Structure-based studies on LOV photoreceptor proteins

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

M. Sc. Vladimir Arinkin

aus Almaty, Kasachstan

Jülich, Juni 2017

aus dem Institut of Complex Systems, Strukturbiochemie (ICS-6)

der Forschungszentrum Jülich

und

Institut für Physikalische Biologie

der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der

Mathematisch-Naturwissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

Referent: PD Dr. Renu Batra-Safferling

Korreferent: Prof. Dr. Dieter Willbold

Externer Korreferent: Prof. Dr. Joachim Heberle

Tag der mündlichen Prüfung: 30.06.2017

Erklärung

Ich versichere an Eides Statt, dass die Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Die Dissertation in der vorgelegten oder in ähnlicher Form wurde noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Jülich, den 20.03.2017

Vladimir Arinkin

List of publications and manuscripts in preparation

Arinkin, V., Granzin, J., Röllen, K., Krauss, U., Jaeger, K.-E., Willbold, D. and Batra-Safferling, R. (2017) 'Structure of a LOV protein in apo-state and implications for construction of LOV-based optical tools', *Scientific Reports*, 7, p. 42971. doi: 10.1038/srep42971.

Arinkin, V., Granzin, J., Krauss, U., Jaeger, K.-E., Willbold, D. and Batra-Safferling, R. (2016) 'Crystal structure of a fast recovering photoreceptor protein PpSB2-LOV: elucidation of kinetic properties and signal transduction mechanism', *Acta Crystallographica Section A* (Conference Abstract), 72(a1), pp. s255–s255. doi: 10.1107/S2053273316096121.

Röllen, K., Granzin, J., Panwalkar, V., **Arinkin, V.**, Rani, R., Hartmann, R., Krauss, U., Jaeger, K.-E., Willbold, D. and Batra-Safferling, R. (2016) 'Signaling States of a Short Blue-Light Photoreceptor Protein PpSB1-LOV Revealed from Crystal Structures and Solution NMR Spectroscopy', *Journal of Molecular Biology*, 428(19), pp. 3721–3736. doi: 10.1016/j.jmb.2016.05.027.

Arinkin, V., Granzin, J., Krauss, U., Jaeger, K.-E., Willbold, D. and Batra-Safferling, R. (2017) 'Light-state crystal structure of extremely slow recovering short LOV protein SBW25-LOV from *Pseudomonas fluorescence*'. *Manuscript in preparation.*

Arinkin, V., Granzin, J., Krauss, U., Jaeger, K.-E., Willbold, D. and Batra-Safferling, R. (2017) 'Crystal structure of a fast recovering short LOV protein PpSB2-LOV from *Pseudomonas putida*'. *Manuscript in preparation.*

Arinkin, V., Granzin, J., Krauss, U., Jaeger, K.-E., Willbold, D. and Batra-Safferling, R. (2017) 'Crystal structure of a chimeric LOV Histidine kinase SB2F1 reveals details on its light-dependent activity'. *Manuscript in preparation.*

Poster presentations

Röllen, K., Granzin, J., **Arinkin, V.**, Willbold, D. and Batra-Safferling, R. (2015) 'Signal propagation in the blue-light-sensitive protein PpSB1-LOV – a comparison between crystal structures in light and dark states', GRC Proteins, Holderness, New Hampshire, USA.

Arinkin, V., Granzin, J., Röllen, K. and Batra-Safferling, R. (2015) 'Crystal structure of a fast recovering full-length LOV protein PpSB2-LOV from Pseudomonas putida', 1st CSSB International Symposium – Systems in Infection Biology From Molecules to Organisms, Hamburg, Germany.

Röllen, K., Granzin, J., Panwalkar, V., **Arinkin, V.**, Krauss, U., Jaeger, K.-E., Willbold, D. and Batra-Safferling, R. (2016) 'Signalling states of a short blue-light photoreceptor protein PpSB1-LOV', GRC Photosensory Receptors & Signal Transduction, Galveston, Texas, USA.

Arinkin, V., Granzin, J., Krauss, U., Jaeger, K.-E., Willbold, D. and Batra-Safferling, R. (2016) 'Crystal structure of a fast recovering photoreceptor protein PpSB2-LOV: elucidation of kinetic properties and signal transduction mechanism', The 30th European Crystallographic Meeting, Basel, Switzerland.

Summary

Proteins containing blue-light absorbing light-oxygen-voltage (LOV) domains participate in a multitude of cellular responses in both eukaryotes and prokaryotes. LOV domains bind non-covalently ubiquitous flavins (e.g. flavin mononucleotide), which are responsible for the blue-light absorption. The primary photochemistry of blue-light absorption contains several intermediate states that lead to the formation of covalent adduct between flavin and highly conserved cysteine residue. The state with formed covalent adduct, referred as the light state, is stable on the wide range of time scales, ranging from seconds to hours. The light state, spontaneously, recovers back to the initial ground state (the dark state) with the characteristic time constant (the adduct lifetime).

Although LOV domains usually form a part of multi-domain protein that contains an effector domain, a considerable number of LOV domains (short LOV proteins) were identified as single proteins, which lacks any additional domains. The physiological role of only a few short LOV proteins, from fungi and bacteria, could be assigned.

In this thesis, the results are presented on short LOV proteins (PpSB2-LOV, SBW25-LOV, Pf5-LOV and W619_1-LOV) from plant and soil colonizing bacteria *Pseudomonas putida* and *Pseudomonas fluorescens*. They show considerable variation of the adduct lifetime, from minutes to even days. Their crystal structures were obtained in the dark and light states. Analysis of the structures revealed the correlation between the adduct lifetime and the cavity volume of the chromophore binding pocket.

Molecular dynamic simulations supported the hypothesis that decrease of the solvent accessibility to the flavin chromophore decelerates the dark recovery. Structural elements, which surround the chromophore binding pocket, were found to be more flexible in proteins with the short adduct lifetime and more rigid in proteins with the long adduct lifetime.

Short LOV proteins of this study contain non-canonical N- and C-terminal extensions, beside the conserved core domain. Such extensions are similar to the typical linker between domains of sensory multi-domain proteins. Often N- and C-terminal extensions form α-helix, which is generally well suitable for mechanical signal transduction. This makes short LOV proteins a prototype system to study not only the signal propagation within LOV domains but also the signal transduction from the LOV domain to the potential effector domain.

Analysis of crystal structures from this work together with previously determined crystal structures of homologous protein PpSB1-LOV, in the dark and light states, revealed the primary structural changes induced by light. The primary structural changes originate

from the formation of covalent adduct between cysteine and flavin-C4a atom, and from the protonation of the flavin-N5 atom. Both events results in the conformational changes of Leu97 and Gln116 residues, which induce further changes in the β-scaffold of the core domain and in the N-terminal α-helix. Altered interactions in the β-scaffold and in the N-terminal α-helix, which largely participate in the dimer interface, result in the rotation of monomers relative to each other. This rotation, thereafter, result in the movement of the C-terminal α-helices, which conceivably can alter the conformation or dynamics of a fused effector domain in a multi-domain protein.

The crystal structure of W619_1-LOV determined in this work is the first structure of LOV domain in the apo form, free of any bound flavin. The structure of W619_1-LOV in the apo form showed changes of the chromophore binding pocket, compared to the FMN-bound structure of homologous PpSB1-LOV protein. Major changes happened in the two α-helices connected by a long loop, which partially form the chromophore binding pocket.

Furthermore, the apo form of W619_1-LOV was stable in solution and was loaded *in vitro* with natural flavins: flavin mononucleotide (FMN), riboflavin (RBF), flavin adenine dinucleotide (FAD) and lumichrome (LC); and with modified flavins: 7-bromo-riboflavin (7-Br-RF) and 8-chloro-riboflavin (8-Cl-RF). Binding of different flavins to W619_1-LOV significantly altered their biophysical and biochemical properties, such as UV-Vis absorbance and fluorescence spectra. Among them, the LC bound to W619_1-LOV has demonstrated remarkable blue-shift of the absorption maximum of 26 nm when compared to the FMN-bound maximum. The LC bound to W619_1-LOV also showed relatively high fluorescence quantum yield of 0.4.

The adduct lifetime considerably varied among different flavins, indicating the importance of specific chromophore-to-protein interactions on the kinetics of dark recovery. For instance, the RBF-bound W619_1-LOV has the adduct lifetime 307 times faster than those of the FMN-bound (1.5 min vs 460 min at 20 °C). Another notable property was found that 8-Cl-RF, forms a covalent bond between the 8-Cl atom and the conserved Cys53 in the binding pocket. The resulting LOV protein provides unique opportunity for its application as an optical tool.

