博士論文番号：0581018

Structural Analysis of the Phototactic Transducer Protein HtrII from *Natronomonas pharaonis*.

（走光性シグナルトランスデューサータンパク質 *p*HtrII の構造学的研究）

林　こころ
奈良先端科学技術大学院大学
バイオサイエンス研究科　生体高分子構造学講座
（児嶋　長次郎　准教授）

平成１９年１２月２５日提出

Contents

Abbreviation

MCPs	methyl-accepting chemotaxis proteins
E. coli	*Escherichia coli*
N. pharaonis	*Natronomonas pharaonis*
Che	chemotaxis protein
HAMP	histidine kinases adenylyl cyclases methyl binding proteins phosphatases
SR	sensory rhodopsin
Htr	halobacterial transducer
*p*pR	*pharaonis* phoborhodopsin
*p*HtrII	*pharaonis* halobacterial transducer II
H. salinarum	*Halobacterium salinarium*
DDM	n-dodecyl-β-D-maltoside
NMR	nuclear magnetic resonance spectroscopy
HSQC	heteronuclear single quantum coherence.
EPR	electron paramagnetic resonance spectroscopy
CD	circular dichroism
SDS-PAGE	sodium dodecylsulfate-polyacrylamide gel electrophoresis

MTSL	(1-oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) methanethiosulfonate
PC	L-α-phosphatidylcholine
CW	continuous wave
ELDOR	electron-electron double resonance
MS	mass spectroscopy
MALDI-TOF-MS	matrix-assisted laser desorption/ionization time-of-flight mass spectroscopy
rms	root mean square
GST	glutathione S-transferase
MBP	maltose-binding protein
Trx	thioredoxin
GB1	protein G B1 domain
TEE	translation-enhancing element
PCR	polymerase chain reaction
LB	Luria Broth media
IPTG	isopropyl β-D-thiogalactoside
HRV3C	human rhinovirus 3C protease

1. General Introduction

1.1 Bacterial taxis

Bacteria can sense and respond environmental changes appropriately. A possible response is to move towards or away from an environment. In the absence of stimuli, a swimming bacterial cell executes a three-dimensional random walk consisting of turns of swimming in a straight line punctuated by tumbles during which the cell briefly stops and randomly reorients before swimming off in a new arbitrary direction. When an attractant stimulus is applied, the sensory pathway detects an attractant stimulus and sends a signal to the flagellar motor which decreases the probability of a tumble event, and thereby the cell swims to the attractant stimulant. Alternatively, if the swimming cell monitors a decrease in the attractant concentration or a repellent stimulus, the tumbling probability is increased so that runs in this direction become shorter.

Most sensory pathways in eukaryotes use serine, threonine or tyrosine protein kinases, whereas most sensory pathways in prokaryotes use a histidine-aspartate phosphorelay (HAP) system. HAP system has a two-component phosphorylation cascade, which consists of the autophorphorylation of a histidine kinase followed by the transfer of the phosphate moiety to an aspartate residue of a protein called the response-regulator [2]. This two-component pathway is also found in many lower

eukaryotes [3] [4] [5] but not identified in animals, and two-component proteins are not encoded by the human, fly or worm genomes.

1.2 The chemosensory system in bacteria

Bacteria can sense a vast number of environmental signals, from the concentrations of nutrients and toxins to oxygen levels, pH, osmolarity, temperature, and the intensity and wavelength of light. Chemotaxis in *Escherichia coli* is one of the best understood sensory systems [6], and it provides an excellent system to explore transmembrane signaling.

Chemotactic signals are detected by transmembrane chemoreceptors, the methyl-accepting chemotaxis proteins (MCPs) [7] and MCP-like protein. Most MCPs consists of periplasmic region, transmembrane region, linker region, methylation region and signaling region. Although the periplasmic region of MCPs consists of four α-helix in *E. coli*, but the fold of periplasmic region from the other species tends to be highly diverse, which reflects the wide range of ligand are sensed by these receptors. Whereas, methylation and signaling region are highly conserved, and X-ray crystal structure shows that methylation and signaling region of MCPs form a 200 Å-long coiled-coil of two antiparallel α-helix structure connected by a 'U-turn' [8]. On the other hand, the

linker region is lacking in crystal structures in MCPs, and it is also called ‘missing link’. Figure 1.1 shows the schematic diagram of the chemosensory system of *E. coli*. Transmission of their external information through the periplasmic region of MCPs is the first step shared by all pathways. The signal is propagated to the cytoplasmic region, and signal is finally transmitted to the flagellar motor via phosphorylation cascades [9-12].

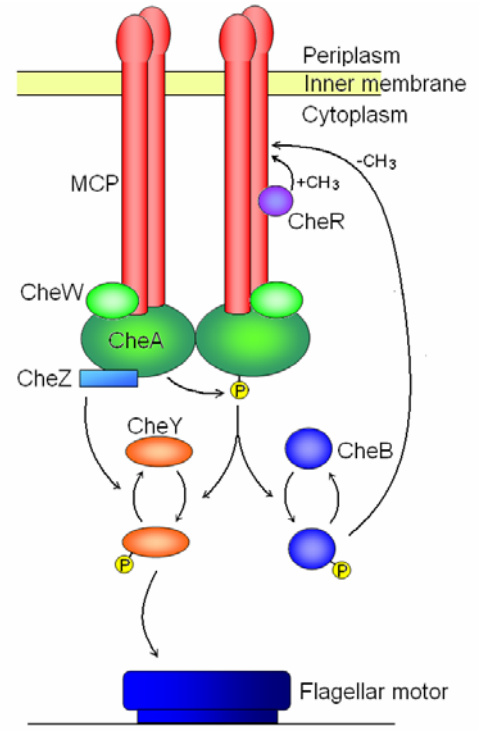


Figure 1.1. Schematic diagram of the chemosensory system of *Escherichia coli*. Two dimeric methyl-accepting chemotaxis proteins (MCPs) are shown. Two chemotaxis proteins CheW and CheA dimer interact with the highly conserved signaling domain of MCPs in the cytoplasm. CheW is an adaptor protein, and CheA is a histidine kinase, respectively. When a repellent ligand binds to periplasmic region of MCPs, the autophosphorylation of CheA dimer was induced. Phosphorylated CheA phosphorylates CheY, which binds to flagellar motor and changes the direction. Phosphorylated CheA also phosphorylates another response regulator, methylesterase CheB. Phosphorylated CheB competes with a constitutive methylesterase, CheR, and controls the degree of methylation of specific glutamates of MCPs. The phosphorylated CheY is dephosphorylated by the phosphatase, CheZ. P, phosphoryl group.

1.3 The photosensory system

Microorganisms including archaea have photosensory systems [13, 14]. In the photosensory systems, MCPs do not have periplasmic region which sense the stimuli, alternatively MCPs form complex with photoreceptor. Sensory rhodopsin I (SRI) and sensory rhodopsin II (SRII) are photoreceptors for phototaxis in *Halobacterium salinarium*. They are members of the seven-transmembrane helical retinal binding proteins which includes rhodopsin and bacteriorhodopsin [13]. SRI mediates attractive responses to the green/orange lights whereas SRII mediates repellent responses to damaging near-UV light. SRI and SRII form complex with membrane integral transducer proteins (HtrI and HtrII, respectively) which belongs to MCPs. M and/or O photo intermediate states of SRII activate signal transduction through HtrII [15, 16].

The extreme halophilic and alkalophilic archaeon *Natronomonas pharaonis* has the SRII homolog, *pharaonis* phoborhodopsin (*p*pR, also called *Np*SRII) [17] and the HtrII homolog, *pharaonis* halobacterial transducer II (*p*HtrII), which is so robust that it can function in photosignaling in *N. pharaonis* [18], *H. salinarum* [19], and *E. coli* [20]. *p*pR is stable both in dark and under illumination, and remains its native absorption spectrum and photocycle in artificial membranes, in a range of detergents, in salt concentrations 25 mM to 4 M, and wide range of pH. Because of these exceptional

stable properties, *p*pR has been studied as a model of photosensory system. It is known that *p*pR [21] and truncated *p*HtrII (amino acid position from 1 to 159) [22, 23] were successfully expressed in *E. coli*.

1.4 Aim to this study

It is known that the linker region is important module for signal transduction in MCPs which transmit the taxis signals from periplasm to cytoplasm across the cytoplasmic membrane [20] [24] [25-27]. The binding of the ligands to the periplasmic region changes the structure of the periplasmic regions, and also changes the structure of the transmembrane four-α-helix bundle of the MCPs dimer. However, it remains unclear how the signal is propagated from periplasmic to cytoplasmic signaling region via linker region, because the structure of linker region is missing.

In this thesis, I investigated the linker region structure of *p*HtrII, which is one of the MCPs and transmit the photosignal in *N. pharaonis*. In the process of this study, I also tried to prepare the linker region of the other chemotaxis related proteins, Tsr and Tar in *E. coli*. But these proteins could not be prepared by existing overexpression system in *E. coli*. To obtain such proteins, I have developed a novel expression system in *E. coli*.

2. Secondary structure analysis of *p*HtrII linker region

2.1 Introduction

Transducer protein *p*HtrII from the extreme halophilic and alkalophilic archaeon *N. pharaonis* is a two-transmembrane protein, belonging to the MCPs [7], and forms a homodimer comprising subunits of ca. 60 kDa and is located in the cytoplasmic membrane. *p*HtrII consists of four distinct areas comprising transmembrane, linker, methylation and signaling regions (Figure 2.1). The phototactic signaling in *N. pharaonis* utilizes *p*HtrII in the form of a complex with photosignal receptor, *p*pR.

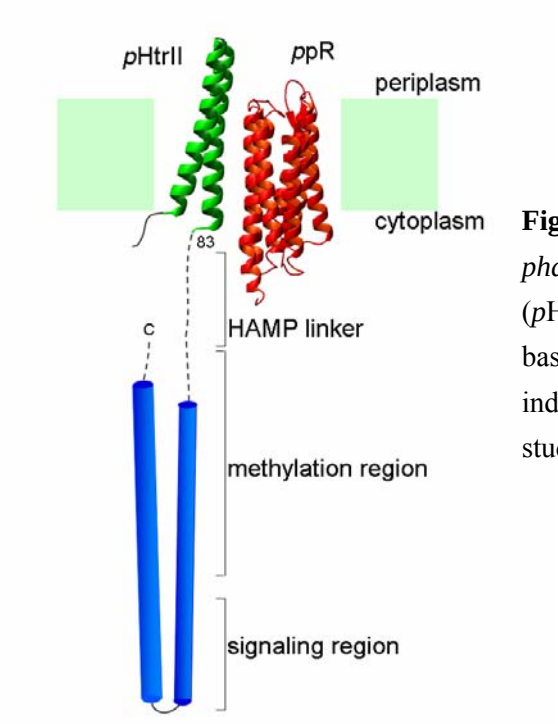


Figure 2.1. Schematic representation of *pharaonis* halobacterial transducer protein (*p*HtrII) and *pharaonis* phoborhodopsin (*p*pR) based on the crystal structures. Dashed line indicates the linker region focused on in this study.

Crystal structures of the two-transmembrane region of *p*HtrII comprising residues 24-82 as complexes with *p*pR in the dark [28] and K and late M states have been reported [29],

although the N-terminal 1-23 residues and C-terminal 83-114 or 83-157 residues were missing. The crystal structure of the methylation and signaling regions has been reported for *Escherichia coli* Tsr, a member of the MCP group [8], although part of the methylation region was missing. Given that *p*HtrII and Tsr share 26% identity in primary sequence, the two proteins may possess similar structures with respect to the methylation and signaling regions. Although a crystal structure of the linker region has yet to be reported, the flexible aspects of the linker region have been reported [30-32]. The functional importance of the linker region in relation to signal transduction has been investigated in *p*HtrII [20, 24, 25] and other MCPs [26, 27]. The linker region contains a HAMP domain, a common structural element which has been identified by sequence comparisons and mutation analyses of sensor kinases and MCPs [33-35].

In this chapter, I describe the identification of structural domain in *p*HtrII linker region based on limited proteolysis. Experimentally, the trypsin resistant fragment has been identified in *p*HtrII linker region and characterized by CD spectroscopy. Total 15 kinds of expression plasmids have been constructed for this domain and characterized by CD and NMR spectroscopy.

2.2 Materials and Methods

2.2.1 Construction of expression plasmids

The expression plasmid of the C-terminal truncated transducer *p*HtrII(1-159) was constructed as previously described [23]. The expression of various length plasmids of *p*HtrII linker region was constructed by TOPO-cloning system (Invitrogen). *p*HtrII genes coding the linker region, 100-159, 100-149, 100-178, 100-189, 100-205, 100-218, 100-229, 100-272, 83-100, 83-159, 83-168, 83-189, 83-178 and 83-229 were amplified by PCR, and they were subcloned into the plasmid vector pET200 D-TOPO by using TOPO cloning system (Invitrogen). Factor Xa cleavage site was inserted just before *p*HtrII gene by quickchange mutagenesis method (Qiagen). The expressed proteins included N-terminal six-residue histidine tag, Xpress epitope tag and cleavage site for Factor Xa just before *p*HtrII protein. *p*HtrII(100-159) and *p*HtrII(100-149) were also subcloned into the plasmid vector pET100-D-TOPO, which includes C-terminal six-residue histidine tag.

2.2.2 Expression and purification

C-terminal truncated transducer *p*HtrII(1-159) with a C-terminal six-residue histidine tag was expressed in *Escherichia coli* BL21 Star (DE3) using M9 minimal medium, and purified by a previously described method with slight modifications [23].

Cells were grown at 37°C. Protein expression was induced by the addition of 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). Cells were harvested 3 h following induction. Harvested cells were resuspended in buffer A [50 mM HEPES buffer (pH 8.0) containing 300 mM KCl and 0.1 mM EDTA]. The suspension was lysed by sonication and crude membranes were collected by ultracentrifugation (138,000 x g for 30 min at 4°C). Membranes were resuspended and solubilized in 50 mM MES buffer (pH 6.5) containing 300 mM KCl, 5 mM imidazole and 1.0% n-dodecyl-β-D-maltoside (DDM) for 12 ~ 16 hours at 4°C. Following ultracentrifugation (138,000 x g for 30 min at 4°C) of the solubilized membranes, the supernatant was loaded onto a Ni^{2+}-NTA agarose (QIAGEN) column. The resin was washed extensively with 50 mM Tris buffer (pH 8.5) containing 300 mM KCl, 50 mM imidazole and 0.1% DDM to remove nonspecifically bound proteins. The histidine-tagged protein was eluted with Tris buffer (pH 8.5) containing 300 mM KCl, 500 mM imidazole and 0.1% DDM (Figure 2.2).

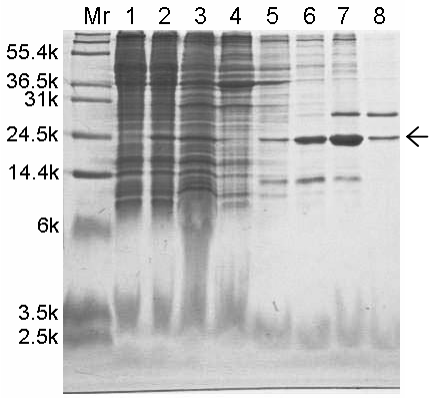


Figure 2.2. Expression and purification of *p*HtrII$^{1-159}$ by Ni-NTA agarose column. Mr, molecular weight marker; Lane 1, lysate of whole cells before induciton; Lane 2, lysate of whole cells after induction, Lane 3, the solubilized fraction from the membrane fraction by 1% DDM; Lane 4, pass; Lane 5, washout; Lane6-8, the elute fraction from the affinity resin. The arrow indicates the molecular weight of *p*HtrII$^{1-159}$ protein.

Water soluble proteins of *p*HtrII linker region were expressed as described above. Cells were lysed in buffer A by sonication, following ultracentrifugation (138,000 x g for 30 min at 4°C), the supernatant was loaded onto a Ni^{2+}-NTA agarose (QIAGEN) column. The resin was washed extensively with buffer A containing 50 mM imidazole, and the histidine-tagged protein was eluted with buffer A containing 500 mM imidazole. The N-terminal six-residue histidine-tag and Xpress epitope tag were digested by Factor Xa protease in 50 mM Tris buffer (pH7.5) containing 50 mM KCl at 20 °C. In the cases of C-terminal histidine-tagged protein, the six-residue histidine-tag was not cleaved and dialyzed with 50 mM Tris containing 50 mM KCl. Target proteins were finally purified by anion exchange chromatography using HiTrapQ HP (GE Healthcare), and dialyzed with buffer B [10 mM citric acid buffer (pH 5.0) 50 mM KCl,].

2.2.3 Trypsin digestion of *p*HtrII(1-159) protein and identification of the proteolytic fragment

Purified 0.48 mM *p*HtrII(1-159) protein was incubated with 1 μg/ml trypsin (SIGMA) in Ni^{2+}-NTA agarose column elution buffer; Tris buffer (pH 8.5) containing 300 mM KCl, 500 mM imidazole and 0.1% DDM for 10 hours at 4 °C. The reaction was stopped by addition of a protease inhibitor (Pefablock SC; Roche applied Science)

and the proteolytic fragment was separated by gel filtration chromatography using HiLoad Superdex 75 column (GE Healthcare) in buffer B containing 0.75 mM Pefabloc SC and 0.1% DDM. The proteolytic fragment was identified by N-terminal protein sequencing analysis and MALDI-TOF MS. Uniformly-labeled *p*HtrII(100-159) proteolytic fragment was prepared for CD and NMR investigation.

