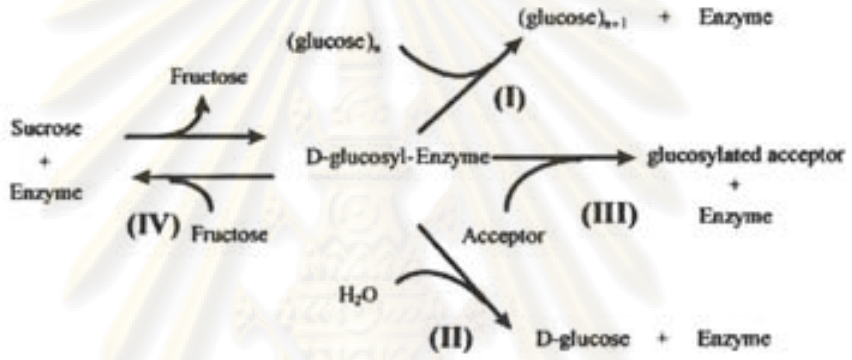


Figure 14. Reaction catalyzed by glucansucrase. (I) glucan synthesis by successive transfer of glucosyl units; (II) sucrose hydrolysis by transfer of the glucosyl unit onto water; (III) oligosaccharide synthesis by transfer of the glucosyl unit onto an acceptor molecule; and (IV) isotopic exchange by reverse reaction of glucosyl-enzyme complex formation (Monchois *et al.*, 1999).

2.1.3 Mechanism of action

The glucan synthesis mechanism was proposed that the reaction occurring in 3 steps: (1) initiation, (2) elongation, (3) termination (Tsuchiya *et al.*, 1953).

(1) Initiation

The initiation of reaction required a primer like glucan, glycogen, or starch for activating the polymerization (Monchois *et al.*, 1999). The

hypothesis has been supported by the fact that addition of exogenous glucan has an activating effect on glucan synthesis (Kobayashi *et al.*, 1986; Germaine *et al.*, 1974; Germaine *et al.*, 1977). However, in the absence of any exogenous primer, the glucansucrases are still active. Binding of glucan to enzyme may promote a change in the catalytic site conformation allowing the enzyme to become more active (Mooser, 1992).

(2) Elongation

The mechanism of autopolymerization and direction of chain growth remains not fully understood (Monchois *et al.*, 1999). There are two mechanisms proposed in the elongation step. The first mechanism occurs at the non-reducing end of glucan chain (Figure 15A). In this mechanism, the aspartic acid or glutamic acid provide a nucleophilic group on the C_1 of the glucosyl moiety of sucrose, leading to the formation of a covalent glucosyl-enzyme complex. Another residue acts as proton donor to give the release of fructose (Sinnot, 1990). Then, another glucosyl residue can also be activated by trapping the hydrogen from the hydroxyl group linked to the C_6 (Monchois *et al.*, 1999). However, the requirement of a primer to start the glucan synthesis making this mechanism is irrelevant (Su and Robyt, 1994).

Another mechanism is the two-site mechanism which elongation occurs at the reducing end of the glucan chain (Ebert and Shenk, 1968). It was found that glucan and glucosyl residues coming from sucrose cleavage were both covalently linked with enzyme through their reducing end (Robyt *et al.*, 1974). This mechanism is composted of two different steps: (1) the two nucleophilic sites attack two sucrose molecules leading to the release of two fructose molecules and giving two glucosyl residues covalently linked to the enzyme; and (2) the OH-C_6 of one of the two glucosyl residues may make a nucleophilic attack on the C_1 of the other one. That promotes the formation of an α(1-6) bond and the release of one or two nucleophilic sites which can attack sucrose in order to create a new covalent glucosyl-enzyme complex (Monchois *et al.*, 1999; Figure 15B).

(3) Termination

Termination step correspond to the dissociation of glucan from the enzyme (Monchois *et al.*, 1999).

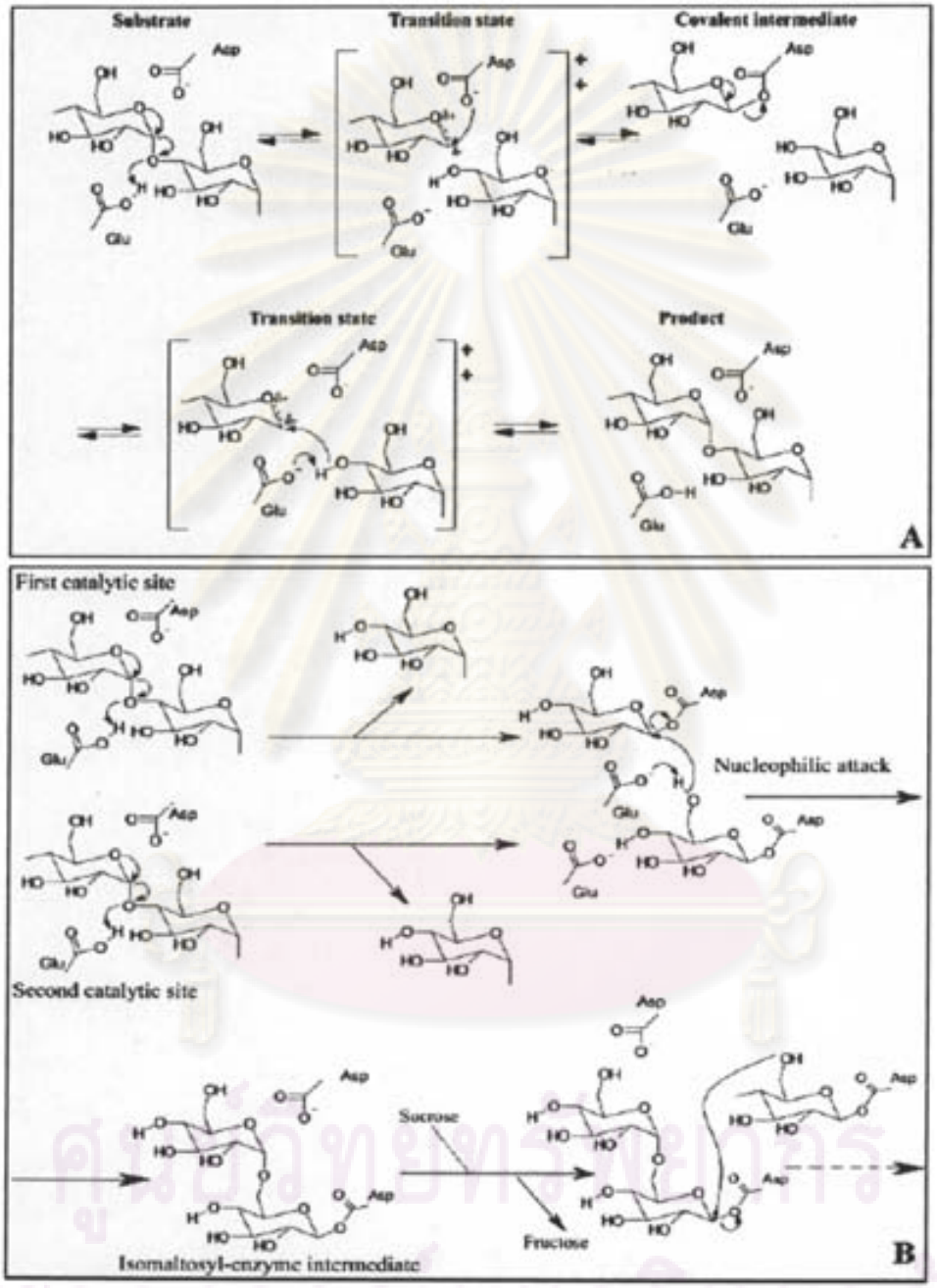


Figure 15. Schematic representation of the two polymerization mechanisms of glucansucrases. A, mechanism involving only one active site, as suggested by

Mooser (Mooser, 1992) and resembling that of α-retaining transglucosidases; B, mechanism involving two active sites, as proposed by Robyt *et al.* (1974) (Source : Moulis *et al.*, 2006a).

However, identification of active site found only one catalytic triad (2 aspartic acids, 1 glutamic acid). Considering the sequential insertion, it is also difficult to understand how the OH-C_6 of one glucosyl group is able to attack the C_1 of the dextransosyl group when it is far away, except if a configuration inversion occurs (Monchois *et al.*, 1999). Moreover, the recent study showed that sucrose and glucose produced by hydrolysis reaction act as initiator for glucan polymerization. Then, elongation occurs by transfer of glucosyl residue coming from sucrose to the non-reducing end of initially formed products (Moulis *et al.*, 2006a). Based on results from Moulis *et al.* (2006a) the first mechanism that involve only one active site and polymerization occurs at non-reducing end of glucan chain seem to have more relevant.

2.1.4 Primary structure

Glucansucrases are large enzymes with an average molecular weight of 160 kDa. Due to its large size, no three-dimensional structure is now available. The sequence comparison of these enzymes showed that they all are compost of four domains: a signal peptide, a variable region, a catalytic domain, and a C-terminal glucan binding domain (Figure 16; Monchois *et al.*, 1999; Remaud-Simeon, 2003). The N-terminal signal peptide is a well conserved region involving in the enzyme secretion. The second motif, variable region, has only few conserved amino acids and seems to have no important role in the enzyme mechanism (Monchois *et al.*, 1999). The highly conserved catalytic domain was proposed to contain $(\beta/\alpha)_8$-barrel of GH family 13 α-amylase enzymes (Moulis *et al.*, 2006a). The last C-terminal domain was predicted to involve in glucan binding as truncation of the C-terminal end of GTF-I was suppressed its glucan binding ability (Ferretti *et al.*, 1987).

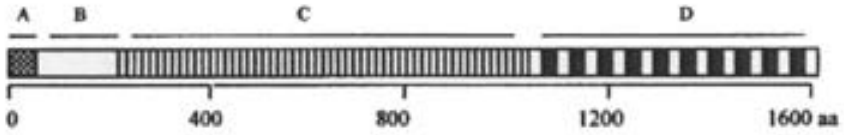


Figure 16. Schematic structure of glucansucrases for which encoding genes have been cloned. A, signal peptide; B, variable region; C, N-terminal catalytic domain; D, C-terminal glucan binding domain (Monchois *et al.*, 1999).

2.1.5 Role of glucansucrases

Glucansucrases were identified more than 50 year ago. They have an industrial value to produce dextrans and oligosaccharides (Monchois *et al.*, 1999). The potential uses of native dextran include secondary recovery of petroleum from oil drilling muds, stabilization of soil aggregates, protective coating for seeds, deflocculants in paper products, metal plating processes, and food stabilizer (Cerning, 1990). Low molecular weight dextrans have their biggest outlet in the pharmaceutical industry, where fractions of various molecular weights and molecular distributions have been used as blood plasma extenders and blood flow improvers (Cerning, 1990). For example, dextran with 70 kDa is used as a blood volume expander and dextran with 40 kDa is used as a blood flow improver. In addition, oligosaccharide containing α-1,2 linkages produced by a dextransucrase from *L. mesenteroides* NRRL B-1299 is used as an prebiotic substance that encourage the growth of a beneficial microflora (Monchois *et al.*, 1999; Remaud-Simeon *et al.*, 2003).

Moreover, the native dextran can be employ as starting material, which is then submitted to acid or enzymatic hydrolysis, solvent fractionation procedures for making pharmaceutical and industrial dextran products (Cerning, 1990). In separation manufacturer, Dextran is also applied as molecular sieves by preparing a cross-linking dextran with epichlorohydrin in the presence of sodium hydroxide (Sephadex®), with bis-acrylamide (Sephacryl®), or with agarose (Superdex®) (Amersham Biosciences).

Glucans or mutans synthesized by *S. mutans* and *S. salivarius* are critically important in dental plaque formation causing a dental caries because they are water-insoluble and possess a marked ability to promote adherence when synthesized *de novo* on various solid surfaces (Cerning, 1990). Initiation of dental plaque occurs by the activity of extracellular GTF of streptococci to form high molecular weight branched extracellular glucans. These glucose polymers provide the aggregation of mutans and other oral streptococci through interaction with bacterial cell-associated glucan-binding proteins. These accumulations of streptococci which produce lactic acid will ensure the dental caries by destroy dental enamel (Smith, 2003).

2.1.6 Dextransucrase S (DSR-S)

Dextransucrase (DSR-S) from *L. mesenteroides* NRRL B-512F is a glucansucrase that mainly produce D-glucose polymer or dextran containing 95% α(1-6) and 5% α(1-3) linkages (Binder *et al.*, 1983; Robyt and Eklund, 1983). As its abilities to produce a wide variety of dextrans and oligosaccharides by addition of different acceptors, it was extensively use for production of industrial polymers such as chromatographic supports, photographic emulsions, iron carriers, and blood plasma substitutes (Kim and Robyt, 1994; Monsan and Paul, 1995). It is also important because of its theoretical and practical aspects in understanding the mechanism of glucan synthesis (Kim and Robyt, 1994). The DSR-S has a molecular weight of 175 kDa and composes of 1,527 amino acid residues which can be divided into four domains like all glucansucrases. In the DSR-S sequences, the three catalytic triads locate at Asp-551, Glu-589, and Asp-662 positions (Moulis *et al.*, 2006a).

The characterization of DSR-S activity revealed that optimal pH and temperature of both soluble and immobilized enzyme were pH 5.2 and 30 ºC respectively (Figure 17; Kaboli and Reilly, 1980). Moreover, addition of Ca^{2+} significantly stabilized the soluble dextransucrase.

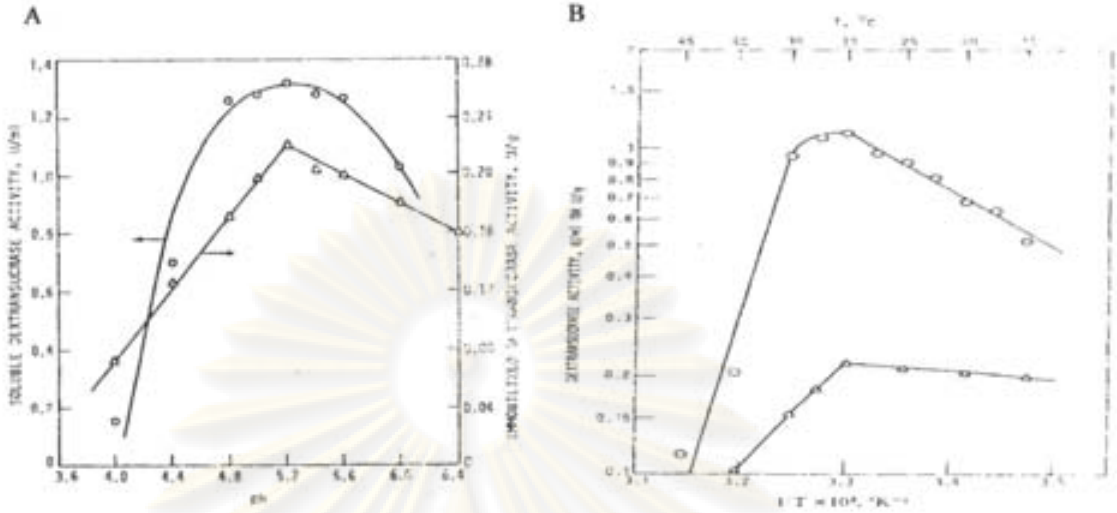


Figure 17. A: Effect of pH on dextransucrase activity at 30 °C, B: Effect of temperature on dextransucrase activity at pH 5.2. (○) soluble DSR-S, (Δ) immobilized DSR-S (Kaboli and Reilly, 1980).

Working with the native enzyme, expression of DSR-S was found to have some problems associated with the synthesis of glucan in the presence of sucrose in culture medium (Kim and Robyt, 1994). The sucrose, however, require for production of the enzyme. The glucan synthesized can cause the enzyme:glucan complex and the high viscosity of the glucan itself which are major problem for the studies of structure-function relationship of the enzymes. Cloning of the *dsr-S* gene to *E. coli* was then carried out to avoid these limitations (Monchois *et al.*, 1997). The gene was cloned into pTrc 99A under the control of the *trc* promoter and induction by IPTG. From this system, 200 U L^{-1} culture of DSR-S were produced overnight at 30 °C in LB supplemented with 100 mM tris HCl pH 6.4. More recently, the expression system for the DSR-S was optimized by producing in pBAD/TOPO Thiofusion vector (Invitrogen) controlled by arabinose operon (Moulis *et al.*, 2006b). The recombinant DSR-S were produced 5850 U L^{-1} culture at 23 °C in 2X YT medium supplemented with 100 mM tris HCl pH 6.4. However, various molecular mass forms of active DSR-S were obtained as also found in previous studies (Figure 18).

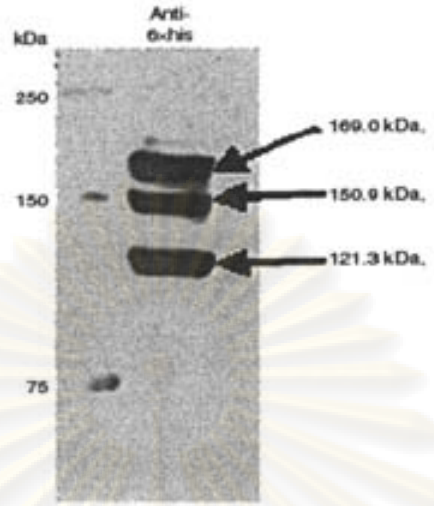


Figure 18. Anti-6x His Western blot from Thioredoxin-DSR-S-His produced at 23 °C (Moulis *et al.*, 2006b).

2.2 The C-terminal glucan binding domain of glucansucrases

Despite to numerous researches that focused on the characterization of N-terminal catalytic domains and enzyme activity, a number of published results were also investigated the structure and biological function of the C-terminal domain. This domain is usually called glucan-binding domain (GBD) as its proposed function. In GTF, the C-terminal domain was supposed to have dual functions: glucan synthesis and glucan-mediated binding of bacterial cells on tooth surfaces (Ferretti *et al.*, 1987; Colby and Russell, 1997). The recent study confirmed that the C-terminal domain of GTF enzymes participated in cellular adherence to tooth surfaces and contribute to the cariogenicity of *S. mutans* (Figure 19; Matsumura *et al.*, 2003). Therefore, this domain has been received attention as vaccines against the dental caries. To date, the GbpB of *S. mutans* has been shown to induce protective immune responses to experimental dental caries. Interestingly, salivas of young children often contain IgA antibody to the GbpB, indicating that initial infection with *S. mutans* can lead to natural induction of immunity to this protein (Smith, 2003).

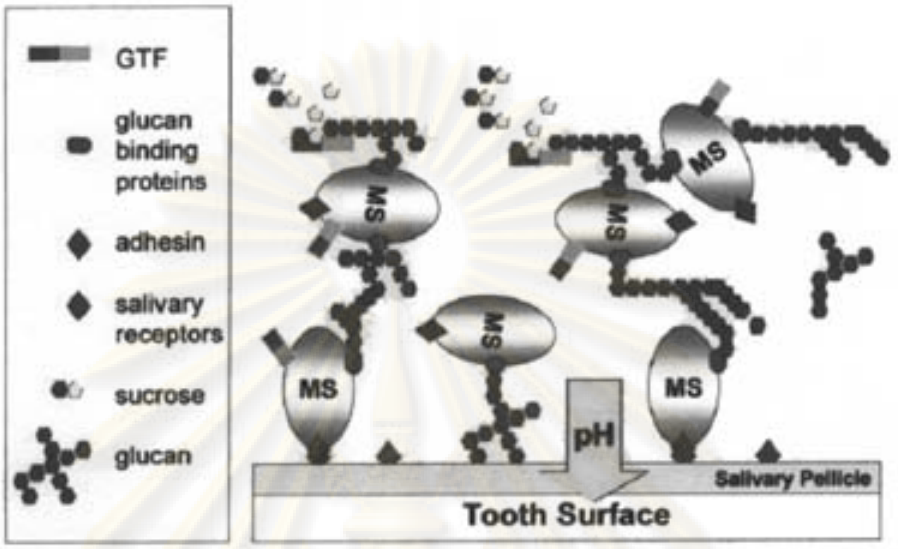


Figure 19. Molecular pathogenesis of mutans streptococci (MS) (Smith, 2003).

In addition, the C-terminal domain of glucansucrases has been interested as a potential affinity tag for recombinant protein purification. The dextran-binding domain (DBD) of GTF from *S. sobrinus* was demonstrated the affinity purification of a green and a red fluorescent protein or GFP and RFP respectively (Kaseda *et al.*, 2001). The purified proteins of about 90 % purity were obtained after the purification using the DBD tag on a cross-linked dextran support of Sephadex G-100 and elution by 50 g/l of dextran 18.1 kDa (Figure 20). The advantage of this system is its low cost compared to other systems because expensive resins are not required.

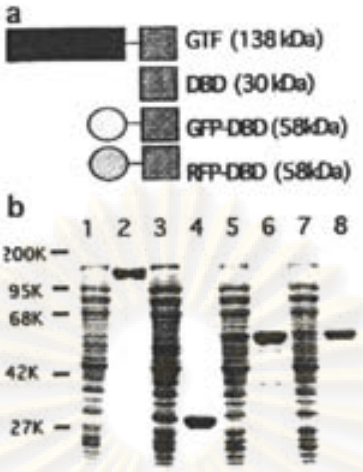


Figure 20. One-step purification of recombinant proteins using DBD as the tag. Schematic representation of the recombinant proteins (a) and SDS-PAGE analysis of the protein preparations (b). The calculated molecular weight of each protein is indicated in parentheses. The parent GTF (lanes 1 and 2), its functional moiety DBD (lanes 3 and 4), the DBD-tagged proteins GFP-DBD (lanes 5 and 6) and RFP-DBD (lanes 7 and 8) were adsorbed with Sephadex G-100 beads and eluted with 25 mg/ml dextran (18.1 kDa). The SDS-PAGE patterns of the proteins before (lanes 1, 3, 5 and 7) and after (lanes 2, 4, 6 and 8) this one-step purification step is shown. After the purification, the SDS-PAGE gel shows mainly single strong bands approximately corresponding to the expected molecular weights. The purities of those proteins were 95%, 96%, 89%, and 92% for GTF, DBD, GFP-DBD, and RFP-DBD, respectively (Kaseda *et al.*, 2001).

2.2.1 Structure of the C-terminal domain

The C-terminal domain of glucansucrases is compost of a series of repeated units which have been divided into 5 classes: A, B, C, D, and N repeats according to the sequence similarity (Table 7; Ferretti *et al.*, 1987; Gilmore *et al.*, 1990; Abo *et al.*, 1991; Giffard *et al.*, 1991; Monchois *et al.*, 1998). The number and distribution of these repeats are specific to each enzyme as shown in Table 8 (Monchois *et al.*, 1999). It was revealed that the A repeat always present in all glucansucrases, whereas the D and N repeats are specific to the enzyme produced by

S. salivarius ATCC 25975 and *L. mesenteroides* NRRL B-512F respectively. They are composed of a common amino acid pattern: an aromatic amino acid residues stretch, sometimes include tyrosine, surrounded by conserved residues (von Eichel-Streiber *et al.*, 1992). The following consensus sequence is IDGYYFD+N+G.

Table 7. The consensus sequences of C-terminal domain (x indicates non conserved amino acid)

Repeat	Consensus sequence	References
A	WYYFNxxDGQAATGLQTIDGQTVxxFDDNxGQxVK	Monchois *et al.*, 1998
B	VNGKTYYFGSDGTAQTQANPKGQTFKDGSGVLRFYNLEGQYVSGSG	Ferretti *et al.*, 1987
C	DGKIxxYFFDPDSGEVVKNRFV	Monchois *et al.*, 1998
D	GGVVKNADGTYSKY	Giffard *et al.*, 1994
N	YYFxAxQGxxxL	Monchois *et al.*, 1998

Table 8. Pattern of repeated units composing the C-terminal domain of glucansucrases (Monchois *et al.*, 1999).

Strain	Gene	Repeating unit
S. mutan GS5	*gtf-B*	A-A-C-A-C-A-C-A-C-A-C-A-C
	gtf-C	A-A-C-A-C-A-C-A
	gtf-D	A-A-A-A-A
S. downei Mfe28	*gtf-I*	A-A-C-A-C-A-C-B-A-C-B-A-C
	gtf-S	A-A-C-A-C-A-C
S. sobrinus 6715 (serotype g)	*gtf-Ia*	A-A-C-A-C-A-C-A-C-A-C
S. sobrinus OMZ176 (serotype d)	*gtf-T*	A-A-C-A-C-C-A-A-C
	gtf-Is	A-A-C-A-C-A-C-A-C-A-C
S. salivarius ATCC 25975	*gtf-J*	A-D-A-D-A-D
	gtf-K	A-D-A-A-A-D-A-D-A-D-A-D
	gtf-L	A-A-C-A-C-A-C
	gtf-M	A-C-A-C-A-A-C-A
L. mesenteroides NRRL B512F	*dsr-S*	A-C-C-A-A-N-C-N-N-N-N
L. mesenteroides NRRL B-1299	*dsr-A*	A-A-A-C-A-C-A-C
	dsr-B	A-C-C-A-A-C

Sequences similar to the A repeats also present in other organisms, notably the C-terminal part of a dextranase inhibitor protein produced by *S. sobrinus* UAB 108 acting like a glucan binding protein (Sun *et al.*, 1994), the choline-binding autolysins (Shah and Russell, 2002), and the toxins A and B of *Clostidium difficile* (Wren *et al.*, 1991) which are also characterized by conserved aromatic amino acids (tyrosine, phenylalanine) and glycine.

Another classification of the repeating unit of the C-terminal domain is YG repeat based on a multiple duplication hypothesis (Giffard and Jacques, 1994). They proposed that the repeat sequences evolve in response to selection for an increase in the number of copies of a particular domain through multiple duplication events occurring at different times (Figure 21). A following consensus sequences is NDGYYFxxxGxxH°x(G/N)xH°H°H° (x, non-conserved amino acid residue; H°, hydrophobic amino acid residue). The YG repeat can be found in both the A-D and N repeated and the outside sequences. Similar elements were also conserved in the cell wall binding domain of the *Clostidium difficile* toxins and the PspA protein of *Streptococcus pneumoniae*, suggesting that similar selective pressures had also been imposed on these sequences (Giffard and Jacques, 1994). The distribution of the YG repeat and their relationship with A-D repeat system was defined in Figure 22.

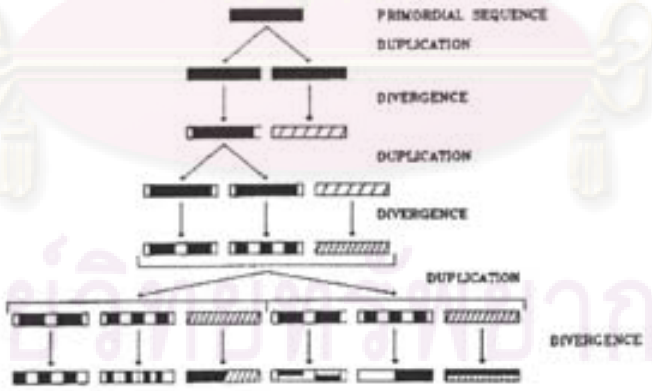


Figure 21. Diagrammatic representation of a hypothetical series of duplication and divergence events consistent with phylogenetic analyses that would give rise

to homologous regions coding for three sequential "YG" repeats within the GBDs of the GtfJ and GtfK of *S. salivarius* (Giffard and Jacques, 1994).

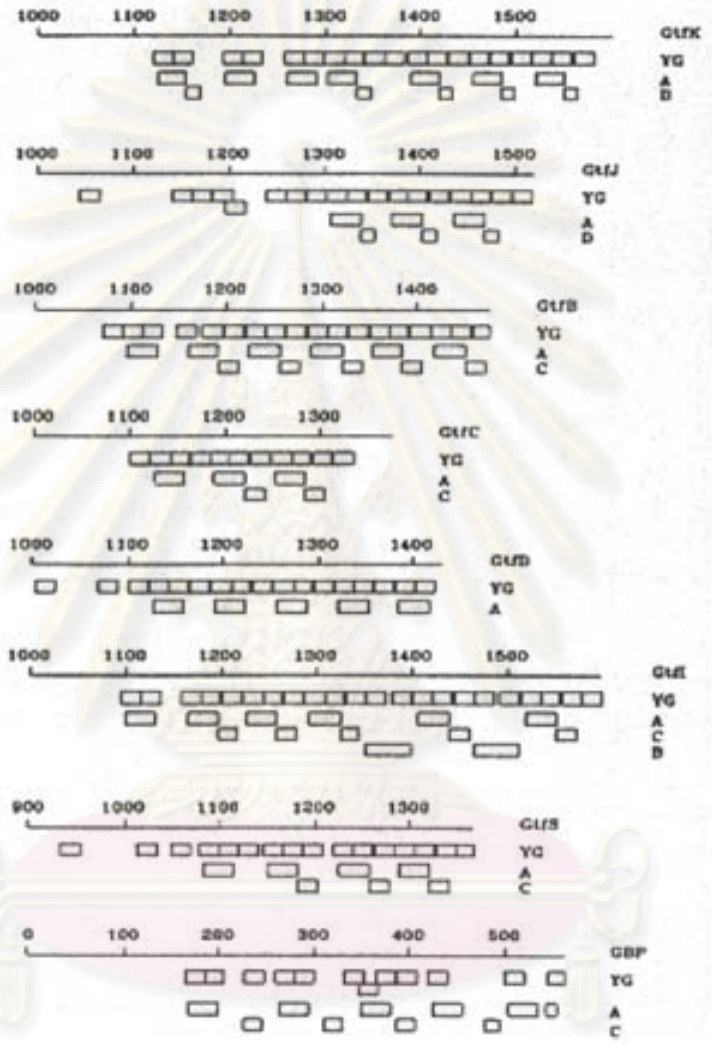


Figure 22. Location of the "YG" repeat in the GBDs of streptococcal Gtfs and the glucan binding protein (GBP) of *S. mutans*, and their spatial relationship with previously defined classes of repeats. Each scale represents the amino acid number of the GBD from an arbitrary beginning to the end of the polypeptide. The only exception is the GBP, where the entire polypeptide is shown (Giffard and Jacques, 1994).

2.2.2 Role in glucan binding

The C-terminal domain has been proposed to responsible for a glucan binding. Conserved amino acid residues may be involved in binding with glucosidic units. The clustered aromatic residues (tyrosine, tryptophane and phenylalanine) may stabilize the binding between sugar and protein by interacting with the sugar unit. The polar (lysine, glycine and phenylalanine) or acid (aspartic acid) residues may allow hydrogen bond forming with hydroxyl residues of sugar (Monchois *et al.*, 1999). Amino acids like lysine, glycine, asparagine, or serine are able to introduce flexibility into the protein structure and may allow the glucosyl residue to be correctly orientated to the binding site (Lemieux, 1989).

The secondary structure prediction for repeated units carried out by von Eichel-Striber and colleagues (von Eichel-Striber *et al.*, 1992) suggested that direct repeating units might be the structure of a functional binding pocket. An antiserum directed against a peptide TGAQTIKGQKLYFKANGQQVKG present in the C-terminal domain of *S. downei* Mfe 28 GTF-I inhibit glucansucrase activity (Smith *et al.*, 1993). A 17-kDa peptide coming from mild trypsic digestion of the C-terminal domain of *S. sobrinus* 6715 GTF-S was able to bind with glucan (Wong *et al.*, 1990).

The minimum number of these repeated units necessary to ensure glucan binding properties was also investigated. For GTF-S from *S. mutans*, the loss of only one C-terminal A repeated unit or of two N-terminal A repeated units is sufficient to suppress its binding capacity (Lis *et al.*, 1995). This difference tends to support the idea that all the units are not required in a similar way. Moreover, it appears that the number of required units is different for enzymes producing a soluble glucan than for those producing an insoluble one, the latter enzymes appearing less sensitive to deletions. (Monchois *et al.*, 1999) The presence of only the first two N-terminal repeated units is sufficient for GTF-I and GTF-Ia to bind glucan (Ferretti *et al.*, 1987; Kato and Kuramitsu, 1990)

More recently, the glucan binding assay of nested series of truncated forms of GTF-I or GBD of *S. downei* revealed that the binding capacity to dextran was proportional to the number of repeats (Figure 23; Shah and Russell, 2002). From this study, a minimal of four repeats was necessary to remain the glucan binding. The results obtained have been confirmed and extended the earlier reports that the C-terminal truncation of GTF results in loss of glucan binding property. Moreover, they proposed that the C-terminal glucan binding region functions as an independent domain as it can efficiently bind dextran as well as the full enzyme (Shah and Russell, 2002).

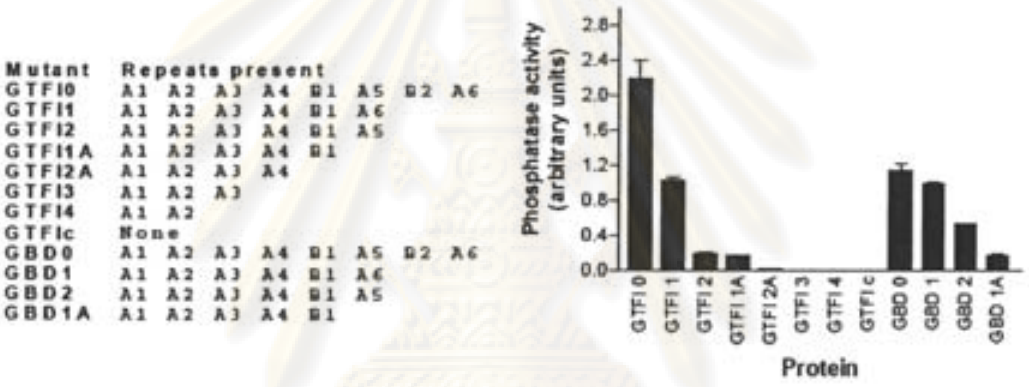


Figure 23. Glucan binding activity. The bar chart shows the binding activity of the di_erent deletion mutants as revealed by alkaline phosphatase activity measured 5 min after addition of substrate. The averaged results of three experiments are shown. The table shows the repeats possessed by each mutant (Shah and Russell, 2002).

Besides many investigations on a C-terminal domain of GTF from oral streptococci, there were only few reports concerning about those of DSR from *L. mesenteroides*. Abilities to bind dextran of different dextransucrases were undertaken by the use of biotin-dextran to detect proteins electroblotted onto nitrocellulose membrane (Shah *et al.*, 2004). Figure 24 showed that neither alternansucrase and dextransucrases of strain B-1355, nor the dextransucrase of B-512F, nor DSR-E of B-1299 bound dextran. The only *Leuconostoc* enzyme to bind

was DSR-B of B-1299, whereas strong binding was shown by the *S. mutan* GTF (Shah *et al.*, 2004). However, they were unsuccessful for expression of C-terminal domain of *Leuconostoc* dextransucrase in *E. coli*. Thus, this lack of binding of the *Leuconostoc* DSR has not been confirmed by glucan binding assay in a microplate.

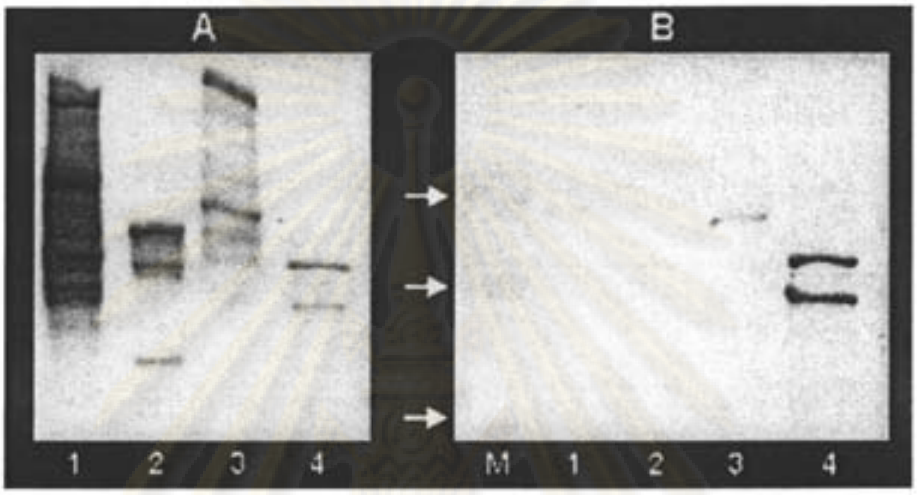


Figure 24. SDS-PAGE analysis of cell extracts from *L. mesenteroides* and *S. mutans* strains. (A) Zymogram of active glucansucrases after SDS-PAGE separation and incubation in sucrose. (B) The same samples electroblotted and developed with biotin-dextran to detect glucan binding proteins. Lane 1, *L. mesenteroides* NRRL B-1355; lane 2, *L. mesenteroides* NRRL B-512F; lane 3, *L. mesenteroides* NRRL B-1299; lane 4, *S. Mutans* UA159; lane M, molecular mass ladder of 250, 150, and 100 kDa (arrows) (Shah *et al.*, 2004).

2.2.3 Role in enzyme activity

GBD is also thought to be involved in the enzyme activity. The presence of the C-terminal glucan-binding domain seems to be necessary to keep an active enzyme. For GTF-D the loss of the C-terminal A repeated unit is sufficient to reduce the activity by of 90% (Lis *et al.*, 1995). The truncation of the last 85 amino acid residues of the C-terminal domain of DSR-S resulted in only about 25% of the initial activity being retained (Monchois *et al.*, 1998). On the other hand, the enzymes

synthesizing predominantly α-1,3 linkages are less affected by deletions at the carboxyl terminus than the enzyme synthesizing α-1,6 linkages (Remaud-Simeon *et al.*, 2000). Some contradictory results show that it is difficult to precisely attribute a clear role of GBD. However, the fact that with some deleted enzymes, hydrolytic activity remains, but glucan binding and synthesis properties disappears (Abo *et al.*, 1991, Kato *et al.*, 1990), shows that the C-terminal domain may be important for the polymer chain growth (Monchois *et al.*, 1999).

The C-terminal domain seems also be involved in glucan structure determination. Deletion of the three C-terminal units (A-C) of GTF-G led to an enzyme being obtained which produced an α(1-6) linked glucan whereas wild-type GTF-G produced a glucan composed by α(1-6) and α(1-3) glucosidic linkaged (Vickerman *et al.*, 1996).

Recent experiment on successive deletions of the C-terminal domain of DSR-S indicated that truncation of one N and C repeat affected 50% of the enzyme activity with no significant change in product formation (Figure 25; Moulis *et al.*, 2006a). On the contrary, deletion of an additional A repeat at the C-terminal end (DSR-S vardel Δ3) resulted in a dramatic decrease of initial activity by 99%. Moreover, shifting of dextran production from high molecular weight dextran to low molecular weight dextran was observed. They proposed that the A repeat may act as the anchor of the growing polymer to the enzyme surface, thus leading to efficient elongation of large size products (Moulis *et al.*, 2006a). It can be conclude that the C-terminal end has the role in elongation step and efficiency of polymer formation.

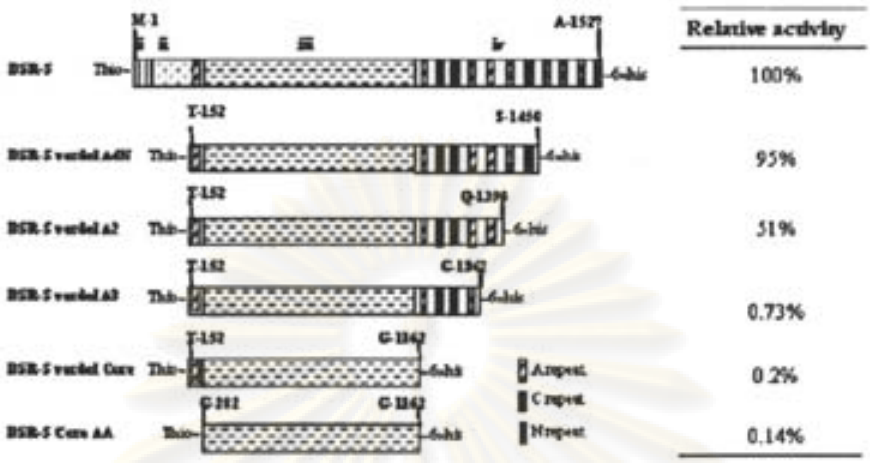


Figure 25. Schematic representation of the DSR-S-truncated variants and their relative activity. The four different domains: (i) signal peptide, (ii) variable region, (iii) catalytic domain, and (iv) C-terminal domain, and the repeated units A, C, and N (shaded boxes) are localized according to Monchois *et al.* (Monchois *et al.*, 1998; Moulis *et al.*, 2006a).