LOV containing histidine kinases (LOV-HK) are the most abundant two component signaling system in bacteria among the LOV containing multi-domain proteins. The SB1F1 and SB2F1 are two engineered LOV-HK, which were previously constructed in the similar fashion as known from the literature YF1 protein. For this, the oxygen-sensitive PAS domain in the histidine kinase FixL from *Bradyrhizobium japonicum* was replaced to the LOV domains of PpSB2-LOV or PpSB1-LOV proteins from *Pseudomonas putida*. The SB1F1 protein crystals resulted in diffraction data that failed to be sufficiently phased. However, several crystal structures of dimeric SB2F1 in the different states and with different nucleotides (ATP,

ADP, AMP-PNP, ATP-γ-S) are presented here. Additionally, the small-angle X-ray scattering (SAXS) data enabled to relate the states observed in solution to the different crystal structures of SB2F1. Analysis of SB2F1 data and comparison to previously determined YF1 structure, suggested large-scale quaternary structural changes upon blue-light illumination. The mechanism of signal transduction from the LOV domain to the histidine kinase is in good agreement with that proposed for the short LOV proteins in this work.

In brief, after formation of the light state, LOV domains rotate relative to each other, which results in the rearrangement of long and helical DHp domain (dimerization and histidine phosphotransfer) of HK. The rearrangement of the coiled-coil-like DHp domains, alters the spatial arrangement of the CA domains (catalytic and ATP binding). Different spatial arrangement of CA domains, in turn, can result in the different autophosprylation rates in the light and dark states. Additional experiments in future are required to prove this hypothesis and elucidate additional details of this mechanism.

Zusammenfassung

Proteine mit Blaulicht-absorbierenden Light-Oxygen-Voltage (LOV) Domänen sind an einer Vielzahl zellulärer Prozesse sowohl in Eukaryonten als auch in Prokaryonten involviert. LOV-Domänen binden nicht kovalent ubiquitäre Flavine (z.B. Flavinmononukleotid), die für die Blaulichtabsorption verantwortlich sind. Die primäre Photochemie der Blaulichtabsorption besteht aus mehreren Übergangszuständen, die u.a. zur Bildung einer kovalenten Bindung zwischen dem Flavin und einem konserviertem Cystein führen. Dieser sogenannte Lichtzustand kann von Sekunden bis hin zu Stunden stabil sein und geht nach einer charakteristischen Adduktlebensdauer (mit spezifischer Zeitkonstante) spontan wieder in den Dunkelzustand (dem Grundzustand) über.

LOV-Domänen können einzeln als sog. kurze LOV-Proteine aber auch Bestandteil von Mehrdomänen-Proteine sein, letztere werden durch zusätzliche Effektor-Domäne ergänzt. Die physiologische Rolle nur weniger kurzer LOV-Proteine etwa aus Pilzen und Bakterien sind bekannt.

In dieser Arbeit werden u.a. Ergebnisse kurzer LOV-Proteine (PpSB2-LOV, SBW25-LOV, Pf5-LOV und W619_1-LOV) aus Pflanzen- und Bodenkolonisierungsbakterien (*Pseudomonas putida, Pseudomonas fluorescens*) präsentiert. Sie zeigen eine beträchtliche Variation in der Adduktlebensdauer. Mittels Röntgenstrukturanalyse konnten die Kristallstrukturen von verschiedenen Lichtanregungszuständen aufgeklärt werden. Die Strukturanalyse ermöglichte es eine Korrelation zwischen der Adduktlebensdauer und dem Volumen der Chromophor-Bindungstasche nachzuweisen.

Molekulardynamische Simulationen unterstützen diese Hypothese, dass eine Verringerung der Lösungsmittelzugänglichkeit zum Flavin-Chromophor die Rückkehr zum Dunkelzustand verlangsamt. Strukturelemente, die die Chromophor-Bindungstasche umgeben, sind in Proteinen mit einer schnellen Adduktlebensdauer flexibler und in Proteinen mit einer langsamen Adduktlebensdauer rigider.

Die hier untersuchten kurzen LOV-Proteine enthalten nicht-kanonische N- und C-terminale Strukturelemente zusätzlich zu der konservierten Kerndomäne. Solche Erweiterungen ähneln den typischen Linker-Domänen in den sensorischen Mehrdomänen-Proteinen. Oft bestehen diese N- und C-terminale Erweiterungen aus α-Helices, die als Teil der mechanischen Signaltransduktion dienen können. Dies macht kurze LOV-Proteine zu einem Referenz-System, für Untersuchungen der Signalausbreitung innerhalb der LOV-Domänen, sowie der Signalweiterleitung zur potentiellen Effektor-Domäne.

Der Vergleich der Kristallstrukturen aus dieser Arbeit mit den Kristallstrukturen des homologen Proteins PpSB1-LOV (s. Literaturdaten), in ihren Dunkel- und Lichtzuständen, zeigen Struktur- und Konformationsänderungen die durch die Belichtung induziert werden. Die ersten Veränderungen ergeben sich aus der Bildung der kovalenten Bindung zwischen dem konservierten Cystein und dem C4a-Atoms des Flavins und aus der Protonierung des N5-Atoms. Beide Ereignisse führen zu Konformationsänderungen der Leu97- und Gln116-Seitenketten, die weitere Veränderungen in den β-Faltblättern der Kerndomäne sowie in der N-terminalen α-Helix nach sich ziehen. Veränderte Wechselwirkungen in den β-Faltblättern und der N-terminalen α-Helix, die weitgehend an der Dimer-Grenzfläche beteiligt sind, führen zu einer Drehung der Monomere relativ zueinander. Diese Rotation initiiert eine Bewegung der C-terminalen α-Helices, die zum Beispiel die Konformation bzw. Dynamik einer fusionierten Effektor-Domäne in einem Mehrdomänen-Proteinen verändern könnte.

Die in dieser Arbeit untersuchte Kristallstruktur von W619_1-LOV zeigt die LOV-Domäne erstmals in der Apo-Form, also Flavin-frei. Es sind Änderungen in der Chromophor-Bindungstasche im Vergleich zur der Flavinmononukleotid-gebundenen Struktur des homologen PpSB1-LOV-Proteins zu beobachten. Im Wesentlichen betrifft dies zwei α-Helices, die teilweise die Chromophor-Bindungstasche ausbilden.

Die Apo-Form von W619_1-LOV ist in Lösung stabil und wurde in vitro mit natürlichen Flavinen beladen: Flavinmononukleotid, Riboflavin, Flavin-Adenindinukleotid und Lumichrom; sowie mit folgenden modifizierten Flavinen: 7-Brom-Riboflavin und 8-Chlor-Riboflavin. Die Bindung der verschiedenen Flavine an W619_1-LOV verändert die biophysikalischen und biochemischen Eigenschaften, wie z.B. das jeweilige UV-Vis- und Fluoreszenzspektrum. Wenn z.B. Lumichrom gebunden ist, kommt es zu einer signifikanten Blauverschiebung, das Absorptionsmaximum ist um 26 nm geschiftet im Vergleich mit dem Absorptionsmaximum bei Bindung von Flavinmononukleotid. Das gebundene Lumichrom zeigte zudem eine erhöhte Fluoreszenzquantenausbeute.

Die Adduktlebensdauer variiert stark zwischen den verschiedenen Flavinen. Die spezifischen Chromophor-Protein-Wechselwirkungen haben Einfluss auf die Kinetik der Rückkehr zum Dunkelzustand. Zum Beispiel hat das Riboflavin-gebundene W619_1-LOV eine 307-fache kurzlebigere Adduktlebensdauer als der Flavinmononukleotid-gebunden W619_1-LOV (1,5 min verglichen mit 460 min bei 20 °C). Weiterhin ist bemerkenswert, dass 8-Chlor-Riboflavin eine kovalente Bindung zwischen dem Cl-Atom und dem konserviertem Cystein in der Bindungstasche ausbildet.

Die am häufigsten vorkommenden Zweikomponenten-Signaltransduktionssysteme in Bakterien sind Kombinationen von LOV-Domänen mit Histidinkinasen (HK). SB2F1 und SB1F1 sind zwei artifiziell konstruierte LOV-HK, die in ähnlicher Weise wie das aus der Literatur bekannte YF1 Protein konstruiert wurde. Hierzu wird die sauerstoffempfindliche

PAS-Domäne der Histidinkinase FixL aus *Bradyrhizobium japonicum* durch die LOV-Domäne des PpSB2-LOV-Proteins bzw. des PpSB1-LOV-Proteins aus *Pseudomonas putida* fusioniert. Verschiedene Varianten von SB2F1-Kristallstrukturen wurden in dieser Arbeit aufgeklärt.

Dies schließt unterschiedliche Lichtanregungszustände sowie verschiedene Nukleotidbeladungen (ATP, ADP, AMP-PNP, ATP-γ-S) mit ein. Zusätzliche Messungen von Kleinwinkelröntgenstreuungs-Daten in Lösung ermöglichten es, die unterschiedlichen Zustände, den verschiedenen Kristallstrukturen von SB2F1 zuzuordnen. Die Analyse der SB2F1-Daten und der anschließende Vergleich mit YF1 lässt Rückschlüsse auf die Gründe für die globalen Veränderungen in der Quartärstruktur bei Blaulicht-Einstrahlung zu. Die Signaltransduktion von der LOV-Domäne zur Histidinkinase ist vergleichbar mit dem in dieser Arbeit vorgeschlagene Mechanismus für kurze LOV-Proteine.

