Structural and functional analysis of a sporulation protein Spo0M from *Bacillus subtilis*.

Yo Sonoda

2016

TABLE OF CONTENTS

ABBREVIATIONS

RNA: ribonucleotide

FBP: fructose-1,6-bisphosphate

DNA: deoxyribonucleotide

ATP: adenosine triphosphate

SPR: surface plasmon resonance

MALDI: Matrix assisted laser desorption/ionization

TOFMS: Time of flight mass spectrometry

Tris: tris(hydroxymethyl)aminomethane

EDTA: ethylenediaminetetraacetic acid

PCR: polymerase chain reaction

IPTG: Isopropyl β-D-1-thiogalactopyranoside

DNase: deoxyribonuclease

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

Hepes: 4-(2-HydroxyEthyl)-1-PiperazineEthaneSulfonic acid

NHS: N-hydroxysuccinimide

EDC: 1-ethyl-3-(3-dimethylaminoproply)carbodiimide

mAb: monoclonal antibody

RU: resonance units

CBB: Coomassie Brilliant Blue

DTT: dithiothreitol

GENERAL INTRODUCTION

Sporulation of Bacillus subtilis

Responses to changes in the external environment are an essential function in the survival of all organisms, and a wide range of strategies is adopted for this purpose. In the bacterial kingdom, the cellular responses to environmental changes include stress responses to heat, organic solvents or high salt concentration, and emergency responses to depletions of particular nutrient sources. The bacterial species *Bacillus subtilis* forms endospores to survive once it senses a nutrient limitation. Endospore formation is one of the ultimate strategies of bacteria to survive environmental assaults that would normally kill the bacterium. Endospores are metabolically inactive, and extremely resistant to heat, toxic chemicals and even radiation [1, 2]. Since *B. subtilis* lends itself to genetic manipulation and is thus a commonly used bacterial model organism, the sporulation mechanism of *B. subtilis* has been well studied [3].

As in Figure 1, the process of endospore formation in *B. subtilis* is complex and highly regulated. Endospore development requires several hours to complete and includes major changes of cellular morphology, which has been used as markers to define development stages; stage 0 - VII. In stage 0, cells undergo normal vegetative growth. Once a *B. subtilis* cell senses a sporulation initiation signal, it stops its normal cell division and divides asymmetrically by forming a septum between the dividing smaller daughter cell and larger mother cell (stage II), resulting the prespore formation [4]. In stage III, the peptidoglycan in the septum is degraded via phagocytic-like process and the forespore is wholly engulfed by the mother cell. In stage IV, a thick layer of peptidoglycan, termed

cortex, is formed between the two forespore membranes. In stage V, more than 10 species of spore coat proteins are synthesized in mother cell to form spore coat. The endospore is matured in stage VI and released in stage VII through programmed cell death of mother cell.

From genetic studies, several dozen of sporulation genes, which are termed *spo* genes, have been identified to control sporulation. In *spo* mutants, the sporulation is interrupted at a specific stage without affecting their vegetative growth [2]. Such *spo* genes were named according to the stage at which the mutation blocked sporulation (eg. *spo0*, *spoII*, *spoIII* etc.). During sporulation, a number of sporulation genes are expressed sequentially and the large portion of those sequential expressions are regulated by the regulatory subunits of RNA polymerase, which are called as σ factors. The σ factors also sequentially appear during spore formation. Five σ factors σ^H, σ^F, σ^E, σ^G and σ^K are known to comprise a regulatory cascade in sporulation. σ^H encoded by *spo0H* gene exists in vegetative deviation stage and the activity is dramatically enhanced when the sporulation is initiated [5]. The expression of *spo0H* gene is regulated by Spo0A protein. Spo0A is a global transcriptional regulator, which is activated by phosphorylation [6]. Phosphorylated Spo0A protein suppresses the expression of *abrB* gene encoding AbrB protein, which is thought to suppress *spo0H* expression [7]. σ^H is reported to induce the expression of sporulation-related genes such as *spo0A*, *spo0F* and *spo0M* [8, 9]. σ^F is encoded by *spoIIAC* gene and induced by σ^H [10]. σ^F induces the expressions of various genes in prespore at the late phase of sptum formation. σ^E is converted from pro-σ^E, which is encoded by *spoIIGB* gene, by the removal of 27 amino acids from the pro-σ^E N terminus and localized in mother cell. The processing is mediated by the sporulation σ^E factor-processing peptidase, which is encoded by *spoIIGA* gene [11]. σ^G is encoded by *spoIIIG*

gene and expressed in forespore. σ^G induces the expression of the genes *sspA*, *sspB*, *sspC*, *sspD*, and *sspE*, which encode a family of small, acid-soluble spore proteins [10]. These proteins are rapidly degraded as a source of amino acids upon spore germination. σ^K is composed of two truncated *spo* genes, *spoIVCB* and *spoIIIC*, which are brought together by site-specific recombination in mother cell during sporulation [12]. σ^K induces the expression of *cotA* and *cotD* genes encoding spore coat proteins. σ^K also induce *gerE* gene expression and the resulting GerE protein induces *cotB* and *cotC* gene expression.

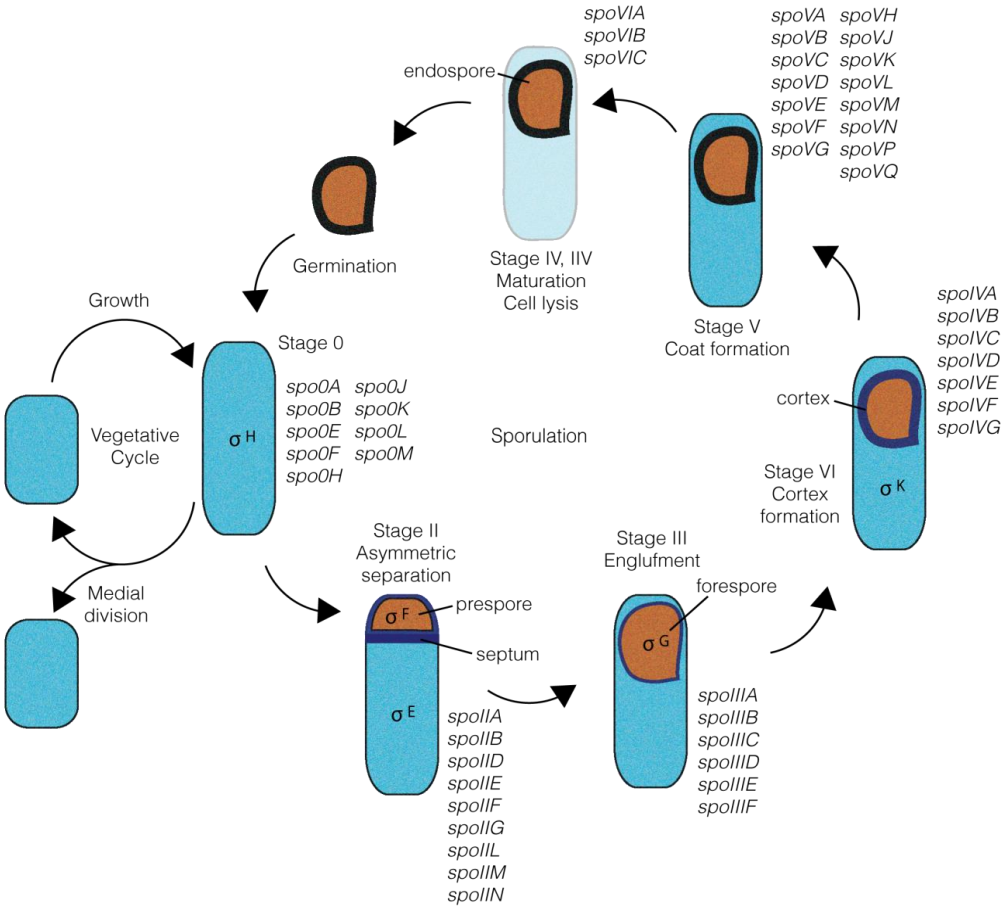

Figure 1 Sporulation cycle of *B. subtilis*. The main sequence of morphological events begins with vegetative cell and ends with the release of a mature endospore. The *spo* genes and σ factors expressed in each sporulation stage are indicated. This figure was modified from Errington (2003) [13]

Carbon catabolite repression in B. subtilis

For many bacteria, glucose is the preferred carbon source, and thus, in the presence of glucose, the expression of genes for the utilization of secondary carbon sources, such as xylose or glycerol, is suppressed by a mechanism called carbon catabolite repression (CCR).[14-16] In *B. subtillis*, the main global regulator for CCR, catabolite control protein A (CcpA), represses the expression of genes by binding to the catabolite-responsive elements (*cre*s). These *cre*s of target genes are highly conserved with the consensus sequence WTGNNARCGNWWWCAW (where W stands for A or T, R for A or G and N for any base).[17] The *cre* element was first reported in the promoter region of *amyE* [18, 19] and thereafter *cre*s were discovered both in the promoter region and open-reading frames of genes in the myo-inositol operon.[20] It has also been reported that regulation of the expression of genes involved in sporulation is at least partially controlled under CCR [21] and that disruption of the *ccpA* gene increases the frequency of sporulation in the presence of glucose,[22] indicating that a portion of genes related to sporulation are controlled by CcpA.

The CCR in *B. subtilis* is illustrated in Figure 2. The phosphoryl group is transferred from phosphoenolpyruvate (PEP) via Enzyme I (EI) to His15 of histidine-containing phosphotransfer protein (HPr). Then the phosphoryl group is transferred from HPr to Enzyme IIA domain ($EIIA^{Glc}$) and is further transferred to the Enzyme IIB domain ($EIIB^{Glc}$), resulting the incorporation of glucose. Glucose molecules are converted into FBP via glycolysis mechanism [14]. The DNA-binding function of CcpA requires the participation of coeffector proteins: the serine-phosphorylated form of HPr, and the catabolite repression HPr-like protein, Crh (HPr-Ser-P and Crh-Ser-P, respectively). Both of the coeffector proteins are activated via phosphorylation by HPr kinase, HPrK.[23] The phosphorylations of these proteins are initiated by high intracellular concentration of FBP and ATP and the resulting

serine-phosphorylated HPr or Crh directly binds to CcpA to form an active complex.[24] The complex binds to *cre* sites of target genes and regulates the gene expression.

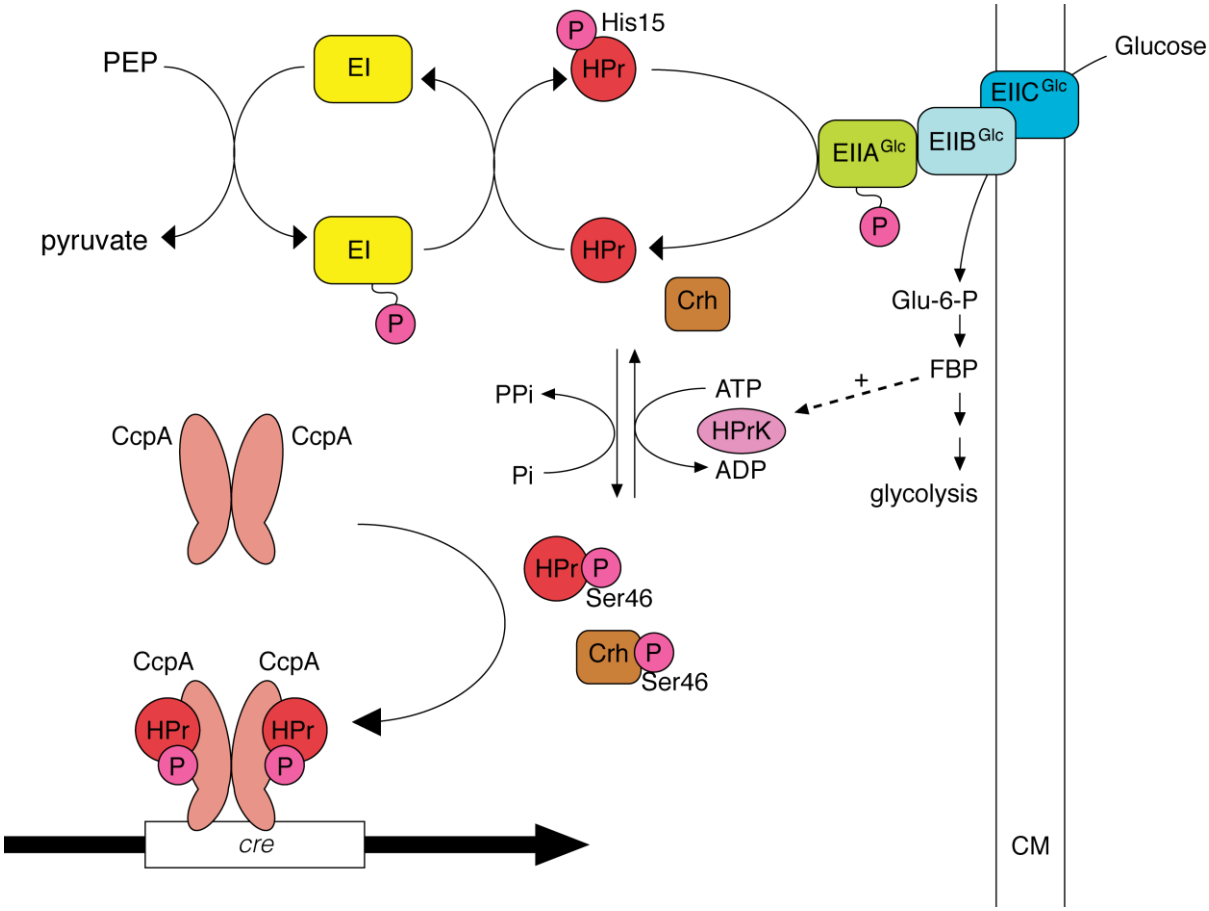

Figure 2 Carbon catabolite repression in *B. subtilis*. Pi: inorganic phosphate. CM: cytoplasmic membrane. This figure was modified from Görke & Stülke (2008). [14]

Spo0M in B. subtilis

Spo0M is one of the sporulation control proteins; it is made up of 258 amino acids and is transcriptionally regulated by σ^H. The *spo0M* gene is not essential for cell viability, but disruption of the *spo0M* gene results in sensitivity to lysis and a defect of the endospore

formation from stage 0 to stage II [8]. The overexpression of the Spo0M protein also negatively affects the sporulation process, indicating that both too little and too much Spo0M protein in the cells will influence the initial step of sporulation [8]. The FtsH protein of *B. subtilis* is an ATP- and Zn^{2+}-dependent membrane metalloprotease that degrades both membrane and cytoplasmic proteins. The *ftsH* null mutant strain has a heat-sensitive pleiotropic phenotype that includes sporulation blockage between stage 0 and stage II [25-27]. Proteome analysis of the *ftsH* null strain revealed that disruption of the *ftsH* gene induces 4-fold higher expression of Spo0M, indicating that Spo0M is one of the substrates of FtsH [28]. FstH may control the amount of Spo0M protein to maintain a level appropriate for sporulation. A recent phylogenetic study revealed that the N-terminal half of the Spo0M protein showed significant homology to that of the consensus sequence of eukaryotic arrestins, which suggested that Spo0M might be a member of the arrestin clan [29]. In addition, in recent interactome analysis of CcpA, it has been reported that the Spo0M protein might directly or indirectly bind to CcpA [30]. Thus it is tempting to speculate that Spo0M might be involved not only in sporulation but also in CCR by regulating CcpA functions via complex formation.

Aim of this study

Although the Spo0M protein is thought to play an important role in the initiation of sporulation as well as in CCR and be involved in arrestin-clan proteins, there is only limited information about the function and structure of Spo0M. In this study, the actual function and structure of Spo0M were examined.

In Chapter 1, *spo0M* gene was cloned and the Spo0M protein was overexpressed and purified. The molecular size of the purified protein in solution was estimated by the analytical

gel filtration method. Using purified Spo0M, the direct interaction of Spo0M to CcpA and the effect of Spo0M on the DNA-binding activity of CcpA was evaluated. Finally, the combined method of SPR and MALDI-TOF MS analysis was performed to find unknown interactions of Spo0M in *B. subtilis* cell.

