농학석사학위논문

Salmonella Typhimurium 유래 effector GtgA 초기결정화 연구

Initial crystallization of GtgA from *Salmonella* Typhimurium

2017년 2월

서울대학교 대학원

농생명공학부 응용생명화학전공

조 예 림

Salmonella Typhimurium 유래 effector GtgA 초기결정화 연구

Initial crystallization of GtgA from *Salmonella* Typhimurium

지도교수 이 상 기

이 논문을 농학석사학위논문으로 제출함

2017년 1월

서울대학교 대학원

농생명공학부 응용생명화학전공

조 예 림

조예림의 석사학위논문을 인준함

2017년 1월

위 원 장 _____ (인)

부 위 원 장 _____ (인)

위 원 _____ (인)

Abstract

Salmonella Typhimurium is a pathogen of some mammals including human. Its pathogenic behaviors are first emerging by delivering its own bacterial proteins, known as effector, into host, but *Salmonella* confronts host defense system, leading to inflammation in epithelial cells of hosts. Inflammation is one of host defense response to pathogens and thus prevents further proliferation of bacteria. In human, a nuclear factor κB (NF-κB) has been known to play a key role in regulating inflammation by mediating transcriptional activation of proinflammatory cytokine production. To circumvent or counteract the NF-κB dependent defense system in human, *Salmonella* deploys three different effector proteins that all belong to members of the PipA family. Proteins in this family are characterized by a zinc-dependent protease. In fact, GtgA, one of *Salmonella* effectors, exhibited a protease activity against NF-κB by cleaving essential residues for DNA binding. This GtgA-mediated NF-κB degradation eventually led to decreasing proinflammatory cytokines. However, molecular details regarding a mode of action by GtgA remain largely uncharacterized. In order to identify structural features of GtgA, I performed extensive experimental approaches including expression of GtgA in various forms, crystallization of those proteins, and X-ray diffraction analysis. Initial experiment indicates that GtgA in its full length remains highly soluble but resists crystallization. Therefore, its hydrodynamic properties were modified by replacing a cluster of hydrophilic residues with hydrophobic ones. Among five constructs, two produced GtgA in insoluble form, but three resulted in a soluble GtgA. Those soluble GtgAs were further subject to purification and crystallization under various conditions. Initial condition for GtgA crystallization was recently

observed and the X-ray diffraction data using those crystals were successfully collected at 2.5 Å-resolution. Currently, a molecular replacement using a template structure with amino acid sequence of 20% identity however failed to deliver a GtgA structure. Now, several experiments are carrying out to determine the phase information of GtgA. Structural properties of GtgA could provide valuable information of host-pathogen interactions.

Key words: GtgA, nuclear factor κB, *Salmonella* Typhimurium, X-ray diffraction, zinc-metalloprotease

***Student Number*:** 2015-21797

Contents

List of Figures

List of Tables

Introduction

Pathogenic bacterium Salmonella enterica infects different hosts depending on its own serotype. Among those various serotypes, Typhimurium selectively invades mammals including human, cattle, swine and horse (Rabsch et al., 2002), and causes gastroenteritis of hosts by penetrating into intestinal epithelial cells (Takeuchi, 1967).

In general, when pathogens reside in human intestine, symbiotic microbes pre-occupied a host intestine and could be a hindrance to pathogen colonization. Surprisingly, under these circumstances, Salmonella was evolved to utilize an inflammatory response of host as a way to suppress the pre-existing symbiotic bacteria (Stecher et al., 2007). Inflammation is one of typical host defense responses against pathogen infection, but Salmonella survives during defense response by attenuating an inflammation process. In previous studies, several Salmonella proteins which promote host inflammation were revealed. One of the ways to cause inflammation is activating transcription factors responsible for producing the proinflammatory cytokines (Hobbie et al., 1997; Stecher et al., 2007). Nuclear factor κB (NF-κB) represents one of those transcription factors. Its activity remained repressed by forming a complex with inhibitory protein. Although immune system leading to inflammation does not work against symbiotic microbes, inflammation is caused by degradation of inhibitory protein upon Salmonella infection. Following that, NF-κB in the complex with inhibitor becomes released and translocated into nucleus to fully function as transcription factor for expressing proinflammatory cytokines (Lawrence, 2009).

Salmonella-dependent inflammation and its subsequent survival during

inflammation are possible in host due to Salmonella's own proteins called effectors. Those effectors are transmitted into host by type III protein secretion system (TTSS) and many TTSS effectors are known to be involved in virulence of pathogen (Alfano and Collmer, 2004). Among Salmonella TTSS effectors, three proteins including PipA, GogA and GtgA, were found to belong to members of the PipA family proteins (Sun et al., 2016). Those proteins in the PipA family are zinc-dependent protease, with a zinc binding motif conserved among other zinc-metalloproteases. It is HExxH motif known as essential residues for catalytic activity. For example, effector PipA cleaves two different NF-κB transcription factors at the nucleus: RelA and RelB. Recent study revealed that a peptide bond between Gly40 and Arg41 is subject to PipA's action (Sun et al., 2016). Decrease of proinflammatory cytokines by degradation of NF-κB results in suppression of host inflammation. Because Salmonella profits on its colonization from inflammation, the suppression of inflammation leads to reduction of virulence. However, it also enables host to preserve its homeostasis rather than die early. This suggests that Salmonella has been evolved to survive longer in the host.

Previous studies showed the structure of TTSS effector from Escherichia coli, NleC. It also cleaves the NF-κB transcription factor. It targets p300 and p50 which are other types of NF-κB as well as RelA and RelB (Baruch et al., 2011; Muhlen el al., 2011; Shames et al., 2011). They showed that NleC cleaves different site of RelA from that of PipA family. NleC cleaves between residues Cys38 and Glu39, whereas GtgA cleaves between residues Gly40 and Arg41. However, NleC has a lot of similarities to PipA. Like PipA, it is also zinc-metalloprotease and shows low sequence similarity to any known structures. In addition, it also contains HExxH motif, a conserved zinc binding motif, which GtgA contains. The surface of NleC near its active site shows electrostatic characteristic due to a lot of

glutamates. This results in complementarity with RelA which has positively charged surface for binding with DNA (Li et al., 2013; Turco and Sousa, 2014). In the same manner, a lot of glutamates in GtgA seem to be related to interaction with RelA and its nonpromiscuous characteristic.

In this study, I approached for crystallization of GtgA protein in various ways. In this process, I made several constructs, assessed their expression, and found the condition of purification. I obtained several crystals in a condition and collected diffraction data. Because attempts to solve the structure of GtgA by molecular replacement was unsuccessful, seleno-methionine labeled GtgA crystallization would help to solve the structure. The atomic structure of GtgA will give a lot of information about mechanism and substrate specificity to the target of GtgA and other PipA family of effectors.