Zusammengefasst sind folgende Strukturänderungen zu beobachten, nach Bildung des Lichtzustandes drehen sich die LOV-Domänen relativ zueinander, was zur Umlagerung der langen helikalen DHp-Domänen (dimerization and histidine phosphotransfer) der Histidinkinase führt. Diese Verschiebung der Coiled-coil-artigen DHp-Domäne ändert dann die räumliche Anordnung der CA-Domänen (catalytic and ATP binding). Die neupositionierten CA-Domänen können dann wiederum zu verschiedenen Autophosphorylierungsraten der Dunkel- bzw. Lichtzustände führen. Zusätzliche Experimente werden erforderlich sein, um diese Hypothese zu überprüfen und auch weitere Details der Signaltransduktion aufzuklären.

Abbreviations

Abbreviation	Full name
7-Br-RF	7-bromo-riboflavin
8-Cl-RF	8-chloro-riboflavin
a.u.	Arbitrary Units
aa	Amino Acid
ADP	Adenosine 5'-diphosphate
AMP-PNP	Adenosine 5'-(β,γ-imido)triphosphate
ATP	Adenosine 5'-triphosphate
ATP-γ-S	Adenosine 5'-(γ-thio)triphosphate
AU	Absorbance Units
BLUF	Blue-Light sensors Using FAD domain
BSA	Bovine Serum Albumin
CA	Catalytic and ATP binding domain
CRYs	Cryptochromes
DHp	Dimerization and Histidine Phosphotransfer domain
DNA	Deoxyribonucleic acid
FAD	Flavin Adenine Dinucleotide (Riboflavin 5'-adenosine diphosphate)
FMN	Flavin Mononucleotide (Riboflavin 5'-monophosphate)
HK	Histidine Kinase
HPLC	High Performance Liquid Chromatography
HTH	Helix-Turn-Helix DNA binding domain
LC	Lumichrome
LOV	Light-Oxygen-Voltage domain
MD	Molecular Dynamics
nf	Nucleotide-free
OD	Optical Density
PAS	Per, Arndt, Sim domain
PYP	Photoactive Yellow Protein
QM	Quantum Mechanics
RBF	Riboflavin
RMSD	Root-Mean-Square Deviation
RMSF	Root-Mean-Square Fluctuation
RR	Response Regulator
SAXS	Small-Angle X-ray Scattering
SEC	Size-Exclusion Chromotography
SeMet	Selenomethionine
SHK	Sensory Histidine Kinase
TCEP	Tris(2-carboxyethyl)phosphin
τ_{rec}	Adduct Lifetime

Content

1 Introduction

1.1 Flavin-binding photoreceptors

Light is the ultimate and essential life giving energy source for all living organisms. Nature has adapted several mechanisms to utilize and respond to light. Generally, eight sensory photoreceptor families can be differentiated, each utilizing a distinct photochemistry. These are: phytochromes, xanthopsins, rhodopsins, cyanobacterial orange carotenoid proteins (OCP), plant UV photoreceptors such as UVR8, as well as the blue-light sensitive and flavin-binding family such as cryptochromes (CRY), BLUF (sensors of blue-light that uses FAD) and light-oxygen-voltage (LOV) domains.

The three major classes of flavoprotein light sensors LOV, BLUF and CRY participate in a multitude of biological responses to light that include phototropism, stomatal opening, the regulation of photosynthesis, stress responses and circadian rhythms (Möglich *et al.*, 2010; Herrou and Crosson, 2011; Zoltowski and Gardner, 2011). LOV, BLUF and CRY bind several natural flavins as a chromophore: flavin mononucleotide (FMN), flavin adenine dinucleotide (FAD) and riboflavin (RBF). Flavins are known for their rich chemistry and photochemistry, and are widely used by flavoproteins/flavoenzymes in different redox states.

BLUF domains

BLUF domains were simultaneously discovered during examination of the photosynthetic gene expression in purple bacteria and during investigation of mechanism of the photo-avoidance response in algae (Masuda, 2013). Various blue-light dependent cellular responses was shown to involve the BLUF domain coupled to different effector domains (Masuda, 2013). Among them, the BLUF domain from the AppA (activation of photopigment and *puc* expression A) protein was extensively studied (Figure 1A). Upon blue-light illumination, AppA undergoes a conformational change and dissociates from the photosynthesis gene-specific transcriptional repressor (Masuda, 2013).

Upon blue-light illumination, the flavin absorption spectra of the BLUF domains becomes ~10 nm red-shifted, indicating conversion from the dark state (ground state) to the light-adapted (activated) state that can persist from seconds to minutes. For instance in AppA, residues Tyr21, Gln63, Met106 and Trp104 form a tetrad of residues that undergo light-induced changes in their electronic state, hydrogen bonding and conformation, which lead to signal propagation. The exact chemical basis of these changes has been the subject of intense

study by time-resolved spectroscopy and combined quantum mechanical / molecular dynamics calculations and remains controversial (Zoltowski and Gardner, 2011; Masuda, 2013).

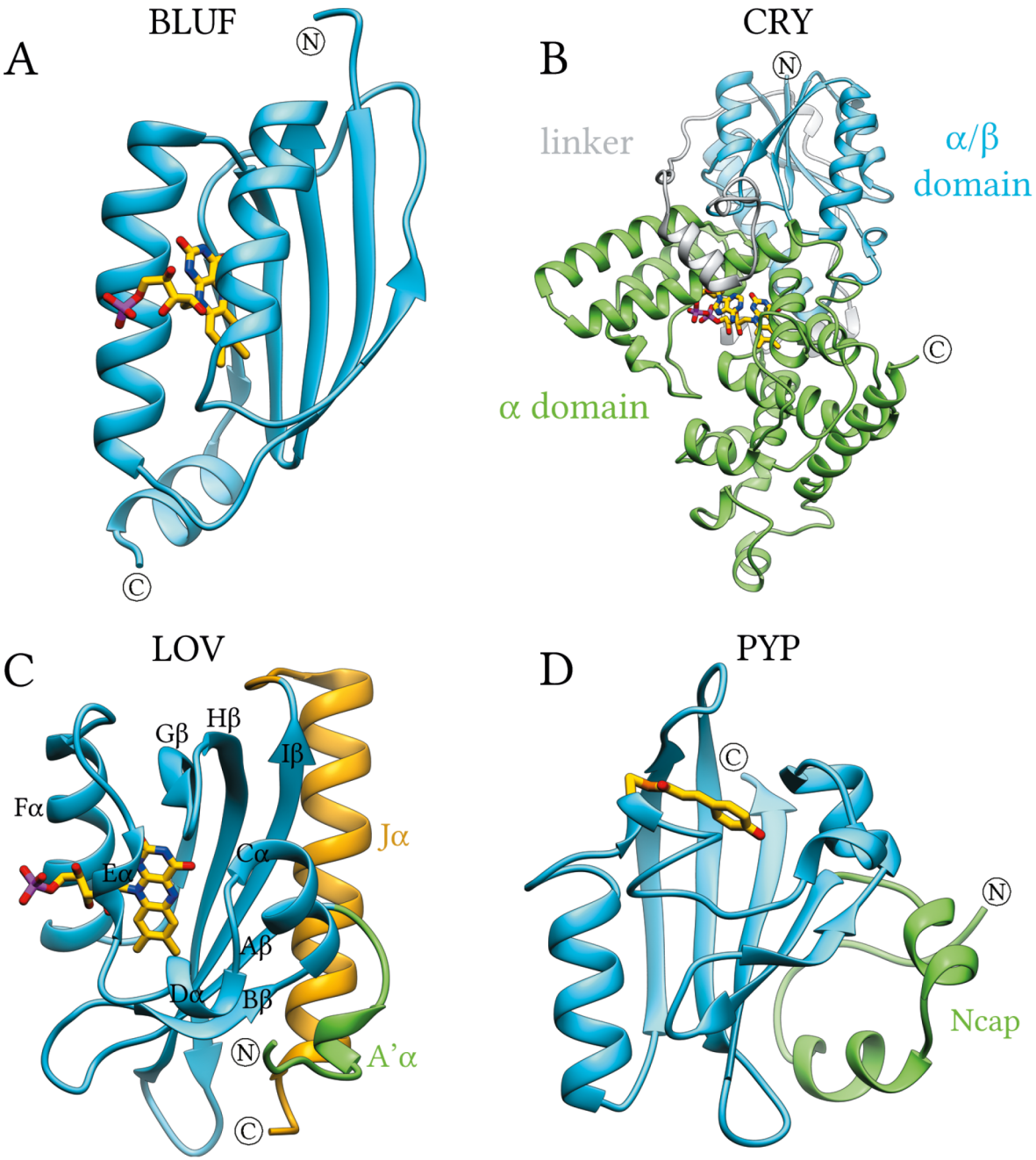


Figure 1. Representative crystal structures of three major flavin binding photoreceptors and prototypical Per-Arnt-Sim domain: BLUF, CRY, LOV and PYP. The structures of BLUF, CRY, LOV and PYP are represented as ribbons. (**A**) *Rhodobacter* BLUF domain AppA (PDB ID: 2IYG) is shown in blue with bound FMN as a stick model. (**B**) *Arabidopsis* CRY1 (PDB ID: 1U3C) with two domains (green – α domain, blue – α/β domain) connected by a linker (gray) is shown with FAD as a stick model. (**C**) *Avena sativa* AsLOV2 (PDB ID: 2V0U) has a typical PAS domain fold with a core domain (blue) flanked by an N-terminal extension A'α (green) (similar to Ncap) and a C-terminal extension Jα (gold). The FMN chromophore is shown as a stick model. (**D**) *Halorhodospira* PYP protein (PDB ID: 2PHY), a prototypic Per-Arnt-Sim (PAS) domain, where the core domain colored in blue, N-terminal extension (Ncap) in green and covalently attached 4-hydroxycinnamic acid chromophore is shown as a stick model.