In Chapter 2, thermostable mutant of Spo0M was constructed to obtain high quality crystals. Purified Spo0M protein with N-terminus 10 amino acid truncation was crystallized and the crystal structure was solved by SAD method. On the basis of this structure, the structures of the N-terminus half of Spo0M was compared to the proteins belonging to the arrestin clan. In addition, using the C-terminal half of the Spo0M structure as a query, potential homologues of the C-terminal half of the Spo0M protein was identified.

CHAPTER 1

Characterization of sporulation control protein Spo0M in *Bacillus subtilis*: the involvement of catabolite repression.

Introduction

Nutrients involving carbon source are necessary for all living organisms. To survive under nutrient-limiting condition, *Bacillus subtilis* forms endospore, which is metabolically inactive and extremely resistant to environmental insults [1]. At the initiation of sporulation, two transcriptional regulators, σ^H and Spo0A have a critical role. σ^H is encoded by *spo0H* gene and the peak of transcription is reached at t_0 (initiation point of sporulation) [7]. A number of genes including *spo0A*, *spo0F* and *spooM* are reported to be at least partially controled by σ^H [8, 9]. Spo0A plays a pivotal role at initiation stage and is activated through phosphorylation via phosphorelay system [6]. Two histidine kinases encoded by *kinA* (as known as *spoIIJ*) and *kinB* genes phosphorylate Spo0F protein and the phosphoryl group on Spo0F is further transferred to Spo0B. Then, phosphorylated Spo0B transfers the phosphoryl group to, and thereby activates, Spo0A. Phosphorylated form of Spo0A (Spo0A-P) acts as an active transcriptional factor that regulates a group of genes essential for subsequent endospore formation. Spo0A-P directly and indirectly regulates more than 500 genes [31, 32]. For example, the expression of various σ-factors is upregulated by Spo0A-P. Some of the σ-factors are individually distributed into the spore or mother cell, resulting in distinct gene

expressions in each cell [3, 33]. Spo0A-P also represses *arbB* gene expression [34]. The ArbB protein acts as a repressor for the *spo0H* gene encoding the σ^H protein [7]. σ^H induces expression of the σ^F protein, which in turn is an essential factor for stage II of sporulation [35].

Spo0M in *B. subtilis* is a member of sporulation control protein functioning at stage 0. The protein was first isolated in the study of ADP-ribosylation and was later found to be transcriptionally induced by σ^H at an initiation stage of sporulation.[8] The *spo0M* deficient mutation blocks the sporulation process from stage 0 to stage II. Also, the overexpression of Spo0M results more than 30 times smaller number of endospore formation. Those data indicate that the amount of Spo0M protein in cell should be strictly regulated. Indeed, later study revealed that Spo0M is a potential substrate of FtsH metalloprotease, which is known to be a key regulator involved in sporulation process.[28]

B. subtilis cells cannot sporulate in the presence of a sufficient amount of glucose. However, once *B. subtillis* senses a nutrient limitation, the expressions of various genes involved in sporulation, such as *spo0* genes, are induced.[13, 31] There are a number of reports about the relations between sporulation and CCR. Wünsche *et al.* reported that more than 300 proteins potentially interact with CcpA.[30] Though Spo0M was reported to be involved in these potential interactions with CcpA, no further investigations into the actual interactions or the potential functional role was performed.[30]

Those findings described above indicate that Spo0M might have an important role not only in sporulation but also in CCR. In this chapter, to characterize Spo0M protein, I have cloned *spo0M* gene from *B. subtilis* genome and the Spo0M protein was overexpressed and purified in *E. coli* using pET vector system. To clarify the interaction between the Spo0M and CcpA protein and to speculate as to the potential role of Spo0M in CCR, I prepared

Spo0M, CcpA and the phosphorylated form of HPr (HPr-Ser-P) and examined the intermolecular interactions between these proteins using the biophysical technique of SPR. Then, the effect of Spo0M on the DNA-binding property of the CcpA/HPr-Ser-P complex was examined. Finally, to identify unknown interaction partner of Spo0M, the combined technique of SPR and MALDI-TOF MS analysis [36] was performed and flagellin protein encoded by *hag* gene was identified as an another potential interaction partner. The Hag protein was purified and examined the interaction with Spo0M.

Materials and methods

Bacterial strains, plasmids, regents and primers

E. coli strain JM109 and BL21-CodonPlus(DE3) were used for plasmid amplification and protein expression, respectively. *B. subtilis* strain 168 (ATCC) was used for genomic DNA preparation. The plasmid pET24a (Novagen) were used to overexpress recombinant proteins. All oligonucleotides were purchased from Thermo Fisher Scientific or Eurofins. The following Table 1.1 summarizes oligonucleotides used in this work.

Preparation of genomic DNA of B. subtilis

B. subtilis strain 168 was grown in 4 ml LB medium overnight. On the next day, cultures were centrifuged (8,000 × g, 3 min, 4 °C) and the pellet was suspended in 400 µl of buffer containing 25 mM Tris-HCl pH 8.0, 10 mM EDTA and 50 mM Glucose. 200 µl of ice cold phenol/chloroform/isoamyl alcohol solution (25:24:1) was added and then the mixture was vortexed for 2 min. The solution was centrifuged at 14,000 × g and 4 °C for 5 min and

the supernatant was transferred into new tube. After repeating this extraction 3 times, the cleared supernatant was obtained as a genomic DNA solution.

Tabel 1.1 Oligonucleotides used in this chapter

oligonucleotide	Sequence (5' – 3')[*1]
Spo0M-F	GGGGGGGGGCATATGTCATTTTTTAAGAAGCTTGCG
Spo0M-R	CCCCTCGAGTTACTCAGCGTATTGGTCTAGG
CcpA-F	CCCCGCTAGCATGAGCAATATTACGATCTAC
CcpA-R	GGGGCTCGAGTGACTTGGTTGACTTTCTAAG
HPr-F	CCCCCCCCCCATATGGCACAAAAAACATTTAAAG
HPr-R	GGGCTCGAGTTACTCGCCGAGTCCTTCG
HPrK-F	CCCCCCCCCCATATGGCAAAGGTTCGCACAAAAG
HPrK-R	GGGCTCGAGCTATTCTTCTTGTTCACCGTC
Hag-F	GGGCCCGGGCATATGATGAGAATTAACCACAATATTGC
Hag-R	CCCCTCGAGTTAACGTAATAATTGAAGTACG
Hag-19F	CCCCATATGAACAACAGTGCGAGCCAAAAG
Hag-31F	CCCCATATGTCTTCAGGTCTTCGCATCAAC
Hag-284R	CCCCTCGAGAGCCTGAGAAAGAATGTTGTTC
xylAcre-F(B)[*2]	(Biotin-) AATCATTTTGAAAGCGCAAACAAAGT
xylAcre-R	ACTTTGTTTGCGCTTTCAAAATGATT
Nonspecific-F(B) [*2]	(Biotin-) AATCATTTATGGCATAGGCAACAAGT
Nonspecific-R	ACTTGTTGCCTATGCCATAAATGATT

*1: Underlined sequence indicates restriction enzyme site.

*2: These oligos are biotinylated at the 5' terminus.

Plasmid construction

The DNA fragments encoding the *spo0M*, *ccpA*, *hpr*, *hprk* and *hag* gene were amplified by PCR using genomic DNA of *B. subtitlis* as a template. The amplified fragments were cloned into pET24a vector. Constructed plasmids were introduced into *E. coli* strain

BL21-CodonPlus(DE3) for overexpression. The plasmids constructed in this chapter were listed in following Table 1.2.

Tabel 1.2 Plasmids constructed in this chapter

Plasmid name	Vector backbone	Primers used	Tag
pET24/Spo0M	pET24	Spo0M-F, Spo0M-R	-
pET24/CcpA	pET24	CcpA-F, CcpA-R	C-terminal His_6 tag
pET24/HPr	pET24	HPr-F, HPr-R	-
pET24/HPrK	pET24	HPrK-F, HPrK-R	-
pET24/Hag^{FL}	pET24	Hag-F, Hag-R	C-terminal His_6 tag
pET24/Hag^{19-304}	pET24	Hag-19F, Hag-R	C-terminal His_6 tag
pET24/Hag^{31-304}	pET24	Hag-31F, Hag-R	C-terminal His_6 tag
pET24/Hag^{1-284}	pET24	Hag-F, Hag-284R	C-terminal His_6 tag
pET24/Hag^{19-284}	pET24	Hag-19F, Hag-284R	C-terminal His_6 tag
pET24/Hag^{31-284}	pET24	Hag-31F, Hag-284R	C-terminal His_6 tag

Protein expression and Purification

- *Growth of bacteria and protein overexpression*

BL21-CodonPlus(DE3), harboring each plasmid was grown in LB media at 37 ºC overnight. The overnight culture is backdiluted 1:1000 and grown to an optical density of 0.6 at 600 nm. The temperature was set to 25 ºC and the cells were cultured further 30 min. 0.1 mM IPTG was added and the cells were cultured at 25 ºC further 16 h.

- *Cell disruption*

Cells pellet overexpressing CcpA was suspended in suitable volume of the Buffer A [20 mM Tris-HCl, pH 7.5, 150 mM NaCl and 25 mM], while cells overexpressing Spo0M, HPr and HPrK are resuspended in the Buffer B [20 mM Tris-HCl, pH8.0]. DNase and 3 mM

$MgCl_2$ were added into the cell suspension prior to disruption by passing through a cell disruptor, Avestin EmulsiFlex-C3, at a pressure of 15,000 psi. The disrupted cell lysates were cleared by centrifugation at 15,000 × g for 30 min. The suspension and solution was kept on ice during the whole procedure.

- *Purification of Spo0M*

The cleared lysate containing Spo0M protein was loaded onto a HiTrap DEAE FF column (GE Healthcare) equilibrated with the Buffer B. The column was washed by 20 column volume (CV) of the Buffer B and the Spo0M proteins were eluted using a linear gradient between 0 to 1,000 mM NaCl in the Buffer B. The fractions containing Spo0M proteins were pooled and dialyzed overnight against the Buffer B. The dialyzed solution was loaded onto HiTrap Q HP coulumn (GE Healthcare) and the column was washed by 20 CV of Buffer B. The protein was eluted using a linear gradient between 0 to 1,000 mM NaCl in the Buffer B. The fractions containing Spo0M proteins were pooled and concentrated to ~ 10 mg/ml by ultrafiltration using a VivaSpin-20 concentrator (Sartorius). The concentrated samples were further purified by gel filtration using a Superdex 200 10/30 column (GE Healthcare) in Buffer C [20 mM Tris-HCl, pH 7.5, 150 mM NaCl] at a flow rate of 0.4 mL/min. The fractions containing Spo0M proteins were confirmed by SDS-PAGE, pooled and concentrated to 15 ~ 20 mg/ml by ultracentrifugation using a VivaSpin-2 concentrator (Sartorius). More than 95% of homogeneity was confirmed by SDS-PAGE.

- *Purification of CcpA-His*

The cleared lysate containing C-terminal His-tagged CcpA protein (CcpA-His_6) was loaded onto a HisTrap FF crude column (GE Healthcare) and the column was washed by 20 CV of the Buffer A. CcpA-His_6 were eluted using a linear gradient between 25 to 500 mM imidazole in the Buffer A. The fractions containing CcpA proteins were pooled, concentrated

to ~ 10 mg/ml by ultrafiltration using a VivaSpin-20 concentrator and further purified by gel filtration by the same method described above for the purification of Spo0M protein. More than 95% of homogeneity was confirmed by SDS-PAGE.

- *Purification and phosphorylation of HPr*

The HPr-Ser-P protein was obtained by the method described before [37]. Briefly, the cleared lysate containing HPr was incubated at 70 ºC for 20 min and centrifuged at 15,000 × g for 30 min to remove precipitate. 10 mM ATP, 10 mM FBP and 10 mM $MgCl_2$ were added into the cleared supernatant and phosphorylation of HPr protein was initiated by adding 1/10 volume of the cleared lysate of the cells overexpressing HPrK protein. After reaction mixture was incubated at 37 ºC for 20 min, the solution was loaded onto a HiTrap DEAE FF column. HPr-Ser-P proteins were eluted using a linear gradient between 0 to 1,000 mM NaCl in the Buffer B. The fractions containing HPr-Ser46-P protein were pooled and further purified by gel filtration in the same method described above. More than 95% of homogeneity and the successful phopsphorylation were confirmed by SDS-PAGE and Native-PAGE, respectively.

- *Purification of Hag derivatives*

Wild type Hag (304 amino acids; Hag^{FL}) and its derivatives (Hag^{19-304}, Hag^{31-304}, Hag^{1-284}, Hag^{19-284} and Hag^{31-284}) with C-terminus His-tag were partially purified by the affinity chromatography as described above for CcpA-His. Among them, Hag^{1-284} and Hag^{19-284} were further purified by gel filtration chromatography.

Estimation of the molecular size of Spo0M in solution by gel filtration

The molecular weight of purified $Spo0M^{FL}$ was estimated by gel filtration using a Superdex 75 10/30 column (GE Healthcare) with a calibration curve drawn by standard proteins including bovine serum albumin (66 kDa), chicken ovalbumin (44 kDa), horse

myoglobin (17 kDa) and vitamin B_{12} (1.35 kDa). 100 µl of $Spo0M^{FL}$ protein slotion and standard protein solutions were analyzed at the flow rate of 0.4 ml/min.

SPR analysis

All SPR measurements were performed by using a Biacore 3000 instrument (GE Healthcare) operated at 25 °C. HBS-EP buffer containing 20 mM Hepes pH 7.5, 150 mM NaCl, 3 mM EDTA and 0.005% Tween-20 was used as a running buffer at a flow rate of 20 µl/min.

- *SPR analysis of HPr-Ser-P and Spo0M proteins binding to CcpA*

Anti His-tag monoclonal antibody (GE Healthcare) was immobilized by amine coupling on the carboxylated dextran matrix of a CM5 sensorchip (GE Healthcare) in flowcell 1 and 2. The immobilization method was followed the instruction of Amine coupling Kit and His capture Kit (GE Healthcare). Briefly, after activation of the carboxylated dextran matrix of a CM5 chip by injecting a mixture containing 50 mM NHS and 200 mM EDC the mAb protein in 10 mM sodium acetate, pH 5.0 was injected, resulting about 14,000 RU of the mAb coupled on the surface. The excess activated carboxyls on surface were blocked by the following injection of 1.0 M ethanolamine-HCl, pH 8.5 solution.

Purified CcpA-His_6 proteins were captured at 200 nM for 60 sec. For all experiments, analyte samples at seven of eight concentrations in HBS-EP buffer, were injected for 30 sec, and subsequent dissociation was monitored for 2 min. The buffer flow for 2 min resulted complete dissociation of analytes. To measure direct binding of Spo0M, Spo0M at eight concentrations with two fold serial dilutions in running buffer, were injected and analysed. To measure the binding of Spo0M in the presence of HPr-Ser-P, Spo0M protein solutions

were prepared in the HBS-EP buffer containg 20 μM HPr-Ser-P and injected. To measure the effect of Spo0M on the interaction between CcpA and HPr-Ser-P, HPr-Ser-P protein solutions were injected in the absence and presence of 20 μM Spo0M.

The curve fitting and *K*d estimation were performed with GraphPad Prism software using a nonlinear regression method with a One-site specific binding model: Y = Bmax*X/(*K*d + X) (where X stands for the concentration of analyte, Y for RU and Bmax for the maximum response of RU).