3. Directed evolution

In the past several years, directed evolution has become an alternative approach to rational design for the improvement of structural and functional properties of proteins or enzymes. The directed evolution does not rely on the understanding of enzyme structure-function relationship, but rely on the simple and powerful Darwinian principles of mutation and selection (Johannes and Zhao, 2006). According to the Darwin's own accounts "The key (of breeding) is man's power of accumulative selection: nature gives successive variations; man adds them up in certain directions useful to them" (Darwin, 1859 cited in Woodyer *et al.*, 2004). As shown in Figure 26, the directed evolution is starting by generation of genetic diversity from a target gene or a family of related gene. This step is typically accomplished by random mutagenesis and gene recombination. The mutated gene library is cloned into an expression vector and then transformed to host cells. The variant proteins are sorted for the desired character by screening or selection and those genes whose encoding

improved proteins, are used as the parents for the next round of evolution. This process of improvement is repeated until the goal is achieved (Woodyer *et al.*, 2004).

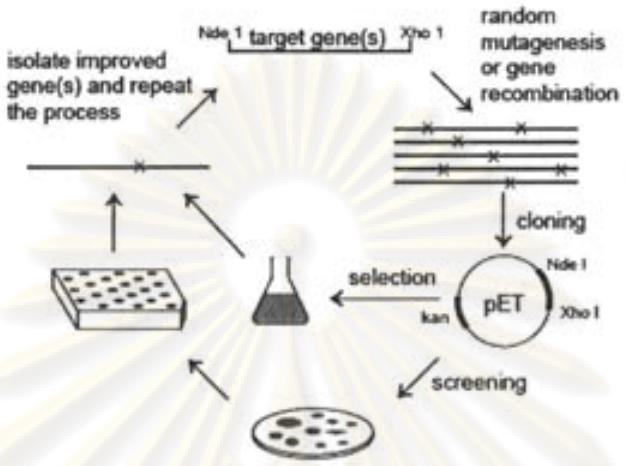


Figure 26. Schematic representation of directed evolution. As with Darwinian natural evolution, genetic diversity is introduced into a target gene through random mutagenesis or gene recombination. Functionally improved mutants are first identified by a selection or high throughput screening method and then used as the parents for next round of evolution. The process will be repeated until the goal is achieved (Woodyer *et al.*, 2004).

3.1 Generation of mutant library

Numerous artificial evolution techniques were developed in the past decade and have enhanced the process of gene manipulation (Ashie, 2003). For random mutagenesis, many effective methods such as UV radiation, mutator strains, and error-prone polymerase chain reaction (EP-PCR). Among these methods, EP-PCR is by far the most common approach (Woodyer *et al.*, 2004). In addition, gene recombination is alternative method for increasing genetic diversity. Compare to the random mutagenesis, the gene recombination can accumulate beneficial mutations while simultaneously removing deleterious mutations (Woodyer *et al.*, 2004). Examples of the gene recombination method are DNA shuffling, staggered extension process (StEP), random priming recombination, random chimeragenesis on transient

templates (RACHITT), and incremental truncation for the creation of hybrid enzymes (ITCHY), etc.

3.1.1 Error-prone PCR (Ep-PCR)

Ep-PCR utilizes a low-fidelity *Taq* polymerase to induce nucleotide substitutions in the gene during PCR amplification. The polymerization is performed at conditions which reduce the fidelity of the polymerase such as unequal concentration of nucleotide bases, inclusion of Mn^{2+}, and varying concentration of Mg^{2+} (Woodyer *et al.*, 2004). Then, the resulted variants are contained random point mutations along the gene. A wide range of substitutions can be obtained by this process, including AT to GC and GC to AT transitions, as well as AT to TA transversions (Fromant *et al.*, 1995). However, the *Taq* polymerase usually inserted nucleotide substitution with bias toward AT to GC changes (Cirino *et al.*, 2003). Nowadays, commercial enzyme with rather equal the nucleotide substitution is also available as Mutazyme II (Strategene). Because of the changes in the Ep-PCR being intragenomic, it is also referred as 'asexual PCR' (Ashie, 2003). One limitation of the Ep-PCR is only small size of fragements (less than 800 bp) can be processed (Ashie, 2003).

3.1.2 DNA shuffling

DNA shuffling is an in vitro recombination of homologous genes firstly invented by W.P.C. Stemmer (1994). The process involves the random fragmentation of parental sequences, assembly of two or more DNA segments into a full-length gene by repeated cycles of denaturation, annealing, and extension in the presence of polymerase (Figure 27). These hybrids are amplified by PCR and then a shuffling library is eventually generated (Ashie, 2003). Before the assembly, the segments are often subjected to random mutagenesis by error-prone PCR, random nucleotide insertion or other methods (Harayama, 1998). Unlike the error-prone PCR, the DNA shuffling is mentioned as "sexual PCR", since it allows mixing of genetic material between different parental sequences (Ahies, 2003).

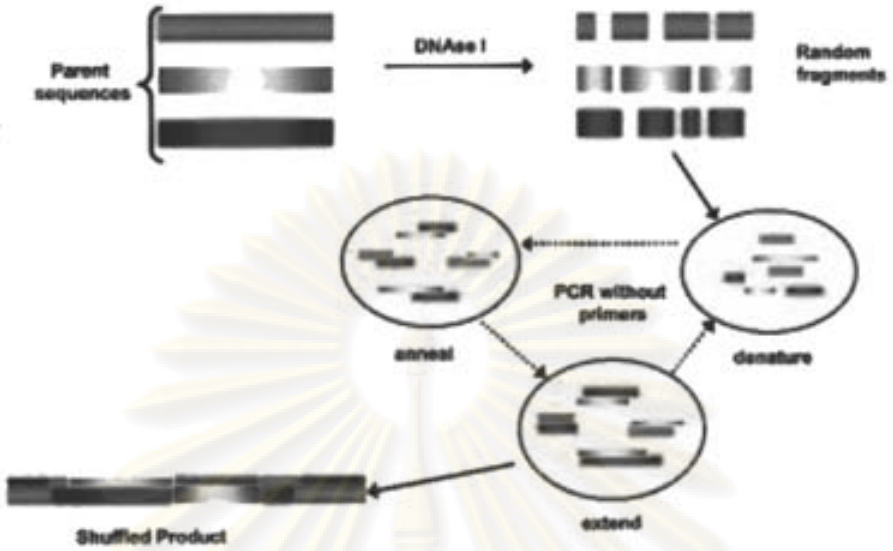


Figure 27. DNA Shuffling (Ashie, 2003).

DNA shuffling, however, usually introduces new point mutations at relatively high rate (0.7%). Even though, these point mutations may useful for increasing the diversity of some applications, they are problematic for others. Much lower mutation rates are desired, for example, during in the in vitro evolution of long genes or whole operon, or during recombination of beneficial mutations already identified previously, or Family shuffling with ssDNA (Zhao and Arnold, 1997). The fidelity of DNA shuffling can enhance by addition of Mn^{2+} instead of Mg^{2+} during the DNase I digestion (Lorimer and Pastan, 1995), the choice of DNA polymerase used during gene assembly, and preparation of DNA by plasmid digestion (Zhao and Arnold, 1997).

3.1.3 Random-priming recombination (RPR)

An alternative method of recombination named random-priming recombination was developed by Shao *et al.* (1998). In this technique, random sequence primers are used to create several short DNA fragments complementary to different sections of the template sequences (Figure 28). Little

point mutations can be occurred owing to base misincorporation and mispriming. Then, these short DNA fragments can prime one another base homology, and be recombined and reassembled into full-length genes by thermocycling in the presence of a thermostable polymerase. The conventional PCR is performed for further amplification of the full-length genes. The following steps are cloning, transformation, expression and then screening or selection for desired protein.

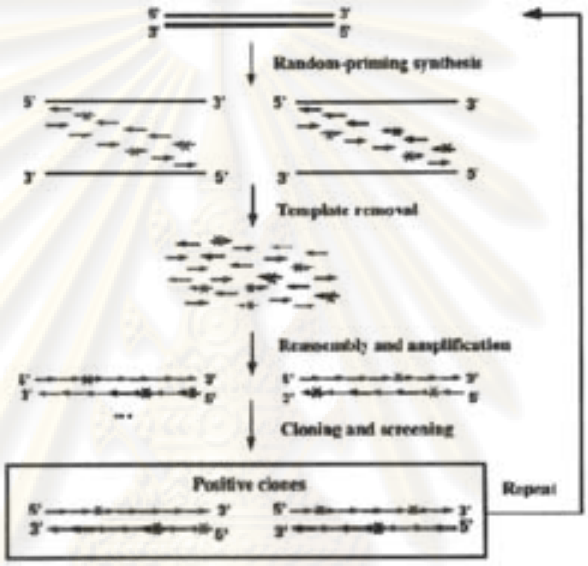


Figure 28. Random-priming in vitro recombination (RPR). (i) Synthesis of short single-stranded DNA fragments from random-sequence primers. X's indicate newly introduced mutations. (ii) Template removal. (iii) Reassembly and amplification. (iv) Cloning and screening (or selection). (v) Cycle is repeated until desired functional improvement is achieved (Shao *et al.*, 1998).

Advantages of RPR over DNA shuffling are: (a) the RPR can directly use single-stranded polynucleotide templates without an intermediated step for synthesizing the whole second strand. Mutations and/or crossovers can be introduced by using *Taq* polymerases. (b) No digestion by DNase I required, so it is not necessary to remove the DNase I before fragment reassembly into full-length sequence. (c) Using synthetic random primer is uniform in their length and

distribution resulting in a low sequence bias. (d) DNA fragment as small as 200 bp can be primed equally as well as large DNA molecules (Shao *et al.*, 1998).

3.1.4 Staggered extension process (StEP)

The staggered extension process (StEP) is based on priming the template sequences followed by repeated cycles of denaturation and extremely shorted annealing/extension step of PCR (Zhao *et al.*, 1998). As illustrated in Figure 28, the growing fragment anneals to different templates according to sequence complementary and extends continually in each cycle. The switching template will be occurred in each short cycle until full-length sequences formed. When the product yield is low, the full-length chimeric genes can be amplified by a conventional PCR. Compared to other recombination techniques, the StEP is much simpler and less labor intensive and can be performed using a pair of flanking primers in a single PCR tube (Zhao, 2004).

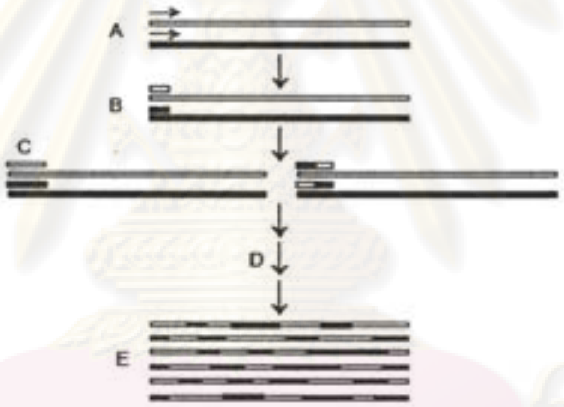


Figure 29. Schematic representation of the StEP in vitro recombination method. Only one primer and single strands from two parental genes (templates) are shown. (A) Denatured template genes are primed with one defined primer. (B) Short fragments are produced by brief polymerase-catalyzed primer extension. (C) Through another cycle of StEP, fragments randomly prime the templates (template switching) and extended further. (D) This process is repeated until full-length genes are produced. (E) Full-length genes are purified and (optionally) amplified in a PCR reaction with external primers. (Woodyer *et al.*, 2004).

3.1.5 Incremental truncation for the creation of hybrid enzymes (ITCHY)

A combinatorial approach to generate fusion libraries between two gene fragments called incremental truncation for the creation of hybrid enzymes (ITCHY) with independent DNA homology was first described by Ostermeier *et al.* (1999). The ITCHY does not rely on the parental sequence homology to create crossovers, but the truncation libraries are generated by digestion of the parental genes with exonuclease III under controlled conditions (Lutz *et al.*, 2001). The process involves protecting one end of a linear DNA containing the gene of interest, while the other end is made susceptible to the exonuclease III digestion (Figure 30; Ashie, 2003). The digestion is preceded at very low rate to produce different truncated forms in the library. The single-stranded DNA was removed by S1 nuclease digestion. The following steps are ligation, transformation, and expression of truncated variants. This approach allows the more diverse set of functional fusion created than DNA shuffling and is able to identify functional structural motifs such as binding and catalytic domains of enzymes (Ostermeier *et al.*, 1999; Ashie, 2003).

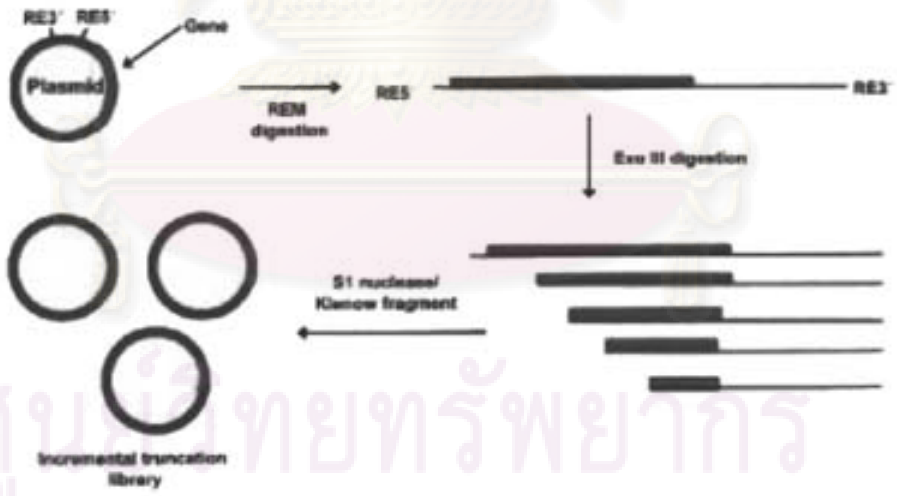


Figure 30. Incremental truncation (Ashie, 2003).

3.1.6 Random chimeragenesis on transient template (RACHITT)

The random chimeragenesis on transient template (RACHITT) is a gene family shuffling technique that aligns donor fragments on a full-length template (Coco *et al.*, 2001). From the Figure 31, one of parent genes is used as a transient scaffold where complementary fragments of the other parents anneal. Unhybridized fragment termini are trimed by endonuclease activity of *Taq* DNA polymerase and then gaps are filled by *Pfu* DNA polymerase. These trimming reactions have a potential to reduce the size of the shuffled fragments and increase the number of recombination per gene (Farinas *et al.*, 2001). In this method, no thermocycling, strand switching, or staggered extension is required, but rather the trimming, gap filling, and ligation of parental gene fragments hybridized on a transient DNA template. Moreover, it can be evolved clones from 4 to 14 parents simultaneously, including mixed, uncharacterized gene and/or significantly more divergent parental genes (Coco *et al.*, 2001). The exploitation of a bottom-strand template from only one parent and only top-strand fragments of other parents can prevents the parental fragment from reannealing to their own complementary stand (Coco *et al.*, 2001). However, the choice of the scaffold will probably influence the outcome of the experiment, as sequences that are similar to the scaffold will be incorporated more frequently into the chimeras (Farinas *et al.*, 2001).

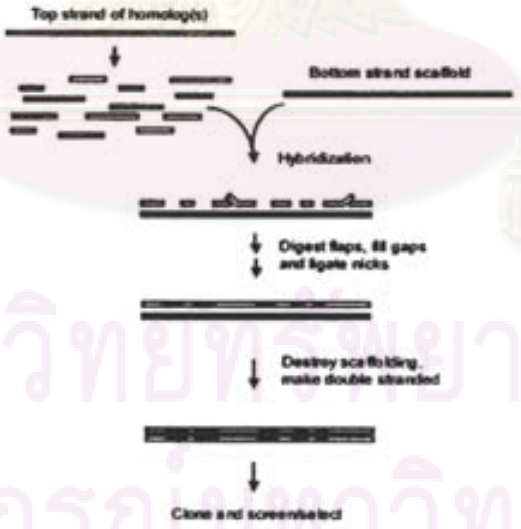


Figure 31. Random chimeragenesis on transient templates (RACHITT) (Coco *et al.*, 2001).

3.2 Screening or selection of the mutant for desired properties

One of the most critical steps of directed evolution approach is finding the desired mutants by screening or selection of the libraries. The screening is the process that all members in the library are assayed individually by using biochemical or biophysical analysis (Woodyer *et al.*, 2004). This process can be performed in either agar plates or microtitre plates (96- or 384-well plates). For the agar plate, the change of color can be detected by direct visualization or by digital image analysis. On the microplate, most screens involve transferring individual colonies to the microplate containing culture medium, growing the cells until stationary phase, and inducing protein expression. An additional cell lysis is carried out, when the desired protein is intracellular expressed to release cellular contents. Eventually, enzymatic activity is assayed with a microplate reader spectrophotometer (Salazar and Sun, 2003).

In a selection, the target protein is linked to the survival or growth of the host organisms where the desired property will grow or grow faster. The selection is often achieved by genetic complementation of host organisms that are deficient in a certain pathway or activity. Using selection, large libraries of protein variants can be performed and the size of the library is limited only by the cell transformation efficiency. Unfortunately, it is very difficult to couple the selected enzymes or proteins to the growth or survival of the microorganism. Moreover, host organisms often create solutions that not related to the target enzyme feature as complexity of genetic regulatory network. Thus, extra care must be taken to ensure the positives are indeed the result of the mutations in the target enzyme (Woodyer *et al.*, 2004).

Another consideration for the screening/selection of desired protein is the number of mutants capable to screen. Designing the experiment for high-throughput variants with a short time period is very challenge. The often screening limits the number of library members approximately 10^4-10^6, even with the use of robotic technology (Woodyer *et al.*, 2004). Therefore, it is beneficial to develop and validate the effective screening or selection process for each desired characteristic. For example, the screening for protein solubility and heterologous protein expression can be done in two ways: screening/selecting for the activity of a fusion partner or a

reporter protein, or screening/selecting for the protein's own function (Roodveldt *et al.*, 2005). Using a C-terminal fusion green fluorescent protein (GFP) as solubility reporter was applied to the screening of TEV protease (van den Berg *et al.*, 2006). In this method, high level of soluble expressed protein resulted in high fluorescent intensities. Moreover, dot blot technique has been successive applied to anti-His antibody to detect the expressed protein. This antibody was conjugated with enzyme producing signal such as anti-His-HRP (Nguyen *et al.*, 2006; Vincentelli *et al.*, 2005).

3.3 Recent applications of directed evolution

Many directed evolution approaches have been established over the part decade. It has been successfully to engineer various proteins for improvement of their natural characteristics such as enzyme activity, substrate selectivity, enzyme secretion and solubility.

3.3.1 Enzyme activity

The activity of enzyme is naturally too low for commercial and therapeutic applications; therefore it is necessary to improve enzyme activity by directed evolution (Johannes and Zhao, 2006). DNA shuffling has been successively applied to the directed evolution of N-acetyltransferase (GAT) that the variant has the activity over 10000 folds compared to the wild-type enzyme after 11 rounds of the shuffling (Castle *et al.*, 2004). In addition, different mutagenesis techniques such as epPCR, DNA shuffling, and DNA family shuffling have also been used to enhance activity of lipase enzyme (Fujii *et al.*, 2005; Qian and Lutz, 2005; Suen *et al.*, 2004). Another recent target of directed evolution is a barley α-amylase. Engineering of the α-amylase by three rounds of epPCR and DNA shuffling was resulted in a 1,000 folds of total activity compared to that of the wild-type enzyme (Wong *et al.*, 2004).

3.3.2 Stability

Many industrial applications require enzymes that can function at high temperature or in non-natural environments such as in organic solvent (Johanne and Zhao, 2006). Combination of epPCR and DNA shuffling was found to increase thermostability of β-glucoronidase in *E. coli* (Flores and Ellington, 2002).

For N-carbamyl-D-amino acid amidohydrolase, the DNA shuffling provided an increase in both thermostability and oxidative stability (Oh *et al.*, 2002).

3.3.3 Solubility and heterologous expression

High level production of stable and functional proteins remains the bottleneck of many scientific researches including structural determination and protein-protein interaction. One example of using directed evolution for increasing the protein solubility is that the solubility of phosphotriesterase from *Pseudomonas diminuta* was improved 20-fold by epPCR and DNA shuffling (Roodveldt and Tawfik, 2005). In addition, Waldo *et al.* (1999) evolved the highly insoluble nucleotide diphosphate kinase from *Pyrobaculum aerophilum* into a 90% soluble by DNA shuffling and active variant that enabled the determination of its crystal structure (Pedelacq *et al.*, 2002). Using epPCR and DNA shuffling was reported to increase the solubility of tobacco etch virus protease for 5 folds of the wild-type with retained activity (van den Berg *et al.*, 2006). Moreover, incremental random N-terminal deletion was enhanced the solubility of 14 from 19 "hard to express" mammalian proteins (Cornvik *et al.*, 2006).

CHAPTER III
MATERIALS AND METHODS

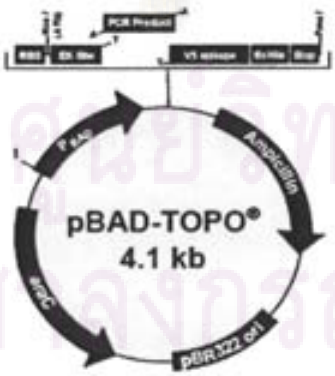
1. Bacterial strain, vectors, and culture conditions

1.1 Bacterial strain

Escherichia coli TOP10 cells (Invitrogen) were used as cloning and expression host in all following experiments. Genotype of the TOP10 strain is F mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 recA1 deoR araD139 Δ0(ara-leu)7697 galU galK rpsL (Str^R) endA1 nupG. The strain is capable to transport L-arabinose, but could not metabolize it. This is compatible to pBAD expression system where enable recombinant protein production in this study.

1.2 Vectors

Table 9. Vector characteristics and source.

Plasmid	Characteristic	Reference
pBAD/TOPO thiofusion 	Resistance to ampicillin, Expression in *E. coli* is driven by the *ara*BAD promoter; Recombinant proteins are expressed as fusions to His-Patch thioredoxin for high-level expression and simple purification.	Invitrogen
pBAD/TOPO TA 	Resistance to ampicillin, Expression in *E. coli* is driven by the *ara*BAD promoter, Providing polyhistidine tag at C-terminus of desired proteins.	Invitrogen

1.3 Culture conditions

Media formula

All bacterial cultures were typically grown in LB (Luria-Bertani) medium or 2XYT rich medium. In case of expression the target protein, the medium was supplemented with 100 mM Tris HCl to pH 6.4. For *E. coli* carrying ampicillin resistant plasmid, 100 μg/ml of filtered-sterilized ampicillin was finally added before inoculation. Solid media were prepared for isolation and propagation of *E. coli* colonies.

LB

Tryptone	10 g/l
Yeast extract	5 g/l
NaCl	10 g/l
Agar (for solid media)	15-20 g/l

2XYT

Tryptone	16 g/l
Yeast extract	10 g/l
NaCl	5 g/l
Agar (for solid media)	15-20 g/l

SOC

Tryptone	10 g/l
Yeast extract	5 g/l
NaCl	5 g/l
$MgSO_4$	10mM
$MgCl_2$	10mM

Adjust to pH 7.0 with NaOH. Autoclave.

Add filter-sterilized glucose to a final concentration of 20mM.

Growth condition

Bacterial cells from glycerol stock or from a single colony were inoculated into vials containing LB-ampicillin medium. The cultivation was preceded at 37 °C overnight with vigorously shaking. These bacterial cultures were generally used as sources for plasmid extraction or as a starter culture for recombinant protein expression.

1.4 Bacterial storage

The bacterium was grown at 37 °C overnight with vigorous shaking. All strains were stored at -20°C and -80°C in 15% (v/v) glycerol or 4 °C in solid media.

1.5 Preparation of competent cells

In particular case, competent *E. coli* TOP10 cells were manually prepared instead of commercial chemically TOP10 competent cells (Invitrogen). The competent cells were prepared by DMSO method as described by Inoue *et al.* (1990). From glycerol stock, starter culture was prepared in 2 ml of LB and growed at 37 °C overnight. Then, Erlenmeyer flask containing 100 ml of SOC was inoculated by the overnight culture to an OD_{600} of 0.025 and incubated at 18 °C until the OD_{600} = 0.6. The culture was placed on ice for 10 min and centrifuged at 3,800 rpm for 10 min at 4 °C. The pellet was resuspended gently in 30 ml of cold TB and incubated on ice for 15 min. After that, the cell solution was centrifuged again at 3,800 rpm for 10 min at 4 °C. The pellet was resuspended gently in 8 ml of cold TB and then added 560 µl of DMSO. That solution was placed on ice for 10 min. Finally, the cells were aliquoted into 50-100 µl and freezed in liquid nitrogen. The competent cells were stored at -80 °C.

TB

Pipes	10 mM
$CaCl_2$	15 mM
KCl	250 mM
$MnCl_2$	55 mM

Adjust pH to 6.7 with KOH prior to adding $MnCl_2$.

Sterilizationn using 0.2 μm filters.

2. Molecular biology techniques

2.1 Polymerase Chain Reaction (PCR)

2.1.1 Primers

Table 10. Primer sequences.

Primer name	Sequence (5' to 3')
ForGBD	3568-ACCATGGATAATAACTATTACTATTTTGAT-3591[1]
RevGBD	4581-GGCTGACACAGCATTTCCATTATTATCAAA-4552
GBD1F	3577-TATTACTATTTTGATAAAACAGGTCATT-3604
GBD3R	3966-TTGATTCTTACTATTTTGCACATAAC-3941
GBD5F	3967-TGGTTCTATTTTGATGGTAATG-3988
GBD7R	4269-TTGCAATACAAACCCTGTTTTCGCACG-4243
GBD9F	4270-GATGGTGTACTAAGATACTTCGATCA-4295
GBD10F	4063-GGTGAATTCATTGATGCAGACGGGGATA-4090
GBD11F	4189-AATCAGTATTATCAATTAGCAGATGGTAAA-4218
GBD12R	4545-AACCTTACCCTGAGCACTTATTAAAG-4519
GBD13R	4431-AGCATCTGTTAAATACCATTTACCTTGATA-4402
GBD14F	4090-ACTTTCTATACGAGTGCCACTGATGGTCGC-4119
GBD15R	4326-AATGATAGCATCTTTCACTTGCTCACCGTT-4297
GBD16R	4389-ATTTTTTACAGCGACACCTTGTGTTGCATT-4360
GBD17F	4195-TATTATCAATTAGCAGATGGTAAATATATG-4224
GBD18R	4470-GTCGTCAACTGCTTTAAAACCTTTGATAAG-4441
NdeI_gbd7F	ACCATGG**CTGCAG**CTGTCT*GAATTC*TGAATCAGTATTATCAATTAGCAG
CdeI_gbd7F	ACCATGGATAATCAGTATTATCAATTAGCA
CdeI_gbd7R	**CTGCAG**GTTAGT*GAATTC*GGCTGACACAGCATTTCCATTAT
CodLeArgF	CAAGATGGTGTACTGCGCTACTTC
CodLeArgR	GTTTTGATCGAAGTAGCGCAGTACACC
CodIleF	ACAAAAGATAGTGCTTTAATTAGTGCTCAG
CodIleR	CTTACCCTGAGCACTAATTAAAGC
pBADF	ATGCCATAGCATTTTTATCC
Trx F	TTCCTCGACGCTAACCTG
pBADR	GATTTAATCTGTATCAGG

*Nco*I site = CCATGG; *Pst*I site = **CTGCAG**; *Eco*RI site = *GAATTC*

2.1.2 Standard PCR

Reaction mixtures

Plasmid DNA	1	µl
Forward primer (10 µM)	1	µl
Reverse primer (10 µM)	1	µl
dNTPs (2.5 mM each)	4	µl
10x enzyme buffer	5	µl
Enzyme	0.5 µl for Expand high fidelity enzyme mix (Roche) or 1 µl for Dynazyme EXT (Finnzymes)	
Sterile water	x	µl to final volume 50 µl

Thermal cycling

Initial denaturation	94 °C	5 min	
Denaturation	94 °C	5 s	x30 cycles
Annealing	55 °C	15 s	
Elongation	72 °C	1 min	
Final elongation	72 °C	7 min	
Cooling	15 °C	∞	

2.1.3 Inverse PCR

Reaction mixtures

5x Phusion GC buffer	10	µl
Plasmid DNA	7	µl
Forward primer (10 µM)	1	µl

Reverse primer (10 μM)	1 μl
dNTPs (2.5 mM each)	4 μl
DMSO	1.5 μl
Phusion DNA polymerase (Finnzyme)	0.5 μl
Sterile water	25 μl

Thermal cycling

Initial denaturation	98 °C	1 min	
Denaturation	98 °C	10 s	x30 cycles
Annealing	56 or 60 °C	15 s	
Elongation	72 °C	3 min	
Final elongation	72 °C	10 min	
Cooling	15 °C	∞	

2.1.4 Colony PCR

Reaction mixtures

Forward primer (10 μM)	1 μl
Reverse primer (10 μM)	1 μl
dNTPs (2.5 mM each)	4 μl
10x enzyme buffer	5 μl
Taq DNA polymerase (5U/μl)(Biolabs)	0.5 μl
Sterile water	38.5 μl

Single colony was picked to the reaction tube by yellow pipette tip before adding *Taq* DNA polymerase.

Thermal cycling

Initial denaturation	94 °C	10 min	
Denaturation	94 °C	5 s	x30 cycles
Annealing	40 °C	15 s	
Elongation	72 °C	1 min	
Final elongation	72 °C	7 min	
Cooling	15 °C	∞	

2.2 Plasmid extraction

Plasmid extraction was performed using QIAprep Miniprep kit (Qiagen) or QIA Maxi prep kit (Qiagen) based on alkaline lysis method followed by DNA adsorption onto silica membrane in the presence of high salt. The plasmids were sometimes purified by alkaline lysis minipreparation (Ausubel *et al.*, 1988).

QIAprep Miniprep kit (Qiagen)

An 1 to 5 ml overnight culture of *E. coli* in LB medium was centrifuged. The pellet was resuspended in 250 µl Buffer P1. Buffer P2 was added and gently invert the tube 4-6 times to mix. 350 µl Buffer N3 was added and inverts the tube immediately but gently 4-6 times. The tube was centrifuged for 10 min at 13,000 rpm. The supernatant was applied to the QIAprep spin column and centrifuged for 1 min. The flow-through was discarded. The column was washed by adding 0.75 ml Buffer PE and centrifuged for 1 min. The flow-through was discarded and then the column was centrifuged for an additional 1 min to remove residual wash buffer. The QIAprep column was placed in a clean 1.5 ml microcentrifuge tube. To elute DNA, 30-50 µl Buffer EB (10 mM Tris-Cl, pH 8.5) was added to the center of each QIAprep spin column, stands for 1 min, and centrifuged for 1 min.

QIA Maxi prep kit (Qiagen)

Plasmid purification using QIAGEN plasmid maxi kit was applied for large scale plasmid preparation. This protocol is based on a modified alkaline lysis procedure, followed by binding of plasmid DNA to anion-exchange resin. Contaminants such as RNA, proteins, dye, and nucleotides are removed by a medium salt wash. Plasmid DNA is eluted in a high-salt buffer and then concentrated and desalted by isopropanol precipitation.

An 100 – 300 ml of overnight culture in LB was centrifuged at 6000 *xg* for 15 min at 4 °C. The pellet was resuspended in 10 ml of buffer P1 which was previously added RNase A and pipetted up and down until no cell clumps remain. Then, 10 ml of buffer P2 was added and mix thoroughly by inverting the tube 4-6 times. The tube was incubated for 5 min at room temperature. 10 ml of pre-chilled buffer P3 was added, mix immediately by vigorously inverting 4 – 6 times, and incubated on ice for 20 min. The tube was centrifuge at 20,000 *xg* for 30 min at 4 °C. Supernatant was placed to a new tube and centrifuge again. The soluble fraction was then transferred to QIAGEN-tip 500 which was equilibrated by applying 10 ml of buffer QBT and allowed it to enter the resin by gravity flow. The column was washed by adding 2x30 ml of buffer QC. The plasmid DNA was eluted with 15 ml of buffer QF and precipitated DNA by applying 10.5 ml of isopropanol. The mixture was mixed and centrifuged immediately at 15,000 *xg* for 30 min at 4 °C. The supernatant was carefully decanted and the pellet was washed with 70% ethanol. The centrifugation step was carried out again at 15,000 *xg* for 10 min and the supernatant was gently decanted. Finally, the plasmid solution was obtained by air-dry the pellet and then dissolved in 1 to 2 ml of TE buffer.

Alkaline lysis minipreparation (Ausubel *et al.*, 1988)

An 1 to 5 ml overnight culture of *E. coli* in LB medium was centrifuged at maximum speed for 20s. Supernatant was then discarded and pellet was resuspended in 250 µl of GTE solution. Let the tube stand for 5 min at room temperature. 250 µl of NaOH/SDS solution was added and mixed by tapping the tube with fingers and incubated on ice for 5 min. 350 µl of potassium acetate solution was added and vortex

the tube for 2 s. The tube was incubated on ice for 5 min. Centrifuge for 10 min and the supernatant was then transfer to a new tube. 0.8 ml of 95% ethanol was added and let the tube stand for 2 min at room temperature. Centrifuge the tube again and discard supernatant. The pellet was washed by 1 ml of 70% ethanol. Air dries the pellet. Redissolve the pellet with 30 µl TE buffer pH 8.0. Plasmid DNA was stored at 4 °C or -20 °C.

Glucose/Tris/EDTA (GTE solution)

Glucose	50 mM
Tris-Cl pH 8.0	25 mM
EDTA	10 mM

Autoclaved and stored at 4 °C

Added 10 µg/ml RNase A.

NaOH/SDS solution

NaOH	0.2 M
SDS	1% (w/v)

Freshly prepared from 10 M NaOH and 10% SDS stock.

5 M potassium acetate solution

Glacial acetic acid	29.5 ml

KOH pellets to pH 4.8

H_2O to 100 ml. Stored at room temperature.

2.3 DNA purification

Both DNA in aqueous solution and agarose gel were isolated using QIAquick Gel Extraction Kit (Qiagen). This system employ silica-gel membrane and buffer to remove comtaminants on column as described above.

2.4 DNA visualization

DNA fragment or plasmid was migrated on agarose gel in 0.5x TAE buffer. Percentage of agarose gel was varied depend on size of target DNA from 0.8 % to 3%. DNA was mixed with loading dye and run together with the molecular weight marker. The electrophoresis was carried out at 100-135V. The gel was immersed in ethidium bromide solution for 15-20 min. To visualize DNA, the gel was exposed under UV light.

50x TAE		**10x loading dye**	
Tris base	242 g/l	Glycerol	50%
EDTA (0.5M, pH 8.0)	100 ml	Bromophenol blue	0.2%
Acetic acid	57.1 ml	Xylene cyanol FF	0.2%

2.5 Restriction enzyme digestion

Different restriction enzymes with various recognition sites were purchased from New England Biolabs. Typically, the restriction enzyme digestion was used for analysis of positive clone. The reaction mixture was prepared in a microcentrifuge tube as described below to a final volume of 30 µl. Then, the tube was incubated at optimal temperature of each enzyme for 1 h. Most of restriction enzymes digested specifically at 37 °C. The digestion was finished by heat inactivation at 65-80 °C for 20 min and subsequently visualized the restriction fragments on agarose gel. Moreover, the restriction enzyme digestion was applied for DNA preparation in further experiment such as cloning. In this case, the total volume and reaction time may increase to 100-500 ml and 3 h respectively.

Digestion mixture

Combine the following in order;

15.5µl sterile H_2O

3 µl 10x Buffer

1 µl BSA (for some restriction enzymes)

10 µl DNA

0.5 µl Enzyme

2.6 Cloning and transformation

Target DNA were cloned into pBAD/TOPO Thiofusion or pBAD/TOPO TA vector to yield fusion proteins with a C-terminal polyhistidine tag and for the former with a N-terminal Thioredoxin tag. These vectors employ topoisomerase I properties to bind the duplex DNA at specific site and cleave the phosphodiester backbone after 5' CCCTT in one stand. Purified DNA fragments were amplified by standard PCR

and followed by DNA extraction on agarose gel using QIAquick gel extraction kit. 0.5 to 4 µl of target DNA was ligated to 1 µl of linearized vector in optimal salt condition. Sterile water was added to a total ligation reaction of 6 µl. The reaction was mixed gently and incubated at room temperature for 30 min. 3 µl of the cloning reaction was pipetted to a vial of chemically competent cells and mixed gently. The vial was placed on ice for 30 min. The cells were heat-shocked for 30s at 42 °C and immediately transfer the tube on ice. 250 µl of SOC medium was applied to the tube. The cells were re-grown at 37 °C for 1 h 30 min with horizontally shaking at 200 rpm. The transformants were spreaded onto room temperature LBA agar plate and incubated at 37 °C overnight. The positive colonies were analyzed by restriction enzyme and DNA sequencing to be sure for orientation as well as insertion.

2.7 DNA Sequencing

All recombinant constructs in this study were sequenced by Macrogen (Korea), Millegen (France), or Genome Express (France).

2.8 Mutagenesis

2.8.1 Unidirectional deletion using Exonuclease III

In order to reduce a size of GBD-7 peptide, nested truncation libraries of GBD-7 was generated using Erase-a-Base system (Promega) which relies on digestion of Exonuclease III (Exo III) from a chosen DNA end. Exo III was able to digested 5' overhang or blunt end DNA but not able to cut 3' overhang end DNA. From this property, the unidirectional deletion variants could be generated by including restriction site of susceptible 5' overhang adjacent to the target gene side and resistant 3' overhang near to protective side of template plasmid (Figure 32).

Both N- and C- terminal directional template have been previously constructed using specific primers in order to position *Pst*I (3' overhang) and *Eco*RI (5' overhang) recognition site. Template plasmid was prepared using QIA Maxi prep kit and then sequentially incubated with *Pst*I and *Eco*RI at 37 °C for 3 h each. 5 µg of double-restricted target DNA was treated with 300 U of Exo III at 10 °C for 7.5, 15,

and 30 min. Subsequently, samples were added to tubes containing 2.25 U of S1 nuclease and stand at room temperature for 30 min for removal of remaining single-stranded regions. Then, the reaction was neutralized by adding S1 stop buffer and heat inactivated the S1 nuclease at 70 °C for 10 min. The DNA samples were precipitated by adding 0.3 volume of 7.5 M ammonium acetate and 2 volume of 100% ethanol. The tubes were well mixed and incubated at -20 °C overnight.

The next day, the tubes were centrifuged at 12,000 *xg* for 5 min. The supernatants were carefully discarded and the pellets were washed with 0.5 µl of 70% ethanol. The tubes were centrifuged again. The pellets were dry under vacuum for 15-30 min and resuspended with appropriate amount of TE buffer.

The samples were placed at 37 °C and added a Klenow reaction mixture containing 0.15 U of the enzyme to fill in the ends. The tubes were incubated for 3 min and then dNTP mix was added. The reactions were preceded for 5 min. Then, the samples were heated at 65 °C for inactivation of the Klenow. The clustered deletions of vector were re-ligated by adding ligase mix containing 0.2 U of T4 DNA ligase and incubated at room temperature for 1 h. 5 µl of the ligation product was transformed to competent *E. coli* TOP10 cells.

