In silico and *in vitro* Characterization of Ten Putative RsbR-like proteins in *Saprospira grandis*

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

MICROBIOLOGY

MAY 2012

By

Sun Ae Kim

Thesis Committee:
Maqsudul Alam, Chairperson
Paul Q. Patek
Hongwei Li

We certify that we have read this thesis and that, in our opinion, it is satisfactory in scope and quality as a thesis for the degree of Master of Science in Microbiology.

THESIS COMMITTEE

_____
Chairperson

_____

_____

ACKNOWLEDGEMENTS

I would like to thank Jesus Christ, my Lord and savior who always guides me with grateful love. I thank to my mom, dad and little brother for their unconditional love and support. I would like to thank Dr. Maqsudul Alam for giving me this opportunity and always believing me. He has shown me as a great role model to be the scientist that I really want to be. I thank to Dr. Paul Q. Patek and Dr. Hongwei Li for serving as members of my graduate committee. I am so thankful for their guidance and support. Many thanks to lab members Duckhyung Lee, Eric Brucks, Jeff Lyford, Jennifer Moss and Nolwenn Phan. I also thank to previous lab members, Dr. Shaobin Hou, Dr. Jennifer Saito, Dr. Xuehua (Ivy) Wan, Aaron Young and Jimmy Saw. It was my blessing to have them as my colleagues and my family members. I would like to thank Dr. Farrukh Jamil for his collaboration and help with crystallization of RsbR1 for this project. I thank to Dr. Yong Soo Kim for his generous advice. I thank to Ms. Ok Sun Kim, Dr. Hye Jin Kim and Dr. Eun Jung Moon for great support and prayer. I thank to the Department of Microbiology including Debbie Morito, Lei Takahashi, Kerwin Lum and all the other people from Microbiology department.

TABLE OF CONTENTS

CHAPTER 3- RESULTS

LIST OF TABLES

page

LIST OF FIGURES

page

ABSTRACT

Globin-coupled sensors (GCSs) are heme-binding proteins that consist of an N-terminal sensor globin domain and a C-terminal transducer domain. Their functions are varied depending on the C-terminal domain and can be classified into two groups such as aerotactic and gene regulation groups. The gene regulating group is further divided into DNA-binding, 2nd messenger and protein-protein interaction group. So far, aerotactic transducers: HemAT-*Hs* from *Halobacterium salinarum* and HemAT-*Bs* from *Bacillus subtilis* are well characterized in a molecular level and a cellular level. In addition, GCSs from the 2nd messenger group within the gene regulating group such as *Bpe*Greg (*Bordetella pertussis*), *Ec*Greg (*E. coli*), *Cv*Greg (*Chromobacterium violaceum*) and *Av*Greg (*Azotobacter vinelandii*) are recently characterized as well.

Recently, genome sequencing of *Saporospira grandis* has been completed and revealed in-depth genomic information that related to interesting traits of this organism such as gliding motility, production of rhapidosome and ixotrophic nutrient obtaining mechanisms. Moreover, the initial machine annotation showed that there are ten putative GCSs which contain C-terminal STAS domains. Based on the machine annotation results, we have ten genes that were predicted to be GCSs and have a C-terminal STAS (Sulfur transporter and *a*nti-sigma antagonist) domain (Table 1.1). The putative function of these proteins is an anti-sigma factor antagonist like SpoIIAA, RsbR and RsbS in *B. subtilis* which involves in nutritional, physical and environmental stress responses. The protein named RsbR1 (SGRA_3210) within *S. grandis* was first discovered and its neighbor genes *rsbS, -T, -V, -X* and *rsbU* shared similarities with *rsbR* and its adjacent genes (*rsbS, -T, -U, -V, -W, sigB, rsbX*) in *B. subtilis*. Also, the structure of RsbR consists of the N-terminal non-heme globin domain and the C-terminal STAS domain. All these genes and their encoded proteins are related to

activate more than 150 stress response genes upon physical stresses through releasing a sigma factor B. Based on these similarities, we postulate RsbR1 and nine other proteins are globin-coupled sensors that are involved in the stress response.

In this study, *in silico* and *in vitro* characterization of ten GCSs were carried out on both the N-terminal globin and the C-terminal STAS domains. Multiple alignments of the globin domains showed that only SGRA_0571 (RsbR2), SGRA_3210 (RsbR1) and SGRA_3852 (RsbR3) aligned their histidine residue with the proximal histidine residue in known globin domains. Two phosphorylation sites from RsbR and its paralogs in *B. subtilis* were aligned with nine GCSs except SGRA_3852 and YtvA which is known not to have any phosphorylation site, instead it has a GTP binding motif. Two amino acid residues (D and G) from the GTP binding motif (DXXG) are all aligned in all GCSs. The 3D structure alignments of the globin domain of *Geobactor sulfurreducens* and RsbR1, RsbR2 and RsbR3 showed around 23% identity but the proximal histidine, where it held the heme, were all aligned with the histidine residue of three GCSs. For *in vitro* characterization, all ten genes were cloned, expressed and purified. The purified proteins were dialyzed and carried UV spectroscopic analysis. Only three GCSs (RsbR1, RsbR2 and RsbR3) exhibited oxygenated myoglobin like peaks. This shows that these proteins are able to bind oxygen. RsbR1 was undergone for crystallization. Initial crystallization screening showed some red crystals.

Initially we found ten GCS genes from the annotation of the genome. *In silico* results predicted them all as GCSs. Multiple alignments of the globin domains and STAS domains of ten GCSs showed only RsbR1, RsbR2 and RsbR3 had the conserved proximal histidine residues which were vital for heme-binding. The STAS domains of ten GCSs were shared high homology with the STAS domains of RsbR and its paralogs from *B. subtilis* including two conserved phosphorylation sites and possible GTP-binding motifs. However, RsbR3 did

not show the conserved phophrylation sites. All ten GCS genes were cloned, expressed and purified. RsbR1, RsbR2 and RsbR3 displayed oxygenated myoglobin-like absorption spectra and they were the only proteins that were validated as GCSs.

CHAPTER 1

INTRODUCTION

1.1. Globin superfamily and globin fold

Globins are porphyrin-containing proteins (Burmester *et al*., 2000) and are found in all three kingdoms of living organisms (Bashford *et al*., 1987; Hardison, 1998; Frey and Kallio, 2003). Possessing globins in unicellular organisms suggest that the presence of globin encoding gene is very ancient and its functions are varied other than transporting and storing oxygen (Lu *et al*., 2008). For many years, globins were limited to myoglobin (Mb), the alpha and beta subunits of hemoglobin (Hb) found in vertebrate and the leghemoglobin discovered in root nodules of legume plants (Vinogradov *et al*., 2005; Appleby, 1984). Following the development of DNA sequencing and bioinformatics tools, globins have been found in all the kingdoms of life forms and have rapidly grown to a large globin superfamily (Frey and Kallio, 2003; Kosmachevskaya and Topunov, 2008; Wajcman *et al*., 2009). The globin superfamily is composed of three lineages. The first lineage includes flavohemoglobins, chimeric proteins (~400 amino acid residues) and related single-domain hemoglobins, which display a canonical 3/3 myoglobin-like α-helical folding with the heme as a prosthetic group (Kosmachevskaya and Topunov, 2008; Vinogradov *et al*., 2007; Wajcman *et al*., 2009). The second lineage consists Globin-coupled sensors (300-700 amino acid residues), chimeric heme proteins composing a N-terminal globin-like heme binding domain and a C-terminal transducer domain and related single-domain protoglobins (Kosmachevskaya and Topunov, 2008; Saito *et al*., 2008). The third lineage includes newly discovered single domain globins called truncated globins whose amino acid sequences are 20 to 40 residues shorter than known full length globins (Lecomte *et al*., 2005; Milani *et al*., 2004).

The typical structure of globins is characterized by a 3 over 3 alpha helical sandwich

fold that consists of eight helices designated A through H (Figure 1.1A) (Wajcman *et al*., 2009). A heme molecule (protoporphyrin IX) is embedded in these helices (A, B, C, and E helices from a distal side of the heme and F, G, and H from a proximal side of the heme) (Kosmachevskay and Topunov, 2009; Wajcman *et al*., 2009). The heme molecule is composed of a porphyrin which contains a nitrogen atom that places toward the center of the ring and an iron atom that is located in the center (Figure 1.1B) (Hardison, 1999). The iron in the heme is held by the proximal histidine residue in the F helix (F8) which is known to the conserved amino acid in all globins and bound to gaseous ligands on the opposite with coordination of other amino acid residues on the distal side of helices (Lecomte *et al*., 2005; Peterson *et al*., 1997).

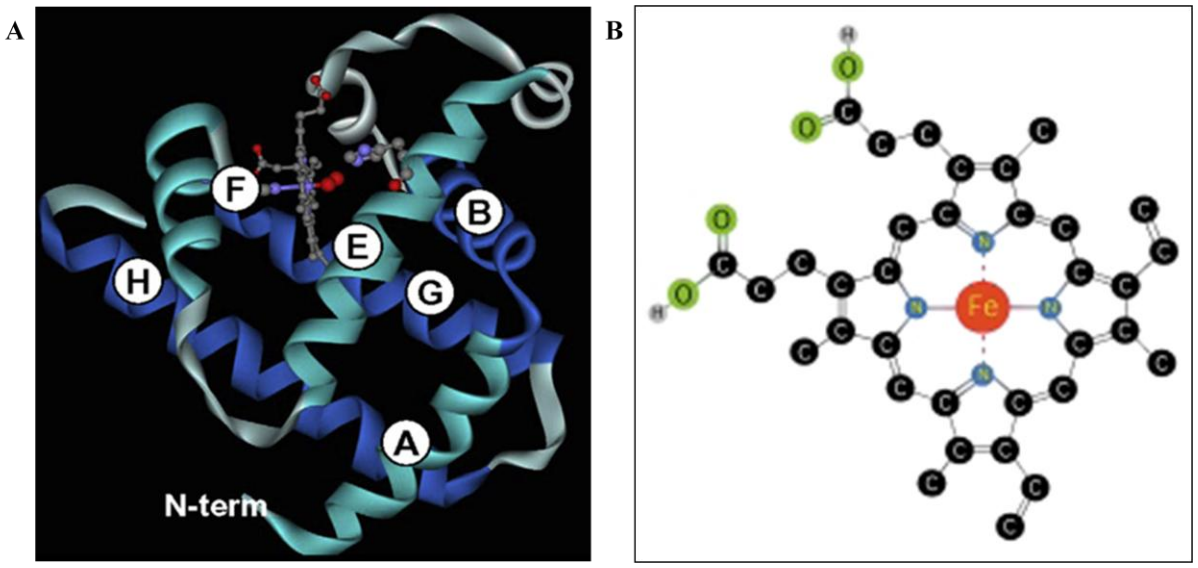


Figure 1.1. Structures of a 3/3 globin and heme. (A) Sperm whale myoglobin is the typical example of a 3/3 globin, the heme group is surrounded by 3 helices on its proximal site (F, G, and H) and 3 helices on its distal site (A, B and E). Figure from Wajaman *et al*., 2009. Structure and function evolution in the superfamily of globins. *C R Biol*, 332**,** 273-82, Figure 1, Page 275. (B) The structure of heme, the ligand-iron porphyrin complex in hemoglobin. The hydrogen atoms saturating the carbon bonds are not shown. Figure from Goldoni, 2001-2002. Porphyrins: Fascinating Molecules with Biological Significance, Elettra highlights, Trieste. 64-66, Figure 2, Page 64.

1.2. Heme-based sensors

Various organisms in all kingdoms of life use heme-based sensor proteins as the key regulators (Gilles-Gonzalez and Gonzalez, 2005) to response changes of gaseous ligands-oxygen, carbon monoxide and nitric oxide (Chan, 2001). These proteins exhibit a heme-binding sensor domain that passes the input signal to a neighboring transmitter domain where it initiates DNA binding or other enzymatic activities (Chan, 2000; Rodger, 1999). Heme-based sensors are divided into six families based on what types of heme-binding domains that they have (Figure 1.2) (Gilles-Gonzalez and Gonzalez, 2007). The adjacent transmitter domains consist of histidine protein kinase, cyclin-dinucleotide phosphodiesterase, nucleotide cyclase, chemotaxis receptor and DNA-binding transcription-factor (Gilles-Gonzalez and Gonzalez, 2005).

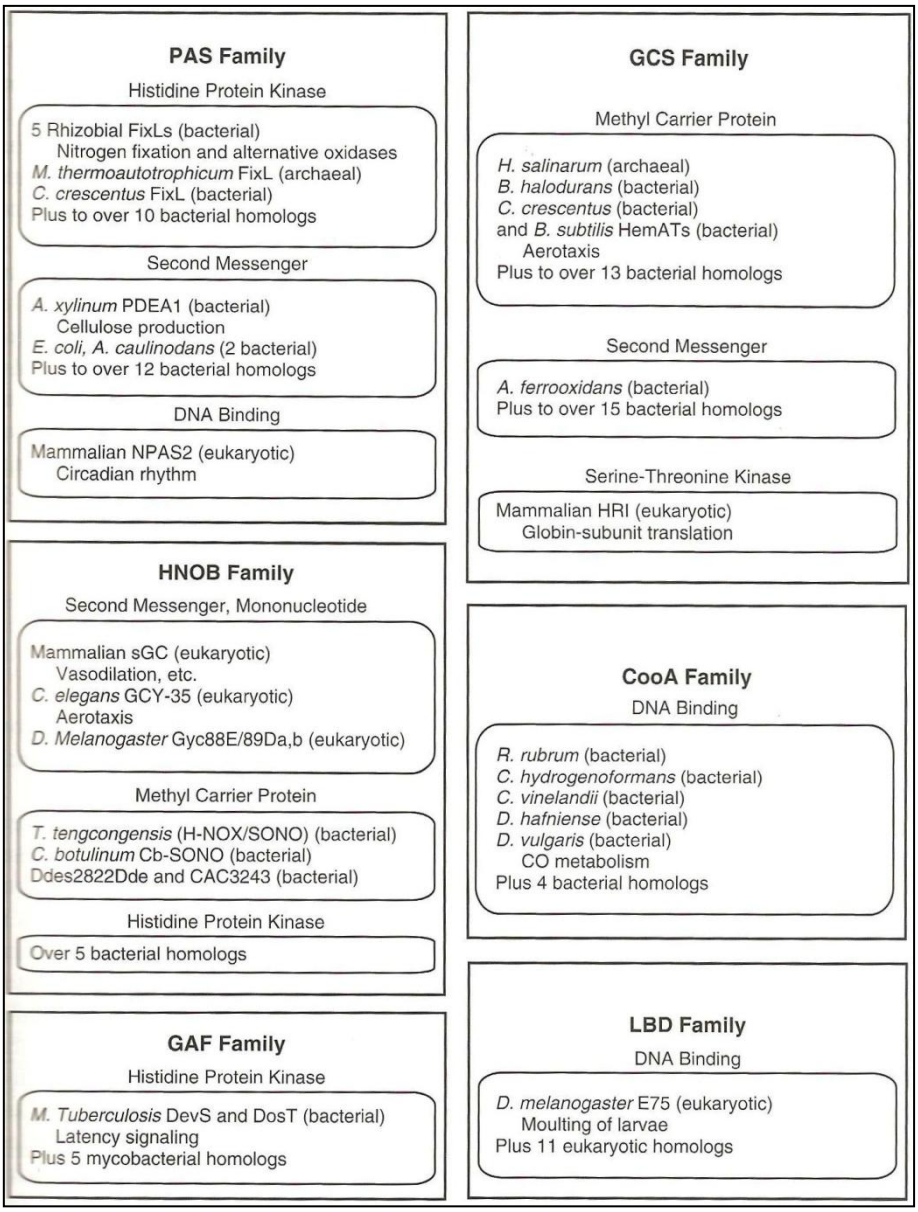


Figure 1.2. Families of heme-based sensors. A distinctive heme-binding domain defines each family of sensors. Subgroups within the families couple their heme-binding domain to different transmitters for signal transduction. Those proteins specifically named are ones that have been purified and established as heme proteins. The physiological functions, if known, are highlighted in green. The last line in each category notes the kingdom membership and approximate numbers of additional members expected from sequence homology. Figure from Gilles-Gonzalez and Gonzalez, 2007. A Surfeit of Biological Heme-based Sensors. *In:* Ghosh, A. (ed.) *The Smallest Biomolecules Diatomics and their Interactions with Heme Proteins.* First Edition ed.Oxford: Elsevier, Figure 1, Page 19.

1.3. Globin-coupled sensors

Globin-coupled sensors (GCSs) are multi-domain sensory proteins (Saito *et al*., 2008), consisting of a globin domain and a signal transduction domain (Desmet *et al*., 2010) (Figure 1.3). The first GCSs, HemAT-*Hs* and HemAT-*Bs* were discovered in archaeon *Halobacterium salinarum* (HemAT-*Hs*) and Gram-positive bacterium *Bacillus subtilis* (HemAT-*Bs*) (Hou *et al*., 2000; Saito *et al*., 2008). These two proteins are aerotactic transducers that contain an N-terminal myoglobin-like heme binding domain and a C-terminal signaling domain which is homologous to the bacterial cytoplasmic signaling domain of methyl-accepting chemotaxis proteins (MCPs) (Hou *et al*., 2001; Saito *et al*., 2008). So far, more than 80 GCSs have been found and there will be more GCSs added as releasing newly sequenced genomic data (Freitas *et al*., 2008).

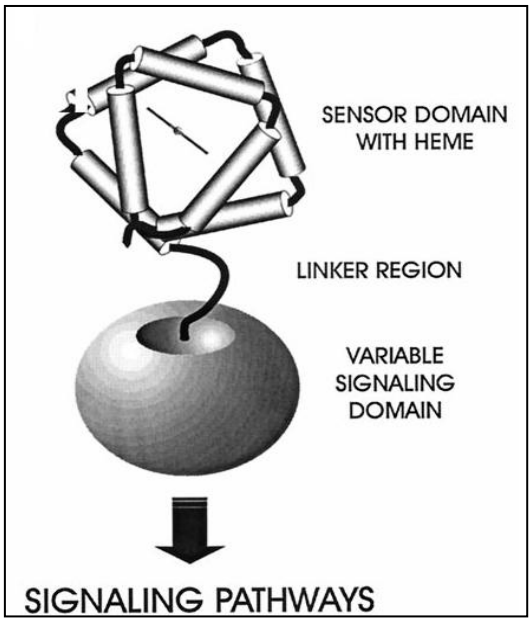


Figure 1.3. Domain structure of a Globin-coupled sensor.

1.4. Functional classification of GCSs

Globin-coupled sensors were categorized into two groups: aerotactic and gene regulation (Freitas *et al.*, 2003). The gene regulation group was further divided into three subgroups: protein-DNA (Gekakis *et al.*, 1998; Hogenesch *et al.*, 1998; Lanzilotta *et al.*, 2000; Reick *et al.*, 2001; Dioum *et al.*, 2002), protein-protein (Gilles-Gonzalez and Gonzalez, 1993; David *et al.*, 1988) and 2nd messenger pathways (Zhao *et al.*, 1999; Delgado-Nixon *et al.*, 2000; Sasakura *et al.*, 2002; Chang *et al.*, 2001). This classification is based on classification of Heme-based sensors (Figure 1.4) (Freitas *et al.*, 2008).

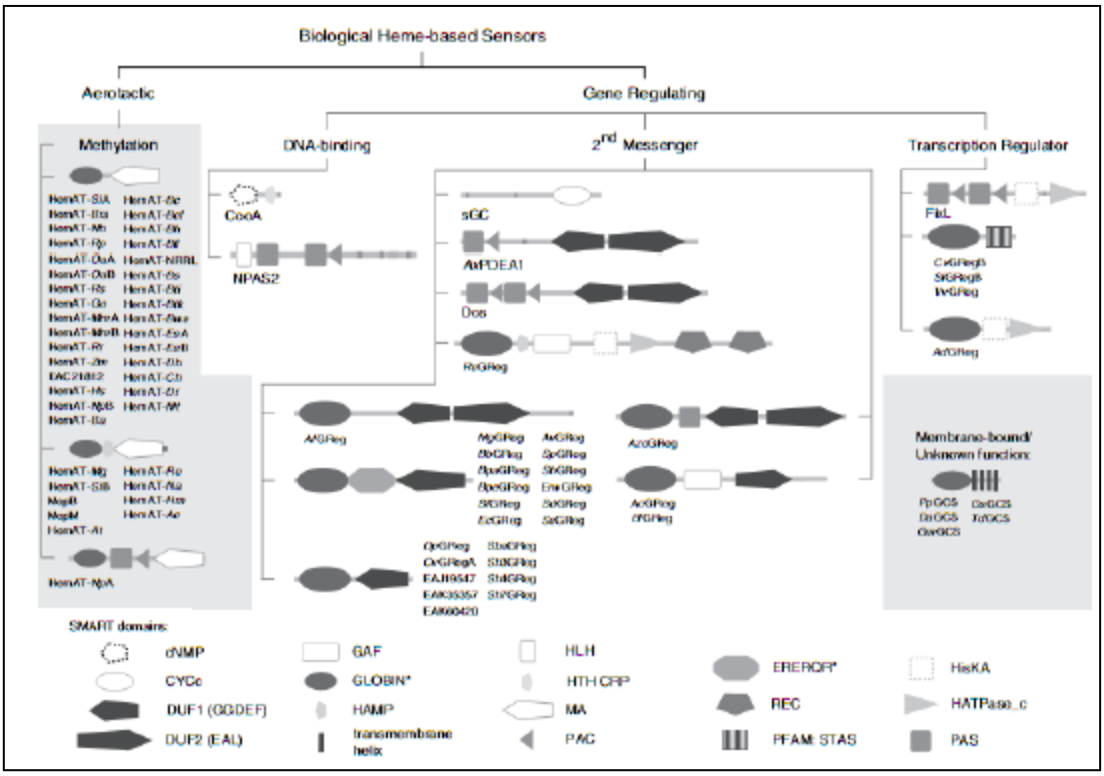


Figure 1.4. Classification schema of biological heme-based sensors. Figure from Freitas *et al.*, 2008. Protoglobin and Globin-coupled Sensors. *In:* Ghosh, A. (ed.) *The Smallest Biomolecules Diatomics and their Interactions with Heme Proteins.* First Edition ed. Oxford: Elsevier, Figure 2, page 186.

1.4.1. Aerotactic

The first discovered GCSs belonging to this group are HemATs from *Bacillus subtilis* and *Helicobactor salinarum* (Hou *et al*., 2000). These two proteins are the only GCSs that have been studied experimentally for their physiological effects (Freitas *et al*., 2008). The HemATs from each organism can mediate a positive response (*B. subtilis*) and a negative response (*H. salinarum*) based on an oxygen level (Gilles-Gonzales and Gonzalez, 2005; Vinogradov *et al*., 2007). These responses were characterized by time-lapsed capillary assays (Hou *et al*., 2000). The HemATs are composed of two domains, an N-terminal globin domain with or without a HAMP (Histidine kinases, Adenylylcyclases, Methyl-accepting chemotaxis Proteins) domain (Alexandre and Zhulin, 2001) and a C-terminal MCP-like domain. They are purified as soluble proteins (Freitas *et al*., 2003). Most of HemATs are found in Gram-negative α-Proteobacteria, Firmicutes and Archaea (Freitas *et al*., 2003; Freitas *et al*., 2008).

1.4.2. Second messenger

Thirty GCSs were categorized in this group based on the functions of their identified domains (Freitas *et al*., 2008). Each GCS includes either the GGDEF (domain containing Gly-Gly-Asp-Glu-Phe motif) domain alone or the combination of the EAL (a conserved Glu-Ala-Leu motif in the domain) domain (Freitas *et al*., 2003; Thijs *et al*., 2007) as a transmitter module (Wan *et al*., 2009). The GGDEF domain is homologous to the adenylyl cyclase catalytic domain and exhibits digualylate cyclase (DGC) activity which synthesizes bis-(3′–5′)-cyclic diguanosine monophosphate (c-di-GMP) (Wan *et al*., 2009). On the contrary, the EAL domain posses phosphodiesterase (PDE) activity and degrades c-di-GMP (Thijs *et al*., 2007). The c-di-GMP is a second messenger in bacteria and plays important roles such as cellulose synthesis (Kitanishi *et al*., 2010), regulation of bacteria motility, and biofilm formation (Wan *et al*., 2009). The GCS from *Bordetella pertussis* which causes whooping

cough exhibits a globin sensing domain with a GGDEF domain. Experimentally, this protein, *Bpe*Greg has been shown that sensing oxygen actually promotes biofilm formation (Wan *et al*., 2009). The heme binding protein, *Ax*PDEA1 from *Gluconacetobacter xylinus* which incorporates with GGDEF and EAL domains displays cellulose production depending on cellular oxygen levels and production of c-di GMP (Freitas *et al*., 2008). In *Escherichia coli*, *Ec*Dos (PDE activity-hydrolysis of c-di GMP) (Sasakura *et al*., 2006) and *Ec*Greg (or YddV) (DGG activity-synthesis of c-di GMP) along with their oxygen sensing domains regulate various physiological functions of bacteria (Kitanishi *et al*., 2010).

1.4.3. Protein-protein interaction

So far, four GCSs have been placed in this subdivision. *Ad*Greg is a GCS which consists of a catalytic HATPase_c domain and a HisKA histidine kinase domain in *Anaeromyxobacter dehalogenans* (Freitas *et al*., 2008). It belongs to the δ-Proteobacterium and is a facultative anaerobe and is able to utilize various metals such as oxidized Fe (III), U (VI) and halogenated compounds for terminal electron acceptors (Cole *et al.*, 1994; Loffler *et al.*, 1999; Sanford *et al.*, 2002; He and Sanford, 2002; He and Sanford, 2003). The function of *Ad*Greg is assumed as a sensory histidine kinase like FixL (Freitas *et al*., 2008). Other three GCSs have been found in *Vibrio vulnificus*, *Chromobacterium violaceum*, and *Silicobacter* sp. TM1040. Their domain organization is composed of an N-terminal globin domain and a C-terminal STAS (Sulfate transporter and antisigma-factor antagonist). It has not been experimentally done to determine their function but it may serve as a regulator of gene expression like an antisigma-factor antagonist which is similar to SpoIIAA in *B. subtilis* spore formation (Aravind and Koonin, 2000; Freitas *et al*., 2008).

1.4.4. Unclassified GCS

Currently, GCSs found in five bacteria are categorized as the unclassified GCS due to

lack of information and partial C-terminal signaling domain or unrecognizable domain with a long sequence. These bacteria are the anaerobic δ-Proteobacteria *Geobacter sulfurreducens*, *Geobacter metallireducens*, *Pelobacter propionicus* DSM 2179 and the ε-Proteobacterium *Thiomicrospira denitrificans*. All GCSs from this group are soluble except for the small portion of memebrane-spanning region that is composed of ~four membrane helices (Freitas *et al*., 2008).

1.5. *Saprospira grandis*

Saprospira grandis is a gram-negative marine bacterium and an obligate aerobe belonging to the phylum of Bacteroides-Flavobacterium-Cytophaga (Dietrich and Biggins, 1971; Mincer *et al*., 2004). It is commonly found on sand of coastlines and other aquatic environments (Lewin, 1997). *S. grandis* is motile but does not swim. Instead it glides across a surface with a mysterious mechanism that is not well understood (Aizawa, 2005). The average movement of this bacterium is approximately two to four micrometers per second and can glide up to 10 micrometer per second on glass surfaces by using its body to form a long helical filament (Aizawa, 2005; McBride, 2001). The interesting traits of *S. grandis* include not only its gliding motility but also its predatory behavior on other marine bacteria (Sangkhobol and Skerman, 1981). It actually captures prey on its sticky surface and gathers them at one place for digestion (Aizawa, 2005). Lewin (1997) defines this preying behavior as “Isotrophy”, which means catching prey on a sticky substance like bird-lime or flypaper (Figure 1.5). Like the R-type pyosin of *Pseudomonas aeruginosa*, *S. grandis* has similar looking proteins called rhapidosomes. The rhapidosomes are rod-shaped particles that are released by lysis of cells from *S. grandis* (Figure 1.6). Morphologically, these particles are similar to particles of tobacco-mosaic virus (TMV) (Correll and Lewin, 1963). However, the particles do not contain any nucleic acids and solely proteins are composed of (Delk and

Dekker, 1972). Thus, it is still not clear what the actual role of rhapidosomes is in *S. grandis*.

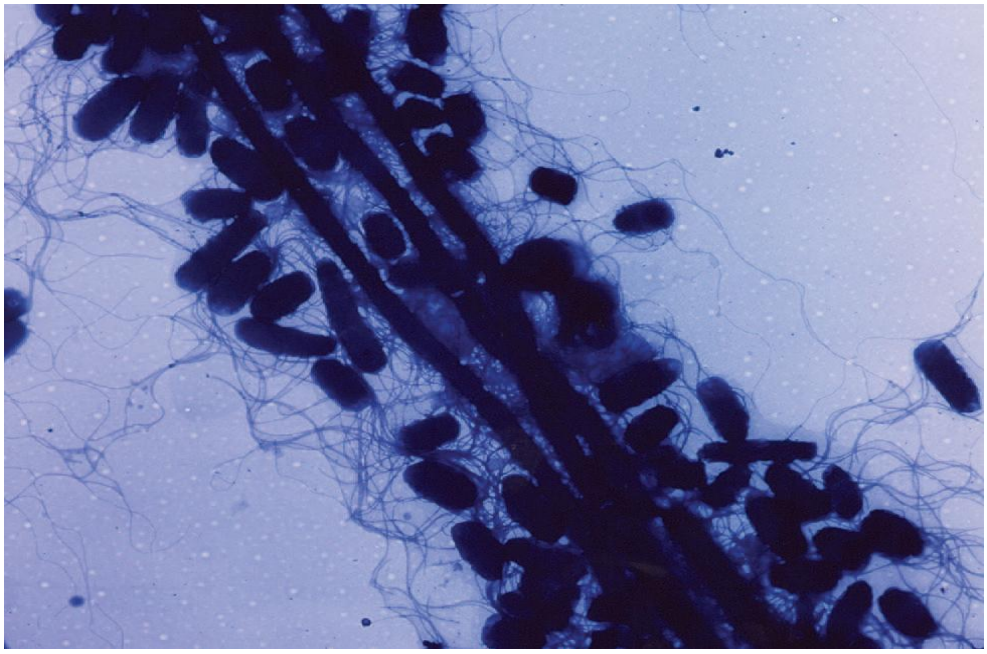


Figure 1.5. Wild-type *Salmonella typhimurium SJW1103* cells were trapped along the filamentous cell surface of *S. grandis*. Figure from Aizawa, S. I. 2005. Bacterial Gliding Motility: Visualizing Invisible Machinery. *ASM News*, 71**,** 71-75, Figure 2.D, Page74.

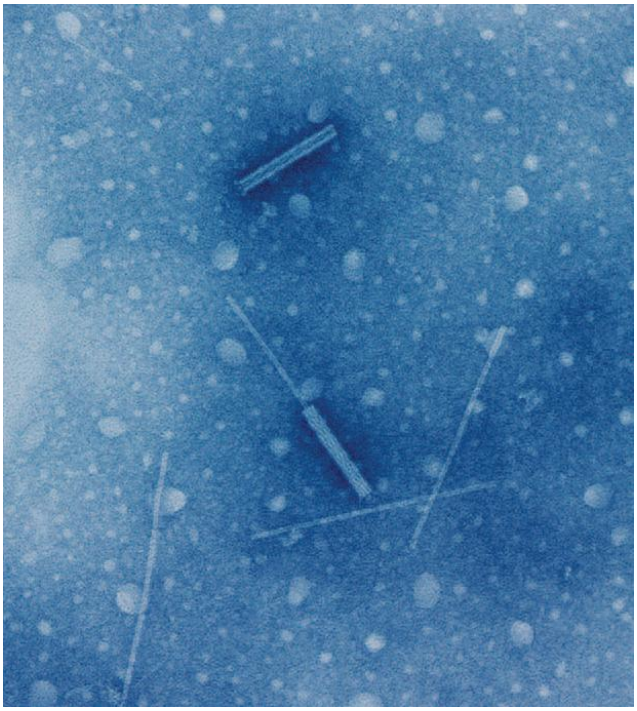


Figure 1.6. Electron micrographs of rhapidosome, a bacteriocin of *S. grandis*. Figure from Aizawa, S. I. 2005. Bacterial Gliding Motility: Visualizing Invisible Machinery. *ASM News*, 71**,** 71-75, Figure 3A, Page75.

1.6. Bacterial stress response

Drastic changes in environmental, physical and nutritional conditions in bacteria switch their usual mode of living into protection mode to prepare future stress (Kim *et al*, 2004). More than 150 general stress response genes are transcribed by the alternative sigma factor called sigma B when induced by general stress in *Bacillus subtilis* (Martinez *et al*., 2010). The sigma B pathway is a signaling transduction cascade initiated via phosphorylation of proteins and protein-protein interactions (Marles-Wright *et al*., 2008). Many proteins are involved in this cascade, especially RsbR, RsbS and RsbT, which form a large complex called a stressosome (Figure 1.7). These are the first proteins to respond to environmental stresses (Pane-Farre *et al*., 2005). They consist as a cluster in a genome and its name of operon is a regulator of sigma B (Rsb) (Marles-Wright *et al*., 2008). RsbR, RsbS and RsbT form the stressosome in the absence of stress. In contrast, RsbR and RsbS are phosphorylated by RsbT which is a serine/threonine kinase. Two phosphorylation sites in RsbR are known as threonine at position 171 and 205 and one phosphorylation site in RsbS (Akbar *et al*., 2001). The phosphorylation events release interaction of RsbR, RsbS and RsbT from the complex. RsbT itself binds to RsbU, and ultimately sigma B is released (Figure 1.8) (Delumeau *et al*., 2006). Based on structural and mutational studies of RsbR, this protein is believed to be a sensor in this whole signal integration (Akbar *et al*., 1997; Delumeau *et al*., 2006). RsbR consists of an N-terminal globin domain and a C-terminal STAS domain. The N-terminal domain of RsbR forms a globin fold and may detect gaseous ligands (Delumeau *et al*., 2006). However, RsbR lacks a heme pocket, and the proximal histidine is replaced by alanine. It has not been clearly described which types of ligand may bind or what signal it senses (Murray *et al*., 2005). In *B. subtilis*, there are six paralog proteins that interact with RsbR and RsbS *in vitro*. Five of these paralogs contain a threonine residue that can be phosphorylated. Of those

five, three contain an additional threonine residue phosphorylation site. One other paralog, YtvA has a light sensing N-terminal LOV (ligh-oxygen-voltage) domain that is a member of PAS (Per-ARNT-Sim) superfamily and a C-terminal STAS domain (Akbar *et al*., 2001). The LOV domain is a photosensor, and is able to sense a blue light. In addition, the C-terminal STAS domain may possess NTP binding ability, and its possible binding site could be located in a NTP binding motif (DXXG) found in the C-terminal STAS domain (Buttani *et al*., 2006).

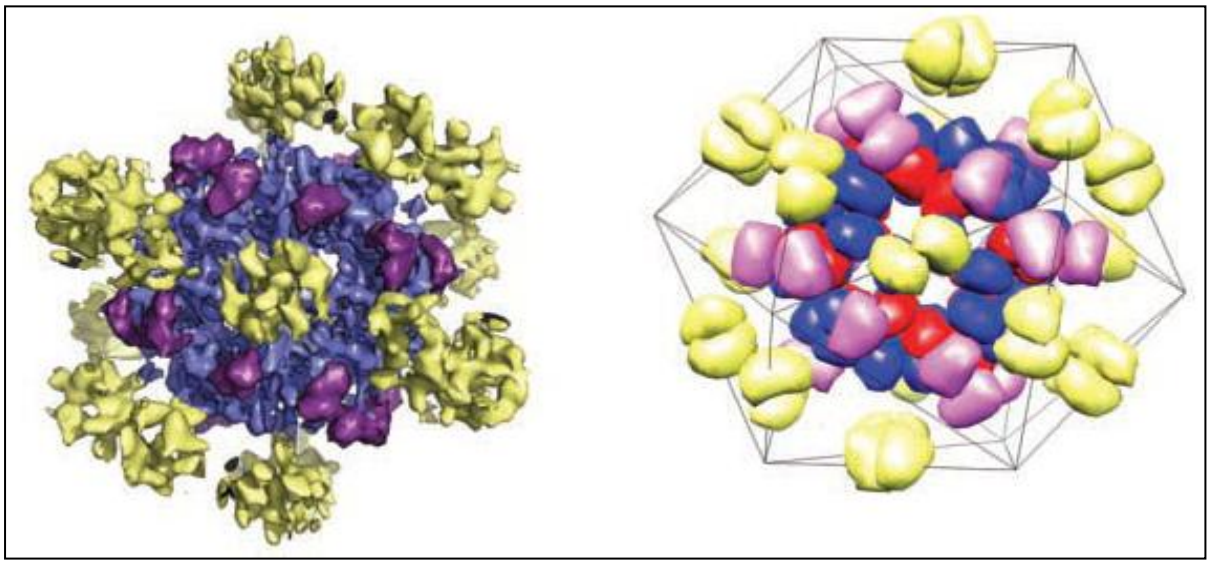


Figure 1.7. Stressosome structure. The N-terminal domain of RsbR is colored in yellow and the C-terminal of RsbR is indicated in blue color. RsbS is colored in red and RsbT is colored in purple. Figure from Marles-Wright *et al*., 2008. Molecular Architecture of the "Stressosome," a Signal Integration and Transduction Hub. *Science*, *322*(5898), 92 -96. Figure 1C, page 93.

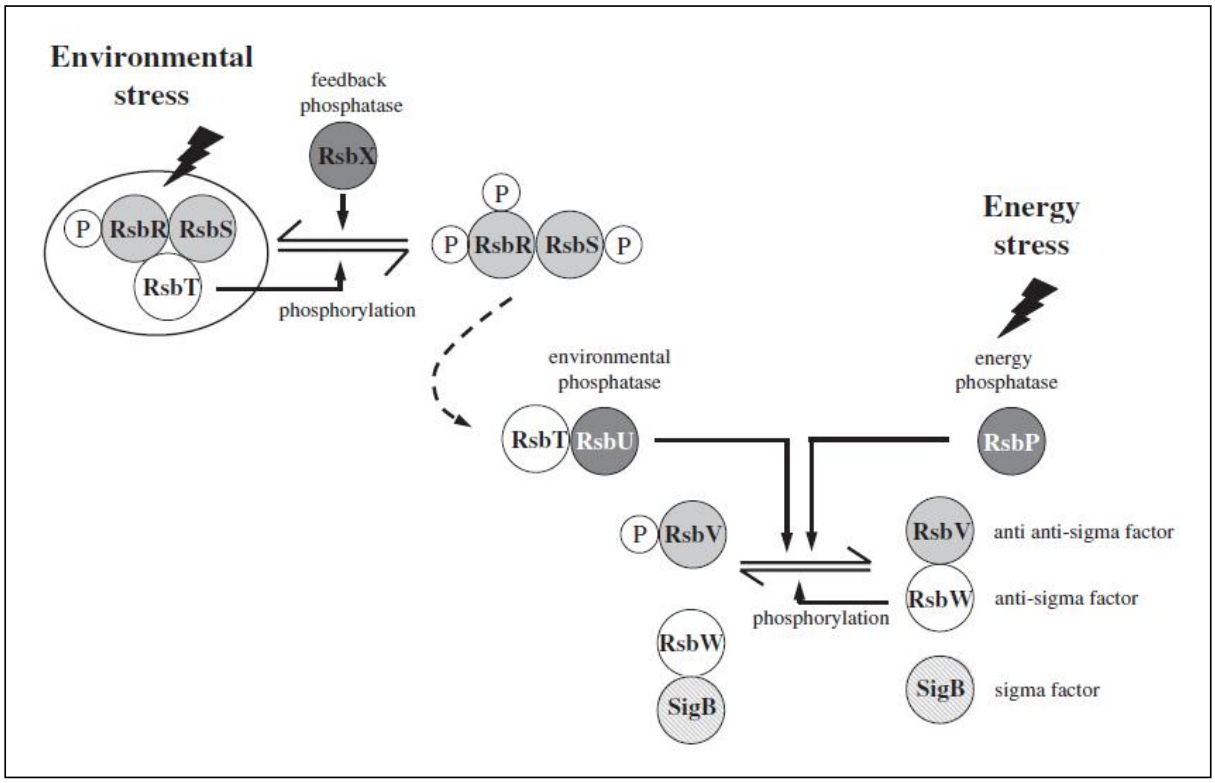


Figure 1.8. The signaling pathways of controlling the activity of sigma B activity in *B. subtilis*. Figure from Pané-Farré *et al*., 2005. The RsbRST Stress Module in Bacteria: A Signaling System That May Interact with Different Output Modules. *J Mol Microbiol Biotechnol*. 9:65-76. Figure 1, page 66.

1.7. GCS in *S. grandis*

Recently a genome project of the marine microbe, *Saprospira grandis* was completed. High throughput bioinformatics analysis revealed many genes that related to its metabolic pathways and adaptations. Based on the marine annotation results, we have found 10 genes that were predicted to be GCSs and have a C-terminal STAS (Sulfur transporter and *a*nti-sigma antagonist) domain (Table 1.1). These findings will put these putative GCSs in fourth member of the GCS with the STAS domain in transcriptional regulator sub-group (Figure 1.4). The STAS domain shows similarities with the SpoIIAA (anti-anti-sigma factor protein) and the C-terminal STAS domain of RsbR protein in *Bacillus subtilis*. RsbR is a general

stress response protein which has two domains: an N-terminal globin fold but without heme and a C-terminal STAS domain. We postulate that putative GCSs in *S. grandis* may be involved in stress responses.