- *SPR analysis of CcpA binding to cre*

To prepare the double strand DNA, equal volumes of both complementary oligonucleotides listed on Table 1.1 at equimolar concentration were mixed in a 1.5 ml microfuge tube. The tubes were heated at 95 °C for 5 min, and then slowly cooled to room temperature for 60 min. The pair of xylAcre-F(B) and xylAcre-R were used for preparing *xylAcre* and the pair of Nonspecific-F(B) and Nonspecific-R were used for *NS*. The DNA sequences of oligonucleotides were referred to those previously reported [38]. About 1,000 RU of streptavidin protein (SIGMA) was immobilized by amine coupling on the CM5 sensorchip in flowcell 1 and 2. The solution containing 50 mM NaOH and 1 M NaCl was injected for 1 min 6 times to wash out unbound streptavidin. About 300 RU of the *NS* and *xylAcre* were captured at 25 nM for 300 sec in flowcell 1 and 2, respectively. To measure the specific interaction between CcpA/HPr-Ser-P complex and *cre*, seven concentrations of the equimolar mixture of CcpA-His_6 and HPr-Ser-P with two fold serial dilutions in running buffer are injected and analyzed. To measure the effect of Spo0M on the DNA-binding of CcpA/HPr-Ser-P, 0.1 μM and 0.3 μM CcpA/HPr-Ser-P binding to *xylAcre* were measures in the presence of Spo0M at the concentration of 1.6 – 200 μM. To confirm direct DNA-binding activity of Spo0M, 200 μM Spo0M was injected in the absence of CcpA/HPr-Ser-P.

- *Identification of protein interacting Spo0M by SPR and MALDI-TOF MS*

About 1,500 RU of purified Spo0M proteins were immobilized on sensor chip CM5 by amine coupling. Reference surface, which was blocked by amine coupling without ligand immobilization, was also prepared. After cleared lysate of *B. subtilis* at a concentration of 10 μg/ml was injected for 2 min, the surfaces were washed with running buffer for 5 min and regenerated with 10 mM HCl solution for 1 min. The HCl solutions were recovered by MICRORECOVER command programmed in Biacore 3000 operation software. The recovered solution was subjected to SDS-PAGE analysis. The single band in the gel stained by CBB was cut out and subjected by MALDI-TOF MS analysis performed by Genomine Inc.

- *SPR analysis of Flagellin binding to CcpA*

The sensor chip with immobilized Spo0M was prepared as in above. Purified Hag^{1-284} and Hag^{19-284} protein at the concentration of 100 μg/ml was injected for 2 min and the sensorgrams were monitored.

Results and discussions

Expression and purification of Spo0M

The data of purification of Spo0M from *E. coli* is summarized in Figure 1.1. To obtain highly pure Spo0M protein, three steps column purification using DEAE sepharose, Q sepharose and Superdex 200 was employed. The treatment of crude lysate by first anion exchange chromatography on DEAE sepharose resulted the peak fractions containing high abundance of ~ 30 kDa protein, which is corresponding to the molecular size of Spo0M. The eluate containing Spo0M was further purified by second anion-exchange chromatography

with Q sepharose. The obtained Spo0M solution was reasonably pure but still contain some contamination proteins in SDS-PAGE gel. The highly pure Spo0M protein was obtained from the third column chromatography with Superdex 200. Typical yield was approximately 50 mg per litter culture.

Estimation of molecular size of Spo0M protein in solution

The purified Spo0M protein was subjected to gel filtration analysis. As shown in Figure 1.2, after comparing the gel filtration markers, the molecular weight of the purified Spo0M protein in solution was estimated to be approximately 51 kDa. The calculated molecular weight of Spo0M based on its amino acid sequence in Protein Calculator v3.4 (http://protcalc.sourceforge.net/) was approximately 30 kDa, suggesting that the Spo0M protein might exist as a dimer in solution. Potential oligomeric state of Spo0M in solution in is further discussed in Chapter 2.

Purification of CcpA and HPr-Ser-P

C-terminally His-tagged CcpA protein was purified by Ni^{2+}-affinity chromatography and size exclusion chromatography. HPr was partially purified by heating at 70 °C. After the precipitant was removed, HPr was phosphorylated by adding the lysate containing HPrK, and HPr-Ser-P protein was further purified by anion-exchange chromatography and gel filtration chromatography. HPr protein without phosphorylation was also purified as a control protein. High purity of CcpA and HPr-Ser-P and the successful phosphorylation of HPr-Ser-P were confirmed by SDS-PAGE and Native-PAGE, respectively (Figure 1.3).

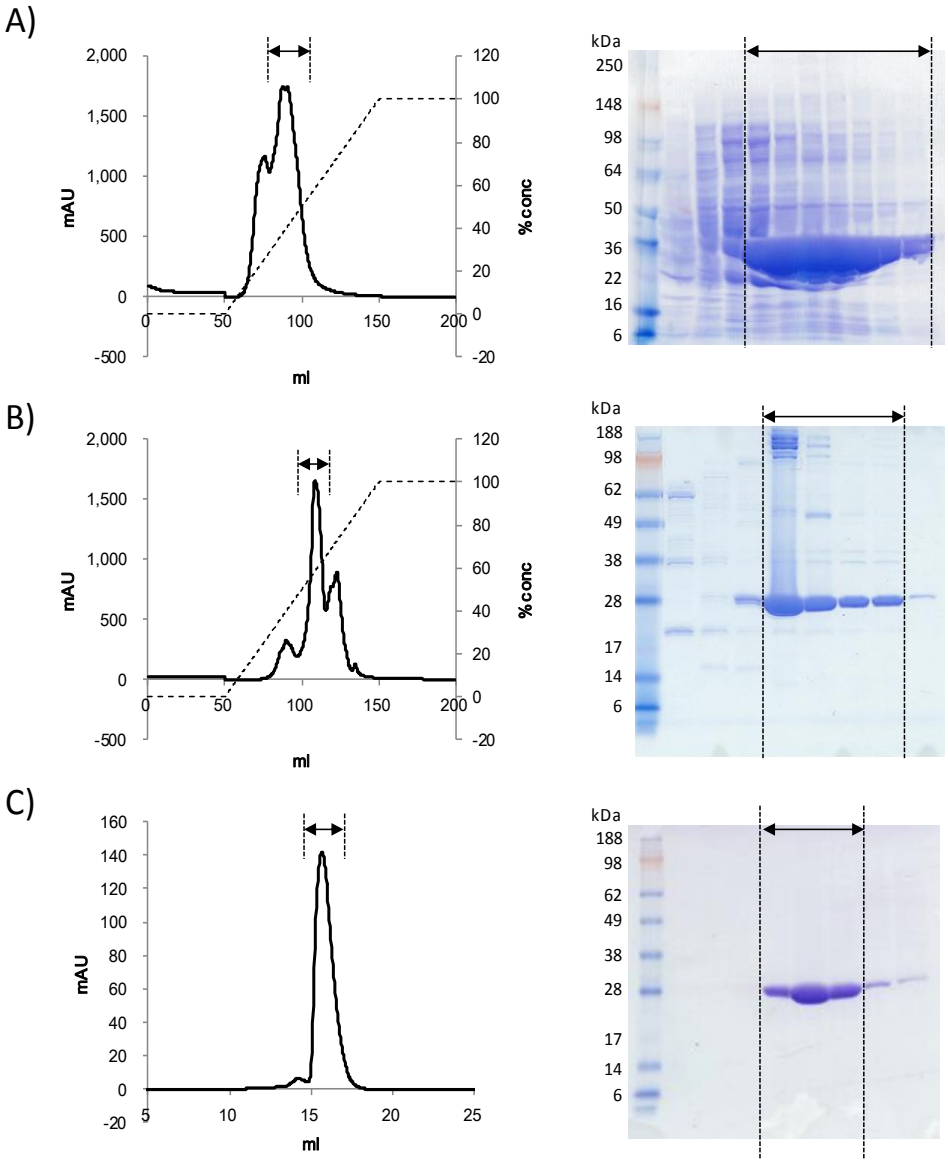

Figure 1.1 Purification of Spo0M with three steps column chromatography. Chromatograms from the purification using the ÄKTA Prime Plus system and SDS-PAGE analysis are indicated on the left and right panel, respectively. HiTrap DEAE FF 5 mL (A), HiTrap Q HP 5mL (B) and Superdex 200 10/30 column (C) were used. The peak fractions containing Spo0M were identified by SDS-PAGE. The fractions collected are indicated by arrows. 4-12% Tris-Glycine gel (A) and 4-12% Bis-Tris gel (B, C) were used for SDS-PAGE. The molecular size (kDa) of SeeBlue® Plus2 Pre-stained Protein Standard (Thermo Fisher scientific) were indicated left on the SDS-PAGE pictures.

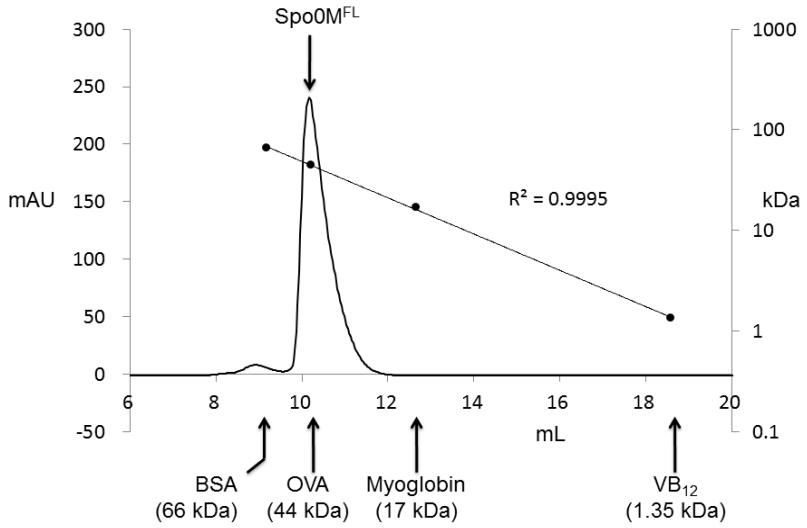

Figure 1.2 Gel filtration chromatography of Spo0M^{FL} and molecular weight standards. Elution profile of Spo0M^{FL} protein from a Surperdex 75 10/30 gel filtration column is indicated. Elution volumes of protein molecular-mass standards, bovine serum albumin (BSA; 66 kDa), chicken ovalbumin (OVA; 44 kDa), horse myoglobin (17 kDa) and vitamin B_{12} (VB_{12}; 1.35 kDa) are indicated at the bottom

.

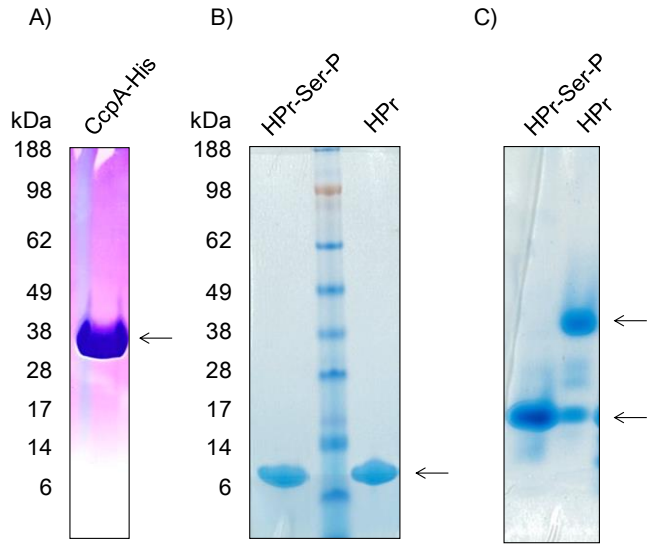

Figure 1.3 SDS-PAGE and Native-PAGE analysis of purified CcpA, HPr and HPr-Ser-P. (A) CcpA-His protein on SDS-PAGE gel. (B) HPr-Ser-P and HPr proteins on SDS-PAGE gel. (C) HPr-Ser-P and HPr proteins on Native-PAGE gel. The arrows indicate CcpA-His, HPr-Ser-P and HPr. 4-12% Bis-Tris gel (A, B) and 12.5% Tris-Glycine gel were used for SDS-PAGE and Native-PAGE, respectively. The molecular size (kDa) of SeeBlue® Plus2 Pre-stained Protein Standard were indicated left on the SDS-PAGE picture.

Intermolecular interaction between CcpA, HPr-Ser-P and Spo0M

Anti-His mouse mAb were chemically immobilized on CM5 sensor chip and CcpA-His_6 was captured by the mAb. Analyte samples were injected over CcpA-His_6 and reference surface (immobilized anti-His mouse mAb but no captured CcpA-His_6). The captured CcpA-His_6 protein was capable of binding with the HPr-Ser-P proteins in a concentration-dependent manner with an estimated affinity *K*d of 9.6 μM (Figuer 1.4A), indicating that the CcpA-His_6 protein and HPr-Ser-P proteins prepared in this study had the functions described previously.[17] Next, the direct interaction between captured CcpA-His_6 and Spo0M in the absence and presence of 20 μM HPr-Ser-P was analyzed. The affinity *K*d values of Spo0M in the absence and presence of HPr-Ser-P were estimated to 54 μM and 13.5 μM, respectively (Figuer 1.4B and 1.4C). These results indicate that Spo0M directly and specifically binds with CcpA and that this interaction is enhanced 4-fold by HPr-Ser-P protein. Since the CcpA protein forms a complex with HPr-Ser-P prior to binding the *cre* box [15], it is suggested that the Spo0M protein might be involved in CCR by affecting the DNA-binding property of CcpA/HPr-Ser-P. The effect of Spo0M on the binding between CcpA and HPr-Ser-P was examined by measuring interaction in the absence and presence of 20 μM Spo0M. While the maximum response of the interaction was increased by approximately double, the affinity *K*d was equivalent to each other (Figuer 1.5). These results indicates that unlike the enhance effect of HPr-Ser-P on the interaction between CcpA and Spo0M, Spo0M does not influence on the binding of HPr-Ser-P with CcpA.

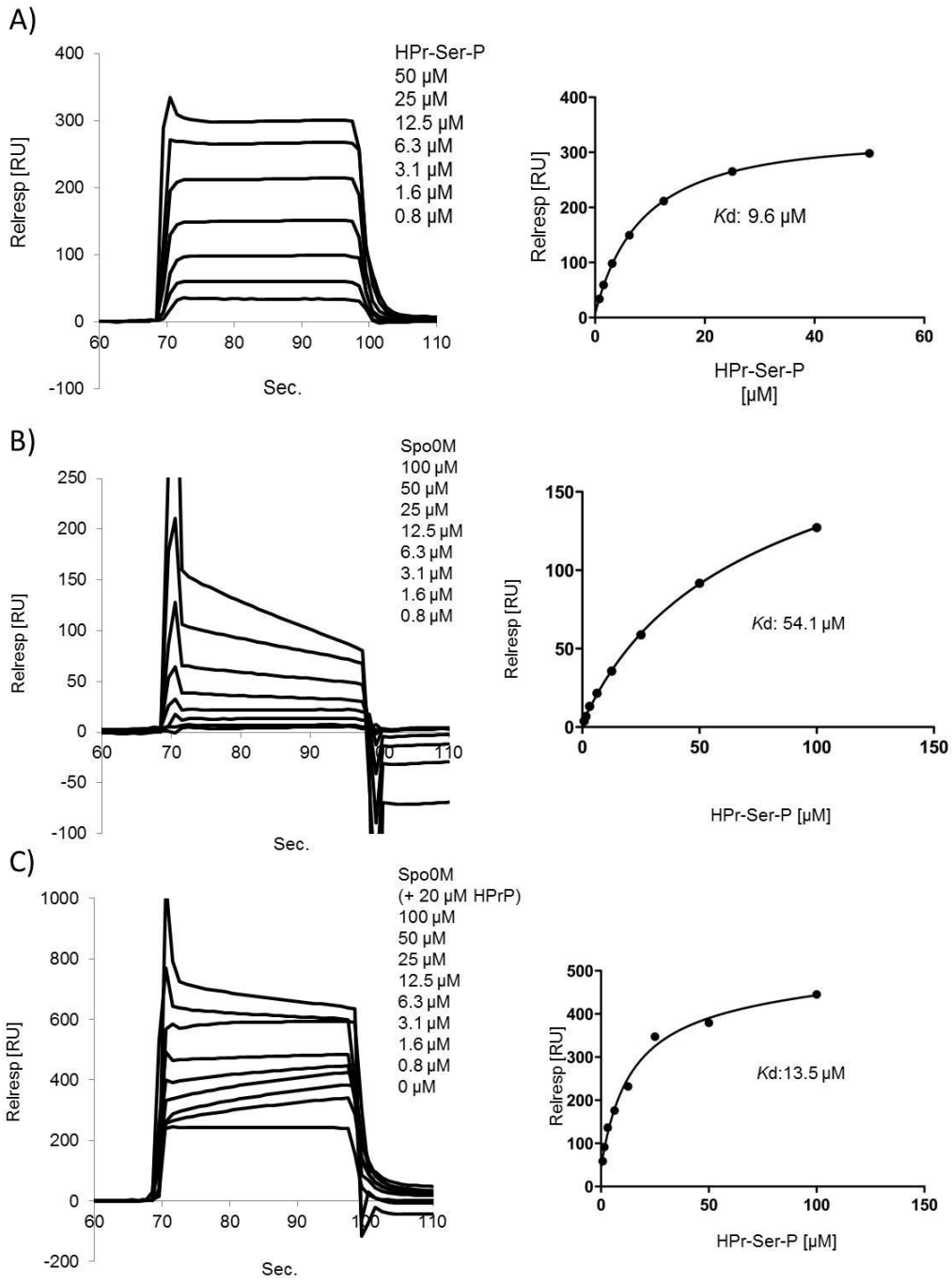

Figure 1.4 SPR analysis of HPr-Ser-P and Spo0M binding to CcpA. The reference subtracted sensorgram (left) and kinetic plot (right) show titration of captured CcpA-His_6 with HPr-Ser-P (A), Spo0M (B) and Spo0M in the presence of 20 µM HPr-Ser-P (C). Concentration of the analyte are depicted on next to each sensorgram. To calculate *K*d of Spo0M in the presence of 20 µM HPr-Ser-P, the bulk response (20 µM HPr-Ser-P alone) was substructed from each sensorgram (C).