Materials and Methods

1. Cloning

Gene of GtgA (GenBank accession number AIE06583.1) was synthesized by requiring to Bioneer. The full length and truncation mutant were amplified using polymerase chain reaction (PCR). Primers were designed for complementary hybridization to 3' and 5' of GtgA gene (Table 1). NdeI restriction site was added to forward primer and XhoI restriction site was added to reverse primer for cloning to expression vector. The template and primers were mixed to 20 μl of Pfu pre-mix (Enzynomics) which contains Pfu DNA polymerase, deoxynucleoside triphosphates and magnesium ion. The annealing temperature was 55°C and elongation was carried out for 41 s and 38 s respectively. PCR products were subcloned into pTOP vector (Enzynomics) and digested with NdeI and XhoI to obtain digested target sequence efficiently. They were obtained by agarose-gel electrophoresis using 1% agarose-gel and were subcloned again into the NdeI and XhoI restriction sites of multicloning site in pET28a vector (Merck). The vector contains five histidine tags and Tabacco Etch Virus (TEV) protease cleavage site on its N-terminus for purification of protein. Three of surface entropy reduction mutants (SERs) (Table 2) were made using QuikChange II site directed mutagenesis method and the template was wild type GtgA within N-terminally histidine tagged pET-28 vector. Primers used in this process are written in the table (Table 1) and elongation of PCR was carried out for 6.5 min. After PCR, DpnI enzyme (enzynomics) was treated for 2 hrs to digest wild type GtgA. Both of ligation and PCR products were transformed to Escherichia coli DH 10B (Invitrogen). From colonies incubated on the Luria-Bertani (LB) media plate,

plasmids could be extracted using Mini-Prep kit (MACHEREY-NAGEL). Then, sequences of gene subcloned to pET28a vector and mutants were identified by requesting to Macrogen.

2. Expression test

pET28 GtgA and its mutants were transformed into *Escherichia coli* BL21 (DE3) cells (Merck) to express the proteins. Cells containing pET28 GtgA were selected by adding 50 μg/ml of antibiotic, kanamycin. It was incubated in two 5 ml tubes of Luria-Bertani media at 37°C until the optical density of the cells at 600 nm became about 0.7. Then, one tube of two was expressed by adding 0.5 mM of isopropyl-β-D-thiogalactopyranoside (IPTG) and incubated at 20°C for 16 hrs. After that, the cells were harvested by centrifugation at 13,500 rpm for 1 min and resuspended with buffer containing 50 mM Tris (pH 8.0) and 100 mM NaCl. Cells were lysed using sonicator at 25% of amplitude for 1 min. The sample without IPTG treatment was used as control, the other with IPTG treatment was used as total fraction, and the supernatant of total fraction at 13,500 rpm for 10 min was used soluble fraction. The expression of proteins was evaluated by 15% of sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE).

3. Purification

Escherichia coli BL21 (DE3) cells containing pET28 GtgA or mutant were incubated in 2.4 L of Luria-Bertani media supplied kanamycin and proteins were expressed with IPTG as the same method of expression test. Cells were harvested by centrifugation at 5,000 rpm for 5 min, resuspended with 80 ml of buffer and lysed by sonication at 35% of amplitude for 30 min. Lysis buffer contained 50 mM Tris (pH 8.0) and 100 mM NaCl. The supernatant obtained from centrifugation of

lysate at 16,560 rpm for 1 hrs was applied to the HisTrap HP (GE Healthcare) and purified using immobilized metal affinity chromatography (IMAC). Additionally, full length of GtgA was also lysed within the buffer containing 1 mM of $ZnCl_2$. During IMAC, the buffer applied for cell lysis was used as A and buffer added with 500 mM of imidazole to buffer A was used as B. Column was washed with 5% of buffer B for 4 cv to remove unbound proteins after sample injection and eluted with 40% of buffer B for 3.8 cv. The N-terminal his tag was cleave by treating 20:1 molar ratio of TEV protease in 4°C for 16 hrs and 25°C for 3 hrs. At the same time, the proteins are dialyzed packed in the cellulose membrane to remove imidazole eluted together. The dialysis buffer was buffer A supplied with 2 mM of dithiothereitol which is necessary for activity of TEV protease. The resulting tag-free proteins are applied to IMAC column again and eluted at 5% of buffer B. After elution, the proteins were concentrated and subjected to gel filtration chromatography using HiLoad 16/600 Superdex 200 pg gel filtration column (GE Healthcare) to remove soluble aggregate and obtain the homologous proteins.

In addition to this standard purification method, the methylation of lysine within protein were also proceeded to alter the property of protein during crystallization. The methylation was applied to full length of GtgA protein. Protein was expressed and purified with IMAC with the same method as above except for buffer. Because Tris inhibits the methylation, buffer used for protein methylation was containing 50mM HEPES (pH8.0) and 100 mM NaCl. The protein purified with IMAC was added with 800 μl of 1 M dimethylamine borane complex and 1600 μl of 1 M formaldehyde in ice. After 2 hrs, 200 μl of 1 M dimethylamine borane complex was added again and leaved the reaction mixture at 4°C for 16 hrs. The reaction was stopped by adding 2.5 ml of 1 M Tris and dialyzed twice against 2 L of buffer containing 50 mM Tris (pH 8.0) and 100 mM NaCl for total 10 hrs.

The dialyzed sample was centrifuged and separated using gel filtration chromatography.

For the phase determination of protein, selenomethionyl protein was prepared which was applied to SER3. The cells were incubated in 50 ml of Luria-Bertani media for 8 hrs and harvested by centrifugation at 3,600 rpm for 10 min. The cells were resuspended with and incubated in M9 minimal medium consisted of M9 salts, 2 mM of $MgSO_4$, 0.1 mM of $CaCl_2$, 0.4% of glucose, and sterile H_2O. The cells grew at 37°C until optical density at 600 nm reached 0.6 and were supplied amino acid mixture. After 30 min, 0.25 mM of selenomethionine and 0.5 mM of IPTG were added and incubated at 20°C for 16 hrs. After that, the protein was purified with the same method as above.

4. Crystallization

The purified proteins were concentrated at 3,400 rpm and concentration was measured using Nanodrop 2000 (Thermo Scientific). The molecular weight and extinction coefficient of the proteins were 25.9 kD and 20400 respectively, which were calculated at the ExPASy server (http://web.expasy.org/protparam). Concentrated proteins were diluted to the range of 10 to 15 mg/ml. They were crystallized using sitting drop vapor diffusion method at 22°C in the six of 96 well sitting drop plates. Six screening kits including Index, PEG/Ion 1, 2 (Hampton research), PACT, pH clear, Protein complex (Qiagen), and Wizard 3, 4 (Rigaku) were used. 2 μl of protein and 2 μl of reagent were mixed in the sample drop well and 60 μl of reagent was applied into the reagent well and plates were sealed with HD CLEAR tape (Shur Tech Brands, LLC). The full length of GtgA was additionally crystallized with 1 mM $ZnCl_2$ containing cofactor ion; zinc. $ZnCl_2$ was dissolved in the buffer adjusted to 10 mM and diluted to 1 mM.

To make reservoirs to reproduce the obtained crystal, HEPES (sigma-aldrich) and $CaCl_2$ (sigma-aldrich) adjusted to 1M, and 100% (v/v) of polyethylene glycol (PEG) 400 (sigma-aldrich) were appropriately diluted and adjusted with diluted water. The pH of HEPES was adjusted to 7.2, 7.5, and 7.8 using NaOH. In the process of reproduce, additive screen which affects solubility or crystallization was used to obtain better crystal. The reservoir was containing 100 mM HEPES (pH 7.5), 30% PEG 400, 200 mM $CaCl_2$. It was diluted with the same stocks to become the concentration after mixing 10X of additive kit (Hampton research).