Table 1.1. List of Ten putative GCS genes

Locus Tag	Size (bp)	Annotation	Domain Structure
SGRA_0571	864	anti-anti-sigma regulatory factor	sensor_globin, STAS
SGRA_1293	873	anti-anti-sigma regulatory factor	sensor_globin, STAS
SGRA_2160	870	anti-anti-sigma regulatory factor	sensor_globin, STAS
SGRA_2161	855	anti-anti-sigma regulatory factor	sensor_globin, STAS
SGRA_2162	891	anti-anti-sigma regulatory factor	sensor_globin, STAS
SGRA_2167	861	anti-anti-sigma regulatory factor	sensor_globin, STAS
SGRA_2168	873	anti-anti-sigma regulatory factor	sensor_globin, STAS
SGRA_2169	852	anti-anti-sigma regulatory factor	sensor_globin, STAS
SGRA_3210	876	anti-anti-sigma regulatory factor	sensor_globin, STAS
SGRA_3852	867	anti-anti-sigma regulatory factor	sensor_globin, STAS

1.8. Objective of study

The goal of this research is designed to validate the putative function by using bioinformatics tools and spectroscopic characterization of recombinant GCS from *S. grandis*. The objectives for this research are as follows:

1. *In silico* characterization of ten putative GCSs within *S. grandis* genome
2. Cloning, expression, and purification of putative GCS genes in *S. grandis*
3. Spectroscopic characterization of purified recombinant GCS proteins
4. Probing protein crystallization of RsbR1 (SGRA_3210) (Collaborative work)

CHAPTER 2

MATERIALS AND METHODS

2.1. *In silico* characterization of ten putative Globin-Coupled Sensors in *S. grandis*

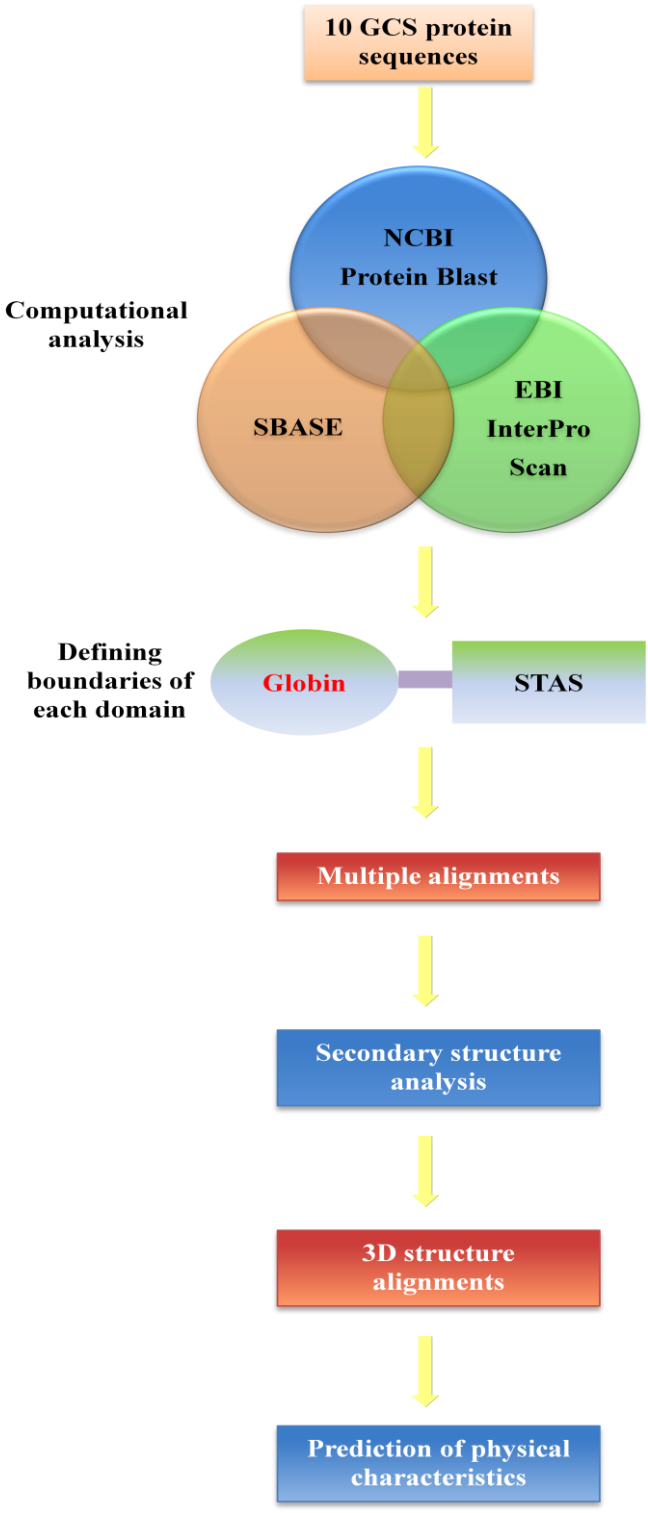


Figure 2.1. Overview of *in silico* characterization of ten GCS genes.

2.1.1. Protein BLAST, domain structure and function predictions

Protein sequences were analyzed with the BLASTP (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The results of each BLASTP led to the conserved domain database to determine any conserved domain present in protein sequences. Addition to protein BLAST, the InterProScan (http://www.ebi.ac.uk/Tools/pfa/iprscan/), and the SBASE (http://hydra.icgeb.trieste.it/servers/protein/sbase/) were performed to identify domain structures and functions of individual proteins.

2.1.2. Multiple alignments

All sequences were aligned by the Clustal W method in MegAlign software (DNASTAR). Globin domains of ten protein sequences were compared to three known globin domains of GCSs from HemAT-*Bs* (*B. subtilis*), HemAT-*Hs* (*H. salinarium*) and *Ec*Greg (*E. coli*). RsbR and four paralogs (YkoB, YojH, YqhA and YtvA) from *B. subilis* were selected to align the C-terminal STAS domains of each protein.

2.1.3. Structure analysis

Protean software from DNASTAR was used to predict secondary structure of each protein sequence. The Garnier-Robson method was chosen as the primary method. The selection of Protein Data Bank (PDB) proteins as database set in Protein BLAST was led to align a query sequence to a similar sequence from 3D structure. Aligned 3D structures were visualized by Cn3D program which was freely available from NCBI.

2.1.4. Prediction of physical characteristics of ten putative GCSs

Molecular weights and pI values of each amino acid sequence were calculated automatically by one of Lasergene application programs, SeqBuilder (DNASTAR). Hydrophobicity plots were also generated by Protean program (DNASTAR). Solubility of a

recombinant protein was calculated by using the web-based Recombinant Protein Solubility Prediction tool (http://www.biotech.ou.edu/).

2.2. *In vitro* characterization of ten putative Globin-Coupled Sensor in *S. grandis*

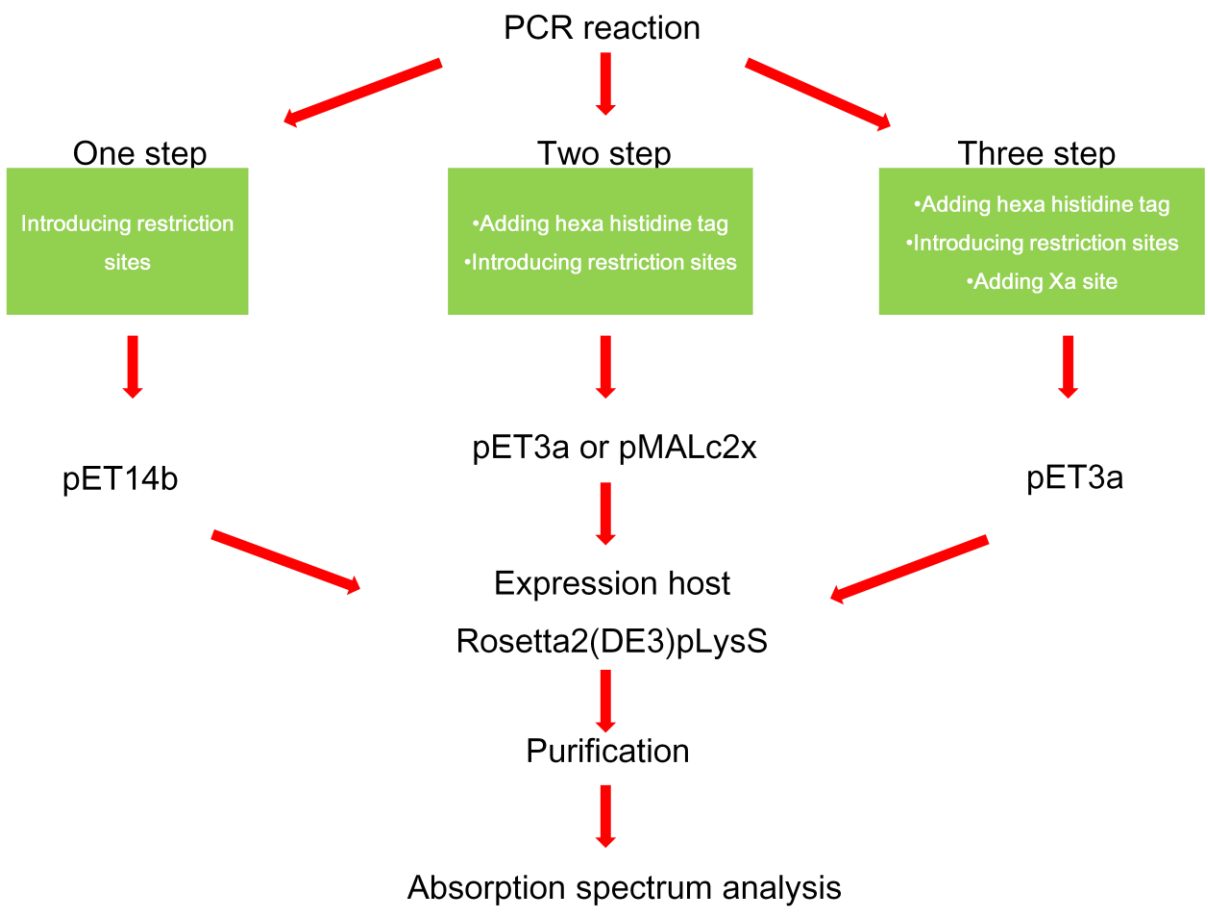


Figure 2.2. Overview of *in vitro* characterization of ten GCS genes.

2.2.1. Polymerase Chain Reaction (PCR)

Polymerase Chain Reactions (PCR) were performed using the high fidelity proofreading *Pfu Turbo* DNA polymerase (Agilent). The reactions were set up according to Table 2.1. Each reaction was started after a heat block temperature reached around 70℃. The general PCR setting used in this study was 94℃ for 2 minutes followed by 25 cycles of 94℃ for 30 seconds, various annealing temperatures (depending on primers) for 30 seconds, 72℃ for 50 seconds, with a final extension at 72℃ for 7 minutes. PCR products were analyzed by agarose gel electrophoresis. The list of primers on this study is on Table 2.2.

Table 2.1. PCR reaction setup

Reagent	Amount
DNA template	1 µl
10 X Pfu buffer	5 µl
dNTPs (2.5 mM or 10 mM)	4 or 1 µl
Primer 1 (100 ng/ml)	1 µl
Primer 2 (100 ng/ml)	1 µl
ddH_2O	37 or 40 µl
Pfu Turbo DNA polymerase (2.5 U/µl)	1 µl
Final volume	50 µl

Table 2.2. Primers for construction of the ten GCSs

Primer	sequence 5'-3'
SGRA_0570 Xa top	AAATT GAA GGA CGA ATG CAG ACA TTT ACA GCC GAA GC
SGRA_0570 his-tag top	AACAT CAC CAT CAC CAT CAC ATT GAA GGA CGA ATG CAG AC
SGRA_0570 NdeI top	AACAT ATG CAT CAC CAT CAC CAT CAC ATT G
SGRA_0570 BamHI bot	AAGGA TCC TTA GGA GGC GGC AAT GCG CAA TC
SGRA_1293 NdeI top	AAA ATA CAT ATG ATT GAT CCT AAC GAA AAA AAT TG
SGRA_1293 BamHI bot	AAA GGA TCC CTA AAA AAT ACT TTG ATT C
SGRA_2160 NdeI top	AAA ATA CAT ATG CAA GAA AGA GTT GTT TG
SGRA_2160 BamHI bot	AAA GGA TCC TTA TTC TTT CTG CTC
SGRA_2161 NdeI top	AAA ATA CAT ATG GCC ATA GTA AAA ATT AAC
SGRA_2161 his-tag bot	GTG GTG GTG GTG GTG GTG TAT TTG TCC AAT TGC
SGRA_2161 BamHI bot	AAA GGA TCC TTA TAT TTG TCC AAT TG
SGRA_2162 NdeI top	AAA ATA CAT ATG AAG AAA GGC GAA ATT TC
SGRA_2162 BamHI bot	AAA GGA TCC CTA GCC TTT TTG CTT TC
SGRA_2167 NdeI top	AAA ATA CAT ATG CTA GCT ATA AAA CG
SGRA_2167 BamHI bot	AAA GGA TCC TTA GGA TTG GGT TTG
SGRA_2168 NdeI top	AAA ATA CAT ATG GAA CTG AAA AAA TGG
SGRA_2168 BamHI bot	AAA GGA TCC TTA ATT TGC AGA TTG
SGRA_2169 NdeI top	AAA ATA CAT ATG AAG GAA ATT AAT ATC
SGRA_2169 BamHI bot	AAA GGA TCC TTA AAG ATT GCG ATT AAG GGC GG
SGRA_3210 NdeI top	ATA ATA CAT ATG TCA CTA ACC TGC TCC AAT TT
SGRA_3210 BamHI bot	ATAGGA TCC CTA TTT TTT AAT TTG TCG AAG TTC
SGRA_3852 EcoRI top	AAGAA TTC ATG CAG CCC CTT AAA GTT TTA AG
SGRA_3852 his-tag bot	GTG GTG GTG GTG GTG GTG AAT CGC CAA TTC TTG
SGRA_3852 BamHI bot	AAGGA TCC TTA GTG GTG GTG GTG GTG GTG

Colony PCRs were performed to verify insert-containing plasmids. A single colony was picked and inoculated in a 50 µl ddH_2O contained microcentrifuge tube. The microcentrifuge tube was vortexed and 2 µl aliquot was dropped on 1.2% agar plate containing appropriate antibiotics for later use. The inoculated colony was heated at 99℃ for 10 minutes and centrifuged at the maximum speed. 10 µl of supernatant was added into a PCR master mix and the reaction was setup as follows:

Table 2.3. Colony PCR reaction setup

Reagent	Amount
DNA template	1 µl
5x GoTag Green buffer	5 µl
dNTPs (2.5 mM or 10 mM)	4 or 1 µl
Primer 1 (100 ng/ml)	1 µl
Primer 2 (100 ng/ml)	1 µl
PCR water	To 50 µl
GoTag Hot Start polymerase (5 U/µl)	0.25 µl
Final volume	50 µl

The PCR setup was the same as above.

2.2.2. PCR purification

PCR products were purified using the QIAquick PCR Purification kit (Qiagen). The PCR sample was mixed with 5 volumes of Buffer PBI, applied to a QIAquick spin column, and centrifuged for 1 minute at 13,000 rpm. The flow-through was discarded and 750 µl Buffer PE was added to wash the column. After 1 minute centrifugation, the flow-through was discarded and additional 2 to 4 minutes centrifugation was done to remove residual

ethanol. The QIAquick column was placed in a clean 1.5 ml microcentrifuge tube and air-dried for 20 minutes. To elute DNA, 50 µl of Buffer EB or ddH_2O was added to the center of the QIAquick membrane, allowed to stand for 1-2 minutes, and then centrifuged for 1 minute.

2.2.3. PCR cloning

The genes of interest were amplified by PCR reactions. The amplified PCR products were sub-cloned into a Blunt DNA cloning vector using the one-step cloning strategy of the Zero Blunt TOPO PCR Cloning Kit (Invitrogen) or the CloneJET™ PCR Cloning Kit (Fermentas) to increase cloning efficiency of downstream applications. The reaction was setup according to Table 2.4 or Table 2.5. The reactions were then placed on ice for 5 minutes and transformed into Mach1-$T1^R$ or DH5α.

Table 2.4. TOPO cloning reaction setup

Reagent	Amount
PCR product	0.5-2 µl
ddH_2O	To 3 µl
Salt	0.5 µl
TOPO vector	0.5 µl
Final volume	3 c

Table 2.5. pJET cloning reaction setup

Reagent	Amount
PCR product	1 µl
ddH_2O	To 20 µl
2x Reaction buffer	0.5 µl
T4 DNA ligase	1 µl
pJET vector	1 µl
Final volume	20 µl

2.2.4. Restriction enzyme digestion

Blunt-end products with specific restriction enzyme sites at the 5' end and 3' end were engineered by PCR. These products were digested with the appropriate restriction enzymes to produce compatible ends for cloning into the appropriate vector. Restriction enzymes were purchased from Promega. The restriction enzyme digestion setup as indicated in Table 2.6 and incubated in a 37℃ water bath for 1 hour. The vector was dephosphorylated by adding 1 µl of Antartic phosphatase (1 U/µl) (New England Biolabs) and 10X Antartic phosphatase buffer and incubated at 37℃ for an additional 1 hour. This was done occasionally to prevent self-ligation of partially digested vector.

Table 2.6. Restriction enzyme digestion setup

Reagent	Amount
DNA	5 to 16.8 µl
10X enzyme buffer	2 µl
Enzyme (s) (10U/µl)	1 µl each
ddH_2O	To 20 µl
Bovine Serum Albumin	0.2 µl
Final volume	20 µl

Alternatively, screening for insert-containing plasmids was done by a mini-digestion. The reaction was setup as follows:

Table 2.7. Mini-digestion setup

Reagent	Amount
DNA	5 µl
10X enzyme buffer	1.5 µl
Enzyme (s) (10U/µl)	0.5 µl each
ddH_2O	To 15 µl
Bovine Serum Albumin	0.15µl
Final volume	15 µl

2.2.5. Extraction of DNA from Agarose gels

DNA was extracted from agarose gels using the Geneclean Spin Kit (Qbiogene). The digested DNA was run on an agarose gel and the desired bands of DNA were excised out and placed in a Geneclean Spin filter. 400 µl of Geneclean Spin Glassmilk was added to the filter and heated at 55℃ to melt the gel, mixing every 1-2 minutes. The tube was centrifuged at 11,000 rpm for 1 minute and the flow-through was discarded. 500 µl of ice-cold Geneclean Spin New Wash was added to the filter, centrifuged at 11,000 rpm for 1 minute and the flow-through was discarded. This washing step was repeated one more time. The filter was centrifuged for an additional 4 minutes, transferred to a new catch tube and heated at 55℃ to remove residual ethanol (until Glassmilk powder became loose). To elute, 15 µl of ddH_2O was added to the filter, allowed to stand for 1-2 minutes and then centrifuged for 1 minute.

2.2.6. Ligation

The DNA purified by Geneclean was ligated overnight at 15℃.

Table 2.8. Ligation reaction setup

Reagent	Amount
Digested DNA (purified)	8 µl
10X T4 DNA ligase buffer	1 µl
T4 DNA ligase	1 µl
Final volume	10 µl

2.2.7. Transformation of *E. coli* competent cells

Either 5 µl of the ligation mixture, 5 µl pJET cloning reaction mixture, or 2 µl TOPO cloning reaction mixture was mixed with 25 µl of Mach1-$T1^R$ (Invitrogen) or 100 µl of DH5α competent cells. In another instance, 1 µl of plasmid was mixed with 25 µl of

Rosetta2(DE3)pLysS (Novagen) and incubated on ice for 10 or 30 minutes. The cells were heat shocked at 42℃ for 20 or 45 seconds and immediately placed on ice for 2 minutes. 250 or 450 or 950 µl of SOC or LB medium was added to the cells and incubated at 37℃ for 1 hour with shaking. After incubation, the cells were centrifuged at 5,000 rpm for 2 minutes. All but 50 µl of residual supernatant was discarded. The 50 µl of residual supernatant was used to resuspend the pellet. The resuspended cells were spread on 1.2% LB agar plates containing the appropriate antibiotics and incubated for 12-16 hours or until colonies grew around 1 mm (diameter)

2.2.8. Plasmid isolation

Two different kits were used to isolate plasmids in this study. One is the QIAprep Spin Miniprep Kit (Qiagen) and another is the illustra plasmidPrep Mini Spin Kit (GE Healthcare). Single colonies were inoculated into 3 ml LB medium and incubated overnight at 37℃ with shaking.

For the QIAprep Spin Miniprep Kit, the cells were harvested by centrifugation at 8,000 rpm for 3 minutes and the cell pellet was resuspended in 250 µl of buffer P1. 250 µl of buffer P2 was added and then gently mixed by inverting 4 to 6 times. Then, 350 µl of buffer N3 was added and mixed again by inverting the tubes. The tubes were centrifuged at 13,000 rpm for 10 minutes. The supernatant was transferred to QIAprep spin columns and centrifuged at 13,000 rpm for 1 minute. The flow-through was discarded. The spin columns were washed by adding 750 µl of buffer PE followed by centrifugation at 13,000 rpm for 1 minute. The spin columns were spun at 13,000 rpm for additional 2 to 4 minutes to remove residual ethanol and transferred to new microcentrifuge tubes. The tubes were air-dried for 20 minutes and 50 µl of buffer EB or ddH_2O was added. The tubes were set for 1 minute and centrifuged at 13,000 rpm for 1 minute to elute DNA.

For the illustra plasmidPrep Mini Spin Kit, the cells were harvested by centrifugation at maximum speed for 1 minute and the cell pellet was resuspended in 175 µl of Lysis buffer

type 7. 175 µl of Lysis buffer type 8 was added and then gently inverted until solution became clear and viscous. Then, 350 µl Lysis buffer type 9 was added and mixed gain until the precipitate was evenly formed. The tubes were centrifuge at 14,000 rpm for 4 minutes. The supernatant was transferred into the spin columns and centrifuged at 14,000 rpm for 1 minute. The flow-through was decanted. 400 µl Wash buffer type 1 was added and centrifuged for 1 minute. After discarding the flow-through, additional 2 minutes centrifugation was done. The columns were transferred to new microcentrifuge tubes. The tubes were air-dried for 20 minutes and 50 µl of Elution buffer was added. The tubes were set for 1 minute and centrifuged at 13,000 rpm for 1 minute to elute DNA.

2.2.9. Small Scale protein expression

To determine optimum conditions of target protein expression and solubility, small scale induction was done before going to large scale induction. Culture grown overnight were inoculated in 50 ml fresh LB medium with appropriate antibiotics and incubated at 37℃ with shaking until the OD_{600} reached 0.6 or 1.0. 1 ml of culture was removed and labeled as uninduced. The remaining culture was transferred into a 50 ml centrifuge tube for 15 ml each. Different concentration of isopropyl-β-D-thiogalactopyranoside (IPTG) was added in each tube for induction. At this point, Heme Precursor containing 100 mg/L $FeSO_4 \cdot 7H_2O$ and 17 mg/L δ-aminolevulinic acid hydrochloride (ALA) was also added to enhance heme biosynthesis. Proteins were induced at 30℃ or 32℃ or 37℃. 1 ml of culture was collected according to a time of induction for 24 hours. Proteins were extracted from the culture using the BugBuster protein extraction reagent (Novagen).

2.2.10. Protein Extraction via BugBuster Protein Extraction Reagents

Culture was harvested by centrifugation at 9,700 rpm for 10 minutes. The cells were resuspended at room temperature in BugBuster reagent (Novagen). Benzonase nuclease (25

units per ml of BugBuster reagent) was added and incubated on a shaking platform at a slow setting for 10 minutes at room temperature. Insoluble debris were removed by centrifugation at 13,000 rpm for 20 minutes at 4℃. The pellet and supernatant were analyzed by SDS-PAGE to check solubility of the protein.

2.2. 11. Protein expression in *E. coli*

500 ml of fresh LB medium containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol was inoculated with 5 ml of an overnight culture (Rosetta2(DE3)pLysS strain, Novagen). The culture was grown at 37℃ with 250 rpm shaking until the OD_{600} reached 1.0. Protein expression was induced with 0.05-0.6 mM IPTG based on expression level and solubility of the fusion protein. The heme precursor was added to the culture at this point. The culture was further induced at 25℃ or 30℃ for 4 hours to 15 hours. Cells were harvested by centrifugation at 4℃ and stored at -20℃ prior to purficiation.

2.2.12. Protein purification of His-tagged Proteins by Gravity-Flow Chromatography

The cell pellet from 500ml culture was thawed on ice and resuspended in 10 ml lysis buffer containing 50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole and 1% glucrose at pH7.6. 125 unit Benzonase (Novagen) and 1:200 dilution of protease inhibitor cocktail EDTA-free (Calbiochem) were added and incubated at room temperature for 10 minutes with a gentle agitation. After incubation, the mixture was sonicated on ice bath for 4 minutes (20 seconds pulses and 30 seconds pauses, Sonic Dismembrator 550, Fisher Scientific). The crude cell extract was then centrifuged at 15000 rpm for 20 minutes to remove insoluble materials at 4℃. The supernatant was filtered through a syringe filter (0.22 μm pore size, Corning). After this step, the total protein quantification was carried by QuickStart Bradford protein assay kit (Bio-Rad). Purification steps were carried out in two ways either a column purification method or a batch purification method for proteins that did not bind a resin.

For the batch purification, the total amount of crude extract was mixed with 50% Ni-NTA agarose slurry (Qiagen) depending on its concentration and a binding capacity of resin. The mixture was spun on a magnetic stirrer at 4℃ for 3 hours to overnight to promote binding of histidine tags onto resins. The incubated crude extract and resin mixture was slowly loaded onto a polypropylene column for gravity flow chromatography (Qiagen). The different concentration of imidazole (20, 50 and 100 mM) containing buffers (300 mM NaCl, 50 mM Na_2HPO_4, with or without 1% glucose at pH7.6 or 8.0) were used for wash buffers in stepwise fashion. The column was washed with at least 4 times to 120 times bed volume of wash buffers. After extensive washing steps, proteins were eluted with 1 time to 2 times bed volume of elution buffers (100 mM or 150 mM or 300 mM imidazole and 300 mM NaCl, 50 mM Na_2HPO_4, with or without 1% glucose at pH7.6 or 8.0). All purification steps were performed at 4℃ and all fractions were analyzed by SDS-PAGE. Eluted proteins were dialyzed against a buffer containing 300 mM NaCl, 50 mM Na_2HPO_4, with or without 1% glucose at pH7.6 or 8.0.

For the column purification, 50% Protino Ni-NTA agarose slurry (Macherey-Nagel) was poured into a polypropylene column and set until it settled. 10 times bed volume of lysis buffer containing 300 mM NaCl, 50 mM Na_2HPO_4, 20 mM imidazole at pH8.0 was applied to equilibrate the column. After equilibration of the column, proper amount of pre-measured crude extract was loaded. The rest of purification steps were the same as above.

2.2.13. Absorption spectrum analysis

Absorption spectra of purified RsbR proteins were measured in the buffer used for dialysis using a Cary Bio 50 Spectrophotometer (Varian). The proteins that displayed oxygenated myoglobin peaks were treated with few grains of sodium dithionite at room temperature until color of protein solutions turned from red to yellow to measure peaks of

deoxygenated proteins.

2.2.14. Initial crystallization of SGRA_3210 (RsbR1)-Collaboration work

The crystallization of SGRA_3210 (RsbR1) was done by Dr. Farrukh Jamil at Universiti Sain Malaysia as part of collaboration. Purified protein sample was filtered through Millipore Centrifugal filters, 0.22 µm to remove precipitate and dust particles. Preliminary crystallization of RsbR was performed by using commercially available kits that are; Crystal Screens 1 and 2 (Hampton Research), Structure Screens 1 and 2 (Molecular Dimensions), Wizard I, II, III and IV (Emerald BioSystems). Vapor diffusion method was used for crystallization in 96-well plate. For each well 1 µl of protein (10 mg/ml) was mixed with 1 µl of the reservoir buffer, after that plate was sealed with sealing tape (sitting drop method) or with siliconised cover slip (for hanging drop). Plates were incubated at 20^{o}C & 4^{o}C. Each drop is visually inspected and scored on the first, second and third day and then every week for one month.

CHAPTER 3

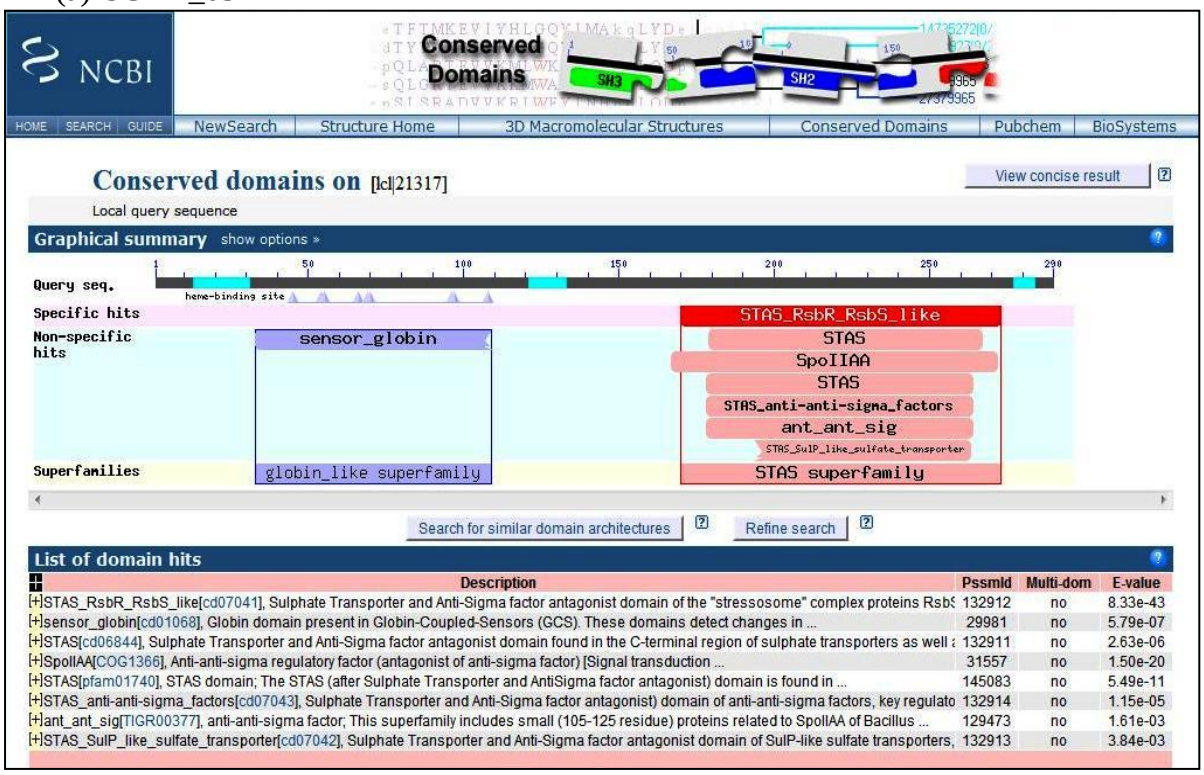
RESULTS

3.1. *In silico* characterization of ten putative GCSs

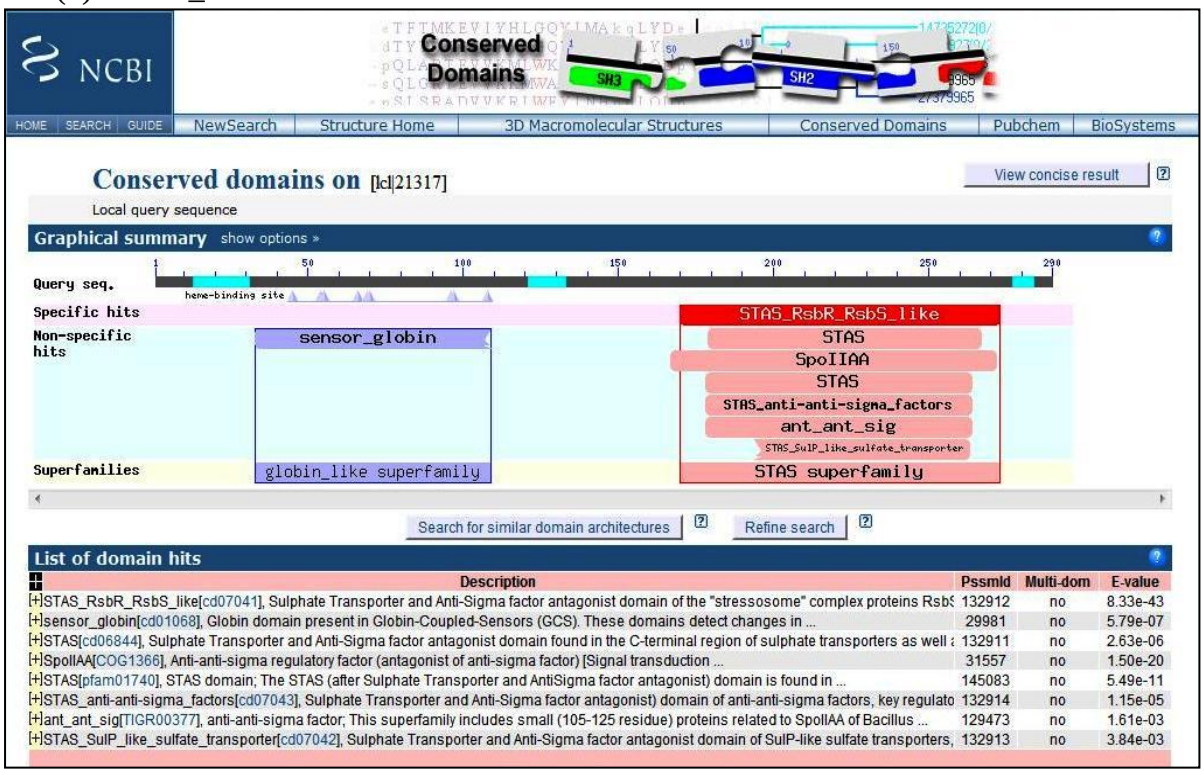
3.1.1. Protein blast, domain structure, putative function and physical locations of ten GCS genes

All ten GCS protein sequences were subjected to computational analysis to identity a domain architecture of a protein and to determine a possible function. Protein BLAST was the first tool to use and it was a bridge to Conserved domain database (CDD) for searching a functional unit that shared a similarity with other organisms. Eight out of ten protein sequences got conserved domains on an N-terminal sensor_globin domain that belongs to globin superfamily and a C-terminal STAS_RsbR_RsbS_like domain which is a member of STAS superfamily. SGRA_2162 and SGRA_2168 only got a hit on STAS_RsbR_RsbS_like domain in the C-terminal region (Figure 3.1). Results of conserved domains showed not only homologous domains but also displayed heme binding sites which were crucial for ligand binding such as diatomic oxygen. However, the proteins without the sensor_globin domain did not show any catalytic binding site.

