Structure and Function of Rice Phytochrome B PAS1 Domain

Studied by Solution NMR Spectroscopy

(溶液 NMR 分光法を用いたイネフィトクロム B PAS1 ドメインの構造および機能研究)

小林　俊達

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　助教授）

平成 17 年 12 月 26 日提出

Contents

Abbreviations and acronyms

C-terminal   carboxyl-terminal

D_2O deuterated water (^2H_2O)

DG   distance geometry

FAD   flavin adenine dinucleotide

GAF   cyclic-GMP specific phosphodiesterase, adenylate cyclase and formate

hydrogen lyase transcription activator

GST   glutathione-S-transferase

HIR   high irradiance responses

HK   histidine kinase

HKLD   histidine-kinase like domain

HR-MAS   high-resolution magic-angle-spinning

HSQC   heteronuclear single quantum colleration

IPTG   isopropyl-1-thio-β-D-galactoside

LFR   low fluence rate responses

L_n artificial linker length of n amino acids

MALDI   matrix-assisted laser desorption ionization

NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
N-terminal	amide-terminal
PAS	PER, ARNT and SIM
PCR	polymerase chain reaction
Pfr	far- red-light-absorbing phytochrome
PHY	phytochrome
Pr	red-light-absorbing phytochrome
PRD	PAS-repeat domain
PΦB	phytochromobilin
rMD	restrained molecular dynamics
SA	simulated annealing
SDS-PAGE	sodium dodecylsulfate - polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
SET	solubility enhancing tag
SG	structural genomics

T_1 spin-lattice longitudinal relaxation time

T_2 spin-spin transverse relaxation time

TOCSY   total correlation spectroscopy

TOF   time-of-flight

Tris   tris-(hydroxymethyl)aminomethane

TRX   thioredoxin

UV   ultraviolet

VLFR   very low fluence responses

WT   wild type

1. General introduction

1.1. Sensor of plants for light condition

Acquiring information about the surrounding environment is crucial to the survival and well-being of living organisms. Because plants are sessile, they are particularly sensitive to environmental changes. Light is arguably the single most important environmental parameter, as it provides the ultimate source of biological energy. Therefore, plants have evolved a remarkable capacity to track and respond to fluctuations in multiple parameters of the light environment in plants. They can monitor the presence, absence, spectral quality (wavelength), fluence rate (intensity), directionality and diurnal duration of the incident light signals, and can modulate their growth and development appropriately towards optimal radiant energy capture, survival and reproduction (Kendrick and Kronenberg, 1994). This process, termed "photomorphogenesis", is facilitated through several informational photoreceptors: UV-B photoreceptors, as yet unidentified at the molecular level, the two blue/UV-A photoreceptor classes, cryptochromes (Cashmore et al., 1999) and phototropines (Briggs et al., 2001), and the red/far-red reversible photoreceptors, phytochromes (Neff et al., 2000; Smith, 2000; Fankhauser, 2001).

1.2. The phytochromes

Phytochromes, which are by the far the most studied of all the plant photoreceptors, were initially purified on the basis of being responsible for the reversible control of night-break of short day flowering plants by red and far-red light. Borthwick et al. (1952) showed that red light stimulates germination of lettuce seeds, and that this induction can be inhibited by a subsequent exposure to far-red light (Borthwick et al., 1952). In fact, the lettuce seeds can be sequentially exposed to red and far-red light with the germination response being determined by the final light-treatment (Kendrick and Kronenberg, 1994). This physiological response allowed the purification of the photoreceptor responsible, later termed phytochrome. The phytochrome (PHY) molecule is a soluble and dimeric chromoprotein, which consists of two ~125 kDa polypeptide covalently attached to a linear tetrapyrrole chromopore (referred as "phytochromobilin (PΦB)" (Cherry et al., 1993; Smith, 2000).

The photosensory activity of PHYs results from its capacity to undergo light-induced, reversible switching between two conformers: the red-light-absorbing (Pr) form and the far-red-light-absorbing (Pfr) form (Smith, 2000). Pfr is generally considered to be the biologically active form of PHY, and can be converted back to the inactive Pr form by exposure to red light. The changes between Pr and Pfr is associated with both a conformational change in structure as well as corresponding changes in the absorption maxima from 666 nm (Pr) to 730 nm (Pfr) (Fankhauser and Chory, 1997).

The absorbance spectra of Pr and Pfr show some overlap (Figure 1-1); consequently if a tissue has been exposed to light, PHY is in the Pr form and far-red light is able to transform a small proportion of these molecules into Pfr. Pure Pfr can not be established even by red-light, because some Pfr molecules absorb red-light and are transformed to Pr (Casal et al., 2003).

There are three phytochrome genes (*PHYA - PHYC*) in the model plant species *Oriza sativa*, and five phytochrome genes (*PHYA – PHYE*) in *Arabidopsis thaliana* (Mathews and Sharrock, 1997). Phylogenetic analysis indicates that a duplication during the evolution of early seed plants gave origin to the *PHYA*/*PHYC* and *PHYB*/*PHYD*/*PHYE* lineages, which in turn diverged into *PHYA* and *PHYC*, and *PHYE* and *PHYB*/*D*, which finally resulted in *PHYB* and *PHYD* (Mathews and Sharrock, 1997). Phytochromes can be classified into two groups based on their stability. PHYA is type I (light-labile) PHY which degrades rapidly on exposure to red or white light, whereas PHYB-PHYE are type II (light-stable) PHY which is stable under red or white light. Type I PHY form a homodimer, whereas type II PHY seems to form a homo- and/or hetero-dimer (Sharrock and Clack, 2004).

All phytochrome species of Pfr formation (signal perception) translocates from the cytoplasm to the nucleus, whereas the accumulation in the nucleus shows different kinetics; accumulation of PHYA in the nucleus occurred within minutes, and

reached its maximum level after 10 min, which was followed by a rapid decline, whereas accumulation of PHYB in the nucleus reached its maximum level after 6h (Kircher et al., 2002). Translocated PHYs into the nucleus initiate altered expression of selected genes that are responsible for directing the morphogenesis, whereas reconversion to Pr can abolish this process. This control of plant growth and development occurs throughout the life cycle from seedling de-etiolation to seed germination (Figure 1-2) (Casal et al., 2003).

1.3. Physiological function of phytochrome

The modes of phytochrome function have been classified into four groups complying with spectral quality and fluence rate: VLFR, LFR, red HIR and far-red HIR. Very low fluence responses (VLFR) could be accounted for by an extraordinary sensitivity to Pfr and a rapid completion of the signaling events downstream PHYA, and are induced by any wavelength between 300 and 780 nm. Low fluence rate responses (LFR) are the classical red/far-red reversible induction responses (i.e., responses that can be induced by a red-light pulse and reverted by a subsequent far-red-light pulse), and are mediated by PHYB. Two further high irradiance responses have also been described, one to continuous red (red HIR), the other to continuous far-red-light (far-red HIR). HIRs are characterized by a dependence on the intensity of light (fluence rate) used in the experiment and by the observation that continuous irradiation can only be

replaced by very frequent pulses of light.

1.4. Domain structure of phytochrome

PHY consists of two major structural domains, referred to as the N- and C-terminal domains (Quail, 2002) (Figure 1-3). The N-terminal domain consists of four subdomains consisting of a short N-terminal extension, a PAS domain, a GAF domain and a phytochrome (PHY) domain (Montgomery and Lagarias, 2002). The GAF domain possesses a covalently linked linear tetrapyrrole chromophore called phytochromobilin (Fankhauser, 2001). The PHY domain is required for the spectral integrity of phytochrome and stabilizes the Pfr form of PHYB (Oka et al., 2004). The C-terminal domain consists of two subdomains referred to as the PAS repeat domain (PRD), consisting of PAS1 and PAS2 domains, and the histidine kinase-like domain (HKLD) (Montgomery and Lagarias, 2002). The PAS domain was first found in PER, ARNT and SIM proteins, and is involved in protein dimerization with other proteins (Taylor and Zhulin, 1999). Histidine kinase is a component of many two-component regulatory systems, and contains the dimerization site (Stock et al., 2000). The N- and C-terminal domains are connected by a proteolytically-sensitive hinge region (Grimm et al., 1988) that includes a phosphorylatable serine in the Pfr (for far-red-light-absorbing) conformation (Lapko et al., 1999; Kim et al., 2004). The C-terminal domain contains the core region of loss-of-function missense mutations, and is important for

dimerization and translocation into the nucleus (Edgerton and Jones, 1992; Cherry et al., 1993; Edgerton and Jones, 1993; Quail et al., 1995; Wagner et al., 1996; Matsushita et al., 2003).

1.5. Aim to this study

PAS domains have been identified in proteins from all three kingdoms of life, and are important signaling modules that combined with a variety of regulatory modules in multidomain proteins. Various cell responses to changes in the environmental and intracellular conditions are controlled *via* PAS-containing receptors, transducers, and regulators. For phytochromes, the PAS domains contain determinants necessary for nuclear translocation, signal transduction and dimerization, however, the underlying molecular mechanism of this activity remains unclear. The purpose of this study is to understand the mechanisms of dimerization and the relation to the signal transduction in terms of the three-dimensional structures and site-directed mutagenesis.

In this thesis, I describe the solution structure of rice PHYB PAS1 domain and secondary structure of PAS1-PAS2 domain (Chapter 2: Structure determination of rice PHYB PAS domains). In chapter 3, I describe homodimerization of the PAS1 domain with the hinge region, and consideration of homodimerization mechanism (Chapter 3: Homodimerization of PHYB PAS1 domain with hinge-region). Some PAS domains are known as ligand-binding domain. To investigate whether the rice PHYB

PAS1 domain has an ability to bind small molecule, I have developed high-throughput

screening method (Chapter 4: Gel-phase NMR: New methods for NMR-based screening

approach).
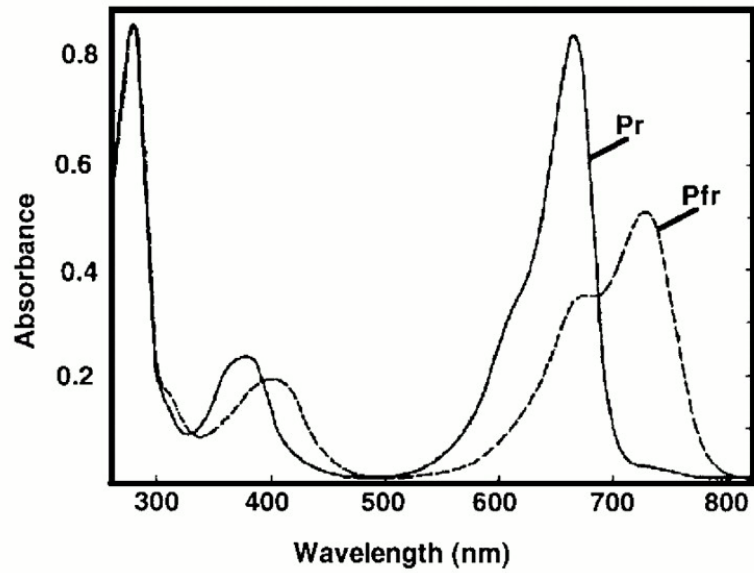
Figure 1-1. Spectral properties of the phytochrome chromoprotein. Absorption spectra of oat phytochrome A as Pr and Pfr (adapted from (Fankhauser and Chory, 1997).
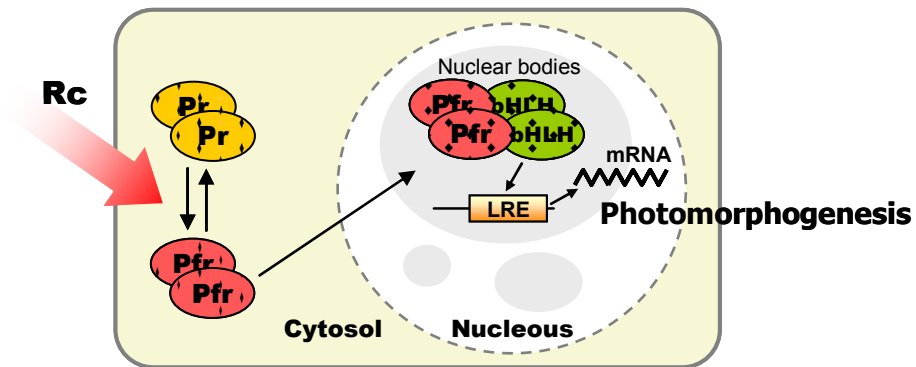
Figure 1-2. Simplified scheam of phytochrome signaling pathways. Upon absorption of red light, PHY translocates from the cytoplasm to nucleus, and regulates gene expression through interaction with transcription factors such as basic-helix-loop-helix proteins.
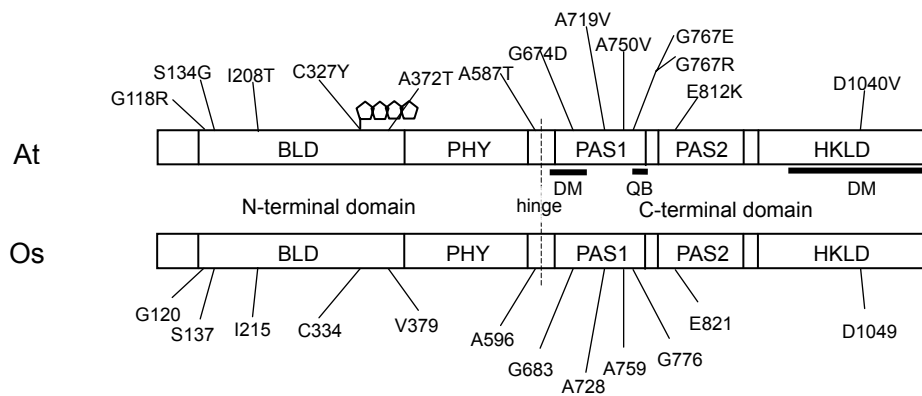
Figure 1-3. Domain structure of PHYB. The subdomains are depicted: BLD (billin lyase domain), PHY (phytochrome domain), PAS1, PAS2 and HKLD (histidine kinase-like domain). The location of loss-of-function missense mutation residues in *Arabidopsis thaliana* (At) PHYB (top) with the corresponding location in *Oriza sativa* (Os) PHYB (bottom) are shown. Dimerization sites (DM) and core region of loss-of-function missense mutations (Quail-box, QB) are represented as a thick bar.

2. Structure determination of rice PHYB PAS domains

2.1 Introduction

2.1.1 PAS domains

PAS domains have been identified in proteins from all three kingdoms of life: *Bacteria*, *Archaea*, and *Eucarya*. Theses include histidine and serine/threonine kinases, chemoreceptors and photoreceptors for taxis and tropism, circadian clock proteins, voltage-activated ion channels. Various cell responses to changes in the environmental and intracellular conditions are controlled *via* PAS-containing receptors, transducers, and regulators (Taylor and Zhulin, 1999).

PAS is an acronym formed from the names of the proteins in which imperfect repeat sequences were first recognized: the *Drosophila* period clock protein (PER), vertebrate aryl hydrocarbon receptor nuclear translocator (ARNT), and Drosophila single-minded protein (SIM) (Nambu et al., 1991). PAS domain comprises a region of approximately 100 to 120 amino residues. It is typical to find PAS domains in pairs in eukaryotic transcriptional activators, such as SIM. Microbial proteins contain single, dual, or multiple (up to six) PAS domains (Taylor and Zhulin, 1999).

Protein-protein interactions mediate signal transduction by some PAS proteins, and the PAS domains may determine the specificity of the interactions (Huang et al.,

1993). PAS domains in PER, FixL and DOS proteins form homodimers *in vitro* (Miyatake et al., 2000; Kurokawa et al., 2004; Yildiz et al., 2005), but PAS domains are usually involved in heterodimer formation. Where present in PAS proteins, such as the aryl hydrocarbon receptor (AHR) and ARNT, the bHLH motif serves as an interface for heterodimerization. However, PAS domains add increased stability and specificity to the dimmers (Pongratz et al., 1998). The ARNT protein forms heterodimers with AHR and mammalian hypoxia-inducible factor (HIF1α) in addition to forming homodimers. Heterodimers are also formed by CLOCK and CYCLE proteins and by PER and TIM proteins in circadian circuits (Gekakis et al., 1998; Schibler, 1998). The specificity of PAS transcriprional enhancers in binding to DNA response elements is determined by the composition of the dimer (Rowlands and Gustafsson, 1997).

PAS domains are also known as a versatile sensor domain, as well as a protein-protein interaction domain (Taylor and Zhulin, 1999). Of the PAS proteins that sense light, PYP is a receptor in which blue light is captured by the 4-hydroxycinnamyl chromophore in the PAS domain (Baca et al., 1994). FixL is an oxygen receptor, in which oxygen binds directly to a heme that is coordinated to a histidine residue within a PAS domain (Miyatake et al., 2000). Other PAS protein, Aer is a transducer that sense oxygen receptor, in which oxygen binds indirectly by sensing redox changes as the electron transport system responds to changes in oxygen concentration. The FAD

cofactor is bound to the PAS domain, where oxidation and reduction of FAD generate the on and off signals for aerotaxis (Taylor et al., 2001). Thus some PAS domains have an ability to bind small molecule to sense environmental conditions.

By mapping a typical PAS domain from the ARNT protein onto the crystallographic structure of the entire PYP, Getzoff and collaborators developed an argument for PYP as a prototypical PAS domain (Pellequer et al., 1998). This generalization is supported by the subsequent determination of the structure of PAS domains in the FixL protein from *B. japonicum* (Gong et al., 1998) and the human HERG protein (Morais Cabral et al., 1998). Four segments have been delineated in the overall PAS fold in PYP: (1) the N-terminal cap or lariat (residues 1-28), including the α1 and α2 helices; (2) the PAS core with the first three β-strands of the central β-sheet (residues 26-69) and the α3 and α4 helices; (3) the helical connector (residues 70-86) with the α5 helix, which diagonally crosses the β-sheet and connects two edge β-strands; (4) the β-scaffold, composed of β4, a connecting loop, and the β5- β6 hairpin that form the second three-stranded half of the central β-sheet (Pellequer et al., 1998).

Three PAS domains are contained in phytochrome protein; the N-terminal domain contains one PAS domain, and the C-terminal domain contains two PAS domains (PAS1 and PAS2 domains). Certain missense mutations within the two PAS domains of the C-terminal domain reduce the biological activity of each phytochrome

without affecting the spectral nature associated with dimerization competence. For example, certain missense mutations within the PRD of PHYB result in inhibition of nuclear translocation (G767R mutation in *Arabidopsis*, G776 in rice) (Matsushita et al., 2003), inhibition of nuclear body formation (G674D, A719V and E812K mutations in *Arabidopsis*, G683, A728 and E821 in rice) (Chen et al., 2003), reduced binding activity for PIF3 (A750V, G767R and E812K mutation in *Arabidopsis*, A759, G776 and E821 in rice) (Ni et al., 1998, 1999) and rapid dark reversion (E812K mutation in *Arabidopsis*, E821 in rice) (Elich and Chory, 1997). A750V, G767R and E812K mutations induce hypocotyl elongation under continuous red-light conditions, whereas G674D and A719V mutations do not (Quail et al., 1995; Wagner and Quail, 1995; Chen et al., 2003; Matsushita et al., 2003). Thus, these mutants suggest the two PAS domains in the C-terminal domain play a crucial role in PHY signaling.

This chapter shows the rice PHYB PAS1 domain determined by solution NMR spectroscopy, and relationship between the structure of PAS1 domain and loss-of-function missense mutations within the PAS1 domain are discussed.

2.1.2 Determination of protein structure using multidimensional NMR experiments

Introduction of stable isotopes into proteins

Nuclear magnetic resonance (NMR) spectroscopy is unique among the methods available for three-dimensional structure determination of proteins and nucleic acids at atomic resolution, since the NMR data can be recorded in solution. In the NMR experiments, solution conditions such as the temperature, pH and salt concentration can be adjusted so as to closely mimic a given physiological fluid. In addition to protein structure determination, NMR applications include investigations of dynamic features of the molecular structures, thermodynamic and kinetic aspects of interaction between proteins and other solution molecules.

Assignment of the spin systems in protein NMR spectra is an essential step in solution structure determination. However, conventional proton NMR experiments are ineffective for structure determinations of proteins with molecular masses greater than approximately 10 kDa. As the size of the protein increases, signal overcrowding because of increasing number of protons and broadening linewidths of peaks becomes more and more severe, up to the point of making it impossible the spectral analysis even by using homonuclear two-dimensional spectroscopy. The introduction of isotropically (^{15}N- and/or ^{13}C-) labeled proteins has enabled heteronuclear NMR spectroscopy which effectively circumvents these problems. Such labeling has become available by the device of high-yield expression systems for proteins in bacteria such as *Eschelichia coli*, which can grow in minimal nutrient media composed of simple organic carbon and

nitrogen sources. Spectral resolution is improved by an increase of the dimensionality of the NMR spectrum. Moreover, the efficiency of coherence transfer is increased by relatively large scalar coupling interactions rather than the small ^{1}H homonuclear three-bond scalar coupling interactions. Recently, even larger proteins consisting of more than 300 amino acids have become accessible by introduction of fractional deuteration. That is because the deuteration reduces the ^{1}H relaxation rates which are otherwise dominated by dipolar interactions with other protons, and the deuteration also reduces the heteronuclear relaxation rates which are otherwise dominated by strong dipolar interactions with directly bonded ^{1}H nuclei (Cavanagh et al., 1996; Krishna and Berliner, 1998; Sattler et al., 1999).

Resonance assignments

Assignment strategies for [15]N-, ^{13}C-labelled proteins employ coherence transfer via ^1J (and in part ^2J) coupling only, which are largely independent of conformation. Also, since the ^1J (^2J) couplings are generally larger than the linewidth ($J > \Delta\nu_{1/2}$) of nuclei under consideration, the transfer via these couplings remains efficient for relatively large molecules with short transverse relaxation times T_2^*. Figure 2-1 shows the spin system of the peptide backbone and indicates the size of the coupling constants used for magnetization transfer in doubly [13]C-, ^{15}N-labelled proteins. The double- and triple-resonance experiments are summarized in Table 2-1 (Cavanagh et al.,

1996; Krishna and Berliner, 1998; Sattler et al., 1999).