5. Data collection

X-ray data was collected at beamline 5C of the Pohang Accelerator Laboratory (Korea). The crystals were put in 20% of ethylene glycol, made by mixing 48 μl of reservoir and 12 μl of ethylene glycol, as the cryo-protectant. The temperature of data collection was 100 K with 1° of oscillation angle. For integration and Scaling of collected data, HKL 2000 (Otwinowski and Minor, 1997) was used and *PHENIX* (Adams et al., 2010) was used for molecular replacement to obtain electron density map.

Results

1. Cloning

For a successful crystallization of GtgA, various plasmids were constructed containing the native, full length GtgA of 228 residues, as well as several modified GtgA. Major modification is a deletion of the possible disordered region(s) or mutation of selected residue(s) for surface entropy reduction. All these changes are intended for altering hydrodynamic properties of GtgA and increasing possibility of crystallization.

Sequence analysis of GtgA was carried out using a program XtalPred (http://ffas.burnham.org/XtalPred-cgi/xtal.pl), in which a possible disordered region(s) was predicted based on twenty-five members of the PipA family proteins (Figure 1a). Specifically, N-terminal twenty-two residues in GtgA are expected to be loop (Figure 1b), and beyond His23 those residues are a part of well-conserved domain in the PipA family. Therefore, those twenty-two residues in the N-terminal region were deleted to produce a modified GtgA, referred as GtgA(H23).

In a second approach, GtgA was subject to mutation that causes changes in hydrodynamic features by reducing its surface entropy. In particular, flexible and charged residues, such as arginine, lysine, glutamate, aspartate, on the surface of protein were mutated into alanine. As a result, three mutants were produced and those proteins having mutations for surface entropy reduction are called SERs (Figure 2a). In SER, candidate residues should be on the surface of GtgA. For that reason, a program SERp SERVER was used and a structure of GtgA was modeled using a SWISS MODEL server (http://services.mbi.ucla.edu/SER), for selecting surface residue(s) for the surface entropy reduction (Figure 2b). Those three

mutants are as follows: SER1 (K194A and E195A), SER2 (K212A and E213A), and SER3 (E138A)

2. Expression test

Full length GtgA, GtgA(H23), and three SERs were expressed in *E. coli* BL21 (DE3) and its expression as a soluble protein was estimated using SDS PAGE analysis. Given that crude lysate contains each protein of 25.9 kD MW, all five proteins were successfully expressed by addition of IPTG (Figure 3a). Detected protein band near 31 kD of marker are expected as GtgA. Moreover, separation of those crude lysate into a soluble fraction by centrifugation allowed a possible estimation of solubility of GtgA. In this comparison, GtgA, GtgA(H23), and SER3 were observed in their soluble fractions (Figure 3c), suggesting that these three proteins are soluble and could be used for further purification and crystallization. Expression level of those three proteins was also high, given that crude lysate and soluble fraction of each construct is almost equal in its expression level (Figure 3 b,c). However, SER1 and SER2 were not soluble protein under our experimental conditions.

3. Purification

Cells harboring each gene for three proteins, such as GtgA, GtgA(H23), and SER3, were cultured in 2.4 L of Luria-Bertani media. The expressed proteins were purified using IMAC. In the first cycle of IMAC, the column was washed with 5% of buffer B and concentration of buffer B was increased to 40% for elution of protein. When the column was washed with 10% of buffer B before protein elution, the protein was eluted (Figure 4b). The sample of peak in 10% of buffer B was examined by SDS PAGE and the band was showed in the same size as GtgA

protein. Therefore, all constructs were purified without wash by 10% of buffer B. When zinc was added to the lysis buffer, the protein does not elute in spite of existence of band in SDS PAGE gel of its soluble fraction (Figure 5) and the result of expression test.

After dialysis and cleavage of histidine tag, the proteins without histidine tag were purified by additional IMAC. The proteins were eluted at 5% of buffer B, the concentration for elution of unbound samples. After elution of target proteins, concentration of buffer B was increased to 60%. At the concentration, TEV protease and target proteins with tag remaining not to be cleaved were eluted (Figure 4c).

The proteins purified by IMAC were concentrated and applied to gel filtration column to remove soluble aggregate. The result of gel filtration of all constructs showed that most of proteins were homologous protein rather than aggregate (Figure 4d).

The results of chromatography showed the same pattern in all constructs. Purified proteins were concentrated and samples in each step were loaded to SDS PAGE to evaluate their purity (Figure 6a). The expression of methylated protein could not be evaluated by expression test, so samples of each step were applied to SDS PAGE for inspection of its purification (Figure 6b).

4. Crystallization

Purified proteins were attempted to be crystallized. However, full length and H23 GtgA does not show any crystal. And as a result of full screening of SER3 which has lower solubility, it formed crystals in a condition (Figure 7a).

The concentration of protein was 12 mg/ml and the condition was containing 100 mM HEPES (pH 7.5), 200 mM $CaCl_2$, and 28% PEG 400. To

determine whether it was protein or salt crystal, ultraviolet microscope was used. Fluorescence of crystal told that the crystal consisted of proteins (Figure 7b,c).

The crystal was tried to reproduce, however, the same condition where crystal observed did not show any crystal. As a result of crystallizing in various concentrations of $CaCl_2$, PEG 400, and protein, reproducing crystal was possible (Figure 8a-d). The variation of $CaCl_2$ was 0, 200, 300, 350, and 400 mM. When there was not $CaCl_2$, the crystal was not formed. The concentrations of PEG 400 used for crystallization were 24, 26, 28, 30, 32, 34, and 36%. The crystals were observed at above 30% and the probability of crystallization was low at 34 and 36%. Therefore, 30 or 32% of PEG 400 was used in the following experiments. The concentrations of protein for crystallization were 8 and 12 mg/ml. 10 and 14 mg/ml of GtgA was also crystallized. 12 mg/ml of GtgA showed the crystals the most. 8 mg/ml of GtgA showed similar result but took long to form crystals. In addition, pH of HEPES was varied to 7.2, 7.5 and 7.8. The crystal showed the different aspect in each batch. The additive screen was executed as well as general reproduce, but the crystal was not formed in all wells.

5. Data collection and scaling

Total 360 frames were obtained from crystal of GtgA SER3 (Figure 9). Using a frame, indexing was executed and the result showed that the crystal formed orthorhombic Bravais lattice. Subsequently, the reflections were integrated and chi^2 value was near 1 at all reflections. Finally, the data was scaled. The log file showed that the set of absences corresponded to that of $P2_12_12_1$, so the space group was determined as $P2_12_12_1$. The final resolution cutoff was 2.5 Å in which I/sigma value was about 1 and $CC_{1/2}$ value was about 0.5 (Table 3).