CRY domains

CRY domains were initially proposed to be a putative blue-light receptor in plants responsible for growth and development (Ahmad and Cashmore, 1996). Later on, they were identified in a wide array of organisms including mammals and insects, functioning both as light-regulated and light-independent regulators of circadian function (Emery *et al.*, 1998; Sancar, 2004). CRY domains are closely related to photolyase family of proteins that repair DNA damaged by UV-light. They contain an N-terminal α/β domain connected by a long disordered linker to a C-terminal α domain (Figure 1B). The C-terminal helical tail binds to the grove along with FAD in similar fashion as DNA in photolyases. Light-induced changes of CRY domains, although still under debate, are nevertheless generally accepted to be related to the transitions between the oxidation states of flavin. The signaling mechanism is very likely to be based on light-induced C-terminal tail dissociation, which enables the binding of its partner protein, for instance, Timeless (TIM) as in *Drosophila melanogaster* (Zoltowski and Gardner, 2011).

LOV domains

LOV domains are a subset of the Per-Arnt-Sim (PAS) sensory domains superfamily (Taylor and Zhulin, 1999). They were originally identified in *Arabidopsis thaliana* as the domains responsible for blue-light absorption in plant photoreceptor proteins known as a phototropins (Huala *et al.*, 1997). Later on, LOV domains were also identified in algae, bacteria and fungi (Huang, Merkle and Beck, 2002; Losi *et al.*, 2002; Schwerdtfeger and Linden, 2003).

The first crystal structures of LOV domains revealed an α+β protein fold similar to the PAS domain, containing five antiparallel β-strands (Aβ, Bβ, Gβ, Hβ, Iβ) with three α-helices (Cα, Dα, Fα) and one 3_{10}-helix (Eα) (Figure 1C,D). α-helices packing against β-scaffold forms a cavity which can non-covalently bind flavins (e.g. FMN, RBF, FAD) as a chromophore. Beside these canonical secondary structure elements, LOV domains are often accompanied by flanking N- and C-terminal extensions. The extensions most often form a helix, which are denoted as A'α helix and Jα helix for N- and C-terminal one, respectively.

LOV domains usually form a part of multi-domain proteins that regulate a myriad of light responses such as phototropism, chloroplast movement, and stomatal opening in plants (Briggs and Huala, 1999; Christie, 2007); circadian rhythms and transcription regulation in fungi (Schwerdtfeger and Linden, 2003); stress response and cell adhesion in bacteria (Avila-Perez, Hellingwerf and Kort, 2006; Purcell *et al.*, 2010). Biological function is mediated by signal transduction from LOV domain to various fused effector domains, such as: histidine (HisK) or serine/threonine kinases (Ser/ThrK), sulfate transporter anti-sigma antagonist (STAS), helix-turn-helix (HTH) DNA-binding domains, GGDEF-EAL domains etc. (Figure 2) (Huala *et al.*, 1997; Losi *et al.*, 2002; Purcell *et al.*, 2010; Zoltowski, Motta-Mena and Gardner, 2013; Glantz *et al.*, 2016).

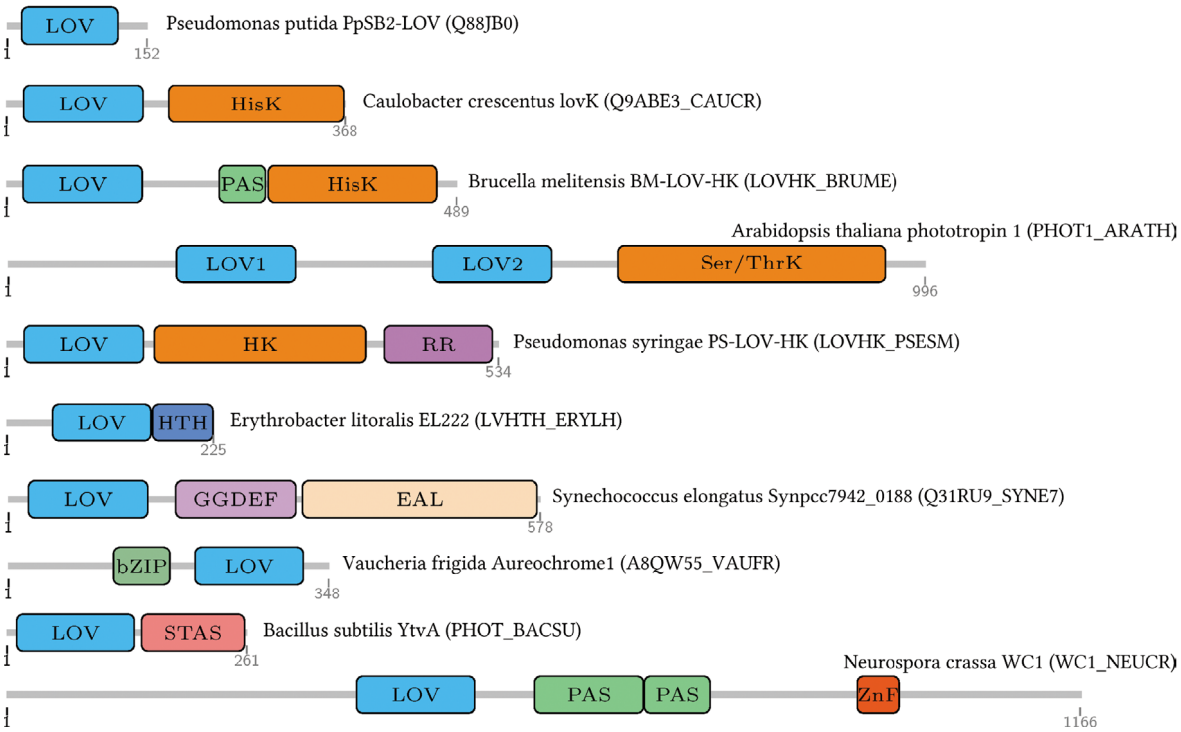


Figure 2. Various LOV-containing proteins utilizing different effector domains. Domains legend: histidine (HisK) or serine/threonine kinases (Ser/ThrK); Per-Arnt-Sim (PAS) sensory or signal transduction domains; response regulator (RR) phosphorylation receiver domains; zinc finger motifs (ZnF); helix-turn-helix (HTH) DNA-binding domains; GGDEF and EAL domains, named after a conserved sequence, synthesizes and degrades cyclic di-GMP; basic leucine zipper (bZIP) DNA-binding domain; sulfate transporter anti-sigma antagonist (STAS). Uniprot accession codes of example proteins are shown in patentees.

However, the so-called short LOV proteins were identified, comprised of only a single LOV domain and lacking a fused effector domain (Losi and Gärtner, 2008). So far, the physiological function could be assigned only for a few short LOV proteins. VVD, a short LOV protein in fungi, for instance, functions as an antagonist by binding to White Collar-1 complex which triggers a multitude of light responses (Schwerdtfeger and Linden, 2003). Another short LOV protein, DsLOV from a marine phototrophic bacterium, was found to mediate light-dependent pigmentation (Endres *et al.*, 2015), whereas the RsLOV protein from purple photosynthetic bacteria apparently regulates not only light-dependent but also redox-dependent processes (Hendrischk *et al.*, 2009; Metz, Jager and Klug, 2012).

Photocycle of LOV domains

The photochemistry of LOV domains has been extensively studied on various isolated LOV domains (Kottke *et al.*, 2003; Alexandre *et al.*, 2009; Pfeifer *et al.*, 2009). In the ground state (henceforth referred to as the dark state), the flavin chromophore is fully oxidized and non-covalently bound to the protein (Figure 3B). This state can be recognized in an UV-Vis absorption spectrum where the flavin absorption peaks are situated at around 360 nm and 447 nm (denoted as LOV447) (Figure 3A). In comparison to the free state, the flavin spectrum, due to the partially hydrophobic environment of the chromophore pocket, show

three narrowed but still not completely resolved peaks at ~447 nm (exact wavelength can vary depending on chromophore or environment of the binding pocket).