The nomenclature established for these experiments is systematic. For example, HNCO experiment used the following coherence transfers:

$^{1}H_N$ {i+1} $\rightarrow$ ^{15}N {i+1} (t$_1$) $\rightarrow$ $^{13}C_O$ {i} (t$_2$) $\rightarrow$ ^{15}N {i+1} $\rightarrow$ $^{1}H_N$ {i+1} (t3)
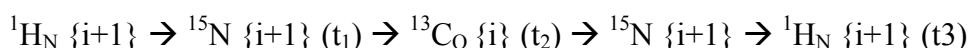
, where {i} represents the amino acid number, i. Magnetizaitons are detected in the t$_1$, t$_2$ and t$_3$ time domains, and transformed to the ω_1, ω_2 and ω_3 frequency domains by Fourier transformations, respectively. Thus, the HNCO experiment provides a cross peak at the cross point of the chemical shifts of the $^{1}H_N/^{15}N$ spins of residue i+1, and the $^{13}C_O$ spin of residue i. On the other hand, HN(CA)CO experiment correlates the chemical shifts of the $^{1}H_N/^{15}N$ spins of residue i, and the $^{13}C_O$ spin of the same residue i. Therefore, if the ^{13}CO{i} shift is unique, the $^{1}H_N/^{15}N${i+1} and $^{1}H_N/^{15}N${i} shifts are linked. In similar ways, almost all of the chemical shifts of ^{1}H, ^{13}C and ^{15}N nuclei in proteins can be assigned (Cavanagh et al., 1996; Krishna and Berliner, 1998; Sattler et al., 1999).

NMR parameters for structure determinations

The principal source of geometric information used in these structure determinations reside in short (< 5 Å) approximate interproton distance restraints derived from the observation of nuclear Overhauser effects (NOE). This distance limit

arises from the fact that the NOE (at short mixing times) is proportional to r^{-6}, where r is the distance between the two protons. The second source of constraints obtained by NMR is the information on ϕ, ψ and χ_1 torsion angles derived from scalar coupling constants. This information can be supplemented and often supplanted by the measurement of heteronuclear couplings by quantitative J correlation spectroscopy. A summary of the heteronuclear quantitative J correlation experiments is showed in Table 2-2 (Cavanagh et al., 1996; Krishna and Berliner, 1998; Sattler et al., 1999).

Structure calculations

NMR based structure calculations essentially involve finding conformations of a macromolecule that are consistent with defined structural constraints such as interproton distance, dihedral angle, and hydrogen bonding constraints. There are varieties of methods which utilize restrained molecular dynamics (rMD) and/or distance geometry (DG) mehods. The DG uses the metric matrix algorithm or the variable target function approach. The rMD uses the simulated annealing (SA) procedure in which pseudoenergy potentials drive a structure toward a conformation that will reduce violations of constraints during a forced heatup-cooldown annealing cycle. The result of an NMR structure calculation is a group of conformers that represents the solution structure of a protein. These structures are often evaluated with root mean square deviation (R.M.S.D.) values which indicate differences between the structures, and

distributions of dihedral angles (e.g., Ramachandran plot) (Cavanagh et al., 1996; Reid,

1997).

2.2. Materials and Methods

2.2.1. Protein expression and purification

The rice PHYB$^{666-782}$ and PHYB$^{666-923}$ were cloned into the pET32c expression vector (Novagen) and over-produced in *E. coli* BL21 (DE3) RIL (Novagen) as thioredoxin and hexa-histidine fusion proteins. ^{13}C/^{15}N- or ^{15}N-labeled protein was induced by the presence of 1 mM IPTG at 25 °C in M9 minimal medium containing ^{15}NH$_4$Cl and [U-^{13}C] glucose or unlabeled glucose, respectively. Cells obtained from M9 cultures were lysed by sonication. The lysate was centrifuged and the supernatant was loaded onto a Ni-NTA agarose resin (Qiagen). Proteins were eluted with imidazole in a stepwise manner. The sample fraction was then passed through a Superdex 26/60 75pg gel filtration column (Amersham). Following removal of the fused tag by enterokinase (Novagen) was performed, Hi-trap Q anion-exchange (Amersham) and Superdex 26/60 75pg columns were used to further purify the samples. The identity and integrity of the protein samples was confirmed by N-terminal sequencing, MALDI/TOF MS and SDS-PAGE.

2.2.2. Site-directed mutagenesis of PHYB$^{666-782}$

Point mutations of G683R, A728V, A759V and G776R of PHYB$^{666-782}$ were introduced into the PHYB$^{666-782}$ expression vector using the QuikChange system (Stratagene). All PCR steps were performed on an Applied Biosystems GeneAmp$^{®}$

thermal cycler and mutated sequences were verified by DNA sequencing. Protein expression and purification was performed as described above.

2.2.3. NMR measurement

Purified PHYB$^{666-782}$ and PHYB$^{666-923}$ were dissolved in 50 mM potassium phosphate buffer (pH 6.8) containing 20 mM KCl and 5 mM DTT in either 93% H_2O, 7% D_2O or 99.8% D_2O. The final concentration of PHYB$^{666-782}$ and PHYB$^{666-923}$ was adjusted to 0.8 mM and 0.2 mM, respectively. Each protein was placed in 5-mm diameter micro-NMR cells (Shigemi, Inc.) for NMR measurements.

All NMR experiments were carried out on a Bruker AV500 spectrometer with a cryogenic probe or a Bruker DRX800 spectrometer at 30ºC. Chemical shifts were referenced to 4, 4-dimethyl-4-silapentane 1-sulfonate. All multidimensional NMR spectra were acquired in a phase-sensitive mode employing a States-TPPI manner. The water flip-back method was employed in several experiments, starting from amide proton magnetization. Mirror image or forward-backward linear prediction was used in indirect time domains. All spectra were processed using the NMRPipe package (Delaglio et al., 1995), and analyzed by Sparky (Goddard and Kneller, 1999).

Sequential assignments of ^1H, ^{13}C and ^{15}N chemical shifts were achieved mainly by through-bond heteronuclear correlations along the backbones and the side-chains with the following 3D or 4D NMR experiments: HNCACB, HN(CO)CACB,

HNCACO, HNCO, C(CO)NH, H(CCO)NH, HC(CO)NH, HCCH-TOCSY and ^{15}N-edited TOCSY. Inter-proton NOEs were derived from 2D, 3D ^{15}N-edited and 3D ^{13}C-edited NOESY spectra recorded with a mixing time of 100 ms. Backbone ϕ and ψ torsion angle restraints were derived from an analysis of H_α, C_α, C_β, C' and backbone ^{15}N chemical shifts using TALOS (Cornilescu et al., 1999). Restraints were used only for those residues that exhibited TALOS reliability scores = 10. The χ_1 angle restraints of aliphatic residues were determined based on the $^3J_{N,H\beta}$ and $^3J_{C',H\beta}$ coupling constants measured in 3D HNHB and HN(CO)HB spectra (Archer et al., 1991; Grzesiek et al., 1992). Aromatic χ_1 angle restraints were determined based on the $^3J_{C',C\gamma}$ and $^3J_{N,C\gamma}$ coupling constants measured in the ^{13}C'-$\{^{13}C_\gamma\}$ spin-echo difference and the ^{15}N-$\{^{13}C_\gamma\}$ spin-echo difference ^{1}H-^{15}N HSQC spectra (Hu et al., 1997). Slowly exchanging amide protons were identified from a series of 2D ^{1}H-^{15}N HSQC spectra recorded following buffer exchange from H_2O to D_2O. Heteronuclear ^{15}N-^{1}H NOE values were determined from a pair of two-dimensional gradient sensitivity-enhanced correlation spectra acquired with and without a 3-s proton saturation period (Farrow et al., 1995).

2.2.4. Structure calculation

The solution structure was calculated based on the analysis of NOEs observed in a ^{15}N-edited NOESY, ^{13}C-edited NOESY and 2D-NOESY spectra. The 1205 NOE cross-peaks were assigned manually using Sparky. A total of 66 ϕ and ψ dihedral angle

restraints were derived with TALOS (Cornilescu et al., 1999), and a total of 8 χ_1 angle

restraints derived from the $^3J_{N,H\beta}$ and $^3J_{CO,H\beta}$ coupling constants on the aliphatic residues

and the $^3J_{CO,C\gamma}$ and $^3J_{N,C\gamma}$ coupling constants on the aromatic residues. A total of 60

hydrogen bond distance restraints (two restraints per bond) were derived from

hydrogen-deuterium exchange experiments and local secondary structure. Additionally,

ambiguous NOEs were collected with CANDID, a module of CYANA, using a standard

protocol consisting of seven cycles of iterative NOE assignments and structure

determination (Herrmann et al., 2002). During the iterative NOE assignments, 2819

NOEs were corrected, yielding a final count of 4024 assigned NOE cross-peaks. Finally,

the structure of PHYB$^{666-782}$ was calculated using a combined distance

geometry/simulated annealing approach using the CNS software package (Brunger et al.,

1998). Of the 100 calculated structures, 30 structures were accepted with no angle

violations greater than 5° and no NOE distance violations greater than 0.5 Å. The

calculated structures were analyzed using PROCHECK-NMR (Laskowski et al., 1996)

and MOLMOL, and the graphics were created using MOLMOL (Koradi et al., 1996).

2.3. Results

2.3.1. Structure determination of PHYB[666-782]

Recombinant rice PHYB[666-782] was overexpressed in *E. coli*, isotopically labeled, and purified by nickel-chelating, anion exchange and gel filtration columns. The elution profile derived from gel filtration chromatography suggested that PHYB[666-782] existed as a monomer in solution. The ^{1}H-^{15}N HSQC spectrum of PHYB[666-782] gave a highly dispersed pattern of cross-peaks, suggesting that the entire protein molecule adopts a stable tertiary conformation in solution.

NMR resonance assignments were obtained by measuring the double and triple resonance NMR spectra of ^{15}N- and ^{15}N, ^{13}C-labeled protein samples. The main-chain assignments were obtained from the HNCACB, HN(CO)CACB, HN(CA)CO, and HNCO spectra. Side-chain assignments were obtained from C(CO)NH, H(CCO)NH, HCCH-TOCSY and 4D-HC(CO)NH spectra. The structure of PHYB[666-782] was determined using 4023 distance constraints obtained from ^{13}C-edited and ^{15}N-edited NOESY spectra, and 8 and 132 torsion angles derived from quantitative *J* experiments and TALOS analyses, respectively. The final 20 structures derived from the NMR data were well defined except for the N- and C-terminal residues (Figure 2-2A). The root mean square deviation (r.m.s.d.) for the averaged structure of the well-defined region was 0.19 Å for the backbone and 0.59 Å for all heavy atoms (residues 666-778). The

statistics of the structures are shown in Table 2-3 online.

2.3.2. Solution structure of PHYB$^{666-782}$

PHYB$^{666-782}$ adopted an α /β fold consisting of six α-helices and four β-strands (Figure 2-2B). The helices were Helix I (residues 667-669), Cα (690- 695), Dα (700- 703), Eα (708- 712), Fα (718- 728) and Helix II (746- 749). The β-strands were Aβ (677- 680), Hβ (753- 763), Gβ (734- 741) and Iβ (769- 774). Nine loops connected each secondary structure element. The first strand, Aβ, was connected to helix Cα by a long loop, L2, and was covered by helices Dα and Eα, which form the PAS core. The helical connector, Fα, linked the PAS core and the β-scaffold consisted of three β-strands, Gβ, Hβ and Iβ (Figure 2-2B). This global structure of PHYB$^{666-782}$ represented the typical PAS fold (Taylor and Zhulin, 1999) except for the missing Bβ strand and the extra helices. Thus residues 677-774 were assigned to the PAS1 domain for rice PHYB. The closest matches derived from a DALI search (Holm and Sander, 1996) were the PAS domains of HIF-2α (1P97), HERG (1BYW) and PYP (3PYP), with Z-scores 11.3, 9.7 and 9.4, respectively. The r.m.s.d. between the PHYB PAS1 domain and the PAS domains of HIF-2α (1P97), HERG (1BYW) and PYP (3PYP) was 1.34, 1.96 and 1.90 Å, respectively, with over 39 coordinates of C$_\alpha$ atoms for residues in secondary structure elements (PAS1 residues 677-680, 700-703, 718-728, 739-742, 753-763 and 769-774).

2.3.3. Conformational analyses of P746 and P752 residues

Interestingly, three regions, region-1 (L695-T696 and L698), region-2 (V709, I713-F714) and region-3 (K740-Q748 and K750-I756), showed minor peaks with substantial chemical shift differences from major peaks in each spectrum, indicating that conformation multiplicities exist in these regions. Notably, region-3 contains two prolines, P746 and P752, while region-1 and region-2 were spatially close to region-3 in the 3D structure, implying that the conformation multiplicity of these regions could have originated from *cis-trans* isomerization of either one or two prolines. NOE cross peaks of G745H$_\alpha$-P746H$_\delta$ and G751H$_\alpha$-P752H$_\delta$ were observed for major peaks on a 3D ^{13}C-edited-NOESY spectrum, indicating that both P746 and P752 adopt a *trans* conformation as the major form. In fact, in the 20 ensemble structures determined, the omega angle for P746 and P752 was -178.40 $\pm$ 0.1 and -179.10 $\pm$ 0.1, respectively.

Although the tertiary structure of the minor conformation was not determined, use of POP software, a conformation prediction program for the Xaa-proline peptide bond based on the chemical shifts and 3D structure database search (Schubert et al., 2002), showed that P746 and P752 adopted a *trans* and *cis* conformation with 100% and 99.9% matches in the database, respectively, as deduced from the C$_\beta$ and C$_\gamma$ chemical shifts of minor peaks. In fact, the chemical shifts of C$_\beta$ and C$_\gamma$ of P752 were differed markedly in the major and minor peaks, (32.7 and 26.1 ppm) and (34.9 and 23.5 ppm),

respectively, in contrast to those of P746, (32.1 and 27.0 ppm) and (32.0 and 27.0 ppm), respectively. Thus, an equilibrium exists between *trans* and *cis* isomerization for the major and minor conformation of P752, respectively, and this represents the origin of the conformational multiplicity within the PAS1 domain. For example, the major conformer contains Helix II within residues P746-S749 consisting of region-3, whereas the secondary structure of the minor conformer is a random coil in this region. The *trans* to *cis* conformational change of P752 could conceivably disrupt Helix II formation.

2.3.4. Structure of PHYB$^{666-923}$

The C-terminal domain of the phytochromes contained the PRD. In order to obtain structural information concerning the PRD, PHYB$^{666-923}$ containing the PAS1 domain and the following ~140 residues was expressed in *E.coli* and purified. Sequential backbone assignments of PHYB$^{666-923}$ have been reported (Kobayashi et al., 2005). Both PHYB$^{666-923}$ and PHYB$^{666-782}$ showed minor peaks, indicating that PHYB$^{666-923}$ had a similar minor conformation to PHYB$^{666-782}$. The secondary structure of PHYB$^{666-923}$ was estimated by TALOS analyses.

In Figure 2-3A, the sequential alignment between residues 666-782 and 783-923 in PHYB and other PAS domains is shown. The secondary structure of residues 666-782 in PHYB$^{666-923}$ was identical to PHYB$^{666-782}$. The secondary structure of residues 666-782 differed from other PAS domains in terms of an additional helix,

Helix II, and the absence of Bβ-strand (Figure 2-3A and B). Helix I, the N-terminal helix, has been found in photoactive yellow protein (PYP) and the sensor domain of *Rhizobium meliloti* FixL (RmFixLH). Residues 810-917 showed secondary structure that was characteristic of the PAS fold (Aβ (810-813), Bβ (819-823), Cα (824- 828), Eα (846-850), Fα (858- 869), Gβ (873-884), Hβ (891-898) and Iβ (909-917)) (Figure 2-3A). Thus, residues 810-917 were assigned to the PAS2 domain, and PHYB$^{666-923}$ adopted the PRD containing the PAS1 and PAS2 domains, although this PAS2 domain had one or two characteristic β-strand(s) instead of the Dα-helix found in typical PAS folds (Figure 2-3A).

2.3.5. No interaction between the PAS1 and PAS2 domains is observed

The ^{1}H-^{15}N HSQC spectra of PHYB$^{666-782}$ and PHYB$^{666-923}$ were compared in an effort to assess the possibility of any interaction between the PAS1 and PAS2 domains (Figure 2-3C). No significant chemical shift differences were observed between PHYB$^{666-782}$ and the corresponding region of PHYB$^{666-923}$. In heteronuclear ^{15}N-^{1}H NOE experiment, residues at the N-terminus (residues 666-670, 672), C-terminus (residues 917-923) and a loop between PAS1-PAS2 domain (residues 779, 780,782-789,791-807) showed low ^{15}N–^{1}H NOE values, indicating these regions are highly flexible (Figure 2-4). Moreover, the elution profile of the analytical SEC for PHYB$^{666-923}$ was independent of the protein concentration. These data indicated that the

PAS1 domain did not interact with the PAS2 domain, either in an intra- nor

inter-molecular manner.

2.4 Discussion

2.4.1. Conserved residues of the PRD

The C-terminal domain contains well conserved residues among all phytochrome species (Clack et al., 1994). To determine the conserved residues in the PAS1 domain, sequence alignments were generated (Figure 2-5A) between the rice PHYB PAS1 domain and those of all other PHYAs and PHYBs where sequence data is available in the SWISS-PROT and TrEMBL (Bairoch and Apweiler, 2000) databases. One cluster of conserved residues is found on the solvent-exposed face on the β-sheet side (Figure 2-5B). This cluster contains the core region of loss-of-function missense mutations (Figure 2-3B) (Quail et al., 1995), and thus, might play a critical role in functions common to phytochromes. The other cluster is found on the N-flanking region of the PAS1 domain, including Helix I (Figure 2-5A and B). This region is necessary to maintain the dimer (see Chapter 3), thus dimer formation might play some important role. On the other hand, the residues in Helix II are not conserved, suggesting Helix II associated with the major-minor conformation change is probably not related to the general functions of phytochromes.

One of the regions related to the conformational changes induced by photoconversion between the Pr and Pfr forms has been mapped to the linker region between the PAS1 and PAS2 domains by limited proteolysis. This suggests that the

linker region is exposed to the environment in the Pfr form (Grimm et al., 1988). In this study, I showed that no secondary structure element was found in the linker region (Figure 2-3A), and that the PAS1 domain did not interact with the PAS2 domain (Figure 2-3C). In addition, the residues on the linker region showed low ^{15}N–^{1}H NOE values (Figure 2-4). These results indicate that the linker represents a flexible linker. On the other hand, the sequence of the linker region is highly conserved among all PHYAs and PHYBs (Figure 2-5A). Thus the linker region might play an important role in terms of conformational flexibility and solvent accessibility in the Pfr form.

2.4.2. Loss of function missense mutants

Four mutations, which reduce a phytochrome function, are found in PHYB$^{666-782}$ (G683D, A728V, A759V and G766R) (Figure 2-6A). These mutants were overexpressed in *E. coli* and purified. The G683D mutant of PHYB$^{666-782}$ was degraded during the purification process. Unlike wild-type (WT) PHYB$^{666-782}$, the A728V, A759V and G776R mutants failed to adsorb onto the anion exchange column under the same buffer conditions, suggesting that the structure or surface charge of these mutants differs from that of PHYB$^{666-782}$. The effect of the mutations on the structure was determined by examining the ^{1}H-^{15}N HSQC spectra of the PHYB$^{666-782}$ mutants.

Figure 2-6B-E shows the ^{1}H-^{15}N HSQC spectra of the WT, and A728V, A759V and G776R mutants of PHYB$^{666-782}$. Changes in tertiary structure were

monitored by comparing backbone amide ^{15}N and ^{1}H shifts for each of the mutants with WT values. For the A728V and A759V mutants, the HSQC spectra are similar to the WT, whereas certain signals shifted markedly (Figure 2-7). Thus, it is likely that the A728V and A759V mutants possess a similar overall structure to the WT in solution. It should be noted that local conformational changes occur at predominantly the N-terminal half of the PAS1 domain containing Aβ and two loops (L1 and L2) in the A728V mutant, and predominantly on the β-sheet in the A759V mutant. On the other hand, many peaks in the HSQC spectra of the G776R mutant are broadened with decreased peak intensity (Figure 2-6E), suggesting that the G776R mutation partially disrupted the tertiary structure compared with PHYB$^{666-782}$. This disruption could conceivably inhibit the nuclear translocation of PHYB.

Four target residues (G683, A728, A759 and G776) participate in the formation of a hydrophobic core within the PAS1 domain, and are conserved among all PHYAs and PHYBs (Figure 2-5A). Two residues found in the core region of loss-of-function missense mutations (A759 and G776) are located on the β-sheet, whereas the other two residues (G683 and A728) are absent (Figure 3-5A). The A750V and G767R mutations in *Arabidopsis* (A759 and G776 in rice) induce hypocotyl elongation under continuous red-light conditions, whereas the G674D and A719V mutations in *Arabidopsis* (G683 and A728 in rice) do not (Quail et al., 1995; Wagner and Quail, 1995; Chen et al., 2003;

Matsushita et al., 2003). These differences in photomorphogenesis might be due to perturbation of the conserved molecular surface on the β-sheet side caused by the A759V and G776R mutations (Figure 2-7). Thus, structural perturbation on the surface derived from core region of loss-of-function missense mutations in the PAS1 domain perturbs phytochrome function.

Table 2-1. Summary of Correlations observed in the 3D double- and triple-resonance experiments used for sequential and side-chain assignments.