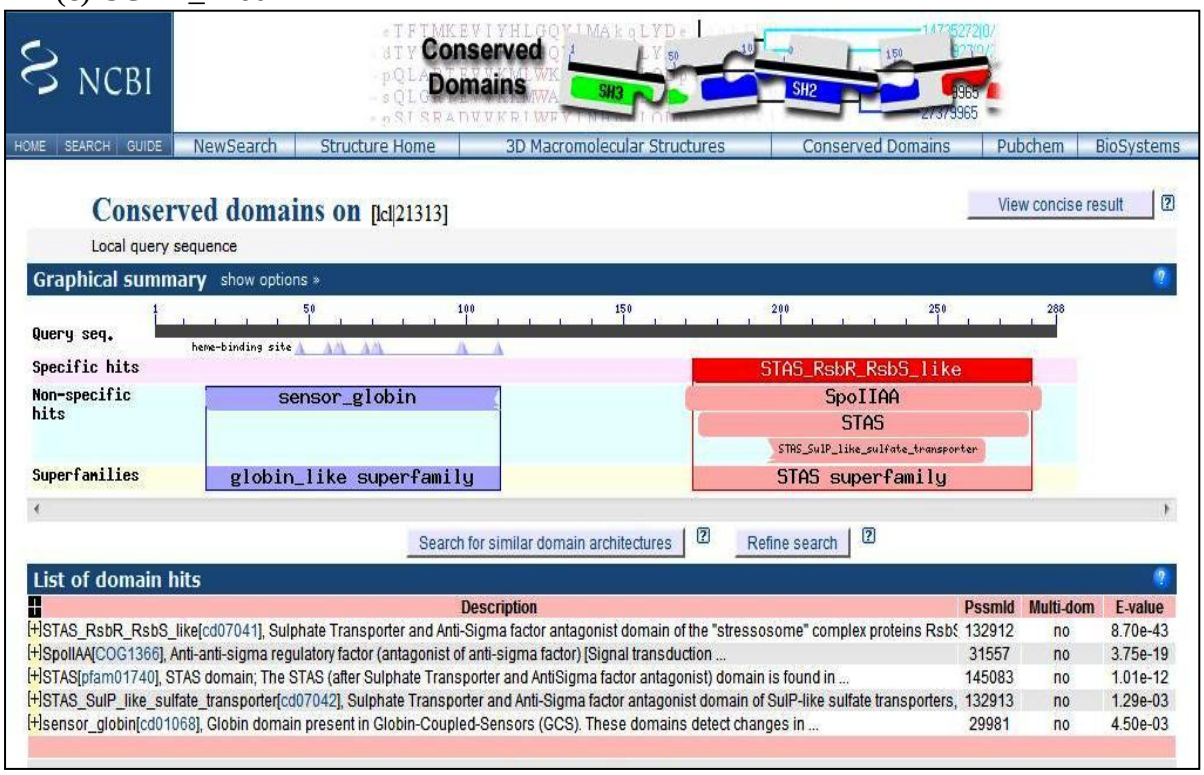
(a) SGRA_0571



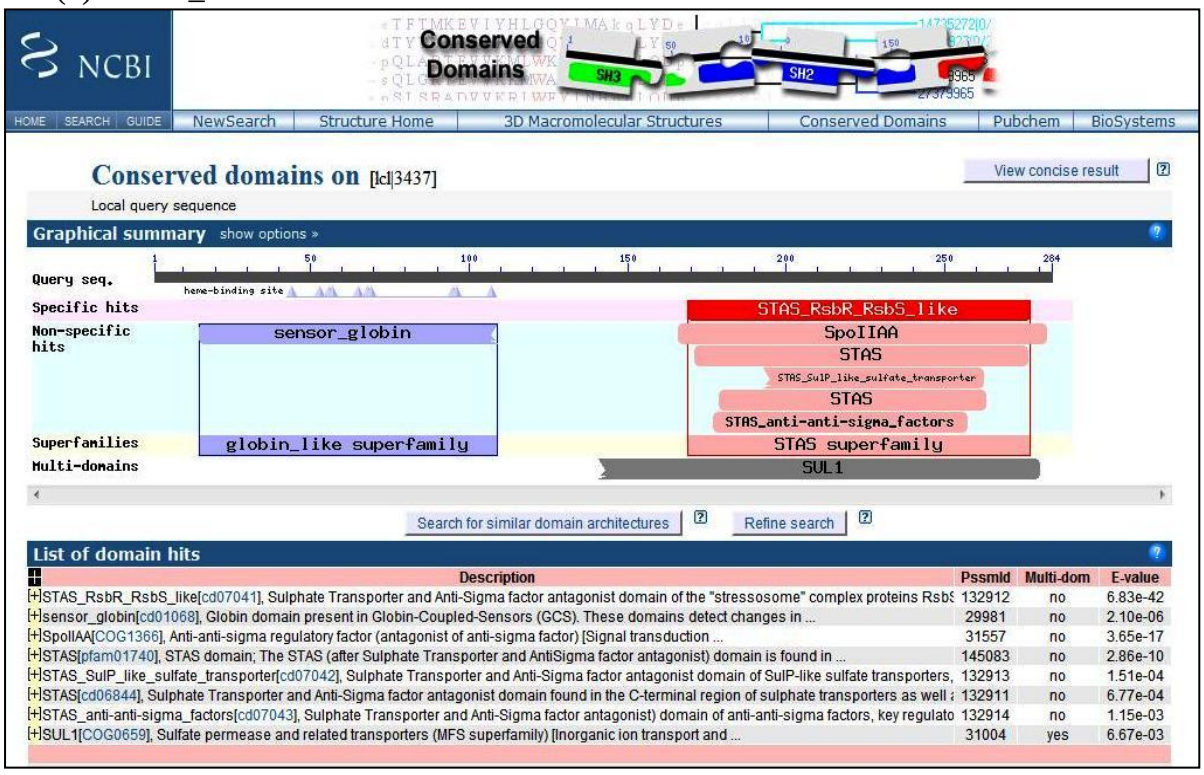
(b) SGRA_1293



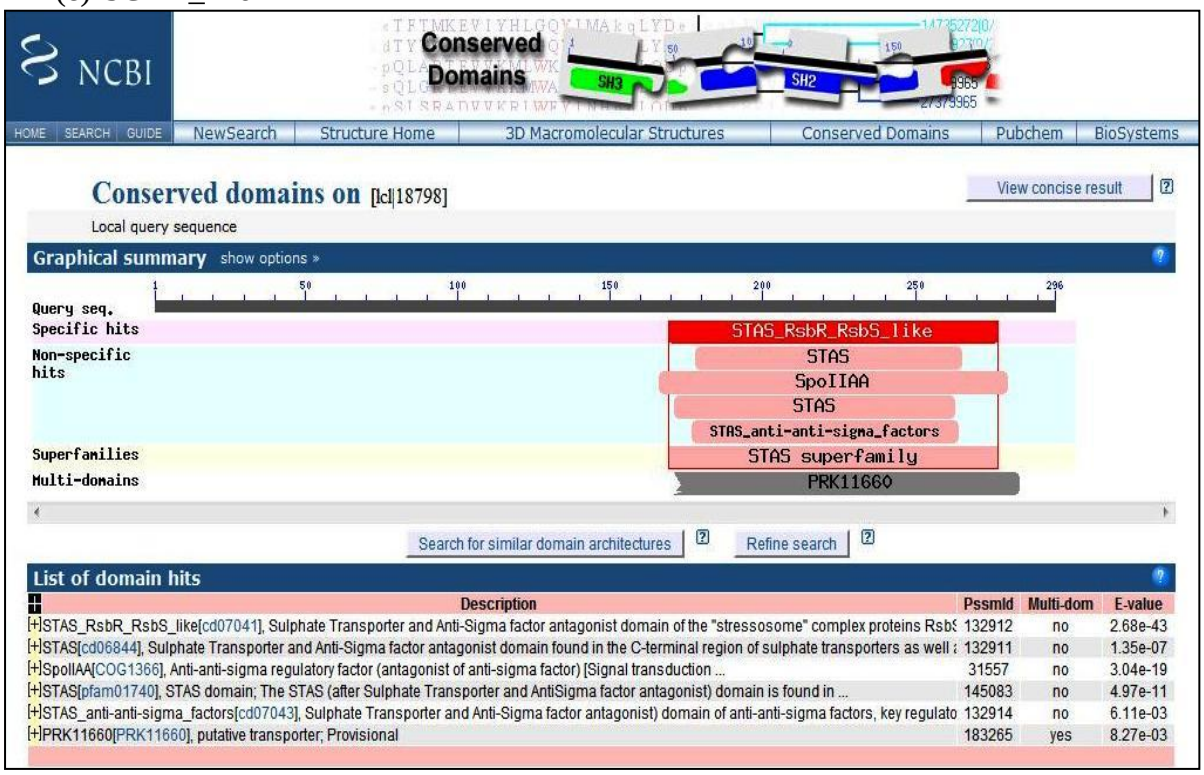
(c) SGRA_2160



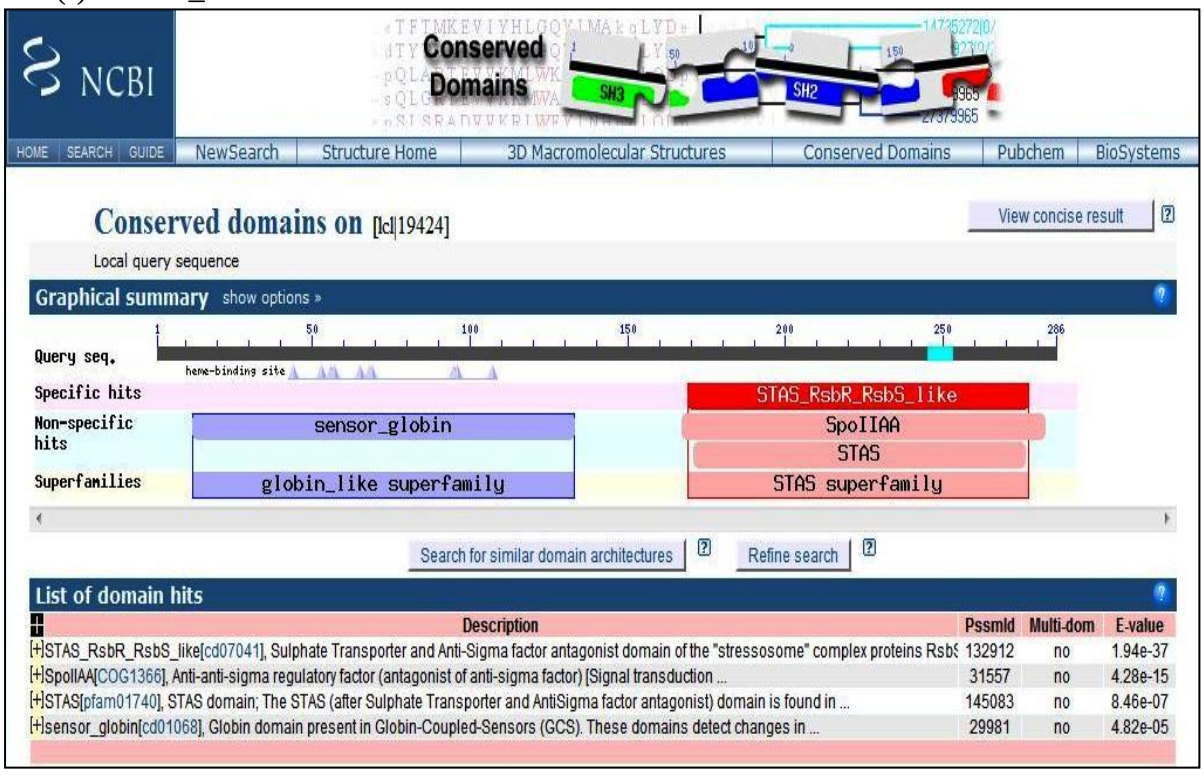
(d) SGRA_2161



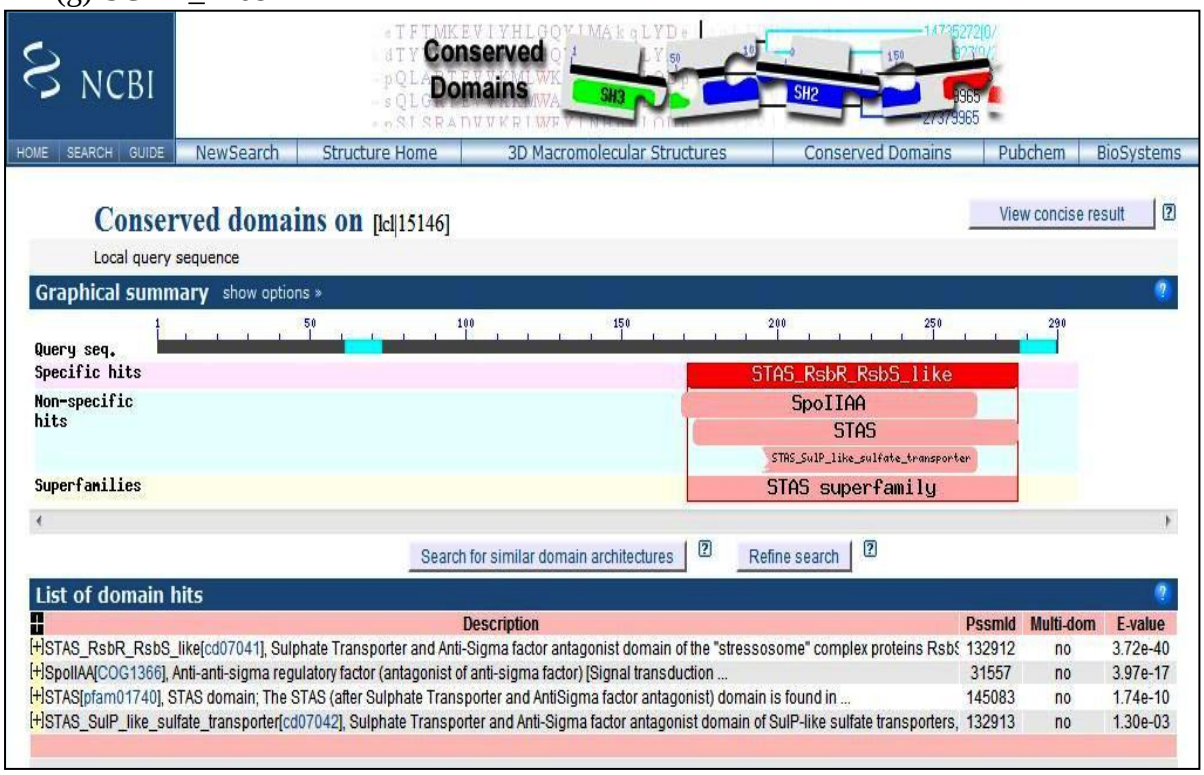
(e) SGRA_2162



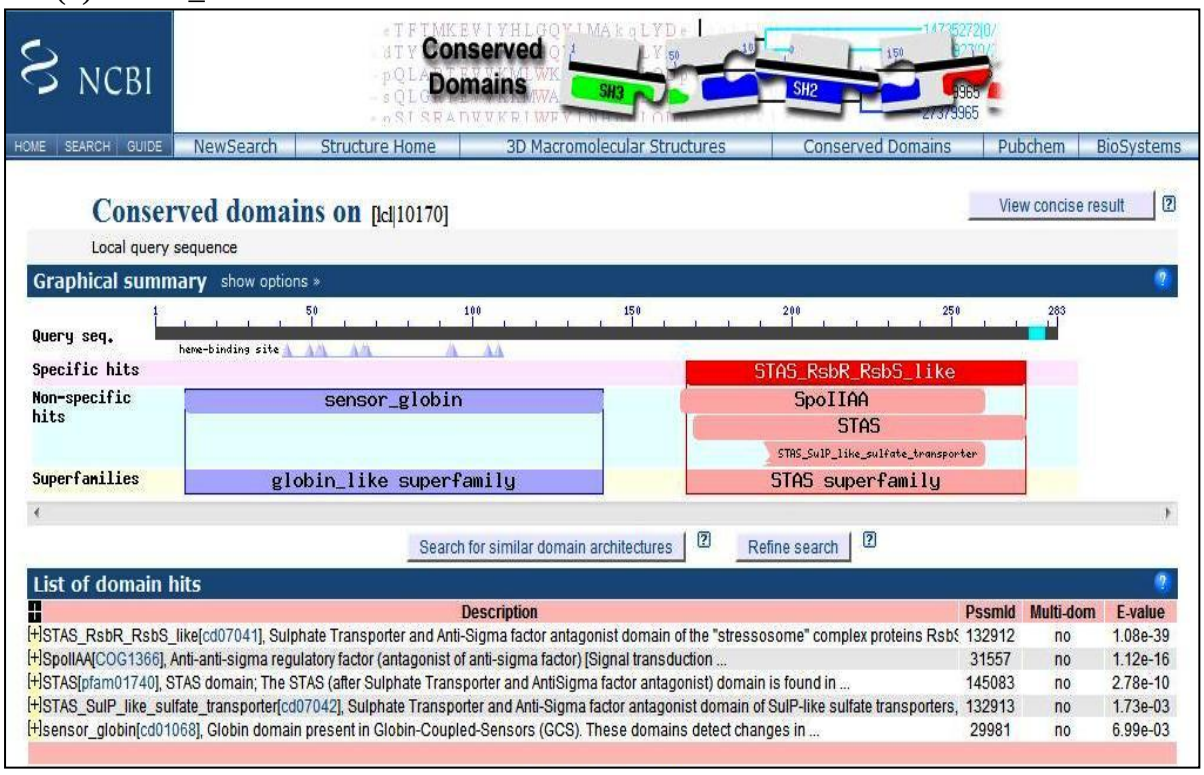
(f) SGRA_2167



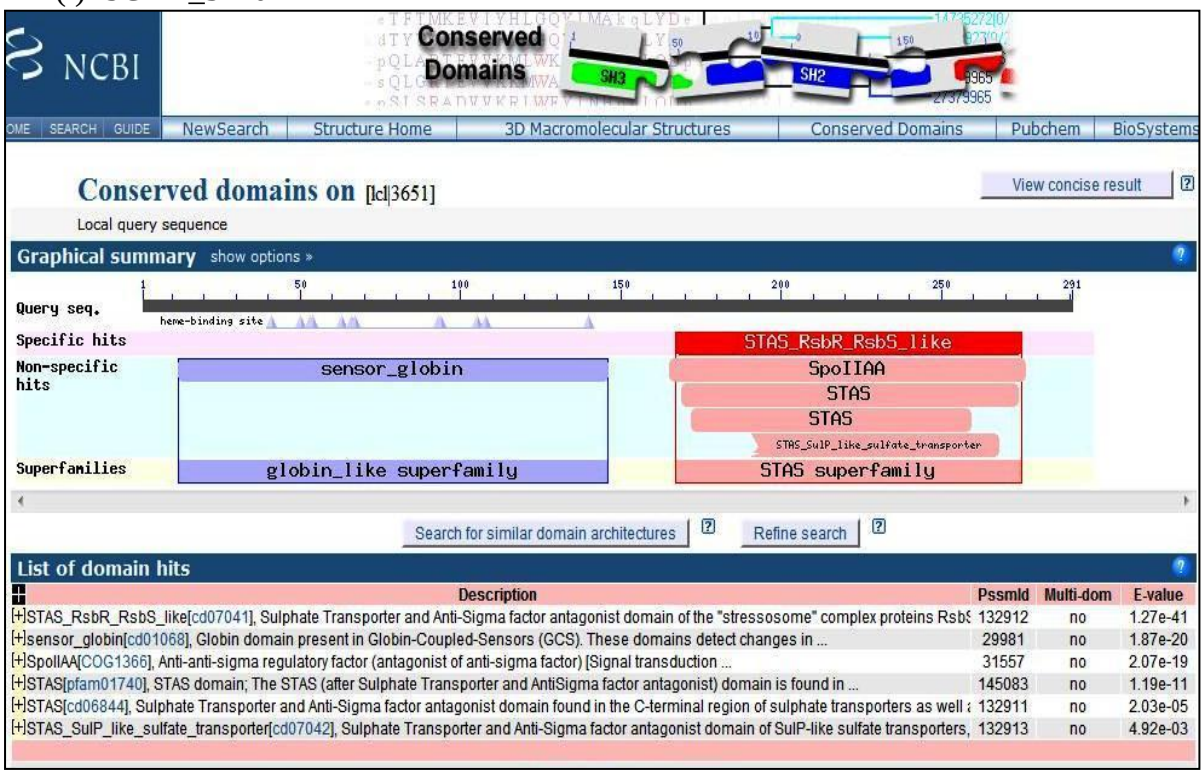
(g) SGRA_2168



(h) SGRA_2169



(i) SGRA_3210



(j) SGRA_3852

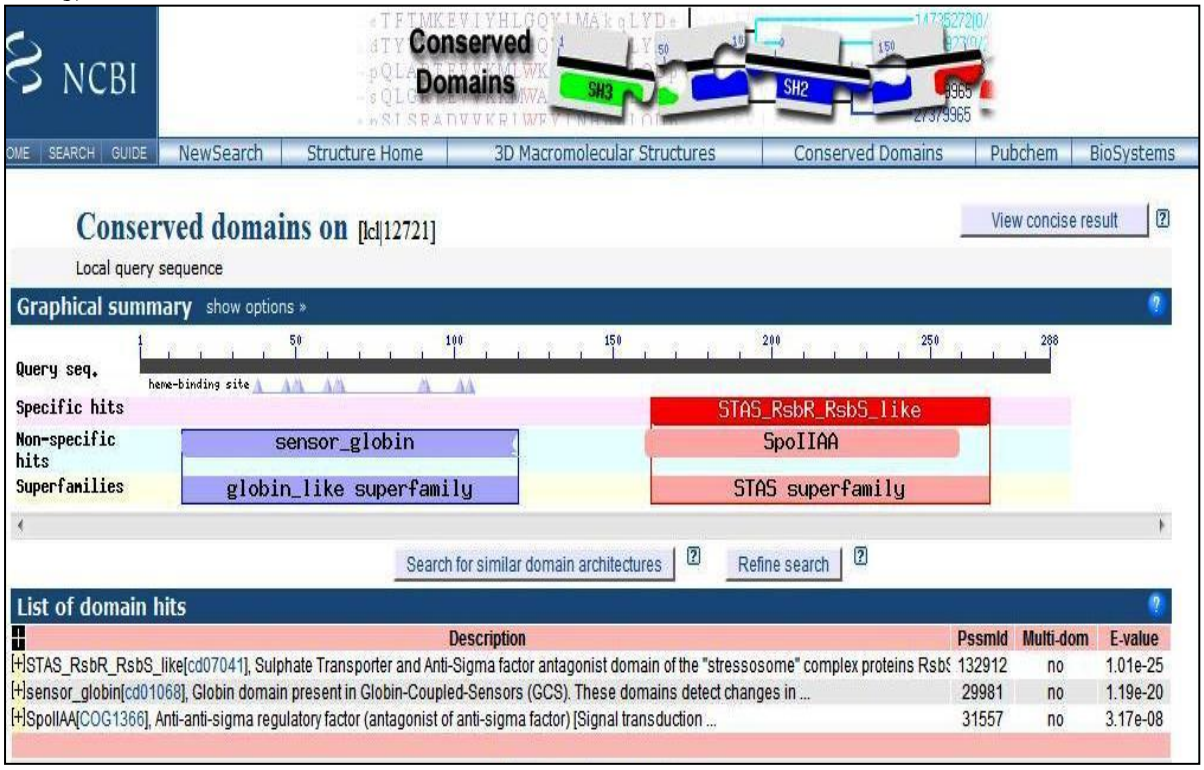
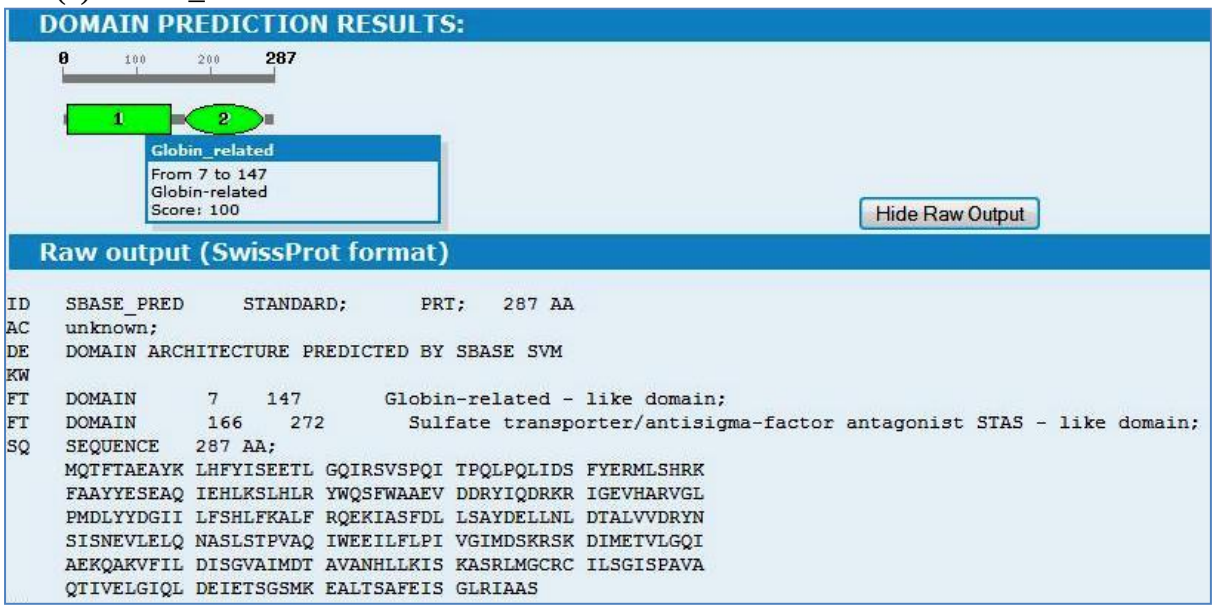


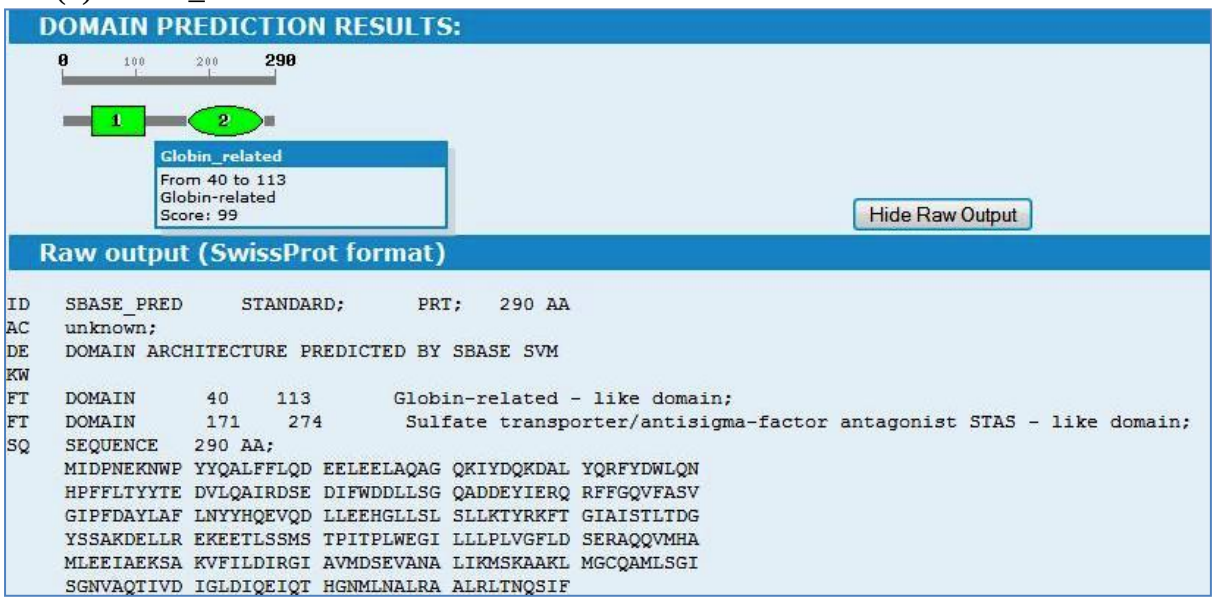
Figure 3.1. Conserved domains of 10 GCSs. (a) SGRA_0571 (b) SGRA_1293 (c) SGRA_2160 (d) SGRA_2161 (e) SGRA_ 2162 (f) SGRA_2167 (g) SGRA_2168 (h) SGRA_2169 (i) SGRA_3210 (j) SGRA_3852.

Since protein BLAST results did not all match with annotation results (Figure 1.1), another bioinformatics tool was used. SBASE is an online protein domain identification tool by analyzing input sequences through support vector machines method. Its data resources are UniProt, PIRSF, InterPro Database, Pfam, SMART and PRINTS (Vlahovicek *et al.*, 2005). Individual sequences were input for analysis. Returned results were on Figure 3.2. Two domains were recognized as Globin-related-like domain and STAS-like domain in all ten protein sequences including SGRA_2162 and SGRA_2168.

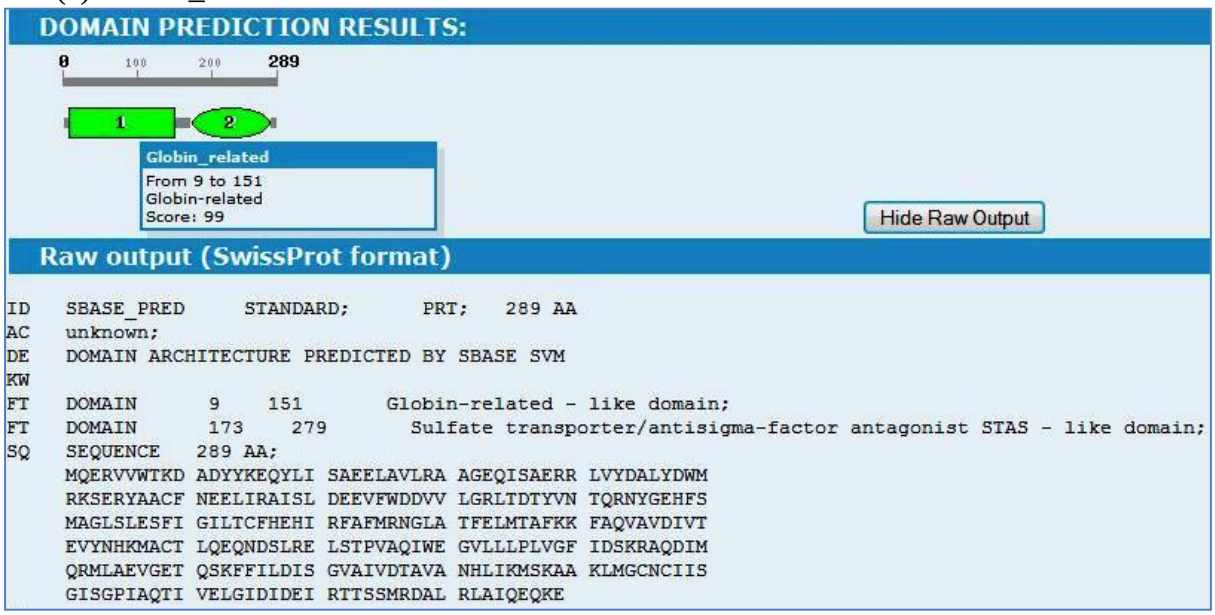
(a) SGRA_0571



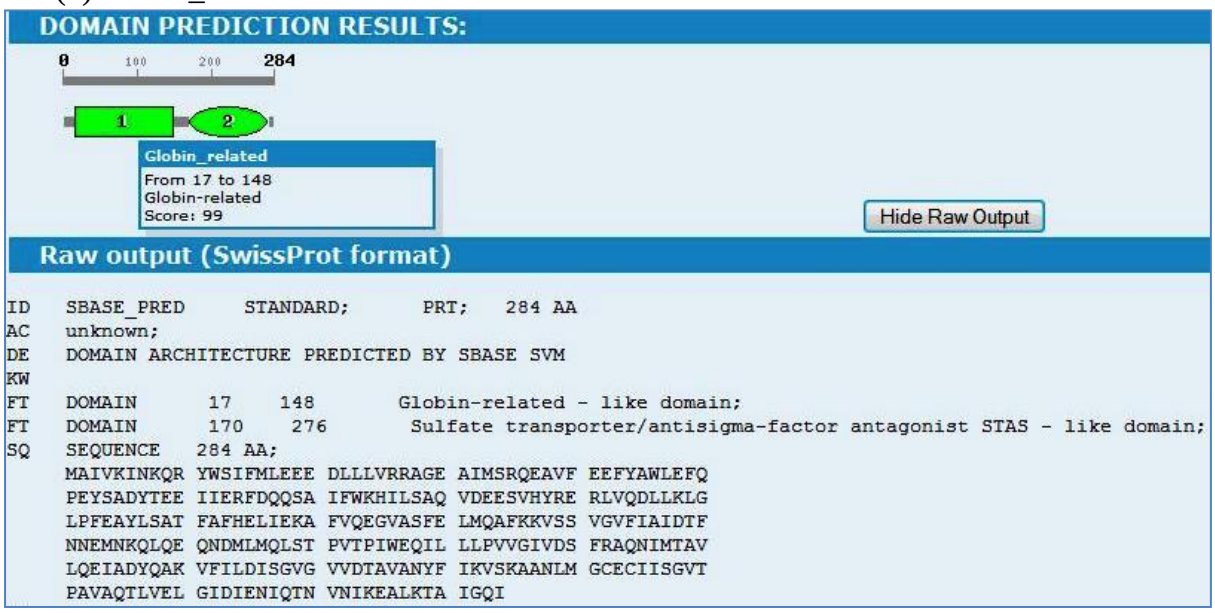
(b) SGRA_1293



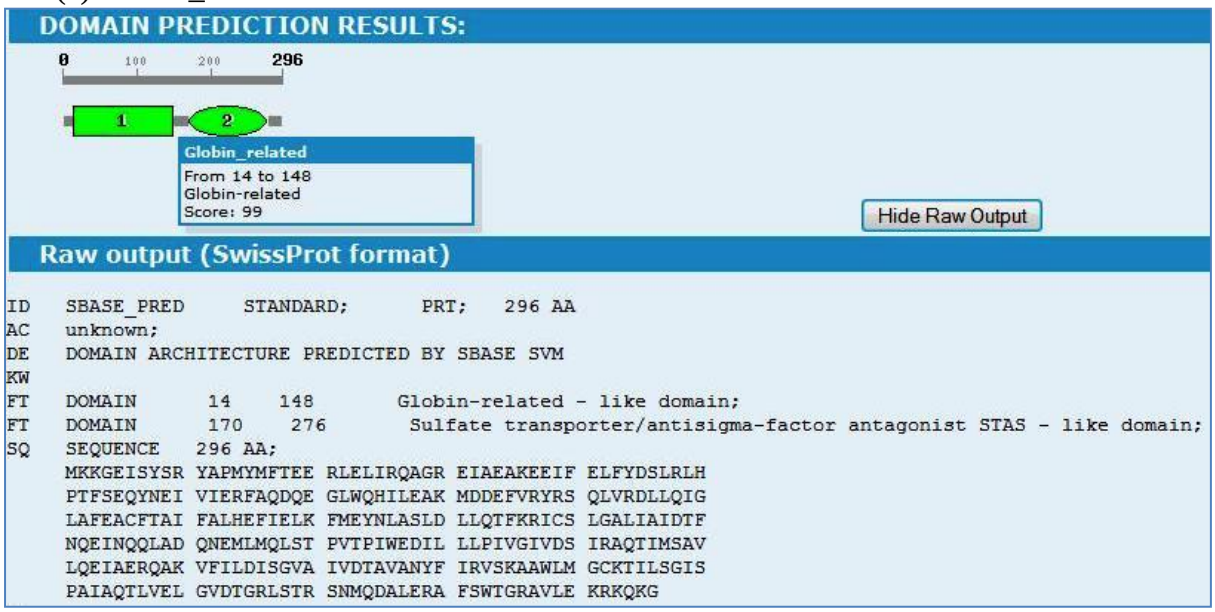
(c) SGRA_2160



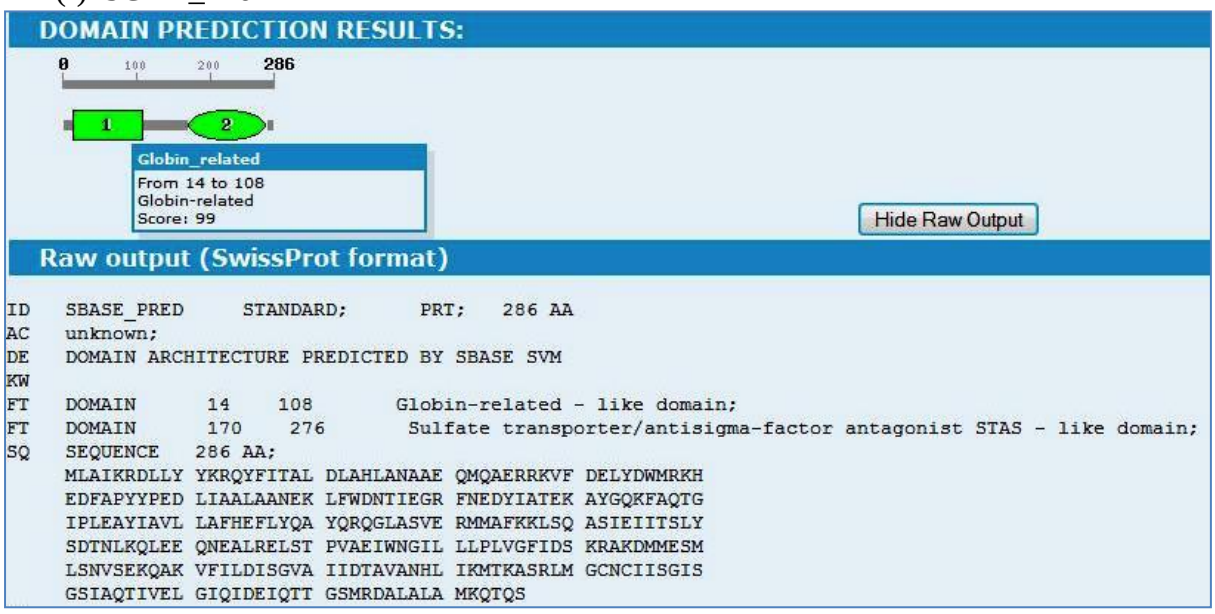
(d) SGRA_2161



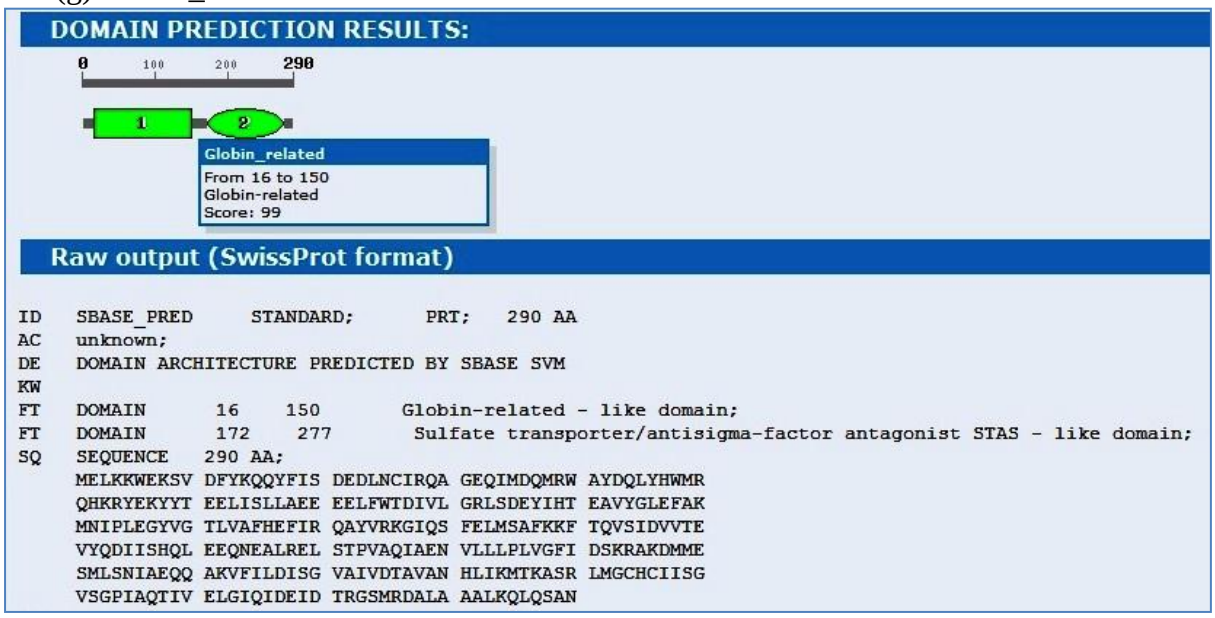
(e) SGRA_2162



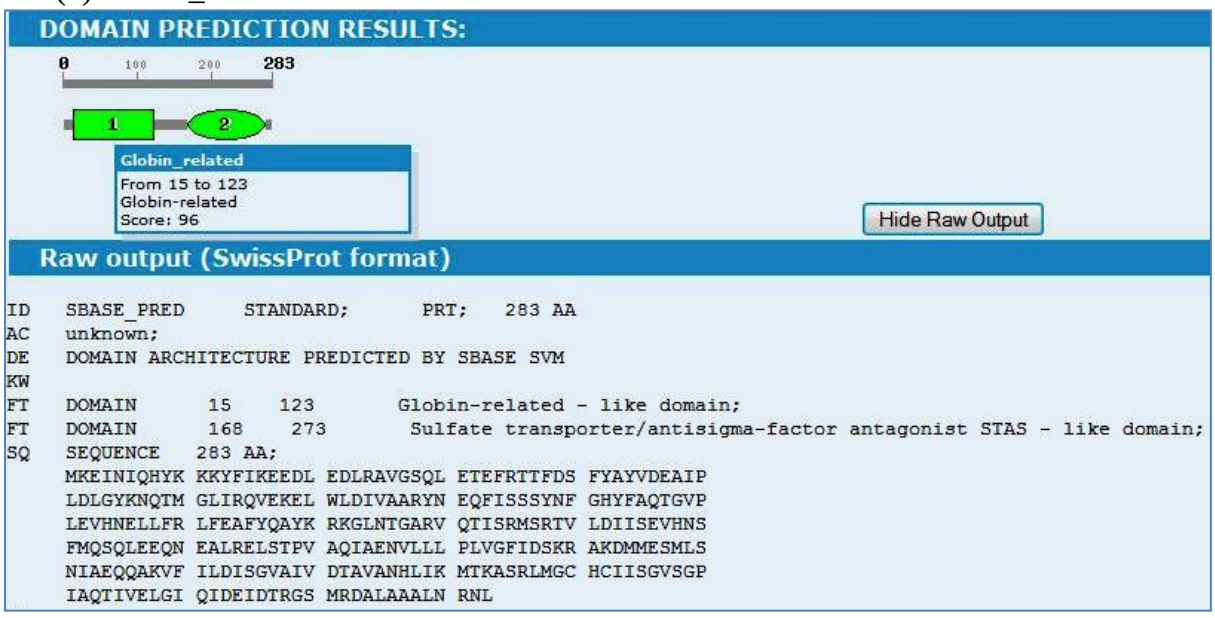
(f) SGRA_2167



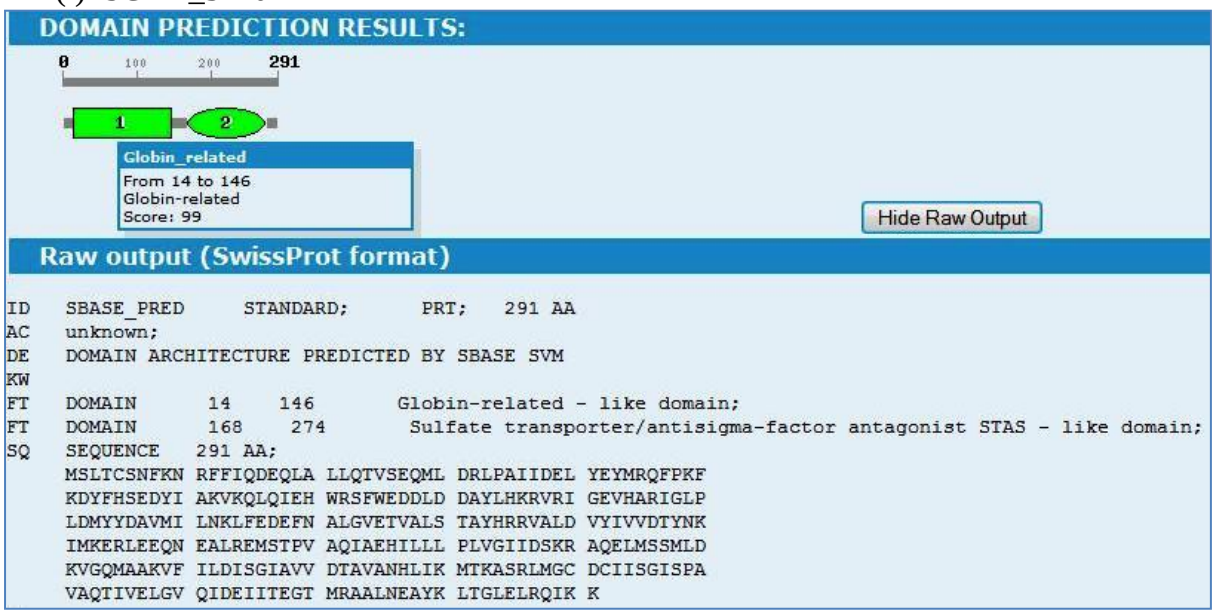
(g) SGRA_2168



(h) SGRA_2169



(i) SGRA_3210



(j) SGRA_3852

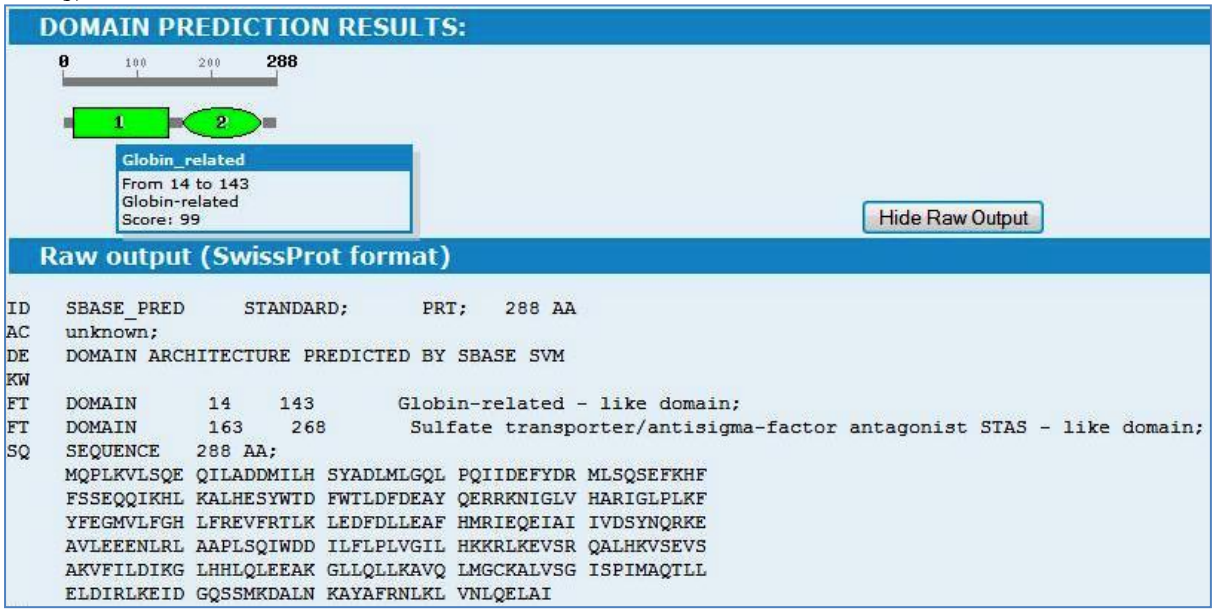
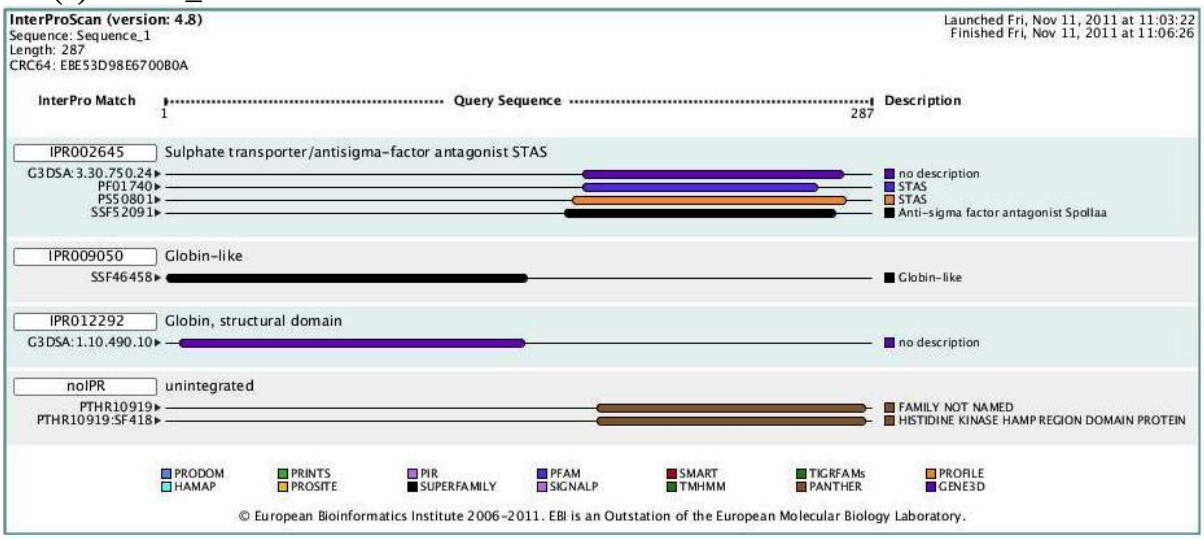


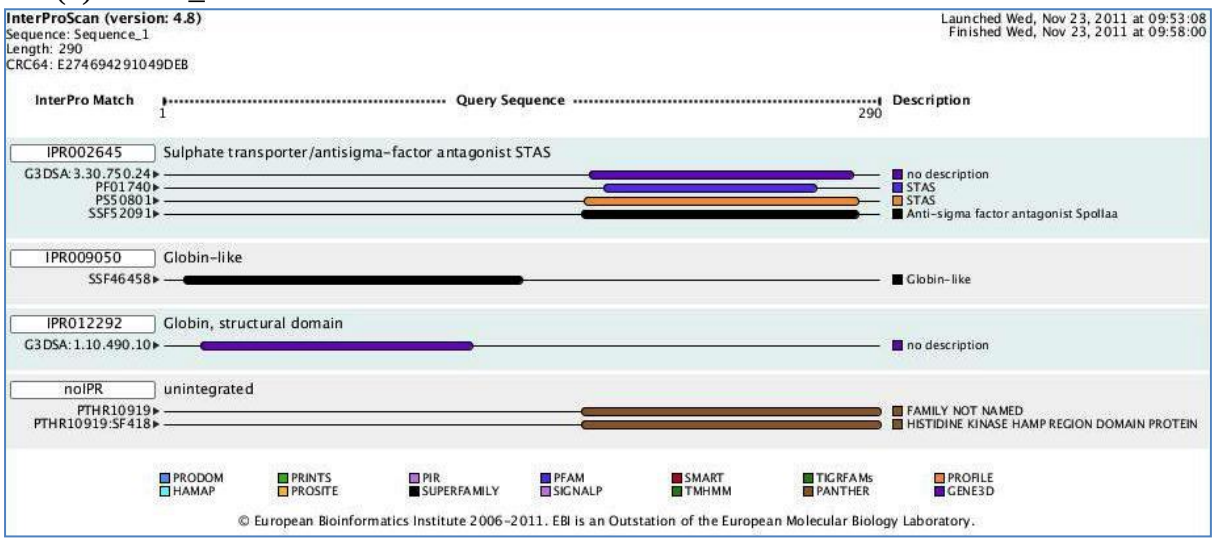
Figure 3.2. Domain predictions of 10 GCSs via SBASE. (a) SGRA_0571 (b) SGRA_1293 (c) SGRA_2160 (d) SGRA_2161 (e) SGRA_ 2162 (f) SGRA_2167 (g) SGRA_2168 (h) SGRA_2169 (i) SGRA_3210 (j) SGRA_3852.