2.2.4 CD spectroscopy

Circular dichroism (CD) spectra were recorded using a J-720W CD spectropolarimeter (JASCO). The CD spectra were recorded between 200 nm and 260 nm (0.1 cm cell, 5 mg/ml protein in 5 mM MES buffer containing 25 mM KCl (pH 5.5)) at 0.1 nm intervals with a time constant of 4 sec and a scan speed of 20 nm/min. The secondary structure was predicted using the Spectra Manager software program, Windows version 1.0 (JASCO).

2.2.5 Secondary structure prediction

Secondary structure prediction was performed by PSIPRED [36], and prediction of coiled coil region was carried out by COILS [37].

2.2.6 NMR measurement

^{1}H-^{15}N HSQC spectra were acquired at 277 K using a DRX800 NMR spectrometer

(Bruker Biospin). Sample was dissolved in buffer B containing 90% H_2O/10% ^2H_2O mixture.

2.3 Results

2.3.1 Domain identification on *p*HtrII linker region.

*p*HtrII(1-159) protein was highly expressed in *E. coli* using M9 minimum medium. The expressed protein was purified from the membrane fraction and subjected to limited proteolysis with low concentrations of trypsin at low temperature in an effort to identify structural domains. Although *p*HtrII(1-159) protein possesses 14 potential trypsin cleavage sites (Figure 2.3A), only one major band was detected by SDS-PAGE of the proteolytic products of *p*HtrII(1-159) protein (Figure 2.3B). The proteolytic product of *p*HtrII(1-159) protein was purified (Figure 2.3C), and then subjected to N-terminal sequence and molecular weight analyses (Figure 2.3D, E). The N-terminal 6 amino acids of the proteolytic product were identical to residues 100-105 of *p*HtrII(1-159) protein. The molecular weight was 7939.49, as determined by MALDI-TOF MS, which was consistent with the theoretical value of 7938.43 for the *p*HtrII(100-159) fragment including the C-terminal six-residue histidine tag. Thus it was concluded that the stable structural domain was included in the trypsin resistant fragment *p*HtrII(100-159).

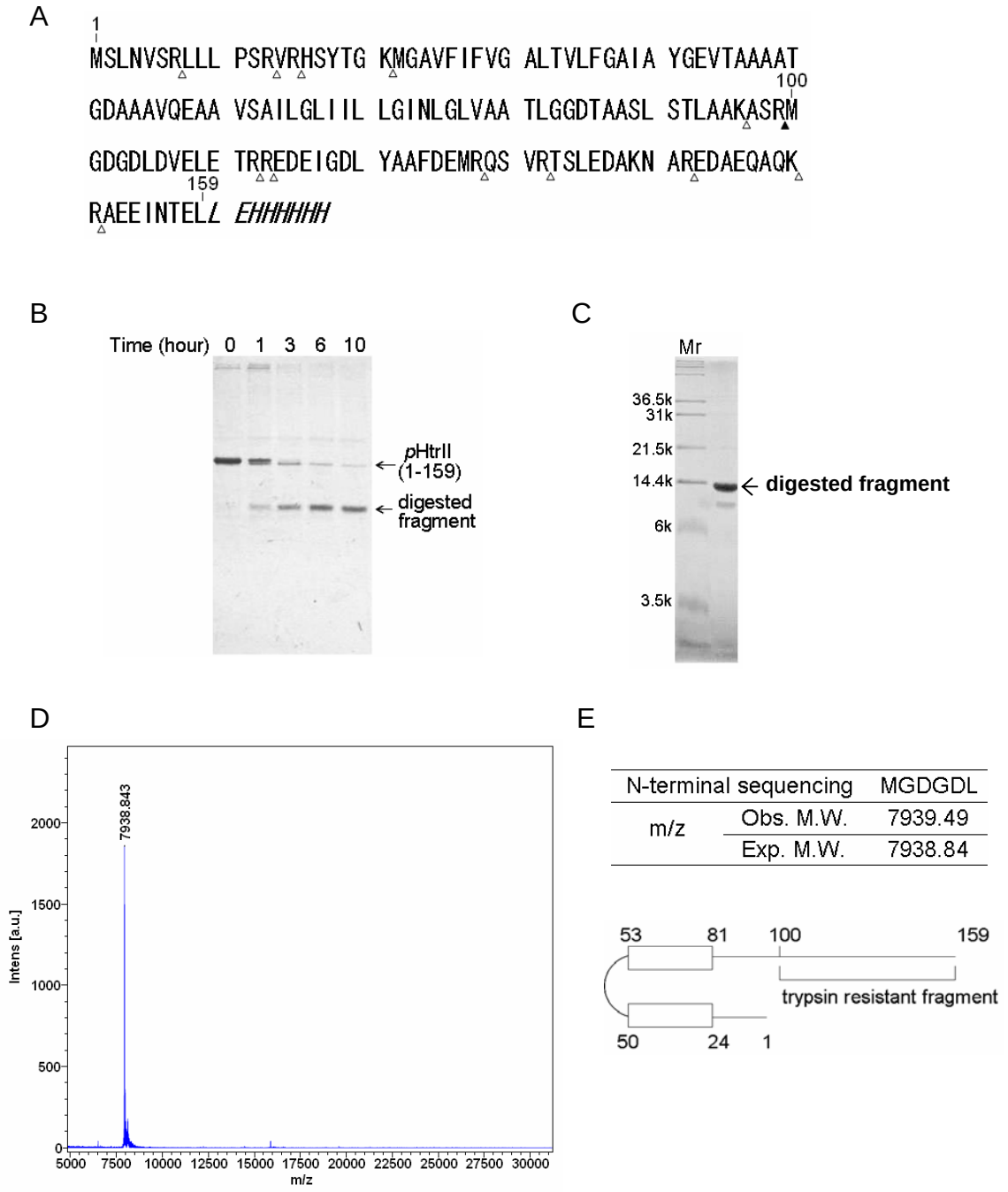


Figure 2.3. Limited trypsin digestion profile of *p*HtrII. (A) Amino acid sequence of the *p*HtrII(1-159) protein. Letters in italics refer to the six-residue histidine tag and additional linker between the tag and the *p*HtrII(1-159) protein. Δ, the putative sites of trypsin cleavage; ▲, trypsin-digested site in this experiment. (B) Time-course of trypsin digestion of the *p*HtrII(1-159) protein at 4°C analyzed by SDS-PAGE. (C) The digested fragment which was purified by gel filtration. (D) Mass spectroscopy of the digested fragment. (E) Identification of the trypsin resistant fragment using MALDI-TOF MS and N-terminal protein sequencing. The digested fragment comprises residues 100-159 of *p*HtrII.

2.3.2 Secondary structure analysis of the digested fragment *p*HtrII(100-159)

CD spectra were measured in an effort to investigate the secondary structure of the digested fragment *p*HtrII(100-159). A negative band at 208 nm with a shoulder at ca. 222 nm was observed at 277 K (Figure 2.4). These features are characteristic of the

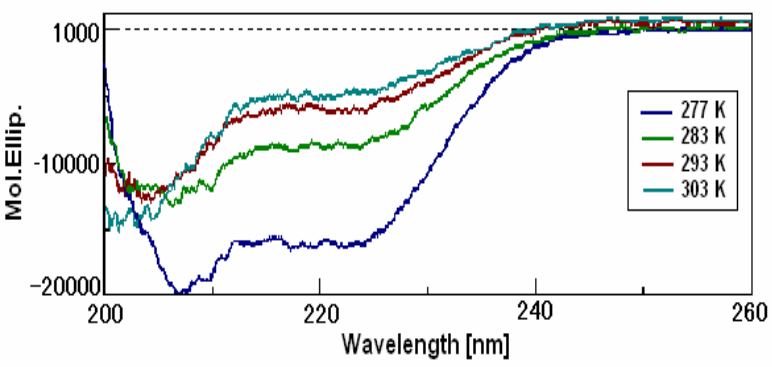


Figure 2.4. CD spectra of the *p*HtrII(100-159) fragment recorded in 5 mM MES buffer (pH 5.5) containing 25 mM KCl at 277 (dark blue), 283 (green), 293 (red) and 303 K (light blue).

presence of α-helical structures. The content of α-helix, β-sheet, turn and random coil regions were estimated to be 40%, 8%, 0% and 52%, respectively. The temperature dependence of the CD spectrum was monitored at 277, 283, 293 and 303 K (Figure 2.4). The magnitude of the negative Cotton band decreased reversibly with increasing temperature. These data indicate that the α-helical structure of the *p*HtrII(100-159) fragment is stable at low temperature. Therefore, all subsequent experiments were carried out at 277K, except for the pulsed ELDOR experiments.

2.3.3 Secondary structure analysis of the expressed *p*HtrII linker

For the preparation of a large amount of *p*HtrII(100-159) fragment, *p*HtrII(100-159) protein was expressed with N-terminal six-residue histidine tag and Xpress epitope tag in *E. coli*. Six-residue histidine tag and Xpress epitope tag were digested by Factor Xa protease and *p*HtrII(100-159) was successfully prepared. The secondary structure of the expressed *p*HtrII(100-159) was analyzed by CD. Unexpectedly, CD spectra showed that the α-helix content of the expressed protein was much less than the trypsin digested fragment. *p*HtrII(100-159). The α-helix structure of the digested fragment was most stable at pH 5 at 277 K, with α-helix content of 40%, whereas the α-helix content of the expressed protein was 22% in the same condition. These data showed that α-helix structure of linker region was unstable in the expressed protein.

To identify more stable structural region, additional expression plasmids were constructed in the linker region. On the basis of secondary structure prediction by PSIPRED and coiled coil region prediction by COILS, twelve kinds of expression plasmids were constructed in the linker region; 100-168, 100-178, 100-189, 100-205, 100-218, 100-272, 83-100, 83-159, 83-168, 83-178, 83-189 and 83-205. Two out of twelve fragments, 100-272 and 83-189 were successfully purified and analyzed secondary structure by CD. CD spectra show that the α-helix content of the expressed

fragments 100-272 and 83-189 were 10% and 9% respectively, and it was indicated that expressed protein does not have stable α-helix structure which was found in digested fragment. Thus, trypsin digested fragment 100-159 was used for structural analysis in chapter 3.

2.3.4. Development of a novel protein expression system

As described above, several kinds of expression plasmids were constructed in this study, and tried to express in *E. coli*. However some proteins were not expressed in *E. coli*. Especially, short proteins such as *p*HtrII(83-100) were degraded during expression step and target protein could not be obtained. To obtain such proteins by overexpression in *E. coli*, a novel protein expression vector, pCold-GST was developed. pCold-GST vector consists of the glutathione S-transferase (GST) tag and the pCold I cold-shock vector for high-level soluble protein expression in *E. coli* (see Appendix).

This pCold-GST vector was applied to ten proteins that were not expressed using the conventional methods. Nine out of ten proteins, including *p*HtrII(83-100) were successfully expressed in soluble fraction (Figure 2.5). These results suggest that pCold-GST expression system is highly effective in preparation of proteins.

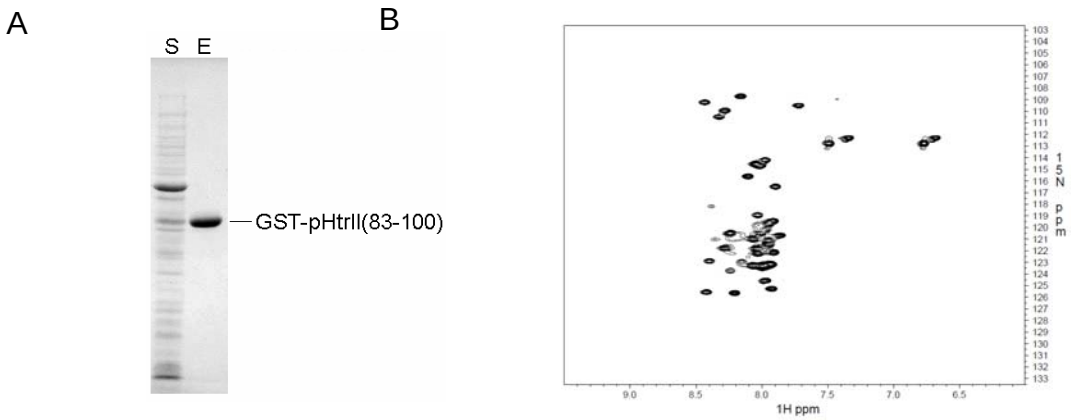


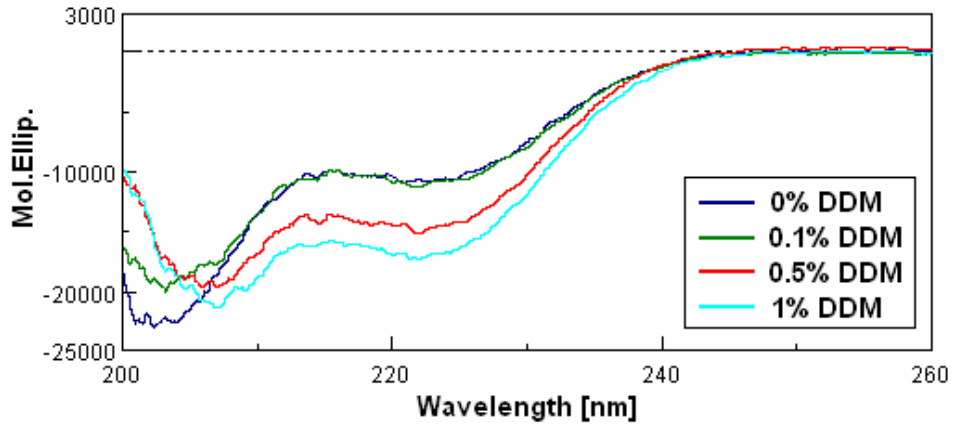
Figure 2.5. Expression and purification of GST-*p*HtrII(83-100) protein. (A) Expression and purification of GST-*p*HtrII(83-100)-His protein by using glutathione sepharose 4B resin. S, the supernatant fraction from the lysate; E, the elute fraction from the affinity resin. (B) ^{1}H-^{15}N HSQC spectrum of 1.1 mM GST-*p*HtrII(83-100)-His is shown. The spectrum was recorded in 20 mM Bis-tris buffer (pH6.8) containing 50 mM KCl at 293 K using a AVANCE800 NMR spectrometer (Bruker Biospin).

2.4 Discussion

2.4.1. α-helix structure in *p*HtrII(100-159)

In this study, the structural domain was identified within *p*HtrII(100-159). This domain was investigated using expressed protein *p*HtrII(100-159), where the amino acid sequence is identical to the digested fragment. Surprisingly, the α-helix structure of the expressed protein was less stable than the digested fragment, as revealed by CD. In our experiment, trypsin digestion was carried out under the presence of detergent (0.1% DDM) to solubilize the transmembrane region of *p*HtrII(1-159). This raises the possibility that the α-helix structure found in digested fragment has been induced by DDM. Then the structure of the expressed protein *p*HtrII(100-159) was analyzed in the presence of DDM. CD spectra were measured in the absence or presence of 0.1%, 0.5% or 1% DDM at 277K, and α-helix content was 12%, 11%, 32%, 35%, respectively (Figure 2.6A). These results showed that DDM could induce the α-helix structure of expressed *p*HtrII(100-159). However, the ^{1}H-^{15}N spectrum of expressed protein in the presence of DDM was different from the trypsin digested fragment (Figure 2.6B). These data show that the expressed protein *p*HtrII(100-159) can form α-helix structure, in the presence of DDM but the induced structure is different from the structure of the trypsin digested fragment.