Experiment	Correlation	J Coupling
^{15}N-edited HOHAHA	C$^{\alpha}$H(i)-^{15}N(i)-NH(i)	$^{3}J_{\text{HN}\alpha}$
	C$^{\beta}$H(i)-^{15}N(i)-NH(i)	$^{3}J_{\text{HN}\alpha}$ and $^{3}J_{\alpha\beta}$
HNHA	C$^{\alpha}$H(i)-^{15}N(i)-NH(i)	$^{3}J_{\text{HN}\alpha}$
H(CA)NH	C$^{\alpha}$H(i)-^{15}N(i)-NH(i)	$^{1}J_{\text{HN}\alpha}$
	C$^{\alpha}$H(i-1)-^{15}N(i)-NH(i)	$^{2}J_{\text{HC}\alpha}$
HNCA	^{13}C$^{\alpha}$(i)-^{15}N(i)-NH(i)	$^{1}J_{\text{HC}\alpha}$
	^{13}C$^{\alpha}$(i-1)-^{15}N(i)-NH(i)	$^{2}J_{\text{HC}\alpha}$
HN(CO)CA	^{13}C$^{\alpha}$(i-1)-^{15}N(i)-NH(i)	$^{1}J_{\text{NCO}}$ and $^{1}J_{\text{C}\alpha\text{CO}}$
HNCO	^{13}CO(i-1)-^{15}N(i)-NH(i)	$^{1}J_{\text{NCO}}$
HCACO	C$^{\alpha}$H(i)-^{13}C$^{\alpha}$(i)-^{13}CO(i)	$^{1}J_{\text{C}\alpha\text{CO}}$
HCA(CO)N	C$^{\alpha}$H(i)-^{13}C$^{\alpha}$(i)-^{15}N(i+1)	$^{1}J_{\text{C}\alpha\text{CO}}$ and $^{1}J_{\text{NCO}}$
HN(CO)CACB	NH(i)-^{15}N(i)- ^{13}C$^{\alpha}$(i-1)/^{13}C$^{\beta}$(i-1)	$^{1}J_{\text{C}\alpha\text{CO}}$, $^{1}J_{\text{NCO}}$ and $^{1}J_{\text{CC}}$
HNCACB	NH(i)-^{15}N(i)- ^{13}C$^{\alpha}$(i)/^{13}C$^{\beta}$(i)	$^{1}J_{\text{NC}\alpha}$ and $^{1}J_{\text{CC}}$
	NH(i)-^{15}N(i)- ^{13}C$^{\alpha}$(i-1)/^{13}C$^{\beta}$(i-1)	$^{2}J_{\text{NC}\alpha}$ and $^{1}J_{\text{CC}}$
HBHA(CO)NH	C$^{\beta}$H(i-1)/ C$^{\alpha}$H(i-1)-^{15}N(i)-NH(i)	$^{1}J_{\text{C}\alpha\text{CO}}$, $^{1}J_{\text{NCO}}$ and $^{1}J_{\text{CC}}$
HBHA(CBCA)NH	C$^{\beta}$H(i)/ C$^{\alpha}$H(i)-^{15}N(i)-NH(i)	$^{1}J_{\text{NC}\alpha}$ and $^{1}J_{\text{CC}}$
	C$^{\beta}$H(i-1)/ C$^{\alpha}$H(i-1)-^{15}N(i)-NH(i)	$^{2}J_{\text{NC}\alpha}$ and $^{1}J_{\text{CC}}$
C(CO)NH	^{13}C^{j}(i-1)-^{15}N(i)-NH(i)	$^{1}J_{\text{C}\alpha\text{CO}}$, $^{1}J_{\text{NCO}}$ and $^{1}J_{\text{CC}}$
H(CCO)NH	H^{j}(i-1)-^{15}N(i)-NH(i)	$^{1}J_{\text{C}\alpha\text{CO}}$, $^{1}J_{\text{NCO}}$ and $^{1}J_{\text{CC}}$
HCCH-COSY	H^{j}(i-1)-^{13}C^{j}(i)-^{13}C$^{j\pm1}$(i)- H(i)$^{j\pm1}$	$^{1}J_{\text{CC}}$
HCCH-TOCSY	H^{j}(i-1)-^{13}C^{j}(i)$\cdots$^{13}C$^{j\pm n}$(i)- H(i)$^{j\pm n}$	$^{1}J_{\text{CC}}$

Table 2-2. Experiments for Determining three-bond coupling constats by quantitative _J_ correlation spectroscopy.

Experiment	Three-bond coupling	Torsion angle
3D HNHA	$^3J_{HN\alpha}$	ϕ
3D (HN)CO(CO)NH	$^3J_{COCO}$	ϕ
2D ^{13}C-$\{^{15}N\}$ spin-echo difference CT HSQC	$^3J_{C\gamma N}$	χ_1 of Thr and Val
2D ^{13}C-$\{^{13}CO\}$ spin-echo difference CT HSQC	$^3J_{C\gamma CO}$	χ_1 of Thr and Val
2D ^{13}C-$\{^{13}C\gamma(aro)\}$ spin-echo difference ^1H-^{15}N HSQC	$^3J_{C\gamma(aromatic)CO}$	χ_1 of aromatics
2D ^{15}N-$\{^{13}C\gamma(aro)\}$ spin-echo difference ^1H-^{15}N HSQC	$^3J_{C\gamma(aromatic)N}$	χ_1 of aromatics
2D ^{13}C-$\{^{13}C\gamma\}$ spin-echo difference ^1H-^{15}N HSQC	$^3J_{C\gamma(aliphatic)CO}$	χ_1 of aliphatics
3D HN(CO)C	$^3J_{C\gamma(aliphatic)N}$	χ_1 of aliphatics
3D HN(CO)HB	$^3J_{COH\beta}$	χ_1
3D HNHB	$^3J_{NH\beta}$	χ_1
3D HACAHB	$^3J_{\alpha\beta}$	χ_1
2D or 3D ^1H-detected long-range C-C COSY	$^3J_{CC}$	χ_2 of Leu and Ile χ_3 of Met
3D ^1H-detected [^{13}C-^1H] long-range COSY	$^3J_{CH}$	χ_2 of Leu and Ile χ_3 of Met

Table 2.3. Structural statistics for PHYB[666-782] [a]

Total number of distance constraints	4023
Long range	1390
Middle range	591
Short range	820
Intraresidue	1222
Hydrogen bond constraints	30×2
Dihedral constraints[b]	
ϕ, ψ	66, 66
χ_1	8
R.m.s. deviation from experimental constraints	
Distance (Å)	0.0282 ± 0.0006
Angle (°)	0.74 ± 0.02
R.m.s. deviation from idealized covalent geometry	
Distance (Å)	0.0032 ± 0.0001
Angle (°)	0.44 ± 0.01
CNS energy terms (kcal/mol)[c]	
E_{bond}	19.5 ± 0.6
E_{angle}	102.7 ± 3.3
E_{imp}	14.3 ± 0.3
E_{vdw}	-211 ± 10
PROCHECK Ramanchandran plot (666-778)	
Residues in most favored regions (%)	80.1
Residues in additional allowed regions (%)	16.8
Residues in generously allowed regions (%)	3.2
Residues in disallowed regions (%)	0.0
R.m.s. deviations to mean structure of the calculated 20 structures	
Backbone (666- 778) (Å)	0.191
All heavy (666- 778) (Å)	0.588

[a] These statistics comprise the ensemble of the 20 lowest energy structures obtained from 100 starting structures. Structure calculations were performed using CNS version 1.1.

[b] None of these structures exhibited distance violations > 0.5 Å or dihedral angle violations > 5°.

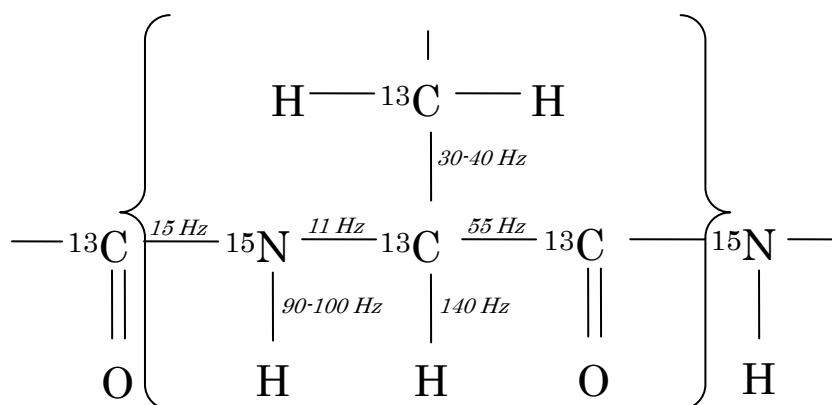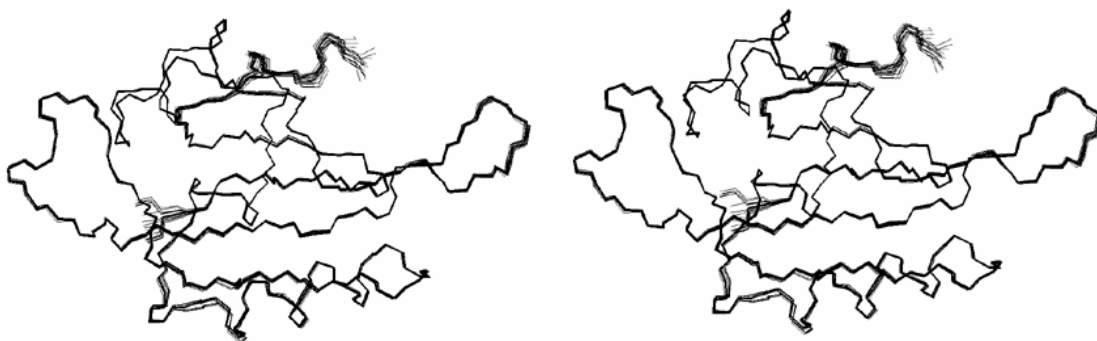[c] E_{vdw} is the Lennerd-Jones energy of CNS energy terms

Figure 2-1. Summary of the one-bond heteronuclear coupling along the polypeptide chain utilized in 3D and 4D NMR experiments.
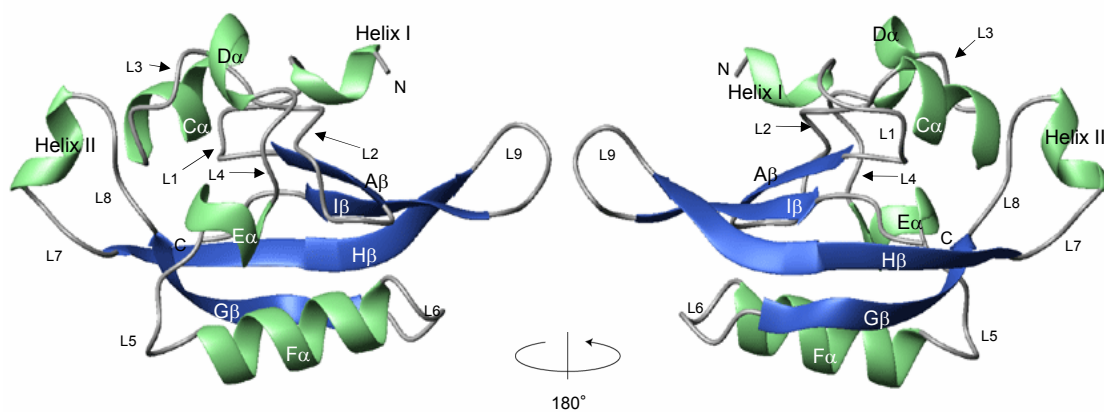
Figure 2-2. The solution structure of rice PHYB$^{666-782}$. (A) Stereo view of backbone superpositions of the 20 structures of PHYB$^{666-782}$. (B) Schematic ribbon drawing of PHYB$^{666-782}$. Helices are shown in green and strands in blue. Six helices (HelixI, Cα, Dα, Eα, Fα and Helix II), four strands (Aβ, Gβ, Hβ and Iβ) and nine loops (L1- L9) are depicted.
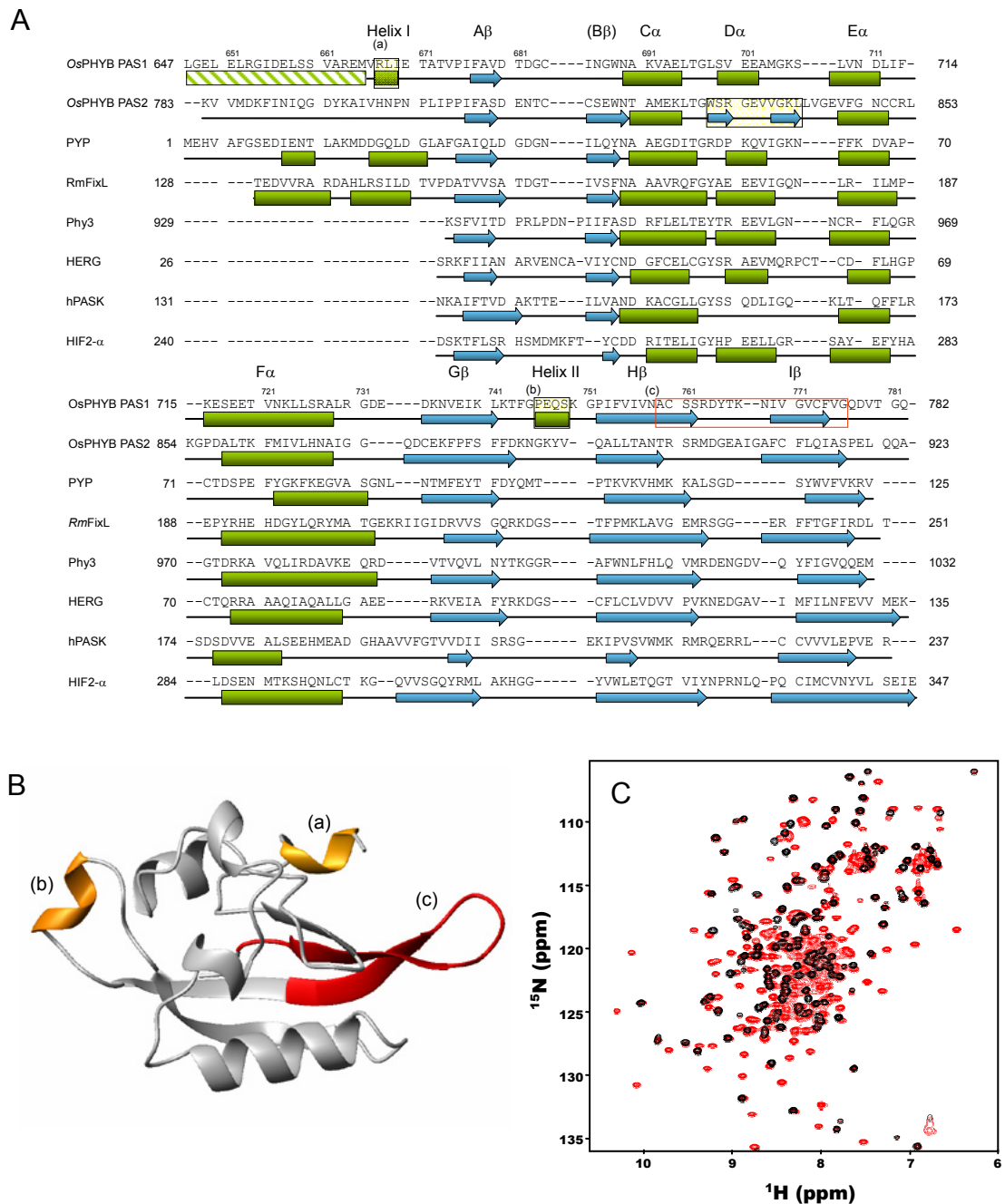
Figure 2-3. Sequence alignment and secondary structure of the PAS domains. (A) Sequence alignment of the rice PHYB PAS1 and PAS2 domains with known PAS domain structures. α-helices and β-strands are drawn in green and blue, respectively. The alignment was initially generated using ClustalW (Pearson and Lipman, 1988) and manually adjusted to match the secondary structure elements. The hatched box indicates the predicted helix estimated using PSIPRED (Jones, 1999). The unique secondary structures found in PHYB are shown in black boxes (a and b). The core region of loss-of-function missense mutations is shown in a red box (c). (B) The unique secondary structures and the core region of loss-of-function missense mutations are represented as yellow and red, respectively, on the ribbon picture

of PHYB$^{666-782}$. (C) Superposition of the ^{1}H-^{15}N HSQC spectra of PHYB$^{666-782}$ (black) and PHYB$^{666-923}$ (red). The PHYB$^{666-923}$ contains both PAS1 and PAS2 domains.
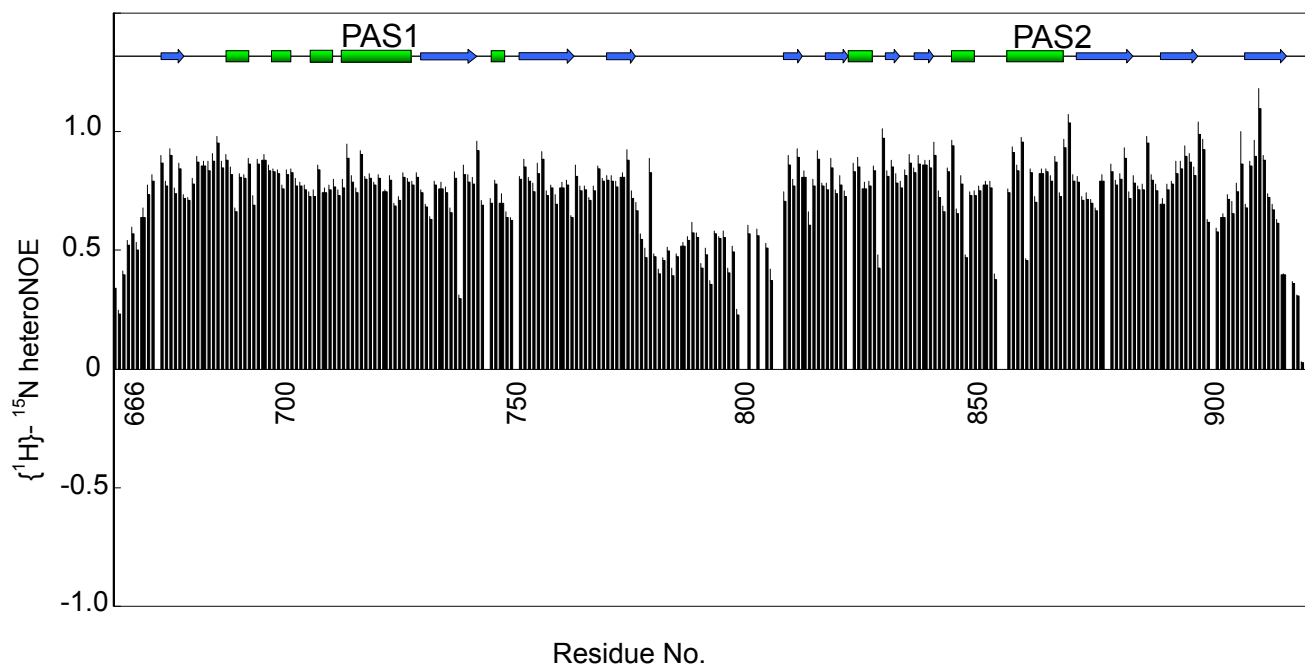
Figure 2-4.   The PAS1 domain was linked to the PAS2 domain by a flexible linker. $\{^{1}H\}$-^{15}N heteronuclear NOE values measured for PHYB[666-923]. Secondary structure elements are indicated. Smaller NOE values implied faster motion of the protein backbone.

Figure 2-5. Conserved residues of the PHYA and PHYB PAS1 domains. (A) Sequence alignment of region 647-820 in rice PHYB with all other PHYAs and PHYBs. The alignment was generated using

ClustalW (Pearson and Lipman, 1988). Completely, highly and moderately conserved residues are denoted "*", ":" and ".", respectively. The charged residues within the hinge-region are drawn in white letters on red (negative) and blue (positive) backgrounds (see Section 3.2.2.). The core region of loss-of-function missense mutations is shown in a red box. (B) Mapping of conserved residues on the surface of the rice PAS1 domain. Completely, highly and moderately conserved residues are in red, deep pink and light pink, respectively. The molecular orientation is essentially the same as in Figure 2-2B.
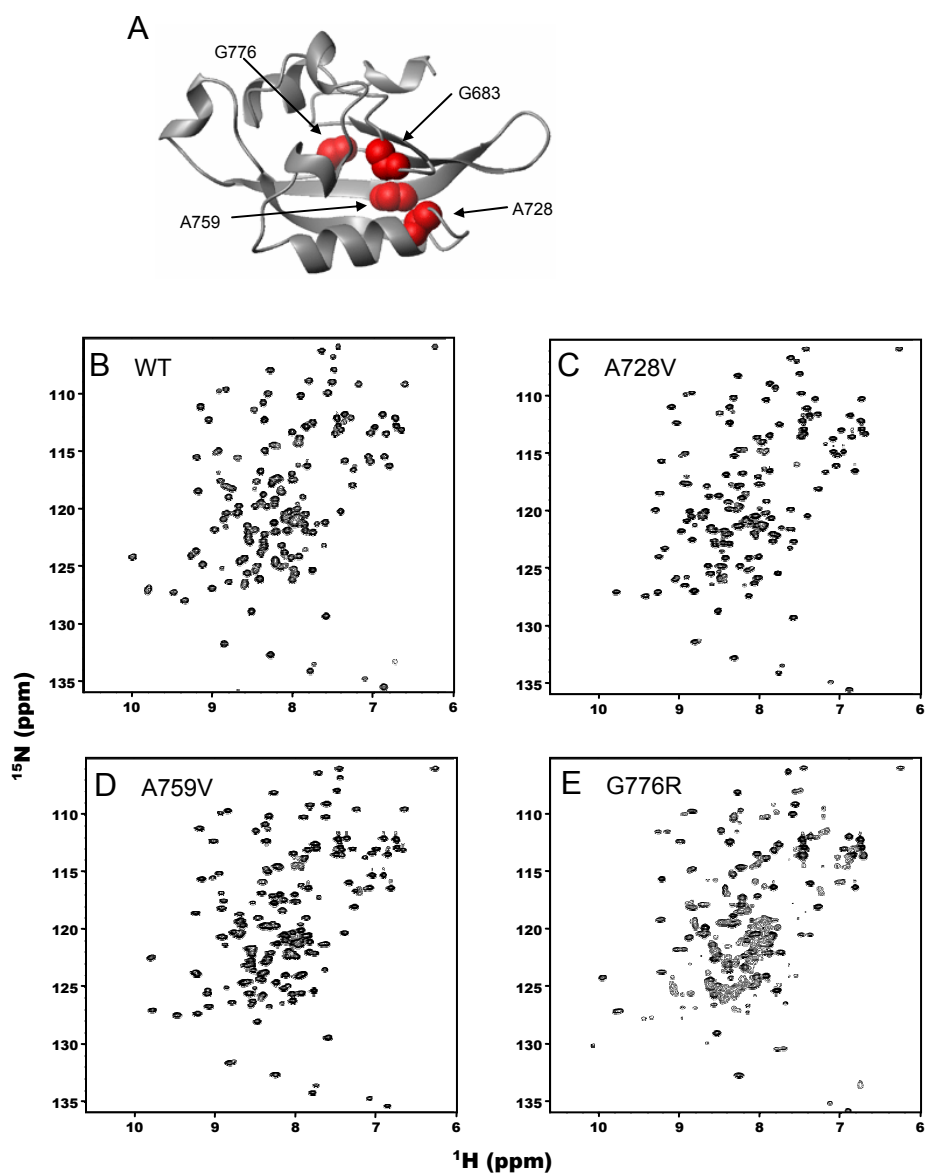
Figure 2-6. NMR analyses of missense mutations of the rice PHYB PAS1 domain. (A) Missense mutations are mapped on the structure of the rice PAS1 domain. (B- E) ^{1}H-^{15}N HSQC spectra of wild-type PAS1 (B), A728V (C), A759V (D) and G776R (E).
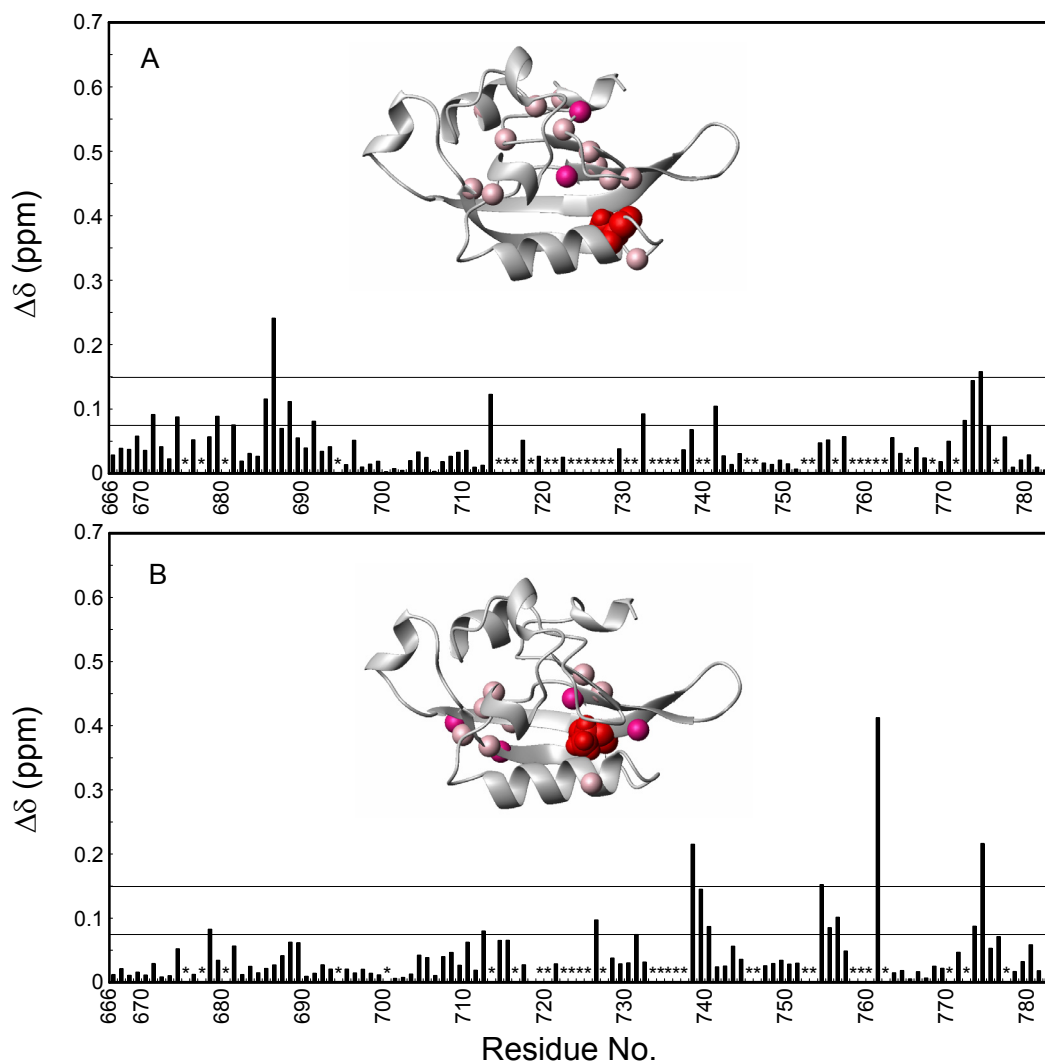
Figure 2-7. Chemical shift differences between WT and mutants of PHYB^666-782. The bars represent combined ^1H and ^{15}N chemical shift differences ($\Delta\delta = [\Delta\delta(^1H)^2 + \{\Delta\delta(^{15}N) / 5\}^2]^{1/2}$) for A728V (A) and A759V (B). Residues with significant chemical shift differences are mapped on the PHYB^666-782 structure: deep pink spheres ($\Delta\delta > 0.150$ ppm), light pink spheres ($0.075 < \Delta\delta < 0.150$ ppm). Asterisks indicate residues not assigned due to signal overlap or broadening.