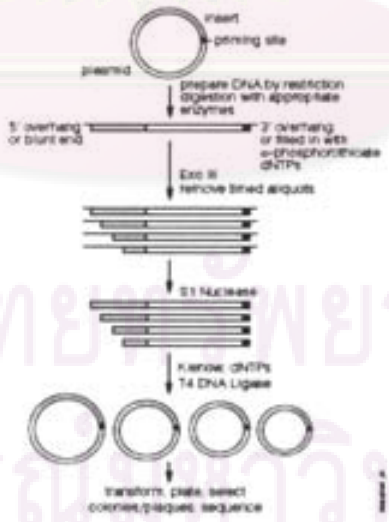


Figure 32. Schematic diagram of the Erase-a-base protocol (Promega).

2.8.2 Site-directed mutagenesis by inverse PCR

In vitro site-directed mutagenesis was accomplished using inverse PCR. Two overlap primers containing a desired mutation were designed and used for synthesis of whole wild-type plasmid using a Phusion DNA polymerase at condition described in Protocol 2.1.3. After the PCR cycling, all PCR mixtures were treated with 20 U of *Dpn*I at 37 °C for 1 h in order to eliminate methylated parental plasmid. The mutated plasmids were discriminated on 0.8% agarose gel and purified using a column kit. The *Dpn*I -treated DNA was transformed into competent *E. coli* TOP10 cells. The positive clones were analyzed by restriction digestion and Sequencing.

3. Recombinant protein production and preparation

3.1 Protein expression

For protein expression, cultivations were performed as described before for recombinant DSR-S expression (Moulis *et al.*, 2006b) at 16, 23, or 30 °C in 2XYT-tris pH 6.4 with overnight preculture to an OD_{600} of 0.2. When the OD_{600} equal to 0.6, the recombinant proteins were induced by adding 0.02 % (w/v) filter-sterilized L-arabinose. Cells were harvested by centrifugation at 6,500 rpm for 15 min at 4 °C. Cells were concentrated to an $OD_{600\ nm}$ of 80 in sonication buffer and sonicated. Cell debris were eliminated by centrifugation at 13,000 rpm 15 min at 4 °C.

PBS buffer pH 7.2		**Sonication buffer**	
Potassium phosphate 50 mM		$CaCl_2$	0.05 g/l
NaCl	150 mM	PMSF	1mM
Adjust pH to 7.2 with 1 N NaOH		In PBS buffer pH 7.2	

3.2 Protein purification

3.2.1 Protein purification onto Ni-NTA matrix

For purification using His tag and Ni-NTA resin (Qiagen), one volume of Ni-NTA slurry was added to 4 volumes of GBD extracts for 2 h of binding at 4°C. Washes were performed with 4 volumes of washing buffer and repeated 3 times.

Proteins were eluted using 8 volumes of elution buffer. The fractions of interest were dialyzed overnight against PBS buffer pH 7.2. The protein content was checked by SDS-PAGE electrophoresis, and quantified by Microbradford assay with BSA as standard.

Washing buffer	**Elution buffer**
Imidazole 20 mM	imidazole 250 mM
NaCl 1 M	in sonication buffer
in sonication buffer	

3.2.2 Protein purification onto Cross-linked dextrans

Sephadex®, Sephacryl® and Superdex® supports were purchased from Amersham Biosciences. One volume of resin was rinsed with 10 volumes of washing buffer to remove ethanol traces and recovered by centrifugation (5,000xg, 2 min). Then, 10 volumes of GBD-7 extracts were added and let stand for 2 h at 4°C. In order to determine the best support to be used, GBD-7 extracts were previously purified by nickel-chelating affinity (protein content: 297 µg/ml). Crude extracts were used in the case of purification of the thioredoxin fused to GBD-7. Unbound proteins were removed by centrifugation (2,000xg, 2 min), and the gel was then washed four times with 10 volumes of washing buffer. GBD-7 tagged proteins were then eluted using 14 volumes of PBS pH 7.2 buffer containing 0.1 to 50 g/l of commercial dextrans ranging from 0.4 to 2,000 kDa (Sigma). Total protein content was measured using Microbradford assay and BSA as standard. Thioredoxin content was measured as described below.

Washing buffer	**Elution buffer**
NaCl 300 mM	Dextran
in sonication buffer	in PBS buffer pH 7.2

3.3 Protein concentrating

Protein extracts were concentrated using a 10 kDa cut-off of Microsep centrifugal device with a low protein binding Omega membrane (Pall Corporation). The spin column containing protein extracts was centrifuged at 5,000 rpm 4 °C until achieve a desired concentration of protein.

4. Protein electrophoresis

4.1 SDS-PAGE

We used NuPAGE electrophoresis system (Invitrogen) for separating a specific range of proteins. This system consists of NuPAGE Novex Bis-Tris pre-cast gels which can separate small to mid size of proteins depend on acrylamide concentration. Appropriate dilution of protein samples were prepared in NuPAGE LDS sample buffer (4x) and then heated at 70 °C for 15 min for denaturation of protein samples. After cool the sample on ice, the 10% or 12% Bis-Tris gel was used to reveal the protein migration by size in NuPAGE MOPS or MES SDS running buffer under denaturing condition. The electrophoresis was performed at 150V for 1 h 30 min. The precision plus protein standard (Bio-rad) comprising 250, 150, 100, 75, 50, 37, 25, 20, 15, and 10 kDa proteins was used to determine the protein size.

4.2 Native-PAGE

Native-PAGE was performed for affinity electrophoresis of GBD variants toward dextran. The native gels supplemented with dextran were prepared using a protocol derived from those of Kataeva *et al.* (2001) or Haas and Banas (2000). Home-made gels were prepared with 7.5 % (w/v) acrylamide in 1.5 M Tris/HCl buffer pH 8.8. Commercial dextran of 2.10^6 (Sigma) was added to the gel at a final concentration of 0.1 % (w/v) before polymerization, resulting in a polyacrylamide-polysaccharide matrix. This matrix retarded the migration of proteins able to bind to the dextran present in the gel. Compared to classical gels, the amounts of ammonium persulfate and TEMED were doubled to avoid incomplete gel polymerization. Purified protein samples were mixed with 2X Native sample buffer (Invitrogen) before loading. Separation was performed at 4 °C and 150 V for 3 h. Native gel

without dextran was performed simultaneously under the same conditions, and both gels were stained with Colloidal Blue. Molecular mass of protein under native gel was addressed by HMW calibration kit for native electrophoresis (Amersham), which corresponds to protein size of 669, 440, 232, 140, and 66 kDa.

4.3 Colloidal blue staining

A Colloidal blue staining kit (Invitrogen) was applied for detection in nanogram amounts of protein band. This method is based on the colloidal properties of Coomassie blue dyes which were created in aqueous or methanolic solutions containing inorganic acids and high salt concentration. The solution was freshly prepared prior to staining as manufacturer's recommendation. The gel was shaked in staining solution for 3-16 h. After staining solution was decanted, deionized water was applied in order to wash the gel until obtain clear background.

4.4 Silver nitrate staining

Silver staining method is a very sensitive for detection of proteins in nanogram level. We used the SilverXpress Silver Staining kit (Invitrogen) based on chemical reduction of the silver ions to metallic silver on a protein bands. This method utilizes ammonium hydroxide to form silver diamine complexes. Then, the bands were visualized by acidification of citric acid. Mark12 Unstained standard (Invitrogen) was applied as a protein standard with 1:20 dilution.

4.5 Western blot analysis

Transfer of proteins from the gel to nitrocellulose membrane was performed using XCell II Blot Module (Invitrogen). While the gel electrophoresis was running, blotting pads, filter papers, and membrane were soaked in 1x NuPAGE transfer buffer with 10% methanol. After the electrophoresis finished, the cassette containing blotting pads, filter papers, gel, membrane, filter papers, and blotting pads was sequentially positioned into a cathode (-) core of the blot module. The anode (+) core was placed on the top of the pads. Then, the transfer module was assembly and filled with the transfer buffer. The transfer was started at 30 V for 1 h.

Then, the membrane was stained by Ponceau's solution for temporally detection of protein bands. After the detection, the membrane was destained by distilled water for several times. The blot was blocked to prevent non-specific binding of antibody by gently shaking the membrane in blocking solution overnight. The membrane was probed by anti-his or anti-thio antibody solution for 90 min at room temperature. After incubation and washing, the second antibody, anti-mouse IgG antibody conjugate with alkaline phosphatase solution, was applied for 90 min at room temperature. For the detection of specific protein bands, BCIP/NBT liquid substrate was added. The purple color appearing on the band indicates his- or thio-tagged protein. The prestained precision plus protein standard (Bio-rad) was used to determine the protein size comprising 250, 150, 100, 75, 50, 37, 25, 20, 15, and 10 kDa.

PBS buffer pH 7.3

NaCl	1.4 M
KCl	27 mM
Na_2HPO_4	101 mM
KH_2PO_4	18 mM

Adjust pH to 7.3 with NaOH

Blocking solution

3% BSA solution or 5% non-fat dry milk in PBS buffer pH 7.3

Washing solution

Tween 20 0.05%

In PBS buffer pH 7.3

Antibody solution

Antibody 0.02%

BSA 0.2%

In PBS buffer pH 7.3

4.6 Gel storage

Gels and membranes were scanned following a high resolution. Then, gels were combined with pre-wet cellophane sheet and dry under vacuum using gel drying machine at 70 °C for 2 h.

5. Characterization of glucan-binding ability

5.1 Glucan binding assay

Glucan binding assays were performed using the method developed by Shah *et al.* (2002). 0.75 mg total proteins of *Thioredoxin*-GBD-*His* extracts were incubated overnight at 4°C, in the Ni-NTA coated 96-well HisSorb plates (Qiagen), and then washed four times with 200 µl of PBST. 200 µl of biotin-dextran (100 µg/ml, Fluka) in PBSA were added into the microwells for 10 min at room temperature, washed 4 times with PBST buffer and incubated during 30 min with 200 µl of Extravidin-alkaline phosphatase conjugate at 1/20,000 dilution (Sigma) in PBSA. 200 µl of phosphatase substrate solution were finally added. After 30 min, the amount of dextran bound to the proteins was determined by the color change monitored at 405 nm using a Multiskan Ascent microplate reader. The assays were always performed together with reagent blank and lysate from *E. coli* carrying the native vector as negative control.

To determine the dissociation constant K_d, different amounts of biotin-dextran ranging from 0.14 nM to 143 nM were added to the *Thioredoxin*-GBD-*His* extracts fixed into the Ni-NTA wells, following the procedure described above. Absorbances measured at 405 nm (representing the amount of dextran bound to the *Thioredoxin*-GBD-*His* ligand) were then plotted as a function of biotin-dextran concentration added into the well. *Kd* values were determined following the one-site saturation ligand binding equation $Abs_{405nm} = \frac{B_{max} \times [Biotin-dextran]}{K_d + [Biotin-dextran]}$ using SigmaPlot 10.0 software. B_{max} corresponds to the maximum absorbance value obtained by the complete saturation of the binding site. For competition studies, Dextran 70 was used as competitor against 1 mM biotinylated dextran.

PBST	**PBSA**
Tween20 0.05 %	BSA 0.2 %
In PBS buffer pH 7.2	In PBS buffer pH 7.2

Alkaline phosphatase substrate solution

p-nitrophenylphosphate (Sigma)	1 mg/ml
$NaHCO_3$	28 mM
Na_2CO_3	22 mM
$MgCl_2$	5 mM

5.2 Binding of His-tagged protein to immobilized Ni-NTA molecules into the HisSorb plate.

The binding between his-tagged proteins to Ni-NTA molecule were determined. The cleared lysates of each GBD derivatives were incubated in the Ni-NTA HisSorb plate at 4 °C overnight. The unbound proteins were removed by washing 4 times by PBST pH 7.2. His-tagged proteins were eluted in PBS buffer pH 7.2 containing 250 mM imidazole solution. The eluted proteins were visualized by SDS-PAGE and Western blot analysis using anti-his antibody.

6. High-throughput screening for GBD mutant libraries

6.1 Platforms

6.1.1 QPix

The QPix (Genetix) bench top colony picking was introduced in directed protein evolution experiment (Figure 33). This robot provides high-accuracy and high-throughput colony picking. The colony grown on solid medium can be picked under the thresholds specified by user such as colony diameter, roundness, axis ratio, etc. Then, the whole tray was imaged by CCD camera. Information of user's define parameters were analyzed to select the colony for picking. The additional application of the QPix is gridding using 96-pin gridding head. This function allows the liquid samples to be replicated onto a membrane for hybridization analysis.

Figure 33. Benchtop colony picking robot, QPix (Genetix).

6.1.2 Biomek 2000

The Biomek 2000 (Beckman) was designed to facilitate pipetting, diluting, and dispending operations (Figure 34). The system was controlled by Biowork software. In this study, we used the Biomek 2000 to pipette the culture media to a 384- well plate.

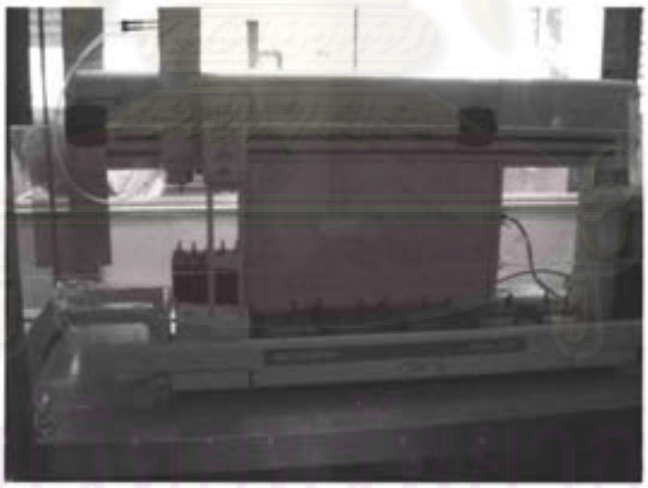


Figure 34. Automated pipetting workstation, Biomek 2000 (Beckman).

6.2 Picking

A 384-well plate was filled with 80 µl of LB+8% glycerol by the Biomek 2000. The GBD variant colonies derived from mutagenesis experiments were picked

to the 384-well plate using QPix, program 'Picking' and 96-pin picking head. The picking was performed under these following parameters; diameter = 7-50 mm, roundness = 0.80, axis ratio = 0.70, proximity = 2, overlap = 20. From Figure 35, the actuator carrying the 96-pin picking head was first move to a water bath, 70 % ethanol bath, and drying. Then, the colony picking was performed according to the defined parameters. After that, the actuator moved to the destination plate and dipped to inoculation 3 times and sterilization the pins. When the picking finished, the plate was shaking at 30 °C overnight. This master plate was used as a source of variants and was stored at -20 °C.

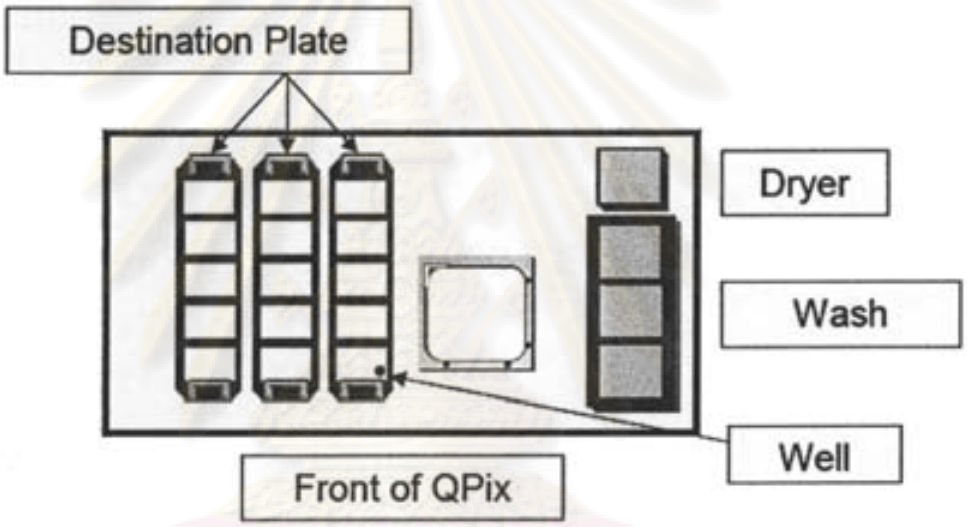


Figure 35. Layout of QPix for program 'Picking' (Genetix).

6.3 Replicating

Using QPix program 'Replicating' and 96-pin gridding head, the bacterial cells from stock culture plate was gridded onto a Durapore membrane (Millipore) overlaid on an Omnitray (Nunc) containing 50 ml of 2XYT agar pH 6.4, 100 μg/ml of ampicillin, and 0.002% of arabinose. The Durapore membrane was previously sterilized under UV light for 2 min. Source plate holder was settled as Genetix 384 well X6000 in which inking depth was indicated at 6,500. Destination plate was specified as Nunc 384-well plate which inking depth equal to 7,500. The gridding was

performed four times per well. These configurations were similar in all experiments. The Omnitray was incubated at 30 °C for 48 h.

6.4 Detection

The detection of positive variants was developed using the biotin-dextran/Extravidin system previously described in the glucan-binding assay coupled with a colony filtration blotting (Dahlroth *et al.*, 2006). Whatman 3MM paper and nitrocellulose were cut to fit with Omnitray size (11.6 x 7.5 cm). Whatman paper was drenched with 10 ml of lysis buffer. Nitrocellulose membrane with pore size 0.22 µm was placed onto the top of Whatman paper. Excess lysis buffer was poured off and then the Durapore membrane containing bacterial colonies was placed on the top of the nitrocellulose membrane with the colonies facing up (Figure 36). The sandwich was incubated for 30 min at room temperature. After that, the membranes were performed the freeze-thaw cycles by placing the container at -80 °C for 10 min and then 37 °C for 10 min for 3 cycles. Only nitrocellulose membrane was kept to further analysis.

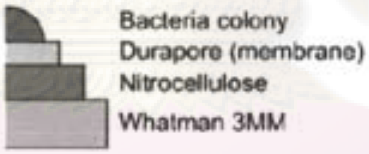


Figure 36. Colony filtration blotting (Dahlroth *et al.*, 2006).

The nitrocellulose membrane was blocked with 40 ml of blocking buffer for at least 1 h at room temperature on a horizontal shaker. The membrane was washed with PBST buffer for 10 min 3 times. Then, 40 ml of 10 µg/ml of biotin-dextran solution was added and incubated for 1 h. Washing step was performed as before. 40 ml of 1/20,000 dilution of Extravidin-alkaline phosphatase solution was added and incubated again for 1 h and washed. The binding of protein to dextran was revealed by adding 5-10 ml of BCIP/NBT (Sigma) and incubated for 30 min. The coloration was stopped by washing with water.

Lysis buffer		**Blocking solution**	
Tris pH 8.0	20 mM	BSA or Casein	3 %
NaCl	100 mM	in PBS buffer pH 7.2	
PMSF	1 mM		
Benzonase	11 Units/µl		
Lysozyme	0.2 mg/ml		

7. Analytical methods

7.1 Protein concentration

Protein concentration was estimated by Microbradford's assay in 96-well plate using BSA as standard.

7.2 Quantification of thioredoxin

Quantitative determination of thioredoxin in each fraction collected during purification onto Sephacryl® was performed using Holmgren's method (1979). Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol, causing the formation of insoluble insulin chains. The amount of thioredoxin present in the sample is then correlated to the turbidity measured at OD 650 nm. Purified *E. coli* thioredoxin was purchased from Sigma and used as standard. Insulin stock solutions were prepared at 10 mg/ml in 50 mM Tris-HCl pH 7.5, adjusted to pH 2-3 by addition of 1 M HCl, and rapidly titrated to pH 7.5 with 1 M NaOH. All experiments were carried out at room temperature. The reaction mixture was prepared in a 96-well plate by addition of 0.1% (w/v) insulin, 2 mM EDTA, protein sample and water in potassium phosphate buffer pH 7.0, to give a final volume of 198 µl. Reaction was started by adding 2 µl of 100 mM DL-dithiothreitol (Sigma) in each well. The microplate was vigorously shaken, and then let stand for 50 min. The increase of turbidity due to the precipitated insulin chains is then followed at OD 650 nm. The non-enzymatic reduction of insulin was recorded in a control well, without any added thioredoxin.

CHAPTER IV
RESULTS AND DISCUSSIONS

In this chapter, results and discussions are separated into three parts. Part I is fulfill the first objective of this study about identification and characterization of a minimal motif responsible for glucan binding of dextransucrase from *Leuconostoc mesenteroides* NRRL B-512F. The second part is application of the shortest GBD fragment to facilitate affinity purification onto commercially available dextran-containing supports. Finally, in the last part, the shortest GBD fragment was improved its soluble expression and reduce the size using mutagenesis techniques.

PART I

RATIONAL ENGINEERING OF THE DSR-S GLUCAN-BINDING DOMAIN

A number of researches have investigated the role of glucansucrases C-terminal domain over the past decades. Most of them highlighted its property to bind the glucan produced by the enzymes (Kobayashi *et al.*, 1989; Lis *et al.*, 1995; Shah and Russell, 2002), and proposed to call it "Glucan binding domain" (GBD). Its presence is often crucial to maintain the enzyme activity (Monchois *et al.*, 1998), and recent investigations showed its involvement in dextran chain elongation (Moulis *et al.*, 2006a). However, in this domain of about 400 amino acids, minimal peptide sequences responsible for glucan binding have not been yet identified, and lack of structural data limits our understanding on GBD-dextran interactions.

To better characterize the role of the DSR-S C-terminal domain, the aim of this part was to identify the minimal determinants responsible for dextran binding. For this purpose, several truncated forms were generated, and their glucan binding properties were determined using a specific glucan binding assay.

1.1 Sequence analysis of the C-terminal domain of DSR-S

Sequence analysis of different glucansucrases revealed that the C-terminal domain was composed of various repeated units. As presented in Literature review, on the basis of sequence similarities, they can be classified as YG repeats (Giffard and Jacques, 1994) or A, B, C, D, and N repeats (Ferretti *et al.*, 1987; Gilmore *et al.*, 1990; Abo *et al.*, 1991; Giffard *et al.*, 1991; Monchois *et al.*, 1998). These repeats were assumed to possess a functional significance.

The C-terminal domain of DSR-S contains 3 A, 3 C and 5 N repeats. The relationship between the A-C-N repeats and the YG repeat of the C-terminal domain of DSR-S is illustrated in Figure 37. At least one YG motif is present in all A, C and N repeats, the last ones covering more amino acids. N repeats are a particularity of this enzyme. They are not highly conserved, but possess the main characteristics of

the YG repeats (Monchois *et al.*, 1998). In order to identify the minimal motif responsible for glucan binding, we designed truncatures of the DSR-S C-terminal domain following the consensus YG repeats. The amino acid sequences of each YG repeat is presented in Table 11. The C-terminal domain of DSR-S contains 15 YG repeats, comprising from 12 to 21 amino acids.

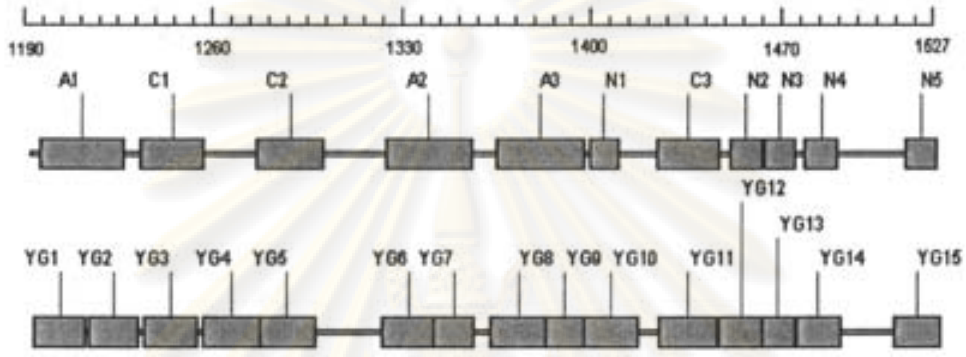


Figure 37. Location and distribution of the A-C-N repeat and the YG repeat in the C-terminal domain of DSR-S (DSR-S numbering).

Table 11. Sequences of the repeats identified in DSR-S C-terminal domain, following the consensus YG motif defined by Giffard and Jacques (1994). Boldface type indicates conserved amino acids; x corresponds to non conserved amino acids, and H° to hydrophobic residues.

Motive	Sequence
consensus	-NDGYYFxxxGxxH°x(G/N)xH°H°H°
YG_1	-NNYYYFDKT--GHLVT-GLQKI
YG_2	-NHTYFFLPN--GIELVKSFLQ
YG_3	DGTIVYFDKK--GHQVFDQYIT
YG_4	NGNAYYFDDA--GVMLKSGLATI
YG_5	DGHQQYFDQN--GVQVKDKFVIG
YG_6	-NQWFYFDGN--GHAVT-GFQTI
YG_7	NGKKQYFYND--GHQSK
YG_8	-DGDTFYTSATDGRLVT-GVQKI
YG_9	NGITYAFDNT--GNLI
YG_{10}	TN-QYYQLAD--GKYMLLDDSGR
YG_{11}	DGVLRYFDQN--GEQVKDAIIVD
YG_{12}	DTNLSYYFNATQGVAV
YG_{13}	--KNDYFEYQ--GNWY
YG_{14}	TD-ANYQLIK--GFKAVDD
YG_{15}	AQGKVYQFDNN-GNAVSA

1.2 Cloning of the C-terminal domain of DSR-S

pBAD/TOPO Thiofusion was selected as cloning and expression vector. Recombinant proteins generated after expression are in fusion with a Thioredoxin tag at the N-terminal extremity, and a His tag at C-terminal end (*Thioredoxin*-GBD-*His* fusion proteins). Fusion with such Thioredoxin tag is well known to improve solubility of linked proteins. It could also be a model of study for protein purification using GBD as affinity tag. The latter fusion partner, His tag, facilitates the binding to immobilized Ni-NTA molecules into 96-well HisSorb plates used in the glucan binding assay and allows the purification of GBD proteins via immobilized metal affinity chromatography (IMAC).

The full-length *gbd* gene (*gbd-0*) was amplified from the recombinant plasmid containing the DSR-S encoding gene (*dsr-S*) kindly given by Moulis *et al.* (2006b) using the Expand High Fidelity PCR system (Roche) with the pair of primers ForGBD and RevGBD (Table 10). The *gbd*-0 gene was then ligated into pBAD/TOPO Thiofusion vector on the basis of TA cloning. After ligation and transformation, positive transformants were analyzed by restriction enzyme digestion and DNA sequencing. Figure 38 showed the restriction map of the recombinant pBAD/TOPO Thio/*gbd*-0 and restriction analysis of the positive clones.

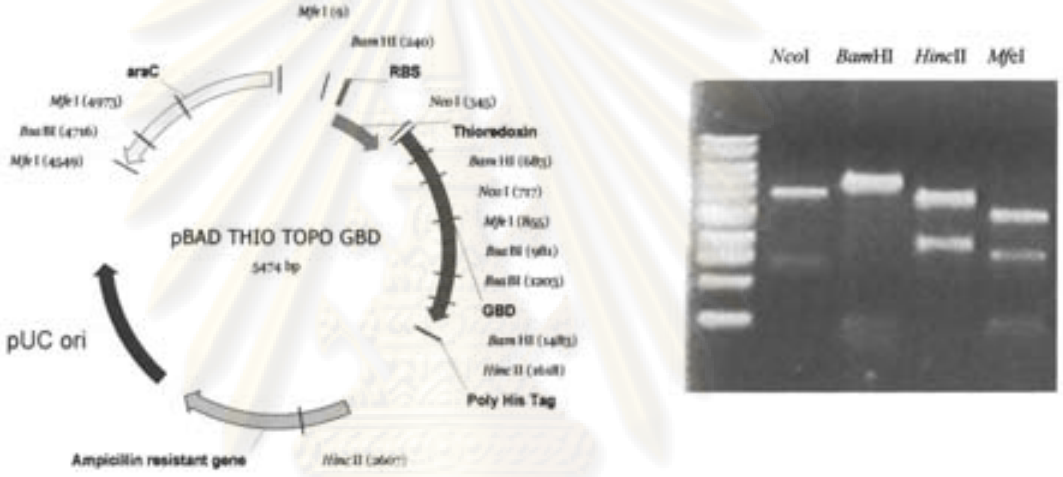


Figure 38. Restriction map of GBD-0: *Nco*I 5102, 372 bp;*Bam*HI 4231, 800, 443 bp; *Hinc*II 4485, 989 bp; *Mfe*I 3694, 846, 510, 424 bp.

Expression of recombinant GBD-0 proteins were performed in LBT (LB supplement with 100 mM tris-HCl pH 6.4) since the DSR-S has been unstable at alkaline pH (Wilke-Douglas *et al.*, 1989, Croux and Garcia, 1992). *E. coli* cells containing *gbd-0* gene were grown in a baffle flask at 23 °C with shaking at 130 rpm. 0.002% of L-arabinose was added to the culture medium when the OD_{600} reached 0.6, as inducer of recombinant protein expression through a tight regulation of the *ara*BAD promoter. Cells were cultivated until an OD_{600} of ≈ 2.5.

Post-sonicated extract was analyzed by SDS-PAGE and Western blot hybridization using anti-thio and anti-his antibody in order to investigate the

expression of GBD-0. The molecular weight of the resulting protein was estimated at 55.7 kDa from Western-Blot analysis (Figure 39). In fusion with both tags, the theoretical molecular weight of only GBD-0 protein is 39.7 kDa, validating our cloning strategy. Presence of both tags at N and C-terminal extremity proves that there is no GBD-0 degradation during the expression. GBD-0 contains the amino acids D1190 to A1527 following DSR-S numbering.

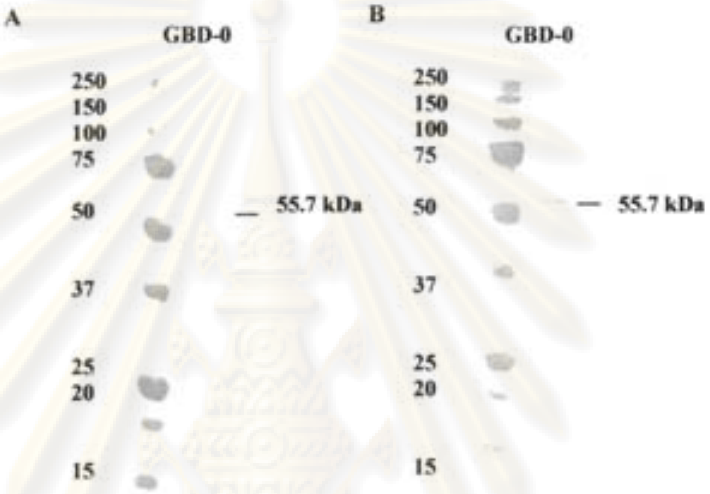


Figure 39. SDS-PAGE of 10% Bis-tris gel (Invitrogen) and Western blot analysis of GBD-0 soluble extracts using anti-Thio (A) and anti-His (B).

To determine the binding ability of GBD-0, an antibody-independent binding assay was first developed by Lis *et al.* (1995). For this assay, protein extracts were fixed to the wells of microtitre plates (Dynatech Laboratories, Chantilly, Va.). After washes and blocking, biotin-dextran was added. Streptavidin-horseradish peroxidase conjugate was then added in order to recognize the biotin. Finally, the GBD:dextran-biotin:streptavidin-horseradish peroxidase complexes were revealed using a color detection.

More recently, Shah and Russell (2002) introduced the used of Ni-NTA coated 96-well HisSorb plates (Qiagen) for this kind of glucan-binding assays. This method exploits a specific binding of the His-tagged protein into Ni-NTA molecules. As

shown in Figure 40, the assay consists, first, in a binding of the *Thioredoxin*-GBD-*His* proteins onto the immobilized Ni-NTA molecules of the wells, followed by washing steps. Biotin-labeled dextran (Molecular weight of 70 kDa) is then added, and after second washes using PBST pH7.2 buffer, the GBD:biotin-dextran complexes are revealed using Extravidin-alkaline phosphatase. We decided to use this protocol for the determination of the glucan binding abilities of the GBD-0 and its derivatives constructed in this study.

Immobilized Ni-NTA molecules ( ) on the HisSorb plate (Qiagen)

Thioredoxin-GBD-*His* fusion protein ( ) binds to Ni-NTA molecule via the His tag

Biotin-labeled dextran ( ) is added into the well, forming GBD:Dextran-biotin complexes

Extravidin-Alkaline phosphatase ( ) recognizes biotin molecule.

Substrate of alkaline phosphatase is added to produce a yellow colored product ( )

Figure 40. Glucan-binding assay developed by Shah and Russell (2002).

Glucan-binding ability of GBD-0 was first compared to that of the whole recombinant DSR-S. Protein extract from *E. coli* harboring the pBAD/TOPO Thiofusion only was used as negative control (NC). As shown in Figure 41, both constructions had the ability to bind the biotin-dextran, confirming the "Glucan Binding Domain" appellation of the DSR-S C-terminal part. When Shah *et al.* (2004) compared the glucan binding ability of various glucansucrases from *Streptococcus* and *Leuconostoc* sp., they found that DSR-S did not present any ability to bind dextran. Their binding test was however carried out with the native enzyme not dissociated from its natural polymer film, which probably interfered during the assay.

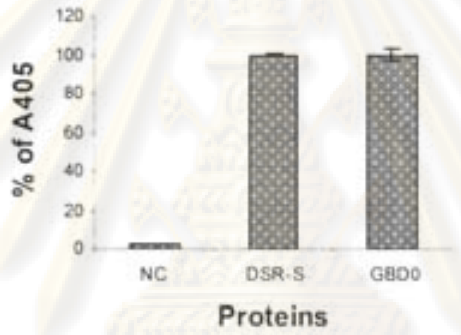


Figure 41. Glucan-binding ability of GBD-0 and whole enzyme DSR-S. Protein extract from *E. coli* carrying the pBAD/TOPO Thiofusion vector only was used as negative control (NC).

To ensure that either the binding test was not due to the mis-binding of the fusion protein to biotin, or not due to biotin interferences between GBD-0 and dextran, the binding assay was performed as a competitive assay with unlabelled dextran of 70 kDa (dextran 70) (Figure 42). The unlabeled dextran 70 can efficiently compete the binding between GBD-0 and biotin-dextran, indicating that both biotin-dextran and unlabel dextran can compete at the same sites on GBD protein. Moreover, the addition of an equal concentration of unlabel dextran 70 could reduced approximately 50% of binding. This result revealed that conjugation of biotin to dextran molecules has no effect on the binding of GBD onto dextran.

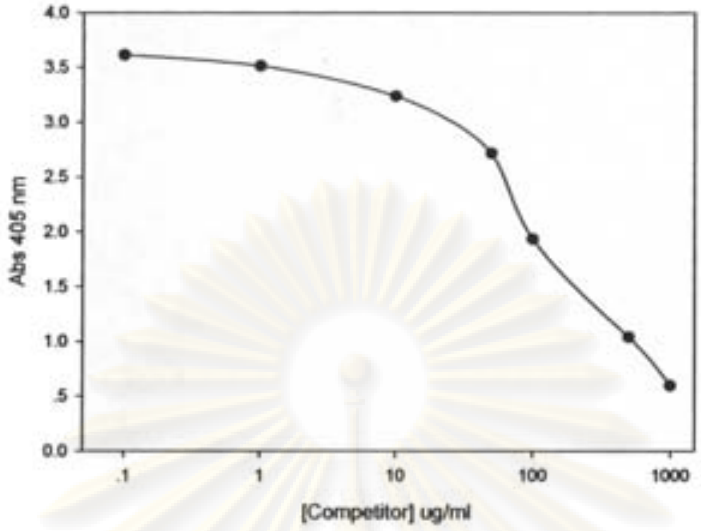


Figure 42. Competitive glucan-binding assay of GBD-0 by dextran 70. The concentration of biotin-dextran was kept constant at 100 µg/ml.

The C terminal domain acts as an independent motif for glucan binding, with an efficiency equivalent to that of the whole DSR-S enzyme. This domain was thus subjected to mutagenesis experiments in order to investigate in more details the GBD structure/function relationships, notably by searching the minimal determinants responsible for glucan binding.

1.3 Mutant construction and characterization

Gbd-0 gene was used as template to amplify several truncated forms by PCR. The truncature positions were chosen on the basis of GBD sequence analysis, considering that this domain is composed of various YG repeated motives as shown in Table 11. Multiple pairs of primers were designed in a way to avoid a cut in the middle of a motif. As for GBD-0 construction, all truncated forms were cloned into the pBAD/TOPO Thiofusion vector, resulting in fusion proteins containing Thioredoxin at the N-terminal extremity, and His tag at the C-terminal end (*Thioredoxin*-GBD-*His* protein, Figure 43).

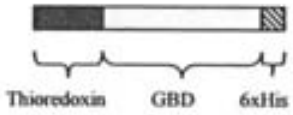


Figure 43. *Thioredoxin*-GBD-*His* fusion protein.

The recombinant proteins were expressed in LBT pH 6.4 at 23°C as previously described by Moulis *et al.* (2006b). Bacterial cell were harvested by centrifugation and the cells were lysed by sonication. Cleared lysates were obtained by centrifugation and filter sterilization. These lysates were used as sources of fusion proteins for Western-blot analyses and glucan binding assays.

Western-blot analyses were performed in order to control the level and the profile of protein expression. After the transfer to the nitrocellulose membrane, both anti-thio and anti-his were used to detect the full-length GBD proteins (Figure 44C, 45C, 46C). The results revealed that all the GBD variants constructed in this study were well-expressed and not degraded.

From the first screening, GBD-1, GBD-2, and GBD-3 were obtained by truncation of GBD-0 into 3 parts (Table 12). GBD-1 contains the first 5 YG repeats (YG_1-YG_5), GBD-2 includes the central repeats YG_6 to YG_{10}, and GBD-3 covers the last 5 YG repeats, from YG_{11} to YG_{15}. As shown in Figure 47, the three variants did not bind dextran. GBD-4 and GBD-5, which correspond to the fusion of GBD-1 and GBD-2, and to the fusion of GBD-2 and GBD-3 respectively, were then constructed. GBD-5, representing the two third ends of the C-terminal part of GBD-0, was the only truncated variant found to be able to bind dextran like GBD-0.