To verify protein architecture and possible function of ten GCS proteins, protein sequences were analyzed by InterProScan. The InterProScan is an integrative protein search tool that actually scans input sequences through InterPro and its secondary databases including ProDom, PRINTS, SMART, TIGRFAMS, Pfam, PROSITE, PIRSF, Superfamily, CATH, PANTHER and Gene 3D (Hunter *et al*., 2009). Ten GCS protein sequences were subjected for analysis and output results from the InterProScan were shown on Figure 3.3. Superfamily recognized Globin-like superfamily in nine proteins except SGRA_2168 and the globin as a structural domain of each protein got a hit on Pfam. The C-terminal region of each sequence was recognized as anti-sigma factor antagonist SpoIIaa superfamily and the C-terminal region was identified as the STAS domain by Pfam and PROSITE databases.

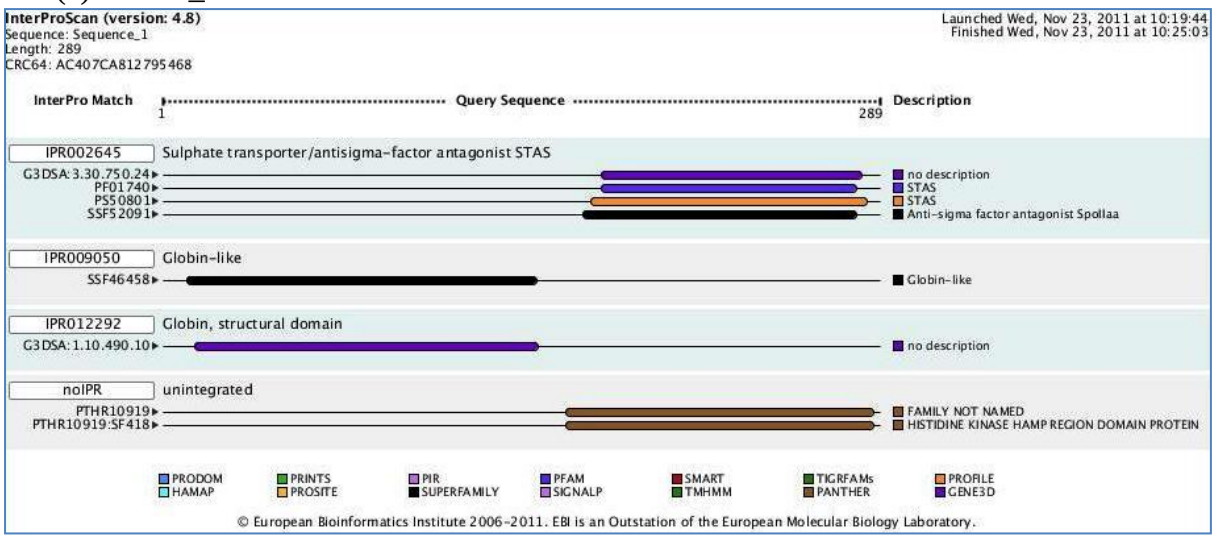
(a) SGRA_0571



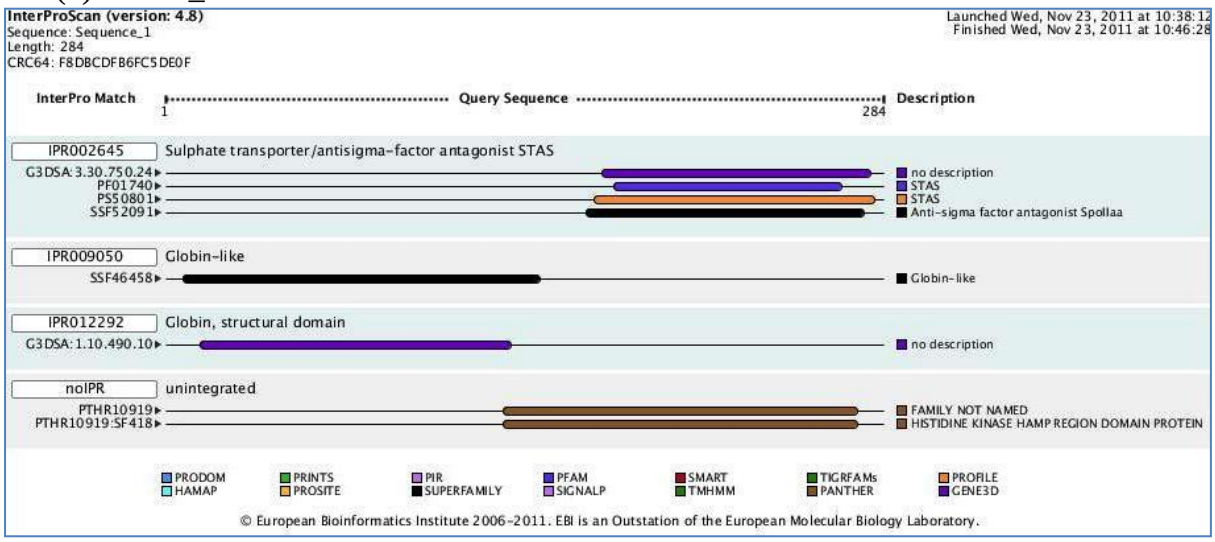
(b) SGRA_1293



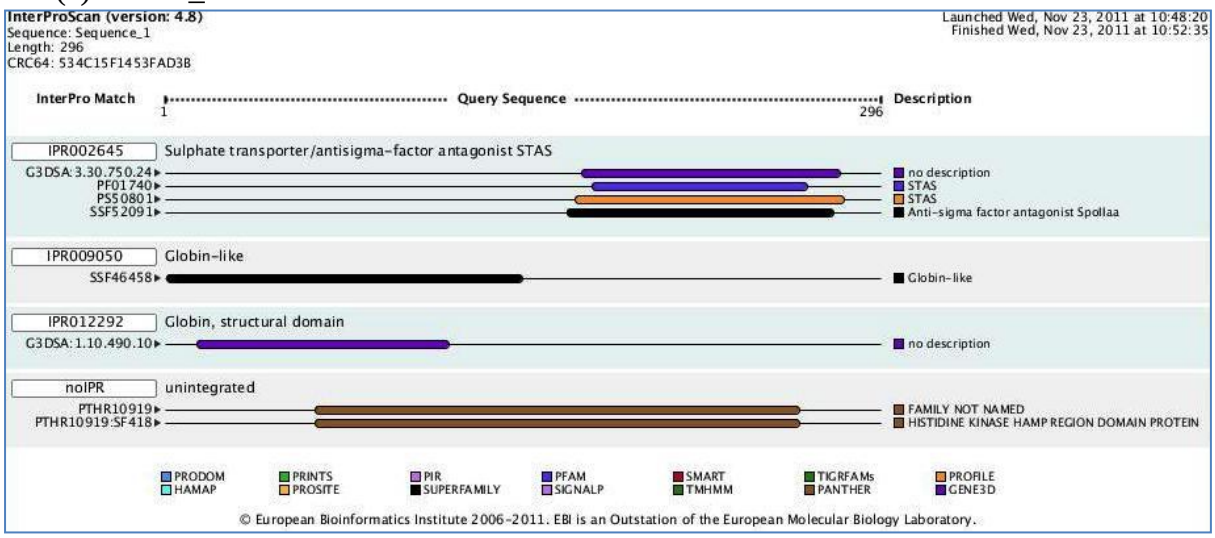
(c) SGRA_2160



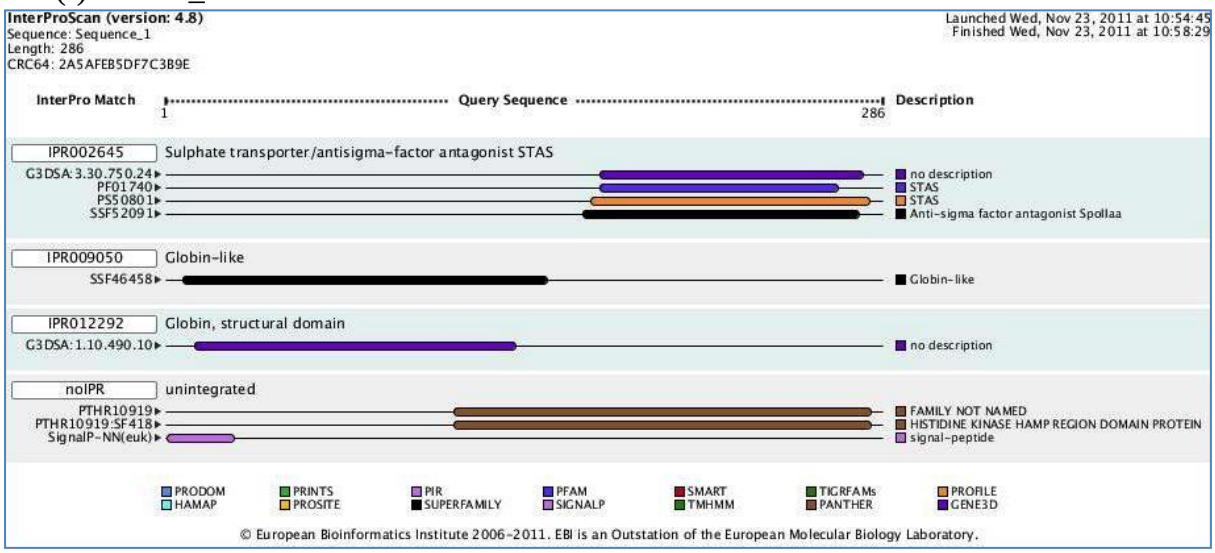
(d) SGRA_2161



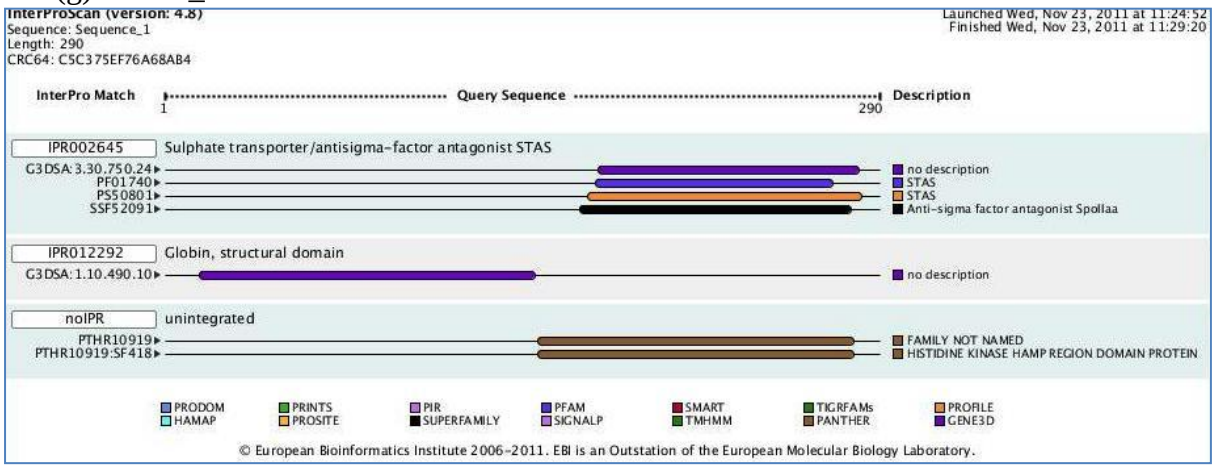
(e) SGRA_2162



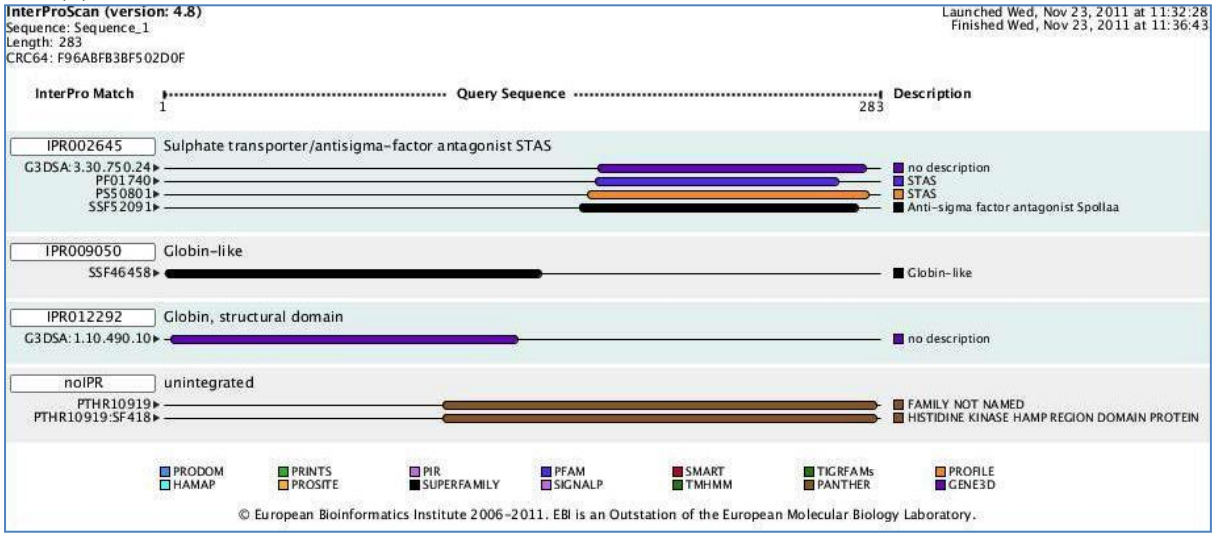
(f) SGRA_2167



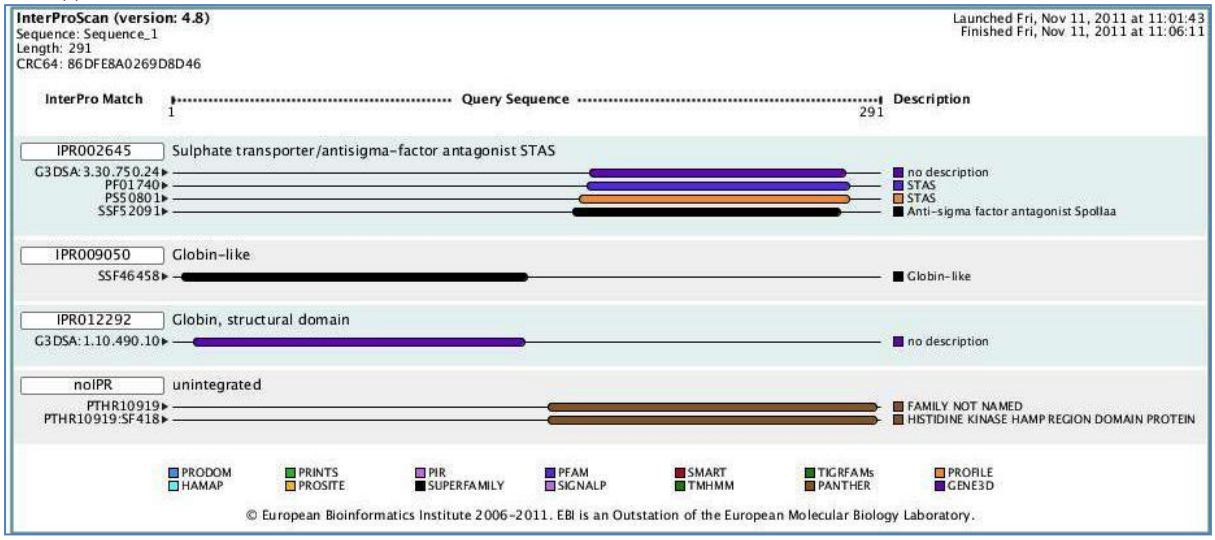
(g) SGRA_2168



(h) SGRA_2169



(i) SGRA_3210



(j) SGRA_3852

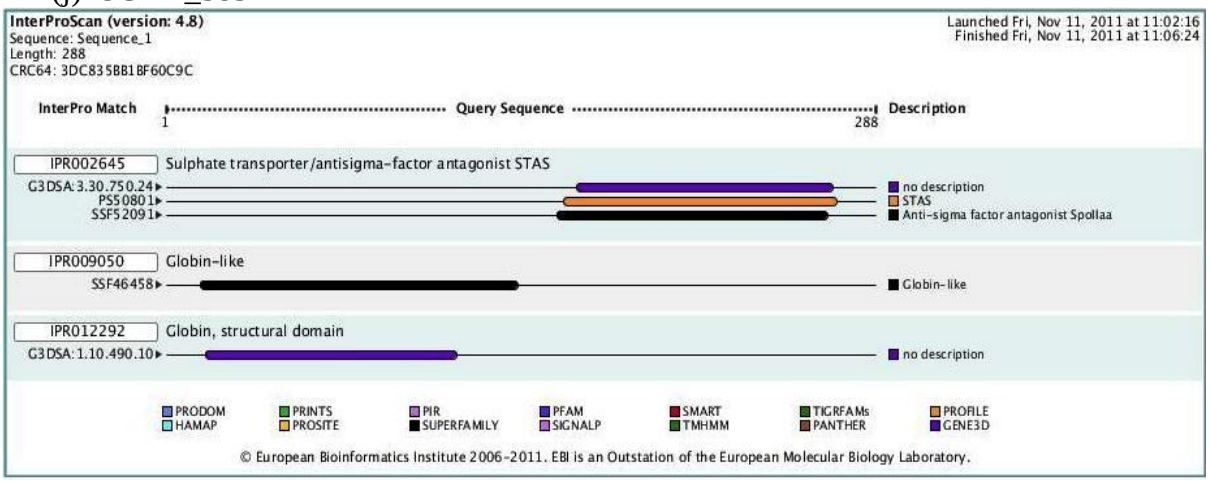


Figure 3.3. InterproScan results of 10 GCSs. (a) SGRA_0571 (b) SGRA_1293 (c) SGRA_2160 (d) SGRA_2161 (e) SGRA_ 2162 (f) SGRA_2167 (g) SGRA_2168 (h) SGRA_2169 (i) SGRA_3210 (j) SGRA_3852.

Table 3.1. Prediction of Globin and STAS domains

Locus Tag	Amino acid length (aa)	Globin		STAS		In this study	
		NCBI (aa)	SBASE (aa)	NCBI (aa)	SBASE (aa)	Globin (aa)	STAS (aa)
SGRA_0571	287	13-144	7-147	165-273	166-272	13-147	165-273
SGRA_1293	290	34-109	40-113	181-271	171-274	34-113	171-274
SGRA_2160	288	18-111	9-151	169-285	173-279	9-151	169-285
SGRA_2161	284	16-109	17-148	169-277	170-276	16-148	169-277
SGRA_2162	296	–	14-148	180-271	170-276	14-148	170-276
SGRA_2167	286	13-133	14-108	169-277	170-276	13-133	169-277
SGRA_2168	290	–	16-150	171-279	172-277	16-150	171-279
SGRA_2169	283	11-141	15-123	167-275	168-273	11-141	167-275
SGRA_3210	291	13-146	14-146	167-275	168-274	13-146	167-275
SGRA_3852	288	17-119	7-147	165-273	166-272	7-147	165-273

CDD from NCBI and SBASE identified boundaries of globin and STAS domains within an amino acid sequence (Table 3.1). Each tool provided slightly different defined domains as little as one amino acid difference to ten amino acids. SGRA_2162 and SGRA_2168 were only predicted the C-terminal STAS domains by analysis from NCBI in contrast to SBASE recognized the globin domains in both genes. In this study, the globin and STAS domains of ten GCS genes were determined by choosing the widest range of each domain that predicted by two different tools.

Physical locations of GCS protein and its paralogs were mapped on Figure 3.4. SGRA_0571, SGRA_1293, SGRA_3210 and SGRA_3852 are all spread in *S. grandis* genome. In contrast, SGRA_2160 to SGRA_2162 and SGRA_2167 to SGRA_2169 are present as clusters in proximal distance. The gene, SGRA_3210 is indicated as *rsbR* because its function is predicted as anti-anti sigma factor antagonist and its orientation of neighbor genes (*rsbS, rsbT, rsbV, rsbU*) are similar to a *sigma B* operon in *B. subtilis* (Hecker *et al*., 2007).

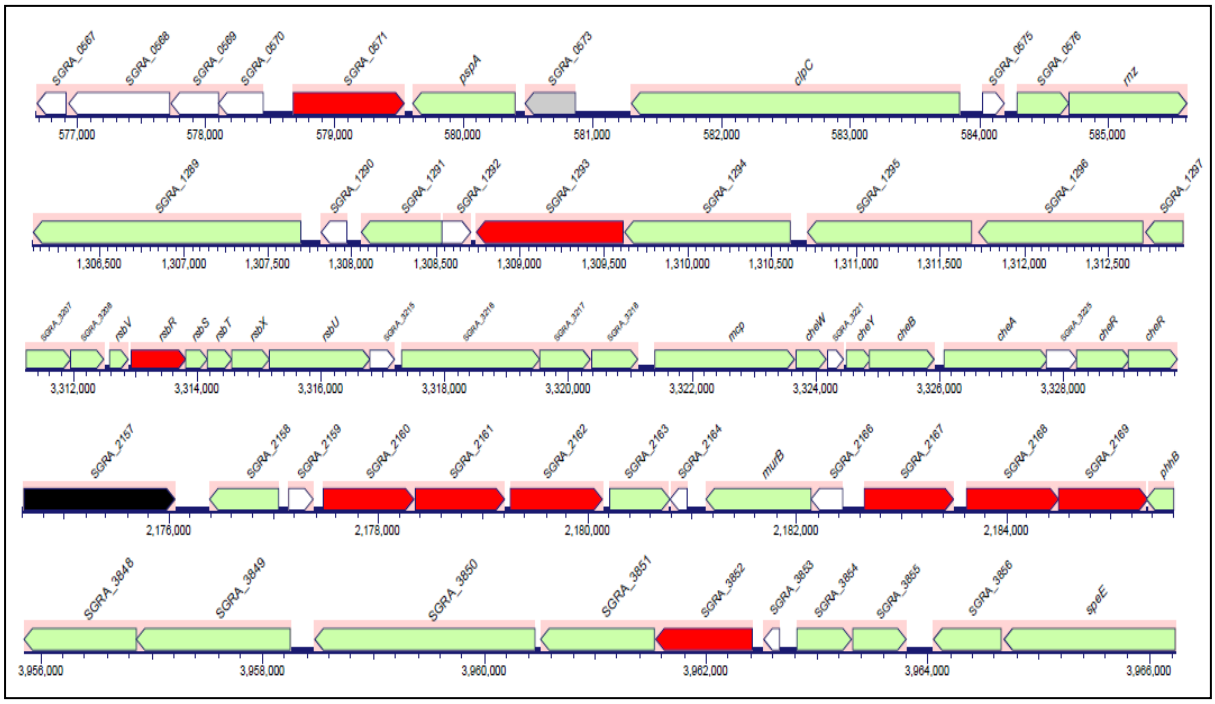


Figure 3.4. Physical locations of SGRA_3210 and its paralogs within *S. grandis* genome. Pentagons represent genes predicted by Pathway Tools. The red pentagons indicate *rsbR* (SGRA_3210) genes with N-terminal globin and C-temrinal STAS domains. The green pentagons indicate genes with predicted functions. Genes that encoded hypothetical proteins were displayed in grey (conserved hypothetical proteins) and white pentagons. The black pentagon indicates transposase.

3.1.2. Multiple alignments

The N-terminal globin domain and the C-terminal STAS domain that were defined by bioinformatics analysis were subjected to multiple alignments against known globin domains from three different GCSs and the C-terminal STAS domain region of RsbR and its paralog proteins from *B. subtilis*. Each alignment was generated by the Clustal W program from Lasergene 9 coresuite (DNASTAR).

According to Saito *et al*. (2008), two regions within a motif are crucial to identify the sensor globin domain in GCSs. These are the B10 tyrosine residue and the proximal histidine residue. The B9 region tends to be occupied by the phenylalanine but it can be placed by other amino acids. The globin regions of ten GCSs were not well lined up with

three known globin domains. The only region that was aligned perfectly was the position B10. Another vital region of the sensor globin, the proximal histidine was lined up only present in the known globin domains and three globins from SGRA_0571 (RsbR2), SGRA_3210 (RsbR1), and SGRA_3852 (RsbR3) (Figure 3.5).

The alignment of the C-terminal region of RsbR and its paralogs from *B. subtilis* with ten GCSs showed highly identical regions. The STAS domain in RsbR and its paralogs (YkoB, YojH and YqhA) are known to have two phosphorylation sites in threonine residues by serine/threonine kinase, RsbT (Akbar *et al*., 2001). The position 171 and 205 threonines of RsbR are exactly lined up with threonine residues of its paralogs. Nine out of ten GCSs from *S. grandis* were aligned with these two regions but mostly serine at the position 171 instead of threonine. The position T205 was aligned mostly with threonine but SGRA_1293 substituted to serine instead of threonine among aligned nine GCSs. YtvA and SGRA_3852 (RsbR3) did not have either serine or threonine residue in these regions. YtvA has a GTP binding property and its motif, DLSG is the site where the GTP binds like other proteins that contain a NTP binding motif, DXXG (Buttani *et al*., 2006; Kjeldgaard *et al*, 1996). Aspartate (D) and glycine (G) residues in the region where the GTP binding motif is in YtvA were all aligned in every proteins and indicated as the red empty box in Figure 3.6.

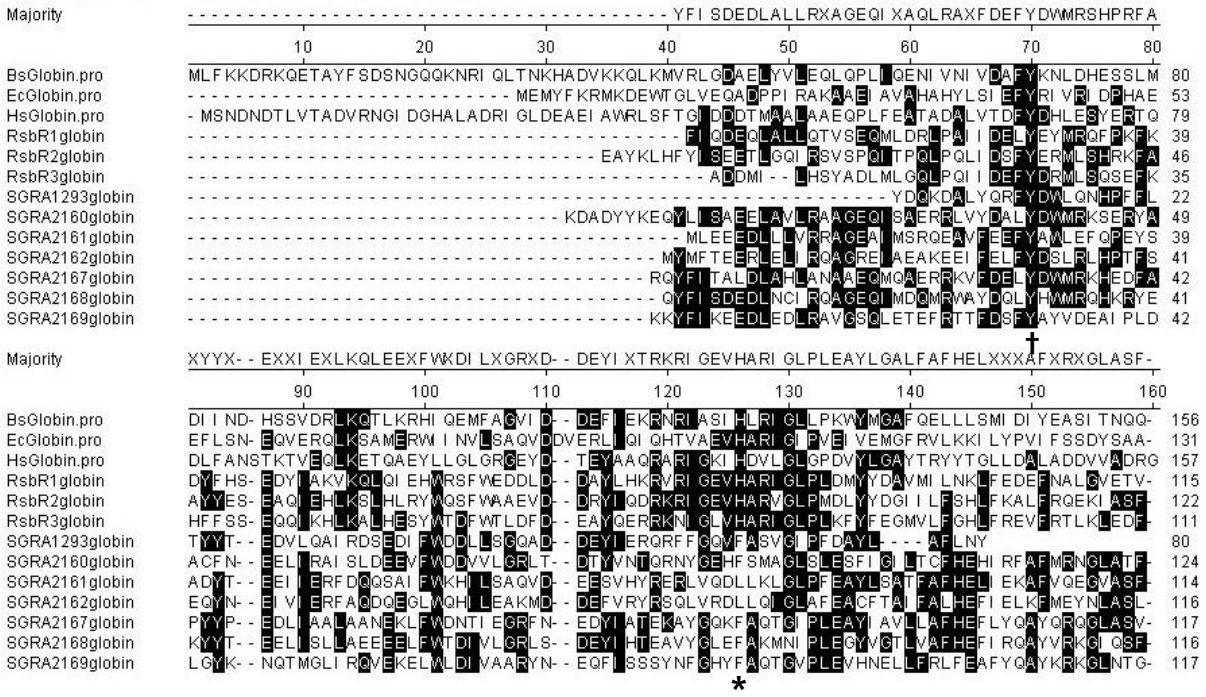


Figure 3.5. Multiple alignment of predicted globin domains of ten GCS protein sequences and protein sequences of three known globin domains. The distal B10 tyrosine residue is indicated with a dagger (†) and the proximal histidine residue with an asterisk (*). Black shading indicates 100% identity.

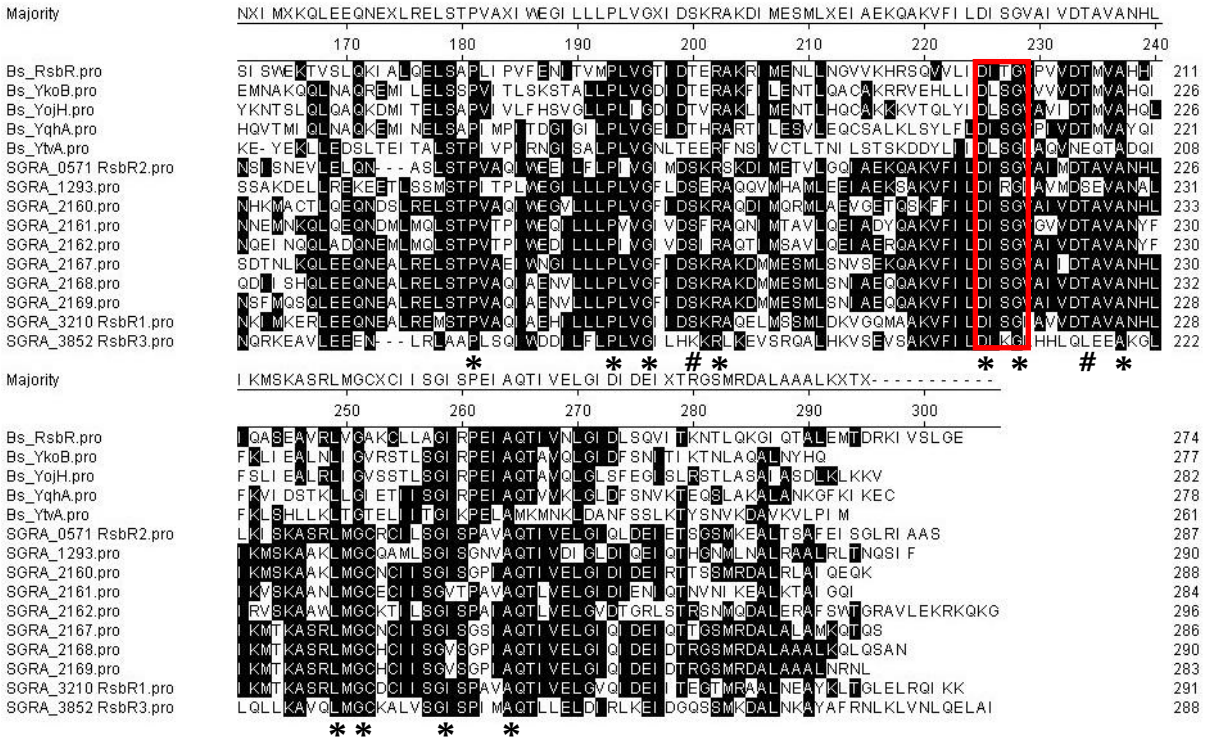


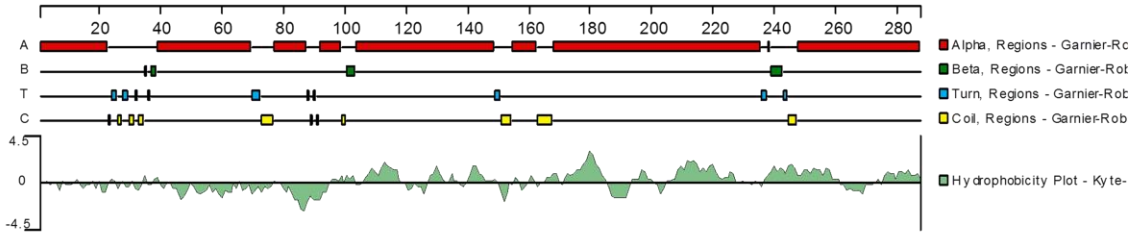
Figure 3.6. Multiple alignment of predicted STAS domains of ten GCSs and STAS domains of RsbR and its four paralogs from *Bacillus subtilis*. An asterisk (*) indicates a conserved residue in all queries and a pound sign (#) indicates Thr/Ser residues at the positions aligned to phosphorylation sites (T171 and T205) of RsbR in *B. subtilis*. A possible NTP binding motif (DXXG) is indicated with a red box. Black shading indicates 100% identity.

3.1.3. Structure prediction

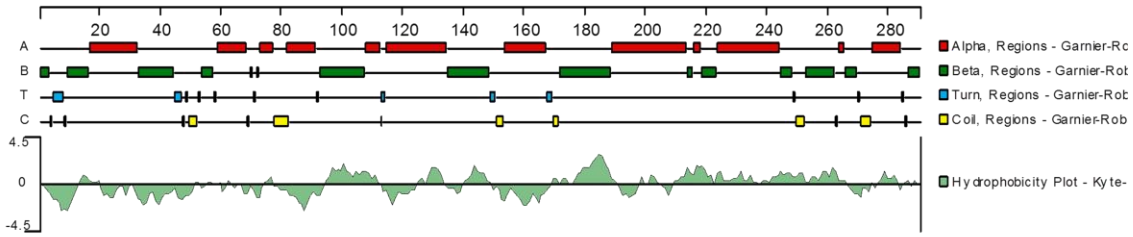
3.1.3.1. Secondary structure and hydrophobicity

A prediction of a protein secondary structure is helpful to determine a potential function of the protein. The Garnier-Robson method was selected and predicted structures were shown in figure 3.7. They were mainly composed of alpha helices. Very short beta sheets or turns and coiled regions were predicted in between helices. More hydrophobic amino acid residues were present toward the C-terminal regions and these were indicated as positive peaks on hydrophobicity plots.