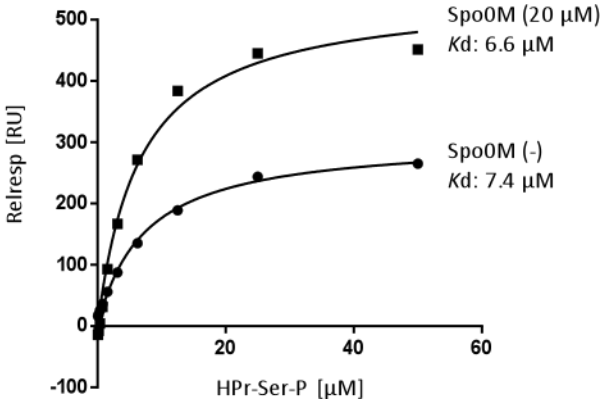

Figure 1.5 SPR analysis of HPr-Ser-P binding to CcpA in the presence of Spo0M. The kinetic plot shows titration of captured CcpA-His_6 with HPr-Ser-P in the presence (filled square) and absence (filled circle) of 20 μM Spo0M. To calculate *K*d in the presence of Spo0M, the bulk response (Spo0M alone) was substructed from each sensorgram.

The influence of Spo0M on the DNA binding of CcpA/HPr-Ser-P

DNA fragment containing *cre* box of *xylA* gene (*xylAcre*) with biotinylation on the 5'-end of forward strand was captured by chemically immobilized streptavidine on sensor chip CM5. The control DNA fragment containing nonspecific DNA sequence (*NS*) was also captured. Various concentrations of an equimolar mixture of CcpA-His_6 and HPr-Ser-P were prepared and injected. The concentration-dependent increase of RU was observed and the apparent *K*d was estimated as 0.94 μM (Figure 1.6A), indicating the specific binding of CcpA-His_6/HPr-Ser-P with *cre* box as described previously.[38] In the presence of 0.3 μM or 0.1 μM of CcpA-His_6/HPr-Ser-P, which produced a sufficient response to evaluate the effect of Spo0M, Spo0M inhibited the DNA binding in a concentration-dependent manner (Figure 1.6B). The *K*i of Spo0M for the DNA-binding of CcpA/HPr-Ser-P was estimated to 45.5 μM from a Dixon plot. No specific binding of 200 μM Spo0M with *xylAcre* was observed (Figure 1.6B), indicating Spo0M has no specific DNA-binding function with *cre*. Because Spo0M binds directly with CcpA as described above, the Spo0M protein might inhibit CcpA/HPr-

Ser-P/*cre* complex formation via a direct interaction with CcpA/HPr-Ser-P, and the Spo0M protein might be involved in the regulation of gene expressions related to catabolite metabolism by CcpA.

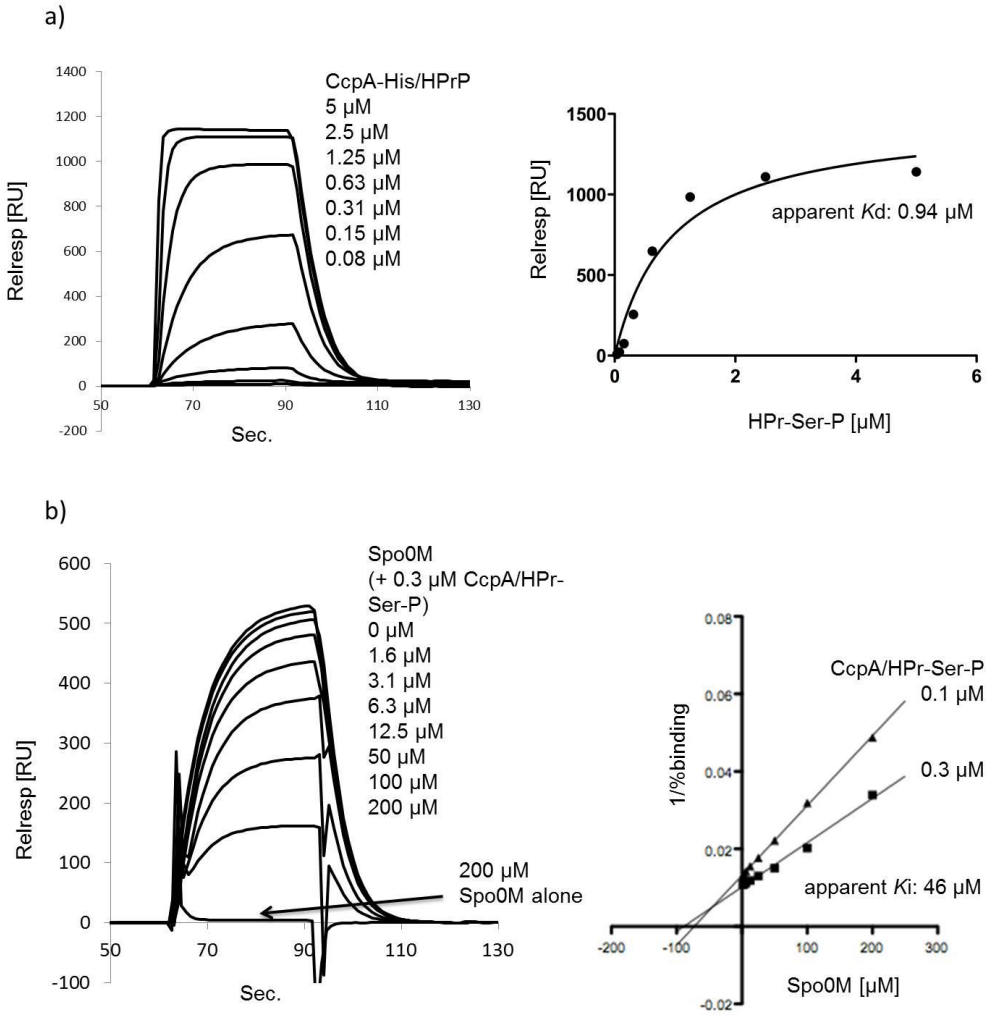

Figure 1.6 SPR analysis of CcpA binding to cre and the effect of Spo0M on the DNA binding property of CcpA. (A) The reference subtracted sensorgram (left) and kinetic plot (right) of CcpA/HPr-Ser-P binding to *xylAcre*. (B) Reference subtracted sensorgram of the binding of CcpA/HPr-Ser-P in the presence of 0.3 μM Spo0M (left) and the Dixon plot of the inverse of the percentage binding of CcpA/HPr-Ser-P (where RU in the absence of Spo0M = 100) (right). The CcpA/HPr-Ser-P concentrations employed were 0.1 μM (triangles) and 0.3 μM (squares). Concentrations of the equimolar mixture (A) and Spo0M (B) are depicted next to the sensorgram.

Identification of flagellin as a potential binding protein with Spo0M

As Spo0M is reported to be involved in arrestin clan [29] and arrestin functions via protein-protein interactions to their binding partner proteins, I speculate that, like other arrestin clan proteins, Spo0M might regulate other protein's function by substantial protein-protein interaction. To identify unknown binding partners to Spo0M, purified Spo0M proteins were chemically immobilized on the dextran surface of sencor chip and *B. subtilis* cell lysate was injected. As shown in Figure 1.7, the increased sensorgram was observed, suggesting the existence of molecule(s) binding to Spo0M in *B. subtilis* lysate. After regeneration of the surface, single protein was observed in SDS-PAGE analysis of the regeneration solution, suggesting that the protein might interact with Spo0M. MALDI-TOF MS analysis was performed and the protein was identified to flagellin homologue called as Hag, the major filament protein in *B. subtilis*.

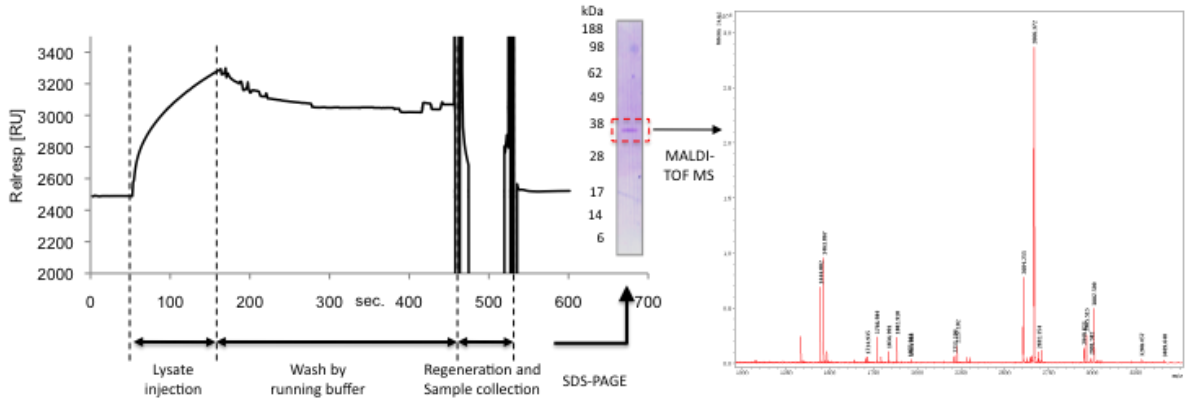

Figure 1.7 Identification of Hag as a potential binding partner by SPR and MALDI-TOS MS analysis. The reference subtracted sensorgram of SPR analysis and MALDI-TOF MS analysis were shown on the left and right, respectively. The molecules binding to the surface was eluted by HCl and analyzed SDS-PAGE. The single protein band was extracted and analyzed by MALDI-TOF MS to identify the protein. The molecular size (kDa) of SeeBlue® Plus2 Pre-stained Protein Standard were indicated left on the SDS-PAGE picture.

Purification of Hag

Fllagelin is known to polymerize to form bacterial flagellum, which is an extracellular organelle that facilitates motility [39]. To clarify the intramolecular interaction between Spo0M and Hag, *hag* gene was cloned and overexpressed. It was reported that flagellin homologue of *Sphingomonas sp. A1*, which is called as p5, polymerizes into a flagellar filament *in vivo* and *in vitro* [40] and that while purified wild type p5 molecule was eluted in the void volume in gel filtration column chromatography, the truncation mutant that lacked the N- and C-terminal regions was unable to form a filament and was eluted as a monomers [41]. We therefore constructed several Hag derivatives with N and/or C terminus truncations referred to as the sequence alignment of Hag and p5 (data not shown). As expected, Hag^{FL} protein with full length is successfully expressed in solution but the large portion of the expressed Hag^{FL} was aggregated and lost during His-tag purification (Figure 1.8). In contrast to Hag^{FL}, Hag^{1-284} and Hag^{19-284} proteins were successfully purified stably. Thus, these proteins were chosen to examine direct binding to Spo0M.

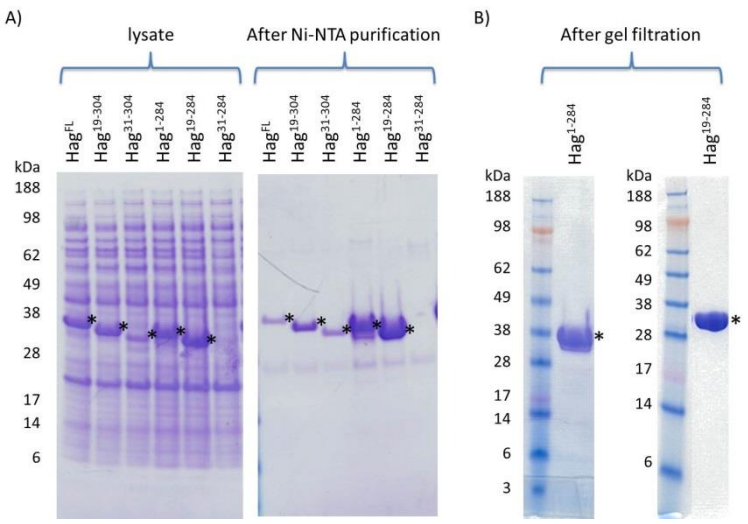

Figure 1.8 Purification of Hag and derivatives. (A) SDS-PAGE analysis of the cell lysate (left) or the eluates from Ni-NTA affinity chromatography (right) of Hag and its derivatives. (B) Purified Hag^{1-284} and Hag^{19-284} after gel filtration chromatography. The molecular size (kDa) of SeeBlue® Plus2 Pre-stained Protein Standard were indicated left on the SDS-PAGE picture. The asterisks indicate bands corresponding to Hag proteins.

SPR analysis of interaction between Hag and Spo0M

In SPR experiments, no obvious interaction were observed between immobilized Spo0M and both Hag^{1-284} and Hag^{19-284} proteins (Figure 1.9). Since wild type flagellin is known to form a polymerized structure, the Hag derivatives used in this study might have different conformation or oligomeric state from wild type Hag in *B. subtilis* lysate, resulting the failed interaction between purified Hag truncation mutants and Spo0M. In *vivo* analysis is necessary to clarify the actual interaction.

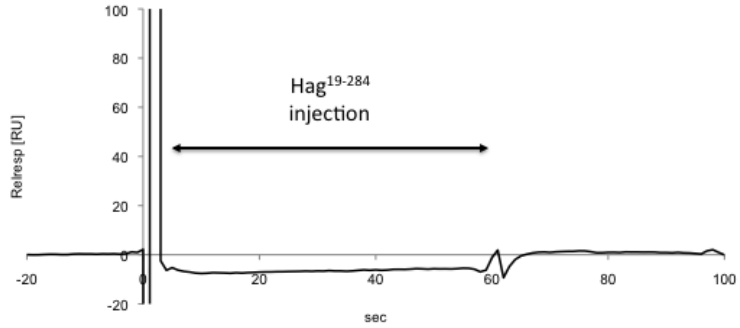

Figure 1.9 SPR analysis of Hag^{19-284} binding to Spo0M. Almost identical sensorgram was obtained from the experiment of Hag^{1-284} (data not shown).

Conclusion

Spo0M in *B. subtilis* has been successfully expressed and purified. The molecular size of purified Spo0M protein was estimated to 51 kDa, suggesting that the Spo0M protein might exist as a dimer in solution. I revealed that Spo0M directly bound to CcpA with a *K*d of 54 μM and the binding affinity was enhanced to 4 folds in the presence of HPr-Ser-P. In addition, Spo0M inhibits DNA binding of CcpA/HPr-Ser-P with a *K*i of 45.5 μM. These data suggested that under glucose-limiting environment, Spo0M might inhibit DNA-binding function of CcpA to release CcpA-mediated catabolite repression, resulting the induction of sporulation related gene expressions. Although flagellin homologue protein called as Hag

was indentified as a potential binding partner of Spo0M, actual interaction was not clarified *in vitro* study. Further functional studies of Spo0M function are necessary to reveal the actual role of Spo0M in *vivo*.