A

GBD-0 YG_1-YG_2-YG_3-YG_4-YG_5-YG_6-YG_7-YG_8-YG_9-YG_{10}-YG_{11}-YG_{12}-YG_{13}-YG_{14}-YG_{15}

GBD-1 YG_1-YG_2-YG_3-YG_4-YG_5

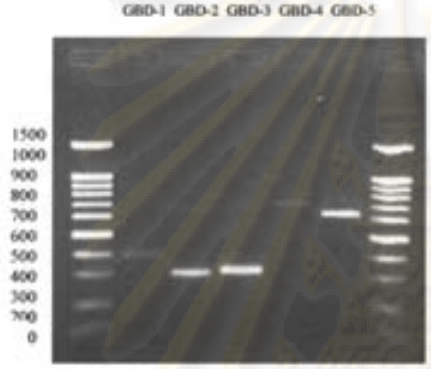
GBD-2 YG_6-YG_7-YG_8-YG_9-YG_{10}

GBD-3 YG_{11}-YG_{12}-YG_{13}-YG_{14}-YG_{15}

GBD-4 YG_1-YG_2-YG_3-YG_4-YG_5-YG_6-YG_7-YG_8-YG_9

GBD-5 YG_6-YG_7-YG_8-YG_9-YG_{10}-YG_{11}-YG_{12}-YG_{13}-YG_{14}-YG_{15}

B



C

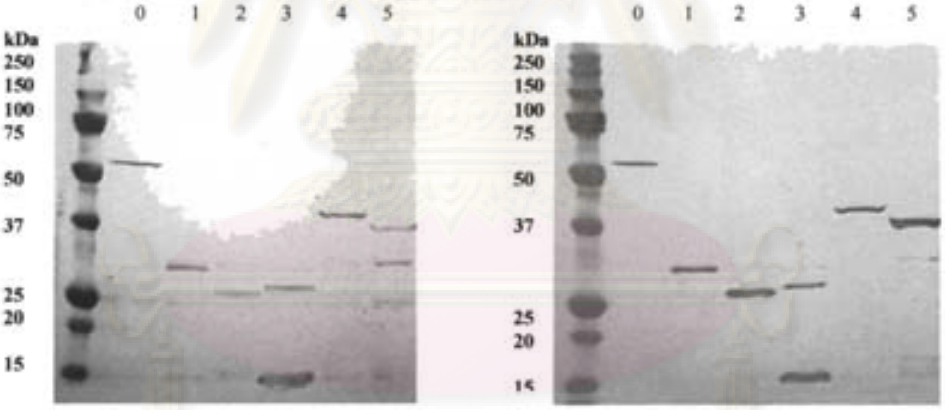
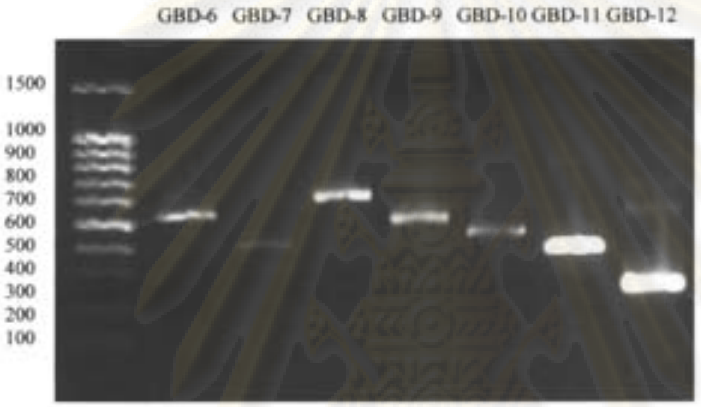


Figure 44. GBD-0 derivatives. A. Rational design of GBD-0 derivatives based on YG repeated unit sequence. B. Purified PCR products GBD-1 to GBD-5 on 2% agarose gel. C. SDS-PAGE and Western blot analysis using anti-thio (Left) and anti-his (Right) antibody.

A

GBD-5	YG6-YG7-YG8-YG9-YG10-YG11-YG12-YG13-YG14-YG15
GBD-6	YG8-YG9-YG10-YG11-YG12-YG13-YG14-YG15
GBD-7	YG10-YG11-YG12-YG13-YG14-YG15
GBD-8	YG6-YG7-YG8-YG9-YG10-YG11-YG12-YG13-YG14
GBD-9	YG6-YG7-YG8-YG9-YG10-YG11-YG12-YG13
GBD-10	YG6-YG7-YG8-YG9-YG10-YG11-YG12
GBD-11	YG6-YG7-YG8-YG9-YG10-YG11
GBD-12	YG8-YG9-YG10-YG11

B



C

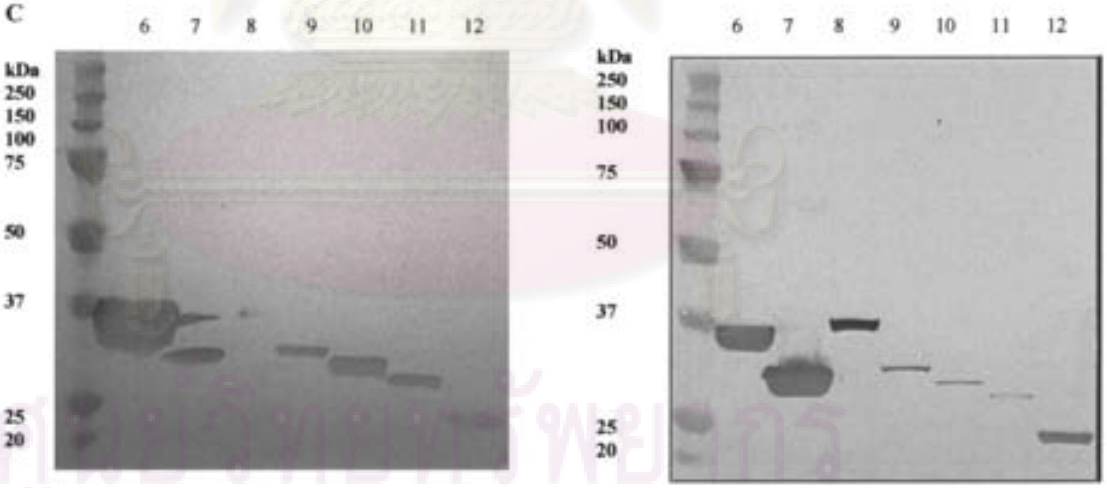


Figure 45. GBD-5 derivatives. A. Rational design of GBD-5 derivatives based on YG repeated unit sequence. B. Purified PCR products GBD-6 to GBD-12 on 2% agarose gel. C. SDS-PAGE and Western blot analysis using anti-thio (Left) and anti-his (Right) antibody.

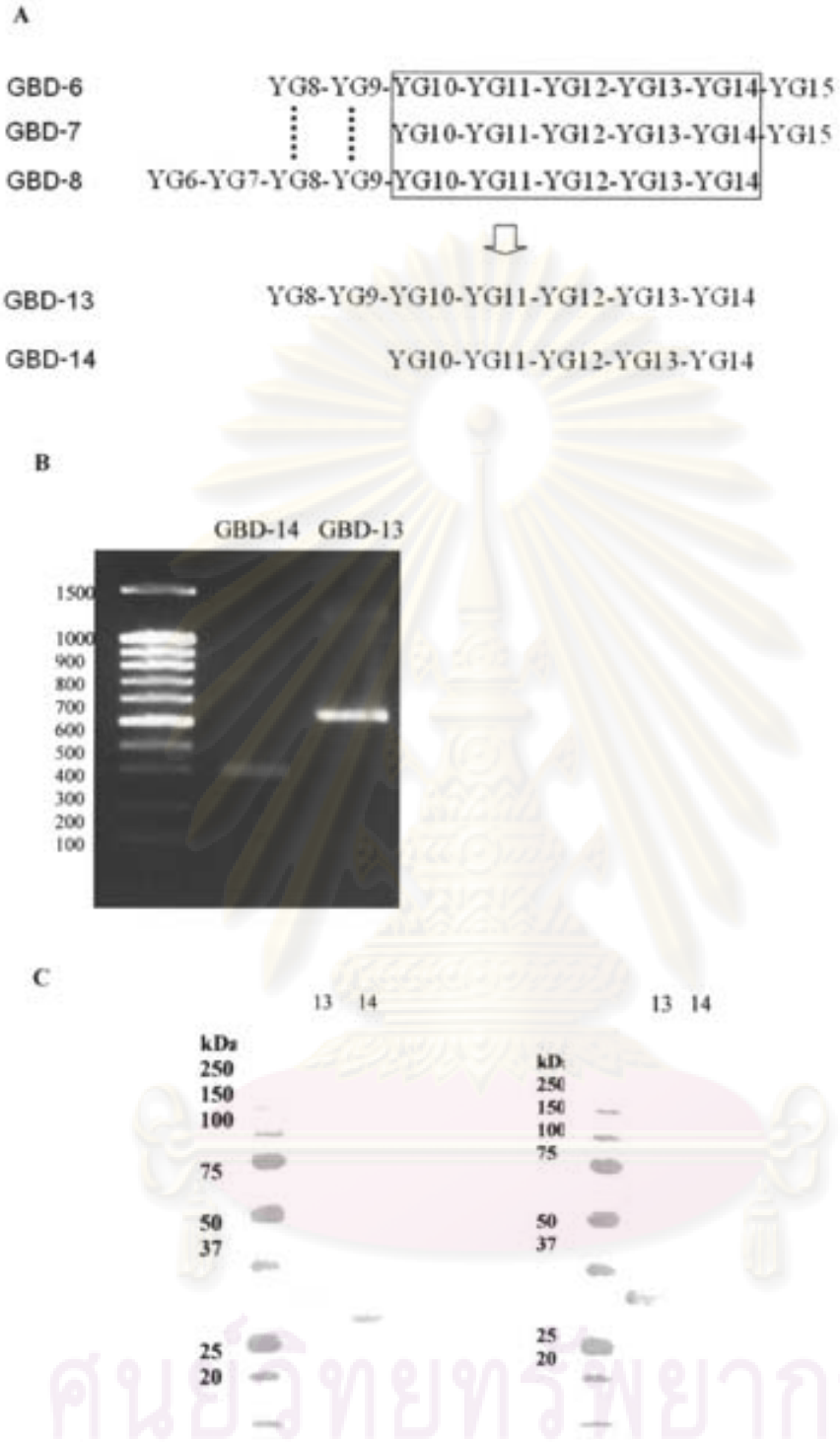


Figure 46. GBD-13 and GBD-14. A. Common motif of GBD-6, GBD-7, GBD-8. B. Purified PCR products GBD-13 and GBD-14 on 2% agarose gel. C. SDS-PAGE and Western blot analysis using anti-thio (Left) and anti-his (Right) antibody.

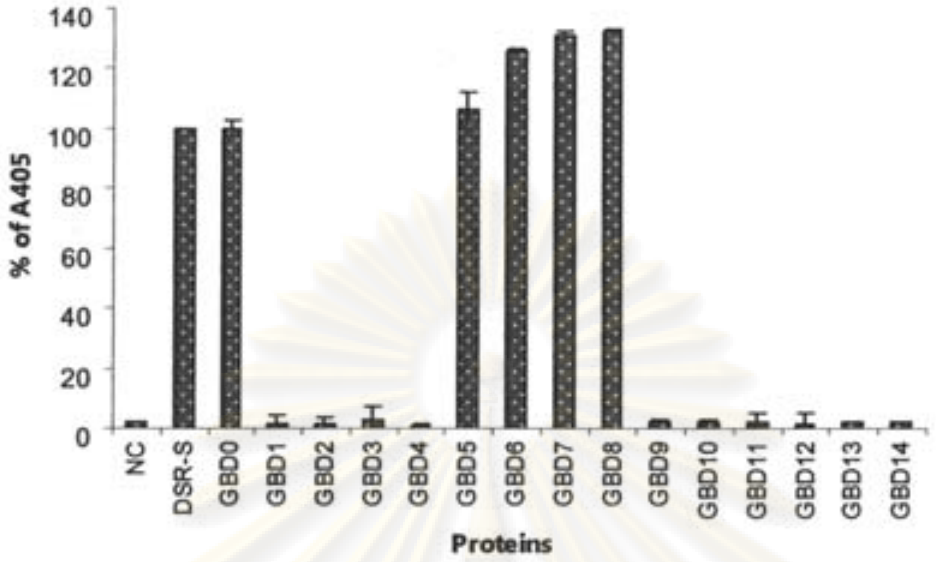


Figure 47. Relative glucan binding ability of GBD-0 derivatives compared to the wild-type. NC = Negative control (production of *E. coli* TOP 10 without GBD derivative plasmid transformation).

GBD-5 (21.5 kDa) includes the motif YG_6-YG_7-YG_8-YG_9-YG_{10}-YG_{11}-YG_{12}-YG_{13}-YG_{14}-YG_{15}. Construction of shorter motives was thus undertaken. Truncatures at N and C-terminal extremities of GBD-5 resulted in the construction of GBD-6, GBD-7 and GBD-8. All these forms revealed dextran binding properties (Figure 47). All other attempts to reduce GBD-7 and GBD-8 constructions entertained a loss of glucan binding ability (GBD-9 to GBD-12 presented Figure 45 and Figure 47).

From GBD-6, GBD-7 and GBD-8 which presented dextran binding properties, the common motif of these forms corresponds to GBD-13 and GBD-14 and is composed of the motives (YG8-YG9)-YG_{10}-YG_{11}-YG_{12}-YG_{13}-YG_{14} (Figure 46). However, both of them did not bind dextran, showing that the addition of either YG_6-YG_7 repeats at the N-terminus (GBD-8) or YG_{15} at the C-terminus (GBD-7) is required to provide binding capacities. Sequence comparison between YG_6 fragment

and YG_{15} revealed the presence of the conserved sequence: FD(G/N)NG(H/N)AV (Figure 48).

```
                    10
YG6     NQW--FYFDGNGHAVTGFQTI
          :   . ::.::.::..
YG15    AQGKVYQFDNNGNAVSA----
                    10
```

Figure 48. Sequence alignment between the repeats A_2 and N_5 of the DSR-S C-terminal domain.

Based on these results, this sequence seems to be crucial for binding activity. It may act directly in binding or be important to ensure a correct folding of the motif and a proper exposition of the amino acids involved in sugar/protein interactions. In the absence of three dimensional structure, it is difficult to choose between these hypotheses because the entire GBD sequence contains many amino-acids usually described as key residues in sugar-protein interactions such as polar residues able to interact with sugar hydroxyl groups (Quiocho, 1986; Fernandez-Tornero *et al.*, 2001), hydrophobic residues responsible for stacking interactions (Shah *et al.*, 2004) and strongly basic and acidic residues involved in salt-bridge formations (Skov *et al.*, 2002). Summary of all constructions in size and dextran-binding ability is presented Table 12.

Table 12. Molecular weight of GBD tag and *Thioredoxin*-GBD-*His* fusion proteins.

Variant	Primers	Amino-acids	Motive	Size (kDa)	Fusion protein size (kDa)	Dextran binding ability
GBD-0	ForGBD- RevGBD-	D1190 to A1527	YG_1-YG_2-YG_3-YG_4-YG_5-YG_6-YG_7-YG_8-YG_9-YG_{10}-YG_{11}-YG_{12}-YG_{13}-YG_{14}-YG_{15}	39.7	55.7	+
GBD-1	GBD1F GBD3R	Y1193 to Q1322	YG_1-YG_2-YG_3-YG_4-YG_5	12.9	28.9	-
GBD-2	GBD5F GBD7R	W1323 to Q1423	YG_6-YG_7-YG_8-YG_9-YG_{10}	8.8	24.8	-
GBD-3	GBD9F RevGBD	D1424 to A1527	YG_{11}-YG_{12}-YG_{13}-YG_{14}-YG_{15}	9.8	25.8	-
GBD-4	GBD1F GBD7R	Y1193 to Q1423	YG_1-YG_2-YG_3-YG_4-YG_5-YG_6-YG_7-YG_8-YG_9	24.5	40.5	-
GBD-5	GBD5F RevGBD	W1323 to A1527	YG_6-YG_7-YG_8-YG_9-YG_{10}-YG_{11}-YG_{12}-YG_{13}-YG_{14}-YG_{15}	21.5	37.5	+
GBD-6	GBD10F RevGBD	G1355 to A1527	YG_8-YG_9-YG_{10}-YG_{11}-YG_{12}-YG_{13}-YG_{14}-YG_{15}	19	35.0	+
GBD-7	GBD11F RevGBD	N1397 to A1527	YG_{10}-YG_{11}-YG_{12}-YG_{13}-YG_{14}-YG_{15}	14	30.0	+
GBD-8	GBD5F GBD12R	W1323 to V1515	YG_6-YG_7-YG_8-YG_9-YG_{10}-YG_{11}-YG_{12}-YG_{13}-YG_{14}	20	36.0	+
GBD-9	GBD5F GBD13R	W1323 to A1477	YG_6-YG_7-YG_8-YG_9-YG_{10}-YG_{11}-YG_{12}-YG_{13}	13.8	29.8	-
GBD-10	GBD5F GBD16R	W1323 to N1463	YG_6-YG_7-YG_8-YG_9-YG_{10}-YG_{11}-YG_{12}	11.8	27.8	-
GBD-11	GBD5F GBD15R	W1323 to I1442	YG_6-YG_7-YG_8-YG_9-YG_{10}-YG_{11}	9.4	25.4	-
GBD-12	GBD14F GBD15R	T1364 to I1442	YG_8-YG_9-YG_{10}-YG_{11}	5.6	21.6	-
GBD-13	GBD10F GBD12R	G1355 to V1515	YG_8-YG_9-YG_{10}-YG_{11}-YG_{12}-YG_{13}-YG_{14}	16.7	32.7	-
GBD-14	GBD17F GBD18R	Y1399 to D1490	YG_{10}-YG_{11}-YG_{12}-YG_{13}-YG_{14}	12.7	28.7	-

1.4 Conformational checking of the deleted-GBD peptides

It was verified that all the variants were able to bind via their His tag onto the Ni-NTA molecules adsorbed on the HisSorb plates. The experiment started with the binding of the cell lysates to the Ni-NTA HisSorb plates at 4 °C overnight. Then unbound proteins were removed by 4 washes usin'g PBST pH 7.2 buffer. Then, 250 mM of imidazole were added into the wells to compete the binding between His-tagged protein and the Ni-NTA molecules (Figure 49). The His-tagged proteins in the eluted fraction were visualized by SDS-PAGE and Western-blot analyses using anti-his antibody. Figure 49 shows that all of GBD derivatives are capable to bind to the immobilized Ni-NTA molecules coated on the wells.

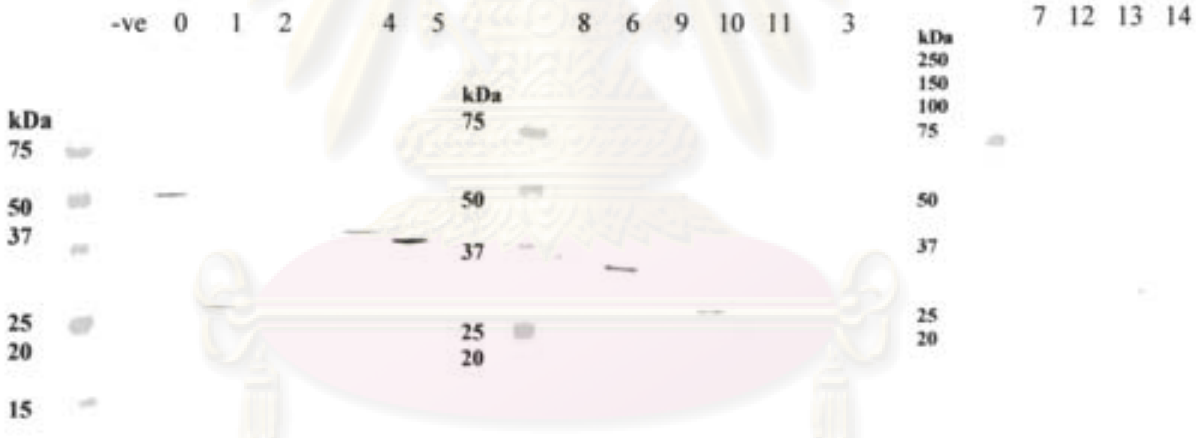


Figure 49. Western-blot analysis of eluted fractions with anti-*His* antibody detection.

1.5 Affinity electrophoresis toward dextran

The electrophoresis behavior of some GBD variants (GBD-6, GBD-0 and GBD-12) in native PAGE gels containing 0.1% (w/v) of dextran of 2000 kDa was investigated in order to confirm their dextran binding ability. GBD-0 and GBD-12

were chosen as positive and negative control respectively. Before starting the electrophoresis, these proteins were purified by Ni-NTA agarose resin to obtain the purified form of these constructs.

1.5.1 His-Ni-NTA protein purification

Culture growth and recombinant protein expression conditions were optimized in order to reach the maximum of proteins of interest (GBD-0, GBD-6 and GBD-12). The *E. coli* cells carrying GBD-0 were grown at 37°C to an $OD_{600} \approx 0.6$. Then, different concentrations of L-arabinose from 0.00002% to 2 % were applied to induce the recombinant protein production. The cells were harvested at 2.5 $OD_{600\ nm}$. The protein samples were analyzed by SDS-PAGE and stained with colloidal blue (Figure 50). The highest expression level was observed using 0.02% (w/v) of L-arabinose as an inducer. However, the majority of the GBD-0 fusion proteins were found to be localized in the insoluble fraction. Determination of the optimal temperature of cultivation was then investigated, as it is known that a diminution of this parameter can increase the protein solubility during heterologous protein expressions (Schein, 1989; Shirano and Shibata, 1990). From Figure 50 (right), the expression at 16°C showed the highest amount of soluble GBD-0 proteins. Therefore, the optimal production of GBD-0 protein was achieved at 16°C cultivations, using 0.02% of arabinose as the inducer.

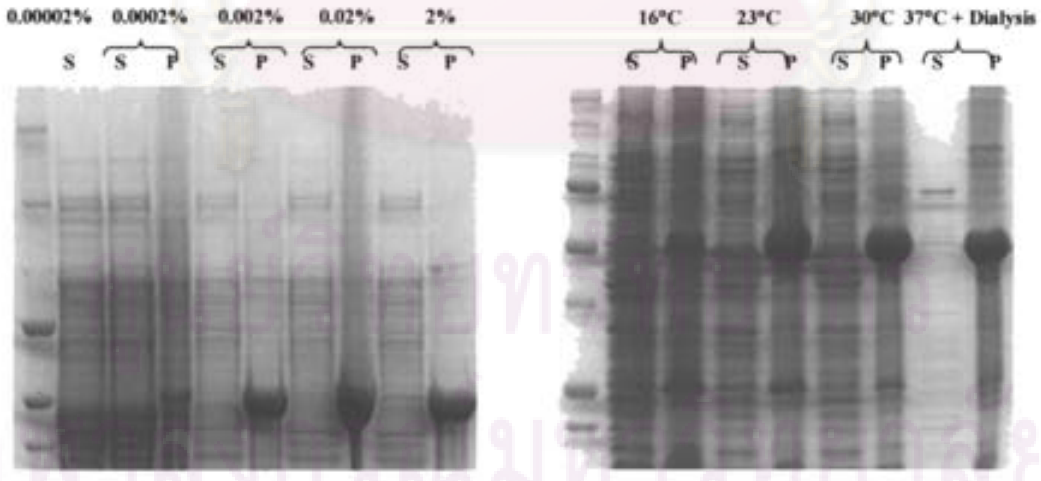


Figure 50. Optimization of GBD-0 expression level: modification of arabinose concentration (left) and temperature (right). S=supernatant, P=pellet.

GBD-0 and GBD-12 proteins were expressed as soluble polypeptides in the conditions described just above. However, the level of expression of GBD-6 was found to be higher at 23°C than at 16 °C, using 0.02 % of L-arabinose.

All proteins were purified from the soluble fractions using Ni-NTA agarose (Qiagen) under native conditions. As many contaminating proteins were observed in the eluates using the standard protocol given by the supplier, washes were performed by increasing the salt concentration to 1M NaCl, and eluted fractions were kept step by step. The purity of each fraction was determined by SDS-PAGE and colloidal blue staining. Elutions obtained in GBD-0, 6, and 12 purifications, respectively are presented in Figures 51 and 52.

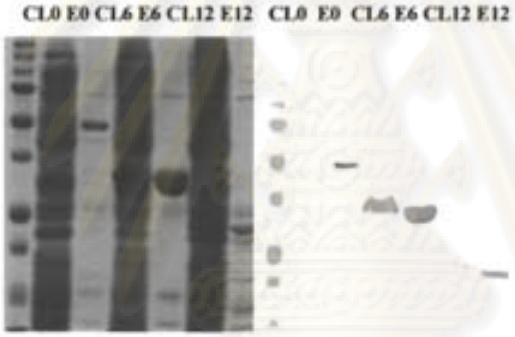


Figure 51. His-tag purification onto Ni-NTA agarose of GBD-0, GBD-6, and GBD-12. CL = cleared lysate; E = eluates.

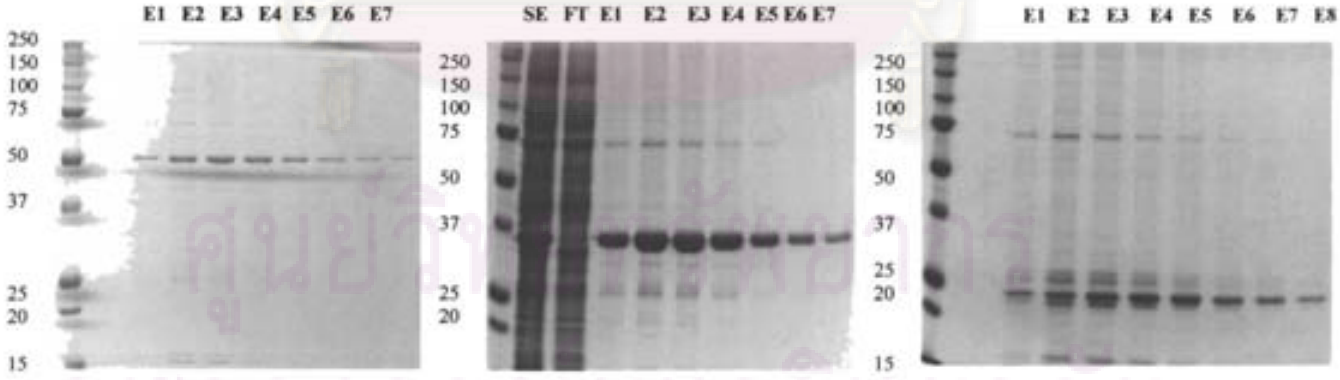


Figure 52. Elution pattern of purification by Ni-NTA agarose for GBD-0 (left), GBD-6 (middle), and GBD-12 (right). SE=sonicated extract, FT=flow through, E=eluate.

1.5.2 Affinity electrophoresis

The purified fractions 7 and 8, for which the purity was estimated superior to 90% were used for the affinity electrophoresis test, as shown in Figure 53. The PAGE without dextran was simultaneously performed as a control. In comparison to the gel without dextran, retardation by dextran was significantly observed for GBD-0 and GBD-6 samples, whereas migration of GBD-12 was not affected by the presence of dextran, confirming the previous results obtained using the glucan-binding assay developed in microplates.

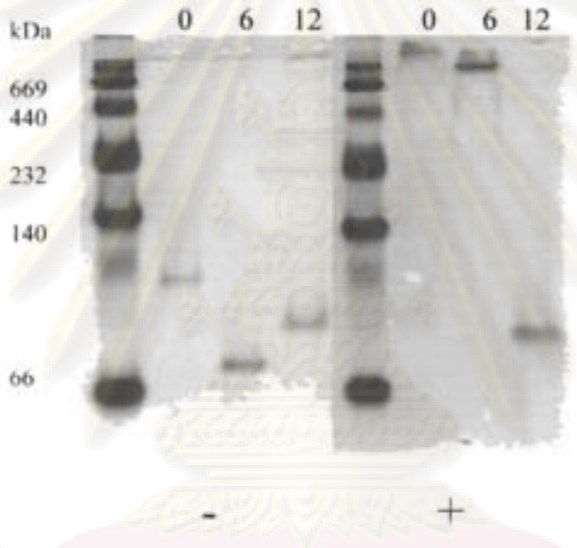
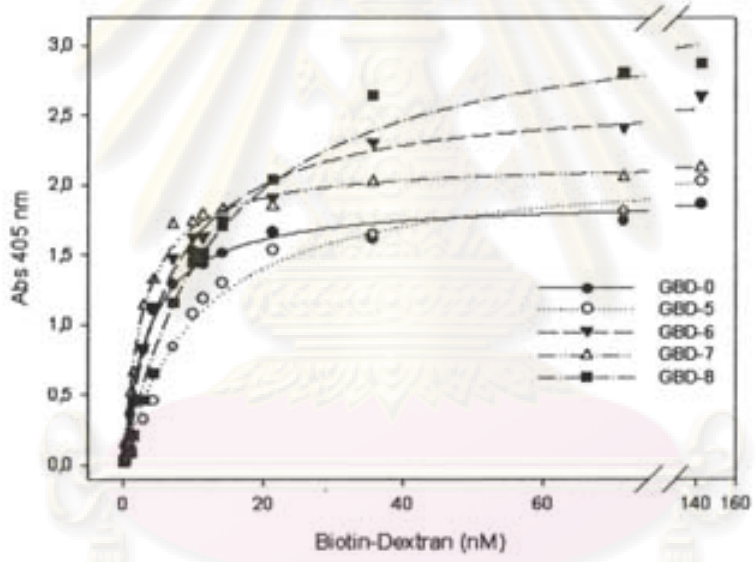


Figure 53. Affinity electrophoresis of GBD-0, GBD-6, and GBD-12. - = gel without dextran, + = gel with 0.1% of 2,000 kDa dextran.

1.6 Determination of binding affinity to dextran

In order to highlight the strength of binding between GBD and dextran, the dissociation constant *Kd* of GBD-0 and its derivatives were investigated. In a glucan-binding assay, an equivalent amount of GBD-0, GBD-5, GBD-6, GBD-7 and GBD-8 sonicated supernatants was incubated with various concentrations of labeled polymer. Using the one-site saturation ligand binding equation

$Abs_{405nm} = \frac{B_{max} \times [Biotin-dextran]}{K_d + [Biotin-dextran]}$, the K_d values obtained are ranging from 2.8 10^{-9} to 13.5 10^{-9} M (Figure 54), which is particularly high for a protein-ligand interaction. In comparison, the C-terminal domain of GTF-I from *S. sobrinus* has the K_d of 11.2 x 10^{-7} M (Shah *et al.*, 2004), and that of GBP1 from the same species has the K_d of 3 x 10^{-7} M (Landale and McCabe, 1987). Then, the GBD of *L. mesenteroides* NRRL B-512F DSR-S seems to show an affinity for dextran 100 times higher than that of glucan-binding domain of other glucansucrases. However, the Kd values obtained from different approaches with native and recombinant protein might not be comparable. Especially, GBD-7 has the highest affinity for dextran (K_d = 2.79 10^{-9} M), and was kept for our further investigations.



	r^2 (nonlinear regression)	B_{max}	K_d (M)
GBD-0	0.9905	1.8984	3.60×10^{-9}
GBD-5	0.9902	2.1757	1.11×10^{-8}
GBD-6	0.9924	2.6653	6.70×10^{-9}
GBD-7	0.9910	2.1719	2.79×10^{-9}
GBD-8	0.9905	3.2908	1.35×10^{-8}

Figure 54. Determination of the dissociation constants K_d of the five GBD constructs following the one-site ligand binding equation.

In this study, we have confirmed the role of the C-terminal domain of DSR-S in the dextran binding, using the glucan-binding assay developed by Shah *et al.* (2002). Different approaches were then performed to ensure these results, showing the high specificity of this system. Many truncated forms of the C-terminal domain were constructed based on the YG motifs present in the primary structure. We identified that the determinants responsible for the dextran binding were located at the C terminal extremity of the GBD. Among them, a minimal motif of 14 kDa has been identified. This small peptide was further studied in order to investigate the possibility to use it as an affinity tag for recombinant protein purification.

PART II
RECOMBINANT PROTEIN PURIFICATION BY GBD-DEXTRAN SUPPORT AFFINITY CHROMATOGRAPHY

Protein purification is a key step for studying proteins with the best accuracy and providing detailed biochemical and biological characterization. To separate a protein of interest from several thousand different protein contaminants may be a very difficult task. Many chromatographic purification methods have been developed to isolate a specific protein from complex mixtures on the basis of their physico-chemical properties such as their size, charge, hydrophobicity. These methods include size-exclusion, ion exchange and hydrophobic chromatography. However, the major limitation of these types of chromatography comes from the lack of specificity. Affinity chromatography overcomes this limitation. It is designed to purify a particular protein from a mixed sample by exploiting the unique interaction with a complementary binding molecule, an affinity ligand. Among the fractionation methods, affinity chromatography is one of the most effective and often allows protein purification in only single affinity step.

Nowadays, the most common use of affinity chromatography is for the purification of recombinant proteins. Indeed, genetic manipulation allows the gene encoding the protein of interest to be fused with a sequence encoding an affinity ligand. The fusion protein produced is tagged and can be directly applied onto an affinity column for fractionation. To date, a number of affinity tag systems are commercially available. Most of them allow (a) one-step adsorption purification, (b) minimal effect on tertiary structure and biological activity, (c) easy and specific removal of the tag to produce the native protein, (d) applicability to a number of different proteins (Terpe, 2003).

However the uses of these commercial tags also present some drawbacks: the affinity resins are sometimes quite expensive; low elution yield may be obtained with incomplete removal of contaminants (Keefe *et al.*, 2001; Clonis, 2006). In addition,

for some constructions, the fusion with a tag may not lead to protein correctly refolded. It is thus often necessary to try different fusions and select the most appropriate for the protein of interest. In this context, the development of a new tag enabling the fraction of proteins on very cheap support was quite challenging.

The use of GBD protein as an affinity tag was firstly proposed by Kaseda *et al.* (2001). They showed that the entire dextran-binding domain (DBD) of the glucosyltransferase GTF-I from *Streptococcus sobrinus* was able to bind onto commercially available cross-linked dextran. Fusion with both green and red fluorescent proteins suggested that this DBD could be used as a tag for purification with over 90% of purity. However, such DBD is of a large size (30 kDa) and displays a relatively low dissociation constant (2×10^{-5} M^{-1}).

In our present work, we have investigated the potential properties of GBD variant derived from the glucan-binding domain of the dextransucrase DSR-S from *L. mesenteroides* NRRL B-512F to be used as an affinity tag for recombinant protein purification. A high affinity of GBD-7 to dextran (2.8 nM) conjugated to a reversible binding to dextran (Part I Figure 41) have opened up the possibility to use the GBD-7 as the efficient affinity tag.

2.1 Affinity of GBD-7 towards commercial cross-linked dextran supports

Among the various truncated forms of GBD that were constructed in part I, GBD-7 was found to be the shortest motif able to bind dextran with a high affinity. Considering that a good affinity tag should be as short as possible to limit interaction with the protein of interest and possible misfolding, we selected GBD-7 for our affinity chromatography assays.

GBD-7 efficiently binds to biotin dextran and free dextran. However, to be used as affinity tag on commercial and cheap cross-linked dextran, we had first to demonstrate that it was able to bind this type of commercial support. Sephadex®, Sephacryl®, and Superdex® which are dextran support harbouring different degrees

of cross linking were purchased from Amersham Biosciences. These supports are reported to show a low non-specific adsorption of peptides. GBD-7 tagged protein was first purified by His-tagged purification onto Ni-NTA agarose resin to estimate with accuracy the column capacity. Equal amounts of GBD-7 proteins (297 µg) were added to the microcentrifuge tube containing 100 µl of each supports previously rinsed by PBS buffer pH7.2. After 2 h of contact at 4°C, unbound proteins were eliminated by successive washes, and bound proteins were eluted by displacement using 50 g/l of exogenous dextran of 1.5 kDa. The amount of GBD-7 tagged proteins adsorbed to 100 µl of each supports was quantitatively estimated (Table 14).

The highest elution yield (60%) was observed using Sephacryl® S300HR gel (Table 13, Figure 55). Generally, Sephacryl® and Superdex® gels showed higher yields of binding than that of Sephadex® gels. Sephacryl S300 HR, which is a copolymer of allyl dextran and N, N-methylenebisacrylamide was thus selected for further investigations. The cross-linking provides a good rigidity and chemical stability of the support while the dextran chain enables the GBD-7 to bind strongly.

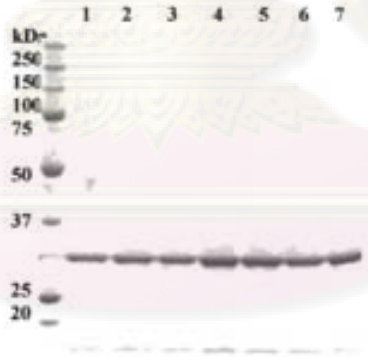


Figure 55. Purity of GBD-7 tag purification onto different dextran-containing supports. Lane1: SephadexG75, Lane2: SephadexG100, Lane3: SephadexG200, Lane4: SephacrylS300S, Lane5: SephacrylS300HR, Lane6: SephacrylS500HR, Lane7: Superdex 75.

Table 13. Binding capacity of GBD-7 fusion proteins previously purified by Ni-NTA affinity (protein concentration: 297 µg/ml) onto various commercial dextran supports.

Matrix	Protein content in flow-through and washes (µg)	Amount of proteins theoretically bound (µg)	Elution protein content (µg)	Elution yield (%)[4]	Total recovery yield (%)[5]
Sephadex[1]					
G75	129	168	85	50	29
G100	94	203	105	52	35
G200	96	201	92	46	31
Sephacryl[2]					
S300S	22	275	156	57	53
S300HR	25	272	163	60	55
S500HR	74	223	112	50	38
Superdex[3]					
G75	17	280	164	59	55

1 Sephadex® is a bead gel prepared by cross-linking dextran with epichlorohydrin
2 Sephacryl® is a cross-linked copolymer of allyl dextran and N, N-methylenebisacrylamide
3 Superdex® is a composite matrix of dextran and agarose
4 Elution yield: Amount of eluted proteins/bound protein
5 Recovery yield: Amount of eluted proteins/proteins added to the resins

2.2 Binding capacity of GBD-7

A maximum adsorption of GBD-7 onto Sephacryl S300HR was defined by addition of purified GBD-7 at different concentrations from 0.05 to 3.5 mg/ml. The GBD-7 proteins were prepared by His-tag purification, and concentrated using a spin filter of 10 kDa cut off until the defined concentration. The relationship between initial amount of loaded GBD-7 proteins and eluted GBD-7 proteins at the end of purification is illustrated in Figure 56. The maximal binding capacity of Sephacryl® S300HR was quantitatively estimated at 12 mg of GBD-proteins per ml of beads.

This binding capacity was comparable to that reported by Kaseda *et al.* (2001) for Sephacryl® S500HR, for which 14 and 5 mg of proteins were bound per ml of beads, for DBD and GTF respectively. In comparison with other affinity tag systems, that of Ni-NTA agarose (His tag) is of 12 mg proteins per ml of beads. The amylose

column yields 4.4 mg of proteins per ml of beads, that of immobilized streptavidin column yields 0.5 mg of proteins per ml of beads (Keefe *et al.*, 2001). Consequently, the capacity obtained is higher than that of most of the commercial supports.

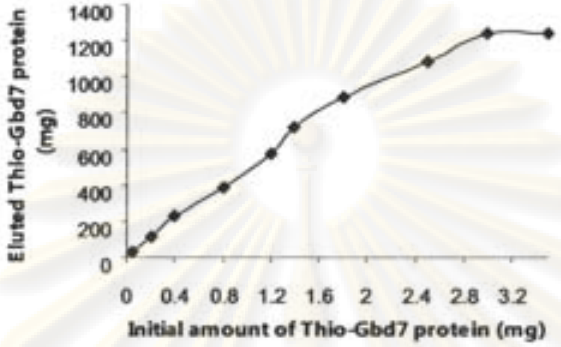


Figure 56. Maximum adsorption of GBD-7 proteins onto Sephacryl® S300HR.

2.3 Optimization of affinity chromatography using GBD-7 tag

2.3.1 Binding time

The contact time necessary to obtain a good interaction of GBD-7 with Sephacryl beads was investigated. Equal amounts of purified GBD-7 proteins were introduced in a test tube containing 100 µl of Sephacryl® S300HR. The amount of bound GBD-7 was assessed at 2 h, 4 h, and 16 h (Figure 57). The amount of bound proteins was quantitatively defined as the amount of initial protein minus the amount of protein recovered in the flow-through. Only 2 h of binding were sufficient to bind more than 95% of GBD-7 onto Sephacryl.

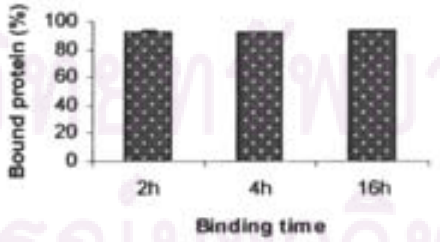


Figure 57. Effect of binding time of GBD-7 onto Sephacryl® S300HR matrix.