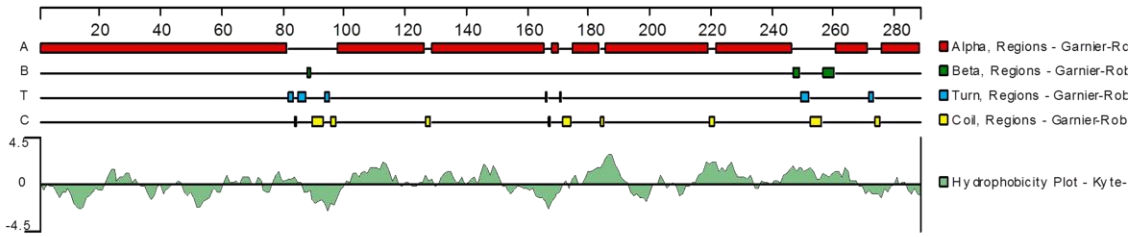
(a) SGRA_0571



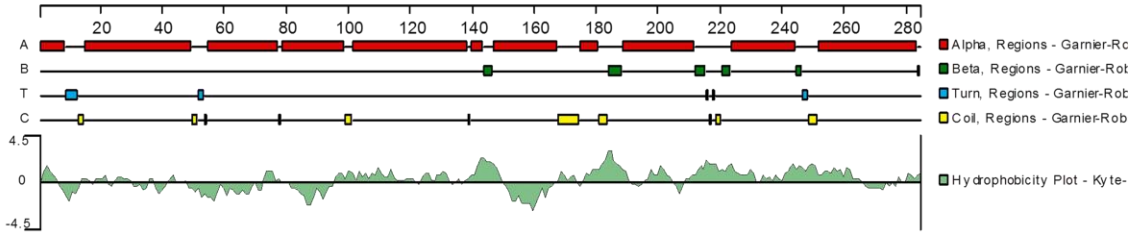
(b) SGRA_1293



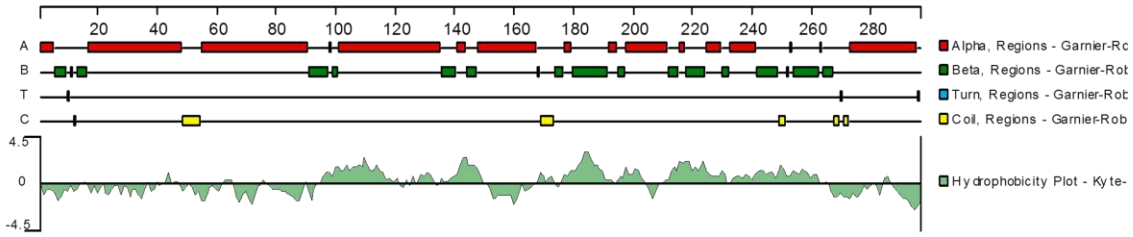
(c) SGRA_2160



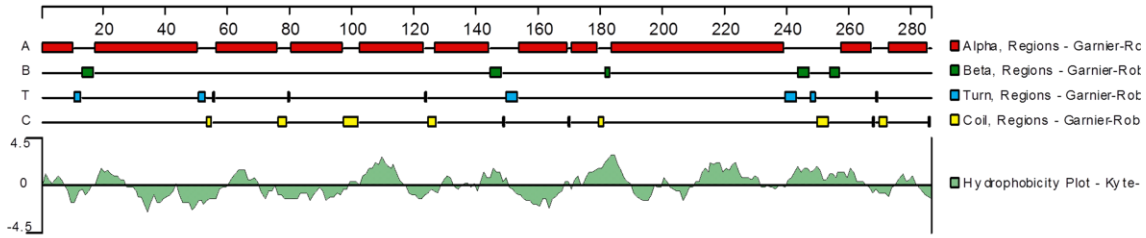
(d) SGRA_2161



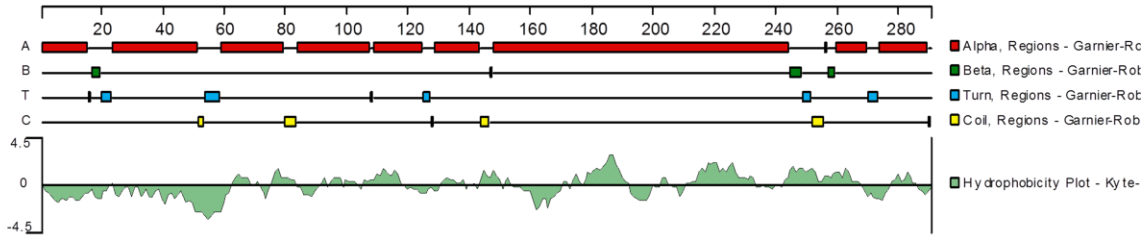
(e) SGRA_2162



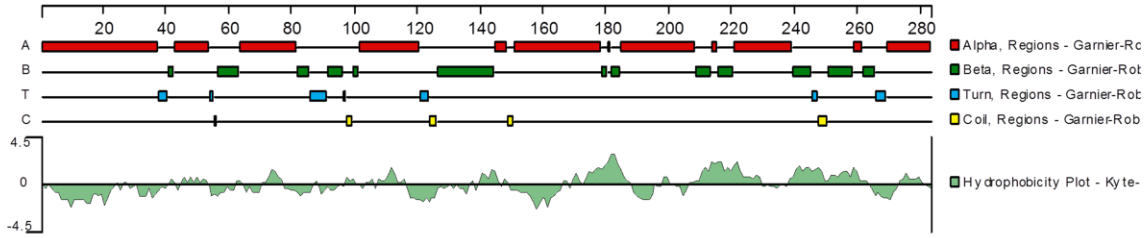
(f) SGRA_2167



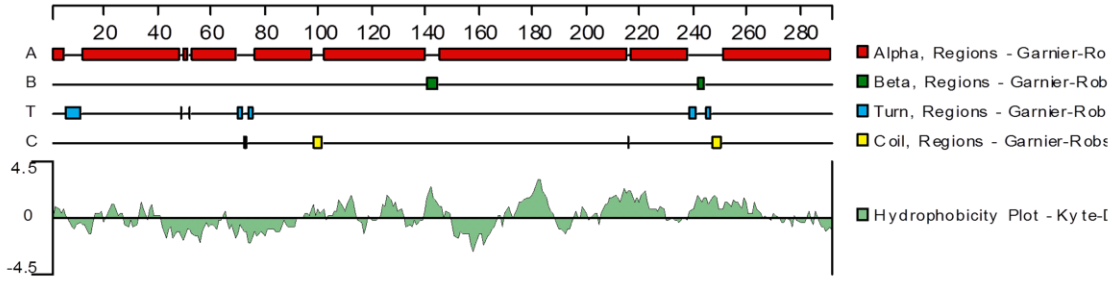
(g) SGRA_2168



(h) SGRA_2169



(i) SGRA_3210



(j) SGRA_3852

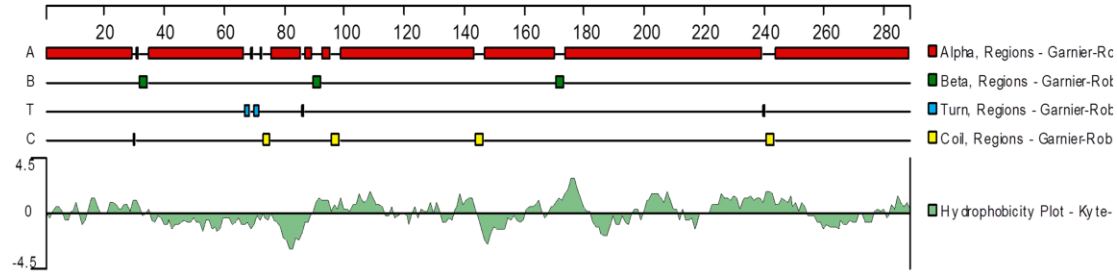
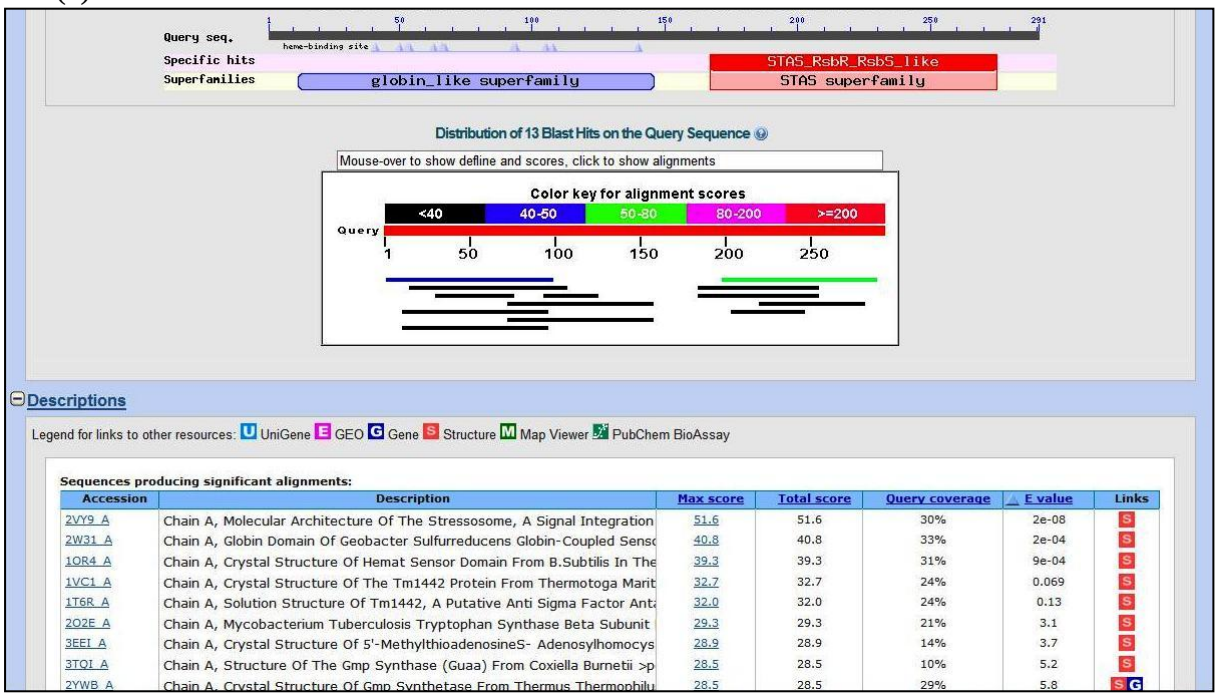


Figure 3.7. Secondary structure predictions of 10 GCSs. (a) SGRA_0571 (b) SGRA_1293 (c) SGRA_2160 (d) SGRA_2161 (e) SGRA_ 2162 (f) SGRA_2167 (g) SGRA_2168 (h) SGRA_2169 (i) SGRA_3210 (j) SGRA_3852.

3.1.3.2. 3D structure alignments of RsbR1 (SGRA_3210), RsbR2 (SGRA_0571) and RsbR3 (SGRA_3852)

The biological function and the evolutionary relationship of proteins can be inferred by 3D structures of proteins. It is also able to surmise the sequence-structure relationships, interactions, active sites and so on (NCBI, n.d.). Entrez's 3D Structure Database, the Molecular Modelling Database (MMDB) provides known 3D structures similar to a given query sequence. Depending on how close they are, homologous relationships may lead us to speculate functional and structural roles of the sequence (Wang *et al*., 2006). Three GCS proteins: RsbR1, RsbR2 and RsbR3 were subjected to find any hit of known 3D structure because only these proteins were aligned to proximal histidine in three known globins (Figure 3.5). One way to find similar 3D structures of a protein sequence was to do Protein BLAST against Protein Data Bank (PDB) as the database. The initial BLAST results of the proteins got hits on the N-terminal globin superfamily and the C-terminal STAS superfamily. The top scored alignments of each sequence were RsbS and the globin domain of *Geobacter sulfurreducens*'s globin-coupled sensor. Expected values (E-values) of these alignments were less than 0.005 which was considered to be significant except the globin alignment of RsbR3 (E-value=0.002) (Figure 3.8a, 3.9a, 3.10a). Alignments of RsbR1, RsbR2 and RsbR3 against proteins with the lowest E-values were visualized in 3D structural view. The aligned regions of proteins were indicated as blue color and identical amino acids were colored in red (Figure 3.8b, 3.9b, 3.10b). Protein sequence identities of the globin domains in RsbR1, RsbR2 and RsbR3 were less than 30%. However, the proximal residues which were viewed on Figure 3.5 were aligned the proximal histidine residues that actually holding the heme molecule on Figure 3.8b, 3.9b and 3.10b. The proximal histidine residue of the globin domain in *G. sulfurreducens* was visualized in yellow.

(a) RsbR1



(b) RsbR1

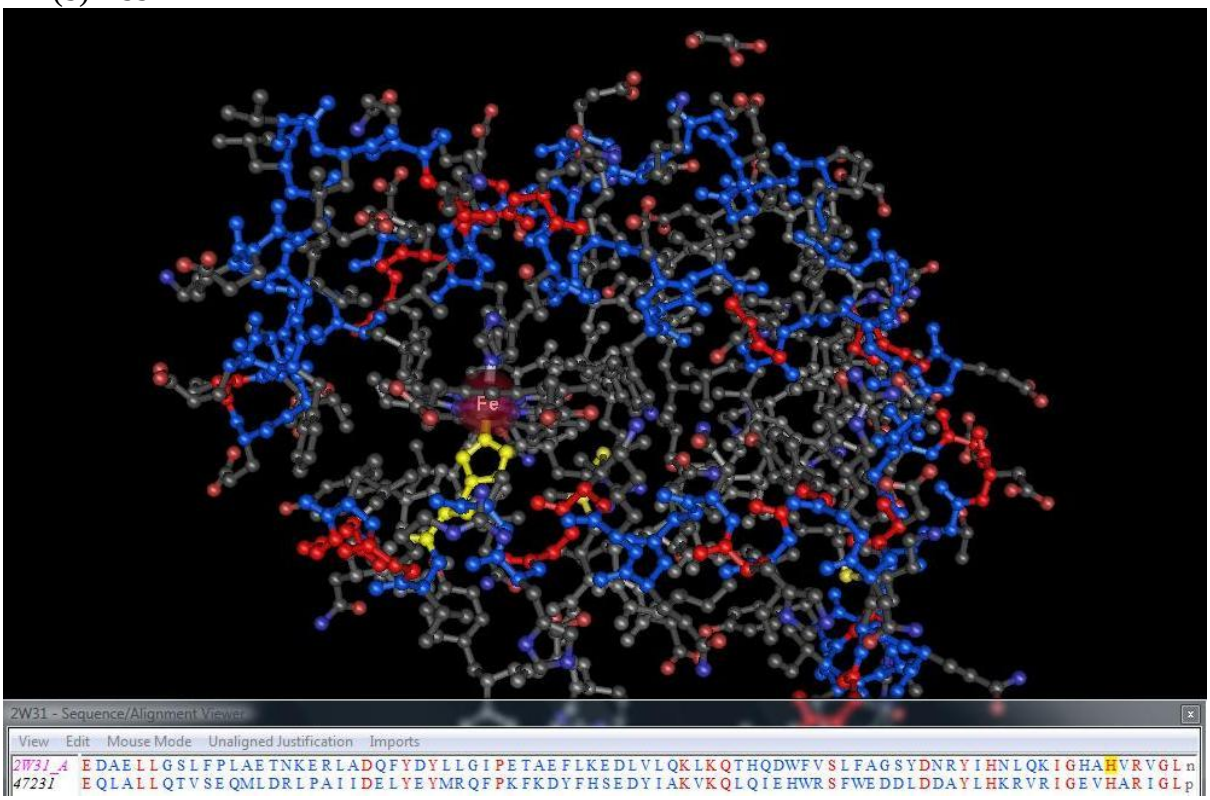
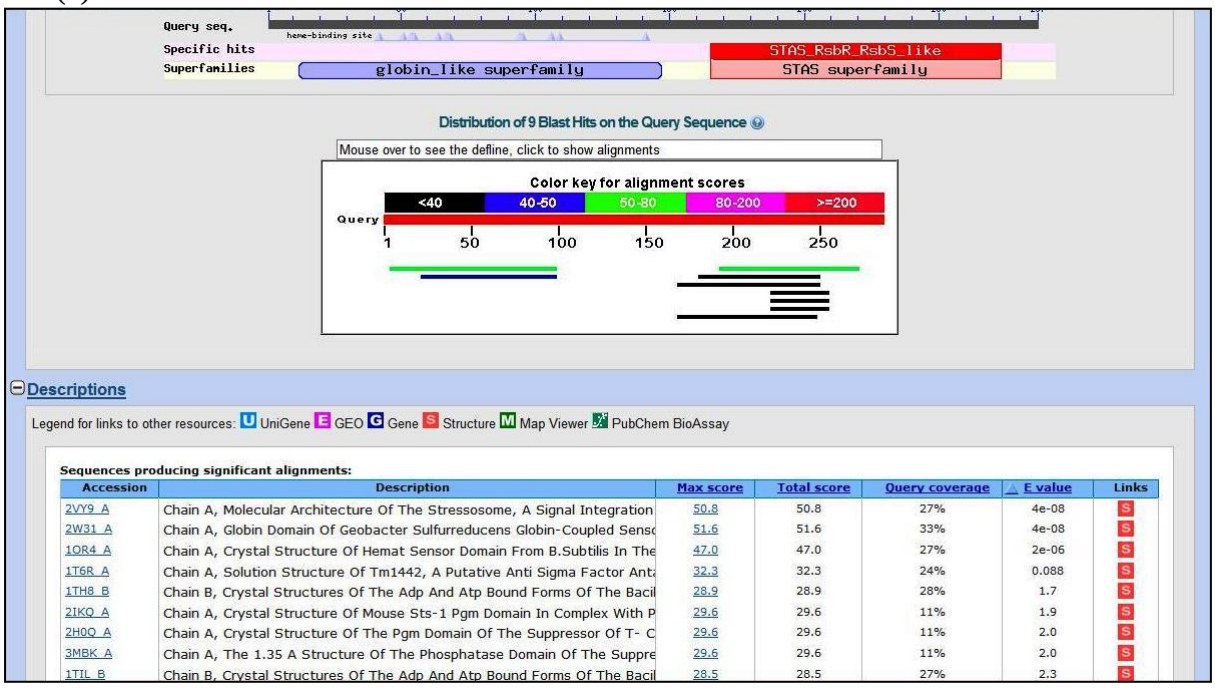


Figure 3.8. Screen shots of the 3D structure protein blast and alignment of the globin domain of RsbR1 and the globin domain of *Geobacter sulfurreducens*'s GCS. (a) Initial protein blast result (b) 3D view of RsbR1 and the globin domain of *G. sulfurreducens* structure alignment.

(a) RsbR2



(b) RsbR2

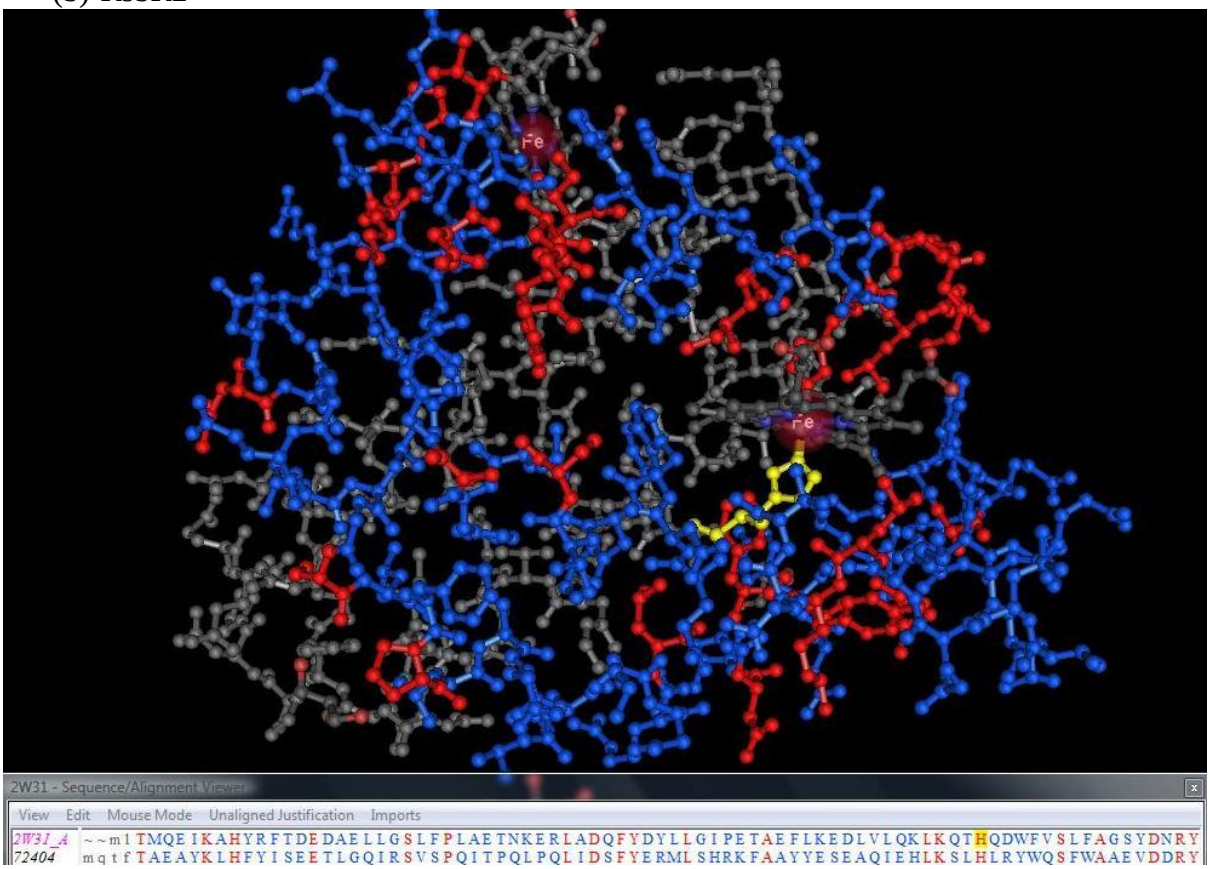
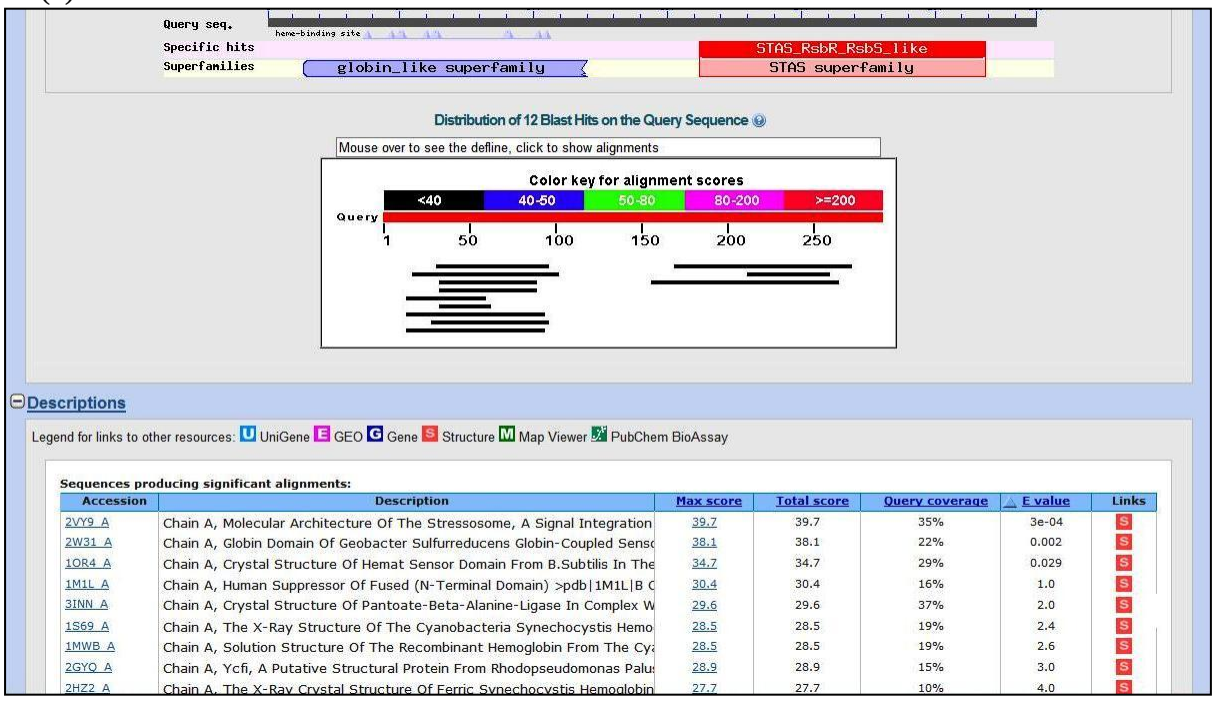


Figure 3.9. Screen shots of the 3D structure protein blast and alignment of the globin domain of RsbR2 and the globin domain of *Geobacter sulfurreducens*'s GCS. (a) Initial protein blast result (b) 3D view of RsbR2 and the globin domain of *G. sulfurreducens* structure alignment.

(a) RsbR3



(b) RsbR3

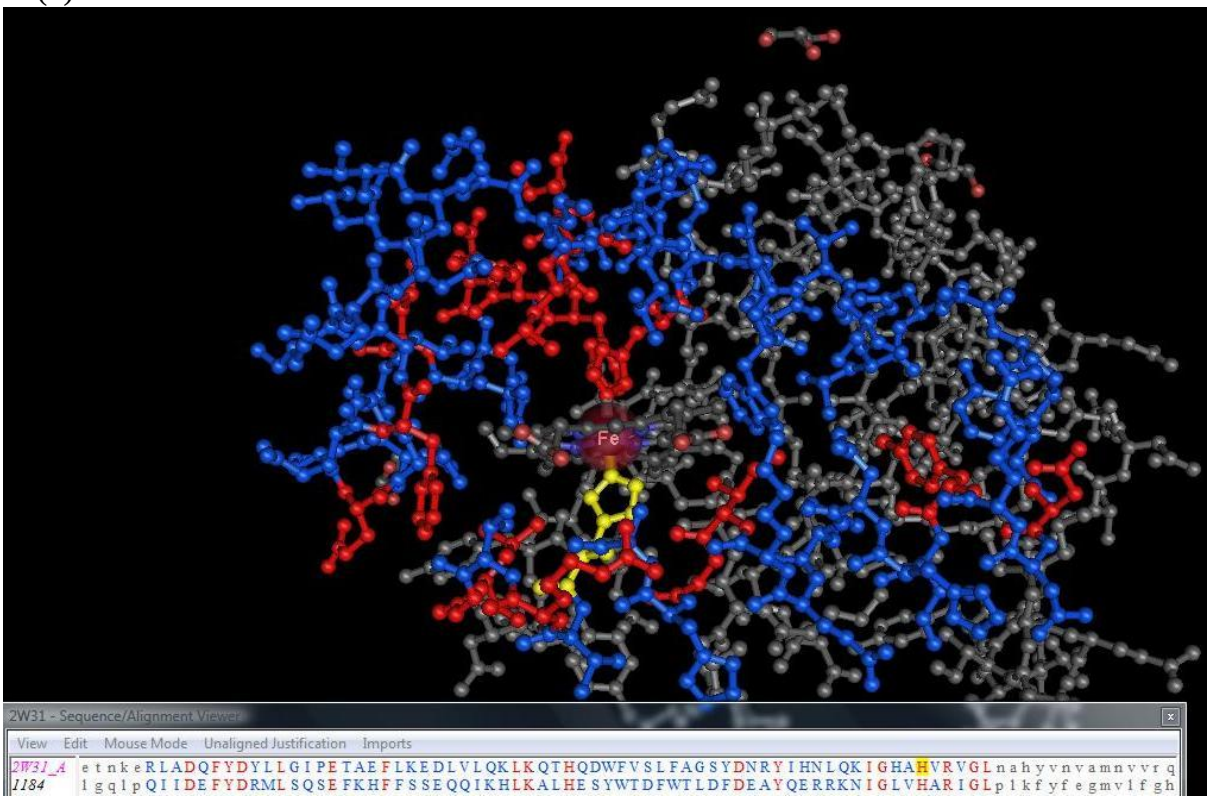


Figure 3.10. Screen shots of the 3D structure protein blast and alignment of the globin domain of RsbR3 and the globin domain of *Geobacter sulfurreducens*'s GCS. (a) Initial protein blast result (b) 3D view of RsbR3 and the globin domain of *G. sulfurreducens* structure alignment.

3.1.4. Prediction of physical characteristics of 10 GCSs

Molecular weights and pI values of the proteins were automatically calculated by SeqMan program in Lasergene 9 coresuit software (DNASTAR). Their MWs were ranges of 32.08 to 33.5 kDa. pI values of each protein were between 4.47 as the lowest to 5.9 as the highest. Since the values were below 7.0, all ten proteins were acidic. Based on amino acid compositions, protein solubility can be predicted by statistical analysis (Wilkinson and Harrison, 1991) and this can be done through the web-based analysis. Predicted protein solubility values were indicated in percentage. Nine out of ten proteins had above 70% chance of being soluble and SGRA_2169 had the lowest (65.7%) (Table 3.2).

Table 3.2. Predicted molecular weights, pIs and solubilities of ten GCS proteins

Locus Tag	MW (kDa)	pI	Solubility (%)
SGRA_0571	32.44	5.23	70.7
SGRA_1293	33.11	4.47	91.4
SGRA_2160	32.68	4.97	84.7
SGRA_2161	32.22	4.52	91.5
SGRA_2162	33.84	4.94	81.2
SGRA_2167	32.35	5.17	74.0
SGRA_2168	33.31	4.93	87.5
SGRA_2169	32.08	5.36	65.7
SGRA_3210	33.30	5.18	84.2
SGRA_3852	33.50	5.90	77.3

3.2. *In vitro* characterization of 10 RsbRs

3.2.1. Cloning and expression

Each gene was amplified by PCR reactions. All ten PCR products were visualized on a 1% agarose gel (Figure 3.11). The sizes of PCR products were around 900 bp but they were slightly different depending on the size of gene and the length of tags. PCR products were purified and they were either cloned into a sub-cloning vector (pJET cloning, Fermentas or TOPO cloning, Invitrogen) or digested by suitable restriction enzymes. Sub-cloned genes were confirmed either by mini-digestions or colony PCRs and underwent restriction enzyme digestions. Colony PCR products showed slightly higher sizes compared to the original PCR products (Figure 3.12) because primers used in PCRs were designed for an upstream and a downstream of multiple cloning sites in a sub-cloning vector. The restriction enzyme digested gene fragments (Figure 3.13) were excised and purified after visualization on the agarose gel. Individual *rsbR* genes were then cloned into proper expression vectors (Table 3.3). The cloned constructs were transformed into E. coli competent cells (Mach1-T1^{R} or DH5α).

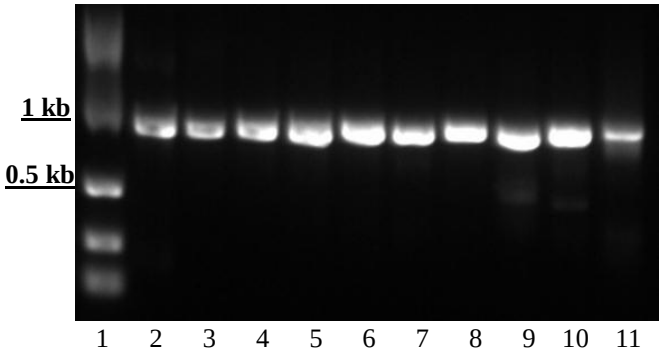


Figure 3.11. PCR products of ten *rsbR* genes. Each amplified product was checked on 1% TAE agarose gel. Lanes: 1, Easy ladder I (Bioline); 2, SGRA_0571; 3, SGRA_1293; 4, SGRA_2160; 5, SGRA_2161; 6, SGRA_2162; 7, SGRA_2167; 8, SGRA_2168; 9, SGRA_2169; 10, SGRA_3210; 11, SGRA_3852.

Table 3.3. Constructions of 10 *rsbR*s in different expression vectors

Locus Tag	Expression Vector	Location of His-tags
SGRA_0571	pET3a	N-terminal
SGRA_1293	pET14b	N-terminal
SGRA_2160	pET14b	N-terminal
SGRA_2161	pET3a	C-terminal
SGRA_2162	pET14b	N-terminal
SGRA_2167	pET14b	N-terminal
SGRA_2168	pET14b	N-terminal
SGRA_2169	pET14b	N-terminal
SGRA_3210	pET14b	N-terminal
SGRA_3852	pMAL-c2x	C-terminal

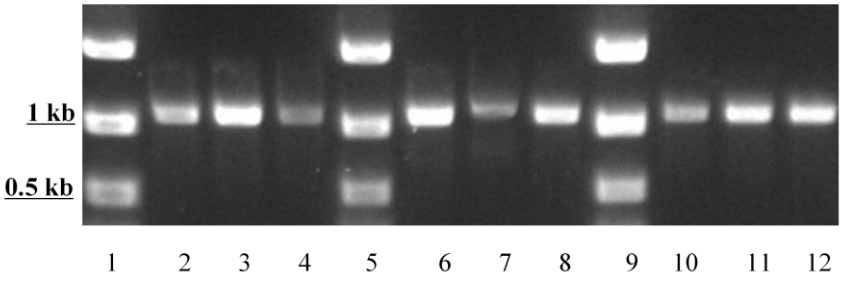


Figure 3.12. Colony PCR of SGRA_1293/pJET, SGRA_2160/pJET, and SGRA_2168/pJET. Three different colonies from each construct were run on colony PCRs and visualized on 1% TAE agarose gel. Lanes: 1;5;9, Easyladder I (Bioline); 2-5, colony 1-3 of SGRA_1293; 6-8, colony 1-3 of SGRA_2160; 10-12, colony 1-3 of SGRA_2168.

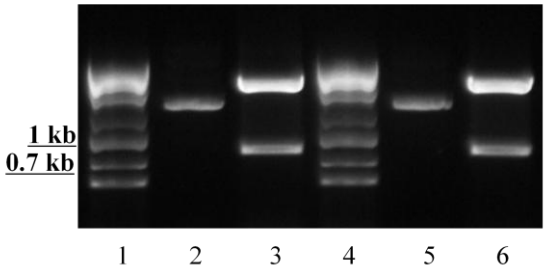


Figure 3.13. Restriction enzyme digestion of SGRA_2161/pJET and SGRA_2169/pJET with *Nde*I and *Bam*HI. Digested products were visualized on 0.7% TAE agarose gel. Lanes: 1;4, Hyperladder III (Bioline); 2, undigest SGRA_2161/pJET; 3, digested SGRA_2161/pJET; 5, undigest SGRA_2169/pJET; 6, digested SGRA_2169/pJET.

In order to confirm presence of an insert in an expression vector, mini-digestions or colony PCRs were done. The confirmed constructs were sequenced to check whether sequences were correct or not. Right constructs were then transformed in Rosetta 2(DE3)pLysS cells and expressed to find optimal conditions for expressing RsbR proteins. Time course analyses were done with different conditions such as optical density, temperature and IPTG) concentration. The heme precursor was also added to the culture at the time of induction for enhancing heme synthesis.

Firstly, SGRA_3210 (RsbR1) was tested in following parameters: (1) temperature (30℃ vs. 37℃), (2) Optical density (OD_{600}) (0.6 vs. 1.0), (3) IPTG concentration (0.05 mM, 0.2 mM, 0.6 mM). Cultures were collected in timely bases for 24 hour and extracted soluble proteins were analyzed on SDS polyacrylamide gels (Figure 3.14-3.17). Induced cultures at 37℃ were not shown any induced soluble protein even though different IPTG concentrations and OD_{600}s were tried (Figure 3.16 and 3.17). In contrast, induced SGRA_3210 proteins at 30℃ were visualized on the gels (Figure 3.14 and 3.15). Induction at 30℃ with OD_{600} at 1.0 and addition of 0.2 mM IPTG yielded more proteins which were visualized as thick bands at ~35 kDa region. At this condition, the culture for 12 hour induction had more soluble protein than others.

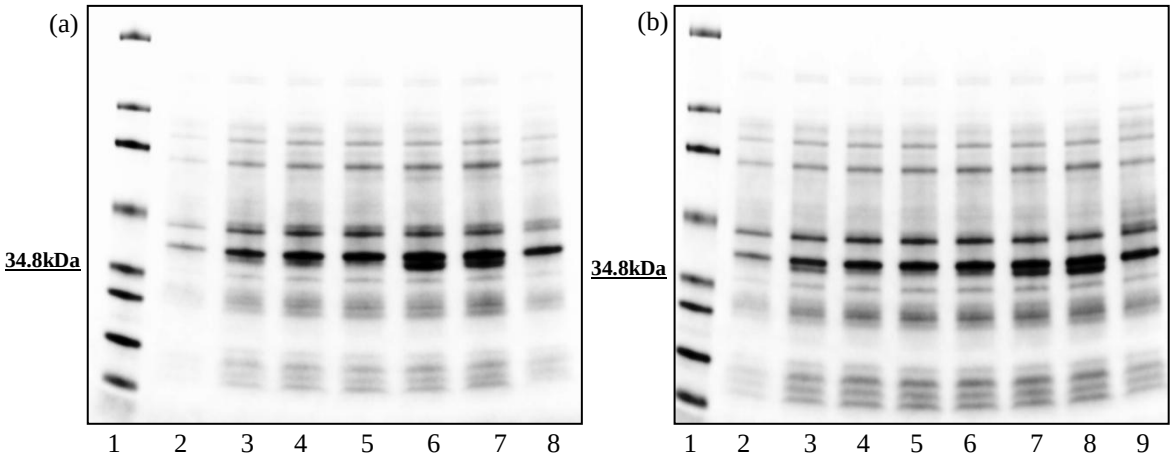


Figure 3.14. Time course analysis of induced SGRA_3210 for 24 hours at 30°C and at 0.6 OD_{600}. IPTG was added when OD_{600} reached to 0.6. (a) 0.05 mM IPTG (b) 0.2 mM IPTG (Lanes: 1, molecular weight marker; 2, uninduced; 3, 1 hour after induction; 4, 2 hours after induction; 5, 3 hours after induction; 6, 4 hours after induction; 7, 6 hours after induction; 8, 12 hours after induction; 8, 24 hours after induction. Soluble proteins were extracted with BugBuster protein extraction reagent (Novagen) and examined on 4-20% SDS polyacrylamide gels.

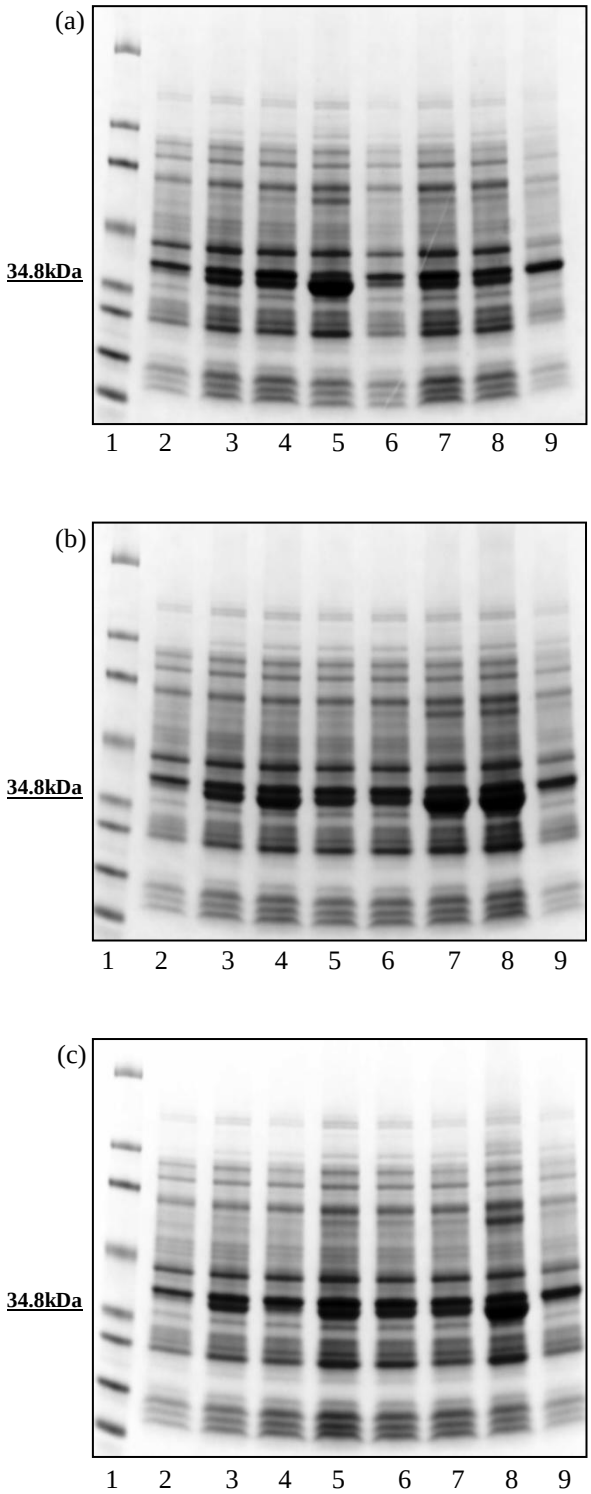


Figure 3.15. Time course analysis of induced SGRA_3210 for 24 hours at 30°C and at 1.0 OD_{600}. IPTG was added when OD_{600} reached to 1.0. (a) 0.05 mM IPTG (b) 0.2 mM IPTG (c) 0.6 mM IPTG. Lanes: 1, molecular weight marker; 2, uninduced; 3, 1 hour after induction; 4, 2 hours after induction; 5, 3 hours after induction; 6, 4 hours after induction; 7, 6 hours after induction; 8, 12 hours after induction; 8, 24 hours after induction. Soluble proteins were extracted with BugBuster protein extraction reagent (Novagen) and examined on 4-20% SDS polyacrylamide gels.