3. Homodimerization of PHYB PAS1 domain with the hinge-region

3.1. Introduction

Phytochrome dimerization is necessary for full activity as it relates to hypocotyl elongation (Matsushita et al., 2003), despite the details of dimerization sites and dimerization mechanisms remain unclear. From modeling studies of phytochrome sequences, the dimerization sites have been proposed to be between residues 730 and 821 (Romanowski and Song, 1992). On the other hand, from experimental studies, Edgerton and Jones reported that the hinge-PAS1 segment [residues 599-683 in oat PHYA (642-727 in rice PHYB)] and a portion of the HKLD [residues 1069-1129 in oat PHYA (1110-1171 in rice PHYB)] are capable of mediating dimerization using analytical size-exclusion chromatography experiments and λ repressor-based *in vivo* assays (Edgerton and Jones, 1993), whereas Cherry and co-workers reported that a portion of the HKLD (residues 919-1093 in oat PHYA (960-1135 in rice PHYB)) was essential for phytochrome dimerization, using analytical size-exclusion chromatography (SEC) experiments (Cherry et al., 1993). These data of two analyses suggest that a portion of HKLD is necessary for phytochrome dimerization, and that the hinge-PAS1 segment is secondary dimerization site. However, mechanisms and physiological functions of these dimerization sites remain unclear. In this chapter, homodimerization

of the PHYB PAS1 domain with the hinge-region is described, and the homodimerization mechanism and a possible role for homodimerization of this region in phytochrome function are discussed.

3.2. Materials and Methods

3.2.1. Protein expression and purification

DNA fragments encoding rice PHYB (residues 621-782, 633-782 and 633-672 designated as PHYB$^{621-782}$, PHYB$^{633-782}$, and PHYB$^{633-672}$, respectively) were amplified by PCR and inserted into a pET32c expression vector (Novagen). A DNA fragment encoding rice PHYB (residues 647-782 designated as PHYB$^{647-782}$) was amplified by PCR and inserted into a pDEST32 expression vector, modified pET32c vector (Novagen). Each PHYB fragment was expressed in *Eschelichia coli* BL21(DE3)RIL with a thioredoxin (TRX)- and hexa-histidine (His$_6$)-tag at the N-terminus. Transformed cells were grown at 37ºC in M9 media containing ^{15}NH$_4$Cl (1 g/l) and unlabeled glucose for ^{15}N-labeled proteins, respectively. Protein expression was induced using 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 25ºC when A$_{600}$ was ~0.6, and cells were harvested for 8 hrs thereafter. Cell pellets were suspended in 50 mM HEPES (pH 7.4), 300 mM KCl, 5 mM imidazole, 0.1 mM EDTA, 0.1 mM dithiothreitol (DTT), 1 mM Pefabloc SC$^{®}$ (Roche) and 10% glycerol and then sonicated. Following centrifugation, the supernatant was loaded onto a Ni-NTA agarose column (Qiagen). Bound proteins were eluted using 50 mM HEPES (pH 7.4), 300 mM KCl, 350 mM imidazole, 0.1 mM EDTA, 0.1 mM DTT, 1 mM Pefabloc SC and 10% glycerol. TRX- and His$_6$-tagged proteins were loaded onto a Superdex 26/60 75pg (Amersham) gel

filtration column equilibrated with 50 mM Tris-HCl, 100 mM NaCl and 5 mM CaCl$_2$. TRX- and His$_6$ tags were then digested using Factor Xa (Novagen) for PHYB$^{647-782}$, or Enterokinase (Novagen) for PHYB$^{621-782}$ and PHYB$^{633-782}$. Hi-trap Q and Superdex 26/60 75pg columns (Amersham) were used for further purification of the proteins. Protein identity and integrity was confirmed by N-terminal sequencing, MALDI/TOF MS and SDS-PAGE.

3.2.2. Site-directed mutagenesis of PHYB$^{647-782}$

Point mutations of R651K, R653D, D656R, R663D and R653D/R663D of PHYB$^{647-782}$ were introduced into the PHYB$^{647-782}$ expression vector using the QuikChange system (Stratagene). All PCR steps were performed on an Applied Biosystems GeneAmp$^{®}$ thermal cycler and mutated sequences were verified by DNA sequencing. Protein expression and purification was performed as described above.

3.2.3. NMR measurement

Purified PHYB$^{666-782}$ and PHYB$^{666-923}$ were dissolved in 50 mM potassium phosphate buffer (pH 6.8) containing 20 mM KCl and 5 mM DTT in either 93% H$_2$O, 7% D$_2$O or 99.8% D$_2$O. The final concentration of PHYB$^{666-782}$ and PHYB$^{666-923}$ was adjusted to 0.8 mM and 0.2 mM, respectively. Each protein was placed in 5-mm diameter micro-NMR cells (Shigemi, Inc.) for NMR measurements.

All NMR experiments were carried out on a Bruker AV500 spectrometer with a cryogenic probe or a Bruker DRX800 spectrometer at 30ºC. Chemical shifts were referenced to 4, 4-dimethyl-4-silapentane 1-sulfonate. All multidimensional NMR spectra were acquired in a phase-sensitive mode employing a States-TPPI manner. The water flip-back method was employed in several experiments, starting from amide proton magnetization. All spectra were processed using the NMRPipe package (Delaglio et al., 1995), and analyzed by Sparky (Goddard and Kneller, 1999).

3.2.4. Analytical size exclusion chromatography (SEC)

Purified $PHYB^{621-782}$, $PHYB^{633-672}$, $PHYB^{647-782}$, $PHYB^{666-782}$ and TRX-His_6-$PHYB^{633-672}$ (0.025 mM to 1.7 mM) were loaded onto a 10/30 HR Superdex-75 analytical gel-filtration column (Amersham) equilibrated with running buffer (50 mM potassium phosphate buffer (pH 6.8) containing 20 mM KCl and 5 mM DTT). The column was eluted at a flow rate of 0.7 ml/min and the absorbance of eluted fractions was monitored at 280 nm. The loading loop volume was 100 μl, and each fraction volume was set to 300 μl. A gel-filtration standard kit (Bio-Rad) was used for calibration (Thyroglobulin: 670 kDa, Bovine gamma globulin: 158 kDa, Chicken ovalbumin: 44.0 kDa, Equine myoglobin: 17.0 kDa and Vitamin B-12: 1.35 kDa).

3.2.5. Homology modeling

The model structure of the PAS2 domain (residues 806-923), a homodimeric core region (residues 926-993) and the ATP-binding domain (residues 1002-1171) of the HKLD were generated using the program MODELLER (Sali and Blundell, 1993) based on the Protein Data Bank (PDB) entries ****, 1JOY and 1R62 as templates, respectively. Sequences were aligned using ClustalW (Pearson and Lipman, 1988).

3.3. Results

3.3.1. PHYB$^{647-782}$ forms homodimer in a concentration-dependent manner

Recombinant rice PHYB$^{647-782}$, containing a portion of the hinge region and the PAS1 domain, was overexpressed in *E. coli* and purified. The self-associative property of PHYB$^{647-782}$ was analyzed by analytical size-exclusion chromatography (SEC). Figure 3-1A shows the apparent molecular weight calculated from the elution volume at different PHYB$^{647-782}$ concentrations ranging from 5 μM to 2.1 mM by reference to a standard plot (gel-filtration molecular mass calibration kit, Bio-Rad). The apparent molecular weights were ca. 25 and 35 kDa at 5 μM and 2.1 mM, respectively. The apparent molecular weight at 2.1 mM (35 kDa) was comparable to the calculated molecular weight of the PHYB$^{647-782}$ dimer (30.2 kDa), indicating that PHYB$^{647-782}$ formed a homodimer. On the other hand, the apparent molecular weight at 5 μM (25 kDa) was larger than the calculated molecular weight of the PHYB$^{647-782}$ monomer (15.1 kDa). This might reflect the non-spherical molecular shape of the PHYB$^{647-782}$ monomer. Thus, PHYB$^{647-782}$ exists in dynamic equilibrium between a monomer and homodimer with dissociation constant of ~200 μM, where the homodimer predominates at higher concentrations.

The ^{1}H-^{15}N HSQC spectra of PHYB$^{647-782}$ were measured with changes in protein concentration from 0.025 to 1.7 mM (Figure 3-1B). Depending on the protein

concentration, certain signals were broadened, and 5 residues showed significant chemical-shift changes on the slow exchange time scale: W688, A693, S761, K767 and F774 (Figure 3-1B and 3-2). In particular, the A693 signal was well separated in the ^{1}H-^{15}N HSQC spectra in a wide concentration range, and so the A693 signal was used as an indicator of the proportion between PHYB$^{647-782}$ monomer and dimer (Figure 3-1). Both peak sets were observed in the protein concentration range of 0.1-0.5 mM. These results were consistent with the analytical SEC experiment. The ^{1}H-^{15}N HSQC spectrum of PHYB$^{647-782}$ at the monomer-predominant concentration (0.025 mM) was similar to that of PHYB$^{666-782}$ (Figure 3-3), indicating that the N-terminal region of PHYB$^{647-782}$ is related to its ability to form a homodimer and does not affect region 666-782 at low concentration. The results derived from analytical SEC and the ^{1}H-^{15}N HSQC spectra indicated that PHYB$^{647-782}$, consisting of the PAS1 domain and a portion of the hinge region, could form a stable homodimer.

3.3.2. Charged residues within the hinge-region provide a critical role for PHYB$^{647-782}$ dimerizaition

PHYB$^{621-782}$, PHYB$^{633-782}$ and TRX-His$_6$-PHYB$^{633-672}$ were expressed and purified in an effort to identify the dimerization site. The self-associative property of each PHYB fragment was analyzed by analytical SEC (Fig. 3-4). The elution profiles of PHYB$^{621-782}$ and PHYB$^{633-782}$ showed concentration-dependent self-dimerization as well

as PHYB$^{647-782}$. TRX- His$_6$-PHYB$^{633-672}$ tended to self-associate at high concentration but the dimer form could not be detected due to aggregation. The apparent molecular weight of PHYB$^{666-782}$ was ca. 17 kDa, suggesting that PHYB$^{666-782}$ (molecular weight 13.6 kDa) adopts the monomeric form. The ^{1}H-^{15}N HSQC spectrum of PHYB$^{666-782}$ was much sharper than PHYB$^{647-782}$, indicating that PHYB$^{666-782}$ is much smaller and highly mobile. These two results were consistent with each other. These data strongly suggest that region 647-665 is crucial for dimer formation.

As I have shown here rice PHYB$^{647-782}$ forms dimer, however, rice PHYA$^{601-740}$ which corresponding to rice PHYB$^{647-782}$ is monomer. Comparing to the hinge-region of PHYB$^{647-782}$ with that of PHYA$^{601-740}$, the hydrophobic residues were well conserved, whereas the charged residues were not (Figure 2-5). To assess the elements necessary for the dimerization of PHYB$^{647-782}$, we characterized dimerization property of several proteins containing from-PHYB-to-PHYA or charge-inversion type point-mutations at the hinge-region using the analytical SEC and the A693 signal on the ^{1}H-^{15}N HSQC spectra (Figure 3-5 and 3-6). From-PHYB-to-PHYA type mutants for two arginine residues (R653D and R663S) gave rise to dimerization with lower binding affinity than WT, and R663S showed much weaker than R653D. Charge-inversion type mutant, R663D leaded to lower affinity than from-PHYB-to-PHYA type mutant, R663S. Furthermore, double mutant (R653D/R663D) could not dimerize in a concentration of 1.7 mM. On the other hand, no significant changes in E651K and D656R mutants were

observed upon the dimerization. These mutations have indicated that the arginine

residues within the hinge-region of PHYB$^{647-782}$ provide a critical role for dimerization.

3.4. Discussion

3.4.1. Dimerization mechanism of the PHYB[647-782]

It has been reported that two regions of the *Avena sativa* PHYA C-terminal domain, residues 599-683 and 919-1129 (residues 642-727 and 960-1171 in rice PHYB, respectively), are capable of mediating dimerization (Cherry et al., 1993; Edgerton and Jones, 1993). One dimerization region (residues 598-683) is composed of the hinge region with the N-terminal half of the PAS1 domain. The other region (residues 919-1029) contains a large portion of the HKLD. However, neither the modeled structures nor the dissociation constants have been reported for these two dimerization regions.

The PAS domain is associated with the protein-protein interaction, and two types of homodimer structures of PAS domains have been reported. One is RmFixL, where two extended helices located upstream of the PAS domain is required for homodimerization (Miyatake et al., 2000). Each extended helix forms a specific interaction site, and the homodimer interface is on the β-sheet side of the PAS domain. A similar structure was reported for the PAS domain of EcDOS (Kurokawa et al., 2004). The other type of structure is dPER, containing two tandemly repeated PAS domains (Yildiz et al., 2005). Both the loop linking the Hα and Iα helices within the first PAS domain and an extended C-terminal helix within the second PAS domain are required for dPER homodimerization. myresults have clearly shown that the N-terminal region

of PHYB$^{647-782}$, upstream of the PAS1 domain, is required for homodimerization (Figure 3-5), and the two tandemly repeated PAS domains, PAS1 and PAS2, do not interact with each other (Figure 2-3C). These data are consistent with the RmFixLH-like dimer model, but not with the dPER-like model.

I have demonstrated that the two positive-charged residues (R653, R663) within the hinge-region of PHYB$^{647-782}$ are required for its dimerization, and the mutations of R663 have distinct effects on the dimerization (Figure 3-5 and 3-6). Although the R653 is conserved only in rice PHYB and *Sorghum bicolor* PHYB, R663 is conserved in all PHYBs (Figure 2-5). In addition, the residue of R663 is not conserved in all PHYAs (Figure 2-5). These experimental data and sequential analyses suggest that R663 is mainly involved in the dimerization, and play a critical role in a difference between PHYA and PHYB in the dimerization ability.

Secondary structure prediction of the hinge-region using PSIPRED (Jones, 1999) displayed that the hinge-region adopted a helical conformation (Figure 2-3A). When arrayed on a helical wheel diagram, two arginine residues (R653 and R663) which alter the dimerization property of PHYB$^{647-782}$ are located on a same face (Figure 3-7). These data indicate that PHYB$^{647-782}$ dimerization occur between the positively-charged residues on the hinge-region of one monomer and negative-charge cluster(s) on the PAS1 domain of the other chain, but not *via* the hydrophilic face of

their hinge regions. PHYB PAS1 domain contains three negative-charge clusters (Figure 3-8; cluster-1, -2 and -3). Each cluster is comprised of five residues; the cluster-1 is comprised of D680, D682, D711, E718 and E719, the cluster-2 is comprised of E670, E694, E701, E702 and E747, and the cluster-3 is comprised of E716, D732, E733, D734 and E738. Experimentally, depending on the PHYB[647-782] concentration, significant peak shifts in the ^{1}H-^{15}N HSQC spectra are observed for A672, W688, A693, L695, D732, V757, S761, K767 and F774 (Figure 3-1B and 3-2). Three of these residues (W688, A693 and A695) are located adjacent to cluster-2, and five residues (D732, V757, S761, K767 and F774) are located adjacent to or in cluster-3, whereas no observed significant peak shifts in the ^{1}H-^{15}N HSQC spectra occur in residues which are spatially close to cluster-1. This result supports the positive-charged residues on the hinge-region of one monomer interact with the negative-charge cluster, either cluster-2 or -3, of the other chain (Figure 3-8).

3.4.2. Dimerization and phytochrome function

Solvent-exposed β-sheet surface of the PAS domain is crucial for biological signaling pathways through intra- and intermolecular protein-protein interaction (Cusanovich and Meyer, 2003; Card et al., 2005). My data clearly showed the core region of loss-of-function missense mutations known as the Quail-box is mapped on the β-sheet (Iβ and Hβ) of the PAS1 domain. For PHYB, 2 mutants are known in the Quail-box, and these mutations induce hypocotyl elongation under continuous red-light

conditions. One of these, the G776R mutant (G767R in *Arabidopsis*), disrupts nuclear translocation (Matsushita et al., 2003), where the ternary structure of the PAS1 domain is partially disrupted, as revealed by NMR (Figure 2-6). These results strongly suggest that the β-sheet surface of PAS1 domain is crucial for phytochrome signaling pathways including phytochrome nuclear-translocation, and structural perturbation on the β-sheet side directly affects photomorphogenesis. On the other hand, in PHYB$^{647-782}$ homodimerization, residues on the solvent-exposed β-sheet surface of one molecule of PAS1 domain are not involved in the interaction with other molecule, although the dimerization interface is mapped on the β-sheet side of the PAS1 domain including Helix I by NMR perturbation experiments (Figure 3-1 and 3-2). These findings raise two possibilities which (1) the homodimerization of PHYB$^{647-782}$ provides a large accessible surface area of a pair of the β-sheets, or (2) a pair of the β-sheets masks each other. Interestingly, accumulation of PHYA in the nucleus occurred within minutes, and reached its maximum level after 10 min, whereas accumulation of PHYB in the nucleus reached its maximum level after 6h (Kircher et al., 2002). I showed PHYB$^{647-782}$ can form homodimerization, whereas PHYA$^{601-740}$ can not. The difference between PHYA and PHYB in the homodimerization ability on the PAS1 domain may participate in kinetics of nuclear translocation *via* accessibility of phytochrome nuclear translocator to the solvent-exposed β-sheet surface of the PAS1 domain.