The absorption of blue-light by flavin results in the excited singlet state that can decay via inter-system crossing into the triplet state on the time scale of 1 – 10 ns. The triplet state has characteristic absorption peaks at ~650 nm and 715 nm (denoted as LOV715). Then the triplet state decays in the order of 0.1 ms to 5 ms and results in the formation of a covalent adduct between the conserved cysteine-Sγ atom and flavin-C4a atom and additionally the protonation of N5 atom in flavin (Pfeifer *et al.*, 2009). This state (henceforth referred to as the light state) leads to a characteristic spectrum change where a broad peak with a maximum at around ~390 nm appears (denoted as LOV390). Several mechanisms for the covalent adduct formation were proposed, for instance, the concerted, ionic or radical pair mechanisms (Crosson and Moffat, 2001; Kay *et al.*, 2003; Kennis *et al.*, 2003; Schleicher *et al.*, 2004). However, an elegant study with time-resolved infrared spectroscopy provided evidence that the adduct formation proceed through the neutral triplet state of flavin, which supported the radical pair mechanism instead of the ionic or concerted one (Bauer *et al.*, 2011; Thöing *et al.*, 2013).

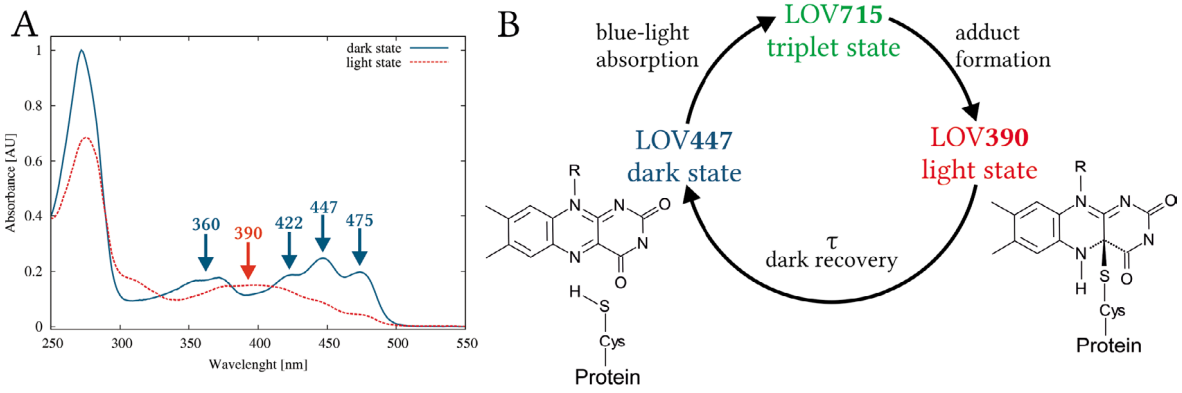


Figure 3. A simplified photocycle diagram of LOV1 from *Chlamydomonas reinhardtii*. (**A**) Typical UV-Vis absorption spectra of LOV domain in the dark and light states. (**B**) A simplified photocycle diagram of the LOV1 domain, where the singlet excited state and the intermediate state between the triplet and light states were omitted. In the dark and light state, only the 7,8-dimethylisoalloxazine ring of flavin is shown, where R – can be a ribityl side chain as in RBF.

The light state can decay back to the dark state spontaneously or after near-UV light exposure (Kottke *et al.*, 2003; Kennis and Alexandre, 2006). The time scale of the thermally-driven decay to the dark state, which is characterized by a time constant (adduct lifetime τ_{rec}), however, varies among LOV proteins from seconds to days (Zoltowski *et al.*, 2007; Jentzsch *et al.*, 2009; Endres *et al.*, 2015). The physiological implications for such variations are not yet understood. The rate-limiting step in the recovery to the dark state was determined to be the deprotonation of flavin-N5 atom. Deprotonation and the associated dark recovery rates are believed to generally depend on solvent access to the active site, the hydrogen bonding

network to flavin and the electronic environment surrounding the flavin (Losi and Gärtner, 2011).

Signaling mechanism of LOV domains

LOV domains are versatile and intrinsically modular light sensor domains. The wide array of effector domains in various organisms raises a question: how such structurally diverse output domains recognize the signal from the LOV domain? A similar question related to the PAS domains superfamily, which are able to bind different ligands, has been previously addressed (Möglich, Ayers and Moffat, 2009b). Analysis of PAS structures have showed the direct involvement of a β-scaffold formed by five structurally conserved antiparallel β-strands in signal propagation. The β-scaffold, together with flanking N- and C-terminal extensions often promote the formation of dimers (mostly homodimers) and higher-ordered oligomers of many proteins (Huang, Edery and Rosbash, 1993; Pongratz *et al.*, 1998; Taylor and Zhulin, 1999). Moreover, various effector domains coupled to PAS domains require its oligomeric state for function, where the activity of effector domains often depends on quaternary structure and dynamics. The common feature among PAS domains seems to be that signal originates primarily in the β-scaffold region and results in structural and dynamic changes which propagate via linkers to mostly C-terminally-fused effector domains (Möglich, Ayers and Moffat, 2009b).

The LOV core domain is also frequently accompanied by N- or C-terminal helical extensions that are either packed against the core or extend from it (Harper, Neil and Gardner, 2003; Circolone *et al.*, 2012; Herman *et al.*, 2013; Endres *et al.*, 2015). Often the LOV domain is N-terminally fused to the effector via a helical linker, similar to the Jα helix in short LOV proteins. Unfortunately, three-dimensional full-length protein structures containing both sensor and effector domains are sparse and apparently elude structural characterization due to structural flexibility or dynamics. However, analysis of the dark and light state structures of LOV domains from different kingdoms, resulted in the proposal of several signal transduction mechanisms.

LOV domains in plant and algae phototropins often come in tandem (LOV1 and LOV2) and fused to the effector domain, like the light-dependent serine/threonine kinase first identified in *Arabidopsis thaliana* (Figure 2) (Huala *et al.*, 1997). LOV1 in full-length proteins seem to modulate the kinase activity but is not obligatory, however, the function of LOV1 is not as well-understood as that of LOV2 (Matsuoka and Tokutomi, 2005). In contrast, LOV2 is sufficient and obligatory for the regulation of kinase activity. Based on the solution and crystal structures of the LOV2 domain from *Avena sativa*, it was shown that in the dark state the C-terminal Jα helix packs against the β-scaffold. (Harper, Neil and Gardner, 2003; Halavaty and Moffat, 2007). However, upon illumination, the signal propagates through the β-scaffold to the Jα helix resulting in undocking and unfolding of this helix. Molecular docking studies suggested that the LOV2 domain inhibits kinase activity by binding to the

catalytic domain (Tokutomi, Matsuoka and Zikihara, 2008). Whereas light-induced unfolding of Jα helix may result in the dissociation from kinase, thus creating access for ATP binding (Tokutomi, Matsuoka and Zikihara, 2008; Pfeifer *et al.*, 2010).

A somewhat similar activation mechanism was proposed for the light-dependent transcription factor protein, Aureochrome, from algae. In the dark state, the Jα helix is packed against the LOV core domain like in the above-mentioned LOV2, however, upon illumination the Jα helix unfolds and facilitates the dimerization of LOV domains (Herman *et al.*, 2013; Hisatomi *et al.*, 2014). This intrinsically promotes the dimerization of DNA-binding effector domains and binding to DNA.

Unlike the phototropin LOV domains, VVD is a short LOV protein from the fungus, *Neurospora crassa*, which lacks a fused effector domain (Heintzen, Loros and Dunlap, 2001; Schwerdtfeger and Linden, 2003). VVD is known to act as an antagonist by binding to the protein, White Collar-1 (WC-1), which controls various light responses. VVD has an N-terminal extension consisting of a short helix, loop and β-strand (altogether referred as Ncap) which packs against the β-scaffold of the core domain, in similar way as the Jα helix in AsLOV2 (Figure 1C) (Zoltowski *et al.*, 2007). Light-induced structural changes disrupt packing of the Ncap and promote a transition from the monomer to a rapidly exchanging dimer (Zoltowski *et al.*, 2007). Although the exact mechanism of WC-1 regulation is not yet elucidated, however, it contains a LOV domain homologous to VVD, which can participate in binding to it.

The first bacterial protein, YtvA, containing LOV domain was identified in *Bacillus subtilis* (Losi *et al.*, 2002). Unlike plant or algae phototropins, YtvA contains a single LOV domain fused to a sulfate transporter anti-sigma antagonist domain (STAS) via a helical linker. Although the protein was identified based on its sequence homology to plant phototropins, crystallographic, biochemical and spectroscopic data have demonstrated that it is capable of different light-induced changes (Bednarz *et al.*, 2004; Möglich and Moffat, 2007; Tang *et al.*, 2010; Engelhard *et al.*, 2013). The isolated LOV domain, YtvA-LOV, was shown to be a homodimer both in solution and in the crystal structure regardless of the photocycle state, where the C-terminal Jα helices extend from the core domain in a coiled-coil-like arrangement (Figure 4A) (Möglich and Moffat, 2007; Jurk *et al.*, 2010). The Jα helix is connected by only one residue, which does not allow it to pack against the hydrophobic patch of the core domain β-scaffold. Instead, this exposed hydrophobic patch forms the dimer interface.

Analysis of the dark and illuminated structures of YtvA-LOV (dark grown crystals were illuminated with blue-light) showed small structural rearrangements present all over the structure. Light-induced structural changes were proposed to be the tilting and rotational motion of monomers relative to each other in the dimer (Möglich and Moffat, 2007). Later on, the crystal structure of the bacterial short LOV protein, PpSB1-LOV, from *Pseudomonas*

putida was obtained in the fully light-adapted state, where the crystals grew under permanent blue-light illumination (Figure 4D) (Circolone *et al.*, 2012).