A



B

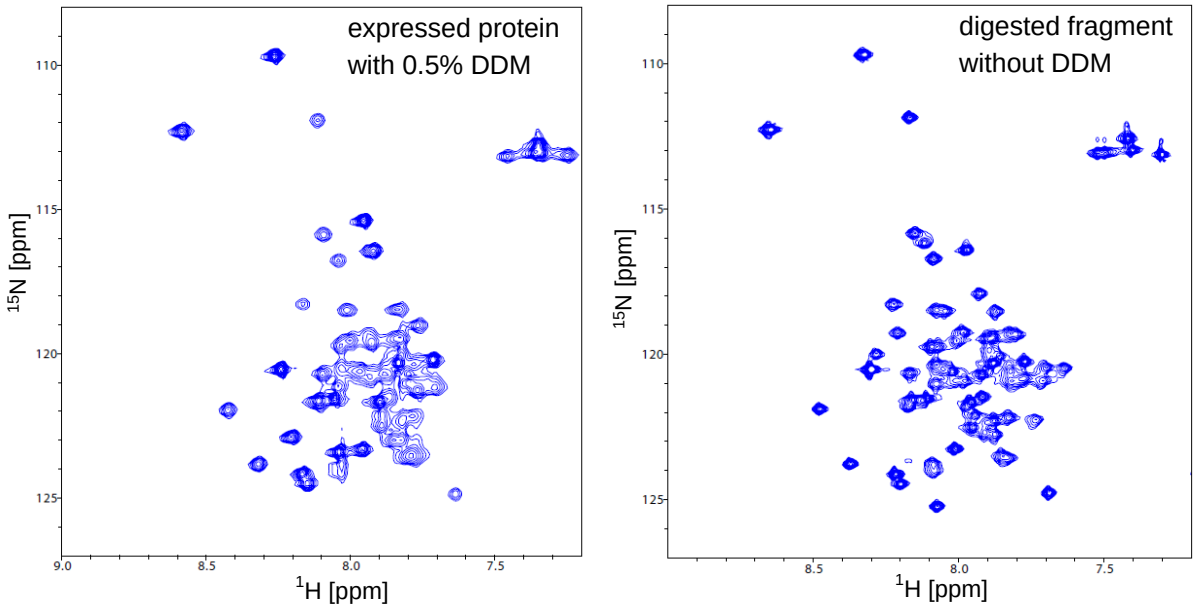


Figure 2.6. CD and NMR spectra of *p*HtrII(100-159) of the expression protein or digested fragment. (A) CD spectra of the expressed protein recorded in 5 mM MES buffer (pH 5.5) containing 25 mM KCl (dark blue), 0.1% DDM (green), 0.5% DDM (red), 1% DDM (light blue) at 277 K. (B) ^{1}H-^{15}N HSQC spectra of the expressed protein (left) and digested fragment (right). spectra were recorded in 10 mM citric acid buffer (pH5.0) containing 50 mM KCl with or without 0.5% DDM at 277 K.

In an effort to identify the extraneous factors that might contribute towards stabilization of the α-helix in the digested fragment, further investigations were

performed using NMR, MS, thin layer chromatography, and SDS-PAGE with silver staining. So far, no extraneous chemical component, such as a peptide, an amino acid and a lipid component, were detected from the digested fragment.

To identify the exact α-helical region in the *p*HtrII(100-159), shorter region *p*HtrII(100-159) was expressed and analyzed by CD. The α-helix content of expressed *p*HtrII(100-149) was 0% under the same condition. This data indicates that the C-terminal region of *p*HtrII(100-159) is important to form a-helical structure, that is the C-terminal region is a part of the α-helix or a critical factor to stabilize the α-helix.

3. Characterization of digested fragment *p*HtrII(100-159)

3.1 Introduction

The functional importance of the linker region has been investigated in *p*HtrII [20, 24, 25] and other MCPs [26, 27] in relation to signal transduction. The linker region contains a HAMP domain which has been identified by sequence comparisons and mutation analyses of sensor kinases and MCPs [33-35]. Recently, the NMR structure of the HAMP domain has been reported for the non-MCP protein Af1503, which was found to contain a HAMP domain as its sole cytoplasmic part [1]. The HAMP domain consists of two helical amphipathic sequences (AS-1 and AS-2) linked by a connector sequence of undefined secondary structure. Unlike the AS-2 helix, the presence of the AS-1 helix in the *p*HtrII linker region is detected by EPR [31]. Almost all MCPs possess a linker region comprising ca. 60 residues, although the *p*HtrII linker is twice as long as many other MCPs and its role is unknown [33]. Long linker regions are also present in *H. salinarium* HtrI and HtrII [38, 39] and is expected to play a unique role in phototactic signal transduction.

In this chapter, the identified structural domain *p*HtrII(100-159) was characterized by NMR and analytical ultracentrifugation. The high-resolution structure of this domain

was determined by NMR and further characterized by CW-EPR using detergent-solubilized and liposome-reconstituted *p*HtrII protein. Finally, the role of the linker region in signal transduction is discussed based on the ELDOR data for the photo-activated *p*pR-*p*HtrII fusion protein and the structural model of the whole linker and transmembrane region of *p*HtrII complexed with *p*pR.

3.2 Materials and Methods

3.2.1 Sample preparation for NMR measurement

For the structure determination, a trypsin digested fragment *p*HtrII(100-159) with a C-terminal six-residue histidine tag was used. For the uniform labeling of proteins with ^{15}N- and $^{15}N/^{13}C$, cells were grown in M9 minimal medium containing $^{15}NH_4Cl$ and [$^{13}C_6$]-glucose as the sole nitrogen and carbon source, respectively. The uniform labeled *p*HtrII(1-159) was expressed and purified by the same method described in chapter 1. *p*HtrII(1-159) was digested by trypsin and purified by gel filtration chromatography by the same method described in chapter 1.

3.2.2 Sedimentation equilibrium by analytical ultracentrifugation

Sedimentation equilibrium experiments were performed using an Optima XL-A analytical ultracentrifuge (Beckman) with a 4-hole An60Ti rotor at 4 °C containing six-sector cells. Concentration profiles of 20 and 800 μM ^{15}N enriched samples in buffer B [10 mM citric acid buffer (pH 5.0) containing 50 mM KCl] were monitored by absorbance at 215 and 280 nm, respectively, at rotor speeds of 21,000, 30,000 and 39,000 rpm. The molecular weight of each sample was determined based on a self-association model using Optima XL-A/XL-I data analysis software version 6.03

(Beckman). Solvent viscosity and density were calculated using SEDNTREP (John Philo, Amgen).

2.2.3 NMR spectroscopy and structure calculation of *p*HtrII(100-159)

Purified *p*HtrII(100-159) fragment was dissolved in buffer B containing 90% H_2O/10% ^2H_2O mixture. NMR spectra were acquired at 277 K using a DRX800 NMR spectrometer (Bruker Biospin). Resonance assignments for ^{1}HN, ^{15}N, ^{13}Cα, ^{13}Cβ and ^{13}C' nuclei for the *p*HtrII(100-159) fragment were obtained through the following triple resonance spectra applied to the ^{15}N/^{13}C-labeled sample: 3D HNCACB, 3D HN(CO)CACB, 3D HN(CA)CO and 3D HNCO spectra [40]. Hα resonances were assigned using 3D H(CACO)NH, and side-chain ^{1}H and ^{13}C resonances were assigned using 3D CCONH, 3D TOCSY-HSQC, 3D NOESY-HSQC and 3D HCCH-TOCSY[40]. 3D ^{15}N-edited NOESY (100 ms mixing time) and 2D ^{1}H NOESY (100 ms mixing time) were measured to obtain distance restraints [40]. All NMR spectra were processed using NMRPipe/NMRDraw [41] and analyzed using Sparky [42]. ϕ and ψ angles were obtained using TALOS[43]. CYANA version 2.1 with the CANDID protocol was used for the structure calculation[44]. The 100 structures obtained were refined using AMBER version 9.0 [45] with 20 ps molecular dynamics simulation using an implicit generalized born solvent model and 1,500 cycle energy minimization. The final 20

lowest-energy structures were checked using PROCHECK-NMR[46]. All graphics were created using MOLMOL [47].

3.2.4 *p*pR-*p*HtrII(1-159) fusion protein expression and purification

PCR was used to introduce 27 nucleotides encoding a 9-residue linker (Ala-Ser-Ala-Ser-Asn-Gly-Ala-Ser-Ala [20] or Ala-Ser-Ala-Ser-Asn-Gly-Arg-Arg-Arg) between the C-terminal residue of *p*pR and the N-terminal residue of *p*HtrII(1-159). This fusion protein, referred to as *p*pR-*p*HtrII(1-159), was expressed and purified as described for the *p*HtrII(1-159) protein except for the addition of 10 mM all-trans retinal during the induction.

3.2.5 Expression and purification of cysteine mutants

For preparation of the cysteine mutants, a Quickchange site-directed mutagenesis kit (Stratagene) was used. Cysteine mutants of the *p*HtrII(1-159) and *p*pR-*p*HtrII(1-159) proteins were expressed and purified as described for the wild-type proteins. except that 15 mM 2-mercaptoethanol was present during the purification. When the cysteine mutants of *p*pR-*p*HtrII(1-159) were induced, 10 μM all-trans retinal was added. The crude membranes were solubilized and the supernatant was collected as described above. The supernatant was loaded onto a Ni^{2+}-NTA agarose column. The resin was washed

extensively with buffer D [50 mM MES buffer (pH6.5) containing 300 mM KCl, 15 mM 2-mercaptethanol and 0.1% DDM] containing 50 mM imidazole to remove nonspecifically bound proteins. The histidine-tagged protein was eluted with buffer D containing 500 mM imidazole (Figure 3.1A).

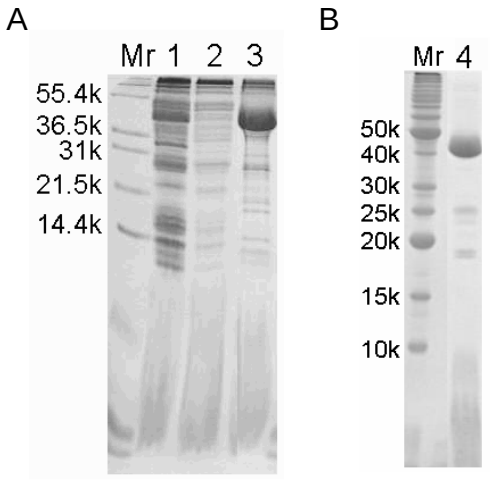


Figure 3.1. Protein expression and purification of *p*pRS154C-*p*HtrII(1-159)S98C chimera protein. (A) Protein expression and purification by Ni-NTA affinity resin. Mr, molecular weight marker; 1, the supernatant fraction after solubilization of membrane fraction; 2, pellet; 3, the elute fraction from the affinity resin. (B) final sample of spin-labeled *p*pRS154C-*p*HtrII(1-159)S98C chimera protein . Mr, molecular weight marker; 4, purified protien.

3.2.6 Spin labeling and reconstitution in PC liposomes

The spin-label (1-oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) methanethiosulfonate (MTSL; Toronto Research Chemicals Inc.) was covalently attached to cysteine residues of solubilized *p*HtrII(1-159) or *p*pR-*p*HtrII(1-159) protein according to the reported protocol [48]. The mutants were treated with 15 mM 2-mercaptoethanol, 3 mM DTT and 0.1% DDM to reduce disulfide bonds or oxidized –SH residues. 2-mercaptoethanol and DTT were removed by passage through a HiPrep

26/10 desalting column (GE Healthcare) equilibrated with 50 mM MES buffer (pH6.8) containing 50 mM KCl and 0.1% DDM. Then 100 mM MTSL in DMSO was added to the protein solution to a molar ratio of 10 (MTSL): 1 (protein). Samples were then incubated at 20°C for 4 hours. Excess MTSL was removed by gel filtration using HiLoad 16/60 Superdex 200 column (GE Healthcare) in buffer B [10 mM citric acid buffer (pH 5.0) containing 50 mM KCl and 0.1% DDM] (Figure 3.1B). Spin-labeled proteins were checked using MS (Table 3.1) and mixed with twice the amount of non-labeled protein. Spin-labeled proteins were mixed with twice the amount of non-labeled protein for the EPR analysis.

Table 3.1: Molecular weight of spin-labeled and non-labeled *p*HtrII$^{1-159}$

position of spin-label		non-label (WT)	134/145	138/145	141/145
m/z	Obs. M.W.	17591.79	18008.31	18024.31	18024.31
	Exp. M.W.	17600.80	18029.40	18019.33	18023.56

For reconstitution into L-α-phosphatidylcholine (PC), 20 mg of egg PC lipid (Avanti Polar Lipids, Inc.) was dissolved in 1 ml of chloroform, and a thin lipid film on the wall of a 5 ml flask was prepared by careful evaporation of the solvent[49]. One ml of buffer B was added to the flask followed by the addition of protein to a molar ratio of

1:70 of protein to lipid. Bio-Beads SM-2 Adsorbent (BIO-RAD) was added to remove the DDM and the sample was then gently stirred overnight at 4°C. The reconstituted protein was pelleted by centrifugation at 20,000 × g for 30 min at 4°C, and then resuspended in 10 mM citric acid (pH5.0) containing 50 mM KCl or 50 mM Tris buffer (pH 7.2 for CW-EPR and pH 7.0 for pulsed ELDOR) containing 50 mM KCl.

3.2.7 EPR measurements

Continuous wave (CW) EPR spectra were recorded using a JES-TE300 EPR spectrometer (JEOL) at 277, 293 or 298 K, with a microwave power of 3.0 mW and modulation amplitude of 0.1 mT. Samples were then loaded into an EPR flat cell (4.2 x 0.4 x 41.7 mm). To avoid any intermolecular effect, 1:2 ratio labeled/unlabeled sample mixtures were used. This helps to ensure the absence or a negligible amount of doubly-labeled dimer or oligomer in the sample. Solubilized samples were prepared in buffer C, and 50 mM Tris buffer (pH 7.2) containing 0.1% DDM and either 50 mM KCl, 4 M KCl, 2M KCl + 2M NaCl, or 4M NaCl.

Pulsed electron-electron double resonance (ELDOR) [50] was measured using an ELEXSYS E580 FT-EPR spectrometer fitted with a second microwave unit (Bruker Biospin). The spectrometer was equipped with a cylindrical dielectric resonator

(ER4118X-MD5-W1, Bruker Biospin) and He gas-flow system (CF935, Oxford Instruments). To avoid any intermolecular effect, 1:2 ratio labeled/unlabeled sample mixtures were used. Solubilized samples were prepared in buffer C containing 10% glycerol or 50 mM Tris buffer (pH 7.0) containing 50 mM KCl, 0.1% DDM and 10% glycerol, and then loaded into EPR capillaries (3-mm inner diameter). Interspin distances were determined using simulation software [51].

3.2.8 Photoactivation of *p*pR

To measure the distances between the M intermediate *p*pR and *p*HtrII linker, samples were placed in an EPR cell and illuminated (Figure 3.2). Illumination with > 480 nm light (VY-50, Toshiba), which converted *p*pR to ppR$_M$, was achieved using a 1 kW halogen-tungsten lamp (Rikagakuseiki) at 253K for 90s [52]. The spin label in *p*pR at position 154 has no affect on the photocycle kinetics[53].

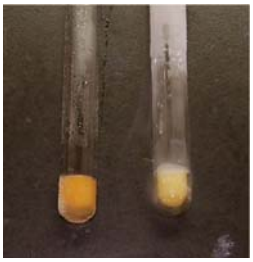


Figure 3.2. Photoactivation of *p*pR-*p*HtrII fusion protein.
Before (left) and after (right) illumination with > 480 nm light.

3.3 Results

3.3.1 Concentration-dependent dimer formation of the *p*HtrII(100-159) fragment

^{1}H-^{15}N HSQC spectra of the *p*HtrII(100-159) fragment were measured with 20, 45, 75, 150, 300, 690 or 800 μM protein. In Figure 3.3A, the chemical shift differences between 20 and 690 μM samples were plotted against residue number. Although the magnitude of the difference was relatively small, the chemical shift differences were localized to residues 105-120. The semi-log plots of these concentration-dependent peak shifts were sigmoidal in shape. The plot of residue R113 is shown in the inset of Figure 3.3A. The sigmoidal curves indicate the presence of a concentration-dependent two-state transition. Assuming monomer-dimer equilibrium, the apparent K_d value was determined to be 115 μM.

This equilibrium was further investigated by sedimentation equilibrium analytical ultracentrifugation at low and high protein concentrations (Figure 3.3B). As shown in Table 3.2, the molecular weights determined of the ^{15}N enriched *p*HtrII(100-159) fragments were 8.1 and 17.1 ± 1.8 kDa at 20 and 800 μM, respectively. These experimental values are consistent with the theoretical molecular weights of monomer and dimer, being 8,044.4 and 16,088.8, respectively. Thus it was concluded that the

*p*HtrII(100-159) fragment forms a homodimer.