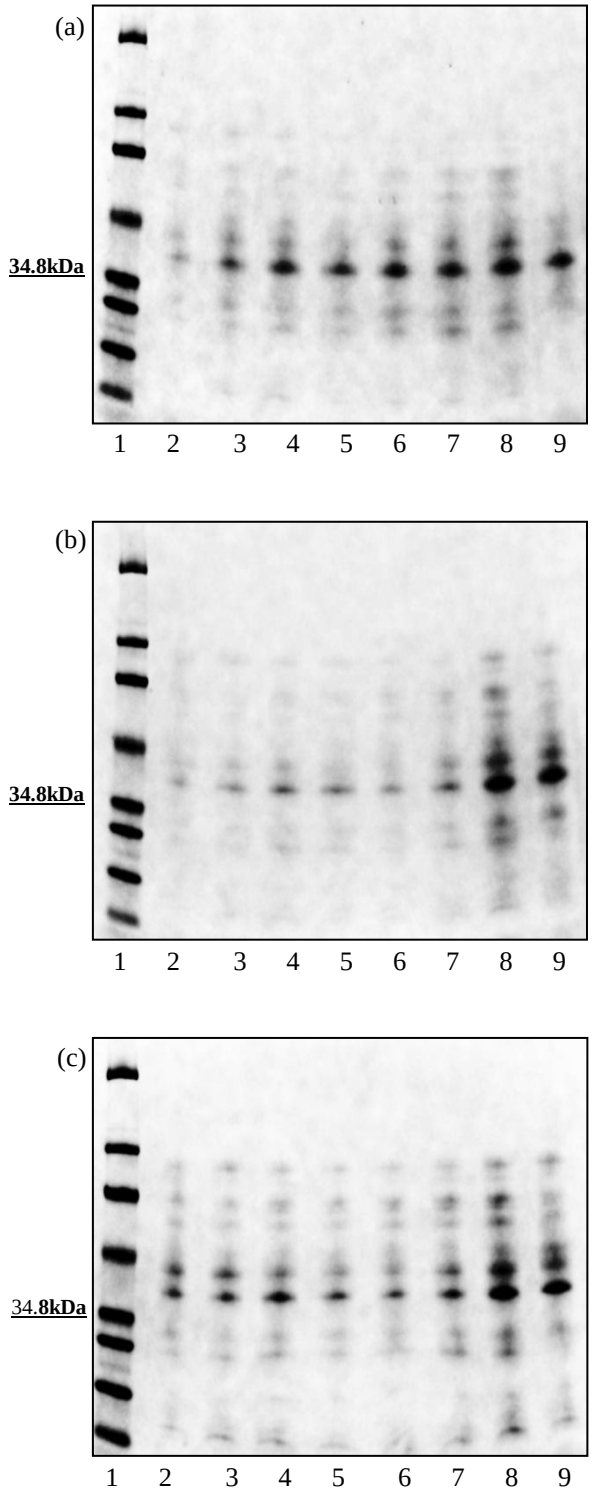


Figure 3.16. Time course analysis of induced SGRA_3210 for 24 hours at 37°C and at 0.6 OD_{600} . IPTG was added when OD_{600} reached to 0.6. (a) 0.05 mM IPTG (b) 0.2 mM IPTG (c) 0.6 mM IPTG. Lanes: 1, molecular weight marker; 2, uninduced; 3, 1 hour after induction; 4, 2 hours after induction; 5, 3 hours after induction; 6, 4 hours after induction; 7, 6 hours after induction; 8, 12 hours after induction; 8, 24 hours after induction. Soluble proteins were extracted with BugBuster protein extraction reagent (Novagen) and examined on 4-20% SDS polyacrylamide gels.

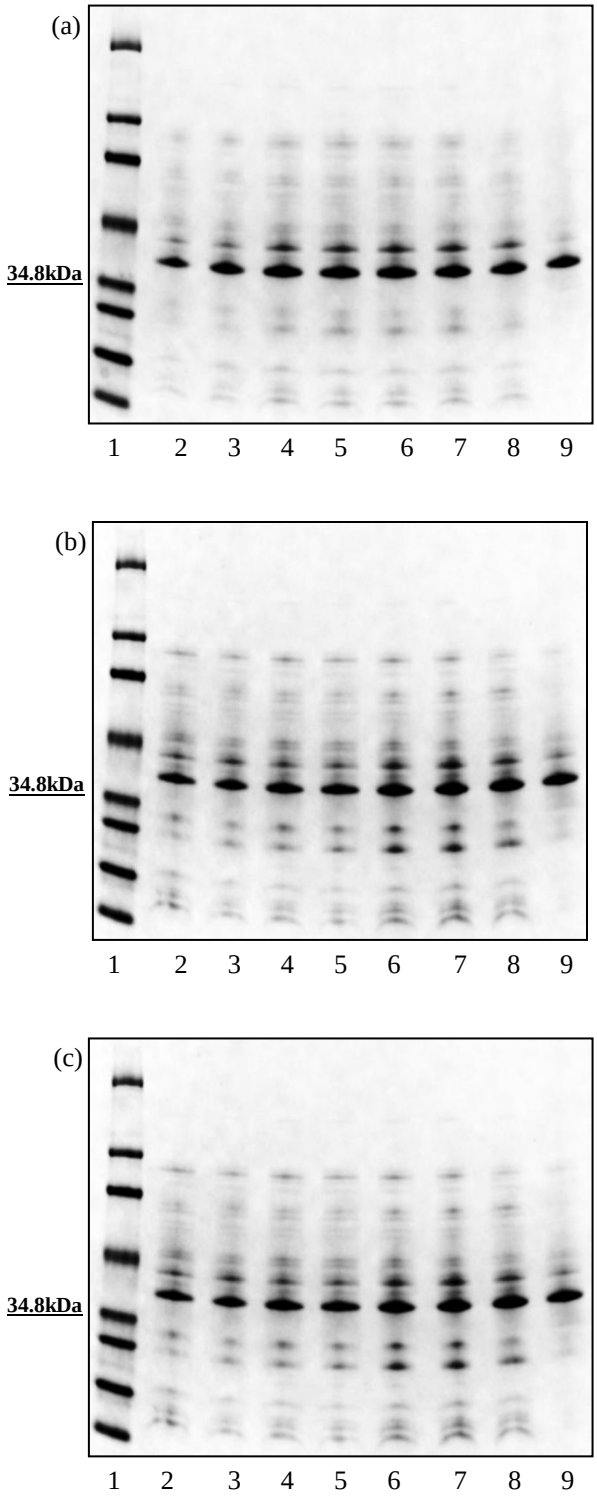


Figure 3.17. Time course analysis of induced SGRA_3210 for 24 hours at 37°C and at 1.0 OD_{600}. IPTG was added when OD_{600} reached to 1.0. (a) 0.05 mM IPTG (b) 0.2 mM IPTG (c) 0.6 mM IPTG. Lanes: 1, molecular weight marker; 2, uninduced; 3, 1 hour after induction; 4, 2 hours after induction; 5, 3 hours after induction; 6, 4 hours after induction; 7, 6 hours after induction; 8, 12 hours after induction; 8, 24 hours after induction. Soluble proteins were extracted with BugBuster protein extraction reagent (Novagen) and examined on 4-20% SDS polyacrylamide gels.

Secondly, SGRA_0571 (RsbR2) protein underwent time course analyses. It was induced at 30℃ and 37℃. Different concentration of IPTG were added at 1.0 OD_{600}. Since the culture at 30℃ was not shown an expected pinkish to reddish pellet due to heme synthesis, this culture was omitted (Data is not shown) for analysis. Instead, induction at 32℃ was done. RsbR2 was induced at 37℃ but its amount was too small to recognize (Figure 3.18). Induction at 30℃ yielded more protein than at 37℃ overall and the bands were able to be seen at around 33 kDa region (Figure 3.19). The thickest band was present on Figure 3.19 (a). This sample was collected after 6 hours of induction with 0.05 mM IPTG.

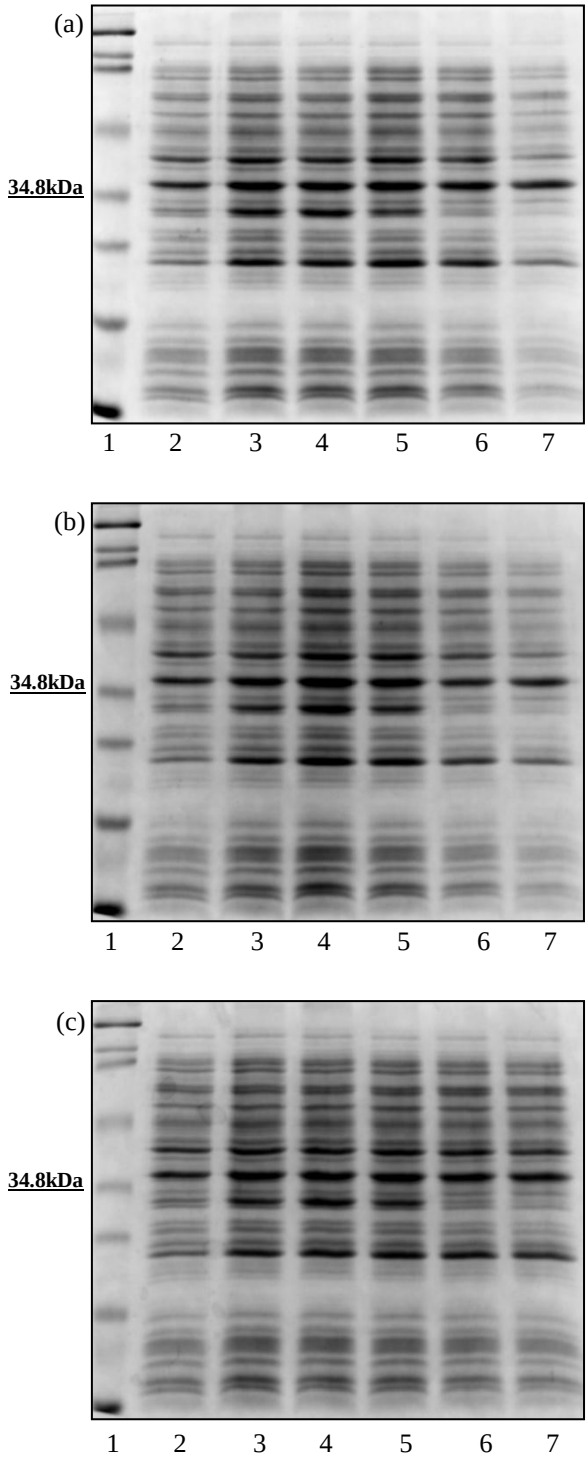


Figure 3.18. Time course analysis of induced SGRA_0571 at 37°C for 24 hours. IPTG was added when OD_{600} reached to 1.0. (a) 0.05 mM IPTG (b) 0.2 mM IPTG (c) 0.6 mM IPTG. Lanes: 1, molecular weight marker; 2, uninduced; 3, 2 hours after induction; 4, 4 hours after induction; 5, 6 hours after induction; 6, 12 hours after induction; 7, 24 hours after induction. Soluble proteins were extracted with BugBuster protein extraction reagent (Novagen) and examined on 10% SDS polyacrylamide gels.

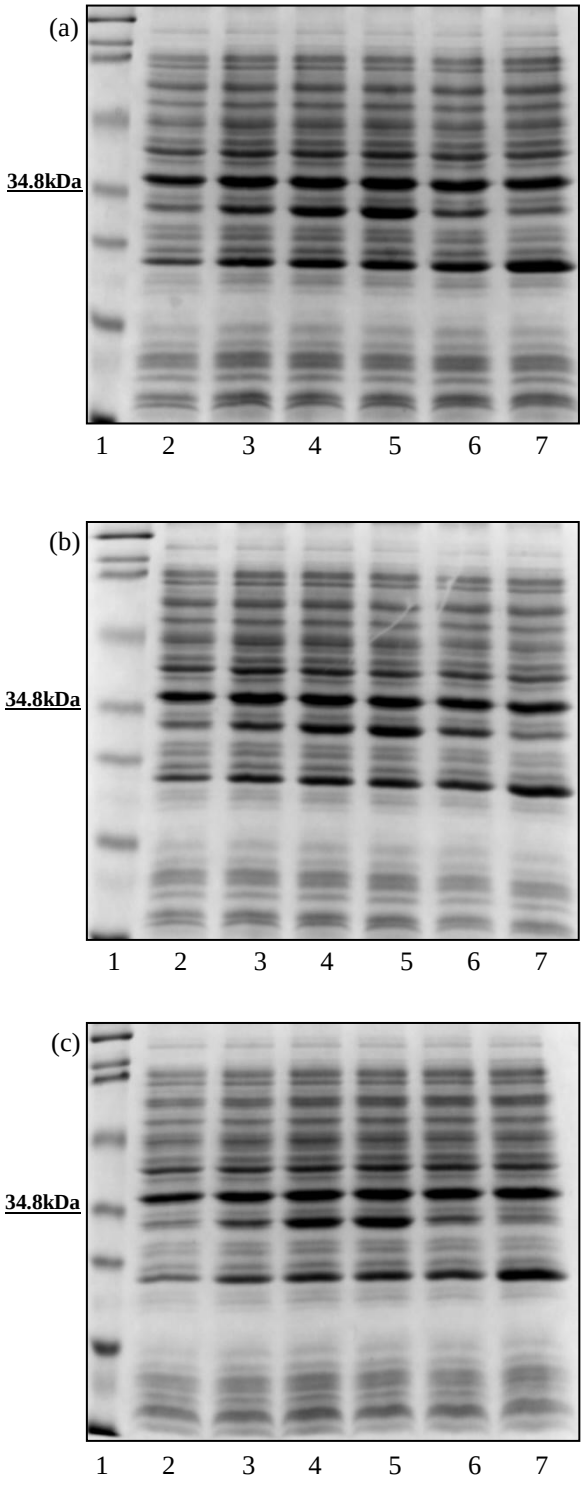


Figure 3.19. Time course analysis of induced SGRA_0571 at 32°C for 24 hours. IPTG was added when OD_{600} reached to 1.0. (a) 0.05 mM IPTG (b) 0.2 mM IPTG (c) 0.6 mM IPTG. Lanes: 1, molecular weight marker; 2, uninduced; 3, 2 hours after induction; 4, 4 hours after induction; 5, 6 hours after induction; 6, 12 hours after induction; 7, 24 hours after induction. Soluble proteins were extracted with BugBuster protein extraction reagent (Novagen) and examined on 10% SDS polyacrylamide gels.

Nine proteins were expressed at 30℃ with OD_{600} at 1.0 except SGRA_0571. Different IPTG concentrations were added to find the best condition to produce more soluble protein. Time course analyses were done and examined on SDS polyarylamide gels (Figure 3.20-3.37). Expression levels of these proteins were higher than SGRA_0571 and SGRA_3210. Based on SDS polyacrylamide gel analyses, expression conditions of ten RsbR proteins were determined (Table 3.4). Temperature at 30℃ except SGRA_0571 (32℃) and OD_{600} at 1.0 were chosen as the common parameters. IPTG concentration and induction time were varied. In addition, shaking at low rpm (150 rpm) produced more soluble protein instead of vigorous shaking (Data is not shown).

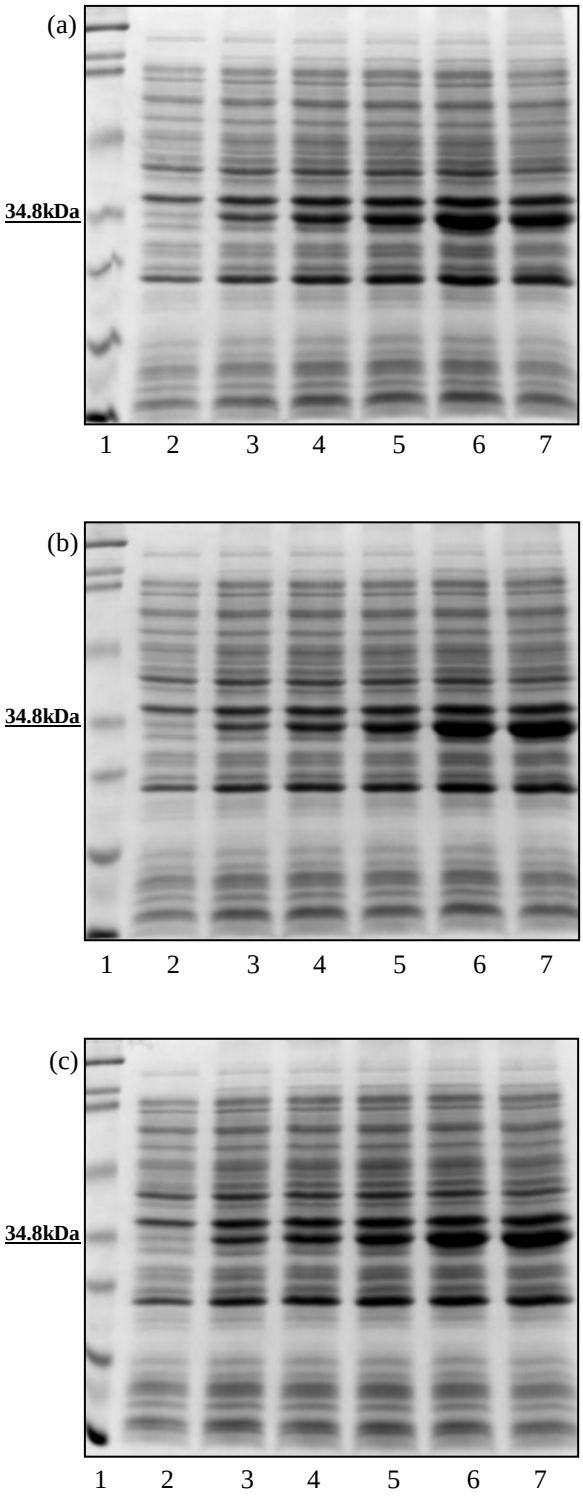


Figure 3.20. Time course analysis of induced SGRA_1293 at 30°C for 24 hours. IPTG was added when OD_{600} reached to 1.0. (a) 0.05 mM IPTG (b) 0.2 mM IPTG (c) 0.6 mM IPTG. Lanes: 1, molecular weight marker; 2, uninduced; 3, 2 hours after induction; 4, 4 hours after induction; 5, 6 hours after induction; 6, 12 hours after induction; 7, 24 hours after induction. Soluble proteins were extracted with BugBuster protein extraction reagent (Novagen) and examined on 10% SDS polyacrylamide gels.

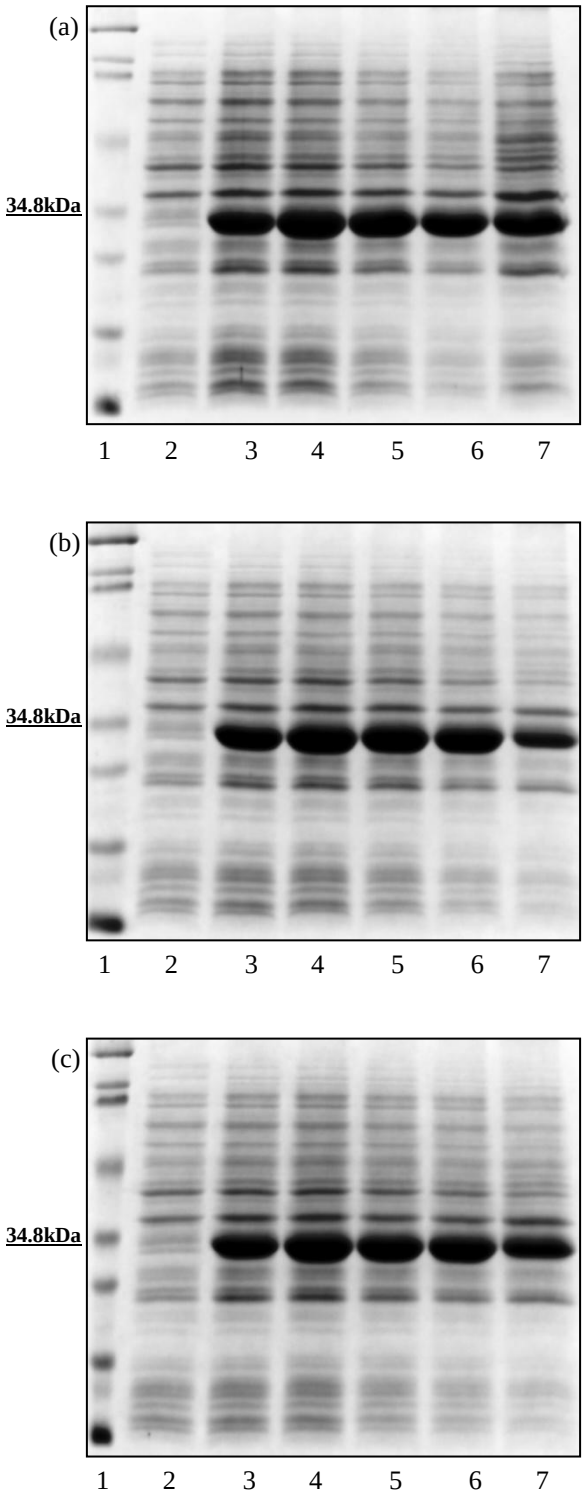


Figure 3.21. Time course analysis of induced SGRA_2160 at 30°C for 24 hours. IPTG was added when OD_{600} reached to 1.0. (a) 0.05 mM IPTG (b) 0.2 mM IPTG (c) 0.6 mM IPTG. Lanes: 1, molecular weight marker; 2, uninduced; 3, 2 hours after induction; 4, 4 hours after induction; 5, 6 hours after induction; 6, 12 hours after induction; 7, 24 hours after induction. Soluble proteins were extracted with BugBuster protein extraction reagent (Novagen) and examined on 10% SDS polyacrylamide gels.

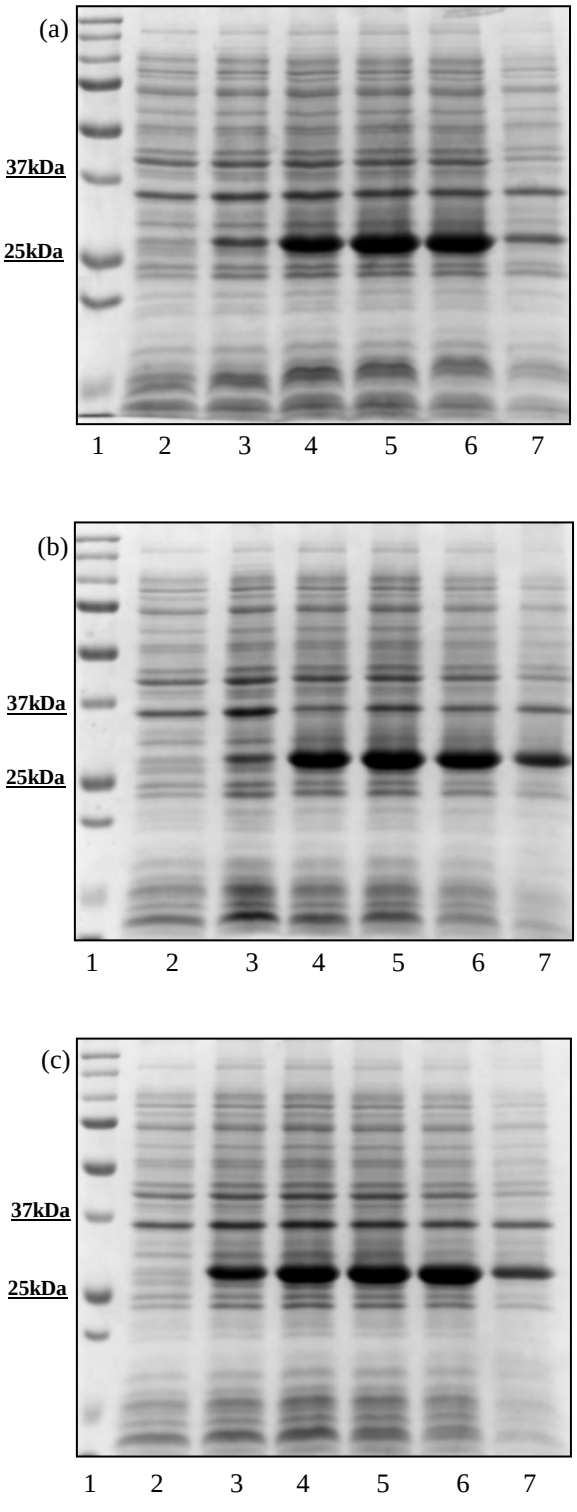


Figure 3.22. Time course analysis of induced SGRA_2161 at 30°C for 24 hours. IPTG was added when OD_{600} reached to 1.0. (a) 0.05 mM IPTG (b) 0.2 mM IPTG (c) 0.6 mM IPTG. Lanes: 1, molecular weight marker; 2, uninduced; 3, 2 hours after induction; 4, 4 hours after induction; 5, 6 hours after induction; 6, 12 hours after induction; 7, 24 hours after induction. Soluble proteins were extracted with BugBuster protein extraction reagent (Novagen) and examined on 12.5% SDS polyacrylamide gels.

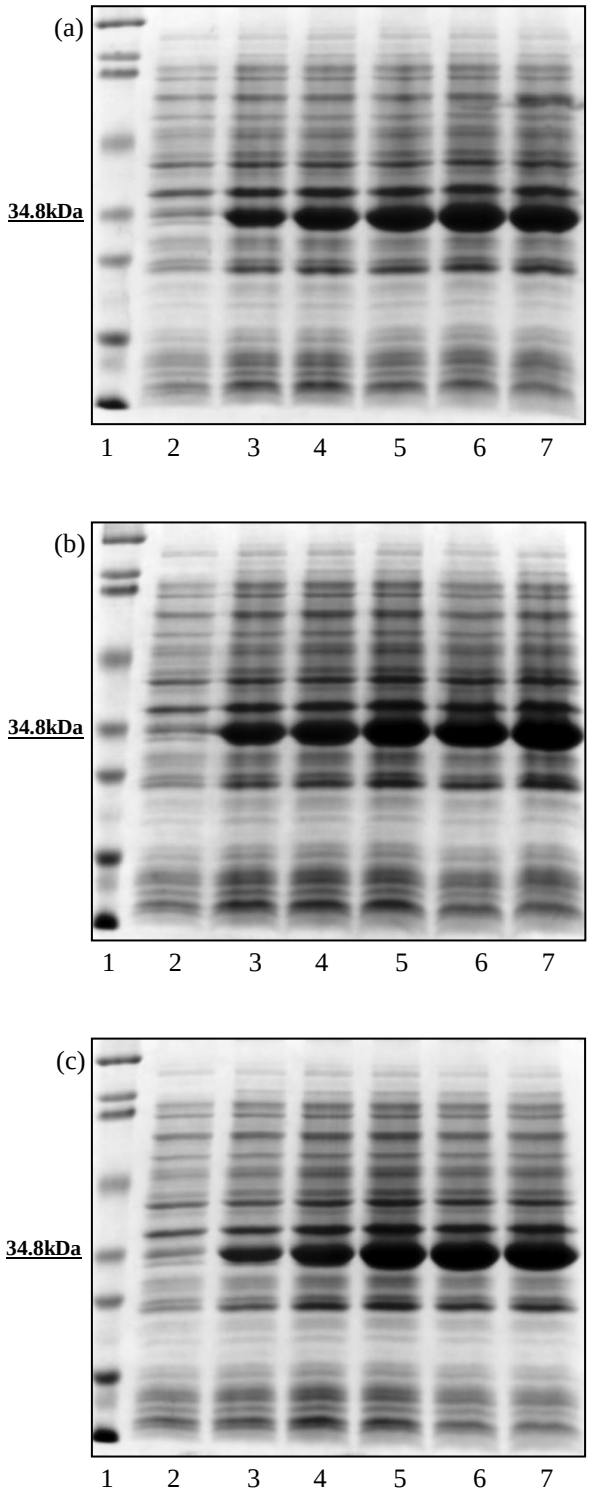


Figure 3.23. Time course analysis of induced SGRA_2162 at 30°C for 24 hours. IPTG was added when OD_{600} reached to 1.0. (a) 0.05 mM IPTG (b) 0.2 mM IPTG (c) 0.6 mM IPTG. Lanes: 1, molecular weight marker; 2, uninduced; 3, 2 hours after induction; 4, 4 hours after induction; 5, 6 hours after induction; 6, 12 hours after induction; 7, 24 hours after induction. Soluble proteins were extracted with BugBuster protein extraction reagent (Novagen) and examined on 10% SDS polyacrylamide gels.

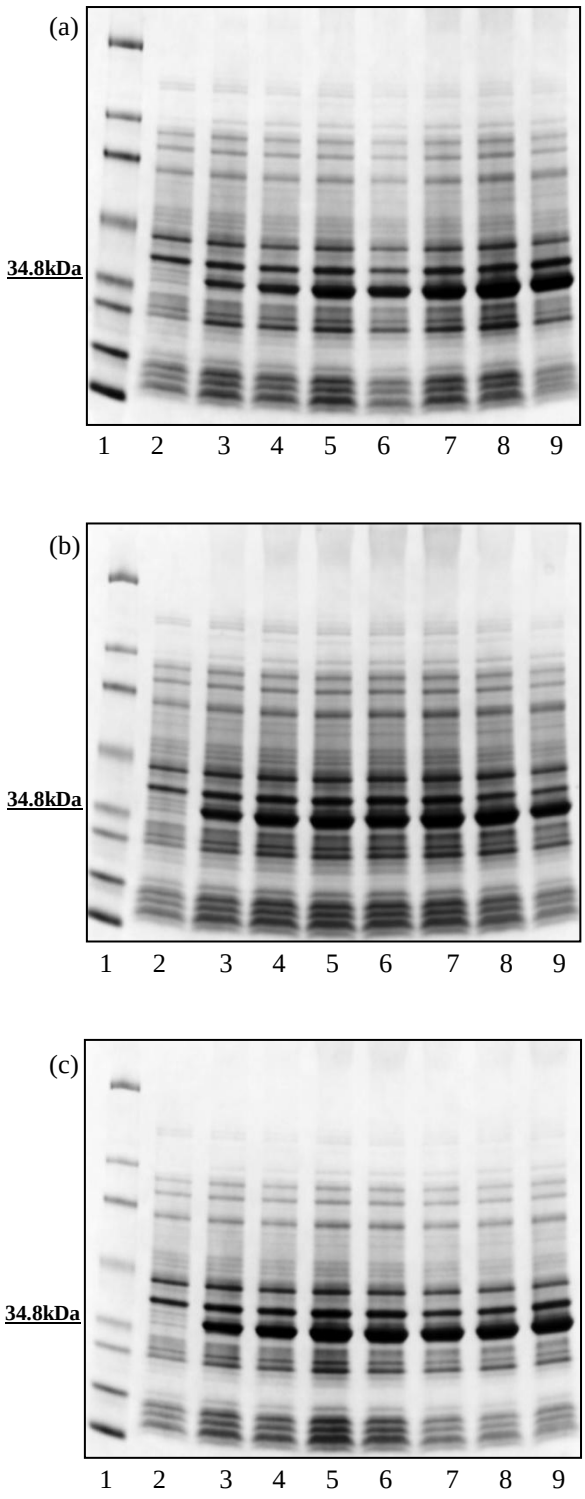


Figure 3.24. Time course analysis of induced SGRA_2167 at 30°C for 24 hours. IPTG was added when OD_{600} reached to 1.0. (a) 0.05 mM IPTG (b) 0.2 mM IPTG (c) 0.6 mM IPTG. Lanes: 1, molecular weight marker; 2, uninduced; 3, 1 hour after induction; 4, 2 hours after induction; 5, 3 hours after induction; 6, 4 hours after induction; 7, 6 hours after induction; 8, 12 hours after induction; 8, 24 hours after induction. Soluble proteins were extracted with BugBuster protein extraction reagent (Novagen) and examined on 4-20% SDS polyacrylamide gels.

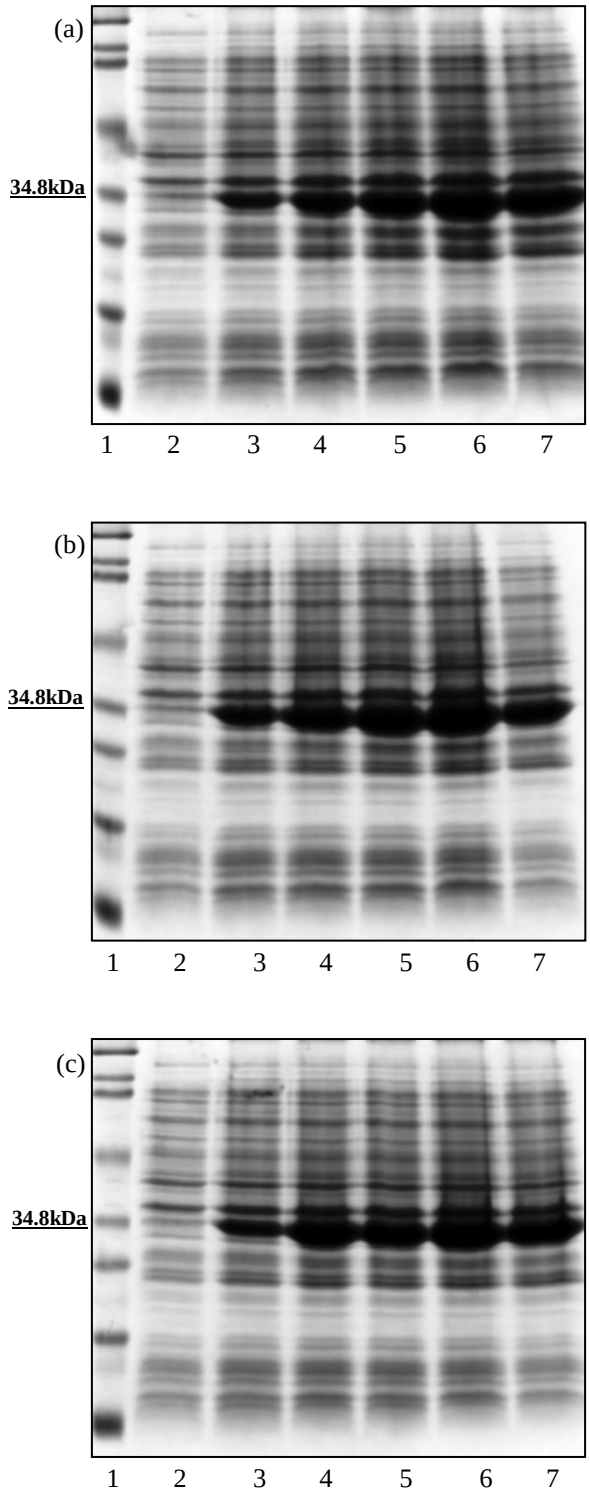


Figure 3.25. Time course analysis of induced SGRA_2168 at 30°C for 24 hours. IPTG was added when OD_{600} reached to 1.0. (a) 0.05 mM IPTG (b) 0.2 mM IPTG (c) 0.6 mM IPTG. Lanes: 1, molecular weight marker; 2, uninduced; 3, 2 hours after induction; 4, 4 hours after induction; 5, 6 hours after induction; 6, 12 hours after induction; 7, 24 hours after induction. Soluble proteins were extracted with BugBuster protein extraction reagent (Novagen) and examined on 10% SDS polyacrylamide gels.

(a)

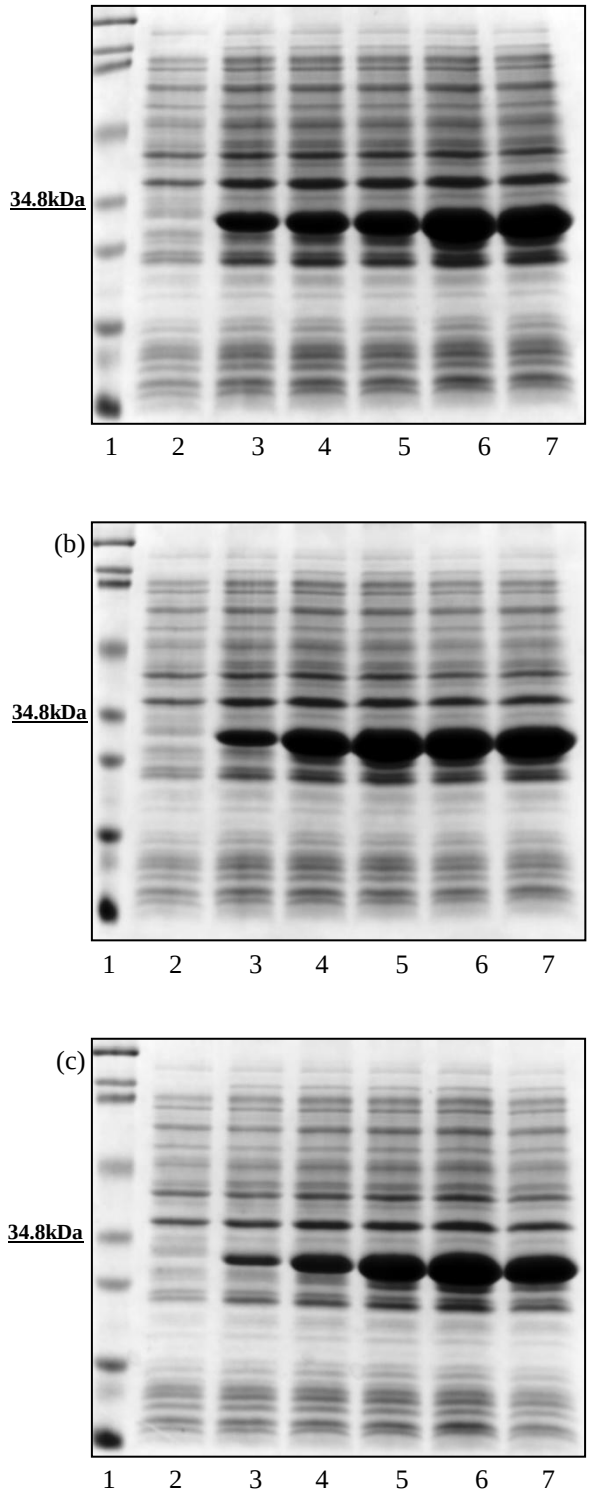


Figure 3.26. Time course analysis of induced SGRA_2169 at 30°C for 24 hours. IPTG was added when OD_{600} reached to 1.0. (a) 0.05 mM IPTG (b) 0.2 mM IPTG (c) 0.6 mM IPTG. Lanes: 1, molecular weight marker; 2, uninduced; 3, 2 hours after induction; 4, 4 hours after induction; 5, 6 hours after induction; 6, 12 hours after induction; 7, 24 hours after induction. Soluble proteins were extracted with BugBuster protein extraction reagent (Novagen) and examined on 10% SDS polyacrylamide gels.

(a)

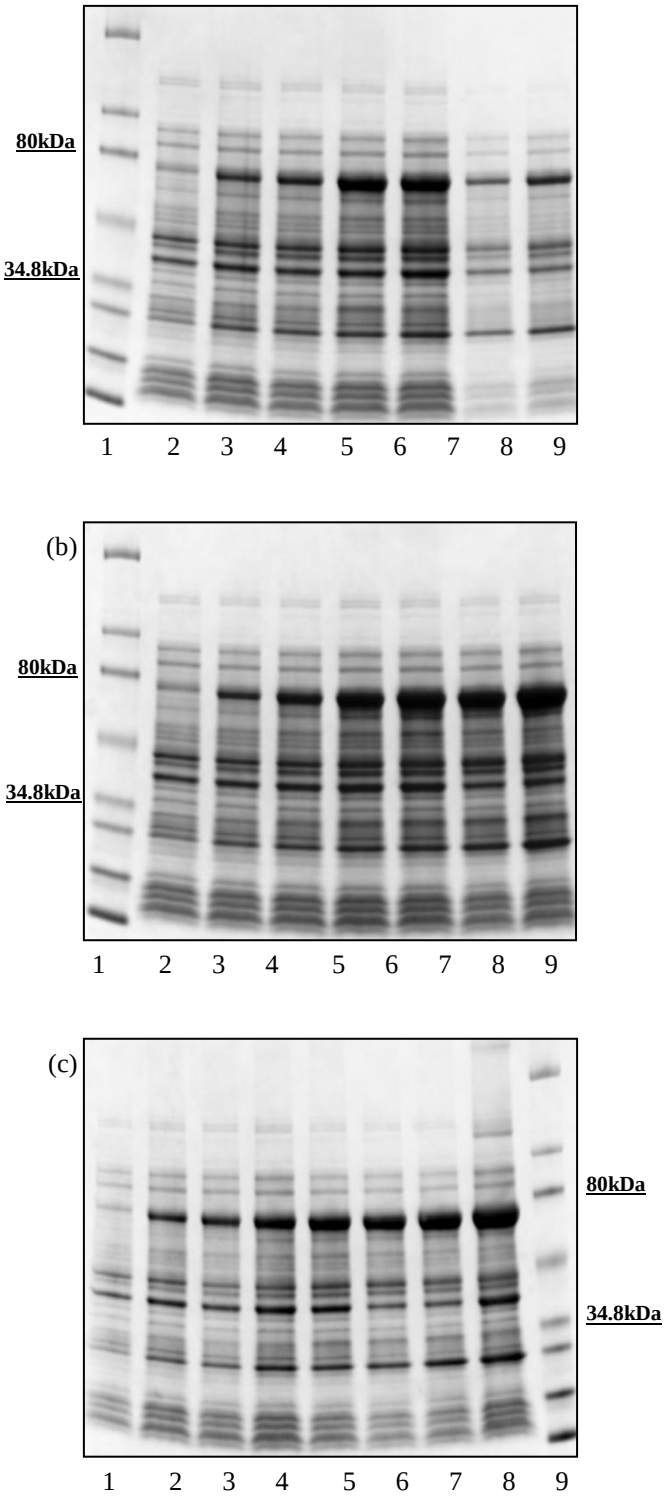


Figure 3.27. Time course analysis of induced SGRA_3852 at 30°C for 24 hours. IPTG was added when OD_{600} reached to 1.0. (a) 0.05 mM IPTG (b) 0.2 mM IPTG (c) 0.6 mM IPTG. Lanes for (a) and (b): 1, molecular weight marker; 2, uninduced; 3, 1 hour after induction; 4, 2 hours after induction; 5, 3 hours after induction; 6, 4 hours after induction; 7, 6 hours after induction; 8, 12 hours after induction; 8, 24 hours after induction. Lanes for (c): 1, uninduced; 2, 1 hour after induction; 3, 2 hour after induction; 4, 3 hour after induction; 5, 4 hour after induction; 6, 6 hour after induction; 7, 12 hour after induction; 8, 24 hour after induction; 9, Molecular weight marker. Soluble proteins were extracted with BugBuster protein extraction reagent (Novagen) and examined on 4-20% SDS polyacrylamide gels.