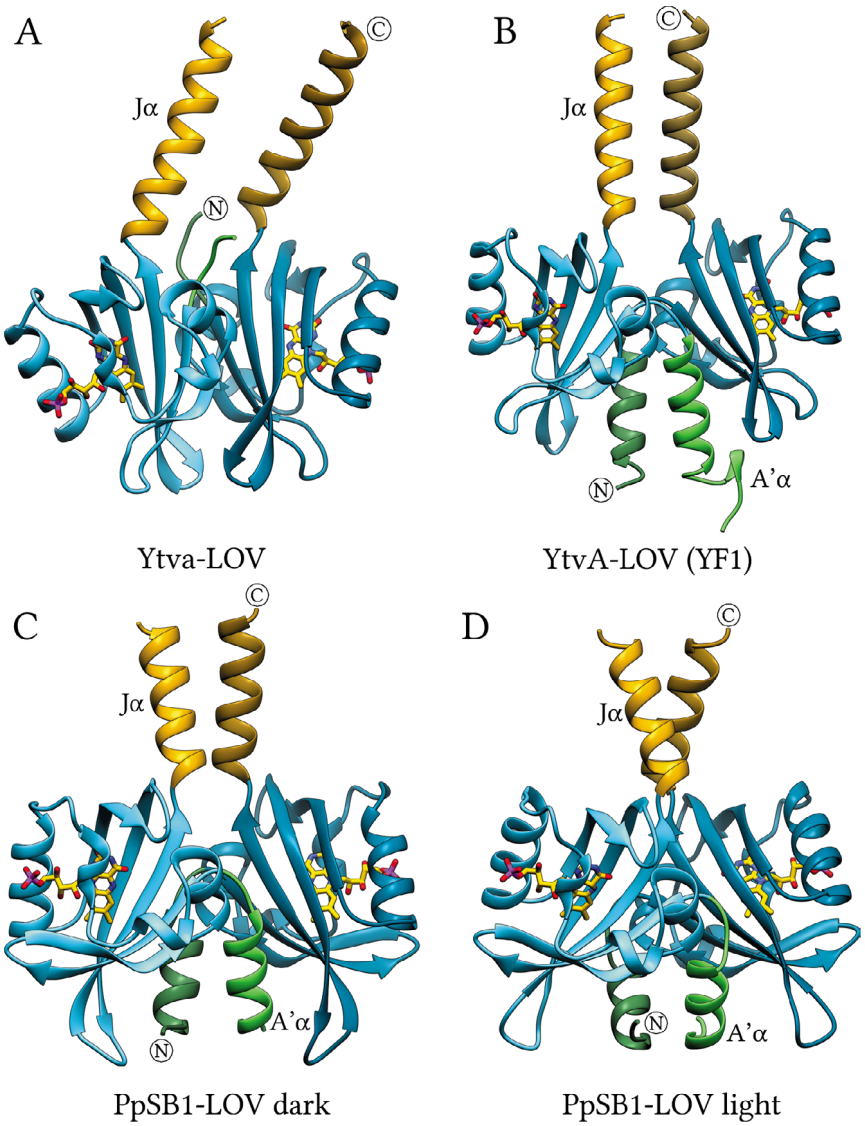


Figure 4. Crystal structures of YtvA-LOV and PpSB1-LOV. Structures are represented as ribbon diagrams where the core domain is colored in blue and the non-canonical elements A'α in green and Jα in gold. Bound FMN molecule are shown as stick models. Chains of a dimer are colored in dark and light shades. (**A**) The dimer structure of YtvA-LOV in the dark state (PDB ID: 2PR5) which was crystallized as an isolated domain. (**B**) The dimer structure of YtvA-LOV in the dark state (PDB ID: 4GCZ) as a part of chimeric protein YF1 (residues 1 – 138). (**C**) The dimer structure of PpSB1-LOV in the dark state (PDB ID: 5J3W). (**D**) The dimer structure of PpSB1-LOV in the light state (PDB ID: 3SW1). The light state structure show a rotation of the dimer interface, together with rearrangement of A'α and Jα helices.

Unlike most LOV proteins studied previously, PpSB1-LOV has very slow dark recovery (τ_{rec} ~40 h at 20 °C) what allowed crystallization of LOV domain in the light state (Circolone *et al.*, 2012). Based on the structural similarity of PpSB1-LOV to YtvA-LOV, a new light-state model of YtvA-LOV was proposed (Diensthuber *et al.*, 2013). It was previously observed that light-induced changes in the YtvA-LOV crystal structure were perhaps restrained by the crystal lattice, whereas the new light state model of YtvA-LOV suggested such motions

with higher magnitude. Later on, the PpSB1-LOV dark state structure confirmed large-scale light-induced changes in the dimer interface (Figure 4C,D) (Röllen *et al.*, 2016). However, despite the similarity between YtvA-LOV and PpSB1-LOV, there are some differences which will be discussed further in Results and Discussion sections.

LOV domain as an optogenetic or optical tool

Light is a versatile medium for non-invasive and the spatiotemporally precise manipulation of biological processes. The field of optogenetic tools is a rapidly developing area, which deals with genetically-encoded light-photoreceptors that are able to perturb and control cellular processes. Optical tools, on the other hand, are not limited to genetically-encoded photoreceptors and deal with *in vitro* applications. Various optogenetic/optical tools have been developed to control gene expression, ion channels and ion pumps with light (Fenno, Yizhar and Deisseroth, 2011) or to report, for instance, the presence of specific ligands, pH or ions with fluorescence (Berg, Hung and Yellen, 2009; Mishin *et al.*, 2015).

Various naturally occurring photoreceptors from different kingdoms are used to design optogenetic/optical tools (Shcherbakova *et al.*, 2015). Green fluorescent protein (GFP) isolated from jelly fish is one of the first optogenetic tools served as *in vivo* fluorescence marker fused to a target protein (Chalfie *et al.*, 1994). Since then, various derivatives of GFP proteins were developed, with excitation wavelengths covering almost the entire visible spectrum (Sample, Newman and Zhang, 2009).

Besides the GFP family, blue-light absorbing LOV photoreceptors have gained attention in the last decade. The increased attention to them lies in the advantageous properties of the LOV protein. LOV proteins use the ubiquitously abundant flavins as a chromophore, in contrast to the GFP protein family, where the chromophore is autocatalytically developed in the presence of molecular oxygen. The intrinsic oxygen-independent fluorescence of bound-flavin in LOV proteins can be used for anaerobic applications (Drepper *et al.*, 2007; Buckley *et al.*, 2016). Moreover, LOV proteins are almost half in size compared to GFP (~12 – 16 kDa vs ~27 kDa), which makes them attractive for applications where the size of the fluorescent label is crucial.

A wide array of optogenetic/optical tools have been recently developed based on LOV proteins, starting from the use of LOV domains as a fluorescence reporter to the development of LOV-based optogenetic tools (Wingen *et al.*, 2014; Pudasaini, El-Arab and Zoltowski, 2015). The photophysical and photochemical properties of LOV-based fluorescence reporters are improving, with the best example being CreiLOV (Mukherjee *et al.*, 2015). Through optimization, CreiLOV has achieved a fluorescence quantum yield of 0.51, which is higher than the typical quantum yield of ~0.2 – 0.3 of the LOV protein (Wingen *et al.*, 2014). Fluorescent reporters with high quantum yield is of special interest for microscopy. Several LOV-based reporters have been successfully used not only for confocal but also for super-

resolution fluorescence microscopy (Buckley *et al.*, 2015; Mukherjee and Schroeder, 2015). Another example exploiting the oxygen-independence of LOV protein fluorescence is development of a FRET-based sensor for molecular oxygen FluBO (Potzkei *et al.*, 2012). FluBO enabled levels of molecular oxygen to be measured *in vivo* for the first time.

Further interesting application exploits the fact that some LOV domains promote dimerization upon blue-light activation. Fusion of LOV domain to tyrosine kinase leads to a light-dependent regulation of cellular responses (Grusch *et al.*, 2014). In a similar fashion, the fusion of the LOV protein to the basic region/leucine zipper (bZIP) DNA-binding domain resulted in a versatile and promising optogenetic system to control gene-expression with light (Nakatani and Hisatomi, 2015).

Despite great progress in the development of LOV-based optogenetic/optical tools there are certain disadvantages. For instance, the excitation and emission wavelengths are so far limited to ~450 nm and ~500 – 520 nm, respectively. Mutagenesis has led only to a maximum of ~10 nm change in wavelength (Wingen *et al.*, 2014; Davari *et al.*, 2016). The use of modified flavin chromophores also failed so far to dramatically affect the excitation wavelength (Mansurova *et al.*, 2011, 2013; Silva-Junior *et al.*, 2013). Further investigations need to be done to provide the wide range of excitation wavelengths as seen in the GFP protein family.