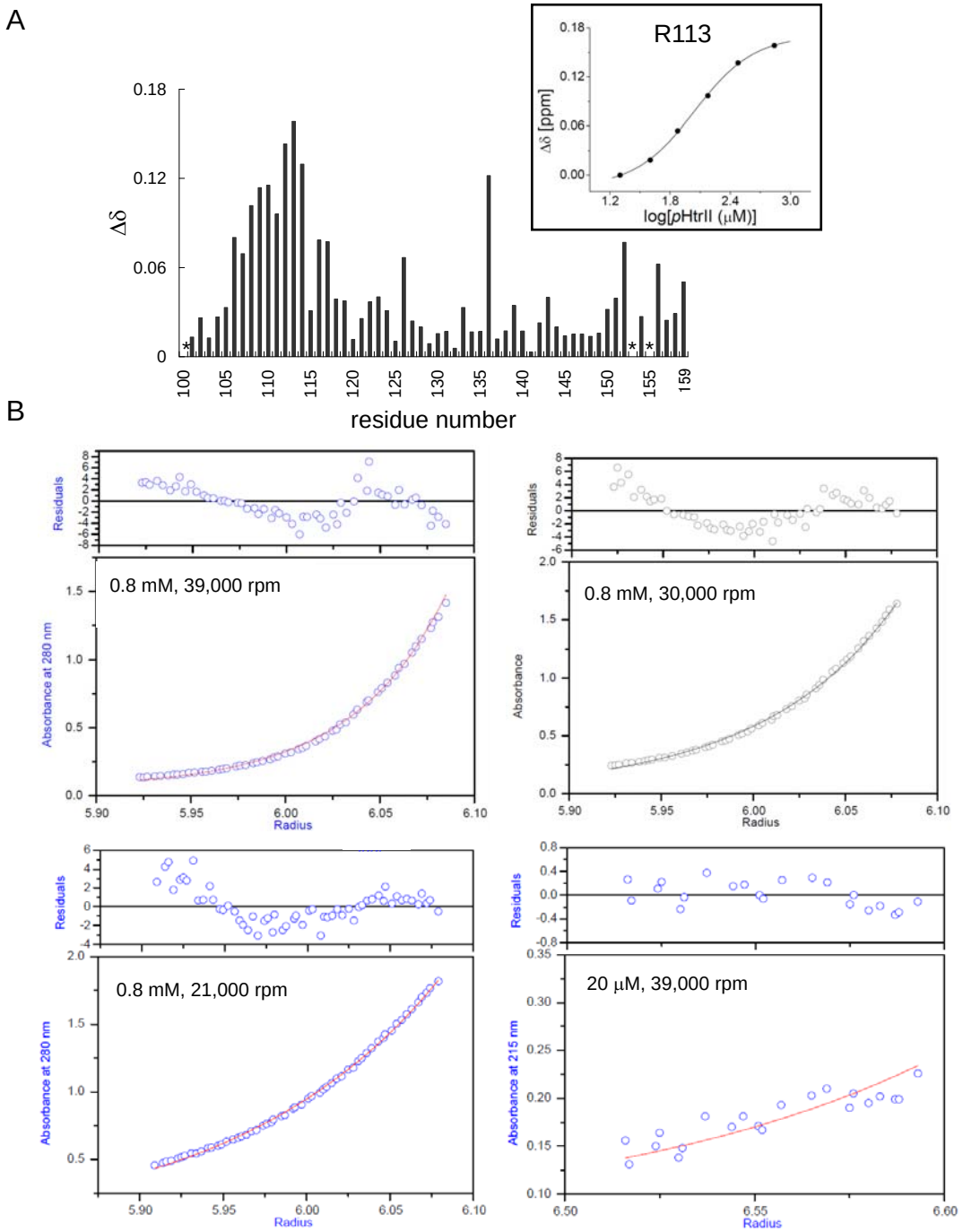


Figure 3.3. (A) Protein concentration-dependence of chemical shifts of the *p*HtrII(100-159) fragment. The bars represent combined ^{1}H and ^{15}N chemical shift differences ($\Delta\delta = [\Delta\delta(^{1}H)^{2} + \{\Delta\delta(^{15}N) / 5\}^{2}]^{1/2}$) between 20 μM and 0.69 mM samples. The asterisks indicate unassigned residues. A plot of the protein concentration against Δδ of the R113 residue is shown in the inset. (B) Ultracentrifugation analysis of the *p*HtrII(100-159) fragment. Sedimentation equilibrium data of 0.8 mM or 20 μM proteins are shown at a rotor speed of 21,000 or 39,000 or 30,000 rpm with the curve fitted to the self-association model.

Table 3.2. Molecular weight of *p*HtrII(100-159)[a]

concentration (μM)	observed value (kDa)	calculated value (Da)
20	8.1	8,044.4 (monomer)
800	17.1 ± 1.8[b]	16,088.8 (dimer)

[a]Molecular weight determined by sedimentation equilibrium analytical ultracentrifugation. Calculated values for monomer and dimer are also shown. For the equilibrium analysis a total of ten data sets were globally fitted to a self-association model at each rotor speed. [b]Averaged for three rotor speeds; 21000, 30000 and 39000 rpm. Error value represents the mean error.

3.3.2 Solution structure determination of the *p*HtrII(100-159) fragment

For the NMR structure determination, isotopically-labeled recombinant *p*HtrII (1-159) protein was overexpressed in *E. coli* and purified by affinity column chromatography. Following trypsin digestion, the *p*HtrII(100-159) fragment was purified by size exclusion and affinity column chromatography, and used for 3D structure determination by solution NMR. The solution comprised buffer B without detergent since the presence of 0.1% DDM did not change the NMR spectrum. NMR resonance assignments were obtained by performing double- and triple-resonance NMR experiments using ^{15}N-labeled and ^{15}N- and ^{13}C-labeled protein samples. Almost all of the main chain and H_α resonances were assigned by conventional triple resonance techniques, although several C_β peaks could not be detected. ^{13}C resonances of 72% of

the side chains could be obtained using 3D CCONH and 3D HCCH-TOCSY, and 58% of the ^{1}H resonance assignments of the side chains including 85% H_{β} and 52% H_{γ} were obtained using 3D TOCSY-HSQC, 3D HCCH-TOCSY and 3D NOESY-HSQC. Most of the sequential d_{NN} and medium-range $d_{\alpha N}$(i, i+2), $d_{\alpha N}$(i, i+3), $d_{\alpha N}$(i, i+4) NOEs were manually assigned using the ^{15}N-edited NOESY spectrum. The observed medium-range NOEs indicated the presence of α-helical content in at least residues 137-147 (Figure 3.4A).

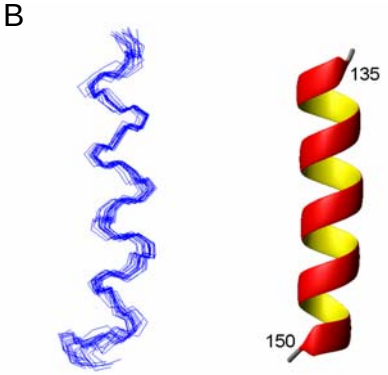


Figure 3.4. NMR structure of the *p*HtrII(100-159) fragment. (A) Summary of the sequential NOE connectivities observed for the *p*HtrII(100-159) fragment. Bars indicate NOESY cross-peaks observed between two residues. Line thickness indicates the strength of the NOE. (B) Superimposition of 20 lowest-energy structures from the final AMBER calculations (left), and ribbon drawing of the representative structure (right).

Unambiguous NOESY peaks were manually identified from the ^{15}N-edited NOESY spectrum and used for initial distance restraints. The dihedral angle constraints for ϕ and ψ were obtained using TALOS, which also supported the presence of α-helical content in residues of 137-152. For the structure calculation, the monomer unit was used to simplify the calculation since no long-range NOE was observed. The first structure was calculated using CYANA version 2.1 and employing the CANDID protocol for ^{15}N-edited NOESY and 2D ^{1}H-^{1}H NOESY. The CANDID protocol provided a total of 266 distance restraints. The final structure was obtained by restrained molecular dynamics using AMBER version 9.0. Figure 3.4B shows the final 20 superimposed structures in ribbon representation, with disordered regions being excluded. The final structure shows good alignment of an α-helix between residues L135 and K150. The rms deviations calculated from the averaged structure were 0.42 and 1.12 Å for backbone and heavy atoms of the converged region (residues 135-150), respectively. The statistics of the structure are given in Table 3.3.

Table 3.3. Structural statistics for *p*HtrII(100-159)[a]

total number of distance constraints	266
short-range (|i-j| ≦1)	224
medium-range (|i-j| < 5)	42
long-range (|i-j| ≧ 5)	0
dihedral constraints (ϕ, φ)	22, 22
hydrogen bonds	9
rms deviation from experimental constraints[b]	
distance (Å)	0.0281 ± 0.0036
angles (deg)	0.485 ± 0.022
rms deviation from idealized covalent geometry	
bond (Å)	0.00362 ± 0.00015
angles (deg)	0.485 ± 0.022
impropers (deg)	0.412 ± 0.020
AMBER energy terms (kcal/mol)	
Eamber	-3229 ± 16
van der Waals energy	-365 ± 12
Electrostatic energy	-3892 ± 130
Constraint energy	5 ± 1
Ramachandran plot[c] (residues 135-150)	
residues in most favored regions (%)	99.4
residues in additionally allowed regions (%)	0.6
residues in generously allowed regions (%)	0.0
residues in disallowed regions (%)	0.0
rms deviation of mean structure derived from 20 calculated structures	
backbone atoms (residues 135-150) (Å)	0.42 ± 0.17
all heavy atoms (residues 135-150) (Å)	1.12 ± 0.12

[a]These statistics comprise an ensemble of the 20 lowest-energy structures obtained from 100 starting structures. Structure was determined using CYANA version 2.1 with the CANDID protocol and refined using AMBER version 9.0. [b]None of these structures exhibited distance violations > 0.5 Å or dihedral angle violations > 5°. [c]Determined using PROCHECK.

3.3.3 Structure analysis of residues 135-150 of *p*HtrII(1-159) protein solubilized in DDM or reconstituted in PC liposomes

Residues 135-150 of the *p*HtrII(100-159) fragment formed an α-helix in water. This structure was investigated for the transmembrane protein *p*HtrII(1-159) using CW-EPR. The following single- or double-cysteine mutants of *p*HtrII(1-159) protein were prepared for the CW-EPR analysis: A145C, A141C, A138C, S134C, A145C/A141C, A145C/A138C and A145C/S134C. The mutants were spin-labeled with MTSL at each cysteine residue. CW-EPR spectra were measured for the samples solubilized in 0.1% DDM or reconstituted into PC liposomes. Figure 3.5 shows the spectra of doubly-labeled sample (gray line) and the summed spectra of singly-labeled samples (black line). Line broadening was observed for all doubly-labeled samples that had either been solubilized or reconstituted. The degree of broadening was consistent with the inter-spin distances of the spin-labels assuming an α-helical structure for residues 135-145 of *p*HtrII(1-159) protein. Under the presence of 1 M guanidine hydrochloride, the line broadening considerably disappeared and much sharper spectra were observed.

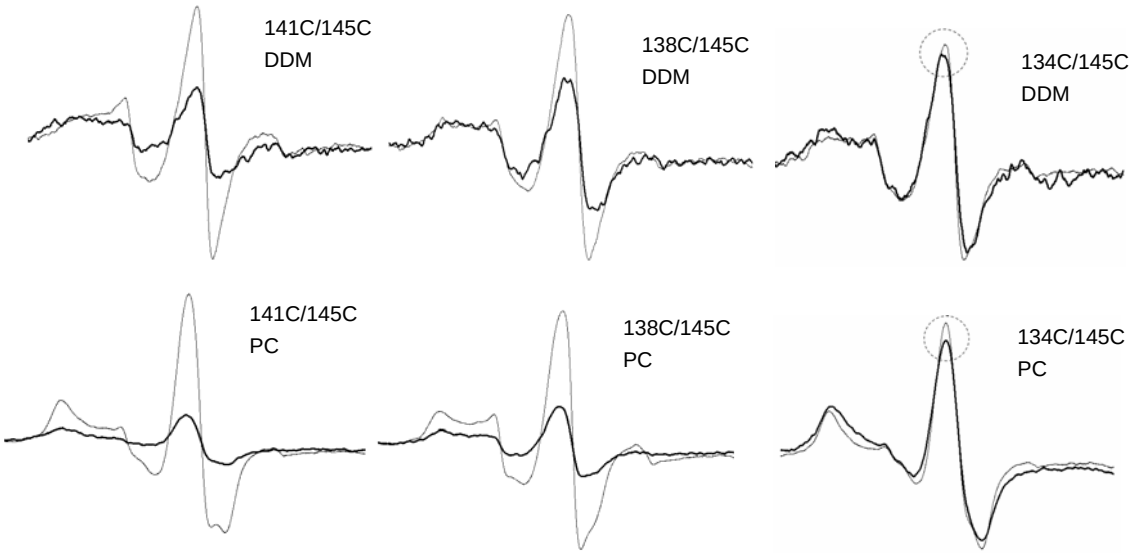


Figure 3.5. CW-EPR spectra of the *p*HtrII(1-159) protein at 277 K. 134C, 138C, 141C and 145C refer to the spin-labeled position. DDM and PC refer to the solubilized sample in 0.1% DDM and the reconstituted sample into PC liposomes, respectively. The summed spectra of two singly-labeled samples (gray) and the spectrum of a doubly-labeled sample (black) are superimposed.

3.4 Discussion

3.4.1 The α-helix structure of residues 135-150

In this study, the α-helical structure was determined by NMR using fragment *p*HtrII(100-159) obtained following trypsin digestion. The presence of an α-helix within the transmembrane protein *p*HtrII(1-159) was confirmed using EPR at pH 5 and 277 K (Figure 3.6). At pH 7.2 and 298 K, dipolar broadening was observed for both the solubilized and reconstituted samples (Figure 3.6), although the degree of broadening

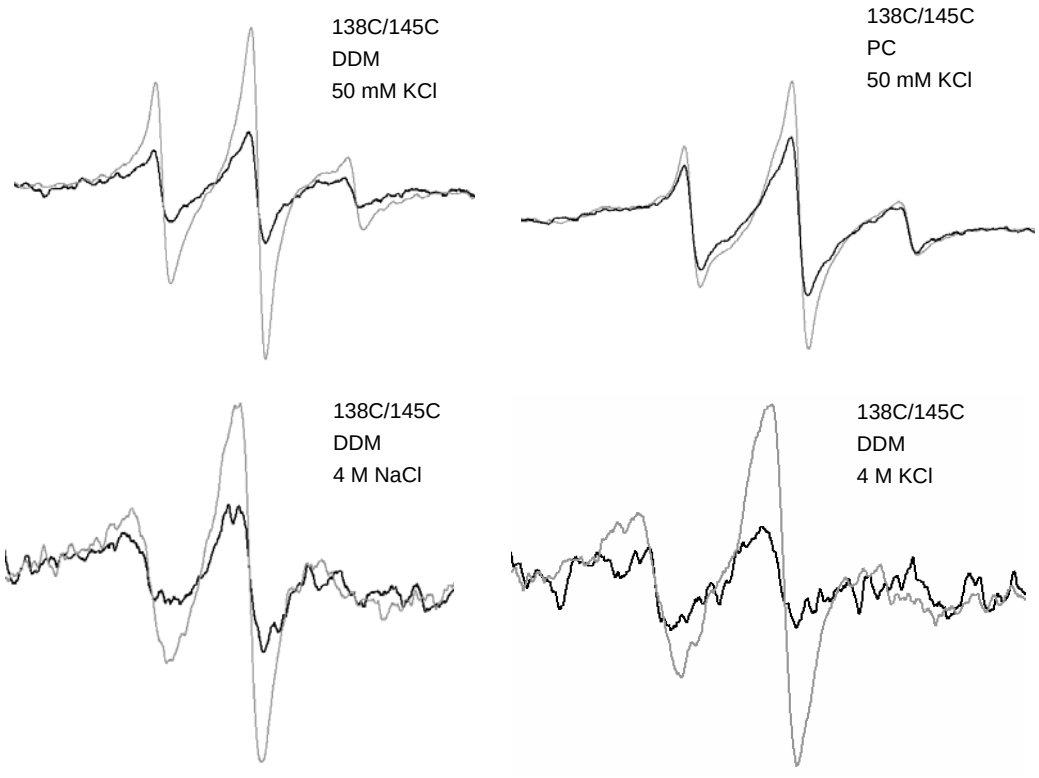


Figure 3.6. CW-EPR spectra of the *p*HtrII(1-159) protein at 277 K. 134C, 138C, 141C and 145C refer to the spin-labeled position. DDM and PC refer to the solubilized sample in 0.1% DDM and the reconstituted sample into PC liposomes, respectively. The summed spectra of two singly-labeled samples (gray) and the spectrum of a doubly-labeled sample (black) are superimposed.

decreased slightly compared to the spectra measured at pH 5.0 and 277 K. It was noted that the α-helical content of digested fragment *p*HtrII(100-159) almost disappeared at 298 K (Figure 2.4). These results indicate that the α-helix found within the digested fragment is significantly stabilized within transmembrane protein *p*HtrII(1-159), possibly because of the homodimer stabilization via the transmembrane region.

3.4.2 Structure of *p*HtrII(1-159) residues 100-134

α-helical content is present only between residues 135-150 of the *p*HtrII(100-159) fragment. Only one major trypsin resistant fragment *p*HtrII(100-159) was detected, although there 4 potential trypsin cleavage sites that lie between residues 100-134 (Figure 2.3A). This resistance to trypsin digestion indicates the presence of protection factors located between residues 100-134, that might relate to the unique structure formed between residues 100-134 and/or intermolecular association. The former possibility is suggested by the experimental data.

The α-helical content of 40%, determined by CD analysis, is significantly higher than the 24% α-helical content calculated from the NMR structure. The chemical shift values and some NOE data suggest the formation of helical structure between residues 100-134. Thus, it is reasonable to assume that the structure adopted by residues 100-134

represents an unstable α-helix or a discontinuous α-helix. Presumably this α-helix is stabilized within the transmembrane protein *p*HtrII, such as with the α-helix located between residues 135-150 shown here, and with stabilization of the α-helix of the EnvZ HAMP domain in the presence of A domain [54]. The presence of intermolecular association is clear since the digestion experiment was performed using 0.48 mM protein, a concentration much higher than the K_d value of dimer formation, 115 μM. However, it remains unclear whether dimer formation can protect the protein from trypsin digestion, or if residues 100-134 can indeed form a dimer interface.