The other dimerization site within the HKLD is similar to histidine kinase

transmitter modules found on bacterial sensor proteins (Schneider-Poetsch et al., 1991; Quail, 1997). The histidine kinase transmitter module is well-known as a dimerization domain (Wadhams and Armitage, 2004). For example, *E. coli* osmosensor EnvZ dimerizes through the homodimeric core region (Tomomori et al., 1999). This region includes one conserved histidine, which is the site of autophosphorylation and a phosphate transfer reaction (Cai and Inouye, 2003), and certain hydrophobic residues required to form a stable dimer with dissociation constant of ~10 μM (Hidaka et al., 1997; Tomomori et al., 1999). In the case of the HKLD, the hydorophobic residues are highly conserved, although the histidine residue is missing. Based on the structure of the EnvZ homodimeric core region, the modeled structure was generated using the homology modeling software MODELLER (Figure 3-9). The nicely packed dimer interface in the modeled structure strongly suggests that the homodimeric core region of the HKLD is associated with homodimerization.
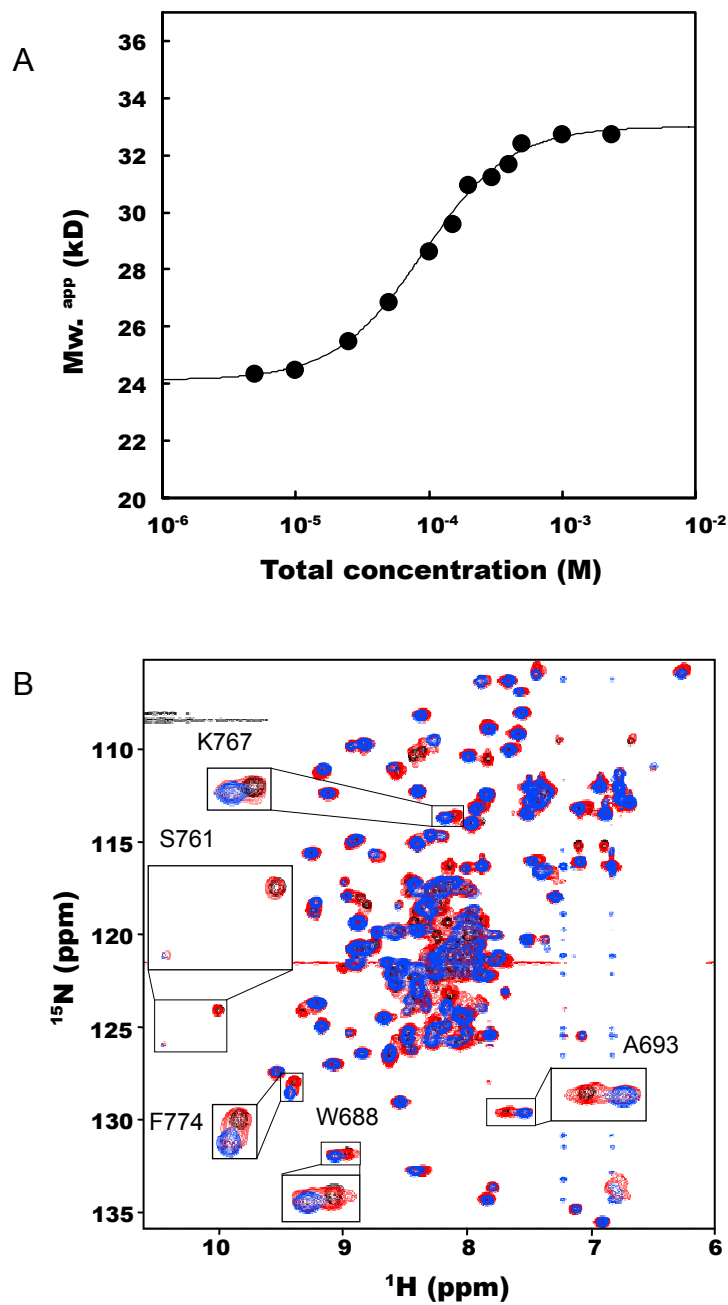
Figure 3-1. Homodimerization of rice PHYB PAS domain. (A) Plot of the protein concentration against the apparent molecular weight estimated from the elution volume in analytical SEC. (B) The ^{1}H-^{15}N HSQC spectra of rice PHYB$^{647-782}$. Protein concentrations are 0.025 (black), 0.2 (red) and 1.7 mM (blue).
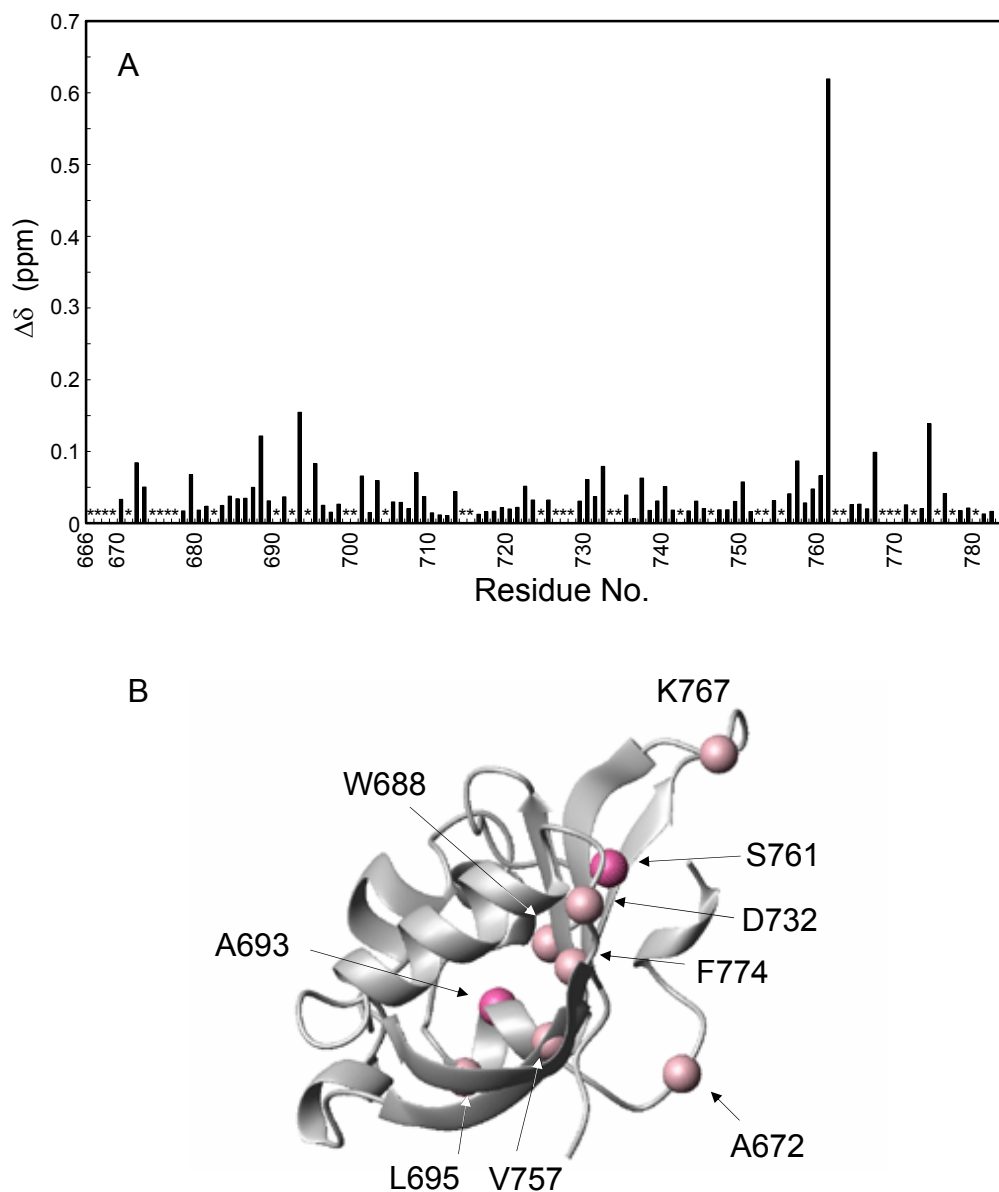
Figure 3-2. Chemical shift differences between the monomer and dimer of PHYB$^{647-782}$. (A) The bars represent combined ^{1}H and ^{15}N chemical shift differences ($\Delta\delta = [\Delta\delta(^1H)^2 + \{\Delta\delta(^{15}N) / 5\}^2]^{1/2}$) between 0.025 and 1.7 mM samples where the monomer and dimer are dominant, respectively. Asterisks indicate residues not assigned due to signal overlap or broadening. (B) The residues with significant chemical shift changes following dimer formation are mapped on the structure of PHYB$^{666-782}$; deep pink spheres, $\Delta\delta >$ 0.150 ppm; light pink spheres, $0.075 < \Delta\delta < 0.150$ ppm. Chemical shift changes are calculated by $\Delta\delta = [\Delta\delta(^1H)^2 + \{\Delta\delta(^{15}N) / 5\}^2]^{1/2}$.
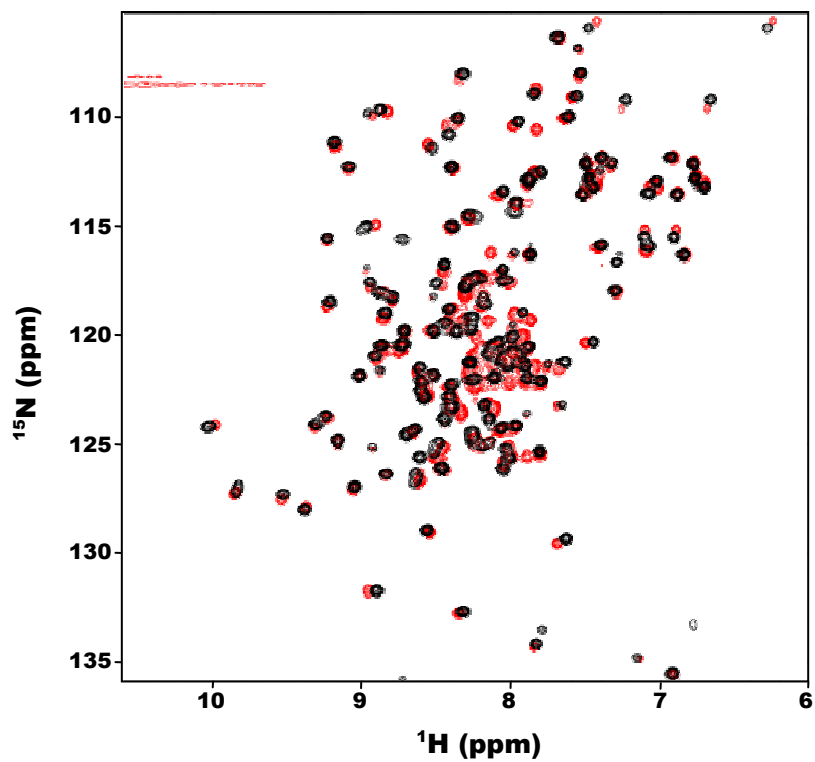
Figure 3-3. Superposition of the ^{1}H-^{15}N HSQC spectra of PHYB$^{666-782}$ (black) and PHYB$^{647-782}$ at 0.025 mM (red).

Constructs		Dimerization
L621–Q782		◯
G633–Q782		◯
L647–Q782		◯
V666–Q782		✕
G633–A672		△

Figure 3-4. Identification of homodimerization sites within the rice PHYB PAS1 domain.
PHYB$^{621\text{-}782}$, PHYB$^{633\text{-}782}$ and PHYB$^{647\text{-}782}$ form homodimers in a concentration-dependent manner, whereas PHYB$^{666\text{-}782}$ is a monomer, as suggested by analytical SEC and NMR spectroscopy. PHYB$^{633\text{-}672}$ tends to aggregate.

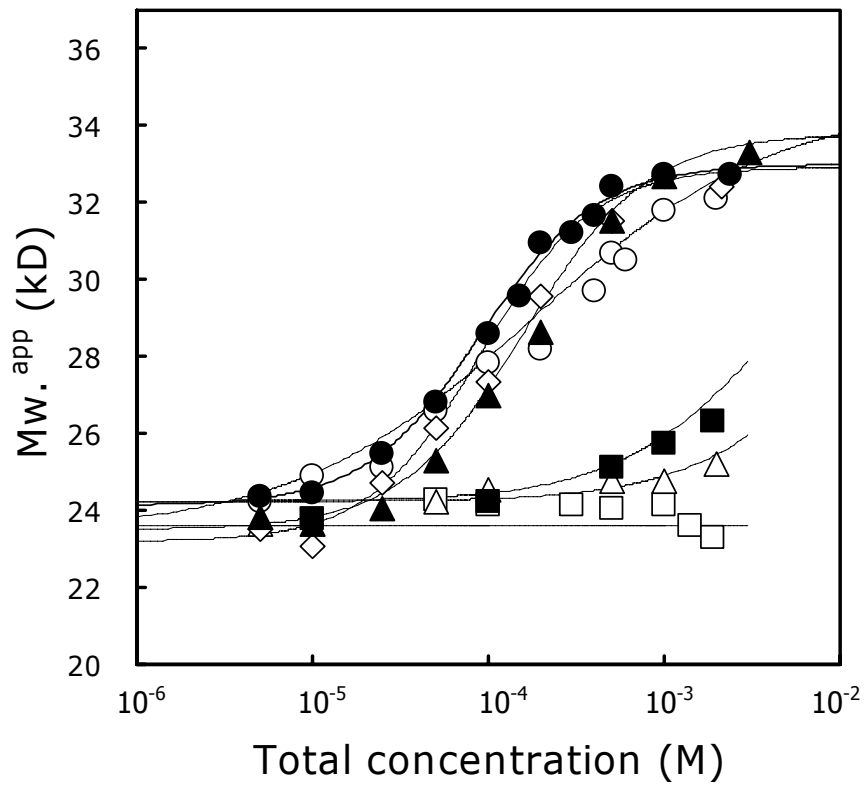**Figure 3-5. Homodimerization property of wild-type and missense mutants of PHYB[647-782].** Plot of the protein concentration against the apparent molecular weight estimated from the elution volume in analytical SEC. Closed circle, closed triangle, closed square, open circle, open triangle, open square and open diamond indicated WT, E651K, R663S, R653D, R663D, R653D/R663D and D656R, respectively.

A

B

Monomer   Dimer

WT (0.5 mM)

E651K (0.5 mM)

Monomer   Dimer

D656R (0.5 mM)   WT (1.7 mM)

R653D (0.5 mM)   R653D (1.7 mM)

R663S (0.5 mM)   R663S (1.7 mM)

R663D (0.5 mM)   R663D (1.7 mM)

R653D/R663D
(0.5 mM)   R653D/R663D
                                  (1.7 mM)

WT (25 μM)
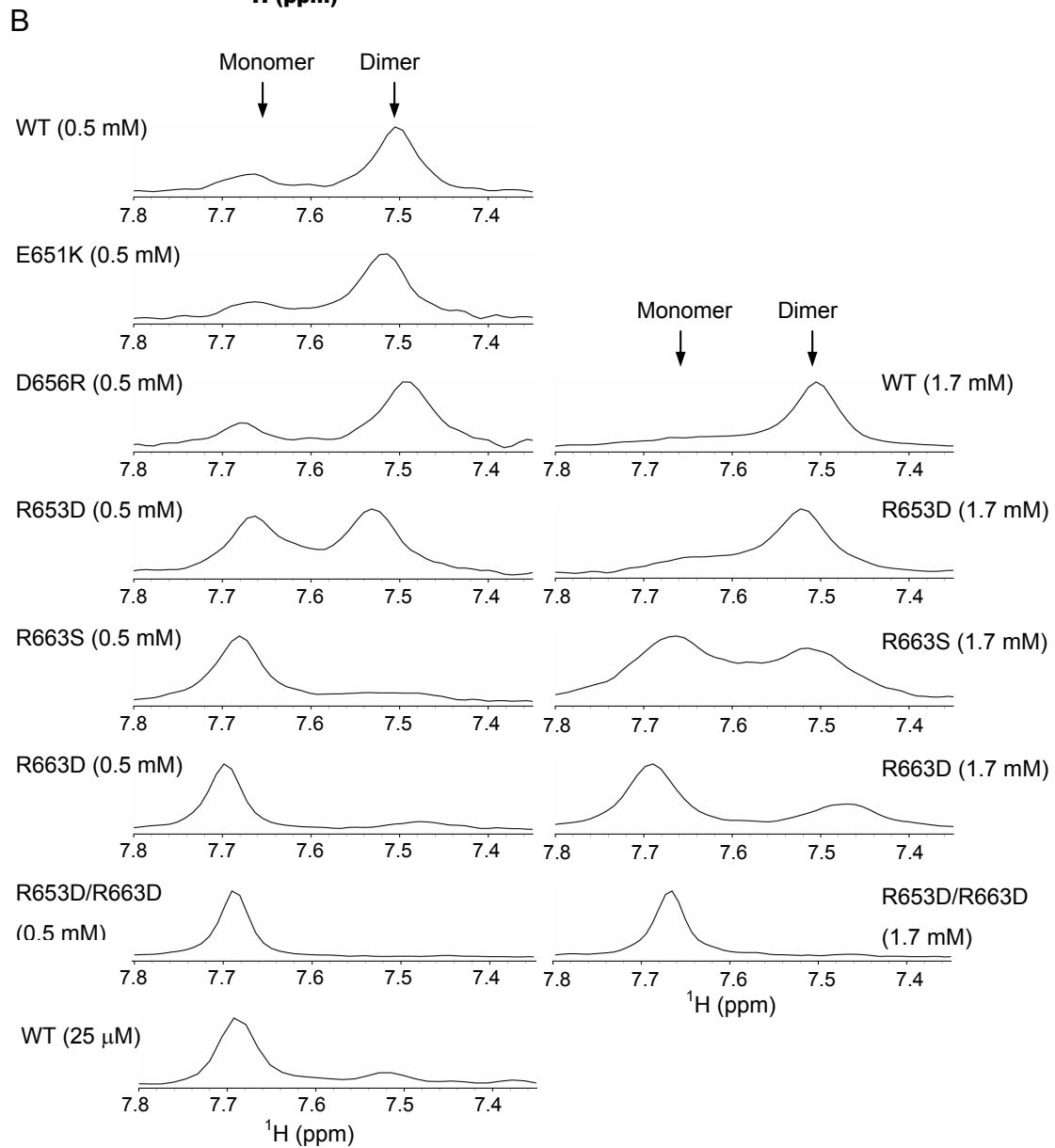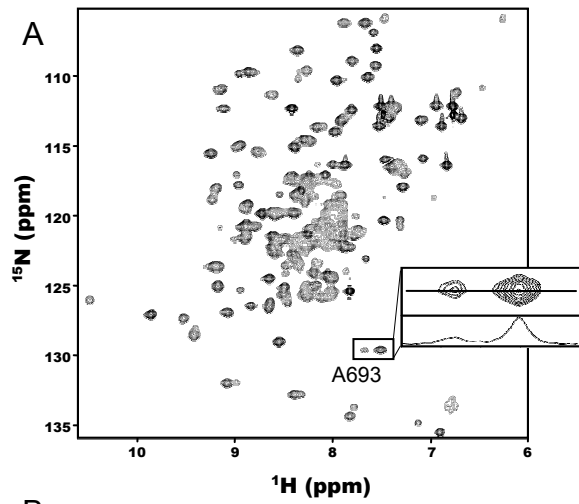
Figure 3-6. A693 signal on the ^{1}H/^{15}N HSQC spectra of PHYB$^{647\text{-}782}$.
(A) The ^{1}H-^{15}N HSQC spectra of rice PHYB$^{647\text{-}782}$ at protein concentration of 0.5 mM. Expanded region of A693 signal on the ^{1}H-^{15}N HSQC and one-dimensional slice at 129.6 ppm (^{15}N) are displayed. (B) One-dimensional slices at 129.6 ppm (^{15}N) of WT and missense mutants of PHYB$^{647\text{-}782}$ are displayed. Left (~ 7.7 ppm) and right (~ 7.5 ppm) indicated monomer and dimer of PHYB$^{647\text{-}782}$, respectively.
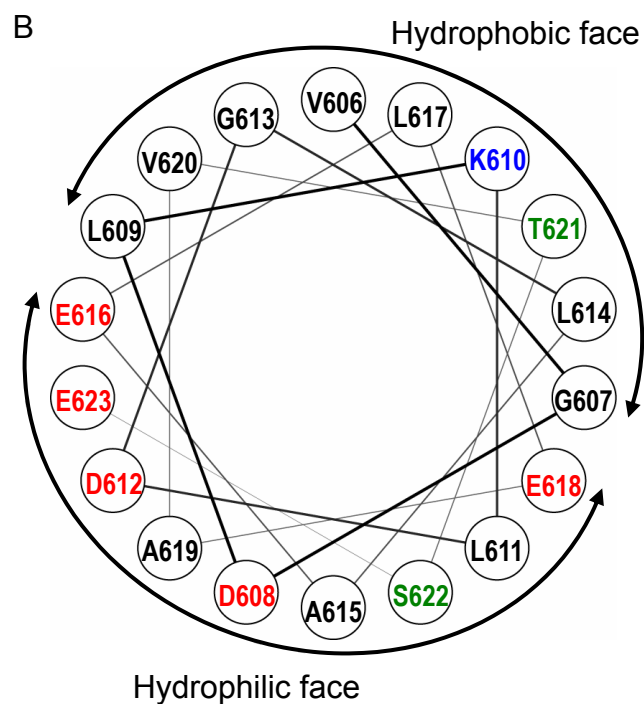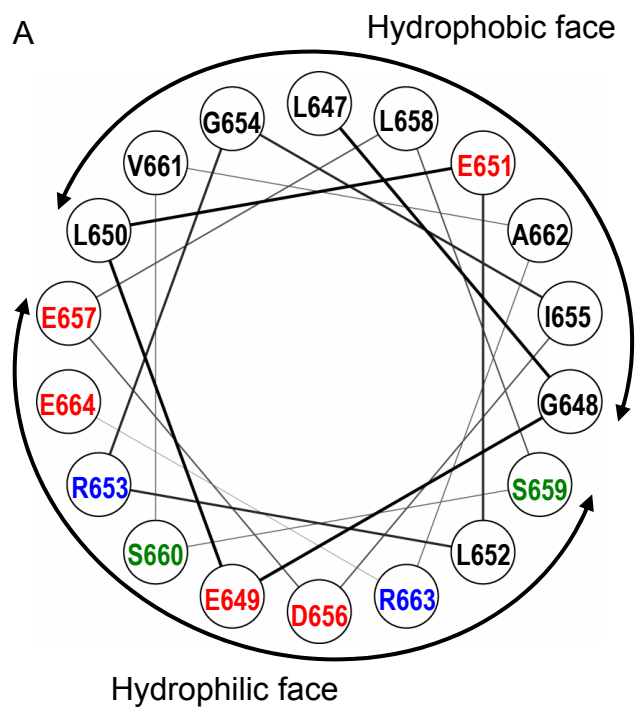
Figure 3-7. Helical wheel alighnment of (A) PHYB[647-664] and (B) PHYA[606-624].
Blue, red, green and black characters indicated positive-charged, negative-charged, non-polar and hydrophilic, and hydrophobic residues, respectively.
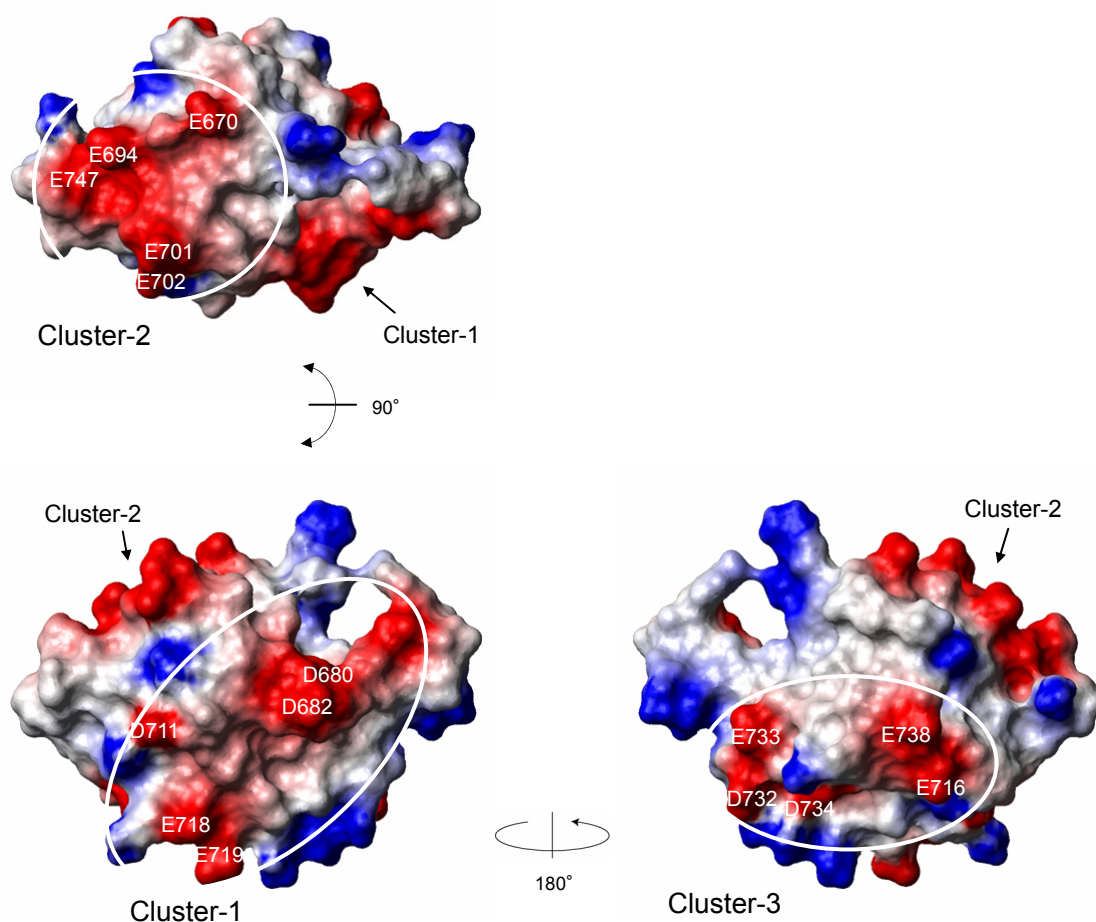
Figure 3.8. Electrostatic potential surfaces of PHYB$^{666\text{-}782}$. Positively and negatively charged residues are in blue and red, respectively. Circles in the figure display two negative charge clusters (L2 and L5 loops, and Cα and Dα helices, respectively) and one small positive charge cluster (Fα helix). Large acidic surfaces are characteristic. Comparing to PHYA, PHYB PAS1 domain contained three oppositely charged residues, K715, K740 and E747. Since rice PHYB PAS1 domain shares as high as 29.9% sequence identity among all PHYAs and PHYBs, these charge inversions may specify the intra- and inter-molecular interaction partners.