6. Molecular replacement

With the scaled data, molecular replacement was tried to determine its structure using *PHENIX* program. By orienting and positioning templates appropriately, the program can calculate the phase of GtgA. Templates were 4o2i, 4lgj and 4q3j of protein data bank which SWISS-MODEL program used for modeling as templates. 4lgj and 4q3j are NleC protein from *Escherichia coli* and had the same function with GtgA (Li et al., 2013; Turco and Sousa, 2014). Sequence alignment with three templates was carried out using clustal omega (Figure 10). GtgA was aligned from 94th residue and sequence identity between GtgA and each model was 21% which belonged to "difficult" among ranges of possibility of molecular replacement. And as a result, the program was failed to solve the structure. Whether it was succeeded or not was determined by the value of log-likelihood gain (LLG) and translation function Z-score (TFZ). LLG value increases as the more component is formed and it should be positive and as high as possible, but resulting LLG value was just 71.3. The resulting TFZ score was about 4.5 and PHENIX documentation tells that TFZ score below 5 means "no".

After failing the standard molecular replacement, sculptor and ensemble of three models were made using *PHENIX*. The sculptor trims the side chain or loop of template comparing with the sequence of target protein, GtgA. In this process, following ranges which are aligned with GtgA was used; 80-222 of 4o2i and 4q3j, and 58-200 of 4lgj. Each products of the sculptor were used as the template of molecular replacement. However, the results showed about 71 of LLG value and about 5.2 of TFZ meaning "unlikely".

As the last trial of molecular replacement, ensembler in *PHENIX* was used. The ensembler is superimposing several models which have the similar structure. Three templates modified with sculptor were made into ensemble, and the

ensemble was used as template for molecular replacement. The resulting LLG value was 73.6 and TFZ value was 5.6 which also meaning “unlikely”.

Discussion

Structure of protein is vital for molecular interaction and mechanism. The structure of GtgA would be informative for learning interaction of *Salmonella* with its hosts.

Based on the previous studies for NleC, protease for NF-κB, a lot of glutamates of GtgA are regarded as residues which constitute the surface of protein near the active site and contribute to binding with its targets, RelA and RelB. Because RelA and RelB have positively charged surface which interacts with DNA, negatively charged residues of GtgA are supposed to give complementarity with target proteins. In addition, although previous research revealed that GtgA cannot cleave p100 and p105 belonging to NF-κB family, there is the possibility that some other NF-κB family proteins could be the target of GtgA, such as NleC. These would be understood by solving the structure of GtgA and executing additional functional analysis. To reveal the structure of GtgA, I made several constructs of GtgA, examined their expression, and purified them for crystallization in this study. All constructs were soluble in the same buffer and showed similar result in each chromatography. A mutant made by reducing a charged residue at surface formed crystals.

Although GtgA is zinc dependent protease, its solubility was lower in buffer including zinc. In the structure of NleC, similar protein used as template of molecular replacement, zinc was binding in metal binding site without supplying zinc. Because GtgA have identical zinc binding motif with that of NleC, GtgA might have zinc within its zinc binding site already.

After scaling the diffraction data, chi^2 value was about 0.967 which was close to 1 and the estimated error was low. The data was cut off at 2.5 Å which had

about 1 of I/sigma value 0.5 of $CC_{1/2}$ value. Even though I/sigma value is not high, 0.5 of $CC_{1/2}$ value is assessed as reliable value enough (Karplus and Diederichs, 2012).

Although the quality of data was high, the electron density map couldn't be obtained. The phases are needed to obtain electron density map, but the phases of each reflection are unknown. The simplest method to obtain the phases is molecular replacement using three structures which had 21% of sequence identity and 30% of similarity. However, the *PHENIX* program failed it due to the low sequence identity, so other method should be applied. Isomorphous replacement and anomalous scattering both of which use the heavy atom are able to obtain the phases of unknown proteins.

For this reason, seleno-methionine labeled GtgA was attempted to crystallize to use anomalous scattering method between two methods. As a result of purifying the selenium-labeled GtgA, the methionine of GtgA protein would have selenium rather than sulfur. If the crystal of seleno-methionine labeled GtgA is obtained, phase of methionines will be acquired by the difference of diffraction data from native protein. It is because that heavy atoms including selenium absorb specific wavelength of X-ray, which is contrary to Friedel's Law (Rhodes, 2010). By comparing data from the native and seleno-methionine labeled protein, the phases of selenium could be clear. Then, based on the sequence and space group, the phases of other residues are able to be acquired.

GtgA is one of the evidence for evolution of interaction between *Salmonella* and its hosts, especially, in respect that it secretes GtgA and other PipA family of effectors in spite of decrease of virulence. It seems that *Salmonella* selected the long-term survival in the host rather than replication (Sun et al., 2016). The mechanism of GtgA is supposed to give the insight for its specificity and

interaction of *Salmonella* with hosts. Therefore, in the basis of this study, virulence and evolution of bacterial pathogen would be learned in depth.

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N, et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, *66*(2), 213-221.

Alfano, J. R., and Collmer, A. (2004). Type III secretion system effector proteins: double agents in bacterial disease and plant defense. *Annu. Rev. Phytopathol.*, *42*, 385-414.

Baruch, K., Gur-Arie, L., Nadler, C., Koby, S., Yerushalmi, G., Ben-Neriah, Y., et al. (2011). Metalloprotease type III effectors that specifically cleave JNK and NF-κB. *The EMBO journal*, *30*(1), 221-231.

Cornelis, G. R., and Van Gijsegem, F. (2000). Assembly and function of type III secretory systems. *Annual Reviews in Microbiology*, *54*(1), 735-774.

Figueroa-Bossi, N., Uzzau, S., Maloriol, D., and Bossi, L. (2001). Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in Salmonella. *Molecular microbiology*, *39*(2), 260-272.

Hobbie, S., Chen, L. M., Davis, R. J., and Galan, J. E. (1997). Involvement of mitogen-activated protein kinase pathways in the nuclear responses and cytokine production induced by Salmonella typhimurium in cultured intestinal epithelial cells. *The Journal of Immunology*, *159*(11), 5550-5559.

Karplus, P. A., and Diederichs, K. (2012). Linking crystallographic model and data quality. *Science*, *336*(6084), 1030-1033.

Lawrence, T. (2009). The nuclear factor NF-κB pathway in inflammation. *Cold Spring Harbor perspectives in biology*, *1*(6), a001651.

Lee, Y. K., and Mazmanian, S. K. (2010). Has the microbiota played a critical role in the evolution of the adaptive immune system?, *Science*, *330*(6012), 1768-1773.

Li, W., Liu, Y., Sheng, X., Yin, P., Hu, F., Chen, C., et al. (2014). Structure and mechanism

of a type III secretion protease, NleC. *Acta Crystallographica Section D: Biological Crystallography*, *70*(1), 40-47.

Morgan, M. J., and Liu, Z. G. (2011). Crosstalk of reactive oxygen species and NF-κB signaling. *Cell research*, *21*(1), 103-115.

Mühlen, S., Ruchaud-Sparagano, M. H., and Kenny, B. (2011). Proteasome-independent degradation of canonical NFκB complex components by the NleC protein of pathogenic Escherichia coli. *Journal of Biological Chemistry*, *286*(7), 5100-5107.

Otwinowski, Z., and Minor, W. (1997). [20] Processing of X-ray diffraction data collected in oscillation mode. *Methods in enzymology*, *276*, 307-326.

Rabsch, W., Andrews, H. L., Kingsley, R. A., Prager, R., Tschäpe, H., Adams, L. G., and Bäumler, A. J. (2002). Salmonella enterica serotype Typhimurium and its host-adapted variants. *Infection and Immunity*, *70*(5), 2249-2255.