3.4.3 Dimer interface of *p*HtrII(100-159) fragment

The trypsin digested fragment *p*HtrII(100-159) forms a homodimer as determined by analytical ultracentrifugation (Figure 3.3). Significant NMR chemical shift changes upon dimer formation are localized to residues 105-120. Chemical shift changes are related changes in chemical environment, such as occurring at an interaction interface or manifested by a conformational change. Thus the dimer interface and/or the conformational change induced by dimer formation is localized to residues 105-120. The intermolecular interaction affects NMR signals as the chemical shift changes and/or line broadening occurs. Since no sequence-specific line broadening was observed upon dimer formation, I can conclude that the dimer interface is localized to residues 105-120.

This conclusion is consistent with previously reported EPR results showing close contacts between residues 115 and 115 (inter-dimer) and residues 122 and 122 (inter-dimer)[31]. The presence of induced conformational change remains uncertain due to a lack of data. The apparent disagreement between the NMR structure and the limited digestion data can be explained by the presence of an unstable α-helix between residues 100-134 and dimer formation involving residues 105-120.

3.4.4 HAMP domain in *p*HtrII

A HAMP domain is found in the *p*HtrII linker region and consists of two helices AS-1 and AS-2 located between residues 83-100 and 117-131, respectively [31]. The solution structure of the Af1503 HAMP domain was recently determined [1]. Sequence alignment of the *p*HtrII linker region and Af1503 HAMP domain was performed using CLUSTALW (Figure 3.7). Based on the secondary structure of Af1503, the first and second helices were estimated to be located between residues 83-102 and 115-136,

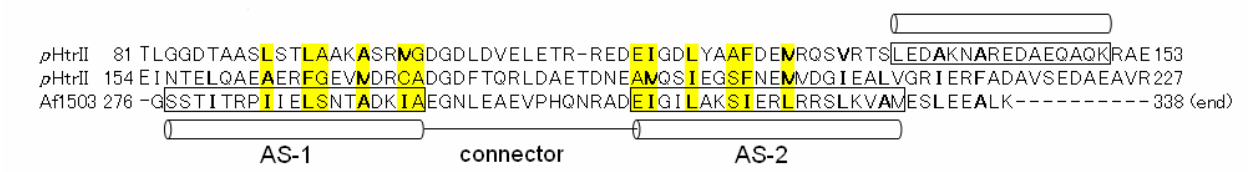


Figure 3.7. Sequence alignment of *p*HtrII and *Archaeoglobus fulgidus* Af1503 HAMP domain. Hydrophobic residues in heptad repeats are shown in bold. The helix region 3D structure determined in this investigation (*p*HtrII) and previously reported (Af1503) [1]are boxed in black. Important regions involved in packing interactions in Af1503 are highlighted in yellow.

respectively, and the HAMP domain was localized to residues 83-136. In Figure 3.7, another HAMP domain is shown in the *p*HtrII linker region. Based on the secondary structure of Af1503, the first and second helices of the second HAMP domain were estimated to be located between residues 157-176 and 190-211, respectively, and the second HAMP domain was localized to residues 157-211 of *p*HtrII. Using PROSITE [55-58], these two HAMP domains within the linker region were also found as the first and second HAMP domains located between residues 84-136 and 157-210, respectively.

In this study only one α-helix, located between residues 135-150, was determined for the digested fragment *p*HtrII(100-159). This region lies outside of the HAMP domain, and no clear structure is observed within the HAMP domain. However, as discussed above, the structure adopted by residues 100-134 of the HAMP domain is assumed to be that of an unstable α-helix. Moreover, residues 105-120 represent the dimer interface of the digested fragment. These characteristic features are consistent with those of the HAMP domain. Assuming a four-helix bundle structure for the first HAMP domain, the presence of an unstable α-helix located between residues 100-134 can easily be accounted for by the absence of the first helix in the first HAMP domain.

The structure of the HAMP domain was examined by pulsed electron-electron double resonance (ELDOR) method using the *p*pR-*p*HtrII(1-159) fusion proteins

*p*pR$_{S154C}$-*p*HtrII(1-159)$_{S98C}$ and *p*pR$_{S154C}$-*p*HtrII(1-159)$_{S134C}$ with an MTSL spin-label at each cysteine residue (Figure 3.8). The distances between the *p*pR E-F loop and the *p*HtrII linker region are shown in Table 3.4. The *p*pR$_{154}$-*p*HtrII$_{98}$ distance of ~ 20 Å for the PC reconstituted sample is slightly shorter than that expected from the reported

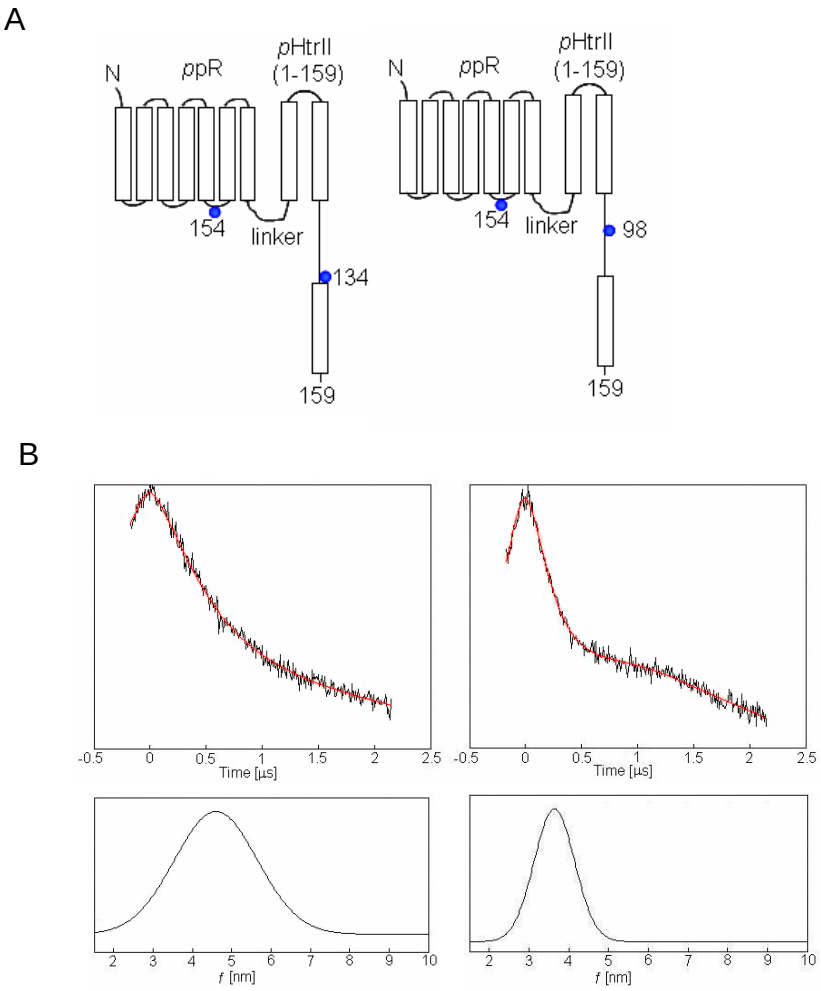


Figure 3.8. Inter-spin distance measurement using ELDOR. (A) Schematic representation of spin-labeled residues *p*pR$_{154}$ and *p*HtrII$_{98}$ or *p*HtrII$_{134}$. Distances between *p*pR$_{154}$ and *p*HtrII$_{98}$ or *p*HtrII$_{134}$ were determined. (B) ELDOR time traces (black line) and simulated spectra (red line) are shown in the upper panels with distance distribution data (lower panels).

values [31]. The *p*pR$_{154}$-*p*HtrII$_{98}$ distance of ~ 35 Å for the DDM-solubilized sample was much longer than that of the reconstituted sample. These distances were independent of pH and linker sequence between *p*pR and *p*HtrII. This difference in distance between the solubilized and reconstituted samples may support the proposition that the detergent-solubilized structure is impaired in the juxtamembrane region [53]. The *p*pR$_{154}$-*p*HtrII$_{134}$ distance for the PC reconstituted sample was not determined due to the fast electron spin relaxation. The *p*pR$_{154}$-*p*HtrII$_{134}$ distance ~ 45 Å for the DDM-solubilized sample was much shorter than that of the extended structure and suggests the presence of a folded structure. Although ELDOR analysis could not clarify the detailed structure of the HAMP domain, these data suggest the presence of some distinct structure, such as a four-helix bundle.

3.4.5 Role of the α-helix located between residues 135-150 in signal transduction

The α-helical structure between residues 135-150 determined in this report together with the crystal structure of the *p*pR-*p*HtrII complex and the modeled structures of two HAMP domains are shown in Figure 3.9. The HAMP domain structures were generated based on the structure of the Af1503 HAMP domain using MODELLER [59]. The *p*pR$_{154}$-*p*HtrII$_{98}$ and *p*pR$_{154}$-*p*HtrII$_{134}$ distances of the modeled structure shown in Figure 3.8 are ca. 20 and 35 Å, respectively. The *p*pR$_{154}$-*p*HtrII$_{98}$ distance is in good agreement

with our ELDOR data for the PC reconstituted sample (Table 3.4). The α-helix located between residues 135-150 is a direct continuation of the second helix of the first HAMP domain which connects the second HAMP domain. More specifically, the α-helix located between residues 135-150 may function as a mechanical joint helix between the two HAMP domains to transfer rotational motion.

Table 3.4: Inter-spin distances between *p*pR and *p*HtrII linker[a]

*p*HtrII position	Inter-spin distances (Å)			
	pH5.0 (dark)	pH5.0 (dark)[f]	pH7.0 (dark)	pH7.0 (light)
98 (DDM[b])	36.4 ± 3.6[d]	35.1 ± 4.5	37.2 ± 3.4	36.8 ± 4.0
98 (PC[c])	20.2 ± 4	19.7 ± 5.4	21.3 ± 3.2	22.2 ± 2.6
134 (DDM)	46.0 ± 7.5	42.9 ± 7.8	47.8 ± 7.4	47.2 ± 7.2
134 (PC)	N.D.[e]	N.D.	N.D.	N.D.

[a]Inter-spin distances of the *p*pR-*p*HtrII fusion protein between *p*pR (position 154) and *p*HtrII (position 98 or 134) determined using the pulsed ELDOR method. [b]DDM: solubilized by DDM. [c]PC: reconstituted in phosphatidylcholine. [d]The deviation represents experimental error as a gaussian peak-width. [e]N.D.: not determined. [f]The *p*pR-*p*HtrII fusion protein with a different linker sequence (ASASNGRRR) between *p*pR and *p*HrII.

Finally, the light-induced structural change was monitored by the ELDOR method. The inter-spin distances of the activated state were measured at pH 7.0 since the absorption maxima of *p*pR remained unchanged following illumination with > 480 nm light at pH 5.0. As shown in Table 3.4, no significant distance change was observed

although a slight increase was recorded for the *p*pR$_{154}$-*p*HtrII$_{98}$ distance within the error bar. These data suggest that the large structural change of > 3 Å is not induced in photo-activated signal transduction. This is in good agreement with previous structural studies [1, 29] and FRET analysis [60] where a 15º clockwise rotation with 0.9 Å displacement of TM2 and 0.3-0.8 Å movement of the linker region, respectively, was observed. In order to transmit these small conformational changes, the structure of the transducer should be relatively rigid. The modeled structure shown in Figure 3.9 can transmit such small motion mechanically, and thus facilitate signal transduction.

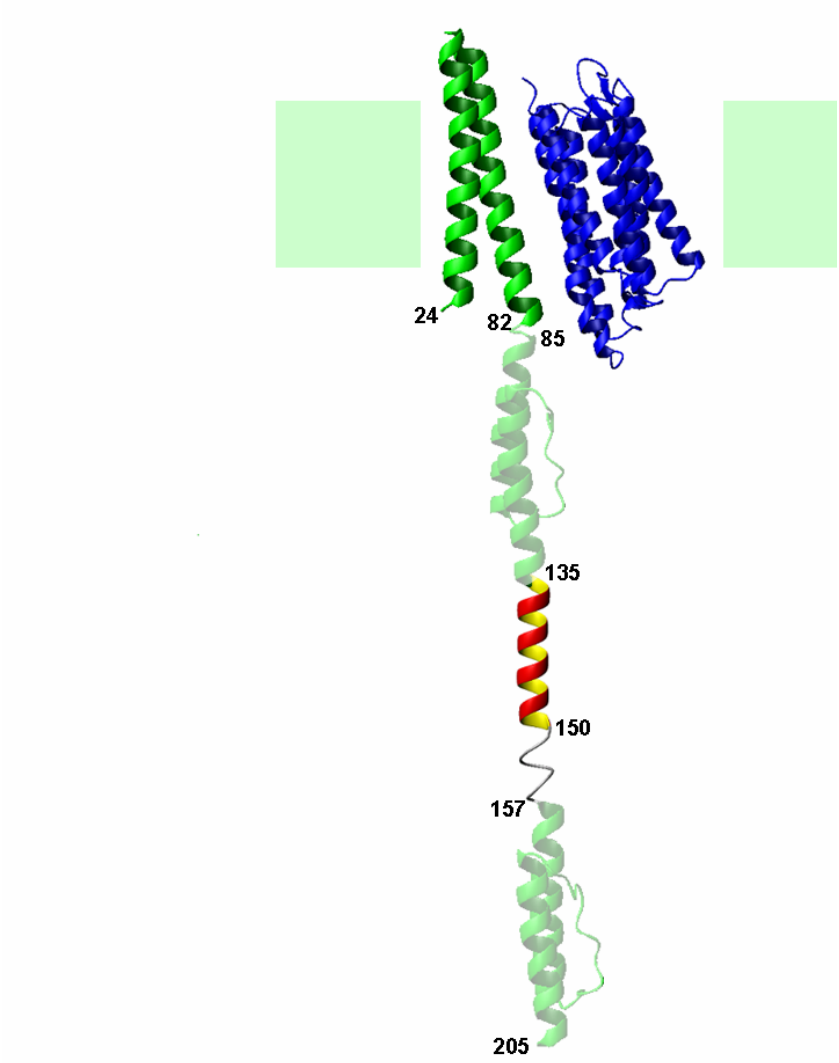


Figure 3.9. Modeled structure of the *p*HtrII linker region in the *p*pR-*p*HtrII complex. Crystal structures of *p*pR and the transmembrane region of *p*HtrII (PDB ID 1H2S) are shown in blue and green, respectively. NMR structure of the joint helix of *p*HtrII (PDB ID 2rm8) is shown in red and yellow. The modeled structures of two HAMP domains of *p*HtrII generated by homology modeling using MODELLER and based on the NMR structure of Af1503 (PDB ID 2ASX) are shown in light green.

4. Conclusion

In this thesis, the structure of the *p*HtrII linker region was investigated biochemically and biophysically. The trypsin resistant fragment *p*HtrII(100-159) within the linker region contained about 40% α-helical structure and formed a homodimer with K_d value 115 μM. The NMR structure of the *p*HtrII(100-159) fragment revealed the presence of the α-helix located between residues 135-150. This α-helix was significantly stabilized within the transmembrane protein *p*HtrII(1-159). Sequence alignment suggested the presence of two HAMP domains located between residues 83-136 and 157-211. The presence of the unstable α-helix between residues 100-134 and the dimer interface between residues 105-120 was supported to formation of Af1503 HAMP domain-like structure. The *p*pR_{154}-*p*H$trII_{98}$ distance determined by ELDOR for the PC reconstituted sample was also in good agreement with the based on the Af1503 HAMP domain-like structure. The light-activation did not change the *p*pR_{154}-*p*H$trII_{98}$ distance and thus, the photo-activation does not induce the large amplitude motion. The α-helix located between residues 135-150 as determined in this report may function as a mechanical joint between two HAMP domains to transmit the photo-activated small motion that facilitates signal transduction.

5. Acknowledgement

The present studies have been performed under the direction of Dr. Chojiro Kojima (Graduate School of Biological Sciences, Nara Institute of Science and Technology). I would like to express my gratitude to him for his guidance and discussions throughout my studies.

I wish to thank Drs. Naoki Kamo (Graduate School of Pharmaceutical Sciences, Hokkaido University) and Yuki Sudo (Division of Biological Science, Graduate School of Science, Nagoya University) for the gifts of the genes of *p*HtrII and *pp*R, and for the precious discussions.

I wish to thank Drs. Hideki Kandori and Yuji Furutani (Nagoya Institute of Technology) for their practical advice of photoactication experiment.

I wish to thank Drs. Masaki Mishima, JunGoo Jee, Toshitatsu Kobayashi and Makoto Nomura (Graduate School of Biological Sciences, Nara Institute of Science and Technology) for their practical advice, helpful discussions, and many setups for NMR measurements and structure calculation.

I wish to thank Dr. Hideyuki Hara (Bruker Biospin K.K.) for pulsed EPR measurements.

Thanks to Dr. Toshio Hakoshima, and all the members of his group and Dr. Chojiro

Kojima's group (Graduate School of Biological Sciences, Nara Institute of Science and Technology), I could carry out my experiments comfortably. I would like to appreciate their kindness.

I wish to thank Ms Junko Tsukamoto (Graduate School of Biological Sciences, Nara Institute of Science and Technology) for technical assistance with the mass spectroscopy and N-terminal protein sequencing. I wish to thank to Ms Hiroko Kinoshita and Ms Akiko Yasuba (Graduate School of Biological Sciences, Nara Institute of Science and Technology) for their technical assistance.