CHAPTER 2

Structure of sporulation control protein Spo0M from *Bacillus subtilis.*

Introduction

A recent phylogenetic study revealed that the N-terminal half of the Spo0M protein showed significant homology to that of the consensus sequence of eukaryotic arrestins, which indicated that Spo0M might be a member of the arrestin clan.[29] The arrestin clan consists of β- and visual (β/v-) arrestins, α-arrestins, and vacuolar protein sorting-associated protein (VPS) 26.[29, 42, 43] The β/v-arrestins are the best characterized members and known as multifunctional scaffolding proteins that regulate G-protein coupled receptor (GPCR)-dependent signaling and membrane trafficking.[44] The β/v-arrestins have structurally well conserved two-lobe, ω-shaped, immunoglobulin-like, β-strand sandwich structures.[45-47] In addition, the crystal structures of other arrestin-clan members, such as VPS26 [48], thioredoxin-interacting protein (TXNIP) [49, 50] and the N-terminus domain of arrestin-domain-containing (ARRDC) 3 [51], which are distantly related to β/v-arrestins on the basis of their amino acid sequences (~ 20% identity), have a β-strand sandwich domain that resembles β/v-arrestins. These structural studies indicate that the 3D structure of mammalian members of the arrestin clan share highly and structurally conserved N- and C-terminal domains in spite of their poor amino

acid homologies. Although Spo0M is thought to be a bacterial member of the arrestin clan, no structural information of any bacterial member of the arrestin clan has been reported so far. Furthermore, in contrast to the N-terminal half, there has been no reliable information or even speculation as to the structure and function of the C-terminal half of the Spo0M protein.

In this chapter, we determined the crystal structure of the truncated form, Spo0M$^{11\text{-}258}$, lacking the 10 amino acids of the N terminus, at a resolution of 2.3 Å. On the basis of this structure, we compared the structures of the N-terminus half of Spo0M and the other proteins belonging to the arrestin clan. In addition, using the C-terminal half of the Spo0M structure as a query, we identified potential homologues of the C-terminal half of the Spo0M protein.

Materials and methods

Bacterial strains, plasmids, regents and primers

The bacterial strains described in chapter 1 were used. The plasmid pET15b (Novagen) was used to overexpress recombinant proteins. All oligonucleotides were purchased from Thermo Fisher Scientific or Eurofins. The following Table 2.1 summarizes oligonucleotides used in this chapter.

Table 2.1 Oligonucleotides used in this chapter

oligonucleotide	Sequence (5' – 3')*
Spo0M-F	GGGGGGGGGCATTAGTCATTTTTTAAGAAGCTTGCG
Spo0M-6F	CCCGGGCCCCATATGAAGCTTGCGGCAAGTGCC
Spo0M-11F	CCCGGGCCCCATATGGCCGGAATTGGAGCTGCA
Spo0M-16F	CCCGGGCCCCATATGGCAAAAGTAGATACAATTTTAG
Spo0M-21F	CCCGGGCCCCATATGATTTTAGAAAAAGACGCATATTTTCC
Spo0M-R	CCCCTCGAGTTACTCAGCGTATTGGTCTAGG
Spo0M-235R	GGGCTCGAGTTAGCTGAAACGGACGCGTAC
Spo0M-245R	GGGCTCGAGTTATAGCTCTTCTGTATCTTC
Spo0M-250R	GGGCTCGAGTTATTCTAGCACCTCTTCTAG
Spo0M-255R	GGGCTCGAGTTATTGGTCTAGGATCTCTTC

*Underlined sequence indicates restriction enzyme site.

Plasmid construction

The DNA fragments encoding the full-length (258 amino acids; $Spo0M^{FL}$) and truncated versions ($Spo0M^{1-255}$, $Spo0M^{1-250}$, $Spo0M^{1-245}$, $Spo0M^{1-235}$, $Spo0M^{6-258}$, $Spo0M^{11-258}$, $Spo0M^{16-258}$ and $Spo0M^{21-258}$) of the *spo0M* gene were amplified by PCR and cloned into the pET15b vector using the *Nde* I and *BamH* I restriction sites, in which an N-terminal His-tag is appended to the coding sequence. All of constructed plasmids were introduced into *E. coli* strain BL21-CodonPlus(DE3) for overexpression. The following Table 2.2 summarizes plasmids constructed in this chapter

Tabel 2.2 Plasmids constructed in this chapter

Plasmid name	Vector backbone	Primers used	Tag
pET15/Spo0M^{FL}	pET15	Spo0M-F, Spo0M-R	C-terminal His_6 tag
pET15/Spo0M$^{6-258}$	pET15	Spo0M-6F, Spo0M-R	C-terminal His_6 tag
pET15/Spo0M$^{11-258}$	pET15	Spo0M-11F, Spo0M-R	C-terminal His_6 tag
pET15/Spo0M$^{16-258}$	pET15	Spo0M-16F, Spo0M-R	C-terminal His_6 tag
pET15/Spo0M$^{21-258}$	pET15	Spo0M-21F, Spo0M-R	C-terminal His_6 tag
pET15/Spo0M$^{1-255}$	pET15	Spo0M-F, Spo0M-255R	C-terminal His_6 tag
pET15/Spo0M$^{1-250}$	pET15	Spo0M-F, Spo0M-250R	C-terminal His_6 tag
pET15/Spo0M$^{1-245}$	pET15	Spo0M-F, Spo0M-245R	C-terminal His_6 tag
pET15/Spo0M$^{1-235}$	pET15	Spo0M-F, Spo0M-235R	C-terminal His_6 tag

Protein expression and purification

- *Protein expression and Cell disruption*

The bacterial culture and protein expression were performed as described in chapter 1. Cells were harvested and resuspended in Buffer D [20 mM Tris, pH 7.5, 50 mM NaCl, 1 mM EDTA, 0.5 mM DTT and 25 mM imidazole]. Cells were disrupted by the method described in chapter 1.

- *Protein purification*

The cell lysates were loaded onto a HisTrap FF crude column equilibrated with Buffer A. The columns were washed with a 10 CV of Buffer D. Proteins were eluted using a linear gradient between Buffer D and Buffer E [20 mM Tris, pH 7.5, 50 mM NaCl, 1 mM EDTA, 0.5 mM DTT and 500 mM imidazole]. The fractions containing Spo0M proteins were pooled and dialyzed overnight at 4 $^{\circ}$C in the presence of thrombin protease (Novagen) against Buffer F [20 mM Tris, pH 7.5, 50 mM NaCl, 1 mM EDTA, 1 mM DTT]. The dialyzed sample was applied to a HiTrap Q column (GE Healthcare) equilibrated with Buffer F and eluted using a

linear gradient between Buffer F and Buffer G [20 mM Tris, pH 7.5, 1 M NaCl, 1 mM EDTA, 1 mM DTT]. The fractions containing Spo0M proteins were pooled and concentrated to ~ 10 mg/ml by ultrafiltration using a VivaSpin-6 concentrator. Finally, the concentrated samples were purified by gel filtration using a Superdex 200 10/30 column in Buffer H [20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM DTT] at a flow rate of 0.4 mL/min. The fractions containing Spo0M proteins were confirmed by SDS-PAGE, pooled and concentrated to 15 ~ 20 mg/ml by ultracentrifugation using a VivaSpin-2.

Crystallization

An attempt was made to crystallize the purified Spo0M proteins using the sitting drop vapor diffusion method. An initial crystallization screening was performed using Hampton Crystal Screen I/II by mixing 0.4 µl of protein with 0.4 µl of reservoir solution. The droplets were equilibrated against 75 µl of reservoir solution at 4 $^{\circ}$C and 20 $^{\circ}$C. To obtain diffraction quality crystals, the crystallization conditions were further refined. The optimal crystallization conditions for Spo0M^{FL} were defined as 0.1 M Mes, pH 6.5, 0.9 - 1.0 M Li_2SO_4 at 20 $^{\circ}$C. The crystallization drops were prepared by mixing 2 µl of protein with 2 µl of buffer. The crystals of Spo0M^{FL} appeared within one week. The crystals were transferred into a solution containing 25% glycerol as a cryoprotectant prior to flash-freezing in a cold nitrogen gas stream (100 K). The optimal crystallization conditions for the N-terminal 10 amino acids truncated version, Spo0M$^{11-258}$, were defined as 0.1 M Tris-HCl, pH 7.0, and 1.0 - 1.1 M Na-citrate at 4 $^{\circ}$C. The crystallization drops were prepared by mixing 2 µl of protein with 2 µl of buffer. The crystals of Spo0M$^{11-258}$ appeared within two days. Crystals were transferred into a solution containing 15% ethylene glycol as a cryoprotectant prior to flash-freezing in a cold nitrogen gas stream (100 K).

Heavy atom derivatization

To screen heavy atoms, the Spo0M$^{11\text{-}258}$ crystals were soaked in a heavy atom solution and then the X-ray diffraction data were collected using an in-house Bruker Hi-star detector coupled with a Bruker M18XHF rotating anode X-ray generator, and K_2PtCl_4 was selected for heavy atom derivatization. A platinum derivative of Spo0M$^{11\text{-}258}$ crystal was obtained by soaking crystals in a solution containing 0.1 M Tris-HCl, pH 7.0, 1.1 M Na-citrate and 10 mM K_2PtCl_4 for 15 min. The crystals were then transferred into a solution containing 15% glycerol as a cryoprotectant prior to flash-cooling by a cold nitrogen gas stream (100 K).

Data collection and structure determination

All X-ray diffraction data were collected at SPring-8 (Hyogo, Japan) after an in-house diffraction check. Diffraction data of Spo0M^{FL} were collected at the BL44XU beamline. The crystal of Spo0M^{FL} was mounted on a nylon loop (Hampton Research) to collect diffraction data using a MAR CCD scanner. The distance between the crystal and detector was set to 200 mm, with an oscillation range of 1 $^{\circ}$. The data were processed using HKL2000 [52], with a final R_{merge} of 6.0% in the resolution range 50 – 3.2 Å. The space group was estimated to *P*1. Although we tried to solve the structure of Spo0M^{FL} by the molecular replacement method using the published structures of arrestin-clan proteins as search models, no reasonable solution was obtained. In addition, the crystal quality of Spo0M^{FL} was not reproducible, which made optimization difficult. To obtain crystals of better quality, truncation mutants were screened and Spo0M$^{11\text{-}258}$ was obtained as a stable mutant (see below).

Diffraction data of Spo0M$^{11\text{-}258}$ were also collected at the BL44XU beamline with the settings described above. The data were processed with a final R_{merge} of 5.6% in the

resolution range 50 – 2.3 Å. The space group was estimated as $P2_12_12_1$. As seen in Spo0M^{FL}, the phase was not determined by the molecular replacement method. To solve the phase problem, a platinum derivative of the Spo0M$^{11\text{-}258}$ crystal was prepared for single-wavelength-anomalous diffraction (SAD) phasing. The wavelength was set at 1.07 Å from X-ray absorption fine structure (XAFS) measurements. The diffraction data were processed with a final R_{merge} of 6.9% in the resolution range 50 – 2.8 Å. The data-collection statistics of native and Pt-derivatized Spo0M$^{11\text{-}258}$ are summarized in Table 2.3.

The crystal structure of Pt derivative of Spo0M$^{11\text{-}258}$ was determined by the SAD method. I identified two Spo0M molecules and seven platinum-binding sites in the asymmetric unit. The initial phase was determined by AutoSol and the initial model was obtained using the AutoBuild in Phenix program.[53] The crystal structure of native Spo0M$^{11\text{-}258}$ was solved by molecular replacement with Phaser from the Phenix program using coordinates from the Pt-derivative structure as a starting model. Further model building and refinement were performed using Coot [54] and Phenix, resulting in a 2.3 Å structure of Spo0M$^{11\text{-}258}$. Noncrystallographic symmetry (NCS) restraints were used in all the stages of refinement. Structure analysis was performed using the MolProbity in Phenix program.[55] Molecular-graphics images were prepared using PyMOL (Schrodinger; http://www.pymol.org). Protein interfaces in the structures were analyzed by PDBePISA.[56]

Table 2.3 Data collection and refinement statistics

	Spo0M$^{11-258}$, native (PDB ID; 5CL2)	Spo0M$^{11-258}$, K_2PtCl_4
Data Collection		
Source	BL44XU, SPring-8	BL26B1, SPring-8
Wavelength (Å)	0.9000	1.0700
Detector	CCD (MAR300HE)	CCD (SaturnA200)
Space group	$P2_1\,2_1\,2_1$	$P2_1\,2_1\,2_1$
Unit cell	a = 66.00, b = 74.23, c = 133.08	a = 64.70, b = 79.06, c = 131.33
Resolution limit (Å)	50 - 2.3 (2.34 - 2.30)	50 - 2.8 (2.85 - 2.80)
Total reflections	588370	405995
Unique reflections	29843 (1428)	17097 (833)
Completeness (%)	99.6 (96.8)	100 (100)
Multiplicity	7.6 (7.1)	8.7 (8.6)
R-merge	0.056 (0.510)	0.069 (0.505)
Mean I/sigma(I)	48.99 (4.90)	52.09 (7.80)
Structure Determination		
Figure of Merit	-	
Prior to density modification		0.25
After density modification		0.77
No. of Pt sites per asymmetric unit	-	7
Refinement		
Program used	*phenix.refine*	
Resolution range (Å)	40 - 2.3 (2.38 - 2.30)	
No. of refelections used	29788 (2837)	
Completeness (%)	99.5 (96.2)	
Protein molecules per asymmetric unit	2	
Contents of aymmetric unit		
Residues	491	
Waters	43	
Average B-factor (Å)		
Protein	62.9	
Waters	54.8	
Bond-length r.m.s. (Å)	0.01	
Bond-angle r.m.s. (^{o})	1.39	
R_{work} (%)	0.23 (0.31)	
R_{free} (%)	0.27 (0.38)	

Thermal shift assay

25 μM of Spo0M proteins in Buffer H containing Protein Thermal Shift^{TM} Dye (Life Technologies) were incubated in the wells of RT-PCR devices (Life Technologies, StepOnePlus^{TM}). The samples were heated from 25 ºC to 90 ºC. Fluorescence intensities were plotted as a function of temperature and accurate fitting to the Boltzmann equation by non-linear regression was performed using GraphPad Prism. The inflection point of the transition curve Tm was calculated using the Boltzmann equation, y = LL+(UL-LL)/(1+exp((Tm-x)/Slope)), where LL and UL are the values of minimum and maximum intensities, respectively.

Results and discussions

Expression and purification of Spo0M truncation mutants

All Spo0M proteins (Spo0M^{FL}, Spo0M$^{1-255}$, Spo0M$^{1-250}$, Spo0M$^{1-245}$, Spo0M$^{1-235}$, Spo0M$^{6-258}$, Spo0M$^{11-258}$, Spo0M$^{16-258}$, Spo0M$^{21-258}$) were overexpressed as N-terminus His_6-tagged proteins in *E. coli*. The expression levels of these proteins were almost the same. However, no proteins of Spo0M$^{1-250}$, Spo0M$^{1-245}$, Spo0M$^{1-235}$ and Spo0M$^{21-258}$ were obtained in soluble fractions of the cell lysate, indicating that these mutant proteins were misfolded and aggregated. The proteins that were successfully overexpressed in soluble fractions, Spo0M^{FL}, Spo0M$^{1-255}$, Spo0M$^{6-258}$, Spo0M$^{11-258}$ and Spo0M$^{16-258}$, were purified to homogeneity by Ni-NTA, anion-exchange and gel filtration column chromatography, resulting in proteins of more than 95% homogeneity as confirmed by SDS-PAGE. All proteins were eluted at similar retention times by using gel filtration column chromatography (data not shown).