Pseudomonas bacterial short LOVs investigated in this work

In sequenced genomes of the *Pseudomonas* genus several bacterial short LOV proteins were identified (Krauss *et al.*, 2005; Rani *et al.*, 2013). Among them, short LOV proteins from *P. putida* and *P. fluorescens* were characterized and proven to undergo a similar photocycle as phototropins or bacterial YtvA-LOV (Krauss *et al.*, 2005; Jentzsch *et al.*, 2009; Rani *et al.*, 2013).

P. putida is a bacterium that is found in most soil and water habitats, and the KT2440 strain is particularly known to colonize plant roots (Espinosa-Urgel, Salido and Ramos, 2000). The surface of the root provides bacteria with nutrients. In turn, the *P. putida* induces plant growth and protects plants from pathogens. Likewise, *P. fluorescens* lives in soil, water and especially on plant roots. The Pf-5 strain, for instance, produces a variety of secondary metabolites, including antibiotics against soil-borne plant pathogens (Paulsen *et al.*, 2005). Strain SBW25 grows not only on plant roots but also on leaves where they can contribute to plant growth (Mavrodi *et al.*, 2007). The presence of light-receptors may be related to the symbiotic interaction between *Pseudomonas* and plants.

Two short LOV proteins PpSB1-LOV and PpSB2-LOV from *P. putida* KT2440 were first identified and characterized (Krauss *et al.*, 2005; Jentzsch *et al.*, 2009). Other short LOV proteins W619_1-LOV and W619_2-LOV from *P. putida* W619 stain together with proteins Pf0-1-LOV, Pf5-LOV, SBW25-LOV from 3 different strains of *P. fluorescens* (Pf0-1, Pf-5 and SBW25, respectively) were subsequently identified and characterized (Rani *et al.*, 2013). In

this work, the results of biophysical, biochemical and structural characterization of PpSB2-LOV, W619_1-LOV, Pf5-LOV and SBW25-LOV proteins will be presented and compared to those of their previously characterized homologues (Table 1).

Table 1. List of short LOV proteins and their adduct lifetimes from *P. putida* and *P. fluorescens*.

Protein	Source	Uniprot ID	Number of residues	Adduct lifetime [min] at 37 °C *
Pf01-LOV	*P. fluorescens* Pf0-1	Q3KHW7	158	810
Pf5-LOV	*P. fluorescens* Pf-5	Q4KI48	152	3.6
PpSB1-LOV	*P. putida* KT2440	Q88E39	142	109
PpSB2-LOV	*P. putida* KT2440	Q88JB0	151	0.5
SBW25-LOV	*P. fluorescens* SBW25	C3K1W0	155	1470
W619_1-LOV	*P. putida* W619	B1J385	146	53
W619_2-LOV	*P. putida* W619	B1JAC4	148	0.9

* Adduct lifetimes from (Rani *et al.*, 2013)

The conserved sequence motif GXN<u>C</u>RFLQ (where X is any residue) of LOV domains with the adduct-forming cysteine were previously identified in plants. However the short LOV proteins from the *Pseudomonas* genus show a slightly different motif – Y(Q/R)D<u>C</u>RFLQG, which places them on a different branch of the phylogenic tree (Krauss *et al.*, 2005; Rani *et al.*, 2013). Additionally, sequence alignment of short LOV proteins in *Pseudomonas* shows that they differ in the length of non-canonical C-terminal extensions (the Jα helix and unstructured terminus) (Figure 5).

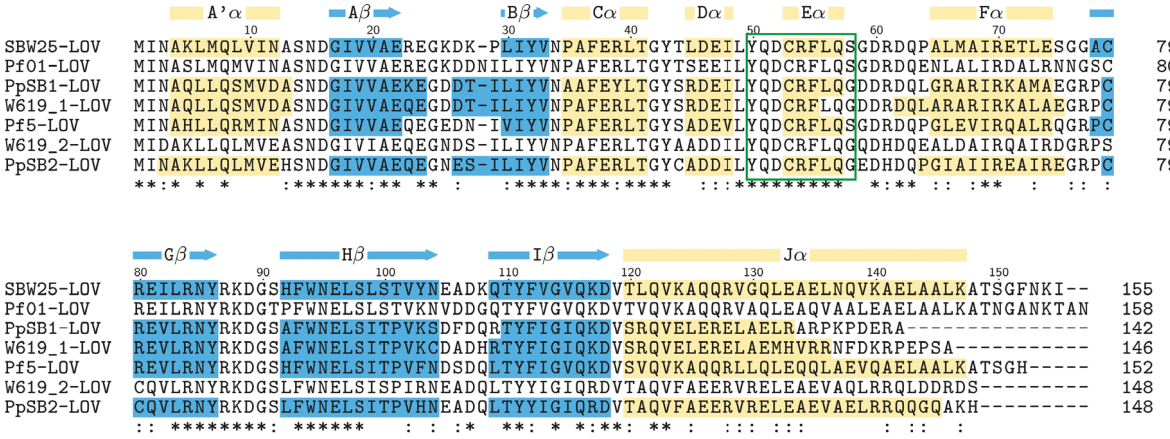


Figure 5. The sequence alignment of short LOV proteins from *P. putida* and *P. fluorescens*. The sequence alignment was done with the Clustal Omega program. The secondary structure is assigned with the DSSP program from crystal structures, where helix (yellow) and strand (blue) are marked in sequence. Secondary structure element labels are shown at the top. The asterisks at the bottom line of the alignment indicate identical residues in a given sequence position, while double dots refer to conserved residues, respectively. Sequence numbering for SBW25-LOV is shown at the top. The conserved sequence identified in *Pseudomonas* LOV proteins is emphasized by the green box.

Their sequence identity matrix shows that several short LOV proteins share high sequence identity (Table 2). For example, the KT2440 and W619_1 strains of *P. putida*, both contain two short LOV proteins, unlike those of *P. fluorescens*. Moreover, PpSB1-LOV and PpSB2-LOV from the KT2440 correspond to W619_1-LOV and W619_2-LOV from the

W619_1 strain with ~89% and ~84% sequence identity, respectively (Table 2). PpSB1-LOV and PpSB2-LOV have long and short adduct lifetimes (109 min vs 0.5 min at 37 °C, respectively) and the corresponding W619_1-LOV and W619_2-LOV also show a similar difference in adduct lifetimes (53 min vs 0.9 min at 37 °C, respectively) (Rani *et al.*, 2013).

The *P. fluorescens* strains contain only single short LOV proteins in their genome, and have also shown a high sequence identity > 70%, especially between SBW25-LOV and Pf01-LOV. Both the proteins demonstrate slow dark recovery (1470 min and 810 min at 37 °C, respectively), in contrast, Pf5-LOV has a short adduct lifetime of only 3.6 min at 37 °C (Rani *et al.*, 2013).

Table 2. The sequence identity matrix of short LOV proteins from *P. putida* and *P. fluorescens*. Values in the table indicate pair-wise sequence identity in percent.

	SBW25-LOV	Pf01-LOV	PpSB1-LOV	W619_1-LOV	Pf5-LOV	W619_2-LOV	PpSB2-LOV
SBW25-LOV	100.0	78.7	57.8	59.6	73.0	61.5	60.8
Pf01-LOV	78.7	100.0	59.9	58.2	71.7	58.8	58.8
PpSB1-LOV	57.8	59.9	100.0	88.7	67.6	62.7	66.9
W619_1-LOV	59.6	58.2	88.7	100.0	66.4	62.3	65.8
Pf5-LOV	73.0	71.7	67.6	66.4	100.0	63.5	66.2
W619_2-LOV	61.5	58.8	62.7	62.3	63.5	100.0	83.8
PpSB2-LOV	60.8	58.8	66.9	65.8	66.2	83.8	100.0

1.2 Sensory histidine kinases

Histidine kinases (HK) are versatile signaling proteins found in most bacteria, archaea, and lower eukaryotic species such as slime molds, fungi, and plants (Inouye and Dutta, 2002). HKs are the major signal transduction protein in prokaryotes as Ser/Thr/Tyr protein kinases are in the eukaryotes. HKs regulate cellular processes such as nutrient acquisition, adaptation to environmental stress, cell motility, development, virulence and intercellular communication (Parkinson, 1992).

HK functions via two-component signal transduction (TCS) pathways. The canonical TCS system, such as the EnvZ protein, consists of two distinct signaling proteins, a mostly membrane-bound sensor histidine kinase (SHK) and a cytoplasmic response regulator (RR) protein (Tomomori *et al.*, 1999). SHK mainly has the periplasmic sensor domain, the transmembrane domain and the linker domain connecting the transmembrane domain and cytoplasmic CA domain (Catalytic and ATP binding). Most of the histidine kinases have periplasmic sensor domains, however, few kinases lack transmembrane domain and have cytosolic sensor instead (Kaserer and West, 2010). PAS, GAF (cGMP-specific phosphodiesterases, adenylyl cyclases and FhlA) and HAMP (Histidine kinases, Adenylate cyclases, Methyl accepting proteins and Phosphatases) are widespread linker domains

(Galperin, Nikolskaya and Koonin, 2001). The PAS domains not only able to transduce signals but also able to integrate intracellular redox signaling when fused with other sensor domains (Parkinson, 2010; Purcell *et al.*, 2010).