Rhodes, G. (2010). *Crystallography made crystal clear: a guide for users of macromolecular models*. Academic press.

Rytkönen, A., Poh, J., Garmendia, J., Boyle, C., Thompson, A., Liu, M., et al. (2007). SseL, a Salmonella deubiquitinase required for macrophage killing and virulence. *Proceedings of the National Academy of Sciences*, *104*(9), 3502-3507.

Shames, S. R., Bhavsar, A. P., Croxen, M. A., Law, R. J., Mak, S. H., Deng, W., et al. (2011). The pathogenic Escherichia coli type III secreted protease NleC degrades the host acetyltransferase p300. *Cellular microbiology*, *13*(10), 1542-1557.

Smith, J. N., and Ahmer, B. M. (2003). Detection of other microbial species by Salmonella: expression of the SdiA regulon. *Journal of bacteriology*, *185*(4), 1357-1366.

Stecher, B., Robbiani, R., Walker, A. W., Westendorf, A. M., Barthel, M., Kremer, M., et al. (2007). Salmonella enterica serovar typhimurium exploits inflammation to compete with the intestinal microbiota. *PLoS Biol*, *5*(10), e244.

Sun, H., Kamanova, J., Lara-Tejero, M., and Galán, J. E. (2016). A Family of Salmonella Type III Secretion Effector Proteins Selectively Targets the NF-κB Signaling Pathway to Preserve Host Homeostasis. *PLoS Pathog*, *12*(3), e1005484.

Takeuchi, A. (1967). Electron microscope studies of experimental Salmonella infection. I. Penetration into the intestinal epithelium by Salmonella typhimurium. *The American journal of pathology*, *50*(1), 109.

Turco, M. M., and Sousa, M. C. (2014). The structure and specificity of the type III secretion system effector NleC suggest a DNA mimicry mechanism of substrate recognition. *Biochemistry*, *53*(31), 5131-5139.

Winter, S. E., Thiennimitr, P., Winter, M. G., Butler, B. P., Huseby, D. L., Crawford, R. W., et al. (2010). Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature*, *467*(7314), 426-429.

Wood, M. W., Jones, M. A., Watson, P. R., Hedges, S., Wallis, T. S., and Galyov, E. E. (1998). Identification of a pathogenicity island required for Salmonella enteropathogenicity. *Molecular microbiology*, *29*(3), 883-891.

Table 1. The primer sequences used in this study.

Full length	Forward	5’- ATGCCATATGCCAACCGGCATCAAACC	3’-
	Reverse	5’- ATGCCTCGAGTTAGTTGCTGAACTCATAGGC	3’-
H23	Forward	5’- CATATGCATGACTCCAAAGTATTCCCGGACC	3’-
	Reverse	5’- CTCGAGTTAGTTGCTGAACTCATAGGCGATGC	3’-
SER 1	Forward	5’- CTGACGCACCTGCAGGAT**GC**GG**C**AGACAGTAACC CGCGC	3’-
	Reverse	5’- GCGCGGGTTACTGTCT**G**CC**GC**ATCCTGCAGGTGCG TCAG	3’-
SER 2	Forward	5’- GTATACCAACATCATTCTG**GC**AG**C**AATGGGCCATAC ATCTCCG	3’-
	Reverse	5’- CGGAGATGTATGGCCCATT**G**CT**GC**CAGAATGATGTT GGTATAC	3’-
SER 3	Forward	5’- GTACCACTGTTACAGACG**C**AGACCTGGAATCTTCG G	3’-
	Reverse	5’- CCGAAGATTCCAGGTCT**G**CGTCTGTAACAGTGGTA C	3’-

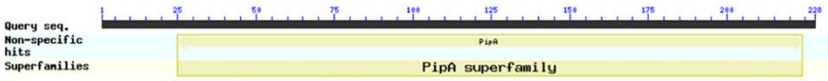
The bold-underlined letters indicate the mutated sequences.

SER: surface entropy reduction mutant

Table 2. Mutated residues for surface entropy reduction mutants.

	Template	Mutant
1	K194, E195	A194, A195
2	K212, E213	A212, A213
3	E138	A138

(a)



(b)

```
1...*...10....*...20....*...30....*...40....*...50....*...60....*...70....*...80....*...90....*..100
MPTGIKPIFINNMMSTYGLSHPHDSKVFPDLPEHQDNPSQLRLQHDGLATDDKARLEPMCLAEYLISGPGGMDPDIEIDDDTYDECREVLSRILEDAYTQ

....*..110....*..120....*..130....*..140....*..150....*..160....*..170....*..180....*..190....*..200
SGTFRRLMNYAYDQELHDVEQRWLLGAGENFGTTVTDEDLESSEGRKVIALNLDDTDDDSIPEYYESNDGPQQFDITRSFIHEVVHALTHLQDKEDSNPR

....*..210....*..220....*...
GPVVEYTNIILKEMGHTSPPRIAYEFSN
```

Figure 1. The result of domain search and prediction of secondary structure.

(a) The result of basic local alignment search tool (BLAST) shows that 25-224 of GtgA belongs to PipA family. (b) The result of secondary structure prediction of GtgA by XtalPred is shown. Red words are predicted as helix, blue is strand, black is loop, and underlined region is disordered region.

(a)

Proposed Mutations:

Cluster #1: **Residues 194 - 195:** KE [?]

- K 194 => A
- E 195 => A

SERp Score: 4.9 [?]

Cluster #2: **Residues 212 - 213:** KE [?]

- K 212 => A
- E 213 => A

SERp Score: 2.89 [?]

Cluster #3: **Residues 138 - 138:** E [?]

- E 138 => A

SERp Score: 2.38 [?]

(b)

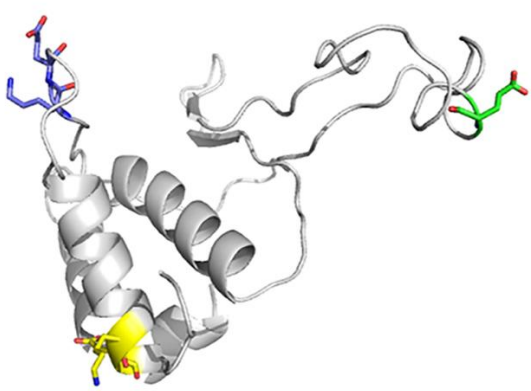


Figure 2. prediction of suitable surface entropy reduction sites.

(a) Surface entropy reduction server proposed three mutations. The score is calculated by predicted coil, entropy, and reference by conserved region. (b) Three selected regions are indicated to model produced by SWISS-MODEL program. Yellow is mutant 1, blue is 2, and green is 3. The model shows the region from 94 to 220 of GtgA.