2.3.2 Elution step

To improve elution efficiency, both stepwise elutions and one-step elution have been investigated. For the one-step elution, 1 ml of 50 g/l of Dextran 1.5 kDa was added to the support and the mix was let stand for 2 h. For the stepwise elutions, the GBD-7 proteins were desorbed by incubating the same amount of support with 0.2 ml of 50 g/l of Dextran 1.5 kDa for 30 min and the procedure was repeated 7 times. Data from Table 14 shows that the multi-step elution was obviously better than the one single step elution. Two cycles of elution permitted a 55 % recovery of pure protein, which is equivalent to the amount of proteins eluted from the single step elution. The full elution program permitted to isolate approximately 85% of pure GBD-7 protein. SDS-PAGE corresponds to these elution steps was shown in Figure 58.

Table 14. One step and stepwise elution in GBD-7 purification onto 0.1 ml of Sephacryl S300HR gel.

Purification step	Volume (ml)	Amount of protein (μg)	Overall yield (%)
One-step			
Ni-purified	1	288.0	100
Flow-through	1	34.7	12.0
Wash	4	9.0	3.1
Eluate	1	158.8	55.1
			70.3
Stepwise			
Ni-purified	1	288.0	100
Flow-through	1	32.8	11.4
Wash	4	6.8	2.4
Eluate 1	0.2	151.6	52.6
Eluate 2	0.2	55.6	19.3
Eluate 3	0.2	22.0	7.6
Eluate 4	0.2	8.5	2.9
Eluate 5	0.2	5.1	1.8
Eluate 6	0.2	1.4	0.5
Eluate 7	0.2	1.1	0.4
			98.9

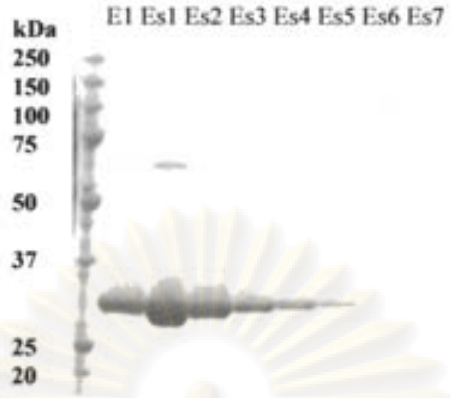


Figure 58. SDS-Page 10% Bis-Tris acrylamide gel showing one step and stepwise elution in GBD-7 purification onto 0.1 ml of Sephacryl S300HR gel. E1 = eluate from one-step elution, Es1-7 = eluates from stepwise elution using 50 g/l Dextran1.5 as an eluent.

2.3.3 Optimization of the size of dextran displacing agent

Isomaltoligosaccharides and dextrans of different sizes (Sigma) were used in order to determine their ability to displace GBD-7 bound proteins from the support. Approximately, yields of about 80% were obtained using dextrans from 1 to 2,000 kDa, no significant differences were observed depending on the dextran size (Figure 59). Interestingly, dextran of 1 kDa that mainly contains isomaltohexaose was as a good competitor as a high molecular weight dextran. In contrast, isomaltotriose which is composed of only 3 glucosyl units was not an efficient competitor. Because of its good ability to displace GBD-7 proteins, its easy removal by simple dialysis, and its reasonable cost, dextran of 1.5 kDa was chosen as the best eluting agent for the purification process.

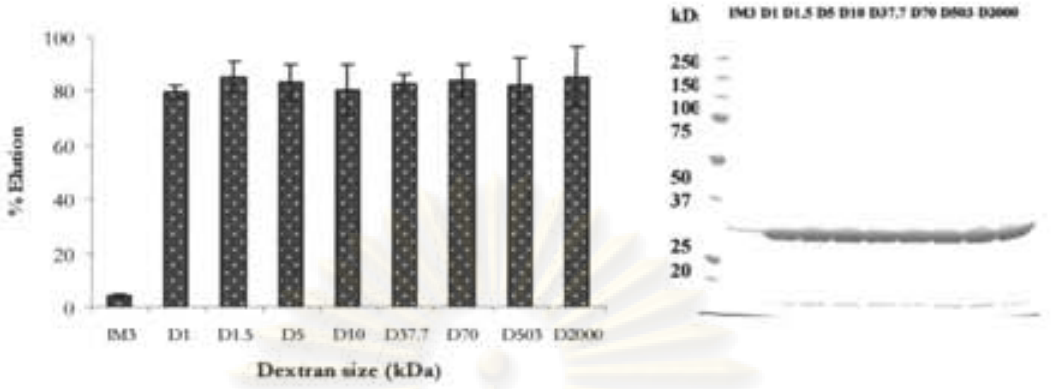


Figure 59. Elution efficiency of GBD-7-tagged protein using isomaltooligosaccharides and dextrans ranging from 0.4 to 2,000 kDa (10 g/l) as competitors.

2.3.4 Dextran concentration

Elution yields are clearly dependent on dextran (1.5 kDa) concentration in the eluent for concentrations ranging from 0.1 and 5 g/l (Figure 60), the maximum yield (80 %) being obtained when using a 5 g/L dextran concentration. The higher concentrations tested (10 to 50 g/l) did not give significant enhancements.

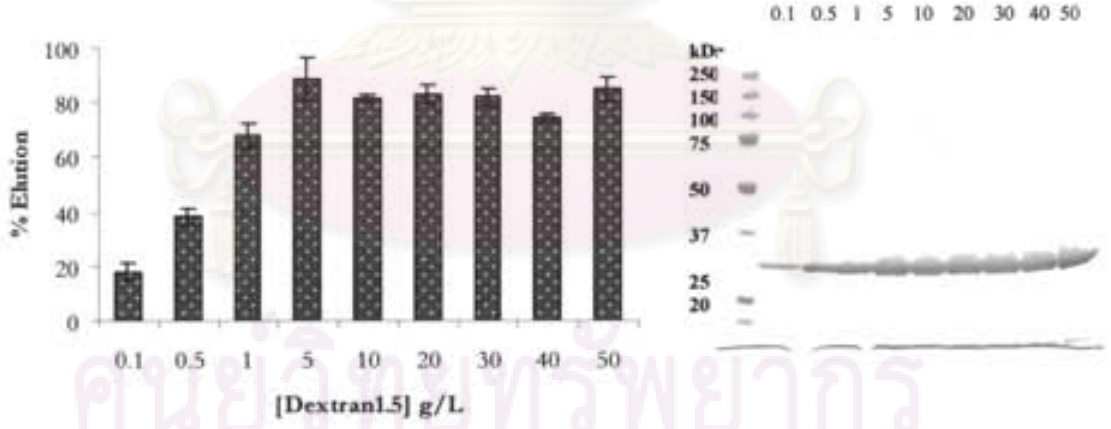


Figure 60. Effect of exogenous dextran concentration on elution efficiency, using dextran 1.5 kDa as displacing agent.

2.4 *Thioredoxin*-GBD-7-*His* affinity purification from cell lysate

In order to confirm the potentialities of this affinity tag system, protein purification was carried out from cell lysate. Thioredoxin was assigned as a model of study for GBD-7 tag purification. One ml of *E. coli* lysate containing *Thioredoxin*-GBD-7-*His* proteins was incubated with 0.1 ml of Sephacryl S300HR. Then, washes and elution with 5 g/l of dextran 1.5 kDa were performed following the protocol optimized (see above). The purity of eluted GBD-7 proteins was estimated at 95% from SDS-PAGE gel, after single step purification (Figure 61).

The Thioredoxin content present in the sonicated extract, flow-through, washes, and eluates was measured following the Hormgren's method (Hormgren, 1989) and the total protein content by Microbradford assay. An advantageous yield of purification of 58% was then determined with a purification factor of 6 (Table 15).

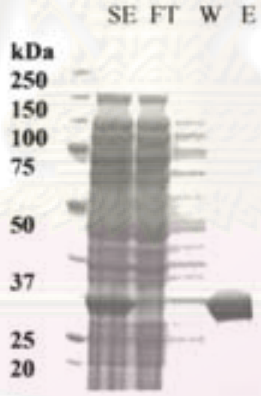


Figure 61. SDS-PAGE of *thioredoxin*-GBD-7-*His* purified onto Sephacryl® S300HR gels. SE = sonicated extract, FT = flow-through, W = pool of washes, E = pool of eluates.

Table 15. *Thioredoxin*-GBD-7-*His* fusion purification by affinity chromatography onto Sephacryl® S300HR gels.

Purification step	Volume (ml)	Overall purification factor	Yield of thioredoxin (%)
Sonicated	1	1.0	100±1.7
Flow-through	1	0.3	28±3.6
Pool of washes	4	2.0	17±5.0
Pool of eluates	1.4	5.9	58±3.6

2.5 Comparison of GBD-7 tag vs His tag affinity purification system

2.5.1 His tag purification of *Thioredoxin*-GBD-7-*His* proteins

Following protein production and extraction, *Thioredoxin*-GBD-7-*His* crude extracts were incubated with Ni-NTA agarose (Qiagen). The resin was then washed with potassium phosphate buffer containing 20-40 mM of imidazole to remove non-specifically bound proteins. The chemical structures of imidazole and histidine are illustrated in Figure 5. Imidazole itself can bind to the nickel ions and disrupt the binding to histidine residues of the His tag. The *Thioredoxin*-GBD-7-His proteins were eluted with phosphate buffer including 250 mM of imidazole, dialyzed overnight againt PBS buffer pH 7.2 and analyzed by SDS-PAGE and protein quantitation assays (Figure 62, Table 16). Purity of the eluted GBD-7 proteins was estimated to be over 95%. The concentration of purified *Thioredoxin*-GBD-7-*His* proteins ranges from 300-360 mg/l.

Figure 62. His tag purification of Thioredoxin-GBD-7-His proteins onto Ni-NTA agarose resins. SE = sonicated extract, FT = flow-through, W = pool of washes, E = pool of eluates.

Table 16. *Thioredoxin*-GBD-7-*His* fusion purification by affinity chromatography onto Ni-NTA agarose gels.

Purification step	Volume (ml)	Overall purification factor	Yield of thioredoxin (%)
Sonicated	1	1.0	100±9.0
Flow-through	1	0.3	25±2.9
Pool of washes	4	3.5	14±5.4
Pool of eluates	1	25.2	54±2.6

2.5.2 Comparison with GBD-7 tag purification

By comparing the GBD-7 tag purification system with the common *His* tag system, we observe that the purification yields are equivalent (Table 16, 17). However, they are much less contaminants in the purified fractions obtained by affinity binding through the GBD-7 tag. This is quite visible when we compare SDS-gels of purified proteins after staining with silver nitrate, a highly sensitive dye (Figure 63). These results indicate that GBD-7 tag is more specific than His-tag.

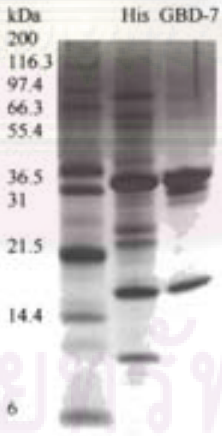


Figure 63. Purity of His tag and GBD-7 purification revealed by SDS-PAGE and SilverExpress staining kit.

2.6 Electroblotting of GBD-7 tag with biotin-dextran

Sonicated lysate and purified GBD-7 proteins were applied to SDS-PAGE gel, with sonicated extracts of GBD-13 proteins as a negative control. The proteins were then transferred to a nitrocellulose membrane, incubated with 3% of BSA as a blocking agent for 3 to 16 h in order to avoid non-specific binding. Then, 10 µg/ml of biotin-dextran were probed to the membrane for 30 min. The membrane was washed 3 times with PBST and probed with Extravidin-alkaline phosphatase for 30 min as a secondary antibody. After 3 washing steps, the GBD-7 tagged proteins were detected by adding BCIP/NBT phosphatase substrate. A second membrane probes with anti-His was simultaneously performed as control.

As shown in Figure 64, the biotin-dextran system can detect GBD-7 proteins, but not GBD-13 which has no dextran binding property. This electroblot with biotin-dextran has been investigated by Shah *et al.* (2004) for *L. mesenteroides*, *S. mutans* glucansucrases. However, neither the alternansucrase of *L. mesenteroides* NRRL B-1355 nor the dextransucrases from *L. mesenteroides* NRRL B-512F nor B-1299 bound to dextran in their study. Nevertheless, they worked with the wild-type enzymes. These enzymes show molecular weight of more than 160 kDa and the membrane transfer was probably less efficient. Here, we confirmed that GBD of DSR-S can be detected onto nitrocellulose membrane with biotin-dextran system. In comparison to anti-His antibody, the biotin-dextran used as an antibody is clearly cheaper.

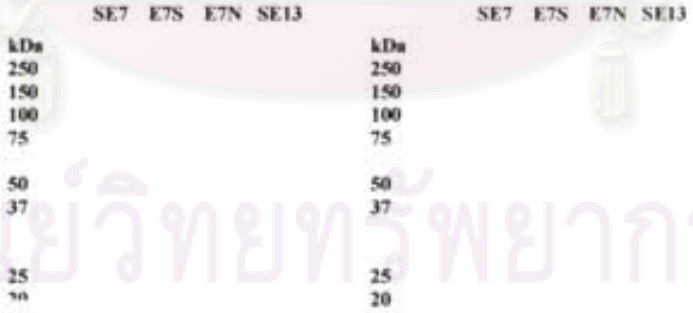


Figure 64. Western blot analysis with biotin-dextran (left) and anti-*His* (right). SE7=sonicated extract of GBD-7; E7S=eluate of GBD-7 proteins from

purification onto S300HR; E7N= eluate of GBD-7 proteins from purification onto Ni-NTA agarose; SE13=sonicated extract of GBD-13.

The opportunity to employ GBD-7 as an affinity tag for recombinant protein purification was demonstrated here. It was observed in these studies that the GBD-7 tag binds tightly to Sephacryl S300HR. Bounded GBD-7 proteins can be desorbed by displacement with soluble dextrans. For this purpose, dextran of 1.5 to 2000 kDa can be used. The elution yields of the purified form attain exquisite values above 80% with up to 98% of recovery.

In conclusion, this purification system permits to obtain high yields of purified proteins (over 95% of purity) in a short process. A second advantage of this approach is that GBD-7 tag can facilitate both purification and detection of recombinant proteins by Western-Blots, as biotin-dextran is commercially available. Moreover, Sephacryl S300HR gel can be cleaned using 0.2 M NaOH and autoclaved for sterilization. After cleaning and re-equilibration, this matrix can be re-used in order to reduce the cost of the purification process.

However, the size of GBD-7 (14 kDa) is relatively large in comparison with that of other commercial tags such as polyhistidine (0.84 kDa), or FLAG (1.01 kDa) tags for instance. Moreover, the tag in fusion to the protein of interest must be very well expressed, must not interfere in the solubility or in the folding of the fusion protein. In our construct GBD-7 was always produced with thioredoxin, to generalize our affinity protocol it is necessary to investigate the production of the tag alone and to further attempt a reduction of its size before testing it with other protein models. Considering that the limit of the rational approach had been reached in part I, we followed a combinatorial approach aiming at improving the expression and purification of the tag and also at reducing its size. Our strategy and results are presented in the following part.

PART III
EVOLVING GBD-7 TOWARDS A SMALLER SIZE AND A BETTER EXPRESSION

For enzyme structure-function relationship studies, a number of biochemical and structural experiments require highly expressed and soluble proteins easy to purify. For example, 5-20 mg/ml of purified proteins are necessary to prepare a crystal for X-ray crystallography (Santarsiero *et al.*, 2002; Mirarefi and Zukoski, 2004). However, heterologous expression of recombinant proteins in *E. coli* is not always efficient. Sometimes, the expression level is low and the protein forms inclusion bodies which are originated from protein misfolding (Baneyx and Mujacic, 2004). To overcome such obstacles, numerous strategies have been accomplished such as codon optimization (Wu *et al.*, 2004), directed evolution (Hart and Tarendeau, 2006; Yun *et al.*, 2006), and utilization of peptide tag (Davis, 1999). Peptide tags fused with the protein of interest are also very useful to facilitate purification by its affinity to a specific ligand. In addition, the presence of some peptide tags such as thioredoxin, maltose-binding protein (MBP), NusA at N-terminal of certain proteins was shown to enhance the expression and solubility of the protein partner (Eliseev *et al.*, 2004). Although, the effect of the tag on protein solubility is tightly linked to the couple formed by the tag and the protein partner, the outcome of the fusion with a tag can not be predicted (Kaseda *et al.*, 2001; Waugh, 2005).

In this context, increasing the choice of commercial affinity tags is of major interest for recombinant protein studies. As we have demonstrated that GBD-7 has a high potential to become an efficient affinity tag, the reduction of its size and the improvement of its expression and/or solubility level would make it more competitive. In addition, no structural data of DSR-S and its C-terminal domain are available. Enhancing expression and solubility of GBD-7 would enable us to attempt in parallel protein crystallization or NMR structural analysis.

In this part, we investigated the expression of *gbd-7* gene cloned into pBAD/TOPO TA vector that does not contain the thioredoxin encoding gene. The

solubility of the protein product was determined. Various strategies have been applied to increase the protein solubility, which include codon usage optimization and generation of randomly designed libraries of truncated GBD-7.

3.1 Expression of GBD-7 in the absence of *Thioredoxin* tag

Thioredoxin was well known to improve solubility and expression of a fusion partner. In the previous study, all GBD proteins were constructed in pBAD/TOPO Thiofusion which generated fusion protein of Thioredoxin tag at N-terminal and His tag at C-terminal of GBD proteins. In order to investigate the expression and solubility of GBD-7 itself, the *gbd*-7 gene was cloned into pBAD/TOPO TA vector (Figure 65). The components of this vector are similar to the former pBAD/TOPO Thiofusion vector except that pBAD/TOPO TA does not show the Thioredoxin tag at N-terminal end of the GBD protein.

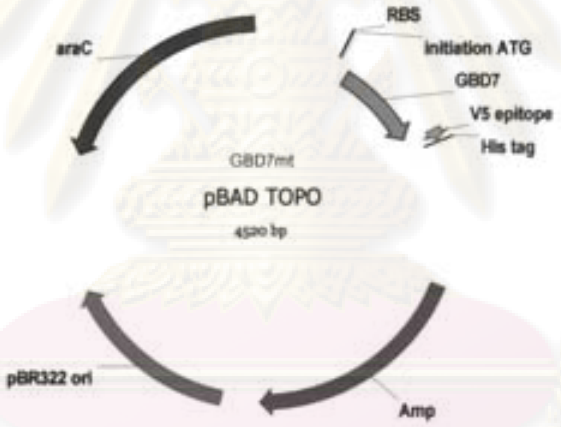


Figure 65. Construction of GBD-7mt in pBAD/TOPO TA vector.

The *gbd*-7 gene was amplified by PCR using primer GBD11F and RevGBD. Pool of PCR products (393 bp) was purified on a column kit after agarose gel electrophoresis. The PCR products were ligated into pBAD/TOPO TA vector and then, were transformed into chemically competent *E. coli* TOP10 cells. Positive clones were analyzed by restriction enzyme digestion and DNA sequencing.

The positive clone, called GBD-7mt, was grown in 2XYT pH 6.4 expression medium at various L-arabinose concentrations and temperatures (Figure 66A, 66B respectively). After cell lysis, both supernatant and pellet were applied to Western blot analysis using anti-His antibody. No protein band was observed at expected size of 18 kDa in both soluble and insoluble fraction at all conditions tested. A glucan binding test in 96-well Ni-NTA HisSorb plate was also performed on the supernatant obtained after expression in 2XYT pH 6.4 at 23 °C with 0.0002% - 0.2% of arabinose. However, no glucan binding signal was detected of GBD-7mt (Figure 66C).

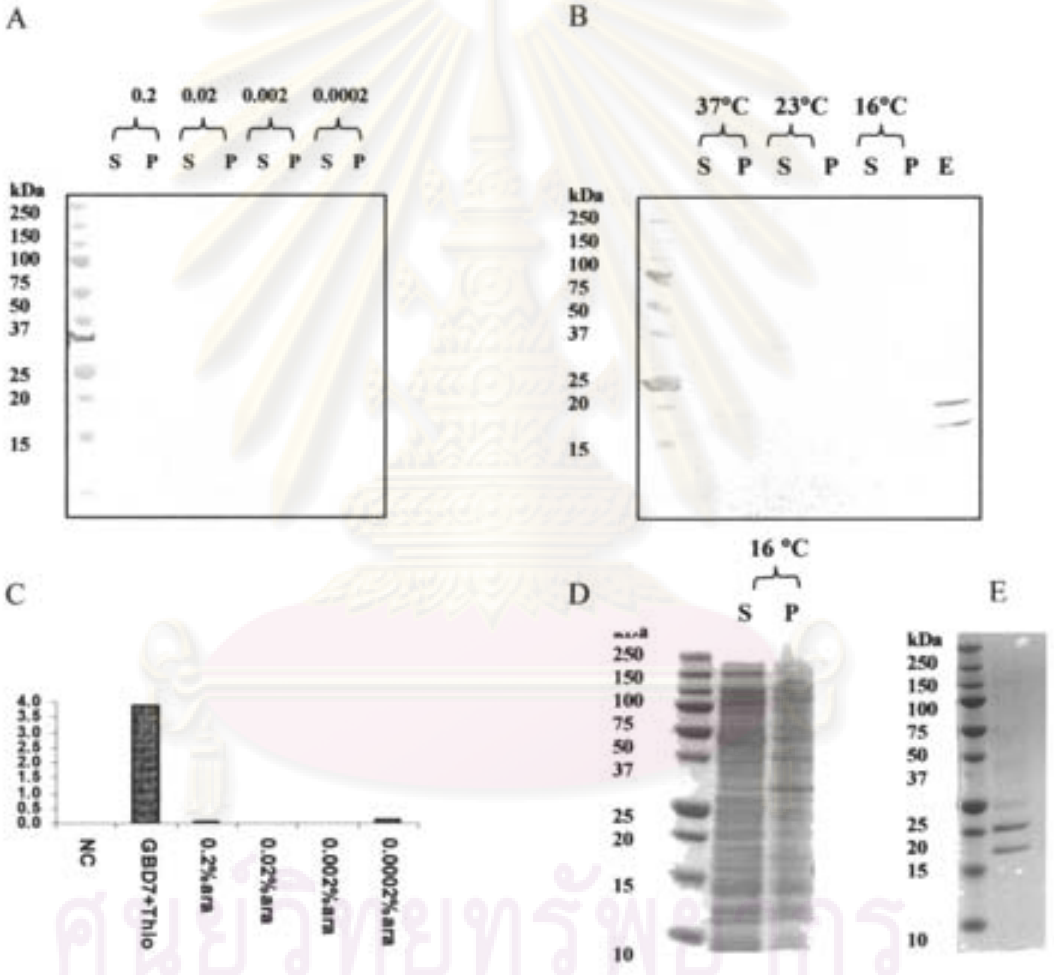


Figure 66. Expression patterns of GBD-7mt at different inducer concentration (A) and temperature (B). Neither protein bands (D) nor glucan-binding properties (C) were observed. After purification onto Sephacryl S300HR using 50 volume of sonicated extract (E), two protein bands were detected. Protein production was estimated at 0.17 mg/L culture.

The only successful mean was to apply 50 volumes of protein extract on Sephacryl S300HR (1 vol) and eluted with 0.4 volume of 5 g/l dextran 1.5 kDa (Figure 66B lane E, 66D). The GBD-7mt expression profile corresponds to two bands of 18 and 20 kDa, able to bind to dextran. The amount of proteins in the eluate was quantified by Microbradford assay, and estimated at 0.17 mg/l culture showing that the expression level is very low.

3.2 Strategy of the study

The goal of this study was both to improve the expression level of soluble GBD-7mt and to reduce the size of the GBD-7 tag. The expression level is sometimes limited by the presence of rare codons. To circumvent such problems, strains that encoded rare codons can be used and another solution is to modify the rare codons by mutagenesis. However, the codon usage is not always responsible for low expression level. The regions surrounding of the ribosome binding site and initiation codons and also those surround the termination have been demonstrated to be highly involved in the expression levels. To increase the protein expression and/or solubility, the presence of rare codons in the *gbd*-7 gene was investigated. The rare codons were first localized by sequence analysis and were then modified by site-directed mutagenesis into codons frequently used by *E. coli*, without changing the amino acid code. A second approach corresponded to the generation of deletion libraries of *Thioredoxin*-GBD-7-*His* proteins, named Ndel and Cdel libraries, in order to decrease the protein size. This method required the development of a new screen permitting to reveal positive clones expressing soluble and functional GBD proteins. The last approach is combination of the soluble expression improvement and protein size decrement. Libraries of truncated GBD-7-*His* proteins, called Ndelmt and Cdelmt, were created using the second approach. The overall strategy used in this study is illustrated in Figure 67.

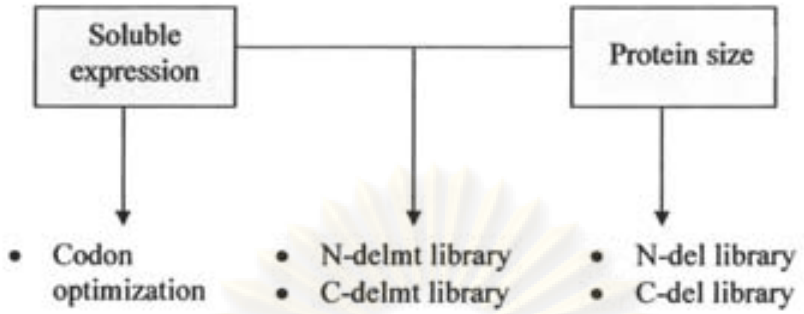


Figure 67. Soluble expression improvement and protein size reduction of GBD-7 strategy.

3.3 Codon optimization of *gbd-7* gene

Codon optimization is a technique that introduces a silent mutation of a rare codon into a frequently used codon in the host strain. From a total of 61 codons that encode 20 amino acids, each codon is unequally present in the gene, depending on the abundance of its correspondent tRNAs, for which the availability differs following the species studied (Gustafsson *et al.*, 2004). This phenomenon is well-known in term of codon bias. Effect of codon bias on heterologous protein expression occasionally leads to translational errors. Rare codons can make ribosomal stalling at a position requiring incorporation of amino acids coupled to minor tRNAs (McNulty *et al.*, 2003).

3.3.1 Search for rare codons represented in *gbd*-7 gene

Rare Codon Calculator (RaCC) online tool available at http://nihserver.mbi.ucla.edu/RACC was applied to determine the number of rare *E. coli* codons in *gbd-7* sequence.

RaCC results:

Red = rare Arg codons AGG, AGA, CGA

Green = rare Leu codon CTA

Blue = rare Ile codon ATA

Orange = rare Pro codon CCC

for the following input sequence: *gbd*-7

```
aat cag tat tat caa tta gca gat ggt aaa tat atg ttg tta gat gat agt
ggt cgt gcg aaa aca ggg ttt gta ttg caa gat ggt gta CTA aga tac ttc
gat caa aac ggt gag caa gtg aaa gat gct atc att gtg gat cca gat act
aac ttg agt tat tat ttc aat gca aca caa ggt gtc gct gta aaa aat gat
tat ttc gag tat caa ggt aat tgg tat tta aca gat gct aat tat caa ctt
atc aaa ggt ttt aaa gca gtt gac gac agc tta caa cat ttt gat gaa gtc
act ggt gta caa aca aaa gat agt gct tta ATA agt gct cag ggt aag gtt
tac caa ttt gat aat aat gga aat gct gtg tca gcc
```

From the RACC result, it was found that *gbd*-7 gene has 3 rare codons: cta, aga, and ata, which encode for leucine, arginine, and isoleucine respectively. The first two codons are located at the GBD-7 amino acids 31 and 32. The last ata codon is located at the amino acid 112 of the GBD-7.

3.3.2 Design for codon optimization

The two first rare codons are located near the N-terminal extremity and the third one is located at C-terminal end of the gene. Previous studies revealed that the presence of rare codons at the N-termimal more severely harmed the translation process than at the C-terminal of the gene, as it lead mRNA instability (Hoekema *et al*., 1987; Deana *et al*., 1998). We began to modify the first adjacent two codons by replacing the low frequently used codons with the more favored, following the codon usage in *E. coli* as shown in Table 17. The triplet code cta was altered to ctg and aga was changed to cgc, providing a silent mutation by site-directed mutagenesis. A pair of primers with overlap sequence named *CodLeArgF* and *CodLeArgR* (Table 2) was designed in order to introduce the mutations by inverse PCR (Figure 68). The mutant

clone containing the desired mutations was selected by DNA sequencing, and named 7LeuArg (Figure 69).

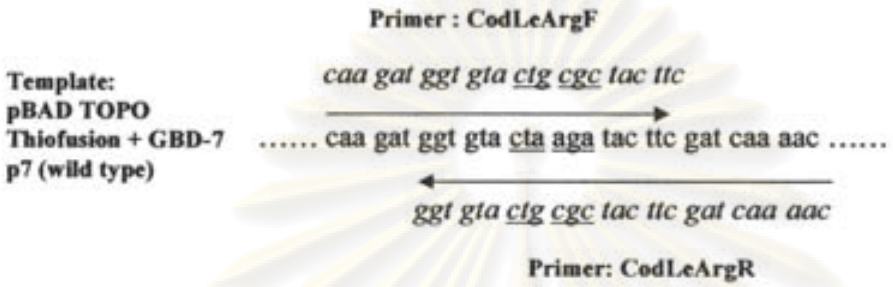


Figure 68. Primer design for changing codons encoding leucine and arginine by site-directed mutagenesis PCR.

Table 17. Codon usage in *E. coli* (Hénaut and Danchin, 1996)

Amino Acid	Codon	Class			Amino Acid	Codon	Class		
		I	II	III			I	II	III
Phe	ttt	55.09	29.08	67.14	Leu	ctt	9.70	5.56	19.00
	ttc	44.91	70.92	32.86		ctc	10.40	8.03	9.04
Leu	tta	10.99	3.44	20.09		cta	3.09	0.83	6.81
	ttg	13.02	5.47	15.05		ctg	52.79	76.67	29.99
Ser	tct	13.26	32.41	19.63	Pro	cct	13.71	11.23	28.30
	tcc	15.02	26.56	11.34		ccc	11.19	1.63	16.26
	tca	10.83	4.79	22.09		cca	18.63	15.25	31.50
	tcg	16.88	7.39	10.60		ccg	56.47	71.89	23.94
Tyr	tat	54.42	35.23	69.60	His	cat	56.80	29.77	61.69
	tac	45.58	64.77	30.40		cac	43.20	70.23	38.31
TER	taa				Gln	caa	33.40	18.65	37.06
	tag					cag	66.60	81.35	62.94
Cys	tgt	40.90	38.85	55.71	Arg	cgt	38.99	64.25	26.05
	tgc	59.10	61.15	44.29		cgc	42.23	32.97	21.94
TER	tga					cga	5.52	1.07	12.80
Trp	tgg	100.00	100.00	100.00		cgg	8.97	0.80	13.62
Ile	att	51.20	33.49	47.57	Val	gtt	23.74	39.77	34.33
	atc	44.37	65.94	26.65		gtc	22.48	13.45	18.95
	ata	4.43	0.57	25.78		gta	14.86	19.97	21.78
Met	atg	100.00	100.00	100.00		gtg	38.92	26.81	24.94
Thr	act	14.85	29.08	26.83	Ala	gct	14.52	27.54	22.86
	acc	46.83	53.60	24.45		gcc	27.62	16.14	23.67
	aca	10.52	4.67	27.93		gca	19.63	24.01	31.27
	acg	27.81	12.65	20.80		gcg	38.23	32.30	22.19
Asn	aat	40.87	17.25	64.06	Asp	gat	62.83	46.05	70.47
	aac	59.13	82.75	35.94		gac	37.17	53.95	29.53
Lys	aaa	75.44	78.55	72.21	Glu	gaa	68.33	75.35	66.25
	aag	24.56	21.45	27.79		gag	31.67	24.65	33.75
Ser	agt	13.96	4.52	18.73	Gly	ggt	32.91	50.84	31.79
	agc	30.04	24.33	17.61		ggc	43.17	42.83	24.51

Arg	aga	1.75	0.62	15.63		gga	9.19	1.97	24.75
	agg	1.54	0.29	9.96		ggg	14.74	4.36	18.95

Genes are clustered by using factorial correspondence analysis into three classes. Class I contains genes involved in most metabolic processes. Class II genes corresponds to genes highly and continuously expressed during exponential growth. Class III genes are implicated in horizontal transfer of DNA. One can see that the distribution of codons in class III genes is more or less even, whereas it is extremely biased in class II genes (in particular, codons terminated in A are selected against).

pBAD TOPO Thio + GBD-7 with mutation (7Leu-Arg)
CTA → CTG (Leu)
AGA → CGC (Arg)

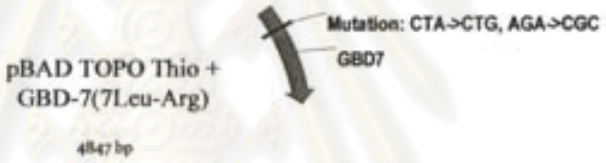


Figure 69. GBD-7 construction by cta and aga codon (7LeuArg) optimization.

Then, we used this plasmid p7LeuArg as a template to modify the last codon mutation from ata to att. Two overlap primers encoding this silent mutation were applied to an inverse PCR using a Phusion high fidelity DNA polymerase (Finnzyme) (Figure 70). After the PCR reaction, DNA template was digested by *Nco*I restriction enzyme for 1 h at 37 °C for digestion of template DNA to remove the methylated parental plasmids from PCR products. The PCR products were purified using a column kit, and transformed into *E. coli* TOP10 cells. Analysis of positive clones was accomplished by DNA sequencing. The mutant which contains the 3 mutations was labeled as 7Ile (Figure 71).

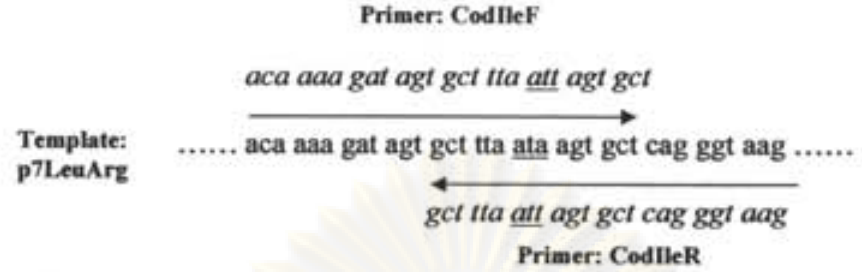


Figure 70. Primer design for changing isoleucine codon by PCR mediated site-directed mutagenesis.

pBAD TOPO Thio + GBD-7 with mutation (7Ile)
ATA → ATT (Ile)

Figure 71. GBD-7 Construction with ata codon (7Ile) optimization.

Moreover, in order to investigate the effect of rare codon on expression of GBD-7 in the absence of *Thioredoxin*, the *gbd*-7 gene containing the two first mutations, (*gbd*-7LeuArg) was cloned into pBAD/TOPO TA vector. After that, we performed transformation and analysis of positive clones. The new construct was named 7mtLeuArg (Figure 72).

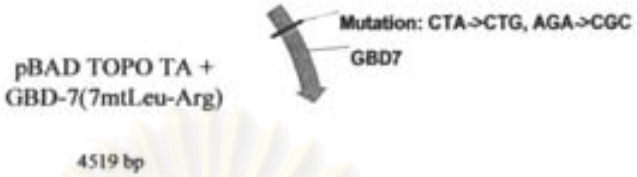


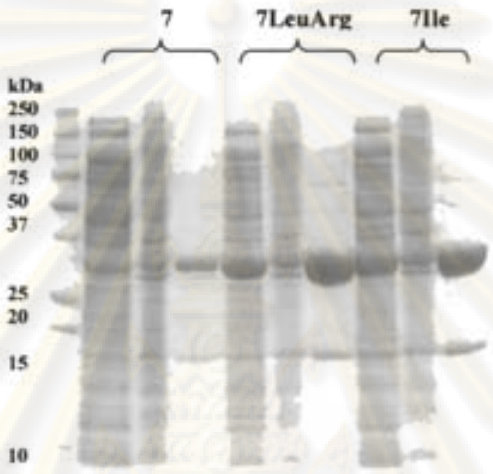
Figure 72. Construction of GBD-7 with optimized cta and aga codon and cloned into pBAD TOPO TA vector (7mtLeuArg).

3.3.3 Protein expression

The protein expression level of the different mutants containing both synthetic codons and wild-type genes were investigated. *E. coli* TOP10 cells carrying wild-type and synthetic *gbd*-7 were grown in 2XYT pH 6.4 at 30 ºC, 0.02% of arabinose. The cells were harvested when the OD_{600} reaches 2.5 and then centrifuged to collect the cells. After cell lysis, the supernatant and pellet fractions were analyzed by SDS-PAGE and colloidal blue staining (Figure 73). Interestingly, the mutants with optimized codons had a higher expression level than that of the control wild-type. The soluble fraction of each strain was purified onto Sephacryl S300HR. Amount of purified GBD-7 proteins were quantified by Microbradford assay. The data obtained showed that the production of GBD-7 was increased by 2.8 or 4.8 folds compared to the wild-type, when changing the codons encoding for lencine-arginine and for isoleucine, respectively (Figure 73).

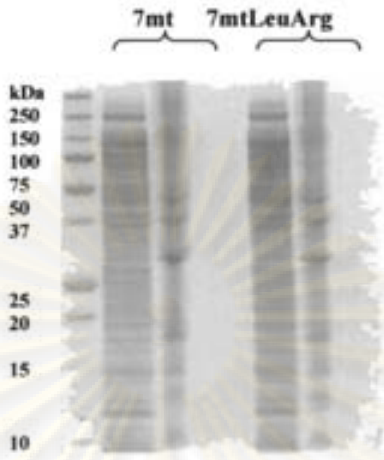
However, in the absence of N-terminal *Thioredoxin* tag, no expression of GBD-7mt and GBD-7mtLeuArg was detected on SDS-PAGE with colloidal blue staining (Figure 74) prior to concentration onto Sephacryl support. The protein band of these untagged constructs could only be revealed by SilverExpress staining, which is a highly sensitive protein detection system (Figure 75). This indicated that the poor

expression of GBD-7mt is probably not due to the codon bias. Other factors such as protein folding, mRNA stability or transcription rate might influence the lack of expression of GBD-7 without *Thioredoxin* tag. Moreover, the present of a 15-kDa protein band shows that the GBD-7 has been produced and binds efficiently without the thioredoxin. This indicates that it is properly folded.



Construct	7 (wild-type)	7LeuArg	7Ile
Protein production (mg purified protein/L culture)	8.5	24.3	40.8

Figure 73. Profile of GBD-7 expression and protein content after codon optimization.



Construct	7mt (wild-type)	7mtLeuArg
Protein production (mg Purified protein/L culture)	0	0

Figure 74. GBD-7mt expression profile after codon optimization.

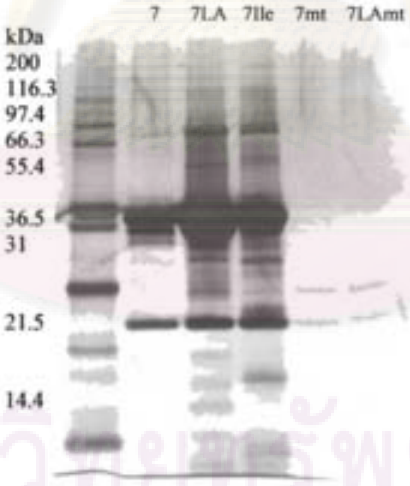


Figure 75. Purification of codon-optimized GBD-7. Each lane corresponds to purified proteins obtained from the GBD-Sephacryl S300HR affinity purification. The protein bands were revealed by SilverExpress staining kit (Invitrogen).