Thermal stability of Spo0M truncation mutants

Although I first attempted to determine the structure by using crystals of the $Spo0M^{FL}$ protein, only crystals that were poorly diffracted, i.e., diffracted up to 3.2 Å, were obtained and the phase was never determined by the molecular replacement method. In addition, the poor reproducibility of the crystal quality led to difficulties in optimizing the crystallization conditions as well as heavy atom derivatization. Since the thermal stability of a purified protein can be a good indicator of the crystal quality [57, 58], I tried to obtain more thermally stable proteins by using N/C-terminus truncations, and the thermal stabilities of each truncation were evaluated by using a differential scanning fluorimetry (DSF) technique. The melting temperatures (Tm, $^{\circ}C$) of each protein are shown in Table 2.4. The Tm of the full-length Spo0M was 41.7 $^{\circ}C$. In terms of C-terminus truncations, 8 or more amino acid truncations from the end completely abolished the stable expression of the Spo0M protein in solution. Although an $Spo0M^{1-255}$ protein that lacks only 3 amino acids from the C-terminal end was overexpressed in the soluble fraction and the protein was successfully purified, the Tm of the purified $Spo0M^{1-255}$ protein decreased to 32.7 $^{\circ}C$, which is 9 $^{\circ}C$ lower than that of $Spo0M^{FL}$. The N-terminus truncation mutants, $Spo0M^{6-258}$, $Spo0M^{11-258}$ and $Spo0M^{16-258}$, but not $Spo0M^{21-258}$, were successfully overexpressed and purified. In contrast to the C terminus truncations, which resulted in unstable proteins, the N-terminus truncation mutants exhibited greater stability than the $Spo0M^{FL}$ proteins with 5 to 10 $^{\circ}C$ higher Tm, indicating that the N-terminus of Spo0M was flexible and negatively influences the thermal stability after purification. Among those mutants, the crystals of the $Spo0M^{11-258}$ protein, which showed the highest Tm among the mutants tested, diffracted to significantly higher resolution than the crystals of $Spo0M^{FL}$, indicating that screening for a thermal stable mutant is a reliable strategy for obtaining good diffraction quality crystals, as described elsewhere.

Table 2.4 Summary of results obtained from thermal shift assay.

	Tm (^{o}C)	Number of experiment
Spo0M^{FL}	41.7 ± 0.76	3
Spo0M$^{6-258}$	46.3 ± 0.42	3
Spo0M$^{11-258}$	50.2 ± 0.56	3
Spo0M$^{16-258}$	46.7 ± 1.44	3
Spo0M$^{1-255}$	32.7	1

Crystallization and X-ray data collection

The pictures of crystals obtained in this study were shown in Figure 2.1. The crystals of Spo0M^{FL} were obtained in one week and reach maximum dimensions of approximately 0.2×0.05×0.05 mm within one month. Although the crystals were diffracted up to 3.2 Å, the phase was never determined by molecular replacement with known arretin structures, which is not surprising due to the extremely poor sequence homology. The crystal shape is dramatically changed by adding 0.01 M $CaCl_2$ in the well solution, but no improvement of diffraction was obtained. In contrast to Spo0M^{FL} crystals, the rod shaped crystals of Spo0M$^{11-258}$ appear in a few days and reach maximum dimensions of approximately 0.5×0.1×0.1 mm within 2 weeks. The crystals were diffracted up to 2.3 Å and phase was successfully determined by SAD method using Pt-derivative.

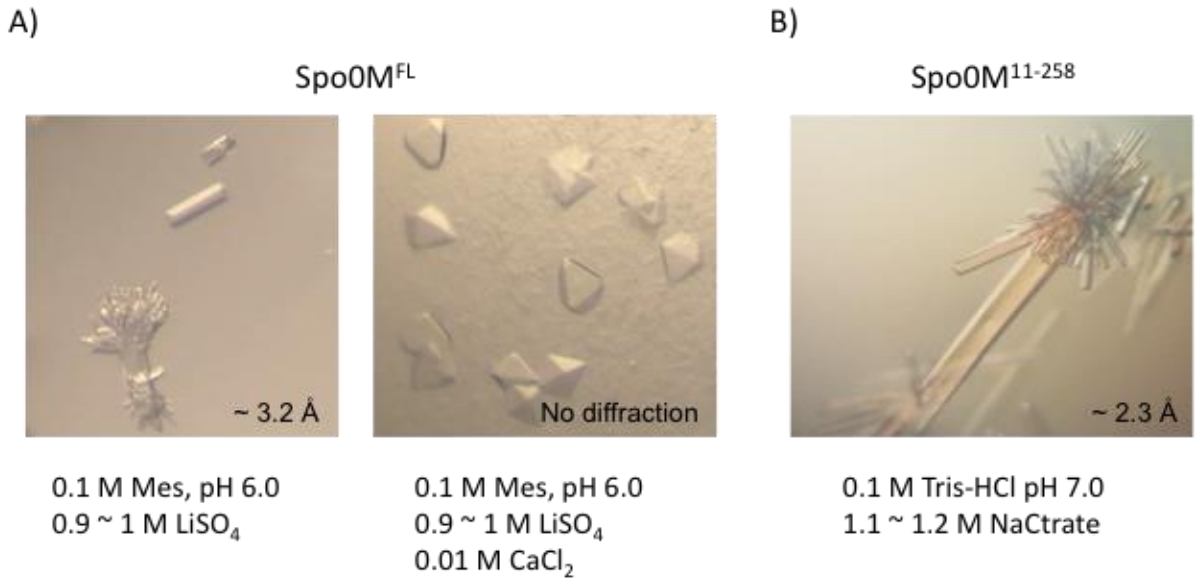

Figure 2.1 Crystals pictures of Spo0M^{FL} (A) and Spo0M$^{11-258}$ (B). The composition of reservoir solution and maximum X-ray diffraction were indicated under and inside the pictures, respectively.

Overall structure

The structure of *B. subtilis* Spo0M$^{11-258}$ was determined by X-ray crystallography using the SAD technique with Pt-soaked crystals. The structure of Spo0M$^{11-258}$ in the $P2_12_12_1$ crystal form was refined to 2.3 Å. The Ramachandran plot of the final model shows 99.4%, 0.6% and 0% of the residues in the preferred, allowed and disallowed regions, respectively. Two molecules, molecules A and B, were found in an asymmetric unit. In the final model, the residues of 4 amino acids derived from the expression vector and 2 amino acids of the N terminus of Spo0M$^{11-258}$ in both molecules and a glutamate residue in the C-terminal end of molecule B could not be assigned. One Na^+ was added to form bonds with the side chains of Asp128 and Asp130 and main chain of Thr20 in each molecule. The two monomers in an asymmetric unit are essentially identical, with a root-mean-square deviation (RMSD) of 0.289 Å for corresponding main chain atoms.

The overall structure of Spo0M is shown in Figure 2.2. The Spo0M$^{11-258}$ molecule consists of two domains and contains fourteen β-strands and six α-helices. The N-terminal domain (N domain) contains residues 13 - 134, which form a β-sandwich structure composed of two antiparallel β-sheets of three (β1a, β2 and β5) and four (β3, β4a, β6 and β7) β-strands. In contrast, the C-terminal domain (C domain) contains residues 138 - 258, which adopt an α/β-fold involving an antiparallel β-sheet of five β-strands (β8 - 12) and six α-helices (α1 - 6) surrounding the β-sheet. There is a short loop consisting of residues 135 - 137, which connects the N and C domains (Figure 2.2A and 2.2B).

Two distinct dimeric arrangements of the Spo0M$^{11-258}$ protein were observed in the crystal lattice, both of which were composed of molecules A and B in the asymmetric unit (Figure 2.2C). In these dimers, two protomers interact with each other through substantial intermolecular interactions. In the dimeric arrangements shown on the left and right in Figure 2.2C, the intermolecular interface is formed between the two N domains (referred to as N-interface) and between the two C domains (referred to as C-interface), respectively. There are four and six potential hydrogen bonds in the N- and C-interfaces, respectively. While both of the N- and C-interfaces possess sufficiently large buried surface areas as the potential dimer interface in solution: approximately 1000 Å^2 and 800 Å^2 (8.1% and 6.4% of the whole surface area), respectively. The Δ^iG P-values [59] of the N- and C-interfaces were 0.587 and 0.671, respectively, which indicates that both the interfaces might be artifacts of crystal packing. As a positive correlation between radius of gyration (Rg) and molecular size of protein is reported [60, 61], Rg of proteins used for the estimation of molecular size in gel filtration study were calculated by using HYDROpro software [62]. The Rg of OVA (PDB ID; 1OVA), Spo0M$^{11-258}$ monomer, Spo0M$^{11-258}$ dimer with N-interface, Spo0M$^{11-258}$ dimer with C-interface and BSA (PDB ID; 4F5S) were calculated to 22.5 Å, 22.6 Å, 26.6 Å, 26.1

Å and 27.6 Å, respectively. The Rg of OVA is equivalent to that of Spo0M monomer, indicating that the estimation of molecular size of Spo0M by gel filtration described in chapter 1 might be inaccurate.

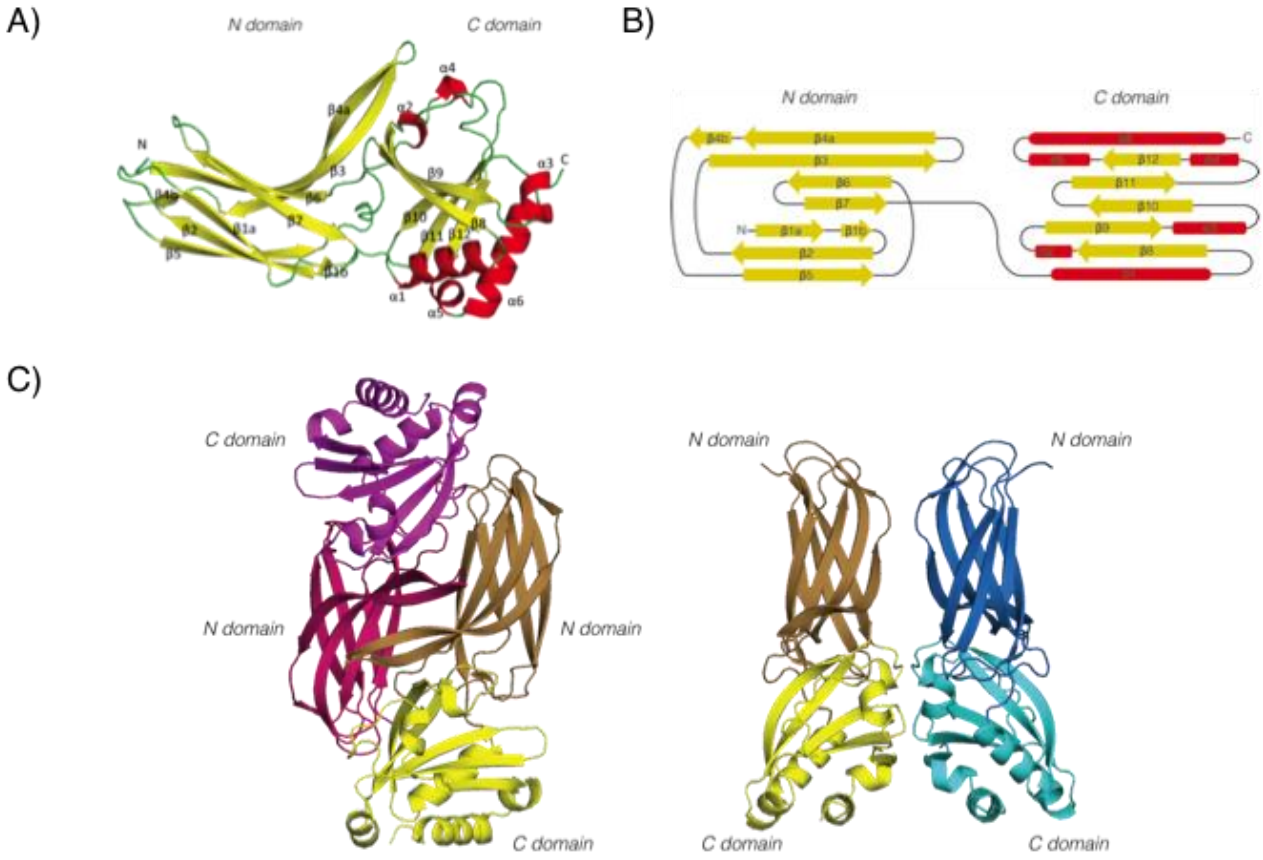

Figure 2.2 Overall structure of the *B. subtilis* Spo0M$^{11\text{-}258}$ protein. (A) Cartoon representation of the *B. subtilis* Spo0M$^{11\text{-}258}$. (B) Topology diagram of the Spo0M$^{11\text{-}258}$ protein. (C) Two distinct dimeric interactions in the Spo0M$^{11\text{-}258}$ crystal. Dimers formed through N and C-interface is shown at left and right panel, respectively.

Although I have attempted to solve the structure of Spo0M^{FL} by molecular replacement with a Spo0M$^{11\text{-}258}$ as a search model, appropriate phase was not determined.

The space group of Spo0M^{FL} crystal is *P*1 and the number of molecule in an asymmetric unit is estimated approximately 16 - 20 molecules, which might cause difficulty solving the Spo0M^{FL} crystal structure. Thus I could not conclude the oligomieric state of Spo0M in solution in this study.

Structural homology of the N domain of Spo0M to arrestin-like folded proteins

The N domain of Spo0M is composed of a seven-stranded β-sandwich (Figure 2.2A and 2.2B), which closely resembles that of the other arrestin-clan members, especially those belonging to α-arrestins. Human α-arrestins are composed of ARRDC1 - 5 and TXNIP. Among the members of α-arrestin, only the crystal structures of the N domain of ARRDC3 and TXNIP and the overall structure of TXNIP have been determined to date.[49-51] The structural and sequence alignments of the N domain of Spo0M, TXNIP and ARRDC3 are shown in Figure 2.3. In spite of their extremely poor sequence similarities, the N domain of Spo0M$^{11-258}$ was superimposed onto those of TXNIP (PDB ID; 4GEJ) and ARRDC3 (PDB ID; 4R7X), with RMSDs of 2.09 and 2.04 Å for 108 and 109 Cα atoms, respectively (Figure 2.3A and 2.3B). In contrast to the β/v-arrestins, α-arrestins are thought to share conserved amino acid sequences, being found many of eukaryotic genomes.[29] The structure of the Spo0M N domain revealed in this study indicates that the α-arrestin-like β-sandwich fold is structurally conserved from bacteria to mammals. One of the best studied functions of eukaryotic arrestin is the regulation of endocytosis of membrane proteins, such as GPCRs.[44] Although there is no endocytosis event on bacterial cell membranes, Spo0M might play a role in membrane trafficking to initiate sporulation that is triggered by environmental change. Eukaryotic arrestins are also known to act as scaffolding proteins facilitating signal transductions and to interact with multiple binding partners. For instance, a

number of proteins related to signal transductions, which include c-Src, MAPK, Mdm2, ASK1 and JNK3, have been reported to bind with β-arrestin.[63] In chapter 1, I have revealed that Spo0M directly binds with CcpA and inhibits the DNA-binding of CcpA to *cre* site, indicating that the Spo0M protein might play a role in the regulation of gene expression related to catabolite metabolism during the initiation step of sporulation by interacting with the CcpA protein. In contrast to the N domain, which has significant structural similarity to other arrestin-clan proteins, no arrestin-like folded protein that bares significant structural homology to the C domain of Spo0M was identified in the Dali server [64] and PDBeFold server.[59] The eukaryotic arrestin structures determined so far comprise two-lobe β-sandwich domains that are related by a pseudo two-fold rotation axis. The Spo0M structure reported here is the first example of an arrestin-clan protein that has only one half of the two-lobe domain, which is linked with a non-arrestin domain. Indeed, a potential arrestin-like folded protein that is assumed to have only half of two-lobe domains like Spo0M was predicted in the *Dictyostelium discoideum* genome.[65]