Table 3.4. Protein expression conditions of ten RsbR proteins

Locus tag	Temperature (°C)	OD_{600}	IPTG (mM)	Time (hour)
SGRA_0571	32	1	0.05	6
SGRA_1293	30	1	0.2	12
SGRA_2160	30	1	0.05	4
SGRA_2161	30	1	0.05	6
SGRA_2162	30	1	0.2	6
SGRA_2167	30	1	0.6	4
SGRA_2168	30	1	0.2	12
SGRA_2169	30	1	0.2	6
SGRA_3210	30	1	0.2	12
SGRA_3852	30	1	0.05	6

3.2.2. Optimization and purification

Ten proteins were expressed in large scale according to conditions on Table 3.4. Induced cultures were harvested by centrifugation and cells were resuspended by appropriate lysis buffer and lysed by sonication. After sonication, raw crude extracts were centrifuged to separate soluble and insoluble proteins. Supernatants which contained soluble proteins were used for purification. Some proteins were required different lysis buffer conditions such as pH and inclusion of additives. Addition of 1% glucose and adjusting pH to 7.6 in lysis buffer increased stability of RsbR1 (SGRA_3210) during extraction (Figure 3.29). RsbR1 induced cell pellets were extracted by adding lysis buffer with pH at 8.0 or 7.6, with or without 1% glucose. The pellets that were extracted with lysis buffer at pH8.0 showed early elution of RsbR1 when they were eluted with elution buffer containing 150 mM imidazole. Presence and absence of 1% glucose did not show much difference (Figure 3.28). In contrast, the pellet that extracted by lysis buffer at pH7.6 with 1% glucose yielded more protein compared to the

pellet without 1% glucose. Majority of RsbR1 was eluted by adding elution buffer with 300 mM imidazole. However, adding 1% glucose in wash and elution buffers did not prevent early elution if protein was not extracted with lysis buffer containing 1% glucose (Figure 3.29).

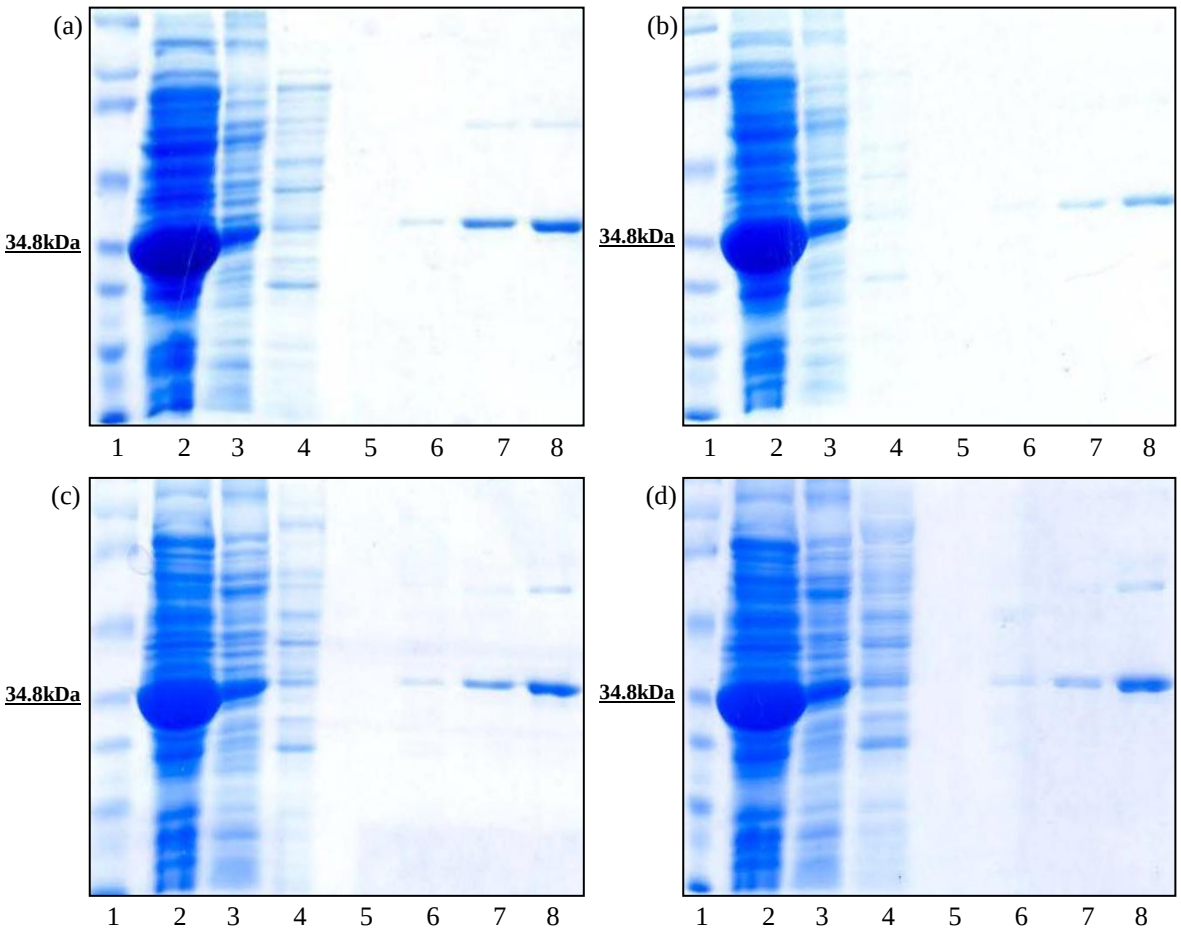


Figure 3.28. Optimization of purifying RsbR1 with and without adding 1% Glucose at pH8.0. (a) Without adding 1% glucose in a lysis buffer and purification buffers (b) Without adding 1% glucose in a lysis buffer and adding 1% glucose in purification buffers (c) Adding 1% glucose in a lysis buffer and without adding 1% glucose in purification buffers (d) Adding 1% glucose in a lysis buffer and adding 1% glucose in purification buffers. Lanes: 1, molecular weight marker; 2, crude extract; 3, flow-through; 4, wash 1; 5, wash 2; 6, elute 1; 7, elute 2; 8, elute 3. Each fraction was examined on a 10% SDS polyacrylamide gel and stained with comassie blue.

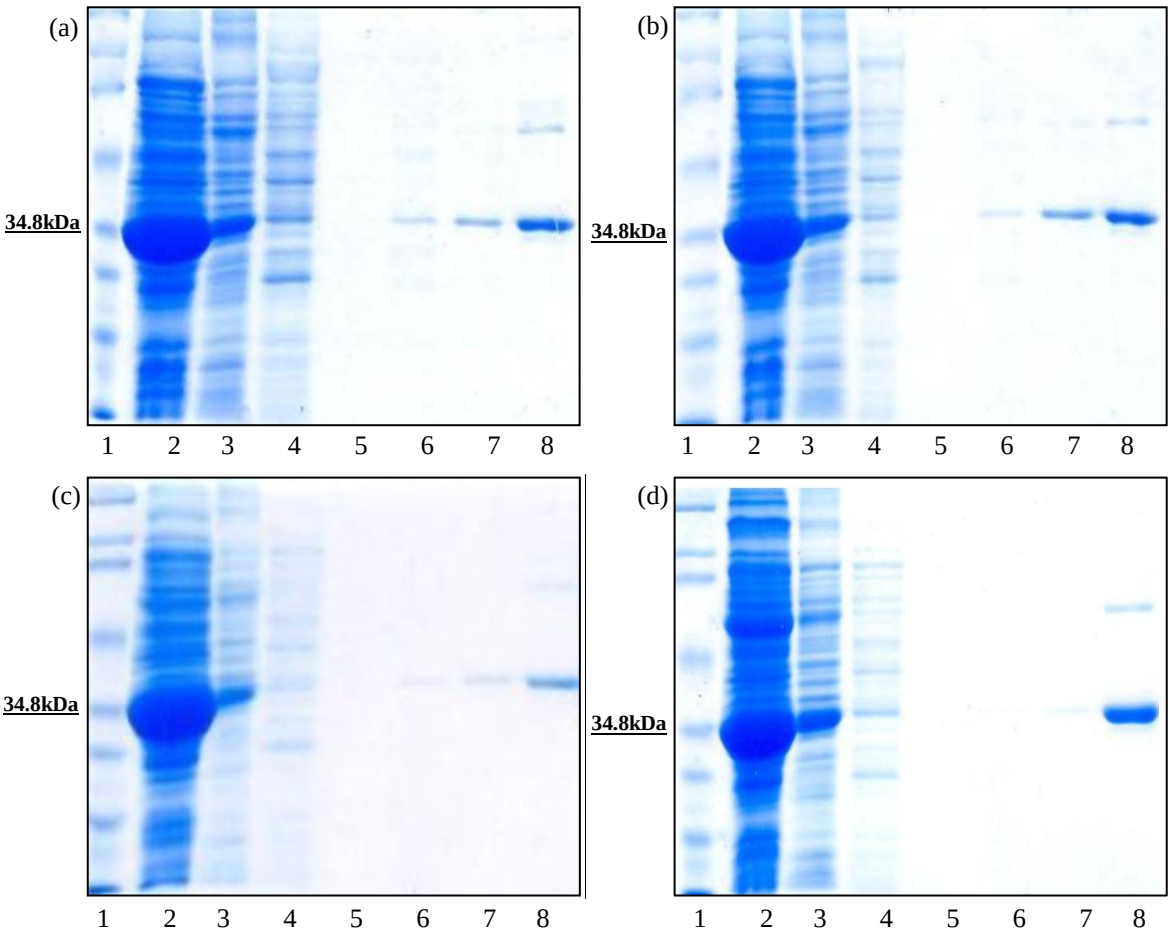


Figure 3.29. Optimization of purifying RsbR1 with and without adding 1% Glucose at pH7.6. (a) Without adding 1% glucose in a lysis buffer and purification buffers (b) Without adding 1% glucose in a lysis buffer and adding 1% glucose in purification buffers (c) Adding 1% glucose in a lysis buffer and without adding 1% glucose in purification buffers (d) Adding 1% glucose in a lysis buffer and adding 1% glucose in purification buffers. Lanes: 1, molecular weight marker; 2, crude extract; 3, flow-through; 4, wash 1; 5, wash 2; 6, elute 1; 7, elute 2; 8, elute 3. Each fraction was examined on a 10% SDS polyacrylamide gel and stained with comassie blue.

Crude extracts were directly poured onto pre-loaded Ni-NTA resins after equilibration of the resins. Most proteins were purified by this column purification method but some proteins including SGRA_0571, SGRA_2161, SGRA _3210 and SGRA _3854 required the alternative way which was batch purification. This was because they did not bind to resins and they collected as a flow-through fraction. Thus, eluted proteins were not enough to visualize on SDS polyacrylamide gels. By mixing crude extracts and resins at 4℃ for 3 hours to overnight, the binding capacity onto the resins were improved and able to get more protein (Figure 3.30).

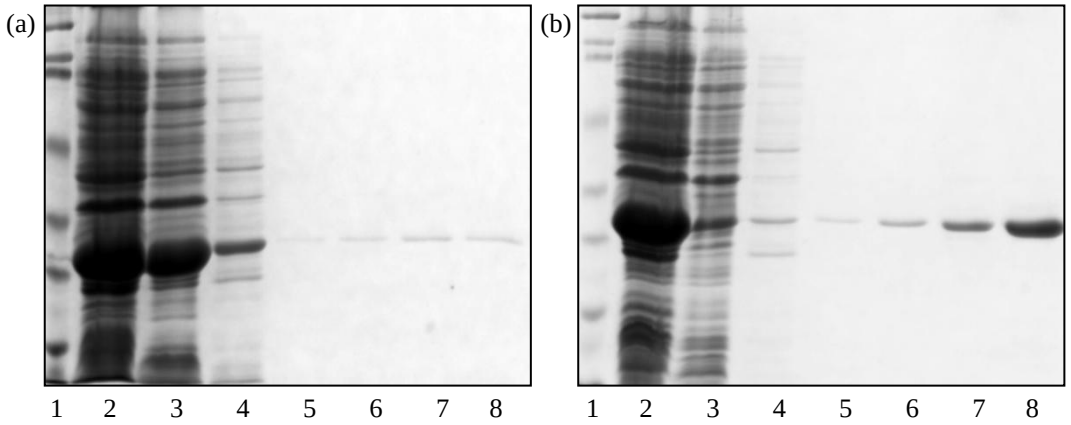


Figure 3.30. Purification of SGRA_2161. (a) Column purification (b) Batch purification. Lanes: 1, molecular weight marker; 2, crude extract; 3, flow-through; 4-5, wash; 6-8, elute. Each fraction was examined on a 10% SDS polyacrylamide gel and stained with comassie blue.

So far, all ten recombinant RsbR proteins were purified by nickel affinity chromatography and their fractions from purification were examined on SDS polyacrylamide gels (Figure 3.30-3.40). Running SDS-PAGE gave us two information, for example, purity and approximate molecular weight of polypeptides (Walsh, 2002). Purity of each protein was varied. Overall, eight out of ten proteins were appeared to be pure greater than 80% and two of them were less than 80% (SGRA_2168 and SGRA_3852) based on SDS-PAGE analyses.

Non-specific bands were present every gels by looking at bare eyes but their presence was not well visualized on pictures.

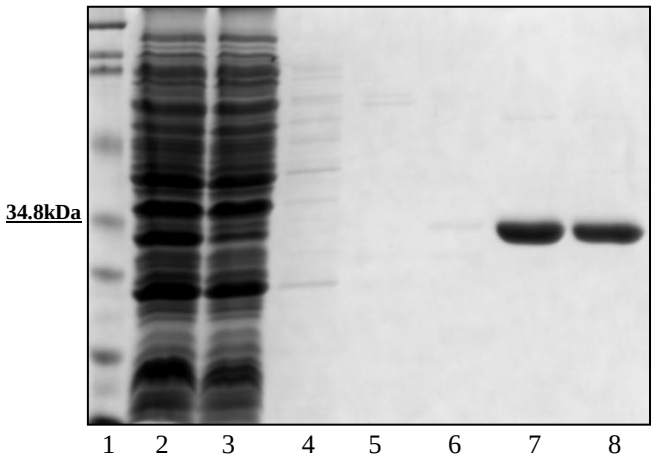


Figure 3.31. Purification of SGRA_0571. Lanes: 1, molecular weight marker; 2, crude extract; 3, flow-through; 4, wash 1 with the wash buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 20 mM imidazole, at pH8); 5, wash 2 with the wash buffer II (50 mM Na_2HPO_4, 300 mM NaCl, 50 mM imidazole, at pH8); 6, wash 3 with the wash buffer III (50 mM Na_2HPO_4, 300 mM NaCl, 100 mM imidazole, at pH8); 7, wash 4 with the wash buffer IV (50 mM Na_2HPO_4, 300 mM NaCl, 150 mM imidazole, at pH8); 8, elute with the elution buffer (50 mM Na_2HPO_4, 300 mM NaCl, 300 mM imidazole, at pH8). Each fraction was examined on a 12.5% SDS polyacrylamide gel and stained with comassie blue.

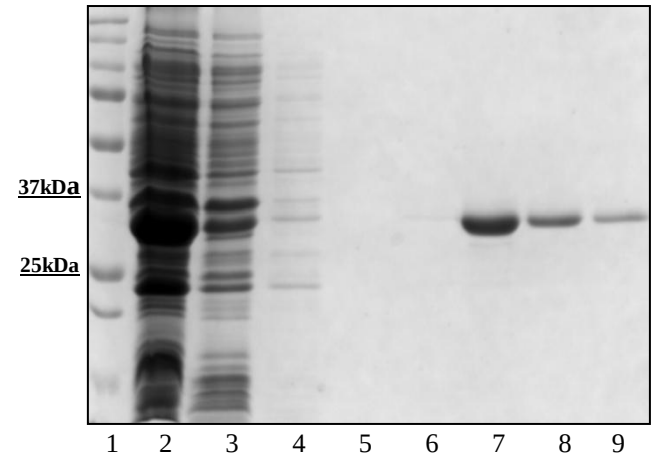


Figure 3.32. Purification of SGRA_1293. Lanes: 1, molecular weight marker; 2, crude extract; 3, flow-through; 4, wash 1 with the wash buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 20 mM imidazole, at pH8); 5, wash 2 with the wash buffer II (50 mM Na_2HPO_4, 300 mM NaCl, 50 mM imidazole, at pH8); 6, wash 3 with the wash buffer III (50 mM Na_2HPO_4, 300 mM NaCl, 100 mM imidazole, at pH8); 7, elute 1 with the elution buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 150 mM imidazole, at pH8); 8, elute 2 with the elution buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 150 mM imidazole, at pH8); 9, elute 3 with the elution buffer II (50 mM Na_2HPO_4, 300 mM NaCl, 300 mM imidazole, at pH8). Each fraction was examined on a 12.5% SDS polyacrylamide gel and stained with comassie blue.

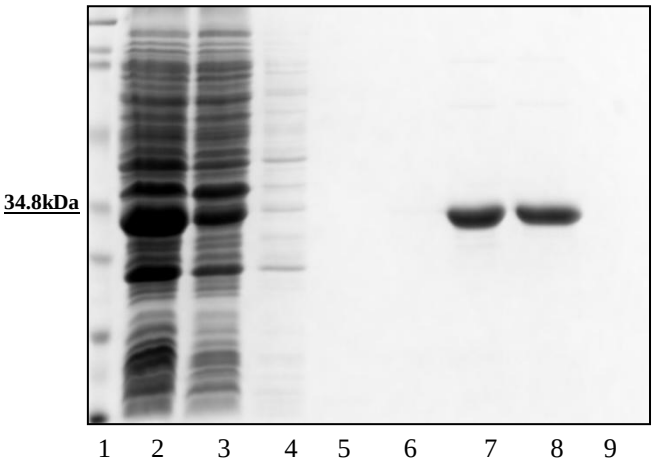


Figure 3.33. Purification of SGRA_2160. Lanes: 1, molecular weight marker; 2, crude extract; 3, flow-through; 4, wash 1 with the wash buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 20 mM imidazole, at pH8); 5, wash 2 with the wash buffer II (50 mM Na_2HPO_4, 300 mM NaCl, 50 mM imidazole, at pH8); 6, wash 3 with the wash buffer III (50 mM Na_2HPO_4, 300 mM NaCl, 100 mM imidazole, at pH8); 7, elute 1 with the elution buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 150 mM imidazole, at pH8); 8, elute 2 with the elution buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 150 mM imidazole, at pH8); 9, elute 3 with the elution buffer II (50 mM Na_2HPO_4, 300 mM NaCl, 300 mM imidazole, at pH8). Each fraction was examined on a 12.5% SDS polyacrylamide gel and stained with comassie blue.

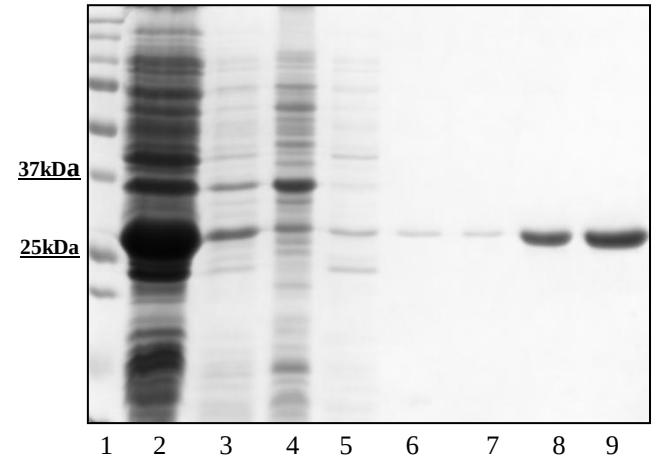


Figure 3.34. Purification of SGRA_2161. Lanes: 1, molecular weight marker; 2, crude extract; 3, pellet after centrifugation; 4, flow-through; 5, wash 1 with the wash buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 20 mM imidazole, at pH8); 6, wash 2 with the wash buffer II (50 mM Na_2HPO_4, 300 mM NaCl, 50 mM imidazole, at pH8); 7, elute 1 with the elution buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 100 mM imidazole, at pH8); 8, elute 2 with the elution buffer II (50 mM Na_2HPO_4, 300 mM NaCl, 150 mM imidazole, at pH8); 9, elute 3 with the elution buffer III (50 mM Na_2HPO_4, 300 mM NaCl, 300 mM imidazole, at pH8). Each fraction was examined on a 12.5% SDS polyacrylamide gel and stained with comassie blue.

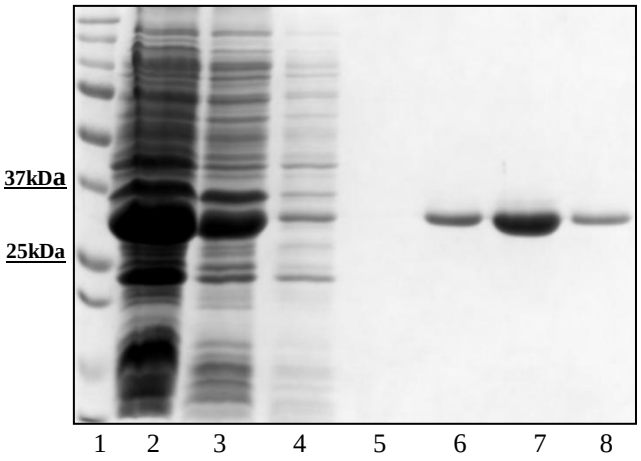


Figure 3.35. Purification of SGRA_2162. Lanes: 1, molecular weight marker; 2, crude extract; 3, flow-through; 4, wash 1 with the wash buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 20 mM imidazole, at pH8); 5, wash 2 with the wash buffer II (50 mM Na_2HPO_4, 300 mM NaCl, 50 mM imidazole, at pH8); 6, elute 1 with the elution buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 100 mM imidazole, at pH8); 7, elute 2 with the elution buffer II (50 mM Na_2HPO_4, 300 mM NaCl, 150 mM imidazole, at pH8); 8, elute 3 with the elution buffer III (50 mM Na_2HPO_4, 300 mM NaCl, 300 mM imidazole, at pH8). Each fraction was examined on a 12.5% SDS polyacrylamide gel and stained with comassie blue.

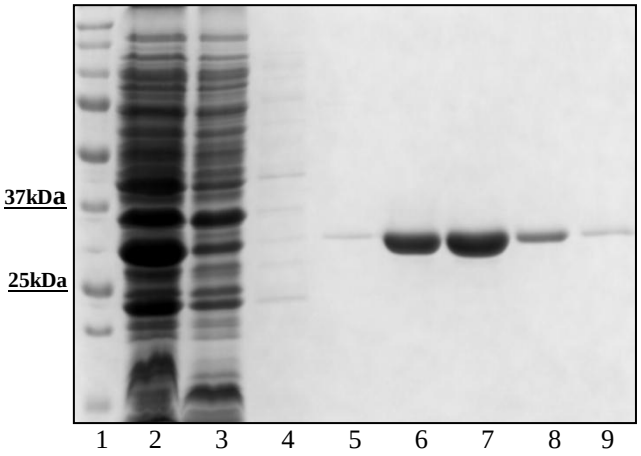


Figure 3.36. Purification of SGRA_2167. Lanes: 1, molecular weight marker; 2, crude extract; 3, flow-through; 4, wash 1 with the wash buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 20 mM imidazole, at pH8); 5, wash 2 with the wash buffer II (50 mM Na_2HPO_4, 300 mM NaCl, 100 mM imidazole, at pH8); 6-9, elute 1-4 with the elution buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 150 mM imidazole, at pH8). Each fraction was examined on a 12.5% SDS polyacrylamide gel and stained with comassie blue.

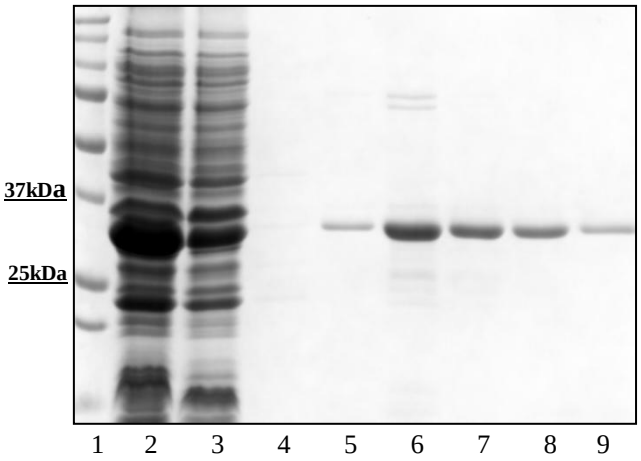


Figure 3.37. Purification of SGRA_2168. Lanes: 1, molecular weight marker; 2, crude extract; 3, flow-through; 4, wash 1 with the wash buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 20 mM imidazole, at pH8); 5-9, elute 1-5 with the elution buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 100 mM imidazole, at pH8). Each fraction was examined on a 12.5% SDS polyacrylamide gel and stained with comassie blue.

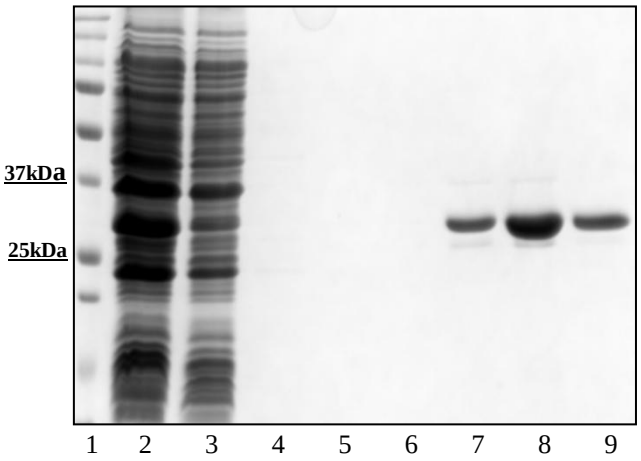


Figure 3.38. Purification of SGRA_2169. Lanes: 1, molecular weight marker; 2, crude extract; 3, flow-through; 4, wash 1 with the wash buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 20 mM imidazole, at pH8); 5, wash 2 with the wash buffer II (50 mM Na_2HPO_4, 300 mM NaCl, 50 mM imidazole, at pH8); 6, wash 3 with the wash buffer III (50 mM Na_2HPO_4, 300 mM NaCl, 100 mM imidazole, at pH8); 7, elute 1 with the elution buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 150 mM imidazole, at pH8); 8-9, elute 2-3 with the elution buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 300 mM imidazole, at pH8). Each fraction was examined on a 12.5% SDS polyacrylamide gel and stained with comassie blue.

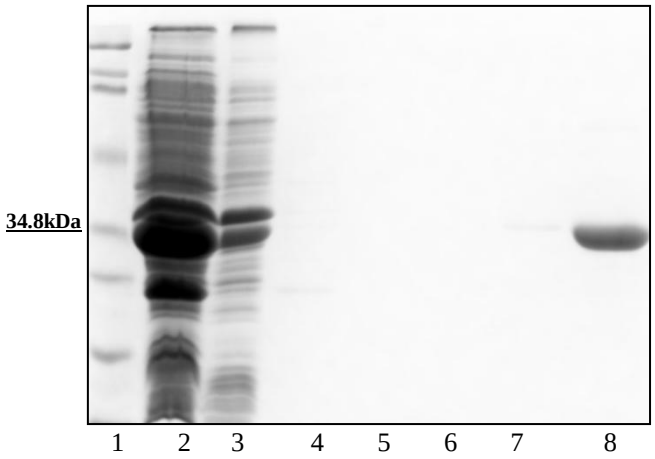


Figure 3.39. Purification of SGRA_3210. Lanes: 1, molecular weight marker; 2, crude extract; 3, flow-through; 4, wash 1 with the wash buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 20 mM imidazole, 1% glucose at pH7.6); 5, wash 2 with the wash buffer II (50 mM Na_2HPO_4, 300 mM NaCl, 50 mM imidazole, 1% glucose at pH7.6); 6, wash 3 with the wash buffer III (50 mM Na_2HPO_4, 300 mM NaCl, 100 mM imidazole, 1% glucose at pH7.6); 7, wash 4 with the wash buffer IV (50 mM Na_2HPO_4, 300 mM NaCl, 150 mM imidazole, 1% glucose at pH7.6); 8, elute with the elution buffer (50 mM Na_2HPO_4, 300 mM NaCl, 300 mM imidazole, 1% glucose at pH7.6). Each fraction was examined on a 10% SDS polyarylamide gel and stained with comassie blue.

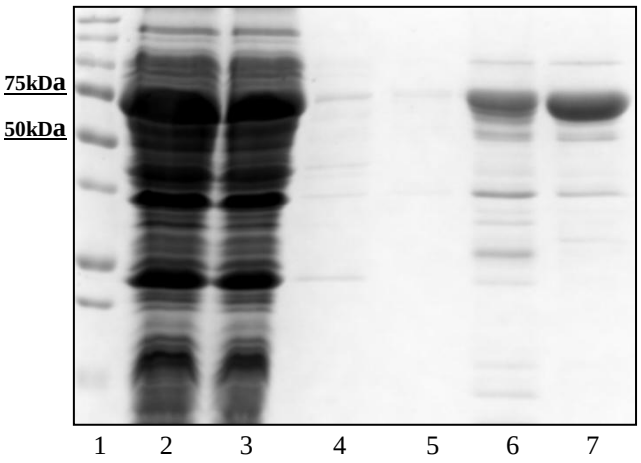


Figure 3.40. Purification of SGRA_3852. Lanes: 1, molecular weight marker; 2, crude extract; 3, flow-through; 4, wash 1 with the wash buffer I (50 mM Na_2HPO_4, 300 mM NaCl, 20 mM imidazole at pH7.6); 5, elute 1 with the elution buffer (50 mM Na_2HPO_4, 300 mM NaCl, 100 mM imidazole at pH7.6); 6, elute 2 with the elution buffer (50 mM Na_2HPO_4, 300 mM NaCl, 150 mM imidazole at pH7.6); 7, elute 3 with the elution buffer (50 mM Na_2HPO_4, 300 mM NaCl, 300 mM imidazole at pH7.6). Each fraction was examined on a 12.5% SDS polyacrylamide gel and stained with comassie blue.

Figure 3.41 shows all ten purified recombinant proteins that visualized on two SDS polyacrylamide gels. Molecular mass of each band that represents purified recombinant proteins was determined by Gel Doc EZ image analyzing software (Bio-Rad). Point-to-Point (Semi-log) regression method was used and determined molecular mass was shown on Table 3.6 (Measured MW). A range of molecular weight of ten proteins is from 27.5 to 67.5 kDa. Due to a maltose binding protein adjacent to SGRA_3852, the mass of SGRA_3853 had 42 kDa more than its own mass. Generally, calculated molecular weight was less than predicted molecular weight of ten RsbR proteins. Protein concentration of the purified proteins was determined followed by dialysis.

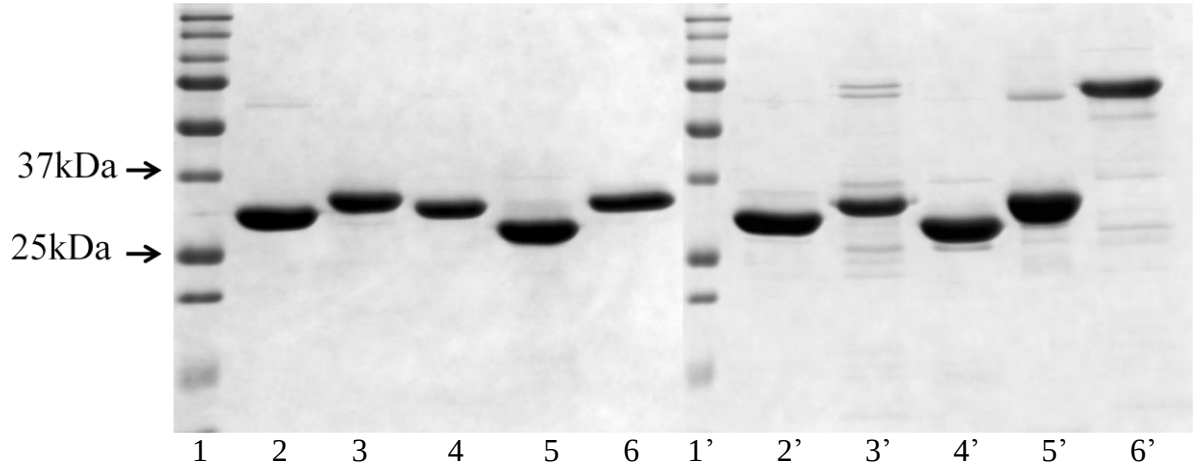


Figure 3.41. Purification of His-tagged ten recombinant RsbR proteins. Proteins were examined on 12.5% SDS polyacrylamide gels and stained with Coomassie brilliant blue. Lanes: 1 and 1', molecular weight markers; 2, SGRA_0571; 3, SGRA_1293; 4, SGRA_2160; 5, SGRA_2161; 6, SGRA_2162; 2', SGRA_2167; 3', SGRA_2168; 4', SGRA_2169; 5', SGRA_3210; 6', SGRA_3852 with a maltose binding protein (42kDa).

Table 3.5. Comparison of Molecular Weight of ten RsbR proteins

Locus Tag	Predicted MW (kDa)	Measured MW (kDa)
SGRA_0571	32.44	30.00
SGRA_1293	33.11	32.20
SGRA_2160	32.68	30.90
SGRA_2161	32.22	27.50
SGRA_2162	33.84	31.50
SGRA_2167	32.35	29.50
SGRA_2168	33.31	31.90
SGRA_2169	32.08	28.30
SGRA_3210	33.30	31.30
SGRA_3852	33.50	67.90*

(*): maltose binding protein (42 kDa) was included

3.2.3. Absorption spectra of 10 purified recombinant RsbRs

To determine whether ten RsbR proteins contain a heme-containing sensor globin domain, absorption spectra of the purified RsbR proteins were measured in 50 mM sodium phosphate buffer. The N-temrinal globin domain of globin-coupled sensors exhibits the oxygen bound myoglobin-like optical property in the near ultravioltet and visible regions. The maximum peaks are occurred at around 410 nm (Soret), 580 nm (α-band) and 540 nm (β -band) (Hou *et al*., 2000 & 2001). Results of absorption spectra were divided into two groups depending on occurrences of absorption maxima. SGRA_0571 (RsbR2), SGRA_3210 (RsbR1) and SGRA_3852 (RsbR3) were categorized in the first group. Their absorption spectra are shown on Figure 3.42 (a,c,e). All three proteins displayed the soret peak at 410 nm

(RsbR2 and 3) or 415 nm (RsbR1). RsbR1 exhibited the α-band at 580 nm and the β- band at 540 nm. The only β- band (535 nm) was detected in RsbR2 and RsbR3. To confirm oxygen binding ability of the heme proteins, a few grains of sodium dithionite, the reducing agent were added in the protein solutions at room temperature. If a heme protein is deoxygenated, the soret peak is shifted to slightly higher wavelength and the α- and β- bands are converged (Hou *et al*., 2000). Upon adding sodium dithionite, the soret peak of RsbR1 shifted to 430 nm and the α- and β- bands were converged to a wide peak at 555 nm. RsbR2 and -3 displayed alterations of the displayed peaks even though they lacked the α- band (Figure 3.42b). The soret peak of RsbR2 shifted to 430 nm and the β- band appeared as a broad peak at 555 nm (Figure 3.42d). RsbR3 displayed the soret peak at 425 nm and another peak at 560 nm (Figure 3.42f). The rest of the proteins were categorized in another group due to lack of the soret peak, the α- and β- bands. The spectra did not show any peak at 400 nm region and beyond. The peaks were only appeared at ~230 nm and ~280 nm regions (Figure 3.43a-g).

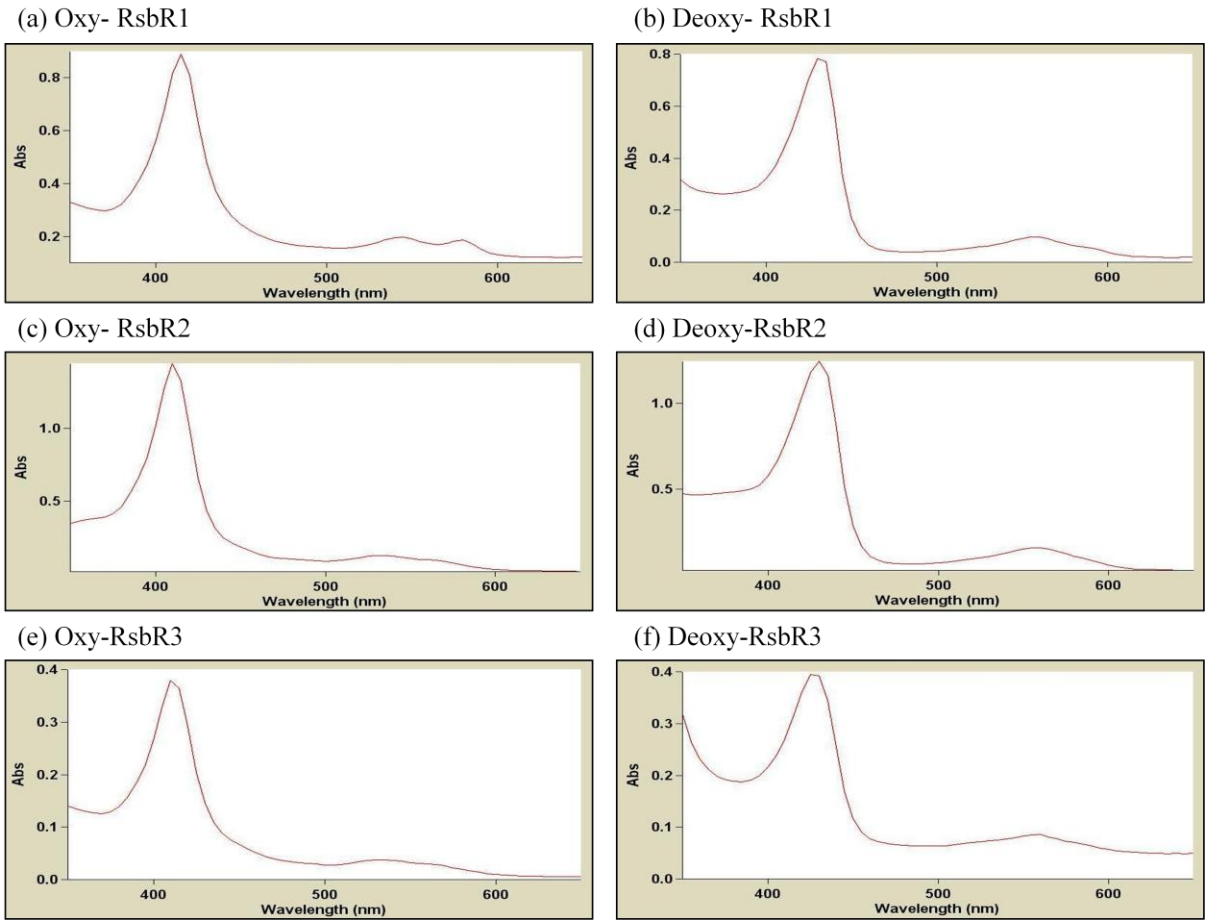
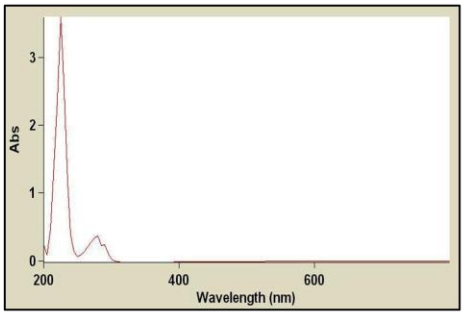
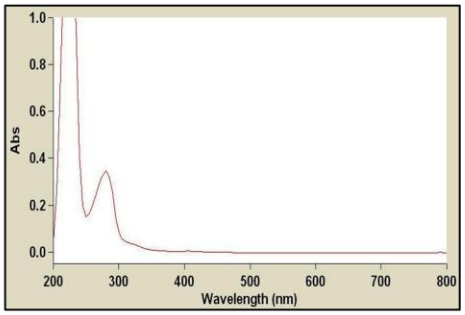


Figure 3.42. Absorption spectra of oxygenated, deoxygenated three his-tagged recombinant RsbRs. (a) oxygenated and (b) deoxygenated RsbR1 (SGRA_3210), (c) oxygenated and (d) deoxygenated RsbR2 (SGRA_0571) (e) oxygenated and (f) deoxygenated RsbR3 (SGRA_3852). Spectra were measured in 300 mM NaCl, 50 mM NaH_2PO_4, pH8.0.