3.4 Directed evolution of GBD-7 by unidirectional deletion

Directed evolution is a powerful tool to improve enzymes or protein properties. This method is composed of 2 major steps which are (a) construction of the mutant library and (b) screening of positive mutants. Effective and reliable screening methods have to be developed to ensure the success of the method. In our study, thousands of truncated variants were generated and screened for their soluble expression and glucan binding ability. The positive clones were characterized in term of size and solubility.

3.4.1 High throughput immunoassay for screening of soluble GBD proteins

A novel high throughput screening for soluble GBD protein was developed. It is based on colony filtration blotting and glucan-binding assay (Cornvik *et al.*, 2005; Shah and Russell, 2002). As shown in Figure 76, our method begins by growing of GBD variant cells onto a Q-tray (22 x 22 cm) containing LBA agar. Following growth, the single colonies are rearranged in a 384-well plate with LB+8% glycerol by the QPixII robotic machine using the 'Picking' program. Parameters for colony picking were settled up in term of diameter, roundness, axis ratio, and overlap. Then, the plate was incubated at 30 ℃ with vigorous shaking overnight. After that, the variant cells were gridded onto a Durapore membrane with 0.22 μm pore size, which was previously laid on Omni-tray containing 2XYT pH 6.4 + 0.002% of L-arabinose. The hydrophilic Durapore membrane provides soluble protein separation from the insoluble fraction. The gridding step was carried out by the QPixII robotic machine using program 'Replicating'. The Omni-tray was incubated at 30 ℃ for 48 h in order to induce recombinant protein expression.

The colony filtration blot was used to isolate the soluble fraction of cell lysate by filtration pass through the Durapore membrane after cell lysis. Then, the proteins were immobilized onto the lower phase nitrocellulose membrane. Glucan-binding phenotype was detected by an immunoassay using biotin-dextran/Extravidin-alkaline phosphatase system on nitrocellulose membrane. It is preferable to use the horizontal shaker to maintain the interaction of those proteins.

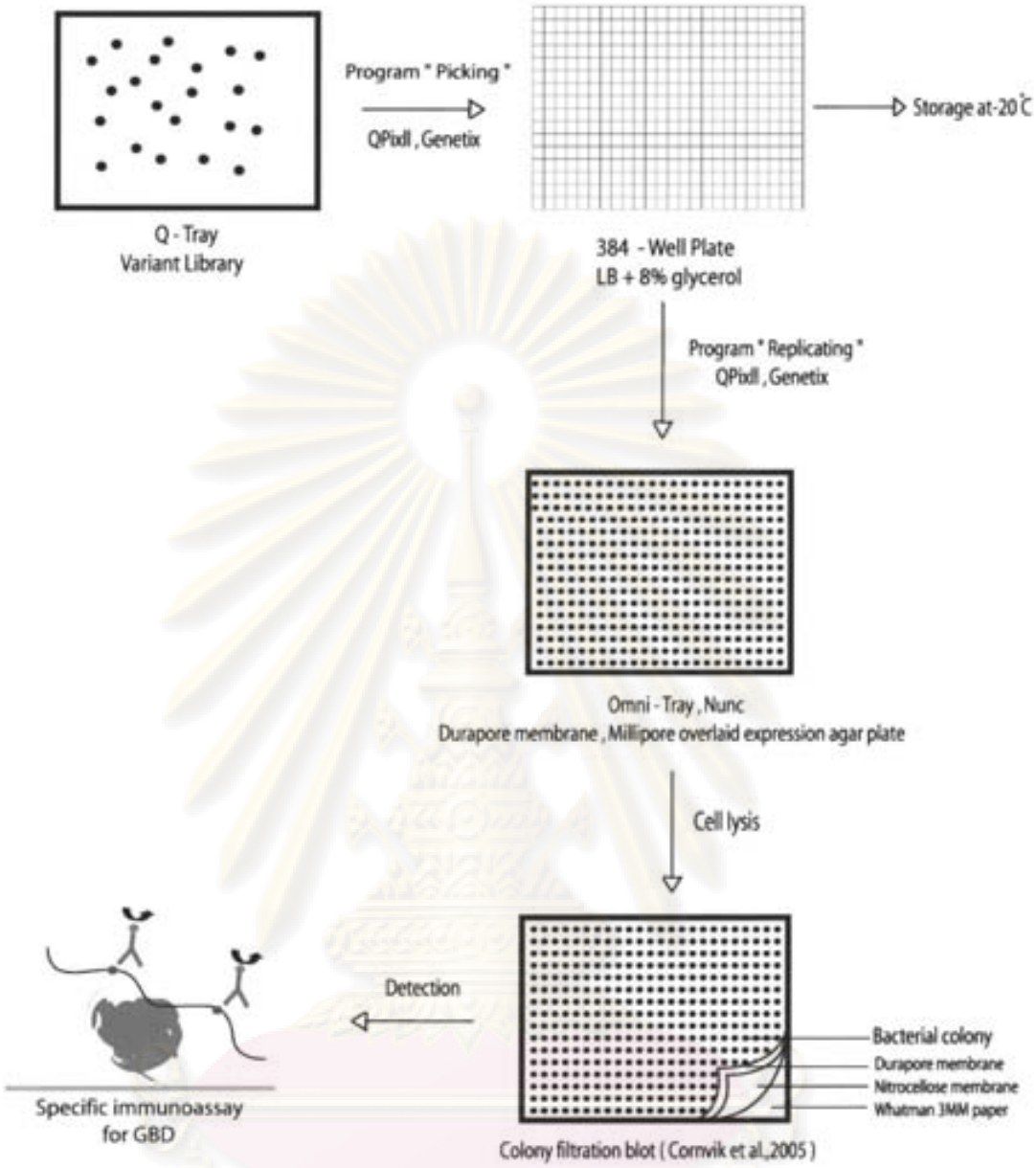
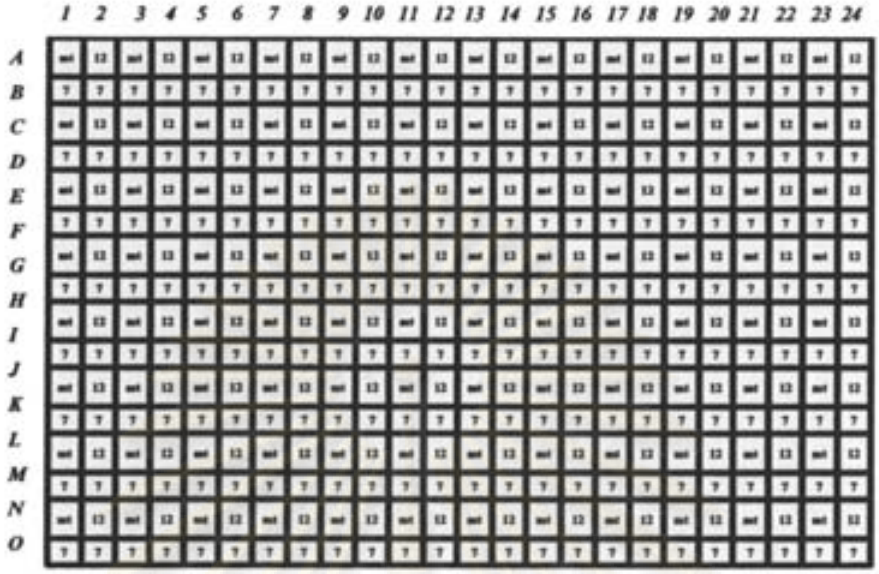


Figure 76. High-throughput screening for functional and soluble GBD proteins.

3.4.2 Evaluation of the screening method

Our screening method was analyzed in term of specificity and error using a standard 384-well plate containing clones expressing GBD-7, GBD-7mt, and GBD-12 as positive, negative expression, and negative dextran-binding control respectively. The pattern of the plates is showed in Figure 77A. Single colonies were picked and used to inoculate the 384-well plate containing LB+8% glycerol.

A



B

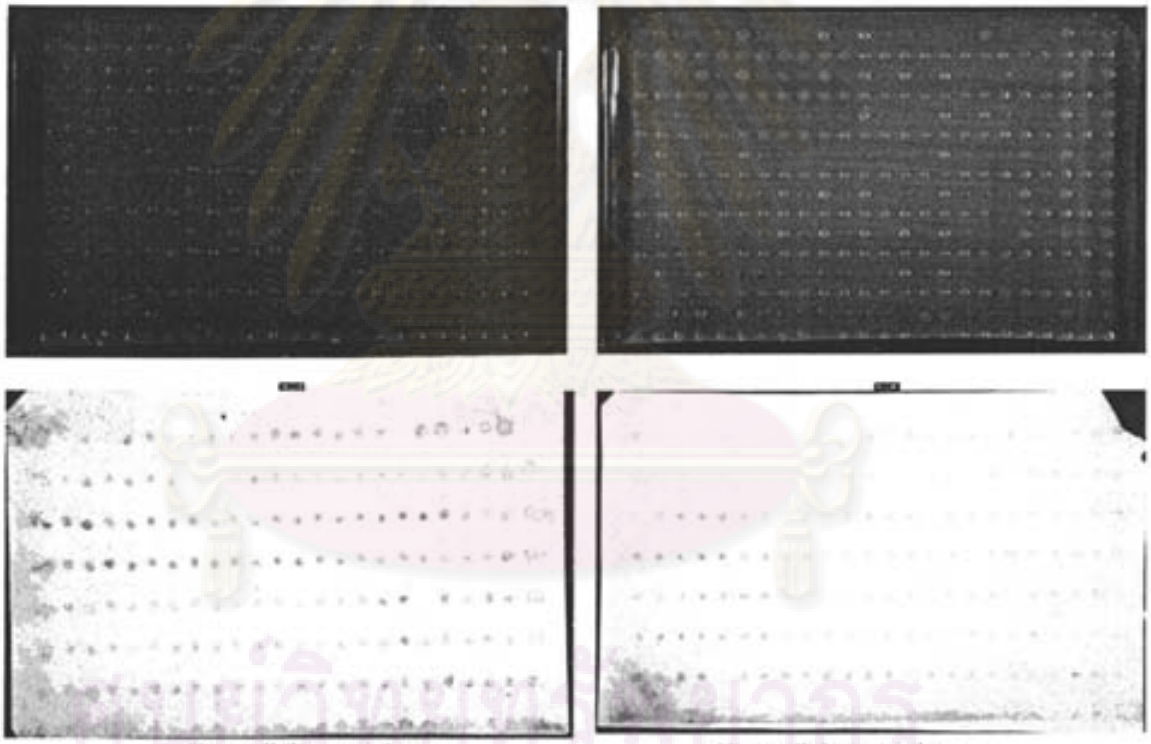


Figure 77. Culture plate pattern (A) and nitrocellulose membrane (B) of control plates containing clones encoding GBD-7 (positive soluble glucan binding protein), GBD-7mt (negative insoluble glucan binding protein), and GBD-12 (negative glucan binding proteins).

An average of 0 % of false positives and 3.8 % of false negatives were calculated from the screening, indicating the high specificity and very good accuracy of biotin-dextran/Extravidin-alkaline phosphatase detection system. This is in contrast to the screening by GFP fusions which had a relatively large number of false positives (Kawasaki and Inagaki, 2001). However, some patchy colonies were detected, but this can be avoided by a removal of excess buffer during the cell lysis.

3.4.3 Generation of nested deletion libraries

Unidirected truncatures of GBD-7 were constructed using the Erase-a-base system relied on ExonucleaseIII (ExoIII) and S1 nuclease digestion into plasmid, at specific positions. This approach was used to attempt the random reduction of GBD-7 size and the improvement the expression of soluble GBD-7.

3.4.3.1 Screening for truncated form of GBD-7

Two libraries of N- and C- terminal deletions of GBD-7 variants were generated from pBAD/TOPO Thiofusion + GBD-7 construct, in order to obtain truncated forms of *Thioredoxin*-GBD-7-*His* proteins. At the beginning, *Eco*RI and *Pst*I site were inserted in front of or behind the *gbd*-7 using a couple of primers containing the restriction sites. After the amplification, PCR products were cloned into pBAD/TOPO Thiofusion vector. Positive clones of N- and C-terminal deletion templates (NdelTP and CdelTP) were analyzed by restriction digestion and DNA sequencing. Figure 78 shows the construction of NdelTP and CdelTP. *Pst*I recognition site, that generated 3' protruding end, was located near the vector side to protect the digestion of ExoIII. In contrast, *Eco*RI which is a 5' overhang DNA producing enzyme, had been addressed just prior to or after the *gbd*-7 gene, permitting the gene deletion by ExoIII.

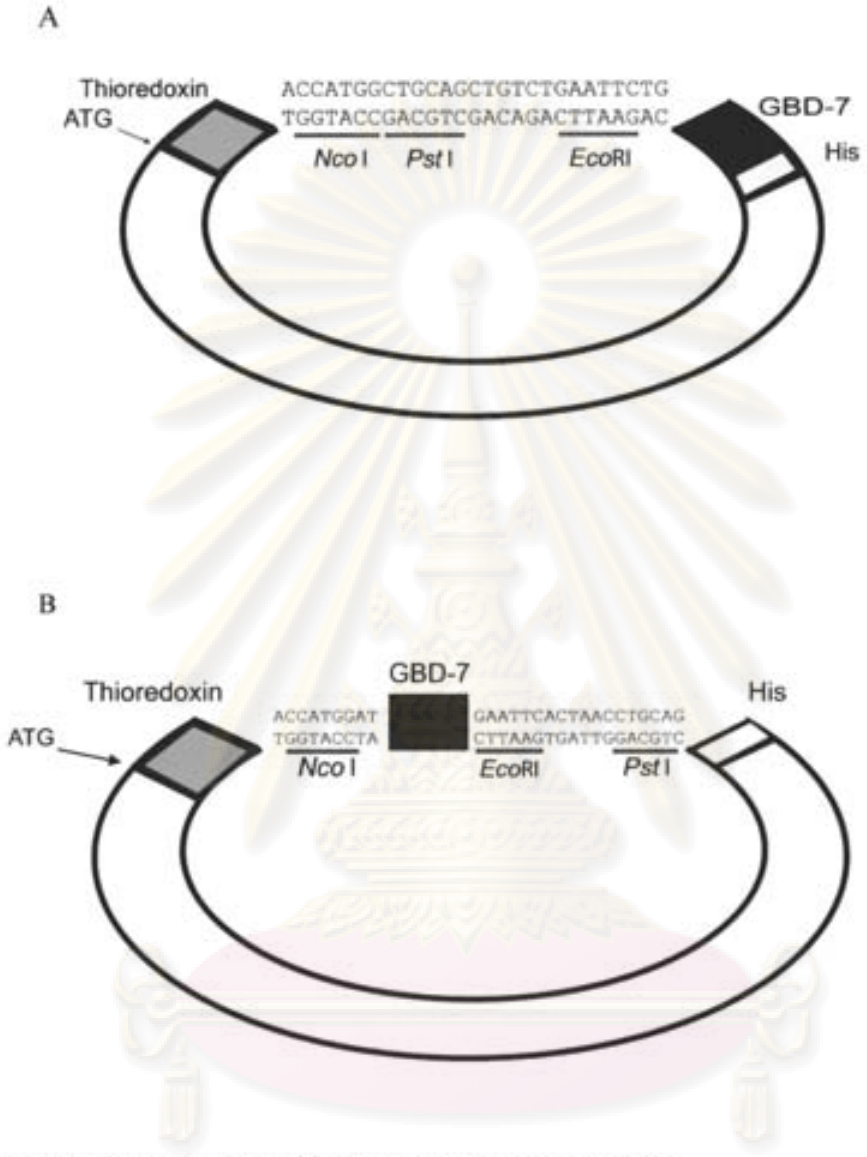


Figure 78. Construction of NdelTP (A) and CdelTP (B).

High quantities of NdelTP and CdelTP plasmids were prepared using the QIA maxi prep column kit. The plasmids were incubated with *Pst*I at 37 ºC until achieving complete digestion. *Pst*I-digested plasmids were purified on a column kit and then digested by *Eco*RI at 37 ºC. When digestion was completed, the *Pst*I- and *Eco*RI-digested plasmids were purified on the column kit. The ExoIII and S1 nuclease digestion were performed at 10 ºC for 15 min following the manufacturer's

instructions. Pool of digestions was transformed into *E. coli* TOP10 cells. In order to explore the digestion reaction, 12 colonies of C-terminal deletion library (Cdel) were randomly subjected to colony PCR, using primers specific for pBAD/TOPO Thiofusion vector, TrxF and pBADR (Table 10). The PCR products were visualized by agarose gel electrophoresis (Figure 79). From a full-length product (637 bp), 5 colony samples corresponded to fragments of size ranging from 500-637 bp (0-137 bp deletion). Fragments of 400-500 bp (137-237 bp deletion) were obtained from 4 samples. No amplified product was obtained for 3 colonies, showing that the deletion might harm over the complementary site of the primers.

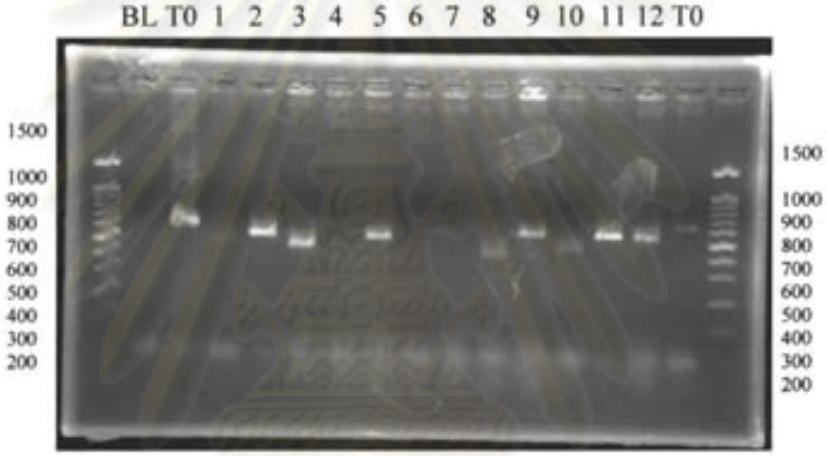
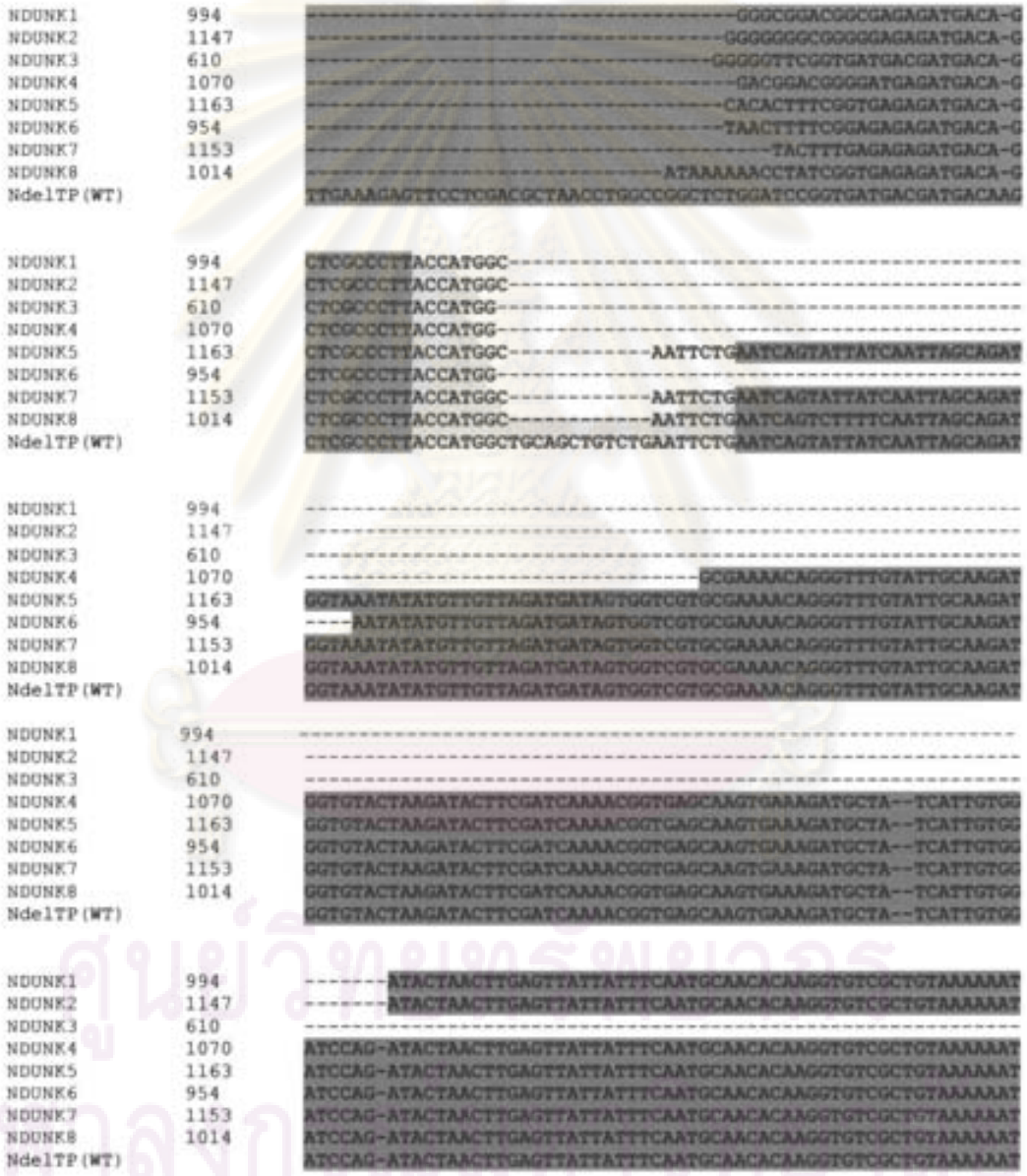


Figure 79. Colony PCR of Cdel libraries using primer TrxF and pBADR. BL= PCR reaction with no bacterial colony; T0=digestion control before ExoIII digestion providing a full-length *gbd*-7 gene.

Therefore, the ExoIII digestion condition was settled at 10 ºC, which can generate deletion fragments differing in size up to 16 bp per min. For Ndel and Cdel libraries, the ExoIII digestions were accomplished at 7.5, 15, and 30 min for creating variable truncated forms of *Thioredoxin*-GBD-7-*His* proteins. We screened 1,392 and 1,476 colonies of the N-del and C-del libraries, repectively. The DNA sequences of ten unknown variants were investigated in order to determine the truncation efficiency of the ExoIII (Figure 80). It was found that five unknown variants were effectively deleted at the *gbd-7* gene range from 28 to 272 bp by the ExoIII, in contrast to other three variants which were not digested by the enzyme. The

last two unknown variants could not be sequenced via primer TrxF. It might be because the complementary sequences of that primer were removed by the ExoIII. According to these results, fragments derived from this digestion protocol had the desired deletion about 50 % of total colony with a moderate variation in length of each construct.



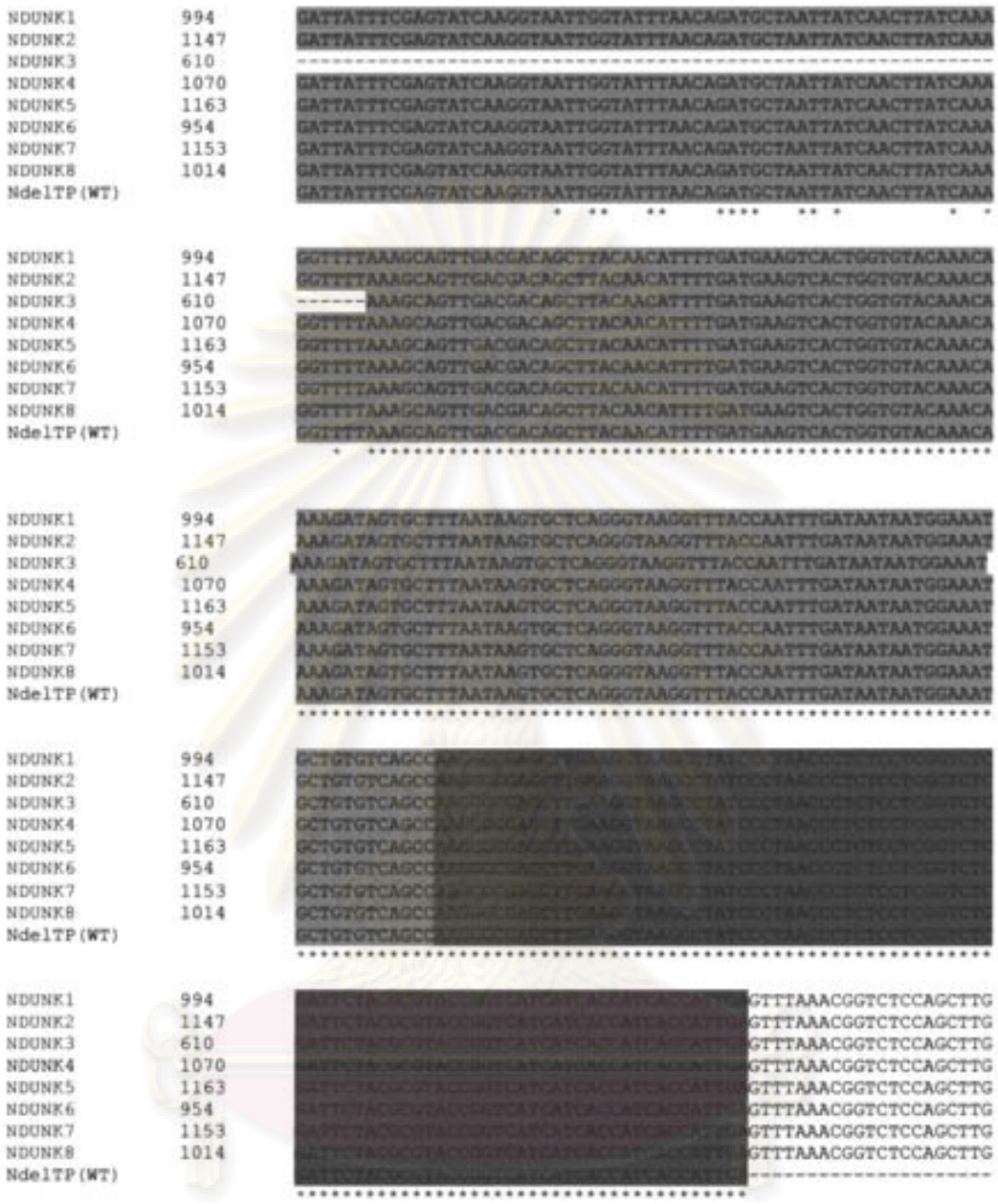


Figure 80. DNA sequence alignment of wild-type GBD-7 and its unknown variants from Ndel library by ClustalW 1.83. Highlight colors indicated different domains: Green=N-terminal Thioredoxin sequence, Yellow=restriction site, Blue=GBD-7, Red=C-terminal V5 epitope and His tag.

Table 18. Summary of soluble glucan-binding mutant screening.

Libraries	Total variants	CoFi blot (1st screening)	Deletion (colony PCR)	Deletion at GBD-7 side
C-terminal deletion (+Thio)	1476	29	15	0
N-terminal deletion (+Thio)	1392	6	2	1

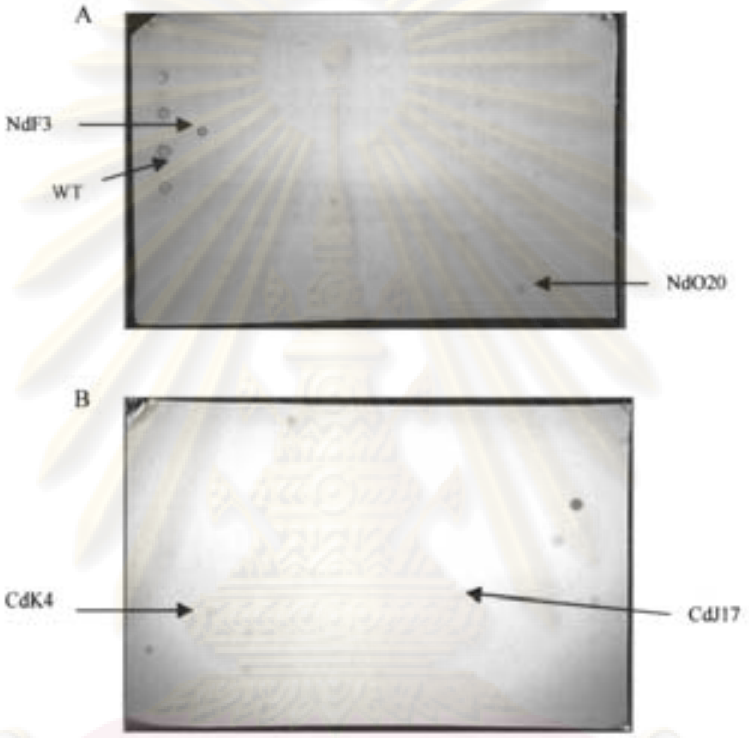


Figure 81. Positive variants from Ndel (A) and Cdel (B) libraries.

From the Ndel library, 6 colonies were found to produce soluble GBD proteins (Figure 81; Table 18), for which 2 colonies, NdO20 and NdF3, were reduced the size. In the Cdel library, 29 colonies expressed soluble GBD, among them 15 variants were shorted by ExoIII. Surprisingly, sequence analysis of these variants showed that all of them entertained deletions not in the gbd-7 gene, but in the pBAD/TOPO vector (Figure 82). Only one variant NdF3, presented deletion in the *gbd*-7 gene. We, therefore, selected 2 variants of each library for solubility test and protein production. Figure 83A showed deletion diagram of selected variants and their

ability to express the soluble GBD proteins. NdO20 and NdF3 variant contained deletions at the N-terminal sequence, which was in the Thioredoxin sequence. As suspected, deletions in the thioredoxin sequences caused significantly reduction of soluble expression level. We found that the variant proteins were more presented in the pellet than that in the soluble fraction (supernatant). It was cleared that Thioredoxin had an important role to enhance GBD-7 solubility as also mentioned in several investigations (Smyth *et al*., 2003; Tsunoda *et al*., 2005; Peisley and Gooley, 2007; Niiranen *et al*., 2007). For CdK14 and CdJ17 obtained from C-del library, deletion occurred at C-terminal vector region of the construction. Surprisingly, both variants dramatically increased their level of expression in supernatant and pellet, compared to the wild-type (Figure 83B). Truncation in the V5 epitope region may have a role in this difference. On the other hand, no evidence had been reported about a role of inhibition of the V5 epitope on recombinant protein expression.

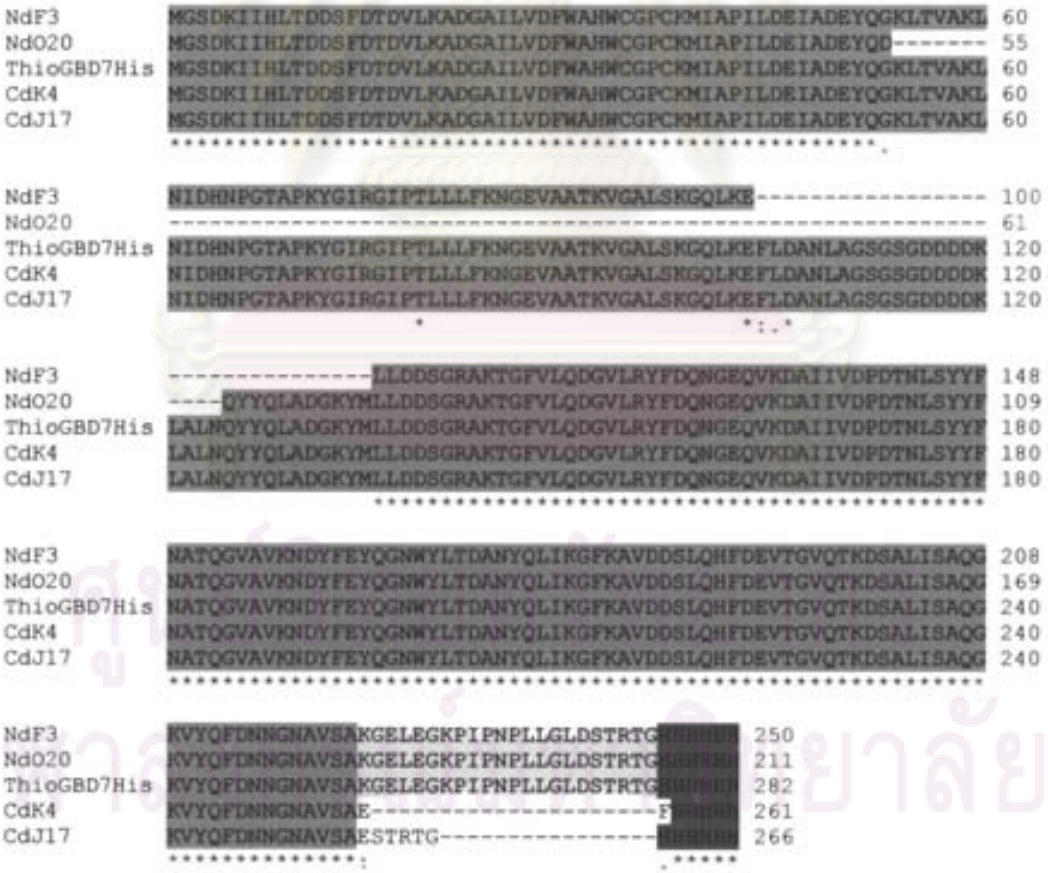


Figure 82. Sequence alignment of wild-type GBD-7 and its variants by ClustalW 1.83. Highlight colors indicated different domains: Green=Thioredoxin tag, Blue=GBD-7, Yellow=V5 epitope tag, Red=His tag.

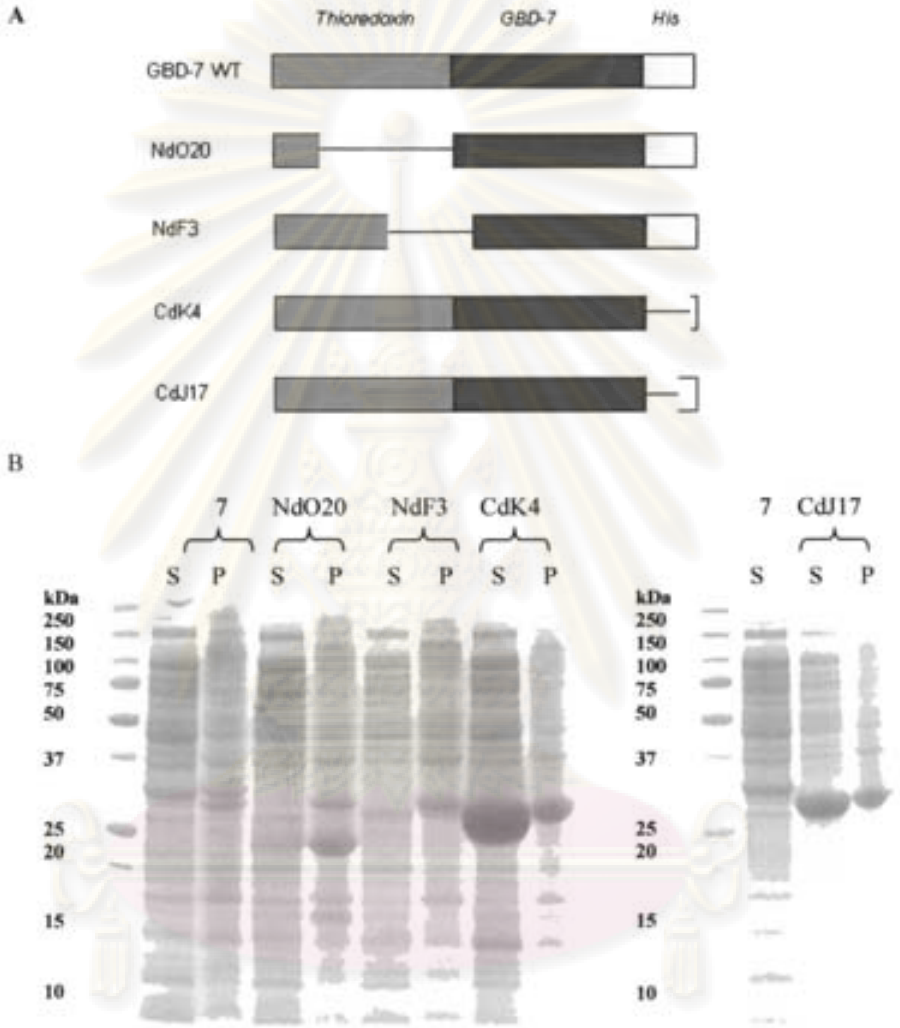
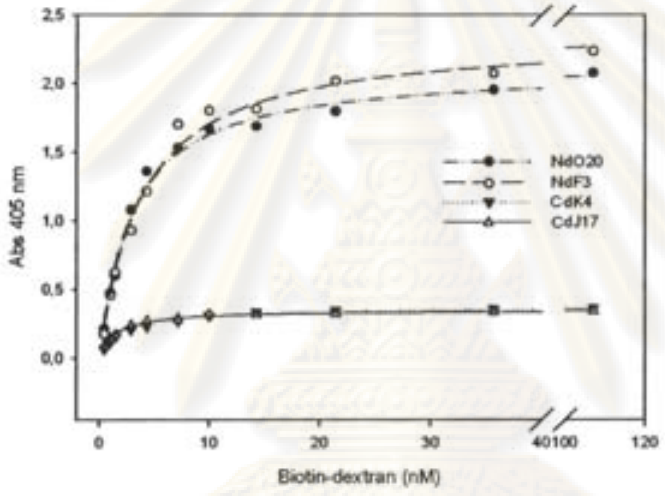


Figure 83. A: Feature map of positive deletion variants. Strength line corresponds to the amino acid deletions. B: Solubility test of all variants. Proteins were expressed at 30 °C and induced by 0.02% of L-arabinose. S=soluble fraction, P=pellet.

The dissociation constant of these variants was estimated following the method described in the part I, showing that they were in the same range than the wild-type (Figure 84). Truncation of the fusion partner (Thioredoxin or V5 epitope) of GBD-7 does not affect its affinity for dextran. Same result was observed for NdF3, for which the N-terminal gbd-7 gene was reduced of 36 bp and the N-terminal thioredoxin encoding gene was deleted for 60 bp. Thus, truncated form of GBD-7 was constructed in Ndel library, whereas no positive gene deletion clone was obtained from Cdel library. This confirmed the important role of the C-terminal end of glucan-binding domain.



	r^2 (nonlinear regression)	B_{max}	K_d (M)
NdO20	0.9912	2.1083	2.93×10^{-9}
NdF3	0.9902	2.3452	3.75×10^{-9}
CdK4	0.9929	0.3769	2.22×10^{-9}
CdJ17	0.9958	0.3475	1.36×10^{-9}

Figure 84. Determination of the dissociation constants K_d of the positive variants following the one-site ligand binding equation $Abs_{405nm} = \frac{B_{max} \times [Biotin-dextran]}{K_d + [Biotin-dextran]}$ **. The K_d value of wild-type GBD-7 is 2.79×10^{-9} M (Part I).**

3.4.3.2 Screening for soluble expression of protein

To improve the expression level of GBD-7 in the absence of thioredoxin tag (GBD-7mt), two libraries of GBD-7mt variants were generated, composed of nested N-terminal deletion (Ndelmt) and C-terminal deletions (Cdelmt). Parental plasmids were prepared from NdelTP and CdelTP by *Nco*I digestion for removal of thioredoxin sequence. The plasmid NdelmtTP and CdelmtTP were then treated with *Eco*RI and *Pst*I before preceding the nested deletion reaction.