Finally I would like to thank my family their support and understanding.

Kokoro Hayashi

December 25, 2007.

6. References

[1] M. Hulko, F. Berndt, M. Gruber, J. U. Linder, V. Truffault, A. Schultz, J. Martin, J. E. Schultz, A. N. Lupas, and M. Coles, The HAMP domain structure implies helix rotation in transmembrane signaling, Cell 126 (2006) 929-940.

[2] A. H. West, and A. M. Stock, Histidine kinases and response regulator proteins in two-component signaling systems, Trends Biochem Sci 26 (2001) 369-376.

[3] I. M. Ota, and A. Varshavsky, A yeast protein similar to bacterial two-component regulators, Science 262 (1993) 566-569.

[4] C. Chang, S. F. Kwok, A. B. Bleecker, and E. M. Meyerowitz, Arabidopsis ethylene-response gene ETR1: similarity of product to two-component regulators, Science 262 (1993) 539-544.

[5] T. Maeda, S. M. Wurgler-Murphy, and H. Saito, A two-component system that regulates an osmosensing MAP kinase cascade in yeast, Nature 369 (1994) 242-245.

[6] G. H. Wadhams, and J. P. Armitage, Making sense of it all: bacterial chemotaxis, Nat Rev Mol Cell Biol 5 (2004) 1024-1037.

[7] J. Rudolph, B. Nordmann, K.-F. Storch, H. Gruenberg, K. Rodewald, and D. Oesterhelt, A family of halobacterial transducer proteins, FEMS Microbiol Lett 139 (1996) 161-168.

[8] K. K. Kim, H. Yokota, and S.-H. Kim, Four-helical-bundle structure of the cytoplasmic domain of a serine chemotaxis receptor, Nature 400 (1999) 787-792.

[9] J. Falke, J., and G. Hazelbauer, L., Transmembrane signaling in bacterial chemoreceptors, Trends Biochem. Sci. 26 (2001) 257-265.

[10] J. Rudolph, and D. Oesterhelt, Deletion analysis of the che operon in the archaeon *Halobacterium salinarium*, J. Mol. Biol. 258 (1996) 548-554.

[11] M. R. Alley, J. R. Maddock, and L. Shapiro, Requirement of the carboxyl terminus of a bacterial chemoreceptor for its targeted proteolysis, Science 259 (1993) 1754-1757.

[12] J. R. Maddock, and L. Shapiro, Polar location of the chemoreceptor complex in the *Escherichia coli* cell, Science 259 (1993) 1717-1723.

[13] J. Spudich, L., C.-S. Yang, K.-H. Jung, and E. Spudich, N., Retinylidene proteins: structures and functions from archaea to humans, Annu. Rev. Cell Dev. Biol. 16 (2000) 365-392.

[14] O. A. Sineshchekov, K. H. Jung, and J. L. Spudich, Two rhodopsins mediate phototaxis to low- and high-intensity light in Chlamydomonas reinhardtii, Proc Natl Acad Sci U S A 99 (2002) 8689-8694.

[15] B. Yan, T. Takahashi, R. Johnson, and J. L. Spudich, Identification of signaling states of a sensory receptor by modulation of lifetimes of stimulus-induced conformations: the case of sensory rhodopsin II, Biochemistry 30 (1991) 10686-10692.

[16] J. P. Klare, V. I. Gordeliy, J. Labahn, G. Büldt, H. J. Steinhoff, and M. Engelhard, The archaeal sensory rhodopsin II/transducer complex: a model for transmembrane signal transfer, FEBS Lett 564 (2004) 219-224.

[17] X.-N. Zhang, J. Zhu, and J. L. Spudich, The specificity of interaction of archaeal transducers with their cognate sensory rhodopsins is determined by their transmembrane helices, Proc Natl Acad Sci U S A 96 (1999) 857-862.

[18] B. Scharf, and E. K. Wolff, Phototactic behaviour of the archaebacterial Natronobacterium pharaonis, FEBS Lett 340 (1994) 114-116.

[19] B. Luttenberg, E. K. Wolff, and M. Engelhard, Heterologous coexpression of the blue light receptor psRII and its transducer pHtrII from *Natronobacterium pharaonis* in the *Halobacterium salinarium* strain Pho81/w restores negative phototaxis, FEBS Lett 426 (1998) 117-120.

[20] K.-H. Jung, E. Spudich, N., V. Trivedi, D., and J. Spudich, L., An archaeal photosignal-transducing module mediates phototaxis in *Escherichia coli*, J. Bacteriol. 183 (2001) 6365-6371.

[21] K. Shimono, M. Iwamoto, M. Sumi, and N. Kamo, Functional expression of pharaonis phoborhodopsin in *Escherichia coli*, FEBS Lett 420 **(**1997) 54-56.

[22] A. A. Wegener, I. Chizhov, M. Engelhard, and H. J. Steinhoff, Time-resolved detection of transient movement of helix F in spin-labelled pharaonis sensory rhodopsin II, J Mol Biol 301 (2000) 881-891.

[23] Y. Sudo, M. Iwamoto, K. Shimono, and N. Kamo, P*haraonis* Phoborhodopsin Binds to its Cognate Truncated Transducer Even in the Presence of a Detergent with a 1:1 Stoichiometry, Photochem. Photobiol. 74 (2001) 489-494.

[24] C. S. Yang, and J. L. Spudich, Light-induced structural changes occur in the transmembrane helices of the *Natronobacterium pharaonis* HtrII transducer, Biochemistry 40 (2001) 14207-14214.

[25] C.-S. Yang, O. Sineshchekov, E. N. Spudich, and J. L. Spudich, The cytoplasmic membrane-proximal domain of the HtrII transducer interacts with the E-F loop of photoactivated *Natronomonas pharaonis* sensory rhodopsin II, J. Biol. Chem. 279 (2004) 42970-42976.

[26] T. Umemura, Y. Matsumoto, K. Ohnishi, M. Homma, and I. Kawagishi, Sensing of cytoplasmic pH by bacterial chemoreceptors involves the linker region that connects the membrane-spanning and the signal-modulating helices, J. Biol. Chem. 277 (2002) 1593-1598.

[27] J. A. Appleman, and V. Stewart, Mutational analysis of a conserved signal-transducing element: the HAMP linker of the *Escherichia coli* nitrate sensor NarX, J. Bacteriol. 185 (2003) 89-97.

[28] V. I. Gordeliy, J. Labahn, R. Moukhametzianov, R. Efremov, J. Granzin, R. Schlesinger, G. Büldt, T. Savopol, A. J. Scheidig, J. P. Klare, and M. Engelhard, Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer

complex, Nature 419 (2002) 484-487.

[29] R. Moukhametzianov, J. P. Klare, R. Efremov, C. Baeken, A. Göppner, J. Labahn, M. Engelhard, G. Büldt, and V. I. Gordeliy, Development of the signal in sensory rhodopsin and its transfer to the cognate transducer, Nature 440 (2006) 115-119.

[30] Y. Sudo, H. Okuda, M. Yamabi, Y. Fukuzaki, M. Mishima, N. Kamo, and C. Kojima, Linker region of a halobacterial transducer protein interacts directly with its sensor retinal protein, Biochemistry 44 (2005) 6144-6152.

[31] E. Bordignon, J. P. Klare, M. Doebber, A. A. Wegener, S. Martell, M. Engelhard, and H.-J. Steinhoff, Structural analysis of a HAMP domain: the linker region of the phototransducer in complex with sensory rhodopsin II, J. Biol. Chem. 280 (2005) 38767-38775.

[32] I. L. Budyak, V. Pipich, O. S. Mironova, R. Schlesinger, G. Zaccai, and J. Klein-Seetharaman, Shape and oligomerization state of the cytoplasmic domain of the phototaxis transducer II from *Natronobacterium pharaonis*, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 15428-15433.

[33] H. Le Moual, and J. Koshland, Daniel, E., Molecular Evolution of the C-terminal Cytoplasmic Domain of a Superfamily of Bacterial Receptors Involved in Taxis, J. Mol. Biol. 261 (1996) 568-585.

[34] L. Aravind, and C. P. Ponting, The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins, FEMS Microbiol Lett 176 (1999) 111-116.

[35] S. B. Williams, and V. Stewart, Functional similarities among two-component sensors and methyl-accepting chemotaxis proteins suggest a role for linker region amphipathic helices in transmembrane signal transduction, Mol Microbiol 33 (1999) 1093-1102.

[36] L. J. McGuffin, K. Bryson, and D. T. Jones, The PSIPRED protein structure prediction server, Bioinformatics 16 (2000) 404-405.

[37] A. Lupas, D. M. Van, and J. Stock, Predicting Coled Coils from Protein Sequences, Science 252 (1991) 1162-1164.

[38] V. J. Yao, and J. L. Spudich, Primary structure of an archaebacterial transducer, a methyl-accepting protein associated with sensory rhodopsin I, Proc Natl Acad Sci U S A 89 (1992) 11915-11919.

[39] W. Zhang, A. Brooun, M. M. Mueller, and M. Alam, The primary structures of the Archaeon Halobacterium salinarium blue light receptor sensory rhodopsin II and its transducer, a methyl-accepting protein, Proc Natl Acad Sci U S A 93 (1996) 8230-8235.

[40] J. Cavanagh, W. J. Fairbrother, A. G. Palmer, and N. J. Skelton, Protein NMR Spectroscopy, Academic Press, San Diego (1996).

[41] F. Delaglio, S. Grzesiek, G. Vuister, W., G. Zhu, J. Pfeifer, and A. Bax, NMRPipe: A multidimensional spectral processing system based on UNIX pipes, J Biomol NMR 6 (1995) 277-293.

[42] T. D. Goddard, and D. G. Kneller, SPARKY3, University of California, Sanfrancisco (1999).

[43] G. Cornilescu, F. Delaglio, and A. Bax, Protein backbone angle restraints from searching a database for chemical shift and sequence homology, J Biomol NMR 13 (1999) 289-302.

[44] T. Herrmann, P. Güntert, and K. Wüthrich, Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA, J. Mol. Biol. 319 (2002) 209-227.

[45] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. I. Cheatham, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman, AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules, Computer Physics Communications 91 (1995) 1-41.

[46] R. A. Laskowski, J. A. Rullmannn, M. W. MacArthur, R. Kaptein, and J. M. Thornton, AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, J Biomol NMR 8 (1996) 477-486.

[47] R. Koradi, M. Billeter, and K. W・hrich, MOLMOL: a program for display and analysis of macromolecular structures, J Mol Graph 14 (1996) 51-55, 29-32.

[48] M. Pfeiffer, T. Rink, K. Gerwert, D. Oesterhelt, and H.-J. Steinhoff, Site-directed Spin-labeling Reveals the Orientation of the Amino Acid Side-chains in the E-F Loop of Bacteriorhodopsin, J. Mol. Biol. 287 (1999) 163-171.

[49] H. Kandori, K. Shimono, Y. Sudo, M. Iwamoto, Y. Shichida, and N. Kamo, Structural changes of pharaonis phoborhodopsin upon photoisomerization of the retinal chromophore: infrared spectral comparison with bacteriorhodopsin, Biochemistry 40 (2001) 9238-9246.

[50] M. Pannier, S. Veit, A. Godt, G. Jeschke, and H. W. Spiess, Dead-time free measurement of dipole-dipole interactions between electron spins, J. Magn. Reson. 142 (2000) 331-340.

[51] M. Pannier, S. Veit, M. Schöps, U. Wiesner, G. Jeschke, and H. W. Spiess, Determination of ion cluster sizes and cluster-to-cluster distances in ionomers by four-pulse double electron electron resonance spectroscopy, Macromolecules 33 (2000) 7812-7818.

[52] Y. Furutani, K. Kamada, Y. Sudo, K. Shimono, N. Kamo, and H. Kandori, Structural changes of the complex between pharaonis phoborhodopsin and its cognate transducer upon formation of the M photointermediate, Biochemistry 44 (2005) 2909-2915.

[53] J. P. Klare, E. Bordignon, M. Doebber, J. Fitter, J. Kriegsmann, I. Chizhov, H.-J. Steinhoff, and M. Engelhard, Effects of solubilization on the structure and function of the sensory rhodopsin II/transducer complex, J. Mol. Biol. 356 (2006) 1207-1221.

[54] R. Kishii, L. Falzon, T. Yoshida, H. Kobayashi, and M. Inouye, Structural and functional studies of the HAMP domain of EnvZ, an osmosensing transmembrane histidine kinase in *Escherichia coli*, J. Biol. Chem. (2007).

[55] N. Hulo, A. Bairoch, V. Bulliard, L. Cerutti, E. De Castro, P. S. Langendijk-Genevaux, M. Pagni, and C. J. A. Sigrist, The PROSITE database, Nucleic Acids Res. 34 (2006) D227-D230.

[56] C. J. A. Sigrist, L. Cerutti, N. Hulo, A. Gattiker, L. Falquet, M. Pagni, A. Bairoch, and P. Bucher, PROSITE: a documented database using patterns and profiles as motif descriptors, Brief Bioinform. 3 (2002) 265-274.

[57] A. Gattiker, E. Gasteiger, and A. Bairoch, ScanProsite: a reference implementation of a PROSITE scanning tool, Appl. Bioinformatics 1 (2002) 107-108.

[58] C. J. A. Sigrist, E. De Castro, P. S. Langendijk-Genevaux, V. Le Saux, A. Bairoch, and N. Hulo, ProRule: a new database containing functional and structural information on PROSITE profiles, Bioinformatics 21 (2005) 4060-4066.

[59] A. Sali, and T. Blundell, L., Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779-815.

[60] Y. Taniguchi, T. Ikehara, N. Kamo, H. Yamasaki, and T. Toyoshima, Dynamics of light-induced conformational changes of the phoborhodopsin/transducer complex formed in the n-dodecyl beta-D-maltoside micelle, Biochemistry 46 (2007) 5349-5357.

[61] P. A. Walker, L. E. Leong, P. W. Ng, S. H. Tan, S. Waller, D. Murphy, and A. G. Porter, Efficient and rapid affinity purification of proteins using recombinant fusion proteases, Biotechnology (N Y) 12 (1994) 601-605.

[62] M. G. Cordingley, P. L. Callahan, V. V. Sardana, V. M. Garsky, and R. J. Colonno, Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro, J Biol Chem 265 (1990) 9062-9065.

[63] M. Mishima, S. Wakabayashi, and C. Kojima, Solution structure of the cytoplasmic region of Na+/H+ exchanger 1 complexed with essential cofactor calcineurin B homologous protein 1, J Biol Chem 282 (2007) 2741-2751.

7. Appendix

— A Novel Protein Expression System in *E. coli* —

Abstract

Production of a recombinant protein in *Escherichia coli* is often hampered by low expression level and low solubility. Many approaches have been developed, such as protein production at low temperature, fusion protein expression using soluble protein tags, and so on. Here, we present a novel cold-shock vector for high-level soluble protein expression in *E. coli* named pCold-GST which is a modified pCold I cold-shock vector including the glutathione S-transferase (GST) tag. This pCold-GST expression system was applied to 10 proteins that were not expressed using the conventional *E. coli* expression methods, and 9 of these proteins were successfully expressed in soluble fraction. Expression and purification of two unstable protein fragments were also demonstrated using the additional C-terminal hexa-histidine tag for purification. The purified proteins were applicable to NMR experiments. These data suggest that pCold-GST expression system is useful and enable to improve the expression and purification of various proteins.

Escherichia coli expression system has many advantages for the high-level production of recombinant proteins. However, not all genes can be expressed efficiently in *E. coli*. Many approaches have been developed to overcome this challenge (S. C. Makrides *et al.*, (1996) *Microbiol Rev*, 60, 512-538). including the modifications of expression conditions, developments of promoter systems and specialized host strains.

An approach to address the low solubility of expressed protein is to fuse the target proteins with a soluble protein tag. Glutathione S-transferase (GST) (D. B. Smith *et al.*, (1988) *Gene*, 67, 31-40), maltose-binding protein (MBP) (C. di Guan *et al*., (1988) *Gene*, 67, 21-30), thioredoxin (Trx) (T. Yasukawa *et al.*, (1995) *J Biol Chem,* 270, 25328-25331), protein G B1 domain (GB1) (W. J. Bao *et al.*, *Protein Expr Purif*, 47 (2006) 599-606) (P. Zhou *et al*., (2001) *J Biomol NMR*, 20 11-14). and other proteins have been used as a soluble protein tag. GST tag is also known as a versatile tool that can be used for protein purification, pull-down assay, and other experiments.

In 2004, a pCold expression vector series, pCold I, II, III and IV, was developed (G. Qing *et al*., (2004) *Nat Biotechnol*, 22**,** 877-882). These expression vectors are cold-shock vectors, in which protein expression is under the control of *cspA* promoter. They contain the *lac* operator sequence immediately upstream of the *cspA* transcription initiation site, and constitutive expression of the *lacI* gene prevents leaky expression of

the cloned genes at 37°C. Cold-shock induction of gene expression is carried out by addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) upon temperature downshift. The pCold I vector includes a translation-enhancing element (TEE) sequence (J. P. Etchegaray *et al*., (1999) *J Biol Chem*, 274, 10079-10085) followed by a hexa His-tag and factor Xa cleavage site (Appendix Figure 1A). These pCold expression systems enable high-level production for many proteins, but still some proteins does not express.