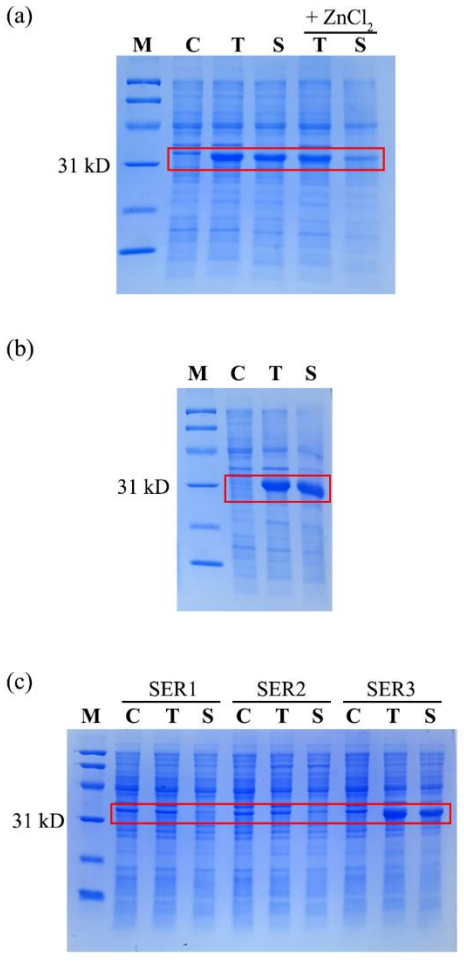

(a)
+ ZnCl2
M C T S T S
31 kD
(b)
M C T S
31 kD
(c)
SER1 SER2 SER3
M C T S C T S C T S
31 kD

Figure 3. SDS PAGE analysis of all constructs used in this study.

M: Marker; C: Control; T; Total fraction; S: Soluble fraction. Red box indicates expressed proteins. The lysis buffer consists of 50 mM Tris (pH 8.0), 100mM NaCl. (a) Full length of GtgA was lysed in the buffer (left) and that added with $ZnCl_2$ (right). (b) H23, truncation mutant, shows that it is soluble. (c) Among three SERs, only the SER3 expressed solubly.

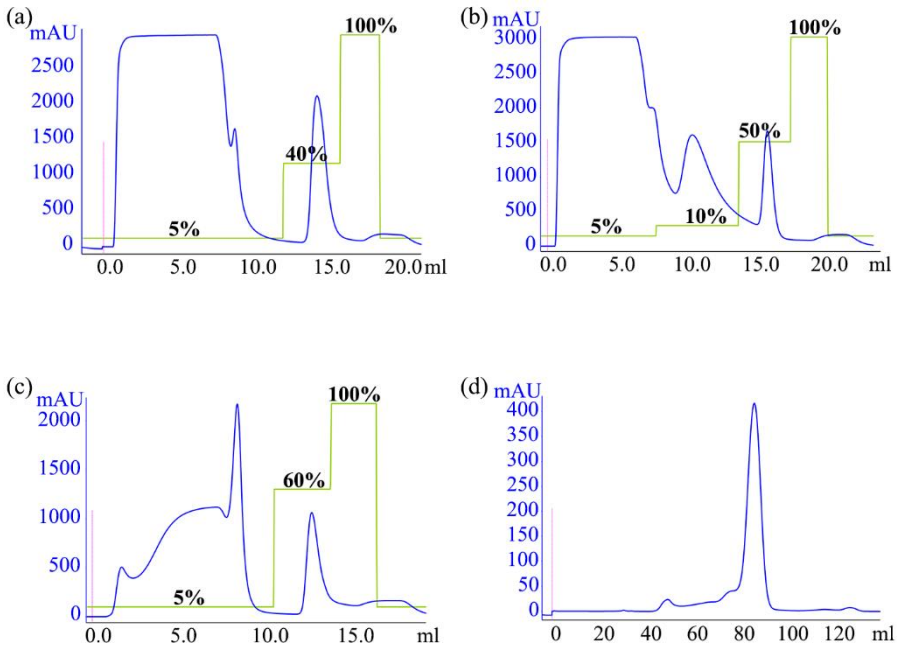

(a)
mAU
2500
2000
1500
1000
500
0
100%
40%
5%
0.0
5.0
10.0
15.0
20.0
ml
(b)
mAU
3000
2500
2000
1500
1000
500
0
100%
50%
10%
5%
0.0
5.0
10.0
15.0
20.0
ml
(c)
mAU
2000
1500
1000
500
0
100%
60%
5%
0.0
5.0
10.0
15.0
ml
(d)
mAU
400
350
300
250
200
150
100
50
0
0
20
40
60
80
100
120
ml

Figure 4. The purification profiles of solubly expressed constructs.

Blue line: absorbance recording at 280nm; Green line: concentration of buffer B. All constructs showed the same aspect of profile. (a) The profile shows IMAC of GtgA with his tag and protein was eluted at 40% of buffer B. (b) The profile shows a result of IMAC when the column was washed 10% of buffer B additionally. (c) The profile shows IMAC of GtgA after cleavage of his tag and protein was eluted at 5% of buffer B. (d) The profile shows GFC of product of IMAC.

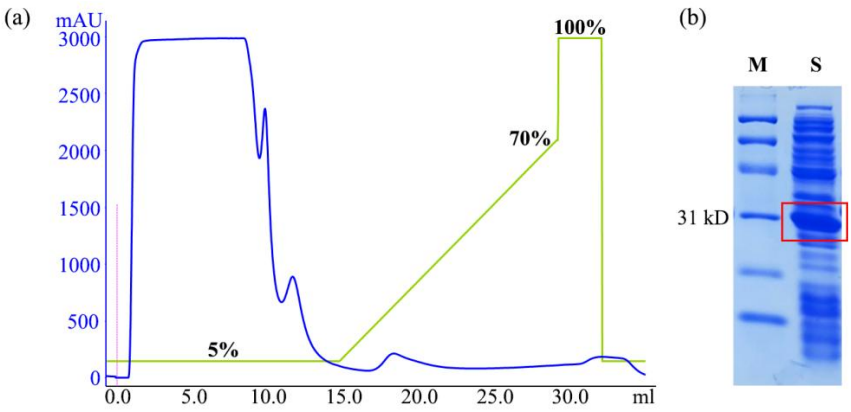
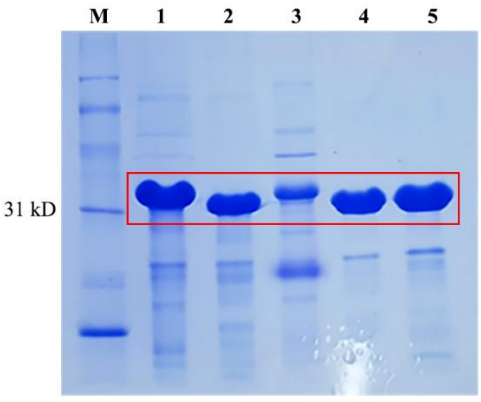

(a)
mAU
3000
2500
2000
1500
1000
500
0
100%
70%
5%
0.0
5.0
10.0
15.0
20.0
25.0
30.0
ml
(b)
M
S
31 kD

Figure 5. The purification profile lysed in buffer containing $ZnCl_2$.

(a) Blue line: absorbance recording at 280nm; Green line: concentration of buffer B. The protein isn't eluted as the concentration of buffer B increase. (b) In the soluble fraction of lysate, GtgA seems to be expressed.