The ExoIII deletion libraries were generated under same conditions with Ndel and Cdel libraries. Total variants of 2,446 and 1,615 colonies from the Ndelmt and Cdelmt were screened respectively (Table 19). Only one positive variant from the Cdelmt library was found by CoFi blotting, and named CdmtA4 (Figure 85). The variant CdmtA4 was checked for deletion by DNA sequencing and compared to wild-type (Figure 86). We found that the deletion was not located in the gene, but at the C-terminal V5 epitope and *His* tag sequences. This deletion caused a frame shift on the C-terminal fusion partner of the GBD-7 to EFTNQLGCFGG in stead of KGELEGKPIPNPLLGLDSTRTGHHHHHH which corresponded to V5 epitope and His tag fusion sequences. Similar to the result from Cdel library, the expression of variant CdmtA4 was significantly higher than that of wild-type, in both supernatant and pellet (Figure 87). Even though, the truncation was not arisen on the gene, the expression problem of GBD-7 without *Thioredoxin* (GBD-7mt) can circumvent.

Figure 85. Positive variants from Cdelmt library.

Table 19. Summary of screening for soluble glucan-binding mutants.

Libraries	Total variants	CoFi blot (1st screening)	Deletion (DNA sequencing)	Deletion at GBD-7 side
N-terminal deletion (-Thio)	2446	0	-	-
C-terminal deletion (-Thio)	1615	1	1	0

```
GBD-7       MGSGSGDDDDKLALNQYYQLADGKYMLLDDSGRAKTGFVLQDGVLRYFDQNGEQVKDAII
60
CdA4mt      MD----------NQYYQLADGKYMLLDDSGRAKTGFVLQDGVLRYFDQNGEQVKDAII
48
            *.          **************************************************

GBD-7       VDPDTNLSYYFNATQGVAVKNDYFEYQGNWYLTDANYQLIKGFKAVDDSLQHFDEVTGVQ
120
CdA4mt      VDPDTNLSYYFNATQGVAVKNDYFEYQGNWYLTDANYQLIKGFKAVDDSLQHFDEVTGVQ
108
            ************************************************************

GBD-7       TKDSALISAQGKVYQFDNNGNAVSAKGELEGKPIPNPLLGLDSTRTGHHHHHH 173
CdA4mt      TKDSALISAQGKVYQFDNNGNAVSAE-------FTNQLGCFGG---------- 144
            **************************:       :.* *  :..
```

Figure 86. Sequence alignment of wild-type GBD-7mt and CdmtA4 by ClustalW 1.83. Highlight colors indicated different domains: Blue=GBD-7, Yellow=V5 epitope tag, Red=His tag.

Figure 87. Comparison of protein expression and solubility of positive CdA4mt variant and the GBD-7mt wild-type. S=supernatant, P=pellet, E=eluate from purification onto Sephacryl S300HR.

3.4.3.3 Protein production of variants

Recombinant protein production of all variants from the 4 libraries was compared to that of the wild-type. Purification of variant proteins onto 0.1 ml of Sephacryl S300HR was performed to calculate their protein production at 30 °C, using 0.02% of L-arabinose. Elution was applied using 5 g/l of dextran 1.5 kDa and eluted fractions were analyzed by SDS-PAGE and Microbradford assay (Figure 88, Table 21). Considering the protein production level, deletion of the Thioredoxin at the N-terminal extremity reduces the level of production approximately 2 times compared to the wild-type. In contrast, deletion at the C-terminal extremity enhances the protein expression level, certainly promoted by a removal of V5 epitope. Especially for the best variant CdmtA4 for which the production level was up to 16 times greater than that of the wild-type 7mt.

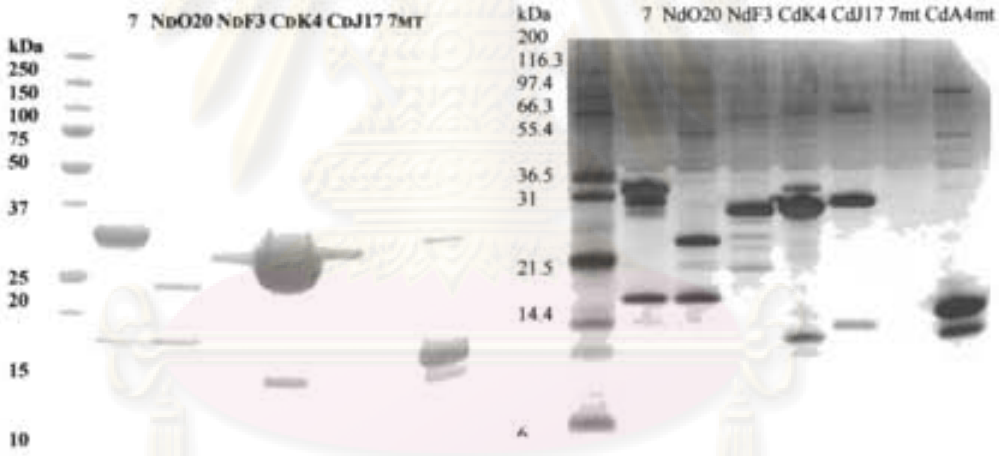


Figure 88. Purification of positive variants onto Sephacryl S300HR. Gels were stained by Colloidal blue (right) and SilverExpress (left).

Table 20. Protein size and production of positive variants at 30 °C.

Variants	7(wild-type)	NdO20	NdF3	CdK4	CdJ17	7mt (wild-type)	A4mt
Protein production (mg Purified protein/L culture)	8.5	3.4	3.9	108.3	26.7	0	15.7
Protein size (kDa)							
Full-length	30	24	29	29	29.5	20	15
Truncated formed	15	15	-	12	13	18	12

In conclusion, in order to develop GBD-7 as a powerful affinity tag, size and soluble expression of the protein is very important. Experiments on the GBD-7 production in the absence of Thioredoxin tag showed very low expression level. To overcome these problems, codon optimization and directed protein evolution strategies were carried out. In *E. coli*, codon optimization typically increased the expression level for 1.2-16 folds (Ejdeback *et al.*, 1997; Makoff *et al.*, 1989; Li *et al.*, 2002; Williams *et al.*, 1988). For GBD-7, its expression was successfully improved by approximately 4.8 folds of the wild-type after rare codons was optimized. Simultaneously, its expression was improved by incremental truncation approach. Only one round of evolution, the expression of the GBD-7 can increase up to 90 times of the wild-type.

We have also highlighted interesting results, as truncations at C-terminal V5 epitope part have large effect on soluble protein expression in with or without Thioredoxin tag, for instance. This might be due to a better protein folding when get rid the V5 epitope, the frame shift mutation may also cause a new amino acid sequence responsible for an increasing protein expression. Engineering of CdmtA4 with a stop codon at the end of the gene should be performed in order to have a better idea of the phenomenon involved here. Moreover, combination of the two strategies will give an advantageous expression level increase of the GBD-7 tag. The high soluble expression of the GBD-7 gained here is increase the feasibility of structural

studies either by NMR or X-ray crystallography to understanding glucan binding mechanism of this enzyme.

From N-terminal deletion libraries, the size of the GBD-7 can be reduced of only 12 amino acids, corresponding to about 1 kDa. However, in comparison with GBD-3, which can not bind to dextran, there are only 15 different amino acids between the CdF3 variant and this GBD-3. These amino acid sequences, L L D D S G R A K T G F V L Q, may have a crucial role in glucan-binding. In contrast to C-terminal libraries, no deletion in the *gbd-7* gene permitted to conserve the dextran binding property. As previously suggested the last N repeats may be involved in the glucan binding or maintain a correct folding. Protection of the vector sequence from ExoIII by 3' protruding end may inefficient, therefore the error on ExoIII digestion was consequently occurred. Another approach like α-phosphorothioate dNTP addition to 3' recessed end (Promega) should be investigated.

Finally, we have developed a new high-throughput screening for glucan-binding ability based on CoFi blotting and glucan-binding assay which is a highly specific to dextran binding. It might be applied for both screening of a fusion partner by used the GBD-7 as a reporter tag, and screening for protein with dextran binding property e.g. glucansucrase enzymes. In response to the growing needs of high-throughput methods, some screening protocols were developed to select the soluble proteins. For example, automated screening of His-tagged protein in 96-well plate (Vincentelli *et al.*, 2005), colony filtration blotting (Cornvik *et al.*, 2005). These methods, however, used anti-His which is relatively expensive. Compare to our methods, the screening for GBD employing biotin-dextran and Extravidin-alkaline phosphatase at dilute concentration which is cheap and could be suitable for high-throughput applications. One limitation of this screening is that it cannot distinguish between full-length and truncated proteins as well as degraded form unless the SDS-PAGE is performed.

CHAPTER V
CONCLUSIONS AND FUTURE PERSPECTIVES

1. Conclusions

In this study, C-terminal domain of a dextransucrase from *L. mesenteroides* NRRL B-512F (DSR-S) was searched for a minimal motif responsible for glucan binding. It was confirmed that the C-terminal domain (39.7 kDa) of DSR-S has the role in glucan binding like other glucansucrases by a glucan binding assay in Ni-NTA HisSorb 96-well plate. From fourteen truncated forms constructed in this study, four motifs were positive for the glucan binding. Sequence comparison of these positive constructs showed the common motif composed of YG_{10}-YG_{11}-YG_{12}-YG_{13}-YG_{14} while either motif YG_6-YG_7 or YG_{15} is required to provide the glucan binding ability at N or C-terminus respectively. Sequence alignment of these two repeats indicated the conserve sequences: FD(G/N)NG(H/N)AV which may interact with a glucan chain or facilitate the correct folding of the GBD. Furthermore, the minimal motif responsible for glucan binding is located between amino acids N1397 to A1527 of DSR-S (GBD-7). With a 14 kDa in size, the Kd of GBD-7 is 2.79×10^{-9} M indicating that the GBD-7 has very high affinity to dextran.

Potential application of GBD-7 tag for protein purification onto cross-linked dextran support was then carried out. The best support for the GBD-7 tag was Sephacryl S300HR that showed the highest bound proteins and elution yield. The binding capacity of the GBD-7 tag to the Sephacryl S300HR was estimated at 12 mg proteins per ml resin. Here, the purification protocol of the GBD-7 tag was optimized to be: binding time 2 h at 4 ℃, wash 15 min 4 times, eluted 15 min 7 times with 5 g/l of dextran 1.5 kDa. At this condition, purification from cell lysates using the GBD-7 tag was given approximately 58% yield of purification with over 95% purity observed by SDS-PAGE. Moreover, experimental design for GBD-7 specific detection onto nitrocellulose membrane was also developed. Using biotin-dextran/Extravidin system, the GBD-7 tagged proteins can be revealed on the membrane.

When engineered *gbd-7* gene into pBAD/TOPO TA vector (no thioredoxin at N-term), only low amount (0.17 mg/l culture) of the GBD-7-His proteins was expressed

at defined condition and two forms of the proteins were observed. Two strategies were performed in order to improve the soluble expression level of the GBD-7 proteins. One of them is codon optimization in which the expression level of the *Thioredoxin*-GBD-7-*His* proteins was increased up to 4.8 folds compared to the wild-type. However, the codon optimization did not enhance the expression of the GBD-7-*His* proteins. On the other hand, incremental truncation at C-terminal end of the GBD-7-*His* protein (CdA4mt) can improved the expression to 15.7 mg/l culture or 92 times compared to the wild-type. This result suggested that V5 epitope region between GBD-7 and His tag may inhibit the protein expression or promote the incorrect protein folding. Similar result was also observed in the *Thioredoxin*-GBD-7-*His* protein, deletion at the V5 epitope sequence can enhance the protein expression up to 108.3 mg/l culture or 13 folds of the wild-type. Finally, we can reduce the size of the GBD-7 by 12 amino acids or about 1 kDa without any change in glucan binding ability.

2. Future perspectives

Considering the GBD-7 tag affinity purification system, we only demonstrated fusion of the GBD-7 tag with small proteins, so it should be try to fuse with other large proteins or enzymes to obtain more information about this purification system. Moreover, engineering of a vector containing the *gbd-7* sequence with multiple cloning site and protease recognition site will provide easy fusion of the GBD-7 tag to other proteins.

In directed evolution experiment, the question of how deletion at C-terminus can enhance the GBD-7 expression is still not clear. Introduction of a stop codon at the end of the gene should provide a better idea about these results. Finally, we can enhance the soluble expression of the GBD-7 as well as its purification process. Therefore, it is highly recommended to investigate a three-dimensional structure of the GBD-7 by X-ray crystallography or NMR analysis for comprehensive understanding of its binding mechanism and further applications.

REFERENCES

Abo, H.; Matsumura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; and Kagawa, H. 1991. Peptide sequences for sucrose splitting and glucan binding within *Streptococcus sobrinus* glucosyltransferase (Water-Insoluble glucan synthetase) J. Bacteriol. 173: 989-996.

Arnau, J.; Lauritzen, C.; Petersen, G.E.; and Pedersen, J. 2006. Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expr. Purif. 48: 1-13.

Ashie, I.N.A. 2003. Bioprocess engineering of enzymes. Food Technol. 57: 44-51.

Ausubel, M.; Brent, R.; Kingston, R.E.; Moore, D.D.; Seidman, J.G.; Smith, J.A.; and Struhl, K. 1988. Current Protocols in Molecular Biology, vol. 1. Media, PA: John Wiley & Sons, Inc.

Baneyx, F.; and Mujacic, M. 2004. Recombinant protein folding and misfolding in *Escherichia coli*. Nature Biotechnol. 22: 1399-1408.

Banki, M.R.; and Wood, D.W. 2005. Inteins and affinity resin substitutes for protein purification and scale up. Microb. Cell Fact. 4: 32.

Beekman, J.M.; Cooney, A.J.; Elliston, J.F.; Tsai, S.Y.; and Tsai, M.J. 1994. A rapid one-step method to purify baculovirus-expressed human estrogen receptor to be used in the analysis of the oxytocin promoter. Gene 146: 285–289.

Binder, T.P.; Coté, G.L.; and Robyt, J.F. 1983. Disproportionation reactions catalyzed by *Leuconostoc* and *Streptococcus* glucansucrases. Carbohydrate Res. 124: 275-286.

Blanar, M.A.; and Rutter, W.J. 1992. Interaction cloning: identification of a helix-loop-helix zipper protein that interacts with c-fos. Science 256: 1014–1018.

Braun, P.; Hu, Y.; Shen, B.; Halleck, A.; Koundinya, M.; Harlow, E.; and LaBaer, J. 2002. Proteome-scale purification of human proteins from bacteria. Proc. Natl. Acad. Sci. U. S. A. 99: 2654 –2659.

Brizzard, B.L.; Chubet, R.G.; and Vizard, D.L. 1994. Immunoaffnity purification of FLAG epitope-tagged bacterial alkaline phosphatase using a novel monoclonal antibody and peptide elution. Biotechniques 16: 730–735.

Cadel S.; Gouzy-Darmon C.; Petres S.; Piesse C.; Pham V.L.; Beinfeld M.C.; Cohen P.; and Foulon T. 2004. Expression and purification of rat recombinant

aminopeptidase B secreted from baculovirus-infected insect cells, Protein Expr. Purif. 36: 19–30.

Carrard, G.; Koivula, A.; Soderlund, H.; and Béguin, P. 2000. Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. Proc. Natl. Acad. Sci. USA 10.1073/pnas.160216697v1.

Carrington, J.C.; and Dougherty, W.G. 1988. A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing. Proc. Natl. Acad. Sci. USA 85: 3391–3395.

Castle, L.A.; Siehl, D.L.; Gorton, R.; Patten, P.A.; Chen, Y.H.; Bertain, S.; Cho, H.J.; Duck, N.; Wong, J.; Liu, D.L.; and Lassner, M.W. 2004. Discovery and directed evolution of a glyphosate tolerance gene. Science 304: 1151-1154.

Cerning, J. 1990. Exocellular polysaccharides produced by lactic acid bacteria. FEMS Microbiol. Rev. 87: 113-130.

Chaga, G.; Bochkariov, D.E.; Jokhadze, G.G.; Hopp, J.; and Nelson, P. 1999. Natural poly-histidine affinity tag for purification of recombinant proteins on cobalt (II)-carboxymethylaspartate crosslinked agarose. J. Chromatogr. A 864: 257–256.

Chant, A.; Kraemer-Pecore, C.M.; Watkin, R.; and Kneale, G.G. 2005. Attachment of a histidine tag to the minimal zinc finger protein of the *Aspergillus nidulans* gene regulatory protein AreA causes a conformational change at the DNA-binding site. Protein Expr. Purif. 39: 152–159.

Chern, J. T.; and Chao, Y.P. 2005. Chitin-binding domain based immobilization of d-hydantoinase. J. Biotechnol. 117: 267-275.

Cirino, P.C.; Mayer, K.M.; and Umeno, D. 2003. Generating Mutant Libraries Using Error-Prone PCR. In Arnold, F.H.; and Georgiou, G. (eds.), Directed Evolution Library Creation Methods and Protocols. pp. 3-9. Totowa, NJ, Humana Press Inc.

Clonis, Y.D. 2006. Affinity chromatography matures as bioinformatic and combinatorial tools develop. J. Chromatography A 1101: 1-24.

Coco, W.M.; Levinson, W.E.; Crist, M.L.; Hektor, H.J.; Darzins, A.; Pienkos, P.T.; Squires, C.H.; and Monticello, D.J. 2001. DNA shuffling method for

generating highly recombined genes and evolved enzymes. Nature Biotechnol. 19: 354-359.

Colby, S.M.; and Russell, R.R.B. 1997. Sugar metabolism by mutans streptococci. J. Appl. Microbiol. Symposium Supplement. 83: 80S-88S.

Connelly, P.R.; Varadarajan, R.; Sturtevant, J.M.; and Richards, F.M. 1990. Thermodynamics of protein-peptide interactions in the ribonuclease S system studied by titration calorimetry. Biochemistry 29: 6108–6114.

Cornvik, T.; Dahlroth, S.L.; Magnusdottir, A.; Flodin, S.; Engvall, B.; Lieu, V.; Ekberg, M.; and Nordlund, P. 2006. An efficient and generic strategy for producing soluble human proteins and domains in *E. coli* by screening construct libraries. Proteins: Struc. Func. Bioinformatics 65: 266-273.

Coutinho, P.M.; and Henrissat, B. 1999. Carbohydrate-active enzymes: an integrated database approach. In Gilbert, H.J.; Davies, G.; Henrissat, V.; and Svensson, B. (eds.), Recent Advances in Carbohydrate Bioengineering, pp. 3-12. Cambridge: The Royal Society of Chemistry.

Croux, C.; and Garcia, J.L. 1992. Reconstruction and expression of the autolytic gene from *Clostridium acetobutylicum* ATCC 824 in *Escherichia coli*. FEMS Miocrobiol Lett. 74: 13-20.

Dahlroth, Li; Nordlund, P. ; Cornvik, T. 2006. Colony filtration blotting for screening soluble expression in *Escherichia coli*. Nature Protocol 1: 253-258.

Dang, C.V.; Resar, L.M.; Emison, E.; Kim, S.; Li, Q.; Prescott, J.E.; Wonsey, D.; and Zeller, K. 1999. Function of the c-Myc oncogenic transcription factor. Exp. Cell Res. 253: 63-77.

Darwin, C. 1859. On the origin of species by means of natural selection, or the preservation of favored races in the struggle for life. London: John Murray.

Davis, G.D.; Elisee, C.; Newham, D.M.; and Harrison, R.G. 1999. New fusion protein systems designed to give soluble expression in *Escherichia coli*. Biotechnol. Bioeng. 65: 382–388.

De Buck, E.; Lebeau, I.; Mellaert, L.V.; Geukens, N.; Anné, J.; and Lammertyn, E. 2004. The use of the cMyc epitope tag can be problematic for pretein detection in *Legionella pneumophila*. J. Microbiol. Methods 59: 131-134.

de Vries, E.G.E.; de Hooge, M.N.; Gietema, J.A.; and de Jong, S. 2003. Correspondence re: C.G. Ferreira et al., Apoptosis: Target of Cancer Therapy. Clin. Cancer Res., 9: 912.

Deana, A.; Ehrlich, R.; and Reiss, C. 1998. Silent mutations in the *Escherichia coli ompA* leader peptide region strongly affect transcription and translation in vivo. Nucleic Acids Res. 26: 4778–4782.

Dreher, M.L.; Gherardi, E.; Skerra, A.; and Milstein, C. 1991. Colony assays for antibody fragments expressed in bacteria. J. Immunol. Methods 139:197–205.

Duplay, P.; Bedouelle, H.; Fowler, A.; Zabin, I.; Saurin, W.; and Hofnung, M. 1988. Sequence of *mal*E gene and of its product, the maltose-binding protein of *Escherichia coli* K12. J. Biol. Chem. 259: 10606–10613.

Einhauer, A.; and Jungbauer, A. 2001. The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. J. Biochem. Biophys. Methods. 49: 455-465.

Einhauer, A.; Schuster, M.; Wasserbauer, E.; and Jungbauer, A. 2002. Expression and purification of homogenous proteins in *Saccharomyces cerevisiae* based on ubiquitin-FLAG fusion. Protein Expr. Purif. 24: 497–504.

Ejdeback, M.; Young, S.; Samuelsson, A.; and Karlsson, B.G. 1997. Effects of codon usage and vector-host combinations on the expression of spinach plastocyanin in *Escherichia coli.* Protein Expr. Purif. 11: 17-25.

Eliseev, R.; Alexandrov, A.; and Gunter, T. 2004. High-yield expression and purification of p18 form of Bax as an MBP-fusion protein. Protein Expr. Purif. 35: 206–209.

Evan, G.I.; Lewis, G.K.; Ramsay, G.; and Bishop, J.M. 1985. Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. Mol. Cell Biol. 5: 3610–3616.

Fatinas, E.T.; Bulter, T.; and Arnold, F.H. 2001. Directed enzyme evolution. Curr. Opi. Biotechnol. 12: 545-551.

Fernandez-Tornero, C.; Lopez, R.; Garcia, E.; Gimenez-Gallego, G.; and Romero, A. 2001. A novel solenoid fold in the cell wall anchoring domain of the pneumococcal virulence factor LytA, Nat. Struct. Biol. 8: 1020-1024.

Ferretti, J.J.; Gilpin, M.L.; and Russell, R.R.B. 1987. Nucleotide sequence of a glucosyltransferase gene from *Streptococcus sobrinus* Mfe28. J. Bacteriol. 169: 4271-4278.

Flores, H.; and Ellington, A.D. 2002. Increasing the thermal stability of an oligomeric protein, beta-glucuronidase. J. Mol. Biol. 315: 325-337.

Fonda, I.; Kenig, M.; Gaberc-Porekar, V.; Pristovaek, P.; and Menart, V. 2002. Attachment of histidine tags to recombinant tumor necrosis factoralpha drastically changes its properties. Sci. World J. 15: 1312–1325.

Fujii, R.; Nakagawa, Y.; Hiratake, J.; Sogabe, A.; and Sakata, K. 2005. Directed evolution of *Pseudomonas aeruginosa* lipase for improved amide-hydrolyzing activity. Protein Eng. Des. Sel. 18: 93-101.

Gaj, T.; Meyer, S.C.; and Ghosh, I. 2007. The AviD-tag, a NeutrAvidin/avidin specific peptide affinity tag for the immobilization and purification of recombinant proteins. Protein Expr. Purif. Doi:10.1016/j.pep.2007.06.010.

Germaine, G.R.; Chludzinski, A.M.; and Schachtele, C.F. 1974. *Streptococcus mutans* dextransucrase: requirement for primer dextran. J. Bacteriol. 120: 287-294.

Germaine, G.R.; Harlander, S.K.; Leung, W.L.S.; and Schachtele, C.F. 1977. *Streptococcus mutans*: functioning of primer dextran and endogenous dextranases in water-soluble and water-insoluble glucan synthesis. Infect. Immun. 16: 637-648.

Giffard, P.M.; and Jacques, N.A. 1994. Definition of a fundamental repeating unit in streptococcal glucosyltransferase glucan-binding regions and related sequences. J. Dent. Res. 73: 1133-1141.

Giffard, P.M.; Simpson, C.L.; Milward, C.P.; and Jacques, N.A. 1991. Molecular characterization of a cluster of at least two glucosyltransferases genes in *Streptococcus salivarius* ATCC 25975. J. Gen. Microbiol. 137: 2577-2593.

Gilmore, K.S.; Russell, R.R.B.; and Ferretti, J.J. 1990. Analysis of the *Streptococcus downei* gtfS gene, which specifies a glucosyltransferase that synthesizes soluble glucans. Infect. Immun. 58: 2452-2458.

Goel, A.; Colcher, D.; Koo, J.S.; Booth, B.J.; Pavlinkova, G.; and Batra, S.K. 2000. Relative position of the hexahistidine tag effects binding properties of a tumor-associated single-chain Fv construct. Biochim. Biophys. Acta 1523: 13–20.

Greenwood, J.M.; Gilkes, N.R.; Kilburn, D.G.; Miller, R.C. Jr.; and Warren, R.A.J. 1989. Fusion to an endoglucanase allows alkaline phosphatase to bind to cellulose. FEBS Lett. 244: 127-131.

Guan, D.C.; Li, P.; Riggs, P.D.; and Inouye, H. 1988. Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. Gene 67: 21–30.

Gustafsson, C.; Govindarajan, S.; and Minshull, J. 2004. Codon bias and heterologous protein expression. Trends Biotechnol. 22: 346-353.

Haas, W.; and Banas, J.A. 2000. Ligand-binding properties of the carboxyl-terminal repeat domain of *Streptococcus mutans* glucan-binding protein A. J. Bacteriol. 182: 728-733.

Hamada, S.; and Slade, H.D. 1980. Biology, immunology and cariogenicity of *Streptococcus mutans*. Microbiolo. Rev. 44: 331-384.

Hammarstrom, M.; Hellgren, N.; van den Berg, S.; Berglund, H.; and Hard, T. 2002. Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. Protein Sci. 11: 313-321.

Harayama, S. 1998. Artificial evolution by DNA shuffling. Trends Biotechnol. 16: 76-82.

Hart, D.J.; and Tarendeau, F. 2006. Combinatorial library approaches for improving soluble protein expression in *Escherichia coli*. Acta Cryst. 62: 19-26.

Hashimoto, M.; Ikegami, T.; Seino, S.; Ohuchi, N.; Fukada, H.; Sugiyama, J.; Shirakawa, M.; and Watanabe, T., (2000) Expression and characterization of the chitin-binding domain of chitinase A1 from *Bacillus circulans* WL-12. J. Bacteriol. 182: 3045–3054.

Hefti, M.H.; Caroline, J.G.; der Toorn, V.V.; Dixon, R.; and Vervoort, J. 2001. A novel purification method for histidine-tagged proteins containing a thrombin cleavage site. Anal. Biochem. 295: 180-185.

Henaut, A.; and Danchin, A. 1996. Analysis of prediction from *Escherichia coli* sequences, or *E. coli* in silico. In Neidhardt, F. C. et al. (eds.), *Escherichia coli* and Salmonella. Second Edition, Vol. 1 pp. 2047–2066. D. C., ASM PressWashington.

Henrissat, B.; and Davies, G.J. 1997. Structural and sequence-based classification of glycoside hydrolases. Curr. Op. Struct. Biol. 7: 637-644.

Henrissat, B.; and Romeu, A. 1996. Updating sequences-based classification of glucosyl-hydrolases. Biochem. J. 316: 695-696.

Hochuli, E.; Bannwarth, W.; Doebeli, H.; Gentz, R.; and Steuber, D. 1988. Genetic approach to facilitate protein purification of recombinant proteins with a novel metal chelate adsorbent. Biotechnology 6: 1321–1325.

Hochuli, E.; Dobeli, H.; and Schacher, A. 1987. New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues, J. Chromatogr. 411: 177–184.

Hoekema, A.; Kastelein, R.A.; Vasser, M.; and de Boer, H.A. 1987. Codon replacement in the PGK1 gene of *Saccharomyces cerevisiae*: Experimental approach to study the role of biased codon usage in gene expression. Mol. Cell Biol. 7: 2914-2924.

Holmgren, A. 1979. Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydrolipoamide. J. Biol. Chem. 254: 9627-9632.

Hope, T.P.: Gallis, B.: and Prikett, K.S. 1996. Metal-binding properties of a calcium dependent monoclonal antibody. Mol. Immunol. 33: 601–608.

Hopp, T.P.; Pricket, K.S.; Price, V.L.; Libby, R.T.; March, C.J.; Ceretti, D.P.; Urdal, D.L.; and Conlon, P.J. 1988. A short polypeptide marker sequence useful for recombinant protein identification and purification. Bio/Technology 6: 1204–1210.

Inoue, H.; Nojima, H.; and Okayama, H. 1990. High efficiency transformation of *Escherichia coli* with plasmids. Gene 96: 23-28.

Janknecht, R.; de Martynoff, G.; Lou, J.; Hipskind, R.A.; Nordheim, A.; and Stunnenberg, H.G. 1991. Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus. Proc. Natl. Acad. Sci. USA 88: 8972-8976.

Janson, J.-C.; and Rydén, L. 1998. Protein purification: principles, high resolution methods and applications. New York: John Wiley & Sons.

Johannes, T.W.; and Zhao, H. 2006. Directed evolution of enzymes and biosynthetic pathways. Curr. Opin. Microbiol. 9: 261-267.

Jones, W.T.; Harvey, D.; Kirk, C.; Rakonjac, J.; Sun, X.; Frearson, N.; and Samarrai, T.A. 2007. A novel peptide tag for detection and purification of recombinant expressed proteins. Protein Expr. Purif. 53: 404-410.

Kaboli, H.; and Reilly, P.J. 1980. Immobilization and properties of *Leuconostoc mesenteroides* dextransucrase. Biotechnol. Bioeng. XXII: 1055-1069.

Kaplan, W.; Husler, P.; Klump, H.; Erhardt, J.; Sluis-Cremer, N.; and Dirr, H. 1997. Conformational stability of pGEX-expressed *Schistosoma japonicum* glutathione S-transferase: A detoxification enzyme and fusion-protein affinity tag. Protein Science. 6: 399-406.

Kapust, R.B.; and Waugh, D.S. 1999. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci. 8: 1668-1674.

Kaseda, K.; Kodama, T.; Fukui, K.; and Hirose, K. 2001. A novel approach for purification of recombinant proteins using the dextran-binding domain. FEBS Lett. 500: 141-144.

Kataeva, I.A.; Seidel III, R.D.; Li, X.L.; and Ljungdahl, L.G. 2001. Properties and mutation analysis of the CelK cellulose-binding domain from the *Clostridium thermocellum* cellulosome. J. Bacteriol. 183: 1552-1559.

Kato C.; and Kuramitsu H.K. 1990. Carboxy-terminal deletion analysis of the *Streptococcus mutans* glucosyltransferase-I enzyme. FEMS Microbiol. Lett. 72: 299-302.

Kawasaki, M.; and Inagaki, F. 2001. Random PCR-based screening for soluble domains using green fluorescent protein. Biochem. Biophys. Res. Commun. 280: 842-844.

Keefe, A.D.; Wilson, D.S.; Seelig, B.; and Szostak, J.W. 2001. One-step purification of recombinant proteins using a nanomolar-affinity streptavidin-binding peptide, the SBP-tag. Protein Expr. Purif. 23: 440-446.

Kellerman, O.K.; and Ferenci, T. 1982. Maltose binding protein from *Escherichia coli*. Methods Enzymol. 90: 459-463.

Kim, D.; and Robyt, J.F. 1994. Production and selection of mutants of *Leuconostoc mesenteroides* constitutive for glucansucrases. Enzyme Microb. Technol. 16: 659-664.

Kim, K.M.; Yi, E.C.; Baker, D.; and Zhang, K.Y. 2001. Post-translational modification of the N-terminal His tag interferes with the crystallization of the wild-type and mutant SH3 domains from chicken src tyrosine kinase. Acta Crystallogr. D Biol. Crystallogr. 57: 759–762.

Kipriyanov, S.M.; Kupriyanova, O.A.; Little, M.; and Moldenhauer, G. 1996. Rapid detection of recombinant antibody fragments directed against cell-surface antigens by flow cytometry. J. Immunol. Methods 196: 51–62.

Kobayashi, M.; Yokoyama, I.; and Matsuda, K. 1986. Substrate binding sites of *Leuconostoc* dextransucrase evaluated by inhibition kinetics. Agric. Biol. Chem. 50: 2585-2590.

Kobayashi, S.; Koga, K.; Hayashida, O.; Nakano, Y.; and Hasegawa, Y. 1989. Glucan-binding domain of a glucosyltransferase from *Streptococcus sobrinus*: isolation of a 55-kilodalton peptide trypsin digest of glucosyltransferase prebound to insoluble glucan. Infect. Immun. 57: 2210-2213.

Kramer, A.; Keitel, T.; Winkler, K.; Stocklein, W.; Hohne, W.; and Schneider-Mergener, J. 1997. Molecular basis for the binding promiscuity of an antip24 (HIV-1) monoclonal antibody. Cell 91: 799–809.

Landale, E.C.; and McCabe, M.M. 1987. Characterization by affinity electrophoresis of an alpha-1,6-glucan-binding protein from *Streptococcus sobrinus*. Infect. Immun. 55: 3011-3016.

LaVallie, E. R.; DiBlasio, E. A.; Kovacic, S.; Grant, K. L.; Schendel, P. F.; and McCoy, J. M. 1993. A Thioredoxin Gene Fusion Expression System That Circumvents Inclusion Body Formation in the *E. coli* Cytoplasm. Bio/Technology 11: 187–193.

Lemieux, R.U. 1989. The origin of the specificity in recognition of oligosaccharides by proteins. Chem. Soc. Rev. 18: 347-374.

Li, Y.; Chen, C.X.; von Specht, B.U.; and Hahn, H.P. 2002. Cloning and hemolysin-mediated secretory expression of a codon-optimized synthetic human interleukin-6 gene in *Escherichia coli*. Protein Expr. Purif. 25: 437–447.

Li, Y.J.; Bi, L.J.; Zhang, X.E.; Zhou, Y.F.; Zhang, J.B.; Chen, Y.Y.; Li, Wei; and Zhang, Z.P. 2006. Reversible immobilization of proteins with streptavidin affinity tags on a surface plasmon resonance biosensor chip. Anal. Bioanal. Chem. 386: 1321-1326.

Lichty, J.J.; Malecki, J.L.; Agnew, H.D.; Michelson-Horowitz, D.J.; and Tan, S. 2005. Comparison of affinity tags for protein purification. Protein Expr. Purif. 41: 98-105.

Lis M.; Shiroza T.; and Kuramitsu H.K. 1995. Role of C-terminal direct repeating units of the *Streptococcus mutans* glucosyltransferase-S in glucan binding. Appl. Environ. Microbiol. 62: 2040-2042.

Lorimer, I.A.; and Pastan, I. 1995. Random recombination of antibody single chain Fv sequences after fragmentation with DNaseI in the presence of Mn^{2+}. Nucl. Acids Res. 23: 3067-3068.

Lu, Q.; Bauer, J.C.; and Greener, A. 1997. Using *Schizosaccharomyces* pombe as host for expression and purification of eukaryotic proteins. Gene 200: 135–144.

Lutz, S.; Ostermeier, M.; and Benkovic, S.J. 2001. Rapid generation of incremental truncation libraries for protein engineering using α-phosphothioate nucleotides. Nucl. Acids Res. 29(4): e16.

Maina, C.V.; Riggs, P.D.; Grandea, A.G. 3rd; Slatko, B.E.; Moran, L.S.; Tagliamonte, J.A.; McReynolds, L.A.; and Guan, C.D. 1988. An *Escherichia coli* vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. Gene 74: 365-373.

Makoff, A.J.; Oxer, M.D.; Romanos, M.A.; Fairweather, N.F.; and Ballantine, S. 1989. Expression of tetanus toxin fragment C in *E. coli*: high level expression by removing rare codons. Nucl. Acids Res. 17: 10191-10202.

Matsumura, M.; Izumi, T.; Matsumoto, M.; Tsuji, M.; Fujiwara, T.; and Ooshima, T. 2003. The role of glucan-binding proteins in the cariogenicity of *Streptococcus mutans*. Microbiol. Immunol. 47: 213-215.

McNulty, D.E.; Claffee, B.A.; Huddleston, M.J.; and Kane, J.F. 2003. Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli*. Protein Expr. Purif. 27: 365–374.

Mirarefi, A.Y.; and Zukoski, C.F. 2004. Gradient diffusion and protein solubility:use of dynamic light scattering to localize crystallization conditions. J. of Crystal Growth 265: 274–283.

Monchois V.; Reverte A.; Remaud-Simeon M.; Monsan P.; and Willemot R.M 1998. Effects of *Leuconostoc mesenteroides* NRRL B-512F dextransucrase gene carboxy-terminal deletions on dextran and oligosaccharide synthesis. Appl. Environ. Microbiol. 64: 1644-1649.

Monchois, V.; Remaud-Simeon, M.; Russell, R.R.B.; Monsan, P.; and Willemot, R.M. 1997. Characterization of *Leuconostoc mesenteroides* NRRL B-512F dextransucrase (DSRS) and identification of amino-acid residues playing a key role in enzyme activity. Appl. Microbiol. Biotechnol. 48: 465-472.

Monchois, V.; Willemot, R.M.; and Monsan, P. 1999. Glucansucrases: mechanism of action and structure-function relationships. FEMS microbiol. Rev. 23: 131-151.

Monsan, P.; and Paul, F. 1995. Enzymatic synthesis of oligosaccharides. FEMS Microbiol. Rev. 16: 187-192.

Monsan, P.; Bozonnet, S.; Albenne, C.; Joucla, G.; Willemot, R.M.; Remaud-Simeon, M. 2001. Homopolysaccharides from lactic acid bacteria. International Dairy J. 11: 675-685.

Mooser, G. 1992. Glycosidases and glycosyltransferases. In The enzymes. Vol XX, pp. 187-221. New York: Academic Press.

Moulis, C.; Arcache, A.; Escalier, P.C.; Rinaudo, M.; Monsan, P.; Remaud-Simeon, M.; and Potocki-Veronese, G. 2006b. High-level production and purification of a fully active recombinant dextransucrase from *Leuconostoc mesenteroides* NRRL B-512F. FEMS Microbiol. Lett. 261: 203-210.

Moulis, C.; Joucla, G.; Harrison, D.; Fabre, E.; Potocki-Veronese, G.; Monsan, P.; and Remaud-Simeon, M. 2006a. Understanding the polymerization mechanism of glycoside-hydrolase family 70 glucansucrases. J. Biol. Chem. 281: 31254–31267.

Munro, S.; and Pelham, H.R.B. 1986. An Hsp70-like protein in the ER: identity with the 78 kd glucose-regulated protein and immunoglobulin heavy chain binding protein. Cell 46: 291–300.

Nagai, K.; and Thogersen, H.C. 1984. Generation of b-globin by sequence-specific proteolysis of a hybrid protein produced in *Escherichia coli*. Nature 309: 810–812.

Negrouk, V.; Eisner, G.; Midha, S.; Lee, H.; Bascomb, N.; and Gleba, Y. 2004. Affinity purification of streptavidin using tobacco mosaic virus particles as purification tags. Anal. Biochem. 333: 230–235.

Nguyen, H.; Martinez, B.; Oganesyan, N.; and Kim R. 2004. An automated small-scale protein expression and purification screening provides beneficial information for protein production. J. Struct. Funct. Genomics 5: 23-27.

Niiranen, L.; Espelid, S.; Karlsen, C.R.; Mustonen, M.; Paulsen, S.M.; Heikinheimo, P., and Willassen, N.P. 2007. Comparative expression study to increase the solubility of cold adapted Vibrio proteins in *Escherichia coli*. Protein Expr. Purif. 52: 210–218.

Nilsson, J.; Stahl, S.; Lundeberg, J.; Uhlén, M.; and Nygren, P. 1997. Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. Protein Expr. Purif. 11: 1-16.

Oh, K.H.; Nam, S.H.; and Kim, H.S. 2002 Directed evolution of N-carbamyl-D-amino acid amidohydrolase for simultaneous improvement of oxidative and thermal stability. Biotechnol. Prog. 18: 413-417.

Ong, E.; Gilkes, N.R.; Warren, R.A.J.; Miller, R.C. Jr.; and Warren, R.A.J. 1989. Enzyme immobilization using the cellulose-binding domain of a *Cellulomonas fimi* exoglucanase. Biotechnol. 7: 604-607.