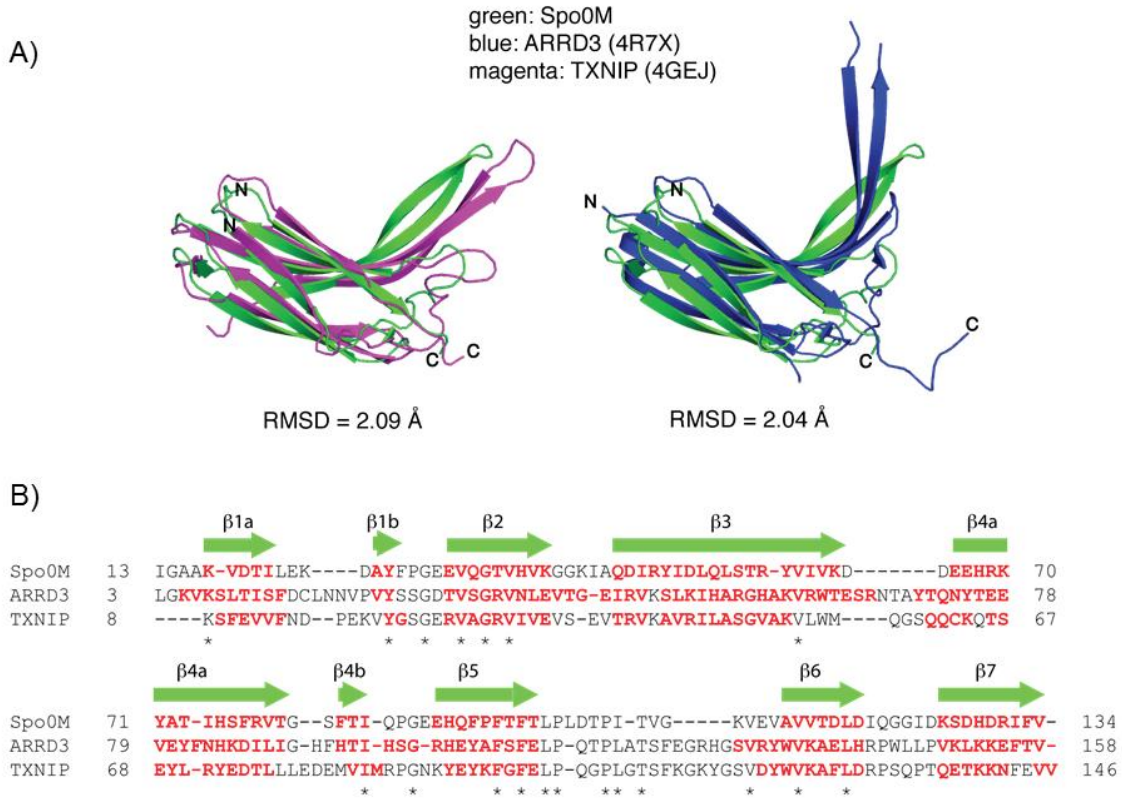

Figure 2.3 Comparison of the N domain of Spo0M$^{11\text{-}258}$ to other arrestin-clan proteins. (A) Superposition of the N domain of Spo0M$^{11\text{-}258}$ (green) with that of human TXNIP (PDB ID; 4GEJ, magenta) and ARDDC3 (PDB ID; 4R7X, blue), respectively. (B) Sequence alignment of the N domains of *B. subtilis* Spo0M to those of human TXNIP and ARRDC3. β-sheet structure elements are indicated in red and by cartoons above the sequences. Conserved amino acids are indicated by asterisks below the sequences.

Structural homology between the C domain of Spo0M and FP domain of PI31

Surprisingly, the dimerization domain, called the FP domain, of the human proteasome inhibitor protein subunit PI31, also known as PSMF1, has been shown to bear significant structural homology to the C domain of Spo0M (Figure 2.4A). The C domain of

Spo0M$^{11\text{-}258}$ was superimposed onto the FP domain of PI31 (PDB ID; 2VT8) with an RMSD of 2.09 Å for 124 Cα atoms (Figure 2.4A). PI31 mediates the maturation of immunoproteasome formation, which is an essential proteasome subtype for the mammalian immune system.[66] Only two proteins, PI31 and F-box only protein 7 (Fbxo7), have been shown to harbor the FP domain so far. From the crystal structure and biochemical studies of the PI31 protein, PI31 is suggested to form a homodimer in solution via the α-helices of its FP domains (α-interface). In addition, PI31 forms a homodimer in crystal via the β-strands of FP domains (β-interface). The β-interface has also been suggested to contribute to the formation of a heterodimer with Fbxo7 in cells. Interestingly, while the α-helix of PI31, which forms an α-interface, was not conserved between PI31 and Spo0M, the β-strand forming a β-interface in PI31 crystal did seem to be structurally conserved in the C-interface in Spo0M$^{11\text{-}258}$ crystal (Figure 2.4B). These findings suggested that the interface was a conserved dimerization interface between *B. subtilis* Spo0M and human PI31 and that a common evolutionary origin might exist and Spo0M might have functions similar to mammalian FP domains.

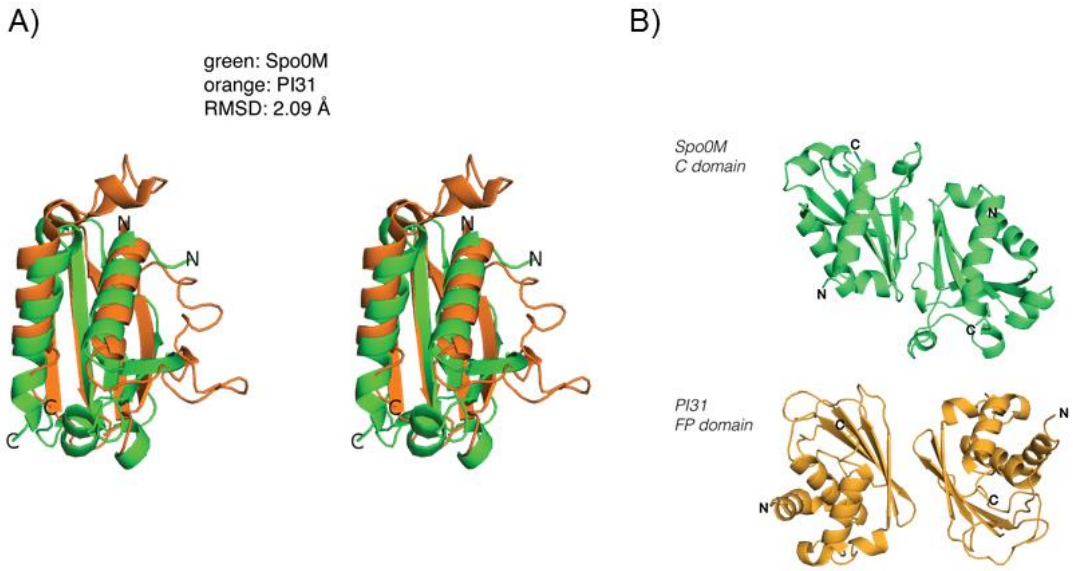

Figure 2.4 Comparison of the C domain of Spo0M$^{11-258}$ to the FP domain of PI31. (A) Superposition of the N domain of Spo0M$^{11-258}$ (green) with the FP domain of human PI31 (PDB ID; 2VT8, orange). In PI31 structure, first nineteen amino acids of N terminus containing an α-helix were removed to clarify the view. (B) C-interface of Spo0M$^{11-258}$ (top) and β-interface of PI31 (bottom) forming a dimer in crystal lattice.

Potential polar core structure of Spo0M

The polar core is a structurally and functionally important feature of arrestin, and is found in the crystal structures of β/v-arrestins and VPS26 [48] but not in those of TXNIP.[50] As shown in Figure 2.5, Spo0M was assumed to harbor a polar core structure located at the interface between the N and C domains, which is represented by charge-charge interactions. The charge-charge interactions include the N domain residues Glu67 (β4), Arg69 (β4) and Asp102 (loop between β5 and β6) and C domain residues Asp158 (β8), Arg168 (loop between α2 and β9), Gln173 (β9) and Arg216 (loop between β11 and α4). The potential polar

core of Spo0M$^{11\text{-}258}$ was derived into two parts as reported in the crystal structure of the VPS26 protein (Figure 2.5A).[48] The first part of the network involves the side-chains of Glu67 and Arg69 of the N domain and Asp158 and Arg216 of the C domain, and main chain of Val60 of the N domain and Cys159 and Gln161 of the C domain (Figure 2.5B). The second part of the network involves the side-chains of Asp102 of the N domain and Arg168 and Gln173 of the C domain, and the main chain of Ile74 and Ile105 of the N domain (Figure 2.5C). As demonstrated in other arrestin-clan members, this potential polar core structure of Spo0M may play a critical role in the function of the protein.

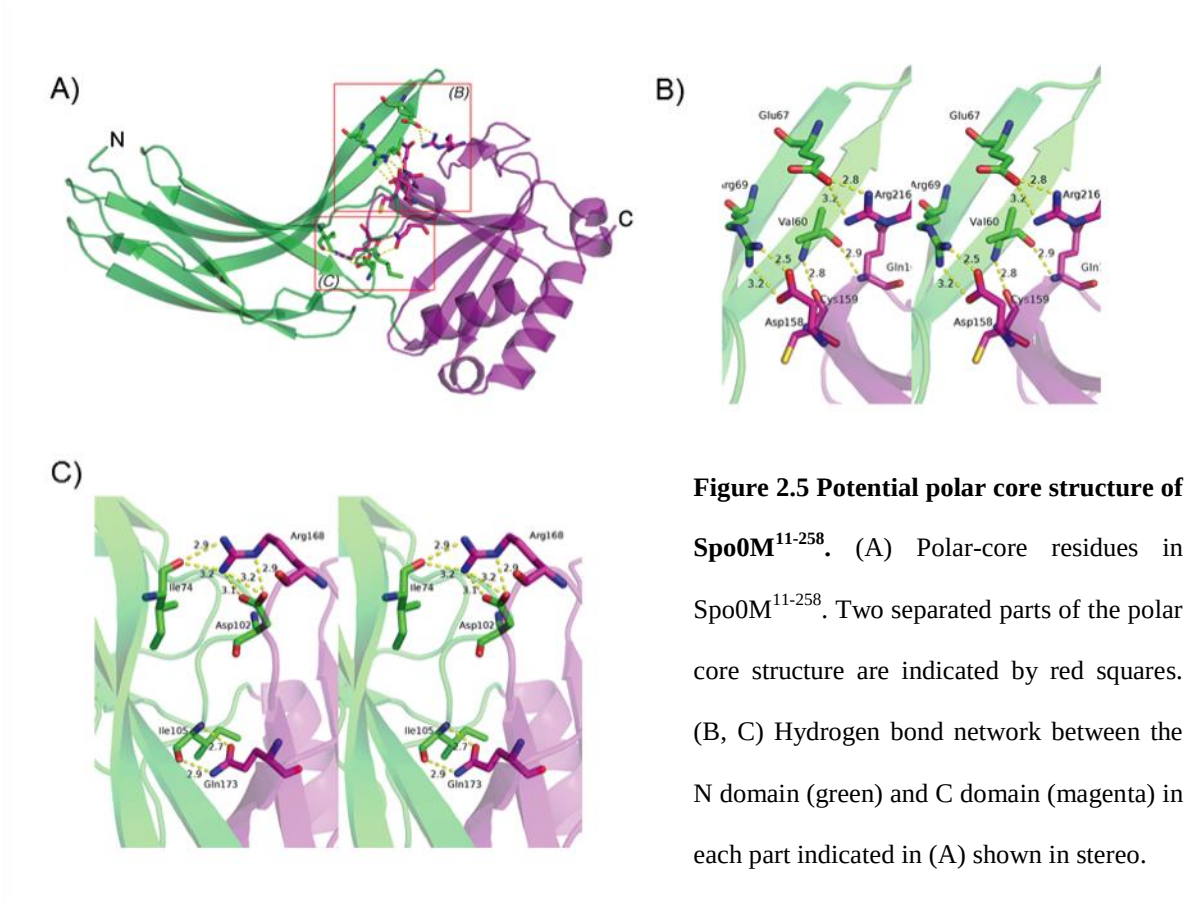

Figure 2.5 Potential polar core structure of Spo0M$^{11\text{-}258}$. (A) Polar-core residues in Spo0M$^{11\text{-}258}$. Two separated parts of the polar core structure are indicated by red squares. (B, C) Hydrogen bond network between the N domain (green) and C domain (magenta) in each part indicated in (A) shown in stereo.

Conclusion

The Spo0M proteins were successfully crystallized. Although the initial attempt to solve the structure of the $Spo0M^{FL}$ protein failed due to phase problems and poor diffraction, the truncation mutant screening using the DSF method resulted in a more stable $Spo0M^{11-258}$ protein lacking 10 amino acids on the N-terminus, which had a Tm 9 $^{\circ}$C higher than that of $Spo0M^{FL}$. The crystal structure of the $Spo0M^{11-258}$ was successfully determined by the SAD method. The N domain of Spo0M adopts a typical arrestin-like fold that in particular resembles α-arrestins, in which human ARRDC1 - 5 and TXNIP are involved, while sharing an extremely poor level of sequence identity. This finding indicates that the β-sandwich fold is a structurally conserved feature in the arrestin-clan from bacteria to mammals and Spo0M may share some functions with mammalian arrestin-clan members. In contrast to the N domain, the C domain of Spo0M had no structural homology to arrestin-clan proteins. These findings about the Spo0M N domain indicate that Spo0M is the first bacterial member found to belong to the arrestin-clan but also that there are arrestin-clan members that have only one β-sandwich domain. Since a potential arrestin-clan protein with one half of the two-lobe domain has also been predicted in other organisms [65], the ancient arrestin might have a monodomain structure. Surprisingly, we found that the FP domain of human PI31 had a significant structural homology to the C domain of Spo0M, suggesting that there may be a common evolutionary origin between the Spo0M C domain and mammalian FP domains.

Arrestin is known to interact directly with various proteins to regulate membrane trafficking as well as signal transduction.[44] Also, the FP domain of PI31 is also reported to mediate protein-protein interactions to form a homodimer and heterodimer with the FP domain of Fbxo7.[66] These facts suggested that, like arrestins and FP domains, Spo0M might function *in vivo* via protein-protein interactions. A functional study will be needed to

clear the role of each domain and determine whether there may also be functional similarities between bacterial Spo0M and mammalian arrestin-clan proteins as well as proteins containing the FP domain.

CONCLUSION OF THIS STUDY

Spo0M in *B. subtilis* has been successfully expressed, purified and crystallized. I have revealed that Spo0M directly bound to CcpA. The binding affinity was enhanced in the presence of HPr-Ser-P. Furthermore, Spo0M inhibits DNA binding of CcpA/HPr-Ser-P. These facts suggest that Spo0M might be involved in CCR and inhibit DNA-binding function of CcpA to release of CcpA-mediated catabolite repression. Although flagellin homologue protein called as Hag was indentified a potential binding partner of Spo0M, actual interaction was not clarified *in vitro* study.

While the structure of wild type Spo0M was failed to be determined, the crystal structure of thermostable truncation mutant $Spo0M^{11-258}$ protein was successfully determined by the SAD method. As suggested previously, the N domain of Spo0M adopts a typical arrestin-like fold that in particular resembles α-arrestins. Surprisingly, I found that the FP domain of human PI31 had a significant structural homology to the C domain of Spo0M. These indicate that Spo0M is the first bacterial member found to have an arrestin-like folded domain and FP-like folded domain, and that Spo0M may have common evolutionary origin with mammalian arrestins and FP domains.

Further functional study will be needed to clear the role of each domain and determine whether there may also be functional similarities between bacterial Spo0M and mammalian arrestin-clan proteins as well as proteins containing the FP domain.