Here, we present a novel cold-shock vector, pCold-GST, where GST-tag was inserted into the pCold I vector. To show the versatility of the pCold-GST vector, 10 genes from bacteria, plant and mammal, which could not be expressed by conventional approaches, were inserted into the pCold-GST vector, and expression level and solubility of each target were checked. Additionally, we show the purification and NMR application of two proteins expressed using the pCold-GST vector.

Materials and Methods

Modification of pCold I plasmid vector. pCold I vector was purchased from TAKARA BIO Inc. (Japan). Glutathione S-transferase (GST) gene from *Schistosoma japonicum* was synthesized by two step PCR reactions (T. Ito *et al*., (2004) *J Biomol*

NMR, 28, 357-367). In the first reaction, all 12 template oligonucleotides were mixed so that the final concentration for each was 0.5 μM, and other required materials (1 × Prime STAR buffer (TAKARA BIO Inc.), 0.2 mM dNTP and 1.0 μL of Prime STAR (TAKARA BIO Inc.) were added. The mixture was heated to 98°C for 1 min, cooled down to 50°C for annealing, and kept at 72°C for 20 min for elongation. In the second reaction, 2 μL of the first reaction product was used as a template in 50 μL mixture, which contained 1 × Prime STAR buffer, 0.3 μM forward and reverse primers, 0.2 mM dNTP, and 1.0 μL of Prime STAR. The mixture was preheated at 98°C for 1 min, cooled down to 50°C, followed by 30 cycles of heating at 98°C for 15 sec, annealing at 50°C for 15 sec, and elongation at 72°C for 1.5 min. The amplified product was purified by MinElute PCR purification kit (QIAGEN), and inserted into *Nde*I site of the pCold I vector. After ligation, the *Nde*I site which appeared just before GST sequence was erased by point mutation using QuickChange mutagenesis (Qiagen).

Construction of expression plasmids. All target gene fragments were amplified by PCR. Most of the PCR products were digested by restriction enzyme reaction, and subcloned into the plasmid vectors. In the other cases, PCR products were subcloned into the plasmid vector pET200 D-TOPO or pET100 D-TOPO by TOPO-cloning system (Invitrogen), or pGEX 6P-3 PRESAT vector by PRESAT system (N. Goda *et al.*,

(2004) *Protein science*, 13, 652-658). Appendix Table 1 shows all expression plasmids used in this study.

Expression check. The proteins with N-terminal six-residue histidine tag and GST-tag were expressed in BL21 (DE3) (Novagen), or Rosetta^{TM} (DE3) (Novagen) or BL21 Star (DE3) strain (Invitrogen) of *Escherichia coli* using 10 mL LB medium containing 100 μg/mL Ampicillin. Each cell was grown at 37°C until the OD_{660} was 0.4, 0.6 or 0.8, respectively. In the cases of pCold I or pCold-GST system, protein expression was induced by addition of 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) after temperature downshift from 37°C to 15°C. Cells were harvested for 16h after induction. In the other systems, protein expression was induced by addition of 1 mM IPTG at 37°C, and cells were harvested 3h following induction at 37°C. Harvested wet cells were resuspended in 50 mM Tris buffer (pH8.0) containing 300 mM KCl, 2 mM 2-mercaptoethanol. The suspension was lysed by sonication and centrifuged (6,000 x g for 30 min at 4°C). Supernatant was loaded onto Glutathione Sepharose 4B (GE Healthcare) and the resin was washed using buffer A. The GST-tagged protein which binds to affinity resin was analyzed by SDS-PAGE.

Protein expression for stable isotope label. Tsr or NHE1 proteins with N-terminal GST-tag and C-terminal six-residue histidine tag were expressed in Rosetta^{TM} (DE3)

strain of *Escherichia coli* (Novagen). In the case of Tsr protein, cells were grown at 37°C in 2 L M9 minimum medium containing 1 mM $MgSO_4$, 50 μM $MnCl_2$, 3.3 μM $FeCl_3$, 0.1 mM $CaCl_2$, 100 μg/mL ampicillin, 20 mg/L thymine, 20 mg/L adenosine, 20 mg/L guanosine, 20 mg/mL cytidine, 20 mg/L thiamine, 20 mf/mL biotin, 0.5g/L $^{15}NH_4Cl$ and 4g/L glucose or 2g/L $^{13}C_6$-glucose. In the case of NHE1 protein, cells were grown at 37°C in 2 L ^{13}C-^{15}N labeled C.H.L. medium (Chlorella Industry Co., Ltd, Japan). When the A_{660} was 0.4-0.6, the protein expression was induced by addition of 1 mM IPTG after temperature downshift from 37°C to 15°C. Cells were harvested for overnight following induction.

Purification of Tsr peptide and NHE peptide. Harvested cells were resuspended in 50 mM HEPES buffer (pH 8.0) containing 300 mM KCl, 0.1 mM EDTA, 1 mM DTT and 1 mg/mL Pefabloc SC (Roche applied Science). The suspension was lysed by sonication and crude membranes were collected by ultracentrifugation (138,000 x g for 30 min at 4°C). The supernatant was loaded onto Glutathione Sepharose 4B column. The resin was washed extensively with 50 mM HEPES buffer (pH8.0) containing 300 mM KCl, 1 mM DTT (Buffer A) to remove nonspecifically bound proteins. The GST-tagged protein was eluted with buffer A containing 30 mM reduced glutathione. The N-terminal GST-tag was digested by GST-tagged human rhinovirus (HRV) 3C

protease (P. A. Walker *et al*.[61, 62], (1994) *Biotechnology* (N Y), 12, 601-605) (M. G. Cordingley *et al*., *J Biol Chem*, 265, 9062-9065) which is supplied as PreScission protease (GE Healthcare). This enzyme recognizes the sequence Leu-Glu-Vla-Leu-Phe-Gln/Gly-Pro and cleaves between the Gln and Gly residues, at 4°C. After digestion, target protein was purified by gel filtration using HiLoad 16/60 Superdex 75 column (GE Healthcare) in 50 mM potassium phosphate buffer (pH6.8) containing 50 mM KCl (buffer B).

NMR experiments. Protein samples were dissolved in buffer containing 95% H_2O/5% ^2H_2O mixture. NMR spectra were acquired at 283 K or 288 K using an AVANCE500 NMR spectrometer (Bruker Biospin). For the resonance assignments for ^{1}HN, ^{15}N, ^{13}Cα, ^{13}Cβ and ^{13}C' nuclei, Tsr-His was dissolved in 10 mM citric acid buffer (pH5.0) containing 50 mM KCl, and NHE1-His was dissolved in 50 mM potassium phosphate buffer (pH6.9) containing 50 mM KCl. The resonance assignments were obtained through the following triple resonance spectra applied to the ^{15}N/^{13}C-labeled sample: 3D HNCACB, 3D HN(CO)CACB, 3D HN(CA)CO and 3D HNCO spectra (J. Cavanagh *et al.*, (1996) *Protein NMR Spectroscopy, Academic Press, San Diego*).

Results and discussion

pCold-GST vector. pCold-GST cold-shock vector was constructed by modifying the pCold I vector (G. Qing *et al*., (2004) *Nat Biotechnol*, 22, 877-882), that is, GST gene and DNA sequence of HRV 3C protease recognition site was inserted into pCold I vector (Appendix Figure 1A). DNA sequences of GST gene and following HRV3C protease recognition sequence were synthesized by two-step PCR method from chemically synthesized long primers (Appendix Figure 1C), and inserted into pCold I vector at *Nde*I site. The *Nde*I site which appeared just before GST sequence was erased by mutation from CATATG to CACATG, thus we can use another *Nde*I site which positioned at multiple cloning sites for construction of expression plasmid. Same as previous pCold systems, protein expression is under the control of the *csp*A promoter and *lac*I gene. Thus cold-shock induction of protein expression is carried out by addition of IPTG upon temperature downshift.

Expression of various proteins in pCold-GST vector. To check the utility of this expression system, we chose 10 proteins including two bacterial proteins, three plant proteins and five mammal proteins, which were not expressed in soluble fraction using conventional expression systems. 3 proteins, Tsr, *p*HtrII and NHE1, were expressed with solubility enhancement tag such as GST, MBP or Trx, however, the degradation

was observed for these proteins during the expression. The rest 7 proteins were not expressed in soluble fraction by using conventional expression systems. The pCold-GST system was applied to all of these 10 proteins, and the protein expression level and solubility were checked. In this experiment, each expression plasmid was transformed into *E. coli* Rosetta ^{TM} (DE3), and cells were harvested in 10 mL scale for overnight following induction. To check the protein expression and its solubility, cells were lysed and purified by Glutathione Sepharose 4B resin. Each fraction of affinity chromatography was analyzed by SDS-PAGE. As shown in appendix figure 2, 9 proteins were expressed in soluble fraction by pCold-GST. These results show that pCold-GST vector is useful to express the proteins that have some difficulties in overexpression in *E. coli*.

The expression level in pCold-GST system was checked under the several conditions, including A_{660} value for induction, host strain and induction time length. Appendix figure 3A and 3B shows the expression results of HY2 and NHE1, respectively. In these experiments, protein expression was induced when A_{660} was 0.4, 0.6 or 0.8 in Rosetta^{TM} (DE3), BL21 (DE3) or BL21 Star (DE3) strain of *E. coli*. The protein expression level slightly depended on A_{660} value for induction and host strains by judging from the band densities of the synthesized

fusion proteins at each condition. In other word, pCold-GST system supplies the stable high protein expression under a wide variety of expression conditions.

Expression and purification of unstable peptides. In the expression check, the synthesis of GST-fused Tsr and NHE1 were maintained high level for overnight after induction (Appendix Figure 2A lower panel). But in these cases, some degraded products were also observed during expression and lysis step (Appendix Figure 2A lower panel and Appendix Figure 3B). Tsr and NHE1 were expressed with additional C-terminal hexa-His tag to achieve purification of these proteins more easily. Isotopically labeled proteins were prepared from 2L medium, and the purification of these proteins was carried out using GST affinity chromatography.

Appendix figure 4A shows the expression and purification of GST-Tsr protein with C-terminal hexa-His tag. Although two degraded products were observed, the target GST-fused Tsr was obtained in the elution fraction. Appendix figure 4B shows the time course of HRV 3C cleavage of GST-fused Tsr analyzed by SDS-PAGE. GST-tag was specifically cleaved within 1.5 hours. Finally, the Tsr was obtained by following size exclusion chromatography (Appendix Figure 4C). The purified protein was analyzed by N-terminal sequencing and MALDI-TOF MS spectroscopy, and no degradation was observed. Purified Tsr was stable and we could measure the NMR spectra for 6 days at

10°C. All of the main chain resonances were assigned by conventional triple resonance experiments (Appendix Figure 4D).

Expression and purification of NHE1 was carried out as described for Tsr. GST-fused NHE1 was obtained by affinity chromatography, and GST-tag was cleaved by HRV 3C protease within 4 hours (Appendix Figure 5 A, B). After NHE1 was purified by Ni affinity chromatography, the sample was applied to size exclusion chromatography (Appendix Figure 5C). Final sample was analyzed by N-terminal sequencing and MALDI-TOF MS. All of the main chain resonances were assigned by conventional triple resonance experiments as reported [63]. These results show that pCold-GST system is highly effective to express and purify the unstable proteins.

In conclusion, we reported the construction of novel protein expression system named pCold-GST. We applied this expression system to 10 proteins that did not express in soluble fraction by conventional expression systems, and succeeded in overexpression of 9 out of 10 proteins in soluble fraction. These expressions were stable under a wide variety of expression condition, although the induction time length should be optimized for unstable proteins to avoid the degradation. Two unstable proteins with chain lengths about 50 amino acids were expressed and purified with additional C-terminal hexa His-tag. These proteins were stable enough to measure NMR spectra

for 6 days. Thus, the pCold-GST expression system is highly effective in preparation of most proteins.

Appendix Table 1: Expression plasmids used in this study

Bacterial proteins

protein	insertion sequence (bp)[a]	vector	restriction sites	tag
Tsr	165	pCold-GST	*Kpn* I, *Hin*d III	N-terminal hexa-His, GST and C-terminal hexa-His
		pGEX 6P-3	*Eco*R I, *Sal* I	N-terminal GST
		pMAL c2E	*Eco*R I, *Sal* I	N-terminal MBP
		pET32c	*Eco*R I, *Sal* I	N-terminal Trx and hexa-His
*p*HtrII	54	pCold-GST	*Nde* I, *Eco*R I	N-terminal hexa-His, GST and C-terminal hexa-His
		pGEX 6P-3	*Bam*H I, *Eco*R I	N-terminal GST

Plant proteins

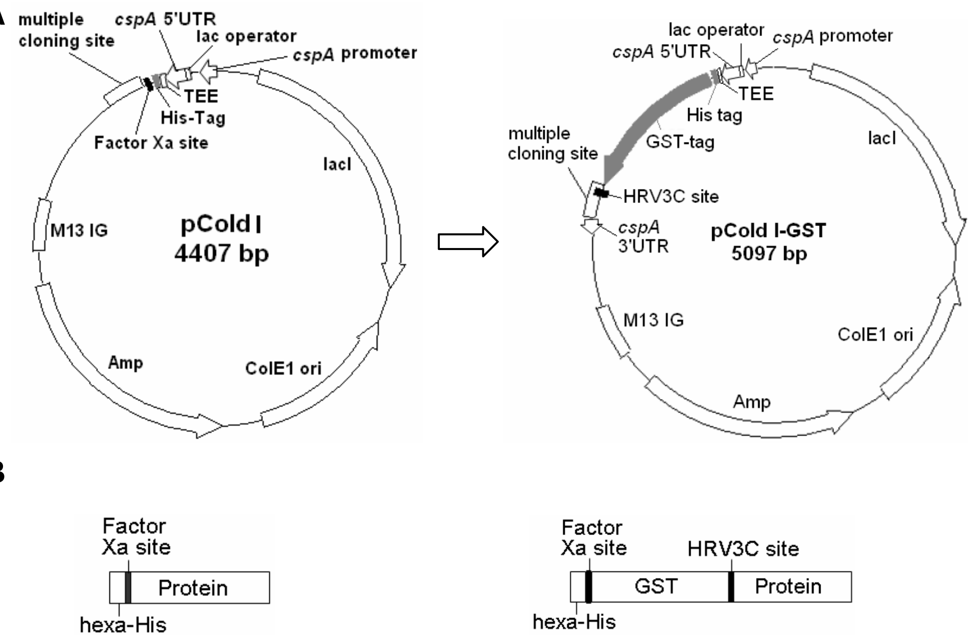
Protein	insertion sequence (bp)	vector	restriction sites	tag
HY2	858	pCold-GST	*Xho* I, *Pst* I	N-terminal hexa-His and GST
		pGEX 6P-3	*Sal* I, *Xho* I	N-terminal GST
FD	429	pCold-GST	*Nde* I, *Eco*R I	N-terminal hexa-His and GST
		pCold I	*Nde* I, *Eco*R I	N-terminal hexa-His
FT	525	pCold-GST	*Nde* I, *Xho* I	N-terminal hexa-His, GST
		pCold I	*Nde* I, *Xho* I	N-terminal hexa-His

Mammal proteins

Protein	insertion sequence (bp)	vector	restriction sites	tag
NHE1	126	pCold-GST	*Nde* I, *Pst* I	N-terminal hexa-His, GST and C-terminal hexa-His
		pCold I	*Nde* I, *Pst* I	N-terminal hexa-His
		pET200 D-TOPO	———	N-terminal hexa-His
		pET100 D-TOPO	———	C-terminal hexa-His
		pGEX 6P-3 PRESAT	———	N-terminal GST
OSBP	1173	pCold-GST	*Xho* I, *Hin*d III	N-terminal hexa-His and GST
		pCold I	*Xho* I, *Hin*d III	N-terminal hexa-His
BAK	636	pCold-GST	*Nde* I, *Eco*R I	N-terminal Hexa-His and GST
		pCold I	*Nde* I, *Eco*R I	N-terminal hexa-His
CaM	450	pCold-GST	*Nde* I, *Xho* I	N-terminal hexa-His and GST
		pET21c	*Nde* I, *Xho* I	C-terminal hexa-His
ESP1	234	pCold-GST	*Nde* I, *Eco*R I	N-terminal hexa-His and GST
		pCold I	*Nde* I, *Eco*R I	N-terminal hexa-His
		pGEX 6P-3 PRESAT	———	N-terminal GST

[a]Insertion sequence does not include affinity tag or linker sequence.