N-terminal
domain

PAS1

PAS2

HKLD

Figure 3-9. Model structure of the rice PHYB C-terminal domain. The structure of the PAS1 domain is determined by solution NMR. The structure of PAS2 and the HKLD are modeled by MODELLER (Sali and Blundell, 1993). Each subdomain is linked by a dotted line.
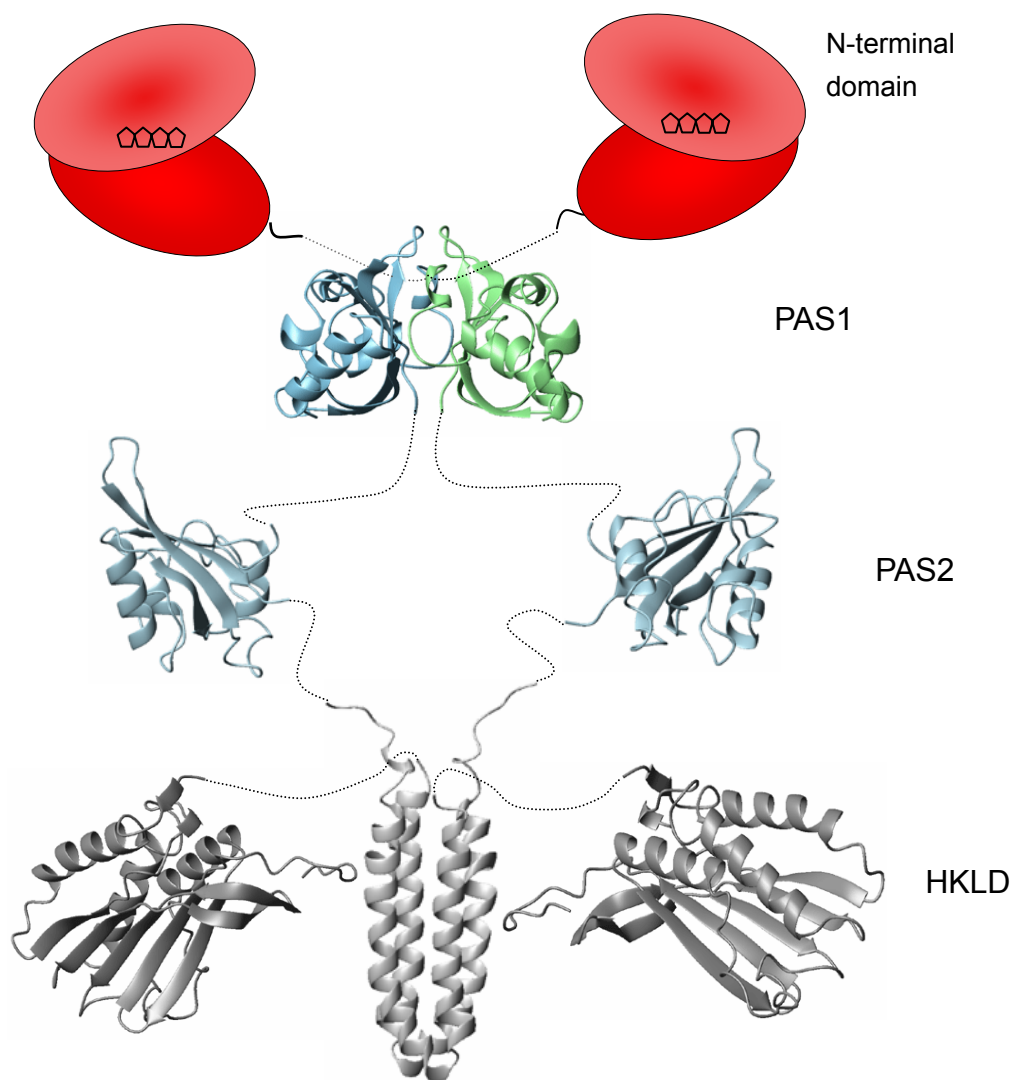
4. Gel-phase NMR: New methods for NMR-based screening approach

4.1. Introduction

NMR has been used for the structural biology, and structural genomics proteomics (SG) projects of small proteins (Adams et al., 2003; Yee et al., 2003), and has made significant contributions (Christendat et al., 2000; Yee et al., 2002; Savchenko et al., 2003). For example, in the SG project of *Methanobacterium thermoautotrophicum*, 17 structures out of 33 are determined by NMR (Yee et al., 2003). The most SG projects are targeting the microorganism, and thus the protein expression is highly successful (~ 95 %) (Lesley et al., 2002; Adams et al., 2003; Yee et al., 2003). However, the preparation of the soluble purified proteins is less efficient (40 % or less of expressed proteins), and this feature seems independent from the amino acid composition and species (Lesley et al., 2002; Yee et al., 2002; Adams et al., 2003). Developing the efficient technique to obtain soluble purified proteins, which is easily applied for the NMR measurement and the crystallization, is crucial for structural studies.

There are many approaches to obtain soluble and correctly folded proteins (Waldo, 2003; Yokoyama, 2003). For example, chemical refolding (Middelberg, 2002), low-temperature expression, different promoters, modified growth media (Makrides, 1996), directed evolution (Pedelacq et al., 2002), solubility enhancing tag (Kapust and Waugh, 1999), and so on. Among them, the solubility enhancing tag (SET) technique has recently applied for NMR study, and at least two solution structures are successfully determined (Zhou et al., 2001b; Vinogradova et al., 2002). They use maltose-binding

protein (MBP) (Vinogradova et al., 2002) and protein G B1 domain (GB1) (Zhou et al., 2001a), though there are many tags which can be used for this SET approach. Glutathione-S-transferase (GST) is one of the most popular tags, since the GST-fusion technique enhances the protein expression, solubility, stability, and used for the efficient protein purification and the binding assay on the glutathione sepharose column resin. However, the SET technique of GST in solution is not established so far.

Protein immobilization to the solid supports is another approach to obtain soluble and correctly folded proteins, and it is a well established technique (Bickerstaff, 1997). Immobilized protein is known to be stabilized (Stempfer et al., 1996b; Lei et al., 2002), and has been used for the refolding of itself since it seems not to be perturbed by aggregation (Stempfer et al., 1996a; Zahn et al., 1997; Rogl et al., 1998; Berdichevsky et al., 1999). NMR method to measure immobilized materials is known as "gel-phase" NMR (Keifer, 1997; Lippens et al., 1999; Shapiro and Gounarides, 1999), and well established for small chemicals and peptides. Most applications are focused on to monitor the chemical reaction process for the solid-phase organic synthesis and the combinatorial chemistry. To the best of myknowledge, the immobilized protein in aqueous solution is not studied by "gel-phase" NMR.

SET technique of GST is expected to have many strong advantages as described, however, some difficulties have been expected, such as signal broadenings due to higher molecular weight, and signal overlapping with GST peaks. I demonstrate here these problems are not crucial when the target protein is small, such as ubiquitin. The GST portion of the GST-tagged protein was not detected in ^{1}H-^{15}N HSQC and the target protein portion was easily assigned by a conventional triple-resonance procedure.

This approach enables the solution NMR measurement of the immobilized GST-fusion protein on the glutathione sepharose resin. The role of these new techniques in structural and biochemical studies has been discussed.

4.2. Materials and Methods

4.2.1. Sample preparation.

The human ubiquitin DNA was chemically synthesized, and inserted to pKT7 plasmid vectors containing Glutathione S-transferase (GST) and PreScission protease recognition site. The expressed protein had an artificial linker between GST and ubiquitin, SDLEVLFQGP-(GGGGS)$_n$, where n is 0, 1, 2, 3, and 4 for the linker length 10, 15, 20, 25, and 30 a. a., respectively. GST-ubiquitin was digested with PreScission protease, and resulted ubiquitin had extra amino acids, GP-(GGGGS)$_n$, at its N-terminus. Yeast ubiquitin hydrolase 1 (YUH1) was prepared as described previously (Sakamoto et al., 1999). The detail of the preparation for GST-ubiquitin and ubiquitin is described below.

Recombinant human ubiquitin was expressed in *Escherichia coli* BL21(DE3) as a GST-fusion protein. Following the addition of isopropyl β-D-thiogalactoside, protein expression was induced for 8 h at 30°C in minimal medium containing ^{15}N ammonium chloride and [u-^{13}C]-glucose. Cells were dispersed in 50 mM Tris HCl (pH8.0), 100 mM KCl and 1 mM Pefabloc SC (Roche) and then sonicated. Following centrifugation, the supernatant was loaded onto a glutathione sepharose 4B column (Amersham) and eluted with 50 mM Tris-HCl (pH 8.0), 100 mM KCl and 30 mM glutathione. The GST-ubiquitin fraction was loaded onto Superdex 26/60 75pg (Amersham) gel filtration column, equilibrated with 50 mM phosphate (pH 6.5), 50 mM KCl and 1 mM Pefabloc SC. For the preparation of ubiquitin, the GST-ubiquitin fraction was digested with PreScission protease (Amersham) and loaded onto the gel filtration

column. Purity was checked by SDS-PAGE (15 %). The final NMR buffer consisted of 50 mM phosphate (pH 6.7) and 200 mM KCl.

4.2.2. Immobilization of the GST-ubiquitin.

For NMR measurement of resin immobilized protein, the purified GST-ubiquitin were admixed to Glutathione Sepharose 4 Fast Flow resin (Amersham) equilibrated buffer C. The sample containing resin and GST-ubiquitin were transposed at 4 °C for 30 min, and then the supernatant was transferred to another tube. To remove free GST-ubiquitin, buffer C washed the resin tenth times.

The 0.07 mM GST-ubiquitin or ubiquitin in the resin were gently suspended by 1mL of buffer C with 10 % D_2O. The suspended solutions were transferred in NMR tube (Shigemi). After the resins settled out, the supernatant was removed, and inner tube was inserted.　The resins with GST-ubiquitin or ubiquitin were add to 1 ml the buffer C containing 10 % D_2O. The solutions containing the resins were gently suspended, and transferred in NMR tube (Shigemi). After the resins settled out, the supernatant was removed, and inner tube was inserted.

4.2.3. NMR spectroscopy.

NMR measurements of ubiquitin in solution condition (1.3 mM), GST-ubiquitin in solution condition (1.3 mM), and on resin condition (approximately 0.5 mM) were performed at 313 K with Bruker DRX800. Two-dimensional ^1H-^{15}N HSQC spectra were acquired with 1024(t1)×80(t2) complex points. The acquisition time is 8 min for ubiquitin and GST-ubiquitin in solution and 8 hr for GST-ubiquitin on

resin condition. All data were processed with the program NMRPipe (Delaglio et al., 1995) and analyzed with the program NMRView .

Spectra for ^{15}N T_1 and T_2 were acquired at 300 K and a ^{15}N frequency of 80.1MHz. Enhanced-sensitivity pulse sequences were used. The T_1 relaxation delays were 19, 94, 184, 289, 414, 574, 779 and 1099 ms and the T_2 relaxation delays were 14.4, 28.8, 43.2, 57.6, 72.0, 86.4, 144.0 and 187.2 ms. The delay between ^{15}N 180° pulses in the CPMG sequence for the T_2 measurements was 900 μs. Each T_1 and T_2 was determined by fitting the measured intensities by a two-parameter single-exponential function using the program CURVEFIT. The uncertainties in the measured peak heights for the T_1 and T_2 measurements were estimated by repeating the experiments [T = 19 ms for the T1 measurements, and T = 14.4 ms for the T_2 measurements]. Rotating correlation time was calculated by TENSOR2 software. The apparent molecular weight,

M_r^{app}, was evaluated from the relation, $\tau_C = \dfrac{4\pi\eta_W}{3k_B T}\left(\sqrt[3]{\dfrac{3\overline{V}M_r^{app}}{4\pi N_A}} + r_W\right)^3$, where τ_C, η_W,

k_B, T, $\overline{V}$, N_A, and r_W are the rotational correlation time, the solvent water viscosity $0.8565*10^{-3}$ cP at 300 K, Boltzmann constant, temperature, the specific volume of protein 0.73 cm^3/g, Avogadro's number, and a hydration layer $3.2*10^{-10}$ m, respectively.

4.3. Results

4.3.1. ^{1}H-^{15}N HSQC of ubiquitin and GST-ubiquitin

^{15}N or ^{13}C/^{15}N enriched GST-fusion ubiquitin was expressed and purified by a conventional procedure. A 10 amino acids linker between GST and ubiquitin, SDLEVLFQGP, was introduced as a PreScission protease (Amersham Biosciences) recognition site. The protease digested ubiquitin had two additional amino acids, GP. In this report, GST-fusion ubiquitin including the 10 amino acids linker was designated as GST-ubiquitin, and the digested product, GP-ubiquitin, as ubiquitin.

^{1}H-^{15}N HSQC spectra of ubiquitin and GST-ubiquitin are shown in Figure 4-1A and 1B, respectively. Based on the conventional triple resonance procedure (Cavanagh et al., 1996), observed peaks were completely assigned for ubiquitin, and mostly for GST-ubiquitin. Though myengineered ubiquitin had two extra residues on its N-terminal, the assignments were similar to those reported previously (Stockman et al., 1993). The spectral differences between ubiquitin and GST-ubiquitin were quite small, where several new peaks were appeared on the GST-ubiquitin spectrum and a few peaks were missing. The new peaks observed in GST-ubiquitin were also observed for GST protein, and thus, assigned for the GST portion of GST-ubiquitin (shown by asterisk in Figure 4-1B). Other GST peaks were much broader, and the conventional assignment procedure did not work for GST at all. The missing peaks were T9, T12, A46, and G47 of ubiquitin, which were found in the loop region. Jeon *et al.* reported similar situation that the flexible C-terminal domain of *E. coli* RNA polymerase alpha subunit exhibited almost the same ^{1}H-^{15}N HSQC spectrum as the isolated C-terminal fragment, though the

other domain was dimmer and not detected such as GST (Jeon et al., 1997). These results indicated that the ubiquitin portion of GST-ubiquitin retained the same conformation as the isolated ubiquitin, and that its motion was independent of that of GST.

The apparent overall correlation times (τ_C) of ubiquitin and the ubiquitin portion of GST-ubiquitin were determined by ^{15}N T_1 / T_2 ratio method assuming the isotropic overall tumbling using 37 independent signals (Table 4-1). Ub and GST-L$_{10}$-Ub in Table 4-1 are ubiquitin and GST-ubiquitin, respectively. The determined correlation time of GST-ubiquitin was about 2 times of ubiquitin, though that of molecular weight was nearly 8 times. As shown in Table 4-1, the apparent molecular weight (Mr^{app}) could be calculated from the correlation time. Theoretically the correlation time of the globular protein was expected to be linearly correlated with the molecular weight. The observed big difference (~ 3 times) between molecular weight and the apparent molecular weight indicated that the ubiquitin portion of GST-ubiquitin had an additional faster motion which could be caused by the amino acids linker between GST and ubiquitin. In fact, it qualitatively agreed with the observed ^{1}H-^{15}N HSQC spectra, where most signals were from the ubiquitin portion.

The linker length between GST and ubiquitin could be a parameter to obtain the high quality NMR spectrum. Three GST-fusion ubiquitin proteins, GST-L$_{10}$-Ub, GST-L$_{20}$-Ub, and GST-L$_{30}$-Ub, with different linker lengths, 10, 20, and 30 amino acids, respectively, were prepared and used for the ^{15}N T_1 and T_2 measurements. The determined correlation times, given in Table 4-1, did not depend on the linker length. No significant difference was found between the observed spectra of three samples,

except for some new peaks assigned to the linker glycines and serines of GST-L$_{20}$-Ub and GST-L$_{30}$-Ub. Thus, it was reasonable to assume that the shortest linker, SDLEVLFQGP, gave the enough flexibility for ubiquitin part of GST-ubiquitin.

4.3.2. ^{1}H-^{15}N HSQC of immobilized GST-ubiquitin

If the motion of the ubiquitin portion of GST-ubiquitin is independent from that of GST, ^{1}H-^{15}N HSQC spectrum of the ubiquitin portion of the immobilized GST-ubiquitin on resin can be measured by solution NMR technique. The sample used for solution NMR is necessary to be highly homogenous in terms of the magnetic susceptibility. Solution-resin mixture sample, the immobilized GST-ubiquitin on resin, is not homogenous definitely, and such NMR measurement causes significant line broadenings without special treatments. In fact the conventional adjustment of the magnetic field homogeneity by shimming did not work well in mysolution-resin mixture system.

The magnetic field inhomogeneity of the solution-resin mixture sample was eliminated using a reference water sample as shown in Figure 4-2A. To evaluate the magnetic homogeneity, the ^{1}H-^{15}N HSQC spectra of ubiquitin was measured in solution condition as well as in solution-resin mixture condition where ubiquitin was not immobilized. The averaged ^{1}H linewidths of 37 well separated peaks were 21.1 ± 2.0 Hz and 27.3 ± 1.1 Hz for solution and solution-resin mixture, respectively. Though about 6 Hz broadening induced by the magnetic field inhomogeneity was observed, the obtained magnetic field was enough homogenous to measure the ^{1}H-^{15}N HSQC spectrum.

The ^{1}H-^{15}N HSQC spectrum of the immobilized GST-ubiquitin was successfully measured under the adjusted homogenous magnetic field (Figure 4-2B). The observed peaks were well distinguished but much broader than those in solution. All peaks from solution GST-ubiquitin (Figure 4-1B) were easily identified in the immobilized GST-ubiquitin spectrum without shifts. These signals were from neither the released solution samples nor the degradation products, revealed by the overloaded SDS-PAGE. The linker length did not change the overall S/N ratio and the apparent linewidth. It is noted that, two missing peaks in solution were observed in the immobilized samples having longer linkers, that is, G47 were observed for GST-L$_{20}$-Ub and GST-L$_{30}$-Ub, and A46 for GST-L$_{30}$-Ub.

4.3.3. Specific protein-protein interaction monitored in solution and on resin

Yeast ubiquitin hydrolase 1 (YUH1) is a cystein protease that catalyzes the removal of ubiquitin C-terminal adducts, and has an important roll for the generation of monomeric ubiquitin. YUH1 interacts with yeast ubiquitin specifically ($K_d \sim 19$ μM), and significant peak shifts for yeast ubiquitin are observed in the presence of YUH1, -0.4 to +0.2, and -2 to 3 ppm for ^{1}H and ^{15}N, respectively (Sakamoto et al., 1999). The differences between yeast and human ubiquitin are only three residues, Ser19 to Pro19, Asp24 to Glu24, and Ser28 to Ala28, respectively. These residues are located at the opposite site of the YUH1 interaction surface, and thus, it is reasonable to assume that YUH1 specifically binds human ubiquitin as well. Figure 4-3A shows the overlaid ^{1}H-^{15}N HSQC spectra of human GST-ubiquitin with (red) and without (black) YUH1 recorded in solution. YUH1 was not ^{15}N-enriched and not observed. Though the molecular ratio of GST-ubiquitin to YUH1 was 10 to 1, significant line broadenings

were observed for Thr7, Leu8, Ile13, Glu34, Gly47, Lys48, His68, Leu69, and Leu71, large peak shifts for Val5, Lys11, Thr14, Gln41, Arg42, Ile44, Gln49, Val70, and Arg74, and small peak shifts for Val26, Ile30, Asp32, Leu50, Asp52, and Leu73. The broadened and shifted residues were similar to those for yeast ubiquitin interacting with YUH1 (Sakamoto et al., 1999). These results indicated that the interaction between human ubiquitin and YUH1 was specific and similar to that of yeast ubiquitin.

The GST-fusion proteins are frequently used for biochemical binding assay, such as GST pull-down experiment, where the GST-fusion protein immobilized on resin was used for the binding experiments. However, it is not clear that the immobilized GST-fusion protein interacts with the target protein on resin with the same intermolecular surface as in solution. In Figure 4-3B, the overlaid ^{1}H-^{15}N HSQC spectra of immobilized GST-ubiquitin with (red) and without (black) YUH1 were shown. Though the apparent linewidth was different, the broadened and shifted residues were quite similar to those in solution (Figure 4-3A). Thus the intermolecular interaction of the immobilized protein on resin is similar to that in solution.