(a)



(b)

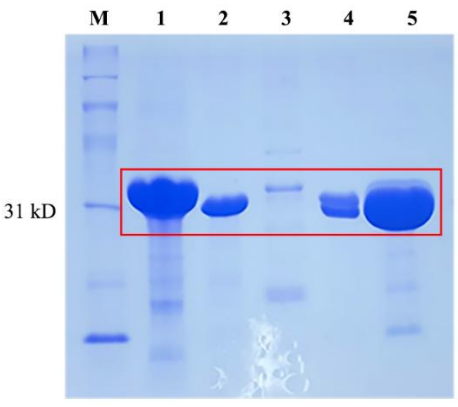
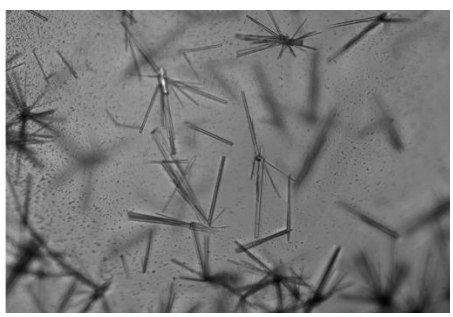


Figure 6. SDS PAGE analysis at each step of purification.

M: Marker; 1: First IMAC with his tag; 2: Second IMAC after cleavage of his tag; 3: 60% of buffer B in second IMAC; 4: GFC; 5: After concentration (a) The SDS PAGE gel shows the purified protein with twice of IMAC, GFC, and concentration. (b) The SDS PAGE gel shows the purified protein with twice of IMAC, lysine methylation, GFC, and concentration.

(a)

(b)



(c)

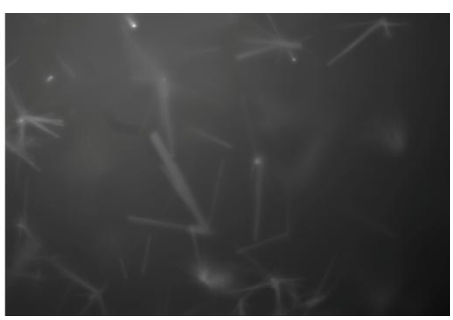


Figure 7. Initial crystal from full screening.

(a) Photograph of initial crystal was taken using general microscope. (b) Photograph of crystal was taken using ultraviolet microscope without ultraviolet. (c) Photograph of crystal was taken using ultraviolet microscope and the crystals shows fluorescence.

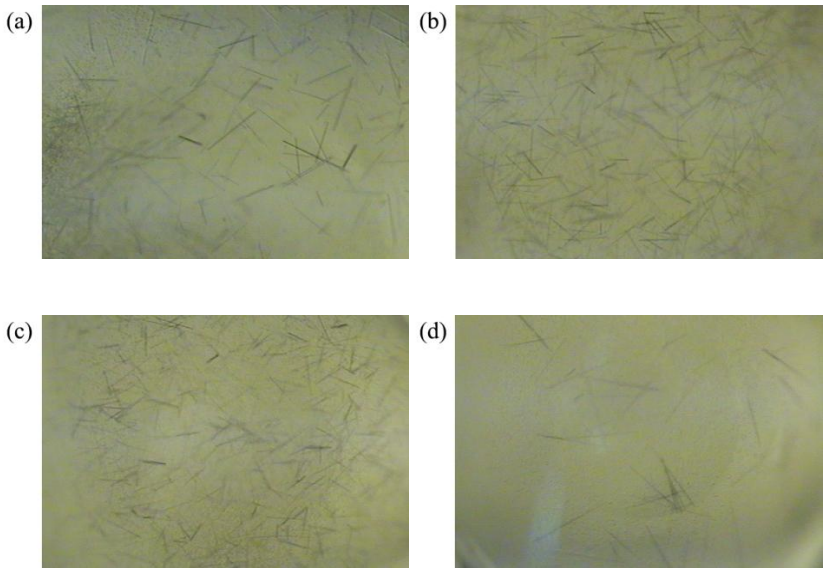

(a)
(b)
(c)
(d)

Figure 8. The crystals reproduced at various conditions.

(a) 12 mg/ml of GtgA was crystallized in the condition containing 100 mM of HEPES (pH 7.5), 32% of PEG400, and 300 mM of $CaCl_2$. (b) 8 mg/ml of GtgA was crystallized in the same condition with (a). (c) The concentration of PEG400 was increased to 34% comparing to (a). (d) The concentration of $CaCl_2$ was increased to 400 mM comparing to (a).

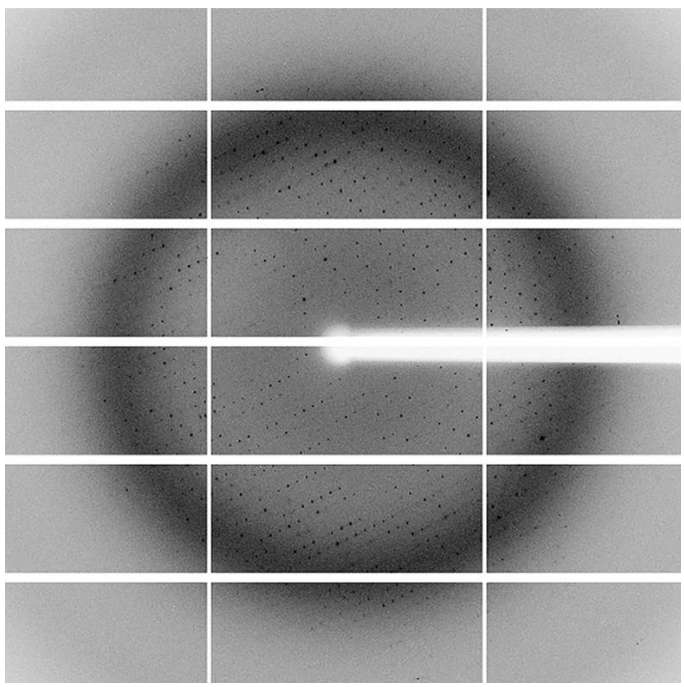


Figure 9. A diffraction pattern of GtgA.

The figure is a diffraction pattern at 90°.

Table 3. Data collection.

Data set	GtgA SER3
Wavelength (Å)	0.97935
Resolution (Å)	50-2.5
Unique reflections	30,387
Multiplicity	6.5 (6.9)
R-merge (%)	13.3 (177.6)
Completeness (%)	99.9 (98.7)
Mean I/sigma (I)	8.8 (1.0)
Space group	$P2_12_12_1$
Unit Cell (Å)	
a, b, c	67.480, 109.565, 115.169
α=β=γ	90

The numbers in parentheses indicate the data in the highest resolution shell.