REFERENCES

1. Nicholson, W.L., et al., *Resistance of Bacillus endospores to extreme terrestrial and extraterrestrial environments.* Microbiol Mol Biol Rev, 2000. **64**(3): p. 548-72.
2. Piggot, P.J. and J.G. Coote, *Genetic aspects of bacterial endospore formation.* Bacteriol Rev, 1976. **40**(4): p. 908-62.
3. Piggot, P.J. and D.W. Hilbert, *Sporulation of Bacillus subtilis.* Curr Opin Microbiol, 2004. **7**(6): p. 579-86.
4. Stragier, P. and R. Losick, *Molecular genetics of sporulation in Bacillus subtilis.* Annu Rev Genet, 1996. **30**: p. 297-41.
5. Dubnau, E.J., K. Cabane, and I. Smith, *Regulation of spo0H, an early sporulation gene in bacilli.* J Bacteriol, 1987. **169**(3): p. 1182-91.
6. Sonenshein, A.L., *Control of sporulation initiation in Bacillus subtilis*. Curr Opin Microbiol, 2000. **3**(6): p. 561-6.
7. Weir, J., et al., *Regulation of spo0H, a gene coding for the Bacillus subtilis sigma H factor.* J Bacteriol, 1991. **173**(2): p. 521-9.
8. Han, W.D., et al., *A novel sporulation-control gene (spo0M) of Bacillus subtilis with a sigmaH-regulated promoter.* Gene, 1998. **217**(1-2): p. 31-40.
9. Predich, M., G. Nair, and I. Smith, *Bacillus subtilis early sporulation genes kinA, spo0F, and spo0A are transcribed by the RNA polymerase containing sigma H.* J Bacteriol, 1992. **174**(9): p. 2771-8.

10. Sun, D.X., P. Stragier, and P. Setlow, *Identification of a new sigma-factor involved in compartmentalized gene expression during sporulation of Bacillus subtilis.* Genes Dev, 1989. **3**(2): p. 141-9.

11. Stragier, P., C. Bonamy, and C. Karmazyn-Campelli, *Processing of a sporulation sigma factor in Bacillus subtilis: how morphological structure could control gene expression.* Cell, 1988. **52**(5): p. 697-704.

12. Kroos, L., B. Kunkel, and R. Losick, *Switch protein alters specificity of RNA polymerase containing a compartment-specific sigma factor.* Science, 1989. **243**(4890): p. 526-9.

13. Errington, J., *Regulation of endospore formation in Bacillus subtilis.* Nat Rev Microbiol, 2003. **1**(2): p. 117-26.

14. Görke, B. and J. Stülke, *Carbon catabolite repression in bacteria: many ways to make the most out of nutrients.* Nat Rev Microbiol, 2008. **6**(8): p. 613-24.

15. Singh, K.D., et al., *Carbon catabolite repression in Bacillus subtilis: quantitative analysis of repression exerted by different carbon sources.* J Bacteriol, 2008. **190**(21): p. 7275-84.

16. Kleijn, R.J., et al., *Metabolic fluxes during strong carbon catabolite repression by malate in Bacillus subtilis.* J Biol Chem, 2010. **285**(3): p. 1587-96.

17. Schumacher, M.A., et al., *Structures of carbon catabolite protein A-(HPr-Ser46-P) bound to diverse catabolite response element sites reveal the basis for high-affinity binding to degenerate DNA operators.* Nucleic Acids Res, 2011. **39**(7): p. 2931-42.

18. Weickert, M.J. and G.H. Chambliss, *Site-directed mutagenesis of a catabolite repression operator sequence in Bacillus subtilis.* Proc Natl Acad Sci U S A, 1990. **87**(16): p. 6238-42.

19. Jacob, S., et al., *Catabolite repression of the operon for xylose utilization from Bacillus subtilis W23 is mediated at the level of transcription and depends on a cis site in the xylA reading frame.* Mol Gen Genet, 1991. **229**(2): p. 189-96.

20. Miwa, Y. and Y. Fujita, *Involvement of two distinct catabolite-responsive elements in catabolite repression of the Bacillus subtilis myo-inositol (iol) operon.* J Bacteriol, 2001. **183**(20): p. 5877-84.

21. Shafikhani, S.H., A.A. Partovi, and T. Leighton, *Catabolite-induced repression of sporulation in Bacillus subtilis.* Curr Microbiol, 2003. **47**(4): p. 300-8.

22. Miwa, Y., M. Saikawa, and Y. Fujita, *Possible function and some properties of the CcpA protein of Bacillus subtilis.* Microbiology, 1994. **140 (Pt 10)**: p. 2567-75.

23. Galinier, A., et al., *New protein kinase and protein phosphatase families mediate signal transduction in bacterial catabolite repression.* Proc Natl Acad Sci U S A, 1998. **95**(4): p. 1823-8.

24. Jault, J.M., et al., *The HPr kinase from Bacillus subtilis is a homo-oligomeric enzyme which exhibits strong positive cooperativity for nucleotide and fructose 1,6-bisphosphate binding.* J Biol Chem, 2000. **275**(3): p. 1773-80.

25. Lysenko, E., T. Ogura, and S.M. Cutting, *Characterization of the ftsH gene of Bacillus subtilis.* Microbiology, 1997. **143 (Pt 3)**: p. 971-8.

26. Deuerling, E., et al., *The ftsH gene of Bacillus subtilis is involved in major cellular processes such as sporulation, stress adaptation and secretion.* Mol Microbiol, 1997. **23**(5): p. 921-33.

27. Wehrl, W., M. Niederweis, and W. Schumann, *The FtsH protein accumulates at the septum of Bacillus subtilis during cell division and sporulation.* J Bacteriol, 2000. **182**(13): p. 3870-3.

28. Thi Nguyen, H.B. and W. Schumann, *The sporulation control gene spo0M of Bacillus subtilis is a target of the FtsH metalloprotease.* Res Microbiol, 2012. **163**(2): p. 114-8.

29. Alvarez, C.E., *On the origins of arrestin and rhodopsin.* BMC Evol Biol, 2008. **8**: p. 222.

30. Wünsche, A., et al., *CcpA forms complexes with CodY and RpoA in Bacillus subtilis.* FEBS J, 2012. **279**(12): p. 2201-14.

31. Fawcett, P., et al., *The transcriptional profile of early to middle sporulation in Bacillus subtilis.* Proc Natl Acad Sci U S A, 2000. **97**(14): p. 8063-8.

32. Molle, V., et al., *The Spo0A regulon of Bacillus subtilis.* Mol Microbiol, 2003. **50**(5): p. 1683-701.

33. Phillips, Z.E. and M.A. Strauch, *Bacillus subtilis sporulation and stationary phase gene expression.* Cell Mol Life Sci, 2002. **59**(3): p. 392-402.

34. Strauch, M., et al., *The SpoOA protein of Bacillus subtilis is a repressor of the abrB gene.* Proc Natl Acad Sci U S A, 1990. **87**(5): p. 1801-5.

35. Wu, J.J., R. Schuch, and P.J. Piggot, *Characterization of a Bacillus subtilis sporulation operon that includes genes for an RNA polymerase sigma factor and for a putative DD-carboxypeptidase.* J Bacteriol, 1992. **174**(15): p. 4885-92.

36. Zhukov, A., et al., *Integration of surface plasmon resonance with mass spectrometry: automated ligand fishing and sample preparation for MALDI MS using a Biacore 3000 biosensor.* J Biomol Tech, 2004. **15**(2): p. 112-9.

37. Seidel, G., et al., *Quantitative interdependence of coeffectors, CcpA and cre in carbon catabolite regulation of Bacillus subtilis.* FEBS J, 2005. **272**(10): p. 2566-77.

38. Kraus, A., et al., *Catabolite repression of the Bacillus subtilis xyl operon involves a cis element functional in the context of an unrelated sequence, and glucose exerts additional xylR-dependent repression.* J Bacteriol, 1994. **176**(6): p. 1738-45.

39. Mukherjee, S. and D.B. Kearns, *The structure and regulation of flagella in Bacillus subtilis.* Annu Rev Genet, 2014. **48**: p. 319-40.

40. Wakabayashi, K., H. Hotani, and S. Asakura, *Polymerization of Salmonella flagellin in the presence of high concentrations of salts.* Biochim Biophys Acta, 1969. **175**(1): p. 195-203.

41. Maruyama, Y., et al., *Crystal structure of a novel bacterial cell-surface flagellin binding to a polysaccharide.* Biochemistry, 2008. **47**(5): p. 1393-402.

42. Kang, D.S., X. Tian, and J.L. Benovic, *Role of β-arrestins and arrestin domain-containing proteins in G protein-coupled receptor trafficking.* Curr Opin Cell Biol, 2014. **27**: p. 63-71.

43. Patwari, P. and R.T. Lee, *An expanded family of arrestins regulate metabolism.* Trends Endocrinol Metab, 2012. **23**(5): p. 216-22.

44. Tian, X., D.S. Kang, and J.L. Benovic, *β-arrestins and G protein-coupled receptor trafficking.* Handb Exp Pharmacol, 2014. **219**: p. 173-86.

45. Hirsch, J.A., et al., *The 2.8 A crystal structure of visual arrestin: a model for arrestin's regulation.* Cell, 1999. **97**(2): p. 257-69.

46. Han, M., et al., *Crystal structure of beta-arrestin at 1.9 A: possible mechanism of receptor binding and membrane Translocation.* Structure, 2001. **9**(9): p. 869-80.

47. Sutton, R.B., et al., *Crystal structure of cone arrestin at 2.3A: evolution of receptor specificity.* J Mol Biol, 2005. **354**(5): p. 1069-80.

48. Shi, H., et al., *The retromer subunit Vps26 has an arrestin fold and binds Vps35 through its C-terminal domain.* Nat Struct Mol Biol, 2006. **13**(6): p. 540-8.

49. Polekhina, G., et al., *Structure of the N-terminal domain of human thioredoxin-interacting protein.* Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 3): p. 333-44.

50. Hwang, J., et al., *The structural basis for the negative regulation of thioredoxin by thioredoxin-interacting protein.* Nat Commun, 2014. **5**: p. 2958.

51. Qi, S., et al., *Insights into β2-adrenergic receptor binding from structures of the N-terminal lobe of ARRDC3.* Protein Sci, 2014. **23**(12): p. 1708-16.

52. Otwinowski, Z. and W. Minor, ***Processing of X-ray Diffraction Data Collected in Oscillation Mode***. 1997, *Methods in Enzymology*. p. 307-326.

53. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.

54. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.

55. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.

56. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* J Mol Biol, 2007. **372**(3): p. 774-97.

57. Ericsson, U.B., et al., *Thermofluor-based high-throughput stability optimization of proteins for structural studies.* Anal Biochem, 2006. **357**(2): p. 289-98.

58. Sonoda, Y., et al., *Benchmarking membrane protein detergent stability for improving throughput of high-resolution X-ray structures.* Structure, 2011. **19**(1): p. 17-25.

59. Krissinel, E. and K. Henrick, *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2256-68.

60. He, L. and B. Niemeyer, *A novel correlation for protein diffusion coefficients based on molecular weight and radius of gyration.* Biotechnol Prog, 2003. **19**(2): p. 544-8.

61. Lobanov, M.I.u., N.S. Bogatyreva, and O.V. Galzitskaia, *[Radius of gyration is indicator of compactness of protein structure]*. Mol Biol (Mosk), 2008. **42**(4): p. 701-6.

62. Ortega, A., D. Amorós, and J. García de la Torre, *Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models.* Biophys J, 2011. **101**(4): p. 892-8.

63. Lefkowitz, R.J. and S.K. Shenoy, *Transduction of receptor signals by beta-arrestins.* Science, 2005. **308**(5721): p. 512-7.

64. Holm, L. and P. Rosenström, *Dali server: conservation mapping in 3D.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W545-9.

65. Aubry, L., D. Guetta, and G. Klein, *The arrestin fold: variations on a theme.* Curr Genomics, 2009. **10**(2): p. 133-42.

66. Kirk, R., et al., *Structure of a conserved dimerization domain within the F-box protein Fbxo7 and the PI31 proteasome inhibitor.* J Biol Chem, 2008. **283**(32): p. 22325-35.

ACKNOWLEDGEMENTS

The author is deeply indebted to Prof. Bunzo Mikami, Gradutate School of Agriculture, Kyoto University for his kind guidance and encouragement during this study.

The author is grateful to Dr. Kimihiko Mizutani, Assistant Professor of Gradutate School of Agriculture, Kyoto University for his valuable suggestions and helps.

Thanks are given all the members of laboratories of Applied Structural Biology, Graduate School of Agriculture, Kyoto University and colleagues of Drug Research Center of Kaken Pharmaceutical Co., Ltd.

X-ray diffraction data were collected at SPring-8 (proposal numbers 2013B6849, 2014A1308, 2014A1279, 2014A1279, 2014B1321, 2014B6948 and 2014B1274), with assistance from beamline scientists.

Yo Sonoda

LIST OF PUBLICATIONS

Original papers

1. Yo Sonoda, Kimihiko Mizutani and Bunzo Mikami. (2015). Crystal structure of Spo0M, a sporulation control protein from *Bacillus subtilis.* ***Acta. Crystallogr. Sect. F Struct. Biol. Cryst. Commun***. **71**(Pt 12), 1488-97.
2. Yo Sonoda, Kimihiko Mizutani and Bunzo Mikami. (2015). Sporulation control protein, Spo0M, inhibits DNA binding of catabolite control protein A via direct interaction. *To be submitted.*

Related papers

1 Norimichi Nomura, Grégory Verdon, Hae Joo Kang, Tatsuro Shimamura, Yayoi Nomura, Yo Sonoda, Saba Abdul Hussien, Aziz Abdul Qureshi, Mathieu Coincon, Yumi Sato, Hitomi Abe, Yoshiko Nakada-Nakura, Tomoya Hino, Takatoshi Arakawa, Osamu Kusano-Arai, Hiroko Iwanari, Takeshi Murata, Takuya Kobayashi, Takao Hamakubo, Michihiro Kasahara, So Iwata and David Drew. (2015). Structure and mechanism of the mammalian fructose transporter GLUT5. ***Nature***. **526**, 397-401.

2 Yusuke Endo, Kentaro Kawai, Takeshi Asano, Seiji Amano, Yoshihito Asanuma, Keisuke Sawada, Yuuta Onodera, Noriko Ueo, Nobuaki Takahashi, Yo Sonoda, Noriyuki Kamei and Tetsumi Irie. (2015). 2-(Isopropylamino)thieno[3,2-d]pyrimidin-4(3H)-one derivatives as selective phosphodiesterase 7 inhibitors with potent in vivo efficacy. ***Bioorg Med Chem Lett.*** **25**, 1910-1914

3 Yusuke Endo, Kentaro Kawai, Takeshi Asano, Seiji Amano, Yoshihito Asanuma, Keisuke Sawada, Keiji Ogura, Naoya Nagata, Noriko Ueo, Nobuaki Takahashi, Yo Sonoda and Noriyuki Kamei. (2015). Discovery and SAR study of 2-(4-pyridylamino)thieno[3,2-d]pyrimidin-4(3H)- ones as soluble and highly potent PDE7 inhibitors. ***Bioorg Med Chem Lett.*** **25**, 649-653.

4 Kentaro Kawai, Yusuke Endo, Takeshi Asano, Seiji Amano, Keisuke Sawada, Noriko Ueo, Nobuaki Takahashi, Yo Sonoda, Mika Nagai, Noriyuki Kamei, and Naoya Nagata. (2014). Discovery of 2-(cyclopentylamino)thieno[3,2-d]pyrimidin-4(3H)-one derivatives as a new series of potent phosphodiesterase 7 inhibitors**.** ***J Med Chem*****.** **57**, 9844-9854

5 Yo Sonoda, Simon Newstead, Nien-Jen Hu, Yilmaz Alguel, Emmanuel Nji, Konstantinos Beis, Shoko Yashiro, Chiara Lee, James Leung, Alexander D. Cameron, Bernadette Byrne, So Iwata and David Drew (2011). Benchmarking membrane protein detergent stability for improving throughput of high-resolution X-ray structures. ***Structure*****. 19**, 17-25.

6 Yo Sonoda, Alex Cameron, Simon Newstead, Hiroshi Omote, Yoshinori Moriyama, Michihiro Kasahara, So Iwata and David Drew. (2010). Tricks of the trade used to accelerate high-resolution structure determination of membrane proteins. ***FEBS Lett.*** **584**, 2539-2547.

7 David Drew, Simon Newstead, Yo Sonoda, Hyun Kim, Gunnar von Heijne and So Iwata. (2008). GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. ***Nat Protoc.*** **3**, 784-798.