4.4. Discussion

4.4.1. Solubility enhancing tag (SET) selection for structural and functional studies

For the proteome-scale protein production, affinity tag is widely used and the solubility of all tagged proteins is quantitatively evaluated (Braun et al., 2002; Hammarstrom et al., 2002; Shih et al., 2002). 60 ~ 90 % of expressed fusion proteins are soluble for MBP-tag, 40 ~ 50 % for GST-tag, and 50 ~ 60 % for transcription termination anti-termination factor (NusA) tag. Other tags are much less effective. Hammarström *et al.* have reported that GB1 tag, thioredoxin (TRX) tag, and protein A Z domain (ZZ) tag give high solubility, 68, 74, and 55 %, respectively, for small proteins (6-19 kDa) (Hammarstrom et al., 2002). At least MBP, GST, NusA, GB1, TRX, and ZZ are solubility enhancing tags and applicable for NMR study. As demonstrated here, GST tag has strong advantage over other tags, including MBP and GB1. That is, GST tag is not observable in ^{1}H-^{15}N HSQC spectrum. In fact MBP and GB1 tags give NMR signals (Zhou et al., 2001a; Vinogradova et al., 2002), and both TRX and ZZ tags are small and assumed to give NMR signals.

Application of the SET technique is not limited for NMR, but used for X-ray crystal studies (Zhan et al., 2001; Smyth et al., 2003). The major difference is the length of the linker between tag and target protein. In X-ray studies, rigidly fused protein with 3 to 5 amino acids short linker is preferred for crystal packing. In NMR studies, loosely fused protein with more than 10 amino acids long linker is preferred for flexibility. Protein NMR measurement on resin demonstrated here is a quite unique approach using

SET technique. Roughly 30 % of soluble proteins are classified as 'aggregated' by NMR spectroscopy (Christendat et al., 2000). Protein NMR measurement on resin potentially suppresses such nonspecific aggregation, and thus it may become one of the major NMR techniques.

4.4.2. "Gel-phase" NMR by HR-MAS method

"Gel-phase" samples often give the very broad signals, 100-300 Hz or more for [1]H (Mazure et al., 1986; Keifer et al., 1996). This linebroadening can arise from either limited motional freedom (strong dipole interaction with resin), chemical shift heterogeneity, and magnetic susceptibility variations within the sample. To overcome these problems, high-resolution magic-angle-spinning (HR-MAS) probe is developed and successfully applied for small molecules in "gel-phase" (Keifer, 1997; Lippens et al., 1999; Shapiro and Gounarides, 1999). That indicates the chemical shift heterogeneity is not serious for "gel-phase" samples, since HR-MAS method can not eliminate it. In view of protein NMR application, the limitation of HR-MAS method is the sensitivity. Typical S/N ratio for HR-MAS probe is about 50, which is one-thirtieth of conventional 5 mm triple resonance probe used in this study.

Here HR-MAS probe is not used, but the NMR measurement of "gel-phase" protein sample is successful. In myapplication, the magnetic susceptibility difference is eliminated using the reference water sample. The dipolar interaction is thought to be eliminated by the fast molecular motion. That is, the ubiquitin protion of GST-ubiquitin is flexible, and the dipolar interaction is averaged out. Additionally the GST portion

gives space between the ubiquitin portion and resin, and it attenuates the magnetic susceptibility differences and the dipolar interaction with resin which is a strong relaxation source. Such contribution of the GST portion could be crucial, since several trials of "gel-phase" NMR measurement using poly histidine tagged proteins did not work well (Kobayashi *et al.*, unpublished results).

4.4.3. Protein-protein interaction in functional studies

Weak protein-protein interactions on the scaffold proteins and membranes seem to be crucial for many signal transduction systems. NMR is one of the most powerful tools to monitor such weak interactions since it can distinguish the significant specific interactions from the non-specific interactions by the structure-based perturbation experiments. Thus, biochemical binding assay of GST-fusion proteins using ^{1}H-^{15}N HSQC spectra in solution and on resin, will be useful for such biologically significant system, and applicable for the large scale systematic studies, such as functional genomics and drug screening.

What is the limitation of GST-fusion protein NMR approach to monitor protein-protein interactions? Molecular weight of the interacting target molecule will not be the limitation. For example, Shimada and co-workers have developed the saturation transfer experiments to be applicable for the larger complex system (Takahashi et al., 2000; Nakanishi et al., 2002; Nishida et al., 2003), and this technique is applicable in GST-fusion protein NMR approach. In solution condition GST-fusion protein should be small (~10 kDa ubiquitin in mysystem) since the signals from the

GST portion of GST-fusion protein are preferred to be negligible. On resin condition, however, the target protein is not limited to be small. Apparent pore size of Sepharose 4 FF resin (Amersham) used for HSQC measurements, is about 45 nm determined by hydrodynamic elution data analyses (Hagel et al., 1996). Molecular diameters of ubiquitin and ubiquitin-YUH1 complex are expected to be less than 3 and 6 nm, respectively, thus this resin has enough space for larger molecules and does not perturb the intermolecular interactions.

4.4.4. High-throughput drug screening using the gel-phase NMR method

NMR has known as a valuable drug screening tool, as well as a protein-structure determination tool. Methods for detection of interaction between protein and ligand are classified into two main categories: monitoring NMR signals from the protein, and the ligand (Pellecchia et al., 2002a). For monitoring NMR signals from protein, a common approach is a chemical-shift mapping. Binding of a ligand alters the chemical environment around the binding site of the protein, and so will perturb the chemical shift of magnetic nuclei at this site. These changes occur mainly by nuclear-spin relaxation and/or chemical exchange, and can be observed using ^{1}H/^{15}N and/or ^{1}H/^{13}C correlation spectra. Other approach is measurement of nuclear Overhauser effects (NOEs). The magnitude of the effects depends strongly on the distance that separates the two dipolar-coupled spins, and so NOEs are sensitive probes of short-range through-space intramolecular and intermolecular interactions.

Recent reports provided some strategies of NMR-based high-throughput screening such as SAR by NMR and competition binding experiments, using the aforementioned approaches (Shuker et al., 1996; Fejzo et al., 1999; Dalvit et al., 2000; Dalvit et al., 2002; Pellecchia et al., 2002b; Pellecchia et al., 2002a). However, large amount of the protein are required for these NMR-based screening methods, because it changes the sample for each compound mixtures. Gel-phase NMR method provides a good solution to this problem since the protein immobilize on resin, and the compounds which are not bound to the protein are easily able to be washed away.

Some PAS domains have an ability to bind small molecule to sense environmental conditions. Thus the rice PHYB PAS1 domain may have the ability to bind small molecule. If PHYB PAS1 domain binding ligands are discovered by gel-phase NMR method, phytochrome function may controlled by ligands artificially.

Table 4-1. Linker length dependence of rotational correlation times and apparent molecular weights in solution

	Mw	τ_C (ns)	Mr^{app} (kDa)
Ub	9028	5.6 ± 0.1	12.7
GST-L$_{10}$-Ub	70218	9.6 ± 0.3	24.1
GST-L$_{20}$-Ub	71478	9.9 ± 0.4	24.9
GST-L$_{30}$-Ub	72738	9.6 ± 0.5	24.1
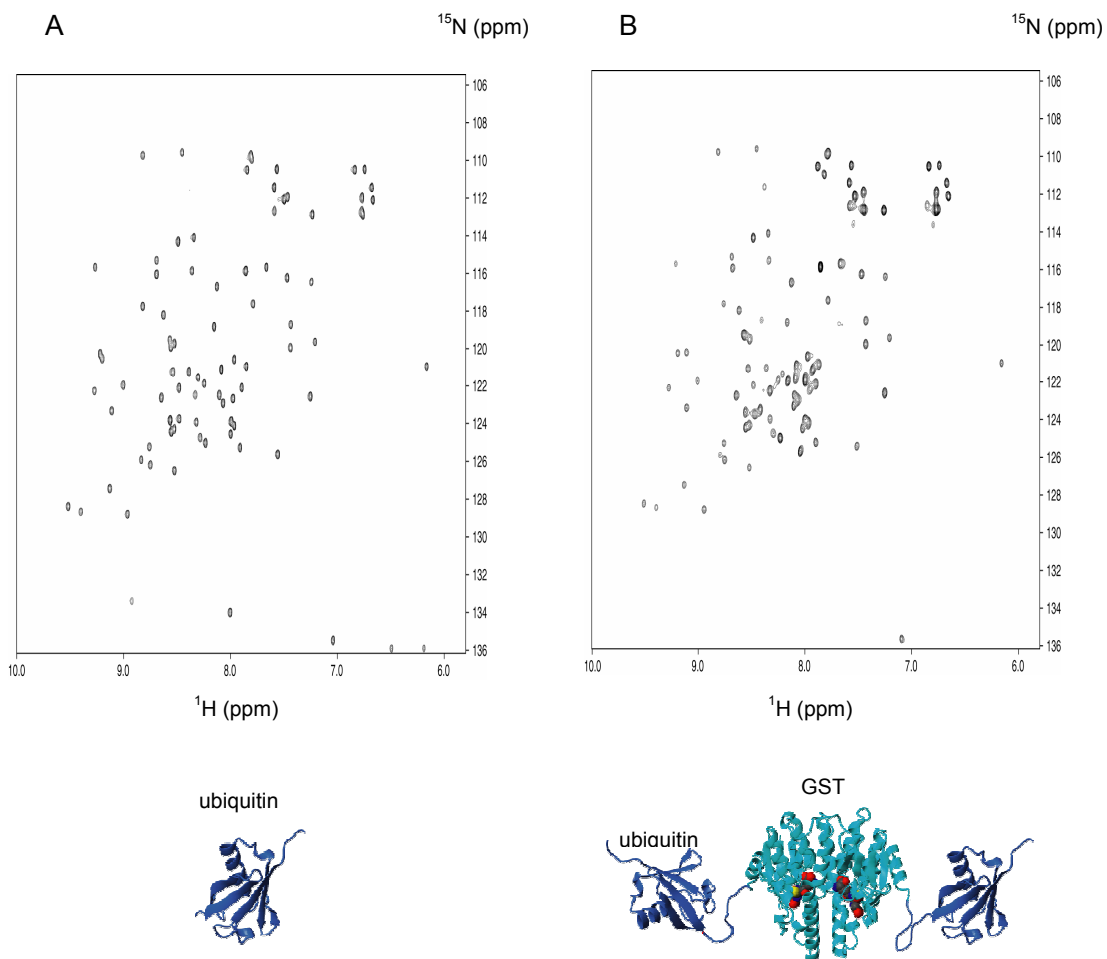
Figure 4-1. **^{1}H-^{15}N HSQC spectra of ubiquitin (A) and GST-ubiquitin (B) in solution.**

Most NMR peaks of GST portion of GST-ubiquitin were not observed though the sample was enriched uniformly.

The schematic diagrams of the sample condition are shown at the bottom of each spectrum.
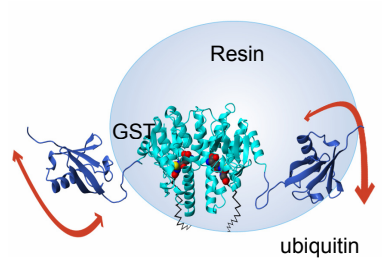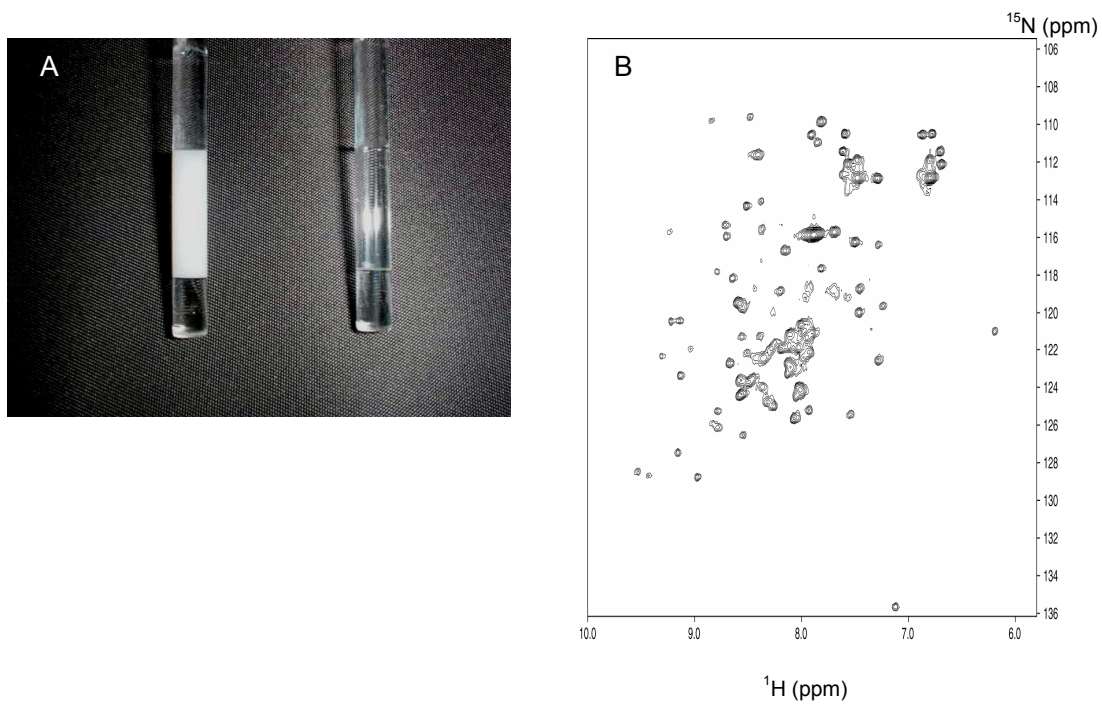
Figure 4-2. (A) NMR sample tubes filled with the glutathione-sepharose column resin (left) and the water (right). Magnetic field inhomogeneity of the resin sample was eliminated using the water sample as a reference. (B) ^{1}H-^{15}N HSQC spectrum of the immobilized GST-ubiquitin on the resin. The schematic diagram of the sample condition is shown at the bottom.
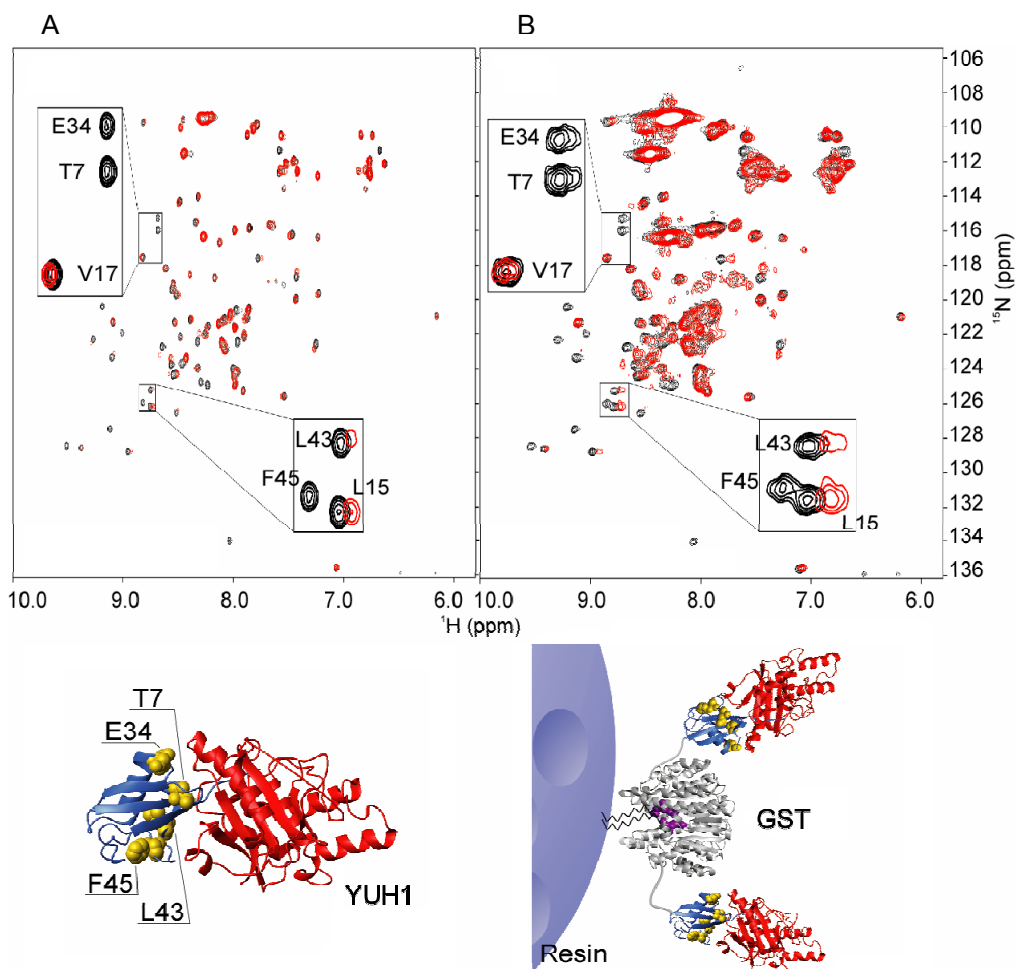
Figure 4-3. **^{1}H-^{15}N HSQC spectra of GST-ubiquitin in the presence of YUH1 in solution.**
(A) and on the resin (B). Selected spectral region is expanded and their peak assignments are given. Based on the
ubiquitin-TUH1 complex structure, the schematic diagrams of the sample condition are shown at the bottom. Amino
acid residues with broadened or largely shifted peaks are drawn by the space filling representation with residue
names.

5. General conclusion

Plants employ PHY proteins to survive and to respond to fluctuations in the red/far-red light environment. The PHYs are dimeric chromoproteins that regulate the expression of a large number of light-responsive genes and thus influence many photomorphogenic events. Although many efforts to obtain crystal structures of PHYs have been made, no atomic-resolution structure has been reported.

In this thesis, a solution structure of a fragment of PHY, the PAS1 domain of rice PHYB was determined (Chapter 2; Figure 2-2). The PAS1 domain (residue 677-774) had a typical PAS-fold structure where 4 α-helices and several loops were wrapped around with one large β-sheet. The core region of loss-of-function missense mutations known as the Quail-box was mapped on the β-sheet (Iβ and Hβ) of the PAS1 domain (Figure 2-6). For phyB, 2 mutants (A759V and G776R) were known in this region that induce hypocotyl elongation under continuous red-light conditions and these mutation sites were locating on the β-sheet, which are consisted of conserved residues among all PHYAs and PHYBs. One of these, the G776R mutant (G767R in *Arabidopsis*), disrupted nuclear translocation, where the ternary structure of the PAS1 domain was partially disrupted, as revealed by NMR (Figure 2-6). These results strongly suggest that structural perturbation on the β-sheet side directly affects photomorphogenesis.

One of the dimerization sites of phyB was found in the N-flanking region of the PAS1 domain (Chapter 3). Two positively charged residues, R653 and R663, in this hinge-region were crucial for dimerization (Figure 3-5 and 3-6). The hinge-region could not form homodimer by itself, and the PAS1 domain was required for stable dimer formation (Figure 3-4). These results suggest the positively charged residues on the hinge-region interact with negatively charged residues on the PAS1 domain. Based on NMR perturbation experiments, the dimerization interface was mapped on the β-sheet side of the PAS1 domain including Helix I, where negatively charged residues were found (Figure 3-2). The location of these residues was consistent with the conserved residues mapped on the PAS1 structure. Stable dimer formation of the β-sheet side of two PAS1 domains may be linked to PHYB function.

Gel-phase NMR method was developed and applied to structural and biochemical studies of proteins (Chapter 4). In the case of GST-tagged ubiquitin, the GST portion was not detected in ^{1}H-^{15}N HSQC and the target protein portion was easily assigned by a conventional triple-resonance procedure (Figure 4-1). This approach enabled the solution NMR measurement of the immobilized GST-fusion protein on the glutathione sepharose resin (Figure 4-2). Furthermore, specific protein-protein interaction was monitored by gel-phase NMR method (Figure 4-3). Thus, biochemical binding assay of GST-fusion proteins using ^{1}H-^{15}N HSQC spectra in solution and on

resin, will be useful for biologically significant systems, and applicable for the large

scale systematic studies based on the structural information, such as functional

genomics and drug screening.

6. Acknowledgements

Finally, I would like to thank my parents for their supports and understanding with endless love.

Toshitatsu Kobayashi

December 26, 2005

7. References

Adams, M.W., Dailey, H.A., DeLucas, L.J., Luo, M., Prestegard, J.H., Rose, J.P., and Wang, B.C. (2003). The Southeast Collaboratory for Structural Genomics: a high-throughput gene to structure factory. Acc Chem Res **36,** 191-198.

Archer, S.J., Ikura, M., Torchia, D.A., and Bax, A. (1991). An alternative 3D NMR technique for correlating backbone 15N with side-chain Hβ resonances in larger proteins. J Magn Reson **95,** 636-641.

Baca, M., Borgstahl, G.E., Boissinot, M., Burke, P.M., Williams, D.R., Slater, K.A., and Getzoff, E.D. (1994). Complete chemical structure of photoactive yellow protein: novel thioester-linked 4-hydroxycinnamyl chromophore and photocycle chemistry. Biochemistry **33,** 14369-14377.

Bairoch, A., and Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res **28,** 45-48.

Berdichevsky, Y., Lamed, R., Frenkel, D., Gophna, U., Bayer, E.A., Yaron, S., Shoham, Y., and Benhar, I. (1999). Matrix-assisted refolding of single-chain Fv- cellulose binding domain fusion proteins. Protein Expr Purif **17,** 249-259.

Bickerstaff, G.F. (1997). Immobilization of Enzymes and Cells. (Totowa: Humana Press).

Borthwick, H.A., Hendricks, S.B., Parker, M.W., Toole, E.H., and Toole, V.K.

(1952). A reversible photoreaction controlling seed germination. Proc Natl Acad Sci U S A **38,** 662-666.

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

Briggs, W.R., Beck, C.F., Cashmore, A.R., Christie, J.M., Hughes, J., Jarillo, J.A., Kagawa, T., Kanegae, H., Liscum, E., Nagatani, A., Okada, K., Salomon, M., Rudiger, W., Sakai, T., Takano, M., Wada, M., and Watson, J.C. (2001). The phototropin family of photoreceptors. Plant Cell **13,** 993-997.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., and Warren, G.L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr **54 (Pt 5),** 905-921.

Cai, S.J., and Inouye, M. (2003). Spontaneous subunit exchange and biochemical evidence for trans-autophosphorylation in a dimer of Escherichia coli histidine kinase (EnvZ). J Mol Biol **329,** 495-503.

Card, P.B., Erbel, P.J., and Gardner, K.H. (2005). Structural basis of ARNT PAS-B dimerization: use of a common beta-sheet interface for hetero- and homodimerization. J Mol Biol **353,** 664-677.

Casal, J.J., Luccioni, L.G., Oliverio, K.A., and Boccalandro, H.E. (2003). Light, phytochrome signalling and photomorphogenesis in Arabidopsis. Photochem Photobiol Sci **2,** 625-636.

Cashmore, A.R., Jarillo, J.A., Wu, Y.J., and Liu, D. (1999). Cryptochromes: blue light receptors for plants and animals. Science **284,** 760-765.