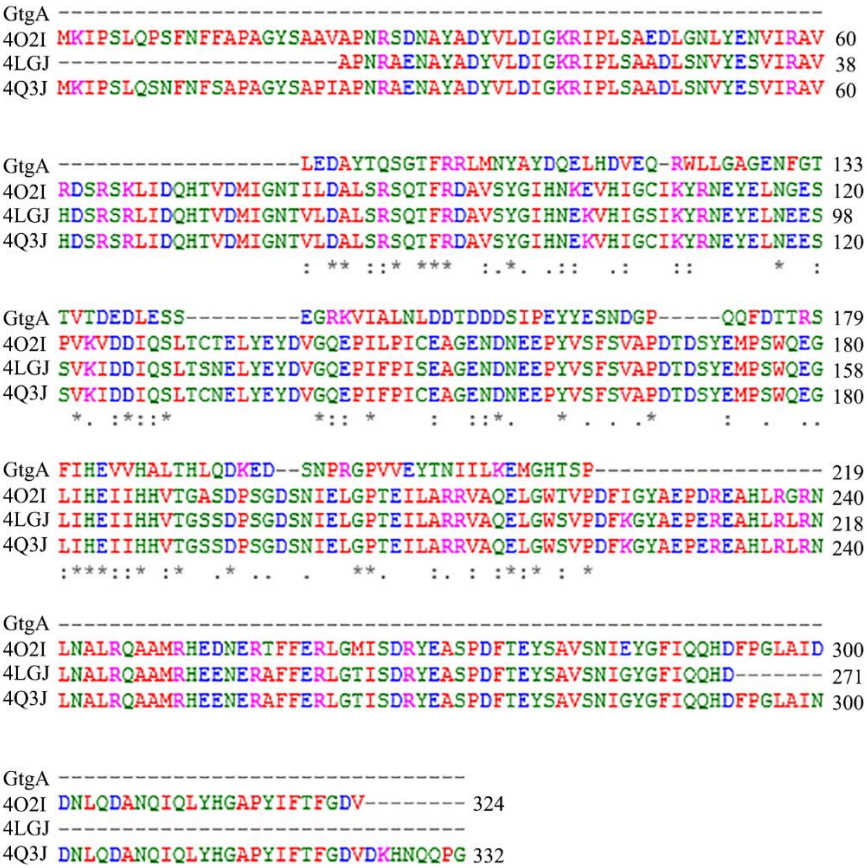

GtgA ----------
4O2I MKIPSLQPSFNFFAPAGYSAAVAPNRSDNAYADYVLDIGKRIPLSAEDLGNLYENVIRAV 60
4LGJ ----------APNRAENAYADYVLDIGKRIPLSAADLSNVYESVIRAV 38
4Q3J MKIPSLQSNFNFSAPAGYSAPIAPNRAENAYADYVLDIGKRIPLSAADLSNVYESVIRAV 60
GtgA ----------LEDAYTQSGTFRRLMNYAYDQELHDVEQ-RWLLGAGENFGT 133
4O2I RDSRSKLIDQHTVDMIGNTILDALSRSQTFRDAVSYGIHNKEVHIGCIKYRNEYELNGES 120
4LGJ HDSRSRLIDQHTVDMIGNTVLDALSRSQTFRDAVSYGIHNEKVHIGSIKYRNEYELNEES 98
4Q3J HDSRSRLIDQHTVDMIGNTVLDALSRSQTFRDAVSYGIHNEKVHIGCIKYRNEYELNEES 120
GtgA TVTDEDLESS---------EGRKVIALNLDDTDDDSIPEYYESNDGP-----QQFDTTRS 179
4O2I PVKVDDIQSLTCTELYEYDVGQEPILPICEAGENDNEEPYVSFSVAPDTDSYEMPSWQEG 180
4LGJ SVKIDDIQSLTSNELYEYDVGQEPIFPISEAGENDNEEPYVSFSVAPDTDSYEMPSWQEG 158
4Q3J SVKIDDIQSLTCNELYEYDVGQEPIFPICEAGENDNEEPYVSFSVAPDTDSYEMPSWQEG 180
GtgA FIHEVVHALTHLQDKED--SNPRGPVVEYTNIILKEMGHTSP---------- 219
4O2I LIHEIIHHVTGASDPSGDSNIELGPTEILARRVAQELGWTVPDFIGYAEPDREAHLRGRN 240
4LGJ LIHEIIHHVTGSSDPSGDSNIELGPTEILARRVAQELGWSVPDFKGYAEPEREAHLRLRN 218
4Q3J LIHEIIHHVTGSSDPSGDSNIELGPTEILARRVAQELGWSVPDFKGYAEPEREAHLRLRN 240
GtgA ----------
4O2I LNALRQAAMRHEDNERTFFERLGMISDRYEASPDFTEYSAVSNIEYGFIQQHDFPGLAID 300
4LGJ LNALRQAAMRHEENERAFFERLGTISDRYEASPDFTEYSAVSNIGYGFIQQHD------- 271
4Q3J LNALRQAAMRHEENERAFFERLGTISDRYEASPDFTEYSAVSNIGYGFIQQHDFPGLAIN 300
GtgA ----------
4O2I DNLQDANQIQLYHGAPYIFTFGDV-------- 324
4LGJ ----------
4Q3J DNLQDANQIQLYHGAPYIFTFGDVDKHNQQPG 332

Figure 10. Sequence alignment with templates used in molecular replacement.

The sign * indicates the same residue in all sequences.

Abstract in Korean

살모넬라 Typhimurium은 인간과 몇 가지 포유류에 침입하는 병원균이다. 살모넬라의 병원성은 effector라고 알려진 자신의 단백질을 숙주에 전달함으로써 발생하지만, 이에 따라 염증 반응으로 이어지는 숙주의 방어 기작이 일어난다. 염증 반응은 병원균에 반응하는 숙주의 방어 기작 중 하나이며, 병원균의 증식을 억제시킨다. 사람은 nuclear factor κB (NF-κB)가 염증전 사이토카인을 생성의 전사를 활성화시켜 염증 반응의 조절에 중요한 역할을 한다는 것이 알려져 있다. 살모넬라는 이 NF-κB에 의한 염증 반응을 피하기 위해 PipA family에 속하는 세 개의 단백질을 이용한다. 이 family의 단백질은 아연을 이용하는 프로테아제이다. 살모넬라의 effector 중 GtgA는 이 family의 단백질 중 하나이며, NF-κB의 DNA에 결합하는 데 중요한 부분을 가수분해한다. 이러한 GtgA의 NF-κB 분해는 결과적으로 염증전 사이토카인 생성을 감소시킨다. 하지만, GtgA의 작용에 대한 분자 수준의 정보는 아직 밝혀지지 않았다. GtgA의 구조적 특성을 밝히기 위해, 본 연구에서는 GtgA를 다양한 형태로 발현시키고, 결정화하여 엑스레이 회절 분석을 포함하여 광범위한 실험을 진행하였다. 초기의 full length GtgA는 매우 가용성이 높지만 결정화가 되지 않다는 점을 암시했다. 그리하여 친수성 잔기들을 소수성을 돌연변이하여 GtgA의 유체 역학적 특성을 바꾸었다. 다섯 가지의 GtgA construct 중 두 가지가 불용성 형태가 되었지만, 세 개는 가용성을 보였다. 이러한 가용성의 GtgA construct들은 정제와 시팅 드롭 증기 확산 방법을 통하여 다양한 조건에서 결정화를 진행하였다. 초기 결정 조건은 최근 엑

스레이 회절 데이터를 성공적으로 얻었다. 그러나 20%의 아미노산 서열 상동성을 가진 단백질들로 molecular replacement가 성공적으로 구조를 밝히지 못 하였다. 현재, GtgA의 위상 정보를 얻기 위하여 추가적인 실험을 진행하고 있다. GtgA의 구조적 특성은 숙주와 병원균과의 상호작용과 관련하여 유익한 정보를 줄 것으로 사려된다.

주요어: GtgA, nuclear factor κB, *Salmonella* Typhimurium, X-ray diffraction, zinc-metalloprotease

학 번: 2015-21797