A



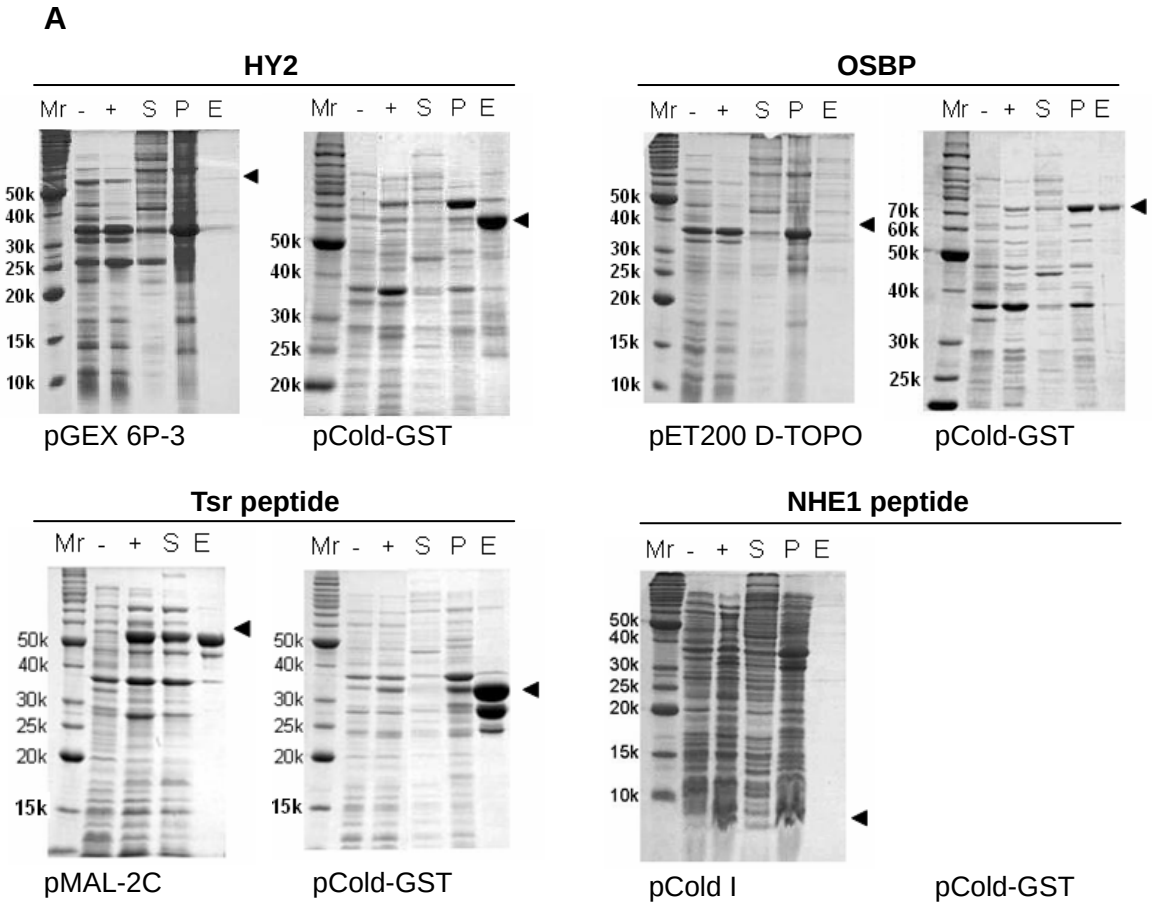
C

								Nde I		*Sac* I		*Kpn* I		*Xho* I		*BamH* I		*Eco*R I	
CTG	GAA	GTT	CTG	TTC	CAG	GGC	CCG	CAT	ATG	GAGCTC		GGT	ACC	CTC	GAG	GGA	TCC	GAA	TTC
Leu	Glu	Val	Phe	Gln	Gly	Pro	Gly	His	Met	Glu	Leu	Gly	Thr	Leu	Glu	Gly	Ser	Glu	Phe

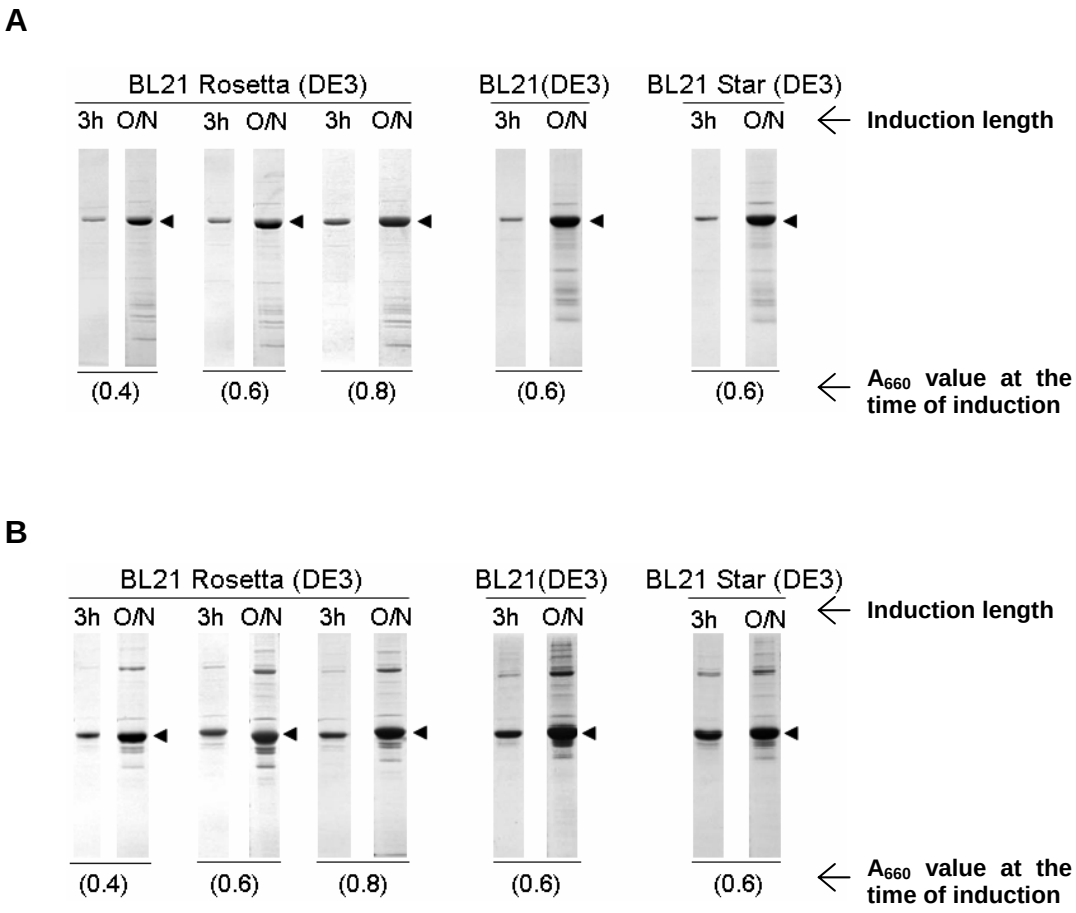
HRV 3C site

Hind III		*Sal* I		*Pst* I		*Xba* I	
AAG	CTT	GTC	GAC	CTG	CAG	TCT	AGA
Lys	Leu	Val	Asp	Leu	Gln	Ser	Arg

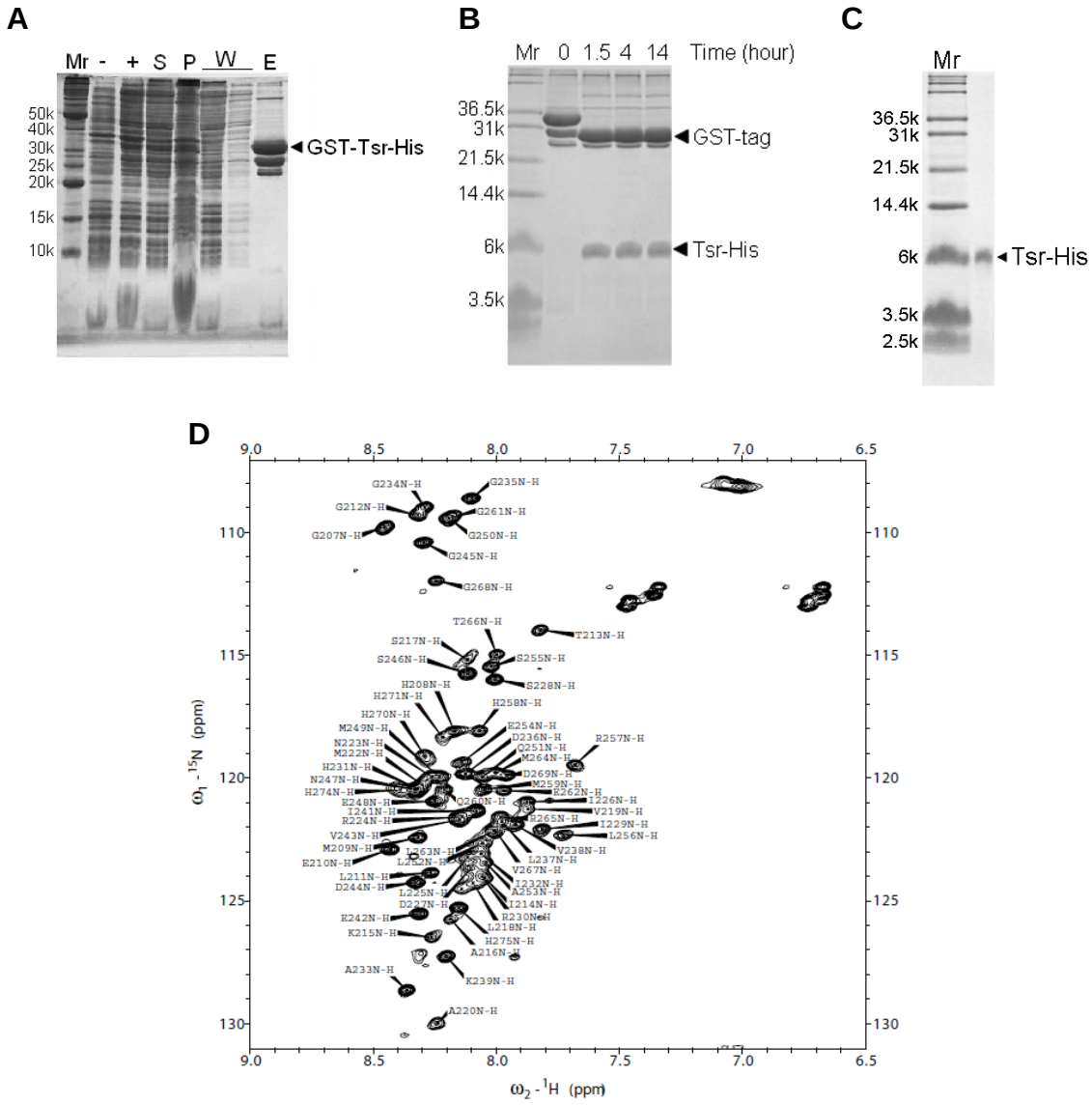
Appendix Figure 1. Structure of pCold-GST vector. (A) Schematic maps of pCold I (left) and pCold-GST (right) vector. The GST gene sequence was inserted into pCold I vector. (B) Schematic representation of expression product in pCold I vector (left) and pCold-GST vector (right). (C) DNA sequence for HRV3C protease site and multiple cloning sites of pCold-GST vector.



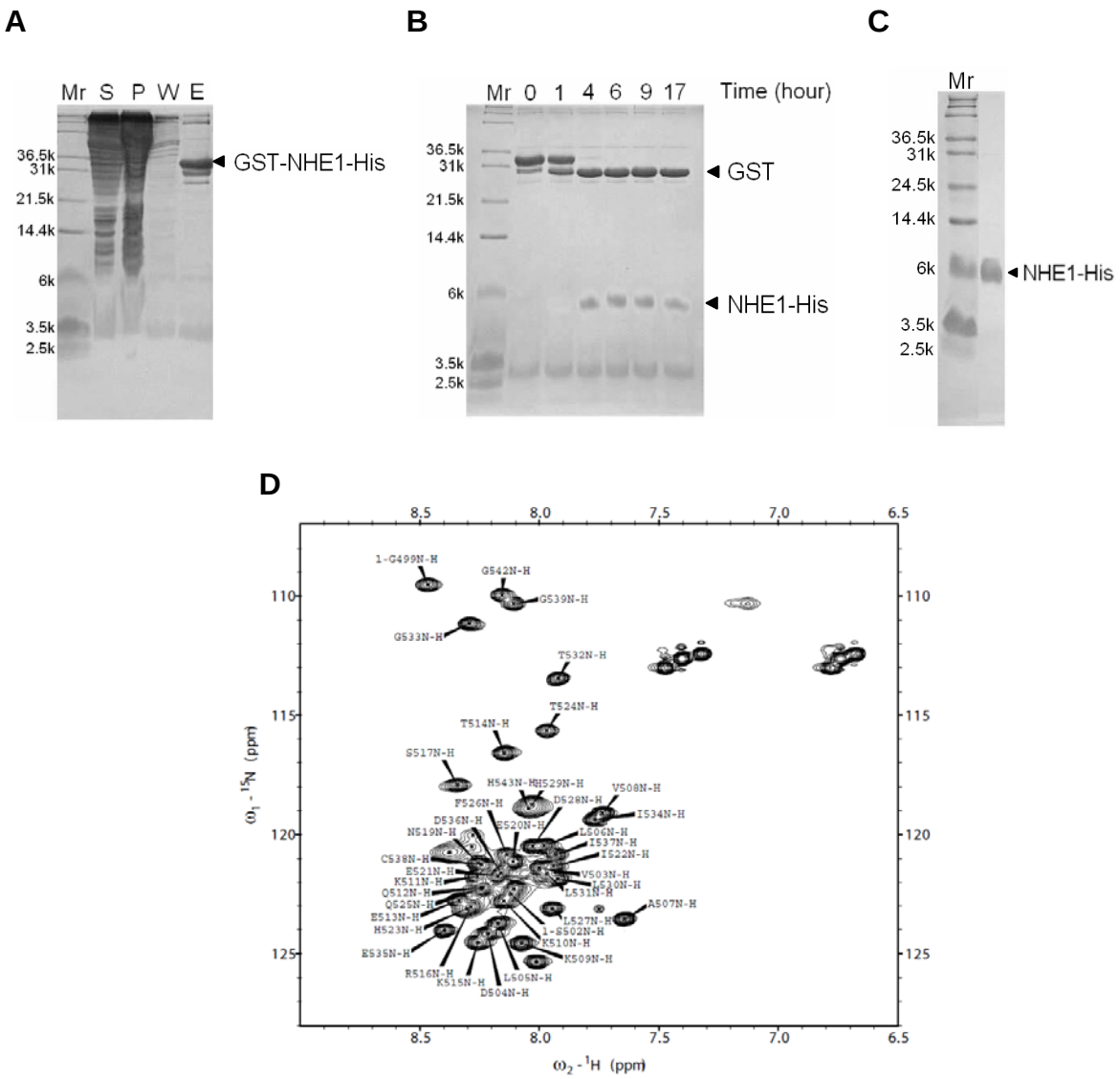
Appendix Figure 2. Protein expression and purification of GST-fused proteins. The arrows indicate the molecular weight of target proteins. (A) Protein expression and purification of GST-HY2, GST-OSBP, GST-Tsr-His, GST-NHE1-His by affinity resin was shown. Mr, molecular weight marker; -, lysate of whole cells before induciton; +, lysate of whole cells after induction, S, the supernatant fraction from the lysate; P, pellet; E, the elute fraction from the affinity resin. Lower line shows the vector systems used in each experiment. (B) The elute fractions from affinity resin in the GST-*p*HtrII, calmodulin (CaM), ESP1, FD and FT protein, which are expressed by using pCold-GST vector.



Appendix Figure 3. Comparison of the expression level with three host cells, BL21 Rosetta DE3, BL21 (DE3) and BL21 (DE3) Star. GST-HY2 (A) and GST-NHE1-His (B) are expressed for 3h or overnight (O/N). Each column shows the eluted fraction from affinity resin. The arrows indicate the molecular weight of target proteins, and lower line shows the A_{660} value for induction.



Appendix Figure 4. Expression and purification of the Tsr-His peptide. (A) Expression and purification of GST-Tsr-His protein. Mr, molecular weight marker; -, lysate of whole cells before induciton; +, lysate of whole cells after induction, S, the supernatant fraction from the lysate; P, pellet; W, flow-through fractions of affinity resin; E, the eluted fraction from the affinity resin. (B) Time-course of HRV3C digestion of the GST-Tsr-His protein. (C) The purified Tsr-His peptide. (D) ^{1}H-^{15}N HSQC spectrum of Tsr-His peptide is shown with backbone amide proton assignments. The spectrum was recorded in 10 mM citric acid buffer (pH5.0) containing 50 mM KCl at 283 K. Backbone signals were assigned by HNCACB, HN(CO)CACB, HNCACO and HNCO spectra.



Appendix Figure 5. Expression and purification of NHE1-His peptide. (A) Expression and purification of GST-NHE1-His protein by using glutathione sepharose 4B resin. Mr, molecular weight marker; S, the supernatant fraction from the lysate; P, pellet; W, flow-through fractions of affinity resin; E, the elute fraction from the affinity resin. (B) Time-course of HRV3C digestion of the GST-NHE1-His protein. (C) The purified NHE1-His peptide. (D) ^{1}H-^{15}N HSQC spectrum of NHE1-His peptide is shown. The spectrum was recorded in 50 mM potassium phosphate buffer (pH6.9) containing 50 mM KCl at 288 K.