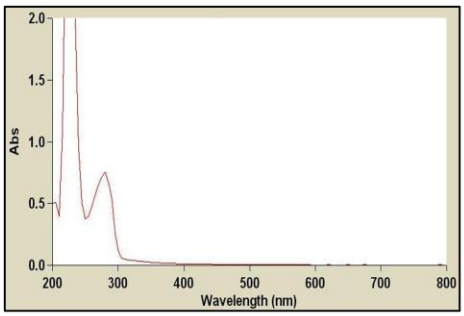
(a) SGRA_1293



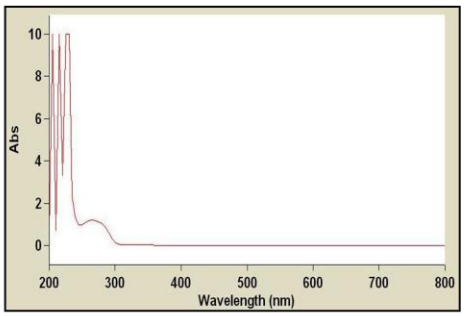
(b) SGRA_2160



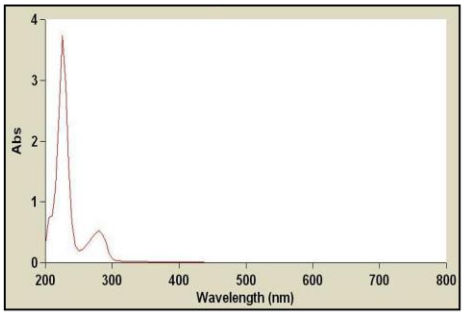
(c) SGRA_2161



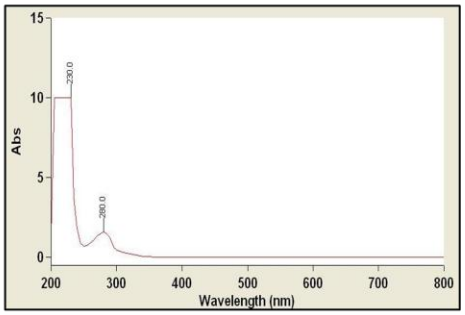
(d) SGRA_2162



(e) SGRA_2167



(f) SGRA_2168



(g) SGRA_2169

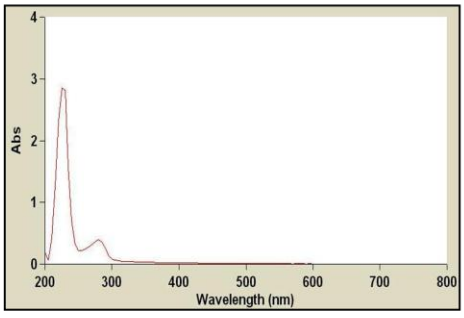


Figure 3.43. Absorption spectra of the seven his-tagged recombinant RsbRs. (a) SGRA_1293 (b) SGRA_2160 (c) SGRA_2161. (d) SGRA_2162. (e) SGRA_2167. (f) SGRA_2168. (g) SGRA_2169. Spectra were measured in 300 mM NaCl, 50 mM NaH_2PO_4, pH8.0.

3.2.4. Initial crystallization of RsbR1 (SGRA_3210)

Initial crystallization of RsbR1 was done by Dr. Farrukh Jamil at Universiti Sain Malaysia. Vapor diffusion method was selected and commercially available kits were used for screening. Based on preliminary trials, MPD based kit (Hampton research) was used for further screening and optimization. Some needle like crystals of RsbR1 was appeared in a solution with 20% PEG 3350, 0.1 M citric acid, 0.8 M ammonium sulphate and 0.1 M KCl at pH4.5 (Figure 3.44). This condition was optimized by changing pH value and the concentration of a constituent. It was observed that 0.1 M citric acid buffer at pH4.5 with 0.7 M ammonium sulphate and 0.1 M KCl was the best condition and was able to produce some thick and rod clustered crystals (Figure 3.45a, b). For further optimization, this solution was used for additive screening. The additives that used on this screening produced some crystals (Figure 3.45c, d). It showed that PEG 3350 and PEG 4000 optimized RsbR1 crystallization in the presence of some salts so far (Figure 3.46). However, a diffraction pattern of the crystals was not observed (Figure 3.47).

Figure 3.44. Needle like RsbR1 crystals in 0.1 M citric acid buffer at pH4.5 with 20% PEG 3350, 0.8 M ammonium sulphate and 0.1 M KCl.

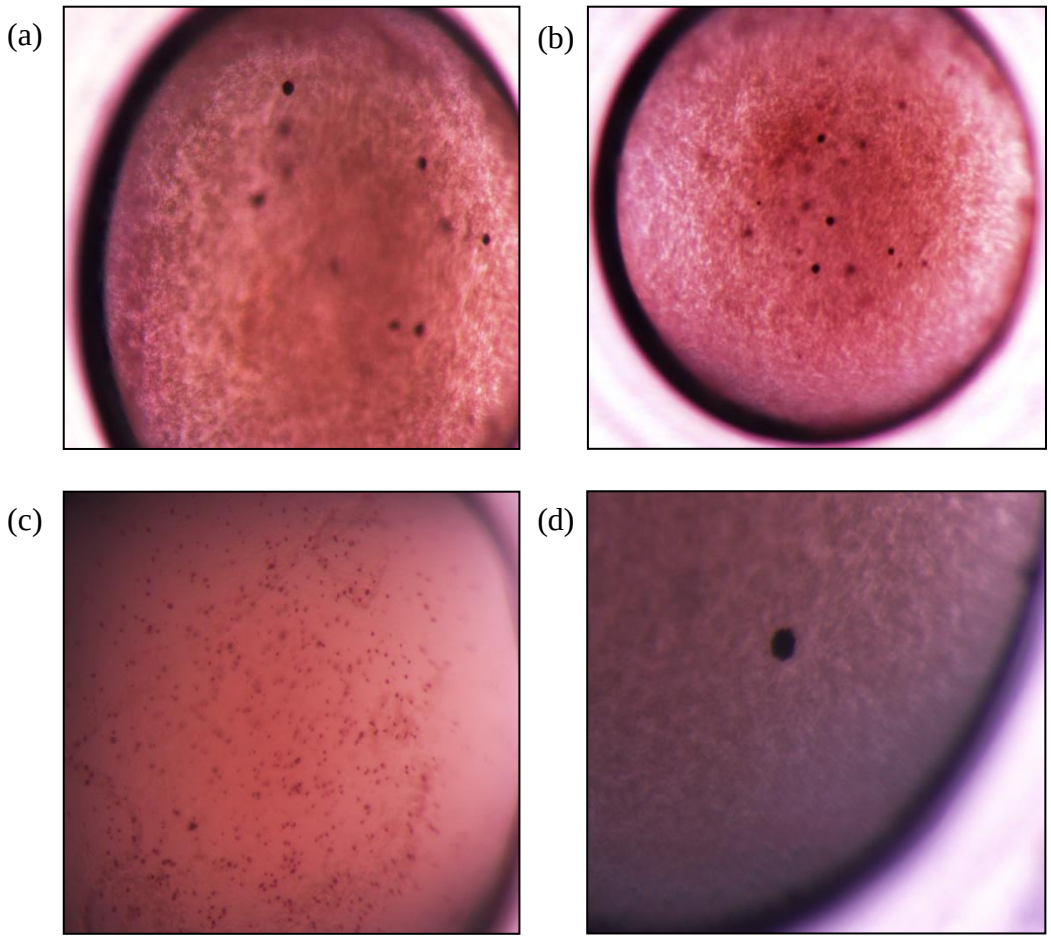


Figure 3.45. Optimized RsbR1 crystals (a) in 0.1 M citric acid buffer at pH4.5 with 20% PEG 3350, 0.7 M ammonium sulphate and 0.01 M $CdCl_2$. (b) in 0.1 M citric acid buffer at pH4.5 with 20% PEG 3350, 0.7 M ammonium sulphate and 0.01 M $MgCl_2$. (c) in 0.1 M citric acid buffer at pH4.5 with 20% PEG 3350, 0.7 M ammonium sulphate and 0. 1 M NaI (d) in 0.1 M citric acid buffer at pH4.5 with 20% PEG 3350, 0.7 M ammonium sulphate and 0.1 M Guanidine hydrochloride.

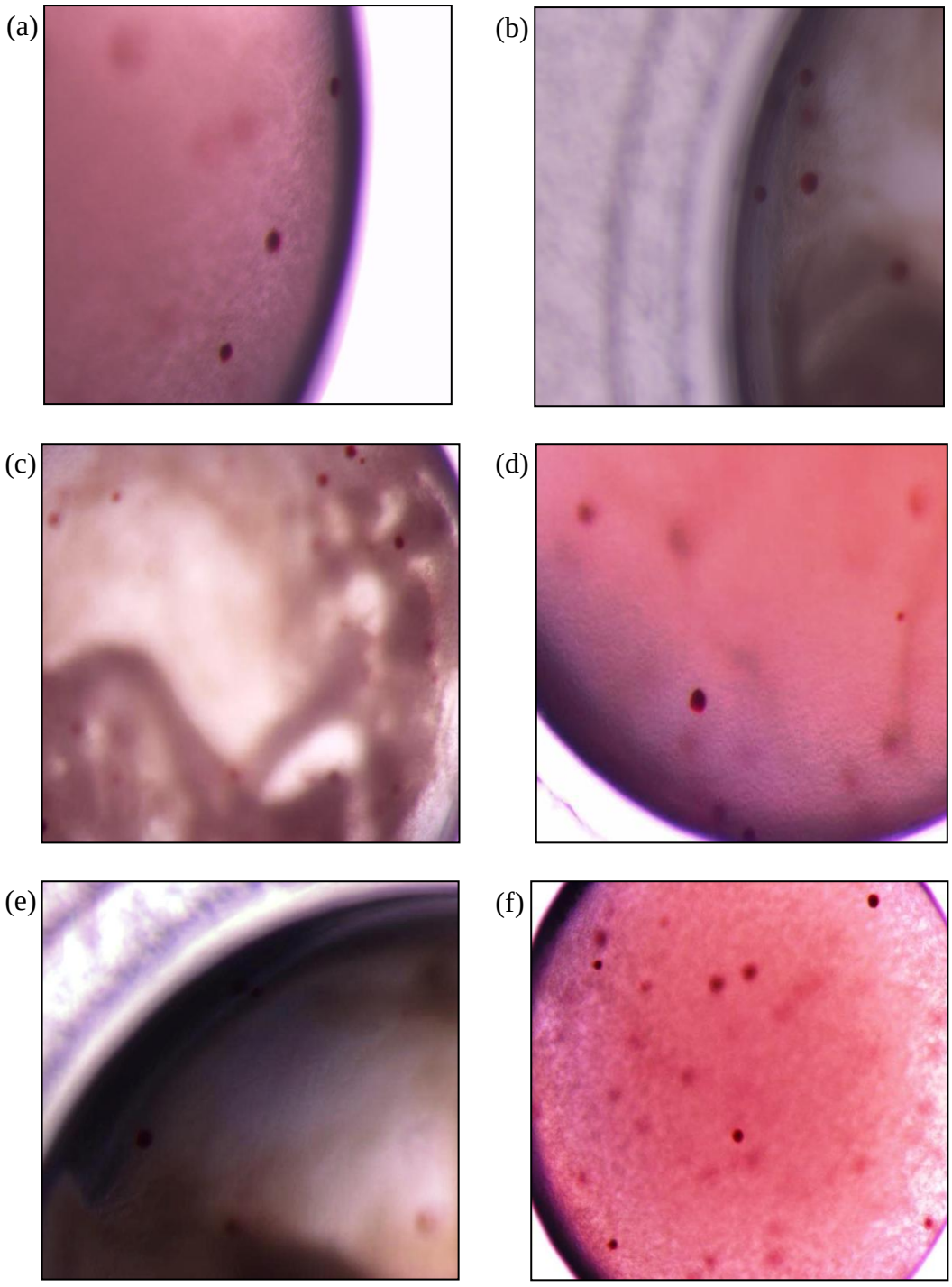


Figure 3.46. Optimization of RsbR1 crystallization with different salts (a) 0.2 M $MgCl_2$ in 20 % PEG 3350, pH5.9 (b) 0.2 M $CaCl_2$ in 20 % PEG 3350, pH5.1 (c) 0.2 M Ammonium chloride in 20 % PEG 3350, pH6.3 (d) 0.2 M Ammonium iodide in 20 % PEG 3350, pH6.2 (e) 0.2 M Magnesium formate dihydrate in 20 % PEG 3350, pH7.0 (f) 0.2 M Ammonium acetate in 20 % PEG 3350, pH7.1.

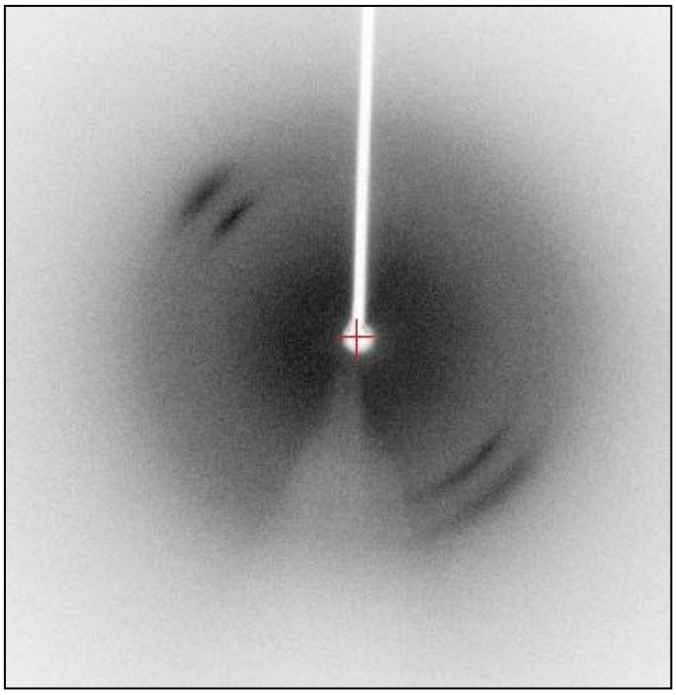


Figure 3.47. The typical X-ray diffraction pattern of RsbR1 crystals from Figure 3.46.

CHAPTER 4

DISCUSSION

Globin-coupled sensors are multiple-domain proteins which consist of an N-terminal sensor globin and a C-terminal transmitter domain. Previously, we discovered Globin-coupled sensors and these proteins which were categorized in the aerotactic group and the 2nd messenger goup within the gene regulation group as the functional classification of GCSs were well characterized in physical and biochemical aspects. The completion of genome sequencing of *S. grandis* revealed presence of GCSs that consisted of sensor_globin domains and STAS domains within the genome based on annotation. This finding would lead us to place them in a new member of the protein-protein interaction group (Figure 1.4). In addition, the predicted functions of them were anti-anti sigma regulatory factors which played important roles in nutritional and environmental stress responses of bacteria. Moreover, the finding of an operon orientation of SGRA_3210 that was similar to the *sigmaB* operon in *B. subtilis* supported our speculation that the encoded proteins of SGRA_3210 might serve as a stress response protein like RsbR in *B. subtilis* including nine other genes.

Fluctuations of environmental and nutritional conditions are life threatening stresses for bacteria. In order to survive, they quickly switched a normal mode of gene expressions to a stress response mode upon stress induction. One way of responding stressful situations is through a stressosome complex and its signaling integration to sigma factor B for the activation against environmental stresses. The activated sigma factor B is bound to the core RNA polymerase that transcribes more than 150 general stress response genes. *S. grandis* has most of genes involved in this response including all genes that form stressosome. However, it does not have *rsbW* and *sigB* genes which suggest that sigma factor B dependent stress responses do not apply to *S. grandis* instead it may present alternative pathways to response

environmental stresses via signal integration of stressosome.

So far, none of GCSs belonging to the protein-protein interaction group have been characterized. In this study, we have characterized ten putative RsbR-like GCSs by approaching *in silico* and *in vitro* methods. Web-based bioinformatics analyses predicted the N-terminal domains of the RsbR-like GCSs as the globin-like domain and the C-terminal STAS domain. There were some different results in terms of defined regions of each domain in the proteins because SBASE, NCBI and InterProScan used different analysis tools to screen given protein sequences. CDD from NCBI did not recognize any homologous domain in the N-terminal regions of SGRA_2162 and SGRA_2168. In contrast, there were hits for globin-like domains in the N-terminal regions of these two proteins from the SBASE and the InterProScan analyses. Based on the results, the defined globin domains and STAS domains of the ten protein sequences were aligned with functionally determined globin and STAS domains. We have found that only three proteins named as RsbR1, RsbR2 and RsbR3 have the conserved histidine residues which were lined up with the proximal histidine residues in three known globin domains and rest of the proteins are substituted with different amino acids. When we selected these proteins for 3D structure alignment, they got hits on the globin structure of the GCS in *Geobacter sulffureducens* and showed in 3D structure alignments with protein sequence alignments. Importantly, the 3D view displayed the histidine residue (the proximal histidine) which held the heme molecule and this residue was aligned with the histidine residues in RsbR1, RsbR2 and RsbR3. This result and multiple alignments of globin protein sequences support that only these proteins may have the heme binding property as globin-couple sensors that are involved in stress responses. The STAS domains of all ten proteins had highly conserved sequences like other STAS domains. Also, multiple alignments of STAS protein sequences showed conserved serine or threonine residue that are known to

be phosphorylated by RsbT (serine/threonine kinase) in RsbR and its paralogs (YkoB, YojH, YqhA) from *B. subtilis* except SGRA_3852 (RsbR3). YtvA is another paralog in *B. subtilis* but it is not phosphorylated. However, it does contain a sensor domain (PAS) which is known to sense light, oxygen and voltage (Akbar *et al*., 2001). Experimentally, it is shown that YtvA is able to detect blue-light and integrates signals to the sigma factor in addition to GTP binding through the GTP binding motif (DLSG) (Butani *et al*., 2006). RsbR3 also showed a similar motif as DIKG at the same region (Figure 3.4). Furthermore, RsbR1, RsbR2 and other seven proteins have conserved amino acids, aspartate and glycine on this region. Serine residues are mostly conserved as well. Taken together, RsbR1 and RsbR2 may sense gaseous ligands such as oxygen and be phosphorylated by RsbT with accompanying GTP binding. RsbR3 does not have phosphorylation sites but it may bind gaseous ligands following with binding of GTP and transmit signals to downstream of signal cascades like YtvA. Based on the *in silico* results, signal cascades of RsbR1 was depicted on Figure 4.1. Both RsbR2 and RsbR3 may play similar role.

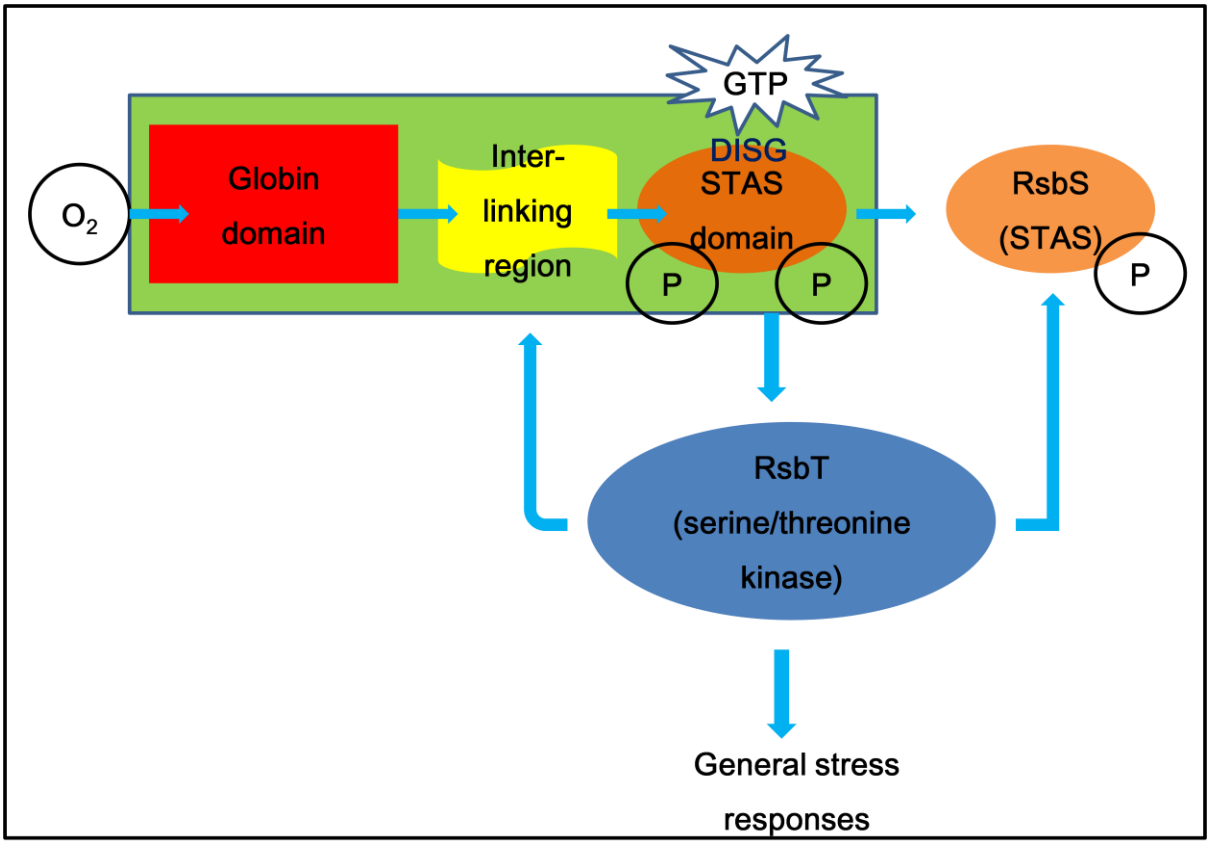


Figure 4.1. Model of signal cascades of RsbR1.

All ten putative RsbR-like proteins have been cloned, expressed and purified. The purified proteins were examined on absorption spectra. Expression and purification of RsbR1, RsbR2 and RsbR3 were more challenge than seven other proteins. Purified recombinant protein, RsbR1 was unstable and was not able to get acceptable purity for further experiment when it was purified by the same condition for other proteins. To improve stability of the protein, several additives were added in purification buffers. We found that adding 1% glucose in the lysis buffer at pH7.6 stabilized RsbR1 protein and achieved about 90% purity. However, we have not understood how glucose actually helped for stabilization of RsbR1.

Oxygenated myoglobin exhibited a distinctive spectrum property due to heme-binding protein conformation. The highest peak at ~410 nm (soret peak) and two small peaks at 580 nm (alpha band) and at 540 nm (beta band) were observed. RsbR1, RsbR2 and RsbR3 displayed similar spectrum results which indicated their heme binding property. Moreover,

treatment with sodium dithionite changed their spectrum patterns which were similar to deoxygenated myoglobin. This suggests that their gaseous ligand such as oxygen in the heme molecules were released. The gaseous ligand for them is the most likely oxygen, based on similar absorption spectrum results. RsbR2 and RsbR3 did not display any peak at expected alpha band. However, their maxima peaks displayed on purified proteins were altered after adding sodium dithionite. This may be because RsbR2 and RsbR3 are able to bind the heme but they might be mis-folded during expression or their structure conformation might be changed by an unknown factor.

The domain structure of RsbR in *B. subtilis* displayed the non-heme binding N-terminal globin domain and the C-terminal STAS domain. The N-termial domain globin domain does not contain the heme molecule but exhibited globin-fold. Based on this, RsbR protein may serve as a sensor, however, its ligand has not been identified yet. In contrast, RsbR1, RsbR2 and RsbR3 from *S. grandis* predicted as the N-terminal globin domains with the heme binding property and it was experimentally validated. These three proteins sense oxygen as a signal and the N-terminal globins are the sensor globin domains. As we don’t have any physiological validation of their role, we speculate that *S. grandis* is an obligate aerobe and the oxygen level for this organism will be vital so it may adapt GCS molecules as sensors and transmit information to downstream in preparation for oxygen depletion.

CHAPTER 5

CONCLUSION

Ten RsbR-like proteins from *S. grandis* were characterized via *in silico* and *in vitro* methods. Both analyses showed SGRA_0571 (RsbR2), SGRA_3210 (RsbR1) and SGRA_3852 (RsbR3) possessed the heme molecule as a prosthetic group and bound oxygen. For *in silico* analysis result, we also found possible phosphorylation and GTP binding properties for all proteins except SGRA_3852. Based on the initial annotation results, we postulated all ten proteins may be GCSs and serve as stress response proteins. However, RsbR1, RsbR2 and RsbR3 are validated as GCSs and rest of proteins that do not bind the heme are not. In this study, we only characterized oxygen binding property of the three GCSs. To understand the actual signaling pathway of *S. grandis*'s stress responses related to these proteins, we will continue to study the proteins at the gene expression level and at the physiological level. Also, crystallization of RsbR1 will be further optimized and structure of RsbR1 will be determined. This will give us a better understanding of signal integrations of RsbR1 upon ligand binding.

REFERENCES

1. Aizawa, S. I. (2005). Bacterial Gliding Motility: Visualizing Invisible Machinery. *ASM News*, 71, 71-75.

2. Akbar, S., Kang, C. M., Gaidenko, T. A., & Price, C. W. (1997). Modulator protein RsbR regulates environmental signalling in the general stress pathway of *Bacillus subtilis*. *Molecular Microbiology*. 24(3), 567-578.

3. Akbar, S., Gaidenko, T. A., Kang, C. M., O'Reilly, M., Devine, K. M., & Price, C. W. (2001). New Family of Regulators in the Environmental Signaling Pathway Which Activates the General Stress Transcription Factor {sigma}B of Bacillus subtilis. *J. Bacteriol.*, 183(4), 1329-1338.

4. Alexandre, G., & Zhulin, I. B. (2001). More Than One Way To Sense Chemicals. *J. Bacteriol.*, 183(16), 4681-4686.

5. Appleby, C. (1984). Leghemoglobin and *Rhizobium* respiration. Annual *Review of Plant Physiology*, 35, 443-478.

6. Aravind, L., & Koonin, E. V. (2000). The STAS domain -- a link between anion transporters and antisigma-factor antagonists. *Current Biology*, 10(2), R53-R55.

7. Burmester, T., Weich, B., Reinhardt, S. & Hankeln, T. (2000). A vertebrate globin expressed in the brain. *Nature*. 407(6803), 520-523.

8. Buttani, V., Losi, A., Polverini, E., & Gärtner, W. (2006). Blue news: NTP binding properties of the blue-light sensitive YtvA protein from Bacillus subtilis. *FEBS Letters*. 580(16), 3818-3822.

9. Chan, M. K. (2000). Heme protein biosensors. *Journal of Porphyrins and Phthalocyanines*, 4(4), 358-361.

10. Chan, M. K. (2001). Recent advances in heme-protein sensors. *Current Opinion in Chemical Biology*, 5(2), 216-222.

11. Chang, A. L., Tuckerman, J. R., Gonzalez, G., Mayer, R., Weinhouse, H., Volman, G., Amikam, D. (2011). Phosphodiesterase A1, a Regulator of Cellulose Synthesis in Acetobacter xylinum, Is a Heme-Based Sensor†. *Biochemistry*, 40(12), 3420-3426

12. Cole, J. R., Cascarelli, A. L., Mohn, W. W., & Tiedje, J. M. (1994). Isolation and characterization of a novel bacterium growing via reductive dehalogenation of 2-chlorophenol. *Appl. Environ. Microbiol.*, 60(10), 3536-3542.

13. Correll, D. L., Lewin, R. A. (1964). Rod-shaped ribonucleoprotein particles from Saprospira. *Canadian Journal of Microbiology*. 10(1), 63-74.

14. David, M., Daveran, M.-L., Batut, J., Dedieu, A., Domergue, O., Ghai, J., Hertig, C. (1988). Cascade regulation of nif gene expression in Rhizobium meliloti. *Cell*, 54(5), 671-683.

15. Delgado-Nixon, V. M., Gonzalez, G., & Gilles-Gonzalez, M.-A. (2011). Dos, a Heme-Binding PAS Protein from Escherichia coli, Is a Direct Oxygen Sensor. *Biochemistry*, 39(10), 2685-2691.

16. Delk, A.S., Dekker, C.A. (1972). Characterization of rhapidosomes of *Saprospira grandis*. *Journal of Molecular Biology*. 64, 287-288.

17. Delumeau, O., Chen, C.C., Murray, J. W., Yudkin, M. D., & Lewis, R. J. (2006). High-Molecular-Weight Complexes of RsbR and Paralogues in the Environmental Signaling Pathway of *Bacillus subtilis*. *J. Bacteriol.* 188(22), 7885-7892.

18. Desmet, F., Thijs, L., El Mkami, H., Dewilde, S., Moens, L., Smith, G., & Van Doorslaer, S. (2010). The heme pocket of the globin domain of the globin-coupled sensor of *Geobacter sulfurreducens* -An EPR study. *Journal of Inorganic Biochemistry*, 104(10), 1022-1028.

19. Dietrich, W. E., & Biggins, J. (1971). Respiratory Mechanisms in the Flexibacteriaceae: Terminal Oxidase Systems of Saprospira grandis and Vitreoscilla Species. *J. Bacteriol.*, 105(3)

20. Dioum, E. M., Rutter, J., Tuckerman, J. R., Gonzalez, G., Gilles-Gonzalez, M.-A., & McKnight, S. L. (2002). NPAS2: a gas-responsive transcription factor. *Science (New York, N.Y.)*, 298(5602), 2385-2387.

21. Freitas, T. A. K., Hou, S., & Alam, M. (2003). The diversity of globin-coupled sensors. *FEBS Letters*, 552(2-3), 99-104.

22. Freitas, T. A., Saito, J.A., Wan, X., Hou, S. & Alam, M. (2008). Protoglobin and Globin-coupled Sensors. *In:* GHOSH, A. (ed.) *The Smallest Biomolecules Diatomics and their Interactions with Heme Proteins.* First Edition ed. Oxford: Elsevier.

23. Gekakis, N., Staknis, D., Nguyen, H. B., Davis, F. C., Wilsbacher, L. D., King, D. P., Takahashi, J. S. (1998). Role of the CLOCK Protein in the Mammalian Circadian Mechanism. *Science*, 280(5369), 1564 -1569.

24. Gilles-Gonzalez, M. A., & Gonzalez, G. (1993). Regulation of the kinase activity of heme protein FixL from the two-component system FixL/FixJ of Rhizobium meliloti. *Journal of Biological Chemistry*, 268(22), 16293 -16297.

25. Gilles-Gonzalez, M.-A., & Gonzalez, G. (2005). Heme-based sensors: defining characteristics, recent developments, and regulatory hypotheses. *Journal of Inorganic Biochemistry*, 99(1), 1-22.

26. Gilles-Gonzalez, M.A. and Gonzalez, G. (2007). A surfeit of biological heme-based

sensors. In Ghosh, A. (Ed.) The smallest biomolecules: diatomics and their interactions with heme proteins. pp. 18-65.

27. Goldoni, A. (2001-2002). Porphyrins: Fascinating Molecules with Biological Significance, Elettra highlights, Trieste. 64-66.

28. Hardison, R. (1999). "The Evolution of Hemoglobin: Studies of a very ancient protein suggest that changes in gene regulation are an important part of the evolutionary story". *American Scientist* 87(2), 126.

29. He, Q. & Sanford, R. A. (2002). Induction characteristics of reductive dehalogenation in the ortho-halophenol-respiring bacterium, Anaeromyxobacter dehalogenans *Biodegradation,* 13,307-16.

30. He, Q., & Sanford, R. A. (2003). Characterization of Fe(III) Reduction by Chlororespiring Anaeromxyobacter dehalogenans. *Appl. Environ. Microbiol.*, 69(5), 2712-2718.

31. Hecker, M., Pané-Farré, J., & Uwe, V. (2007). SigB-Dependent General Stress Response in Bacillus subtilis and Related Gram-Positive Bacteria. *Annual Review of Microbiology*, 61(1), 215-236.

32. Hou, S., Larsen, R. W., Boudko, D., Riley, C. W., Karatan, E., Zimmer, M., Ordal, G. W. (2000). Myoglobin-like aerotaxis transducers in Archaea and Bacteria. *Nature*, 403(6769), 540-544.

33. Hou, S., Belisle, C., Lam, S., Piatibratov, M., Sivozhelezov, V., Takami, H., & Alam, M. (2001). A globin-coupled oxygen sensor from the facultatively alkaliphilic Bacillus halodurans C-125. *Extremophiles*, 5(5), 351-354.

34. Hou, S., Freitas, T., Larsen, R. W., Piatibratov, M., Sivozhelezov, V., Yamamoto, A., Meleshkevitch, E. A. (2001). Globin-coupled sensors: A class of heme-containing sensors in Archaea and Bacteria. *PNAS*, 98(16), 9353-9358.

35. Hogenesch, J. B., Gu, Y.-Z., Jain, S., & Bradfield, C. A. (1998). The basic-helix–loop–helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors. *Proceedings of the National Academy of Sciences*, 95(10), 5474 -5479.

36. Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P. (2009). InterPro: the integrative protein signature database. *Nucleic Acids Research*, 37(Database), D211-D215.

37. Kjeldgaard, M., Nyborg, J., & Clark, B. (1996). The GTP binding motif: variations on a theme. *The FASEB Journal*, 10(12), 1347 -1368.

38. Kim, T. J., Gaidenko, T. A., & Price, C. W. (2004). In Vivo Phosphorylation of Partner Switching Regulators Correlates with Stress Transmission in the

Environmental Signaling Pathway of *Bacillus subtilis*. *J. Bacteriol.* 186(18): 6124-6132.

39. Kitanishi, K., Kobayashi, K., Kawamura, Y., Ishigami, I., Ogura, T., Nakajima, K., Igarashi, J. (2010). Important Roles of Tyr43 at the Putative Heme Distal Side in the Oxygen Recognition and Stability of the Fe(II)−O2 Complex of YddV, a Globin-Coupled Heme-Based Oxygen Sensor Diguanylate Cyclase. *Biochemistry*, 49(49), 10381-10393.

40. Kosmachevskaya, O. V., & Topunov, A. F. (2009). Hemoglobins: Diversity of structures and functions. *Applied Biochemistry and Microbiology*, 45(6), 563-587.

41. Lanzilotta, W. N., Schuller, D. J., Thorsteinsson, M. V., Kerby, R. L., Roberts, G. P., & Poulos, T. L. (2000). Structure of the CO sensing transcription activator CooA. *Nat Struct Mol Biol*, 7(10), 876-880.

42. Lecomte, J. T., Vuletich, D. A., & Lesk, A. M. (2005). Structural divergence and distant relationships in proteins: evolution of the globins. *Current Opinion in Structural Biology*, 15(3), 290-301.

43. Lewin, R.A. (1997). *Saprospira grandis*: A Flexibacterium That Can Catch Bacterial Prey by ``Ixotrophy". *Microbial Ecology*. 34: 232-236.

44. Loffler, F. E., Tiedje, J. M., & Sanford, R. A. (1999). Fraction of Electrons Consumed in Electron Acceptor Reduction and Hydrogen Thresholds as Indicators of Halorespiratory Physiology. *Appl. Environ. Microbiol.*, *65*(9), 4049-4056.

45. Lu, C., Egawa, T., Batabyal, D., Mukai, M., & Yeh, S. R. (2008). Mocrobial Hemoglobins: Structure, Function, and Folding. *In:* GHOSH, A. (ed.) *The Smallest Biomolecules Diatomics and their Interactions with Heme Proteins.* First Edition ed. Oxford: Elsevier.

46. Marles-Wright, J., & Lewis, R. J. (2008). The *Bacillus subtilis* stressosome. *Commun Integr Biol*. 1(2), 182-184.

47. Martinez, L., Reeves, A., & Haldenwang, W. (2010). Stressosomes Formed in *Bacillus subtilis* from the RsbR Protein of Listeria monocytogenes Allow {sigma}B Activation following Exposure to either Physical or Nutritional Stress. *J. Bacteriol.*, 192(23), 6279-6286.

48. McBride, M. J. (2001). Multiple Mechanisms for Cell Movement over Surfaces. *Annual Review of Microbiology*. 55(1), 49-75.

49. Milani, M., Pesce, A., Nardini, M., Ouellet, H., Ouellet, Y., Dewilde, S., Bocedi, A. (2005). Structural bases for heme binding and diatomic ligand recognition in truncated hemoglobins. *Journal of Inorganic Biochemistry*, 99(1), 97-109.

50. Mincer, T. J., Spyere, A., Jensen, P. R., & Fenical, W. (2004). Phylogenetic Analyses and Diterpenoid Production by Marine Bacteria of the Genus Saprospira. *Current Microbiology*, 49(4), 300-307.

51. National Center for Biotechnology Information (n.d.). NCBI: Structure, retrieved December 4, 2011, from NCBI Web site: http://www.ncbi.nlm.nih.gov/sites/entrez?db=structure.

52. Pané-Farré, J., Lewis, R.J. & Stülke, J. (2005). The RsbRST Stress Module in Bacteria: A Signalling System That May Interact with Different Output Modules. *J Mol Microbiol Biotechnol. 9:65-76.*

53. Peterson, E. S., Huang, S., Wang, J., Miller, L. M., Vidugiris, G., Kloek, A. P., Goldberg, D. E.. (1997). A Comparison of Functional and Structural Consequences of the Tyrosine B10 and Glutamine E7 Motifs in Two Invertebrate Hemoglobins (Ascaris suum and Lucina pectinata)†. *Biochemistry*, 36(42), 13110-13121

54. Reick, M., Garcia, J. A., Dudley, C., & McKnight, S. L. (2001). NPAS2: an analog of clock operative in the mammalian forebrain. *Science (New York, N.Y.)*, 293(5529), 506-509.

55. Rodger, K.R. (1999). Heme-based sensors in biological systems. *Current Opinion in Chemical Biology*, 3(2), 158-167.

56. Saito, J. A., Wan, X., Lee, K. S., Hou, S., & Alam, M. (2008). Globin-coupled sensors and protoglobins share a common signaling mechanism. *FEBS Letters*, 582(13), 1840-1846.

57. Saito, J. A., Freitas, T. A. K., & Alam, M. (2008). Cloning, Expression, and Purification of the N-terminal Heme-Binding Domain of Globin-Coupled Sensors. *Globins and Other Nitric Oxide-Reactive Proteins, Part B* (Vol. 437, pp 163-172). Academic Press.

58. Sangkhobol, V., Skerman, V.B. D. (1981). Saprospira species—Natural predators. *Current Microbiology*. 5(3), 169-174.

59. Sasakura, Y., Hirata, S., Sugiyama, S., Suzuki, S., Taguchi, S., Watanabe, M., Matsui, T. (2002). Characterization of a Direct Oxygen Sensor Heme Protein from *Escherichia coli*. *Journal of Biological Chemistry*, 277(26), 23821 -23827.

60. Sasakura, Y., Yoshimura-Suzuki, T., Kurokawa, H., & Shimizu, T. (2006). Structure−Function Relationships of *Ec*DOS, a Heme-Regulated Phosphodiesterase from *Escherichia coli*. *Acc. Chem. Res.*, 39(1), 37-43.

61. Sanford, R. A., Cole, J. R., & Tiedje, J. M. (2002). Characterization and Description of Anaeromyxobacter dehalogenans gen. nov., sp. nov., an Aryl-Halorespiring Facultative Anaerobic Myxobacterium. *Appl. Environ. Microbiol.*, 68(2), 893-900.

62. Thijs, L., Vinck, E., Bolli, A., Trandafir, F., Wan, X., Hoogewijs, D., Coletta, M. (2007). Characterization of a Globin-coupled Oxygen Sensor with a Gene-regulating Function. *Journal of Biological Chemistry*, 282(52), 37325 -37340.

63. Vinogradov, S. N., Hoogewijs, D., Bailly, X., Mizuguchi, K., Dewilde, S., Moens, L., & Vanfleteren, J. R. (2007). A model of globin evolution. *Gene*, 398(1-2), 132-142.

64. Vlahovicek, K., Kajan, L., Agoston, V., & Pongor, S. (2005). The SBASE domain sequence resource, release 12: prediction of protein domain-architecture using support vector machines. *Nucleic Acids Research*, 33, D223-D225.

65. Wajcman, H., Kiger, L., & Marden, M. C. (2009). Structure and function evolution in the superfamily of globins. *Comptes Rendus Biologies*. 332(2-3): 273-282.

66. Walsh, G. (2002). *Proteins: biochemistry and biotechnology*. John Wiley and Sons.

67. Wan, X., Tuckerman, J. R., Saito, J. A., Freitas, T. A. K., Newhouse, J. S., Denery, J. R., Galperin, M. Y. (2009). Globins Synthesize the Second Messenger Bis-(3'-5')-Cyclic Diguanosine Monophosphate in Bacteria. *Journal of Molecular Biology*, *388*(2), 262-270.

68. Wang, Y., Addess, K. J., Chen, J., Geer, L. Y., He, J., He, S., Lu, S. (2007). MMDB: annotating protein sequences with Entrez's 3D-structure database. *Nucleic Acids Research*, *35*(Database), D298-D300.

69. Wilkinson, D. L., & Harrison, R. G. (1991). Predicting the Solubility of Recombinant Proteins in Escherichia coli. *Nat Biotech*, 9(5), 443-448.

70. Zhao, Y., Brandish, P. E., Ballou, D. P., & Marletta, M. A. (1999). A molecular basis for nitric oxide sensing by soluble guanylate cyclase. *Proceedings of the National Academy of Sciences*, 96(26), 14753 -14758.