Ostermeier, M.; Shim, J.H.; and Benkovic, S.J. 1999. A combinatorial approach to hybrid enzymes independent of DNA homology. Nature Biotechnol. 17: 1205-1209.

Pedelacq, J.D.; Piltch, E.; Liong, E.C.; Berendzen, J.; Kim, C.Y.; Rho, B.S.; Park, M.S.; Terwilliger, T.C.; Waldo, G.S. 2002. Engineering soluble proteins for structural genomics. Nat. Biotechnol. 20: 927-932.

Peisley, A.A.; and Gooley, P. R. 2007. High-level expression of a soluble and functional fibronectin type II domain from MMP-2 in the *Escherichia coli* cytoplasm for solution NMR studies. Protein Expr. Purif. 53 : 124-131.

Pestovl, N.B.; and Rydström, J. 2007. Purification of recombinant membrane proteins tagged with calmodulin-binding domains by affinity chromatography on calmodulin-agarose: example of nicotinamide nucleotide transhydrogenase. Nature protocol 2: 198-202.

Porath, J.; Carlsson, J.; Olsson, I.; and Belfrage, G. 1975. Metal chelate affinity chromatography, a new approach to protein fractionation. Nature 258: 598-599.

Puig, O.; Caspary, F.; Rigaut, G.; Rutz, B.; Bouveret, E.; Bragado-Nilsson, E.; Wilm, M.; and Seraphin, B. 2001. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods 24: 218–229.

Qian, Z.; and Lutz, S. 2005. Improving the catalytic activity of *Candida antarctica* lipase B by circular permutation. J. Am. Chem. Soc. 127: 13466-13467.

Quiocho, F.A. 1986. Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. Annu. Rev. Biochem. 55: 287-315.

Remaud-Simeon, M.; Albenne, C.; Joucia, G.; Fabre, E.; Bozonnet, S.; Pizzut, S.; Escalier, P.; Potocki-Veronese, G.; and Monsan, P. 2003. Glucansucrases: structural basis, mechanistic aspects, and new perspectives for engineering. Oligosaccharides. Food Agric. 849: 90–103.

Remaud-Simeon, M.; Willemot, R.M.; Sarçabal, P.; Potocki de Montalk, G.; and Monsan P. 2000. Glucansucrases: molecular engineering and oligosaccharide synthesis. J. Mol. Catal.B. 10: 117-128.

Rigaut, G.; Shevchenko, A.; Rutz, B.; Wilm, M.; Mann, M.; and Seraphin, B. 1999. A generic protein purification method for protein complex characterization and proteome exploration. Nat. Biotechnol. 17: 1030–1032.

Robyt, J.F.; and Eklund, S.H. 1983. Relative, quantitative effects of acceptors in the reaction of *Leuconostoc mesenteroides* B-512F dextransucrase. Carbohydrate Res. 121: 279-286.

Robyt, J.F.; Kimble, B.K.; and Walseth, T.F. 1974. The mechanism of dextransucrase action. Direction of dextran biosynthesis. Arch. Biochem. Biophys. 165: 634-640.

Roodveldt, C.; Aharoni, A.; and Tawfik, D.S. 2005. Directed evolution of proteins for heterologous expression and stability. Current opinion in structural biology 15: 50-56.

Roodveldt, C.; and Tawfik, D.S. 2005. Directed evolution of phosphotriesterase from *Pseudomonas diminuta* for heterologous expression in *E. coli* results in stabilization of the metal-free state. Protein Eng. Des. Sel. 18: 51-58.

Routzahn, K. M.; and Waugh, D. S. 2002. Differential effects of supplementary affinity tags on the solubility of MBP fusion proteins. J. Struct. Funct. Genomics 2: 83-92.

Rudert, F.; Visser, E.; Gradl, G.; Grandison, P.; Shemshedini, L.; Wang, Y.; Grierson, A.; and Watson, J. 1996. pLEF, a novel vector for expression of glutathione S-transferase fusion proteins in mammalian cells. Gene 169: 281–282.

Sachdev, D.; and Chirgwin, J.M. 2000. Fusions to maltose-binding protein: control of folding and solubility in protein purification. Methods Enzymol. 326: 312–321.

Salazar, O.; and Sun, L. 2003. Evaluating a screen and analysis of mutant libraries. In directed enzyme evolution: screening and selection methods. Edited by Arnold, F.H.; and Georgiou, G. Vol 230. Totowa, NJ: Humana Press Inc..

Santarsiero, B. D.; Yegian, D. T.; Lee, C. C.; Spraggon, G.; Gu, J.; Scheibe, D.; Uber, D. C.; Cornell, E. W., Nordmeyer, R. A.; Kolbe, W. F.; Jin, J.; Jones, A. L.; Jaklevic, J. M.; Schultza, P. G.; and Stevensc, R. C. 2002. An approach to rapid protein crystallization using nanodroplets. J. Appl. Cryst. 35 : 278-281.

Sassenfeld, H.M.; and Brewer, S.J. 1984. A polypeptide fusion designed for purification of recombinant proteins. Bio/Technology 2: 76–81.

Scheich, C.; Sievert, V.; and Büssow, K. 2003. An automated method for high-throughput protein purification applied to a comparison of His-tag and GST-tag affinity chromatography. BMC Biotechnol. 3: 12.

Schein, C.H. 1989. Production of soluble recombinant proteins in bacteria. Bio/Technology 7: 1141-1149.

Schmidt, T.G.M.; and Skerra, A. 1993. The random peptide library assisted engineering of a C-terminal affinity peptide, useful for the detection and purification of a functional Ig Fv fragment. Protein Eng. 6: 109–122.

Schmidt, T.G.M.; Koepke, J.; Frank, R.; and Skerra, A. 1996. Molecular interactions between the *Strep*-tag affinity peptide and its cognate target, streptavidin. J. Mol. Biol. 255: 753–766.

Shah D.S.H.; and Russell R.R.B. 2002. Glucan binding domain of streptococcal glucosyltransferases. Biologia 11: 129-136.

Shah D.S.H.; Joucla G.; Remaud-Simeon M.; and Russell R.R.B. 2004. Conserved repeat motifs and glucan binding by glucansucrases of oral streptococci and *Leuconostoc mesenteroides*. J. Bacteriol. 186: 8301-8308.

Shao, Z.; Zhao, H.; Giver, L.; and Arnold, F.H. 1998. Random-priming in vitro recombination effective tool for directed evolution. Nucleic Acids Res. 26: 681-683.

Sharma, S.S.; Chong, S.; and Harcum, S.W. 2006. Intein-mediated protein purification of fusion proteins expressed under high-cell dencity conditions in *E. coli*. J. Biotechnol. 125: 48-56.

Sharrocks, A. 1994. A T7 expression vector for producing N- and C-terminal fusion proteins with glutathione S-transferase. Gene 138: 105-108.

Shirano, Y.; and Shibata, D. 1990. Low temperature cultivation of *Escherichia coli* carrying a rice lipoxygenase L-2 cDNA produces a soluble and active enzyme at a high level. FEBS Lett. 271: 128-130.

Sinnot, M.L. 1990. Catalytic mechanisms of enzymic glycosyl transfer. Chem. Rev. 90: 1171-1202.

Skov, L.K.; Mirza, O.; Sprogoe, D.; Dar, I.; Remaud-Simeon, M.; Albenne, C.; Monsan, P.; and Gajhede, M. 2002. Oligosaccharide and sucrose complexes of amylosucrase : Structural implications for polymerase activity. J. Biol. Chem. 277: 47741-47747.

Smith, D.B.; and Johnson, K.S. 1988. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione-S-transferase. Gene 67: 31–40.

Smith, D.J. 2003. Caries vaccines for the twenty-first century. J. Dent. Educ. 67: 1130–1139.

Smith, D.J.; Taubman, M.A.; Holmberg, C.F.; Eastcott, J.; King, W.J.; and Ali-Salaam, P. 1993. Antigenicity and immunogenicity of a synthetic peptide derived from a glucan binding domain of mutans streptococcal glucosyltransferase. Infect. Immun. 61: 2899-2905.

Smith, P.A.; Tripp, B.C.; DiBlasio-Smith, E.A.; Lu, Z.; La Vallie, E.R.; and McCoy, J.M. 1998. A plasmid expression system for quantitative in vivo biotinylation of thioredoxin fusion proteins in *Escherichia coli*. Nucleic Acids Res. 26: 1414–1420.

Smyth, D.R.; Mrozkiewicz, M.K.; McGrath, W.J.; Listwan, P.; Kobe, B. 2003. Crystal structures of fusion proteins with large-affnity tags. Protein Sci. 12: 1313–1322.

Stemmer, W.P.C. 1994. DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. Proc. Natl. Acad. Sci. USA 91: 10747-10751.

Stofko-Hahn, R.E.; Carr, D.W.; Scott, J.D. 1992. A single step purification for recombinant proteins. Characterization of a microtubule associated protein

(MAP 2) fragment which associates with the type II cAMP-dependent protein kinase. FEBS Lett. 302: 274–278.

Su, D.L.; and Robyt, J.F. 1994. Determination of the number of sucrose and acceptor binding sites for *Leuconostoc mesenteroides* B-512FM dextransucrase, and the conformation of the two site mechanism for dextran synthesis. Arch. Biochem. Biophys. 308: 471-476.

Su, X.A.; Prestwood. A.K.; and McGraw, R.A. 1992. Production of recombinant porcine tumor necrosis factor alpha in a novel *E. coli* expression system. Bio Techniques 13: 756–762.

Suen, W.C.; Zhang, N.Y.; Xiao, L.; Madison, V.; and Zaks, A. 2004. Improved activity and thermostability of *Candida antarctica* lipase B by DNA family shuffling. Protein Eng. Des. Sel. 17: 133-140.

Sun, J.W.; Wanda, S.Y.; Camilli, A.; and Curtiss III, R. 1994. Cloning and DNA sequencing of the dextranse inhibitor gene (dei) from *Streptococcus sobrinus*. J. Bacteriol. 176: 7213-7222.

Terpe, K. 2003. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl. Microbiol. Biotechnol. 60: 523-533.

Tomme, P.; Warren, A. J.; Miller, R. C. J.; Kilburn, D. G.; and Gilkes, N. R. 1995. Enzymatic Degradation of Insoluble Carbohydrates. Saddler, J. N. and Penner, M. H. (eds.), Vol. 618, pp. 145–163. San Diego: ACS.

Tsuchiya, H.M.; Hellman, N.N.; and Koepsell, H.J. 1953. Factors affecting molecular weight of enzymatically synthesized dextran. J. Am. Chem. Soc. 75: 757-758.

Tsunoda, Y.; Sakai, N.; Kikuchi, K.; Katoh, S.; Akagi, K.; Miura-Ohnuma, J., Tashiro, Y., Murata, K., Shibuya, N.; and Katoh, E. 2005. Improving expression and solubility of rice proteins produced as fusion proteins in *Escherichia coli*. Protien Expr. Purif. 42: 268-277.

van den Berg, S.; Löfdahl, P.A.; Härd, T.; and Berlund, H. 2006. Improved solubility of TEV protease by directed evolution. J. Biotechnol. 121: 291-298.

van Hijum, S. A. F. T.; Kralj, S.; Ozimek, L.K.; Dijkhuizen, L.; and van Geel-Schutten, I.G.H. 2006. Structure-function relationships of glucansucrase and

fructansucrase enzymes from lactic acid bacteria. Microbiol. Mol. Biol. Rev. 70: 157-176.

Vaughan, T.J.; Williams, A.J.; Pritchard, K.; Osbourn, J.K.; Pope, A.R.; Earnshaw, J.C.; McCafferty, J.; Hodits, R.A.; Wilton, J.; and Johnson, K.S. 1996. Human antibodies with sub-nanomolar affinities isolated from large non-immunized phage display library. Nat. Biotechnol. 14: 309–314.

Vickerman M.M.; Sulavik M.C.; Minick P.E.; and Clewell D.B. 1996. Changes in the carboxy-terminal repeat region affect extracellular activity and glucan products of *Streptococcus gordonii* glucosyltransferase. Infect. Immun. 64: 5117-5128.

Vincentelli, R.; Canaan, S.; Offant, J.; Cambillau, C.; and Bignon, C. 2005. Automated expression and solubility screening of His-tagged proteins in 96-well format. Analytical Biochem. 346: 77-84.

von Eichel-Striber C.; Saueborn M.; and Kuramitsu H.K. 1992. Evidence for a modular structure of the homologous repetitive C-terminal carbohydrate-binding sites of *Clostridium difficile* toxins and *Streptococcus mutans* glucosyltransferases. J. Bacteriol. 174: 6707-6710.

Voss, S.; and Skerra, A. 1997. Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. Protein Eng. 10: 975–982.

Waldo, G.S.; Standish, B.M.; Berendzen, J.; and Terwilliger, T.C. 1999. Rapid protein-folding assay using green fluorescent protein. Nat. Biotechnol. 17: 691-695.

Wang, J.Y.; and Chao, Y. P. 2006. Immobilization of cells with surface-displayed chitin-binding domain. Appl Environ Microbiol. 72: 927-931.

Watanabe, T.; Ito, Y.; Yamada, T.; Hashimoto, M.; Sekine, S.; and Tanaka, H. 1994. The role of the C-terminal domain and type III domains of chitinase A1 from *Bacillus circulans* WL-12 in chitin degradation. J. Bacteriol. 176: 4465–4472.

Waugh, D.S. 2005. Making the most of affinity tags. Trends Biotechnol. 23: 316-320.

Wen, C.M.; Tseng, C.S.; Cheng, C.Y.; and Li, Y.K. 2002. Purification, characterization and cloning of a chitinase from *Bacillus* sp. NCTU2. Biotechnol. Appl. Biochem. 35: 213–219.

Wilke-Douglas, M.; Perchorowicz, J.T.; Houck, C.M.; and Thomas, B.R. 1989. Methods and compositions for altering physical characteristics of fruit and fruit products. PCT patent WO 89/12386.

Williams, D.P.; Regier, D.; Akiyoshi, D.; Genbauffe, F.; and Murphy, J.R. 1988. Design, synthesis and expression of a human interleukin-2 gene incorporating the codon usage bias found in highly expressed *Escherlichia coli* genes. Nucl. Acids Res. 16: 10453-10467.

Wilson, D.S.; Keefe, A.D.; and Szostak, J.W. 2001. The use of mRNA display to select high-affinity protein-binding peptides. PNAS 98: 3750–3755.

Wong C.; Hefta S.A.; Paxton R.J.; Shively J.E.; and Mooser G. 1990. Size and subdomain architecture of the glucan-binding domain of sucrose:3-alpha-D-glucosyltransferase from *Streptococcus sobrinus*. Infect. Immun. 58: 2165-2170.

Wong, D.W.; Batt, S.B.; Lee, C.C.; and Robertson, G.H. 2004. High-activity barley α-amylase by directed evolution. Protein J. 23: 453-460.

Woodyer, R.; Chen, W.; and Zhao, H. 2004. Outrunning nature: directed evolution of superior biocalysts. J. Chem. Education 81: 126-133.

Wren, B.W.; Russell, R.R.B.; and Tabaqchali, S. 1991. Antigenic cross-reactivity and functional inhibition by antibodies to *Clostridium difficile* toxin A, *Streptococcus mutans* glucan binding protein and a synthetic peptide. Infect. Immun. 59: 3151-3155.

Wu, X.; Jörnvall, H.; Berndt, K.D.; and Oppermann, U. 2004. Codon optimization reveals critical factors for high level expression of two rare codon genes in Escherichia coli: RNA stability and secondary structure but not tRNA abundance. Biochem Biophys Res Commun. 313: 89-96.

Xu, M.Q.; Paulus, H.; and Chong, S. 2000. Fusions to self-splicing inteins for protein purification. Methods Enzymol. 236: 376–418.

Yun, C.; Matsuda, H.; and Kawamura, M. 2006. Directed evolution to enhance secretion efficiency and thermostability of chitosanase from *Mitsuaria chitosanitabida* 3001. Biosci. Biotechnol. Biochem. 70: 559-563.

Zhang, X.K.; Wills, K.N.; Hunsmann, M.; Hermann, T.; and Pfahl, M. 1991. Novel pathway for thyroid hormone receptor action through interaction with jun and fos oncogene activities. Mol. Biol. 11: 6016–6025.

Zhao, H. 2004. Staggered extension process in vitro DNA recombination. Methods Enzymol. 388: 42-49.

Zhao, H.; and Arnold, F.H. 1997. Optimization of DNA shuffling for high fidelity recombination. Nucl. Acids Res. 25: 1307-1308.

Zhao, H.; Giver, L.; Shao, Z.; Affholter, J.A.; and Arnold, F.H. 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. Nature Biotechnol. 16: 258-261.

Zheng, C.F.; Simcox, T.; Xu, L.; and Vaillancourt, P. 1997. A new expression vector for high level protein production, one step purification and direct isotopic labeling of calmodulin-binding peptide fusion proteins. Gene 186: 55–60.

APPENDICES

APPENDIX A: dextransucrase-S sequence (for numbering).

```
    attgtgtattgaaatttttactgtttttacaaataaattaagtta
    taatttttcaatgtaatagaagagaatattataaggagaaaattt
  1 atgccatttacagaaaaagtaatgcggaaaaagctttataaagtt
    M  P  F  T  E  K  V  M  R  K  K  L  Y  K  V
 46 gggaaaagttgggtagttggtggggtttgtgcttttgcattaacc
    G  K  S  W  V  V  G  G  V  C  A  F  A  L  T
 91 gcctcatttgctttagcaacaccaagtgttttaggagacagtagt
    A  S  F  A  L  A  T  P  S  V  L  G  D  S  S
136 gtacctgatgtgagtgcgaataacgttcaatctgcttcagataat
    V  P  D  V  S  A  N  N  V  Q  S  A  S  D  N
181 acaacggatacgcagcagaacactacggttaccgaagaaaatgat
    T  T  D  T  Q  Q  N  T  T  V  T  E  E  N  D
226 aaagtacagtctgcagctactaatgacaatgtaacaacagctgca
    K  V  Q  S  A  A  T  N  D  N  V  T  T  A  A
271 agcgacacaacacaatctgctgataataatgtgacagaaaaacag
    S  D  T  T  Q  S  A  D  N  N  V  T  E  K  Q
316 tcagatgatcatgcacttgataatgaaaaagtcgataacaaacaa
    S  D  D  H  A  L  D  N  E  K  V  D  N  K  Q
361 gatgaagtcgctcaaaccaatgttactagcaaaaatgaggaatca
    D  E  V  A  Q  T  N  V  T  S  K  N  E  E  S
406 gcagttgcttcaactgacactgatcctgctgaaacgacaactgac
    A  V  A  S  T  D  T  D  P  A  E  T  T  T  D
451 gaaacacaacaagttagcggcaagtacgttgaaaaagacggtagt
    E  T  Q  Q  V  S  G  K  Y  V  E  K  D  G  S
496 tggtattattattttgatgatggcaaaaatgctaaaggtttatca
    W  Y  Y  Y  F  D  D  G  K  N  A  K  G  L  S
541 acgatagacaacaatattcaatatttttacgagagtggtaaacaa
    T  I  D  N  N  I  Q  Y  F  Y  E  S  G  K  Q
586 gccaaaggacagtatgtcacaattgataatcaaacatattatttt
    A  K  G  Q  Y  V  T  I  D  N  Q  T  Y  Y  F
631 gataagggctcaggtgatgagttaactggtctgcaaagcattgat
    D  K  G  S  G  D  E  L  T  G  L  Q  S  I  D
676 gggaacatagttgcttttaacgatgaagggcaacaaatttttaat
    G  N  I  V  A  F  N  D  E  G  Q  Q  I  F  N
721 caatattaccaatctgaaaatggtacaacatactactttgatgat
    Q  Y  Y  Q  S  E  N  G  T  T  Y  Y  F  D  D
766 aaaggacacgctgctaccggtattaagaatatcgagggcaaaaat
    K  G  H  A  A  T  G  I  K  N  I  E  G  K  N
811 tattattttgataatcttgggcaactaaaaaaaggcttctctggt
    Y  Y  F  D  N  L  G  Q  L  K  K  G  F  S  G
856 gtgattgatggtcaaataatgacatttgatcaggaaacagggcaa
    V  I  D  G  Q  I  M  T  F  D  Q  E  T  G  Q
901 gaagtttctaacacaacttctgaaataaaagaaggtttgacgact
    E  V  S  N  T  T  S  E  I  K  E  G  L  T  T
946 caaaacacggattatagcgaacataatgcagcccacggtacggat
    Q  N  T  D  Y  S  E  H  N  A  A  H  G  T  D
991 gctgaggactttgaaaatattgacggctatttaacagctagttca
    A  E  D  F  E  N  I  D  G  Y  L  T  A  S  S
1036 tggtatcgtccaacaggtattttacgtaacggaacagactgggaa
    W  Y  R  P  T  G  I  L  R  N  G  T  D  W  E
1081 ccttctacagatacagatttcagaccaatattgtcagtgtggtgg
    P  S  T  D  T  D  F  R  P  I  L  S  V  W  W
1126 ccagataagaacacccaggtcaattatttaaattacatggctgat
```

```
     P  D  K  N  T  Q  V  N  Y  L  N  Y  M  A  D
1171 ttagggtttatcagtaatgcggacagttttgaaactggggatagc
     L  G  F  I  S  N  A  D  S  F  E  T  G  D  S
1216 caaagcttattaaatgaagcaagtaactatgttcaaaaatcaatt
     Q  S  L  L  N  E  A  S  N  Y  V  Q  K  S  I
1261 gaaatgaaaattagtgcgcaacaaagtacagagtggttaaaggat
     E  M  K  I  S  A  Q  Q  S  T  E  W  L  K  D
1306 gcaatggcggccttcattgtcgcgcaaccacagtggaatgaaact
     A  M  A  A  F  I  V  A  Q  P  Q  W  N  E  T
1351 agtgaagatatgagcaatgaccatttacaaaatggcgcattaact
     S  E  D  M  S  N  D  H  L  Q  N  G  A  L  T
1396 tatgtcaacagtccactgacacctgacgctaattcaaactttaga
     Y  V  N  S  P  L  T  P  D  A  N  S  N  F  R
1441 ctacttaatcggacaccaacaaaccagactggtgaacaagcgtat
     L  L  N  R  T  P  T  N  Q  T  G  E  Q  A  Y
1486 aatttagataattcaaaaggtggttttgaattgttgttagccaat
     N  L  D  N  S  K  G  G  F  E  L  L  L  A  N
1531 gacgttgataattcaaaccctgtagtacaagcagaacaattgaat
     D  V  D  N  S  N  P  V  V  Q  A  E  Q  L  N
1576 tggttatattatttaatgaattttggtacgattacggccaacgac
     W  L  Y  Y  L  M  N  F  G  T  I  T  A  N  D
1621 gcggatgctaattttgatggtattcgtgtagatgcagtcgacaat
     A  D  A  N  F  D  G  I  R  V  D  A  V  D  N
1666 gtggatgctgatttgttacaaattgctgccgattatttcaaacta
     V  D  A  D  L  L  Q  I  A  A  D  Y  F  K  L
1711 gcttacggtgttgatcaaaatgatgctactgctaatcagcatctt
     A  Y  G  V  D  Q  N  D  A  T  A  N  Q  H  L
1756 tcaattttggaagattggagtcacaatgatcctttgtatgtaaca
     S  I  L  E  D  W  S  H  N  D  P  L  Y  V  T
1801 gatcaaggaagcaatcaattaaccatggatgattatgtgcacaca
     D  Q  G  S  N  Q  L  T  M  D  D  Y  V  H  T
1846 caattaatctggtctctaacaaaatcatctgacatacgaggtaca
     Q  L  I  W  S  L  T  K  S  S  D  I  R  G  T
1891 atgcagcgcttcgtggattattatatggtggatcgatctaatgat
     M  Q  R  F  V  D  Y  Y  M  V  D  R  S  N  D
1936 agtacagaaaacgaagccattcctaattacagctttgtacgtgca
     S  T  E  N  E  A  I  P  N  Y  S  F  V  R  A
1981 cacgacagcgaagtgcaaacggttattgcccaaattgtttccgat
     H  D  S  E  V  Q  T  V  I  A  Q  I  V  S  D
2026 ttgtatcctgatgttgaaaatagtttagcaccaacaacagaacaa
     L  Y  P  D  V  E  N  S  L  A  P  T  T  E  Q
2071 ttggcagctgctttcaaagtatacaatgaagatgaaaaattagca
     L  A  A  A  F  K  V  Y  N  E  D  E  K  L  A
2116 gacaaaaagtacacacaatataatatggctagtgcttatgcgatg
     D  K  K  Y  T  Q  Y  N  M  A  S  A  Y  A  M
2161 ttgctaaccaataaggatactgttcctcgtgtctattatggcgat
     L  L  T  N  K  D  T  V  P  R  V  Y  Y  G  D
2206 ttatatacagatgatggtcaatatatggcaacaaagtcaccatac
     L  Y  T  D  D  G  Q  Y  M  A  T  K  S  P  Y
2251 tatgatgcgattaacactttgctaaaggctagagttcagtatgtt
     Y  D  A  I  N  T  L  L  K  A  R  V  Q  Y  V
2296 gctggtggccaatcgatgtccgttgatagtaatgacgtgttaaca
     A  G  G  Q  S  M  S  V  D  S  N  D  V  L  T
2341 agtgttcgctatggtaaagatgccatgacagcttctgacactgga
     S  V  R  Y  G  K  D  A  M  T  A  S  D  T  G
2386 acatctgagacgcgtactgaaggtattggagtcatcgtcagcaat
```

```
     T  S  E  T  R  T  E  G  I  G  V  I  V  S  N
2431 aacgcggagctacaattagaggatgggcatactgtcacattgcat
     N  A  E  L  Q  L  E  D  G  H  T  V  T  L  H
2476 atgggggcagctcataagaaccaagcttatcgtgctttgttatca
     M  G  A  A  H  K  N  Q  A  Y  R  A  L  L  S
2521 acaactgcagatggattagcttattatgatactgatgaaaatgca
     T  T  A  D  G  L  A  Y  Y  D  T  D  E  N  A
2566 cctgtggcgtacacagatgctaacggcgatttgatttttacgaat
     P  V  A  Y  T  D  A  N  G  D  L  I  F  T  N
2611 gaatcaatttatggtgtacaaaatccacaagtttctggttacttg
     E  S  I  Y  G  V  Q  N  P  Q  V  S  G  Y  L
2656 gcagtttgggttccggtaggtgcgcaacaagatcaagatgcacga
     A  V  W  V  P  V  G  A  Q  Q  D  Q  D  A  R
2701 acggcctctgatacaacaacaaacacgagtgataaagtgttccat
     T  A  S  D  T  T  T  N  T  S  D  K  V  F  H
2746 tcaaacgctgctcttgattctcaagtcatctacgaaggtttctca
     S  N  A  A  L  D  S  Q  V  I  Y  E  G  F  S
2791 aacttccaagcatttgctacagacagcagtgaatatacaaacgta
     N  F  Q  A  F  A  T  D  S  S  E  Y  T  N  V
2836 gtcatcgctcagaatgcggaccaatttaagcaatggggtgtgaca
     V  I  A  Q  N  A  D  Q  F  K  Q  W  G  V  T
2881 agcttccaattggcaccacaatatcgttcaagtacagatacaagt
     S  F  Q  L  A  P  Q  Y  R  S  S  T  D  T  S
2926 ttcttggattcaattattcaaaacgggtatgcattcacggatcgt
     F  L  D  S  I  I  Q  N  G  Y  A  F  T  D  R
2971 tatgacttaggttatggcacaccgacaaaatatggaactgctgat
     Y  D  L  G  Y  G  T  P  T  K  Y  G  T  A  D
3016 cagttgcgcgatgctattaaagccttacatgctagcggtattcaa
     Q  L  R  D  A  I  K  A  L  H  A  S  G  I  Q
3061 gccattgccgattgggtgccggaccaaatttataatttgccagag
     A  I  A  D  W  V  P  D  Q  I  Y  N  L  P  E
3106 caagaattagctactgtcacaagaacaaattcatttggagatgac
     Q  E  L  A  T  V  T  R  T  N  S  F  G  D  D
3151 gatacagattctgatattgacaatgccttatatgttgtacaaagt
     D  T  D  S  D  I  D  N  A  L  Y  V  V  Q  S
3196 cgtgggggtggtcaatatcaagagatgtatggtggtgccttctta
     R  G  G  G  Q  Y  Q  E  M  Y  G  G  A  F  L
3241 gaagagttacaggcactctatccatccctatttaaagtgaatcaa
     E  E  L  Q  A  L  Y  P  S  L  F  K  V  N  Q
3286 atctcaactggcgttccaattgatggcagtgtaaagattactgag
     I  S  T  G  V  P  I  D  G  S  V  K  I  T  E
3331 tgggcggctaagtacttcaatggctctaacatccaaggtaaaggt
     W  A  A  K  Y  F  N  G  S  N  I  Q  G  K  G
3376 gctggatacgtattgaaagatatgggttctaataagtactttaag
     A  G  Y  V  L  K  D  M  G  S  N  K  Y  F  K
3421 gtcgtttcgaacactgaggatggtgactacttaccaaaacagtta
     V  V  S  N  T  E  D  G  D  Y  L  P  K  Q  L
3466 actaatgatctgtcagaaactggctttacacacgatgataaagga
     T  N  D  L  S  E  T  G  F  T  H  D  D  K  G
3511 atcatctattatacattaagtggttatcgtgcccaaaatgcattt
     I  I  Y  Y  T  L  S  G  Y  R  A  Q  N  A  F
3556 attcaagatgatgataataactattactattttgataaaacaggt
     I  Q  D  D  D  N  N  Y  Y  Y  F  D  K  T  G
3601 catttagtaacaggtttgcaaaagattaataaccatacctacttc
     H  L  V  T  G  L  Q  K  I  N  N  H  T  Y  F
3646 ttcttacctaatggtatcgaactggtcaagagcttcttacaaaac
```

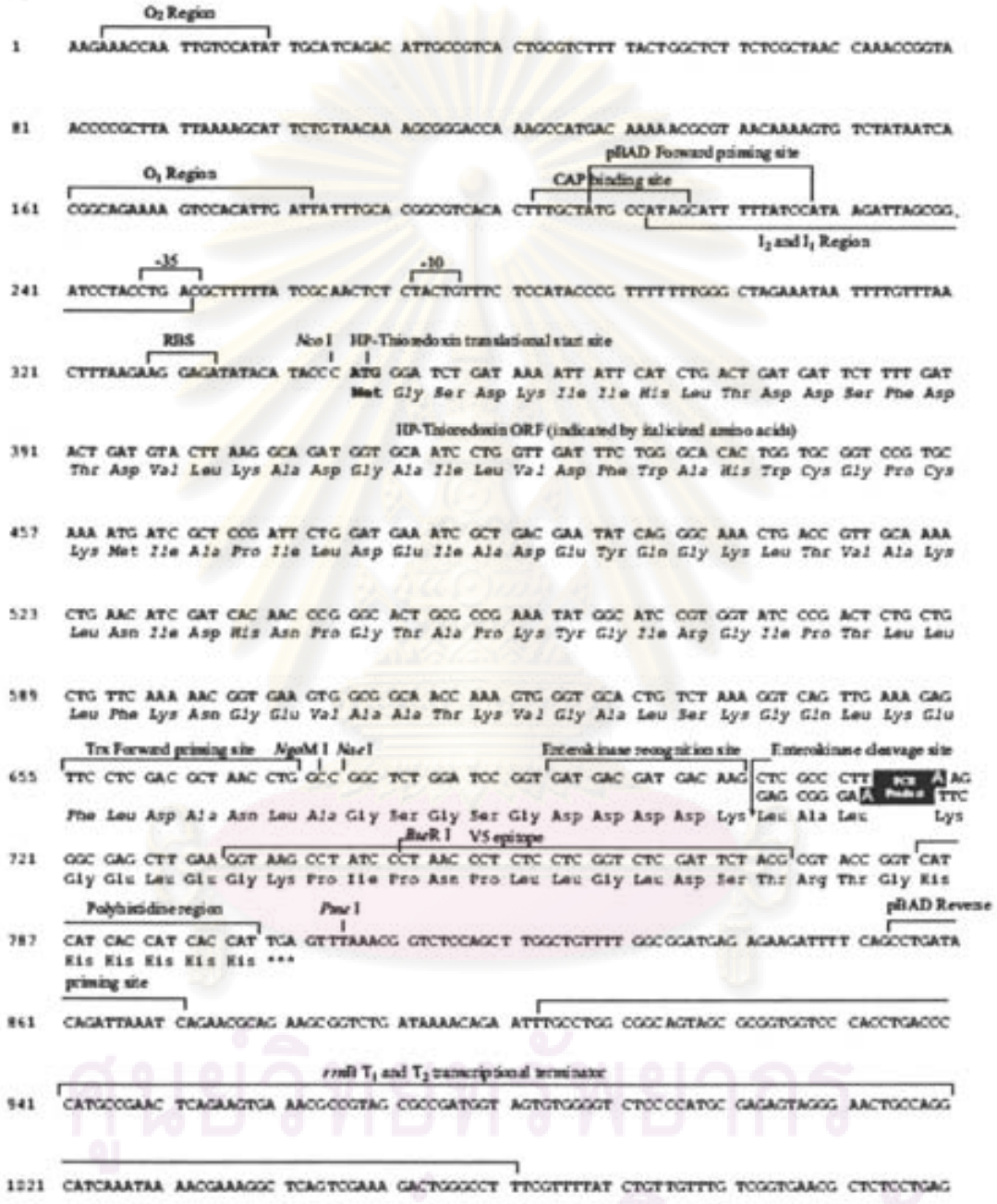
```
     F  L  P  N  G  I  E  L  V  K  S  F  L  Q  N
3691 gaagatggtacaattgtttatttcgataagaaaggtcatcaagtt
     E  D  G  T  I  V  Y  F  D  K  K  G  H  Q  V
3736 tttgatcaatatataactgatcaaaatggaaatgcgtattacttt
     F  D  Q  Y  I  T  D  Q  N  G  N  A  Y  Y  F
3781 gatgatgctggtgtaatgcttaaatcagggcttgcaacgattgat
     D  D  A  G  V  M  L  K  S  G  L  A  T  I  D
3826 ggacatcaacagtattttgatcaaaatggtgtgcaggttaaggat
     G  H  Q  Q  Y  F  D  Q  N  G  V  Q  V  K  D
3871 aagtttgtgattggcactgatggttataagtattactttgaacca
     K  F  V  I  G  T  D  G  Y  K  Y  Y  F  E  P
3916 ggtagtggtaacttagctatcctacgttatgtgcaaaatagtaag
     G  S  G  N  L  A  I  L  R  Y  V  Q  N  S  K
3961 aatcaatggttctattttgatggtaatggccatgctgtcactggt
     N  Q  W  F  Y  F  D  G  N  G  H  A  V  T  G
4006 ttccaaacaattaatggtaaaaaacaatatttctataatgatggt
     F  Q  T  I  N  G  K  K  Q  Y  F  Y  N  D  G
4051 catcaaagtaaaggtgaattcattgatgcagacggggatactttc
     H  Q  S  K  G  E  F  I  D  A  D  G  D  T  F
4096 tatacgagtgccactgatggtcgcctagtaactggtgttcagaag
     Y  T  S  A  T  D  G  R  L  V  T  G  V  Q  K
4141 attaatggtattacctatgcttttgataacacaggaaatttgatc
     I  N  G  I  T  Y  A  F  D  N  T  G  N  L  I
4186 acaaatcagtattatcaattagcagatggtaaatatatgttgtta
     T  N  Q  Y  Y  Q  L  A  D  G  K  Y  M  L  L
4231 gatgatagtggtcgtgcgaaaacagggtttgtattgcaagatggt
     D  D  S  G  R  A  K  T  G  F  V  L  Q  D  G
4276 gtactaagatacttcgatcaaaacggtgagcaagtgaaagatgct
     V  L  R  Y  F  D  Q  N  G  E  Q  V  K  D  A
4321 atcattgtggatccagatactaacttgagttattatttcaatgca
     I  I  V  D  P  D  T  N  L  S  Y  Y  F  N  A
4366 acacaaggtgtcgctgtaaaaaatgattatttcgagtatcaaggt
     T  Q  G  V  A  V  K  N  D  Y  F  E  Y  Q  G
4411 aattggtatttaacagatgctaattatcaacttatcaaaggtttt
     N  W  Y  L  T  D  A  N  Y  Q  L  I  K  G  F
4456 aaagcagttgacgacagcttacaacattttgatgaagtcactggt
     K  A  V  D  D  S  L  Q  H  F  D  E  V  T  G
4501 gtacaaacaaaagatagtgctttaataagtgctcagggtaaggtt
     V  Q  T  K  D  S  A  L  I  S  A  Q  G  K  V
4546 taccaatttgataatattggaaatgctgtgtcagcataa 4584
     Y  Q  F  D  N  I  G  N  A  V  S  A  *
```

APPENDIX B: amino acid sequence of the C-terminal domain of DSR-S (GBD-0).

NNYYYFDKTGHLVTGLQKINNHTYFFLPNGIELVKSFL
QNEDGTIVYFDKKGHQVFDQYITDQNGNAYYFDDAGV
MLKSGLATDGHQQYFDQNGVQVKDKFVIGTDGYKYY
FEPGSGNLAILRYVQNSKNQWFYFDGNGHAVTGFQTIN
GKKQYFYNDGHQSKGEFIDAQGDTFYISATDGRLVTG
YQKINGITYAFDNTGNLITNQYYQLADGKYMLLDDSGR
AKTGFVLQDGVLRYFDQNGEQVKDAIIVDPDTNLSYYF
NATQGVATKNDYFEYQGNWYLTDANYQLIKGFKAVD
DSLQHFDEVTGVQTKDSALISAQGKVYQFDNNGNAVS
A

Hightlight indicated YG repeat motif.

APPENDIX C: Cloning site of pBAD TOPO Thiofusion vector (Invitrogen).



APPENDIX D: Cloning site of pBAD TOPO TA vector (Invitrogen).

APPENDIX E: Standard amino acid abbreviations (Voet and Voet, 2004).

Amino acid	1-letter	3-letter
Alanine	A	Ala
Arginine	R	Arg
Asparagine	N	Asn
Aspartic acid	D	Asp
Cysteine	C	Cys
Glutamic acid	E	Glu
Glutamine	Q	Gln
Glycine	G	Gly
Histidine	H	His
Isoleucine	I	Ile
Leucine	L	Leu
Lysine	K	Lys
Methionine	M	Met
Phenylalanine	F	Phe
Proline	P	Pro
Serine	S	Ser
Threonine	T	Thr
Tryptophan	W	Trp
Tyrosine	Y	Tyr
Valine	V	Val

BIOGRAPHY

Miss Surisa Suwannarangsee was born on the January 13th, 1980 in Prachuap Khiri Khan, Thailand. In 2002, she graduated with the Bachelor of Science in Genetics from faculty of Science, Chulalongkorn University, Thailand. After that, she continued to study in Ph.D. program at Biological Science Program, Faculty of Science, in the same university.

Publication:

Suwannarangsee, S.; Moulis, C.; Potocki-Veronese, G.; Monsan, P.; Remaud-Simeon, M.; and Chulalaksananukul, W. 2007. Search for a dextransucrase minimal motif involved in dextran binding. FEBS Lett. 581: 4675-4680.

Award and scholarships:

1998-2001	Development and Promotion of Science and Technology Talents Project
2001	Awarded Excellent Student from Faculty of Science (Professor Tab's Gold Medal), Chulalongkorn University
2002-2008	Royal Golden Jubilee Ph.D. Scholarship
2005	First prize in oral presentation entitled "Identification of peptides specific for dextran binding and their use as affinity tag for purification onto Sephadex® gel" in 10th Biological Science Graduate Congress, NUS, Singapore