Cavanagh, J., Fairbrother, W.J., Palmer III, A.G., and Skelton, N., J. (1996). Protein NMR spectroscopy. (San Diego: Academic Press).

Chen, M., Schwab, R., and Chory, J. (2003). Characterization of the requirements for localization of phytochrome B to nuclear bodies. Proc Natl Acad Sci U S A **100,** 14493-14498.

Cherry, J.R., Hondred, D., Walker, J.M., Keller, J.M., Hershey, H.P., and Vierstra, R.D. (1993). Carboxy-terminal deletion analysis of oat phytochrome A reveals the presence of separate domains required for structure and biological activity. Plant Cell **5,** 565-575.

Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J.R., Booth, V., Mackereth, C.D., Saridakis, V., Ekiel, I., Kozlov, G., Maxwell, K.L., Wu, N., McIntosh, L.P., Gehring, K., Kennedy, M.A., Davidson, A.R., Pai, E.F., Gerstein, M., Edwards, A.M., and Arrowsmith, C.H. (2000). Structural proteomics of an archaeon. Nat Struct Biol **7,** 903-909.

Clack, T., Mathews, S., and Sharrock, R.A. (1994). The phytochrome apoprotein

family in Arabidopsis is encoded by five genes: the sequences and expression of PHYD and PHYE. Plant Mol Biol **25,** 413-427.

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

Cusanovich, M.A., and Meyer, T.E. (2003). Photoactive yellow protein: a prototypic PAS domain sensory protein and development of a common signaling mechanism. Biochemistry **42,** 4759-4770.

Dalvit, C., Pevarello, P., Tato, M., Veronesi, M., Vulpetti, A., and Sundstrom, M. (2000). Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. J Biomol NMR **18,** 65-68.

Dalvit, C., Flocco, M., Knapp, S., Mostardini, M., Perego, R., Stockman, B.J., Veronesi, M., and Varasi, M. (2002). High-throughput NMR-based screening with competition binding experiments. J Am Chem Soc **124,** 7702-7709.

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

Edgerton, M.D., and Jones, A.M. (1992). Localization of protein-protein interactions between subunits of phytochrome. Plant Cell **4,** 161-171.

Edgerton, M.D., and Jones, A.M. (1993). Subunit interactions in the carboxy-terminal

domain of phytochrome. Biochemistry **32,** 8239-8245.

Elich, T.D., and Chory, J. (1997). Biochemical characterization of Arabidopsis

wild-type and mutant phytochrome B holoproteins. Plant Cell **9,** 2271-2280.

Fankhauser, C. (2001). The phytochromes, a family of red/far-red absorbing

photoreceptors. J Biol Chem **276,** 11453-11456.

Fankhauser, C., and Chory, J. (1997). Light control of plant development. Annu Rev

Cell Dev Biol **13,** 203-229.

Farrow, N.A., Zhang, O., Forman-Kay, J.D., and Kay, L.E. (1995). Comparison of

the backbone dynamics of a folded and an unfolded SH3 domain existing in

equilibrium in aqueous buffer. Biochemistry **34,** 868-878.

Fejzo, J., Lepre, C.A., Peng, J.W., Bemis, G.W., Ajay, Murcko, M.A., and Moore,

J.M. (1999). The SHAPES strategy: an NMR-based approach for lead

generation in drug discovery. Chem Biol **6,** 755-769.

Gekakis, N., Staknis, D., Nguyen, H.B., Davis, F.C., Wilsbacher, L.D., King, D.P.,

Takahashi, J.S., and Weitz, C.J. (1998). Role of the CLOCK protein in the

mammalian circadian mechanism. Science **280,** 1564-1569.

Goddard, T.D., and Kneller, D.G. (1999). SPARKY 3 (University of California, San

Francisco).

Gong, W., Hao, B., Mansy, S.S., Gonzalez, G., Gilles-Gonzalez, M.A., and Chan,

M.K. (1998). Structure of a biological oxygen sensor: a new mechanism for

heme-driven signal transduction. Proc Natl Acad Sci U S A **95,** 15177-15182.

Grimm, R., Eckerskorn, C., Lottspeich, F., Zenger, C., and Rudiger, W. (1988).
Sequence analysis of proteolytic fragments of 124-kilodaloton phytochrome
from etiolated *Avena sativa* L. : Conclusions on the conformation of the native
protein. Planta **174,** 396-401.

Grzesiek, S., Ikura, M., Clore, G.M., Gronenborn, A.M., and Bax, A. (1992). A 3D
triple-resonance NMR technique for qualitative measurement of carbonyl-Hβ J
coplings in isotopically enriched protein. J Magn Reson **96,** 215-221.

Hagel, L., Ostberg, M., and Andersson, T. (1996). Apparent pore size distributions of
chromatography media. J Chromatogr A **743,** 33-42.

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

Herrmann, T., Guntert, P., and Wuthrich, K. (2002). Protein NMR Structure
Determination with Automated NOE Assignment Using the New Software
CANDID and the Torsion Angle Dynamics Algorithm DYANA. J Mol Biol **319,**
209-227.

Hidaka, Y., Park, H., and Inouye, M. (1997). Demonstration of dimer formation of the
cytoplasmic domain of a transmembrane osmosensor protein, EnvZ, of
Escherichia coli using Ni-histidine tag affinity chromatography. FEBS Lett **400,**

238-242.

Holm, L., and Sander, C. (1996). Mapping the protein universe. Science **273,** 595-603.

Hu, J.S., Grzesiek, S., and Bax, A. (1997). Two-Dimentional NMR Methods for

Determining χ_1 Angles of Aromatic Residues in Proteins from Three-Bond $J_{C'C\gamma}$

and $J_{NC\gamma}$ Coupling. J. Am. Chem. Soc. **119,** 1803-1804.

Huang, Z.J., Edery, I., and Rosbash, M. (1993). PAS is a dimerization domain

common to Drosophila period and several transcription factors. Nature **364,**

259-262.

Jeon, Y.H., Yamazaki, T., Otomo, T., Ishihama, A., and Kyogoku, Y. (1997). Flexible

linker in the RNA polymerase alpha subunit facilitates the independent motion

of the C-terminal activator contact domain. J Mol Biol **267,** 953-962.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific

scoring matrices. J Mol Biol **292,** 195-202.

Kapust, R.B., and Waugh, D.S. (1999). Escherichia coli maltose-binding protein is

uncommonly effective at promoting the solubility of polypeptides to which it is

fused. Protein Sci **8,** 1668-1674.

Keifer, P.A. (1997). High-resolution NMR techniques for solid-phase synthesis and

combinatorial chemistry. Drug Discov Today **2,** 468-478.

Keifer, P.A., Baltusis, L., Rice, D.M., Tymiak, A.A., and Shoolery, J.N. (1996). A

comparison of NMR spectra obtained for solid-phase-synthesis resins using

conventional high-resolution, magic-angle-spinning, and high-resolution

magic-angle-spinning probes. J Magn Reson **119,** 65-75.

Kendrick, R.E., and Kronenberg, G.H.M. (1994). Photomorphogenesis in Plants, 2nd

edn. (Dordrecht, The Netherlands: Kluwer).

Kim, J.I., Shen, Y., Han, Y.J., Park, J.E., Kirchenbauer, D., Soh, M.S., Nagy, F.,

Schafer, E., and Song, P.S. (2004). Phytochrome Phosphorylation Modulates

Light Signaling by Influencing the Protein-Protein Interaction. Plant Cell **16,**

2629-2640.

Kircher, S., Gil, P., Kozma-Bognar, L., Fejes, E., Speth, V., Husselstein-Muller, T.,

Bauer, D., Adam, E., Schafer, E., and Nagy, F. (2002). Nucleocytoplasmic

partitioning of the plant photoreceptors phytochrome A, B, C, D, and E is

regulated differentially by light and exhibits a diurnal rhythm. Plant Cell **14,**

1541-1555.

Kobayashi, T., Mishima, M., Akagi, K., Sakai, N., Katoh, E., Takano, M., Yamazaki,

T., and Kojima, C. (2005). 1H, 15N and 13C backbone and side-chain

assignments of the rice phytochrome B PAS1 domain and backbone assignments

of the PAS1-PAS2 domain. J Biomol NMR **31,** 269-270.

Koradi, R., Billeter, M., and Wuthrich, K. (1996). MOLMOL: a program for display

and analysis of macromolecular structures. J Mol Graph **14,** 51-55, 29-32.

Krishna, N.R., and Berliner, L.J. (1998). Modern Techniques in Protein NMR. (New

York: Kluwer Academic).

Kurokawa, H., Lee, D.S., Watanabe, M., Sagami, I., Mikami, B., Raman, C.S., and Shimizu, T. (2004). A redox-controlled molecular switch revealed by the crystal structure of a bacterial heme PAS sensor. J Biol Chem **279,** 20186-20193.

Lapko, V.N., Jiang, X.Y., Smith, D.L., and Song, P.S. (1999). Mass spectrometric characterization of oat phytochrome A: isoforms and posttranslational modifications. Protein Sci **8,** 1032-1044.

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

Lei, C., Shin, Y., Liu, J., and Ackerman, E.J. (2002). Entrapping enzyme in a functionalized nanoporous support. J Am Chem Soc **124,** 11242-11243.

Lesley, S.A., Kuhn, P., Godzik, A., Deacon, A.M., Mathews, I., Kreusch, A., Spraggon, G., Klock, H.E., McMullan, D., Shin, T., Vincent, J., Robb, A., Brinen, L.S., Miller, M.D., McPhillips, T.M., Miller, M.A., Scheibe, D., Canaves, J.M., Guda, C., Jaroszewski, L., Selby, T.L., Elsliger, M.A., Wooley, J., Taylor, S.S., Hodgson, K.O., Wilson, I.A., Schultz, P.G., and Stevens, R.C. (2002). Structural genomics of the Thermotoga maritima proteome implemented in a high-throughput structure determination pipeline. Proc Natl Acad Sci U S A **99,** 11664-11669.

Lippens, G., Bourdonneau, M., Dhalluin, C., Warrass, R., Richert, T., Seetharaman, C., Boutillon, C., and Piotto, M. (1999). Study of Compounds Attached to Solid Suppourts Using High Resolution Magic Angle Spinning NMR. Current Organic Chemistry **3,** 147-169.

Makrides, S.C. (1996). Strategies for achieving high-level expression of genes in Escherichia coli. Microbiol Rev **60,** 512-538.

Mathews, S., and Sharrock, R.A. (1997). Phytochrome gene diversity. Plant, Cell Environ **20,** 666-671.

Matsushita, T., Mochizuki, N., and Nagatani, A. (2003). Dimers of the N-terminal domain of phytochrome B are functional in the nucleus. Nature **424,** 571-574.

Mazure, M., Calas, B., Cave, A., and Parello, J. (1986). Solid phase peptide synthesis. Structural characterization by proton NMR of a peptide immobilized on a polyacrylic resin. CR Acad Sci **303,** 553-556.

Middelberg, A.P. (2002). Preparative protein refolding. Trends Biotechnol **20,** 437-443.

Miyatake, H., Mukai, M., Park, S.Y., Adachi, S., Tamura, K., Nakamura, H., Nakamura, K., Tsuchiya, T., Iizuka, T., and Shiro, Y. (2000). Sensory mechanism of oxygen sensor FixL from Rhizobium meliloti: crystallographic, mutagenesis and resonance Raman spectroscopic studies. J Mol Biol **301,** 415-431.

Montgomery, B.L., and Lagarias, J.C. (2002). Phytochrome ancestry: sensors of

bilins and light. Trends Plant Sci **7,** 357-366.

Morais Cabral, J.H., Lee, A., Cohen, S.L., Chait, B.T., Li, M., and Mackinnon, R. (1998). Crystal structure and functional analysis of the HERG potassium channel N terminus: a eukaryotic PAS domain. Cell **95,** 649-655.

Nakanishi, T., Miyazawa, M., Sakakura, M., Terasawa, H., Takahashi, H., and Shimada, I. (2002). Determination of the interface of a large protein complex by transferred cross-saturation measurements. J Mol Biol **318,** 245-249.

Nambu, J.R., Lewis, J.O., Wharton, K.A., Jr., and Crews, S.T. (1991). The Drosophila single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. Cell **67,** 1157-1167.

Neff, M.M., Fankhauser, C., and Chory, J. (2000). Light: an indicator of time and place. Genes Dev **14,** 257-271.

Ni, M., Tepperman, J.M., and Quail, P.H. (1998). PIF3, a phytochrome-interacting factor necessary for normal photoinduced signal transduction, is a novel basic helix-loop-helix protein. Cell **95,** 657-667.

Ni, M., Tepperman, J.M., and Quail, P.H. (1999). Binding of phytochrome B to its nuclear signalling partner PIF3 is reversibly induced by light. Nature **400,** 781-784.

Nishida, N., Sumikawa, H., Sakakura, M., Shimba, N., Takahashi, H., Terasawa, H., Suzuki, E.I., and Shimada, I. (2003). Collagen-binding mode of vWF-A3

domain determined by a transferred cross-saturation experiment. Nat Struct Biol **10,** 53-58.

Oka, Y., Matsushita, T., Mochizuki, N., Suzuki, T., Tokutomi, S., and Nagatani, A. (2004). Functional analysis of a 450-amino acid N-terminal fragment of phytochrome B in Arabidopsis. Plant Cell **16,** 2104-2116.

Pearson, W.R., and Lipman, D.J. (1988). Improved tools for biological sequence comparison. Proc Natl Acad Sci U S A **85,** 2444-2448.

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

Pellecchia, M., Sem, D.S., and Wuthrich, K. (2002a). NMR in drug discovery. Nat Rev Drug Discov **1,** 211-219.

Pellecchia, M., Meininger, D., Dong, Q., Chang, E., Jack, R., and Sem, D.S. (2002b). NMR-based structural characterization of large protein-ligand interactions. J Biomol NMR **22,** 165-173.

Pellequer, J.L., Wager-Smith, K.A., Kay, S.A., and Getzoff, E.D. (1998). Photoactive yellow protein: a structural prototype for the three-dimensional fold of the PAS domain superfamily. Proc Natl Acad Sci U S A **95,** 5884-5890.

Pongratz, I., Antonsson, C., Whitelaw, M.L., and Poellinger, L. (1998). Role of the PAS domain in regulation of dimerization and DNA binding specificity of the

dioxin receptor. Mol Cell Biol **18,** 4079-4088.

Quail, P. (1997). An emerging molecular map of the phytochromes. Plant Cell Envioron. **20,** 657-665.

Quail, P.H. (2002). Phytochrome photosensory signalling networks. Nat Rev Mol Cell Biol **3,** 85-93.

Quail, P.H., Boylan, M.T., Parks, B.M., Short, T.W., Xu, Y., and Wagner, D. (1995). Phytochromes: photosensory perception and signal transduction. Science **268,** 675-680.

Reid, D.G. (1997). Protein NMR Techniques. (New Jersey: Human Press).

Rogl, H., Kosemund, K., Kuhlbrandt, W., and Collinson, I. (1998). Refolding of Escherichia coli produced membrane protein inclusion bodies immobilised by nickel chelating chromatography. FEBS Lett **432,** 21-26.

Romanowski, M., and Song, P.S. (1992). Structural domains of phytochrome deduced from homologies in amino acid sequences. J Protein Chem **11,** 139-155.

Rowlands, J.C., and Gustafsson, J.A. (1997). Aryl hydrocarbon receptor-mediated signal transduction. Crit Rev Toxicol **27,** 109-134.

Sakamoto, T., Tanaka, T., Ito, Y., Rajesh, S., Iwamoto-Sugai, M., Kodera, Y., Tsuchida, N., Shibata, T., and Kohno, T. (1999). An NMR analysis of ubiquitin recognition by yeast ubiquitin hydrolase: evidence for novel substrate recognition by a cysteine protease. Biochemistry **38,** 11634-11642.

Sali, A., and Blundell, T.L. (1993). Comparative Protein Modelling by Satisfaction of Spatial Restraints. J Mol Biol **234,** 779-815.

Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, Progress in Nuclear Magnetic Resonance Spectroscopy. Progress in Nuclear Magnetic Resonance Spectroscopy **34,** 93-158.

Savchenko, A., Yee, A., Khachatryan, A., Skarina, T., Evdokimova, E., Pavlova, M., Semesi, A., Northey, J., Beasley, S., Lan, N., Das, R., Gerstein, M., Arrowmith, C.H., and Edwards, A.M. (2003). Strategies for structural proteomics of prokaryotes: Quantifying the advantages of studying orthologous proteins and of using both NMR and X-ray crystallography approaches. Proteins **50,** 392-399.

Schibler, U. (1998). Circadian rhythms. New cogwheels in the clockworks. Nature **393,** 620-621.

Schneider-Poetsch, H.A., Braun, B., Marx, S., and Schaumburg, A. (1991). Phytochromes and bacterial sensor proteins are related by structural and functional homologies. Hypothesis on phytochrome-mediated signal-transduction. FEBS Lett **281,** 245-249.

Schubert, M., Labudde, D., Oschkinat, H., and Schmieder, P. (2002). A software tool

for the prediction of Xaa-Pro peptide bond conformations in proteins based on
13C chemical shift statistics. J Biomol NMR **24,** 149-154.

Shapiro, M.J., and Gounarides, J.S. (1999). NMR methods utilized in combinatorial
chemistry research. Prog Nucl Magn Reson Spectros **35,** 153-200.

Sharrock, R.A., and Clack, T. (2004). Heterodimerization of type II phytochromes in
Arabidopsis. Proc Natl Acad Sci U S A **101,** 11500-11505.

Shih, Y.P., Kung, W.M., Chen, J.C., Yeh, C.H., Wang, A.H., and Wang, T.F. (2002).
High-throughput screening of soluble recombinant proteins. Protein Sci **11,**
1714-1719.

Shuker, S.B., Hajduk, P.J., Meadows, R.P., and Fesik, S.W. (1996). Discovering
high-affinity ligands for proteins: SAR by NMR. Science **274,** 1531-1534.

Smith, H. (2000). Phytochromes and light signal perception by plants--an emerging
synthesis. Nature **407,** 585-591.

Smyth, D.R., Mrozkiewicz, M.K., McGrath, W.J., Listwan, P., and Kobe, B. (2003).
Crystal structures of fusion proteins with large-affinity tags. Protein Sci **12,**
1313-1322.

Stempfer, G., Holl-Neugebauer, B., and Rudolph, R. (1996a). Improved refolding of
an immobilized fusion protein. Nat Biotechnol **14,** 329-334.

Stempfer, G., Holl-Neugebauer, B., Kopetzki, E., and Rudolph, R. (1996b). A fusion
protein designed for noncovalent immobilization: stability, enzymatic activity,

and use in an enzyme reactor. Nat Biotechnol **14,** 481-484.

Stock, A.M., Robinson, V.L., and Goudreau, P.N. (2000). Two-component signal

transduction. Annu Rev Biochem **69,** 183-215.

Stockman, B.J., Euvrard, A., and Scahill, T.A. (1993). Heteronuclear

three-dimensional NMR spectroscopy of a partially denatured protein: the

A-state of human ubiquitin. J Biomol NMR **3,** 285-296.

Takahashi, H., Nakanishi, T., Kami, K., Arata, Y., and Shimada, I. (2000). A novel

NMR method for determining the interfaces of large protein-protein complexes.

Nat Struct Biol **7,** 220-223.

Taylor, B.L., and Zhulin, I.B. (1999). PAS domains: internal sensors of oxygen, redox

potential, and light. Microbiol Mol Biol Rev **63,** 479-506.

Taylor, B.L., Rebbapragada, A., and Johnson, M.S. (2001). The FAD-PAS domain as

a sensor for behavioral responses in Escherichia coli. Antioxid Redox Signal **3,**

867-879.

Tomomori, C., Tanaka, T., Dutta, R., Park, H., Saha, S.K., Zhu, Y., Ishima, R., Liu,

D., Tong, K.I., Kurokawa, H., Qian, H., Inouye, M., and Ikura, M. (1999).

Solution structure of the homodimeric core domain of Escherichia coli histidine

kinase EnvZ. Nat Struct Biol **6,** 729-734.

Vinogradova, O., Velyvis, A., Velyviene, A., Hu, B., Haas, T., Plow, E., and Qin, J.

(2002). A structural mechanism of integrin alpha(IIb)beta(3) "inside-out"

activation as regulated by its cytoplasmic face. Cell **110,** 587-597.

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

Wagner, D., and Quail, P.H. (1995). Mutational analysis of phytochrome B identifies a small COOH-terminal-domain region critical for regulatory activity. Proc Natl Acad Sci U S A **92,** 8596-8600.

Wagner, D., Koloszvari, M., and Quail, P.H. (1996). Two Small Spatially Distinct Regions of Phytochrome B Are Required for Efficient Signaling Rates. Plant Cell **8,** 859-871.

Waldo, G.S. (2003). Genetic screens and directed evolution for protein solubility. Curr Opin Chem Biol **7,** 33-38.

Yee, A., Pardee, K., Christendat, D., Savchenko, A., Edwards, A.M., and Arrowsmith, C.H. (2003). Structural proteomics: toward high-throughput structural biology as a tool in functional genomics. Acc Chem Res **36,** 183-189.

Yee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A., Le, B., Ramelot, T., Lee, G.M., Bhattacharyya, S., Gutierrez, P., Denisov, A., Lee, C.H., Cort, J.R., Kozlov, G., Liao, J., Finak, G., Chen, L., Wishart, D., Lee, W., McIntosh, L.P., Gehring, K., Kennedy, M.A., Edwards, A.M., and Arrowsmith, C.H. (2002). An NMR approach to structural proteomics. Proc Natl Acad Sci U S A **99,** 1825-1830.

Yildiz, O., Doi, M., Yujnovsky, I., Cardone, L., Berndt, A., Hennig, S., Schulze, S., Urbanke, C., Sassone-Corsi, P., and Wolf, E. (2005). Crystal structure and interactions of the PAS repeat region of the Drosophila clock protein PERIOD. Mol Cell **17,** 69-82.

Yokoyama, S. (2003). Protein expression systems for structural genomics and proteomics. Curr Opin Chem Biol **7,** 39-43.

Zahn, R., von Schroetter, C., and Wuthrich, K. (1997). Human prion proteins expressed in Escherichia coli and purified by high-affinity column refolding. FEBS Lett **417,** 400-404.

Zhan, Y., Song, X., and Zhou, G.W. (2001). Structural analysis of regulatory protein domains using GST-fusion proteins. Gene **281,** 1-9.

Zhou, P., Lugovskoy, A.A., and Wagner, G. (2001a). A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. J Biomol NMR **20,** 11-14.

Zhou, P., Lugovskoy, A.A., McCarty, J.S., Li, P., and Wagner, G. (2001b). Solution structure of DFF40 and DFF45 N-terminal domain complex and mutual chaperone activity of DFF40 and DFF45. Proc Natl Acad Sci U S A **98,** 6051-6055.