In response to extracellular stimuli, the SHK first catalyzes the autophosphorylation of its conserved histidine residue by transferring a phosphoryl group from the bound ATP. Second, upon the binding of RR to SHK, the phosphoryl group is transferred from histidine to a conserved aspartate residue within the receiver domain of the RR protein. In turn, RR often regulates the specific gene expression by direct interaction with the promoter region of the gene. The response, however, depends on the concentration of phosphorylated RR, which is often also regulated by HK as a feedback loop through the dephosphorylation of RR. Thus HK often act as a self-consistent system which regulates cellular responses (Russo and Silhavy, 1993).

In addition, a hybrid HK with a multi-step phosphorelay system possesses a C-terminal receiver domain, a histidine-containing phosphotransfer (HPt) domain/protein, and an RR domain/protein. The phosphoryl group from the histidine residue is first transferred to an aspartate residue within the C-terminal receiver domain. It is then transferred to a histidine residue of the HPt domain/protein and eventually to an aspartate residue of RR, activating the signaling system. Such a multi-step phosphorelay system is more complex yet it allows for a more sophisticated modulation of signaling which enables other proteins to interact with the HPt domain/protein.

Crystal structures of isolated sensors, dimerization and histidine phosphotransfer (DHp) domains, catalytic and ATP binding (CA) domains and the complete kinase domains (DHp + CA), have greatly contributed to the understanding of SHKs (Tomomori *et al.*, 1999; Marina, Waldburger and Hendrickson, 2005; Ayers and Moffat, 2008). The crystal structure of VicK SHK from *Streptococcus mutans* have demonstrated for the first time the entire cytoplasmic portion of SHK that comprises: the signal transducer HAMP, the sensor domain PAS, the DHp and the CA domains, as illustrated in Figure 6 (Wang *et al.*, 2013). VicK has short N-terminal transmembrane helices connected to the HAMP domain for anchoring of VicK to the membrane, which were truncated for crystallization. A long, rod-shaped dimer showed an arrangement of domains connected by long helical segments. Based on comparison to the related SHK HL853, a model of activation was proposed (Marina, Waldburger and Hendrickson, 2005). In this model, sensory input results in the bending of the DHp helix which repositions the CA domain close to the phospho-accepting histidine residue. Since VicK is a dimeric protein, it was suggested that autophosporylation happens in an only one-side-at-a-time manner. This also agrees well with observations on other SHK (Bhate *et al.*, 2015). However, a detailed understanding of SHK function, especially at different steps, remains elusive and requires further intensive investigations.

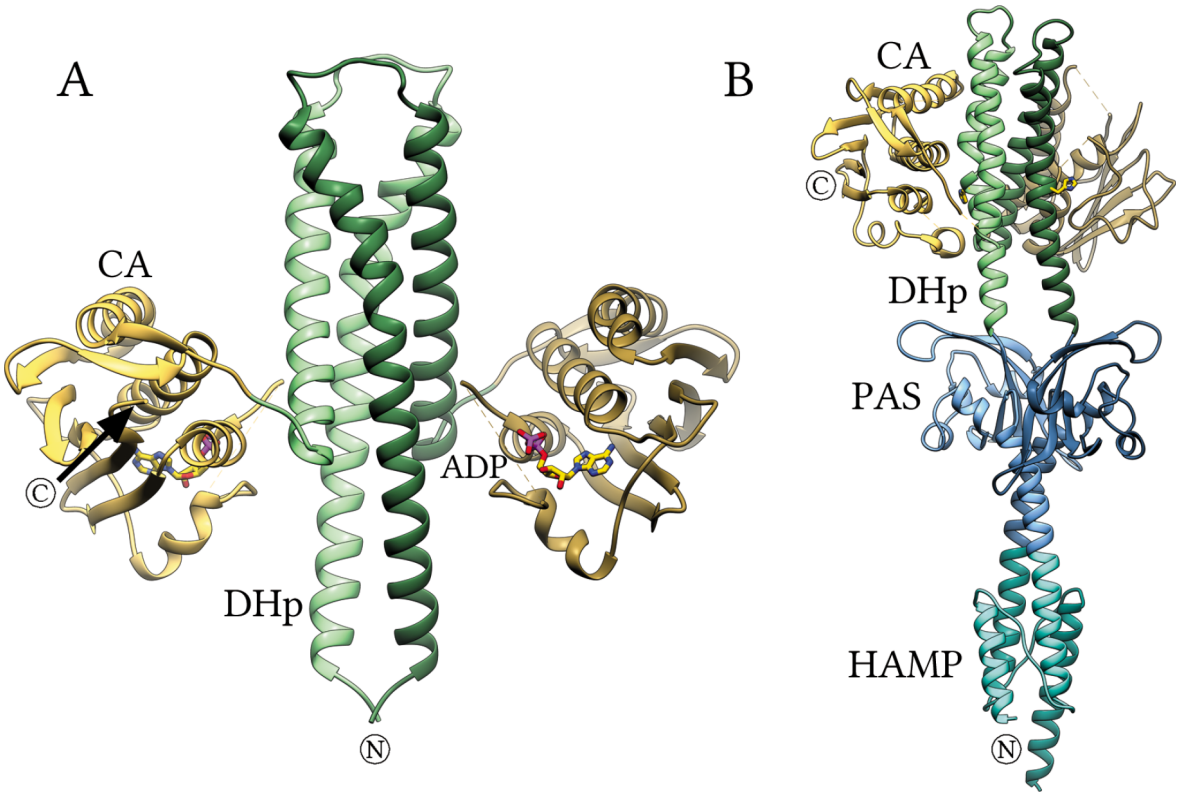


Figure 6. Crystal structures of the cytoplasmic portions of SHK HK853 and VicK. Structures of HK853 (PDB ID: 2C2A) and VicK (PDB ID: 4I5S) are represented as ribbons where chains are colored in light and dark shades. (**A**) The crystal structure has one molecule which by application of crystallographic symmetry results in symmetric-dimer. The CA domains (gold) are flanked laterally from the DHp domain (green) helix bundle. The bound ADP nucleotide is shown as a stick model. (**B**) The structure of VicK dimer contains the following cytoplasmic domains: HAMP (cyan), PAS (blue), DHp (green) and CA (gold).

LOV Histidine Kinases

LOV histidine kinases (LOV-HK) form a subgroup of SHKs, with the LOV domain acting as a sensor domain. LOV-HKs are the most abundant class among the bacterial LOV-containing proteins (Briggs, 2007; Glantz *et al.*, 2016). Several bacterial LOV-HK proteins have been biochemically and biophysically characterized, as listed in Table 3, yet only one three-dimensional structure of a natural LOV-HK has been obtained so far (Rivera-Cancel *et al.*, 2014). Elucidation of the physiological functions of LOV-HK has shown that they mostly regulate host and pathogen interactions between the bacteria and both plants and animals, such as virulence or cell attachment to host (Purcell *et al.*, 2007; Swartz *et al.*, 2007; Bonomi *et al.*, 2012). In plant-pathogenic bacterium, PS-LOV-HK is used for discrimination of the soil or leaf environment which differ in ambient light conditions (Moriconi *et al.*, 2013).

There are several domain architectures observed in characterized bacterial LOV-HKs: two-domain proteins (LOV–HK) and three-domain proteins (LOV–HK–RR or LOV–PAS–HK). According to the recent bioinformatical analysis of LOV protein diversity, LOV–HK–RR is the most common architecture found in bacteria (Glantz *et al.*, 2016). The HK domain mostly contains the dimerization and phosphotransfer domain (DHp) which includes

phospho-accepting histidine residue and promotes the dimerization of HK. Among SHK proteins, dimerization was determined to be essential for kinase activity (Bhate *et al.*, 2015).

Table 3. Characterized HK-LOV proteins from bacteria and algae.

Name	Source	Function	Domain architecture	Reference
BM-LOV-HK BA-LOV-HK	*Brucella melitensis* *Brucella abortus*	virulence, survival within macrophages	LOV–PAS–HK	(Swartz *et al.*, 2007)
EL346-LOV-HK	*Erythrobacter litoralis*	putative general stress response RR identified	LOV–HK	(Swartz *et al.*, 2007)
LOVK	*Caulobacter crescentus*	cell attachment	LOV–HK	(Purcell *et al.*, 2007)
O. tauri LOV-HK	*Ostreococcus tauri***	circadian clock regulation	LOV–HK–RR	(Djouani-Tahri *et al.*, 2011)
PS-LOV-HK	*Pseudomanas syringae*	virulence, soil and leaf discrimination	LOV–HK–RR	(Swartz *et al.*, 2007; Moriconi *et al.*, 2013)
R-LOV-HK	*Rhizobium leguminosarum*	cell attachment, flagella synthesis, EPS* production	LOV–HK	(Bonomi *et al.*, 2012)
Xac-LOV	*Xanthomonas axonopodis* pv. citri	cell attachment, virulence	LOV–HK–RR	(Kraiselburd *et al.*, 2012)

* extracellular acidic exopolysaccharide; ** algae

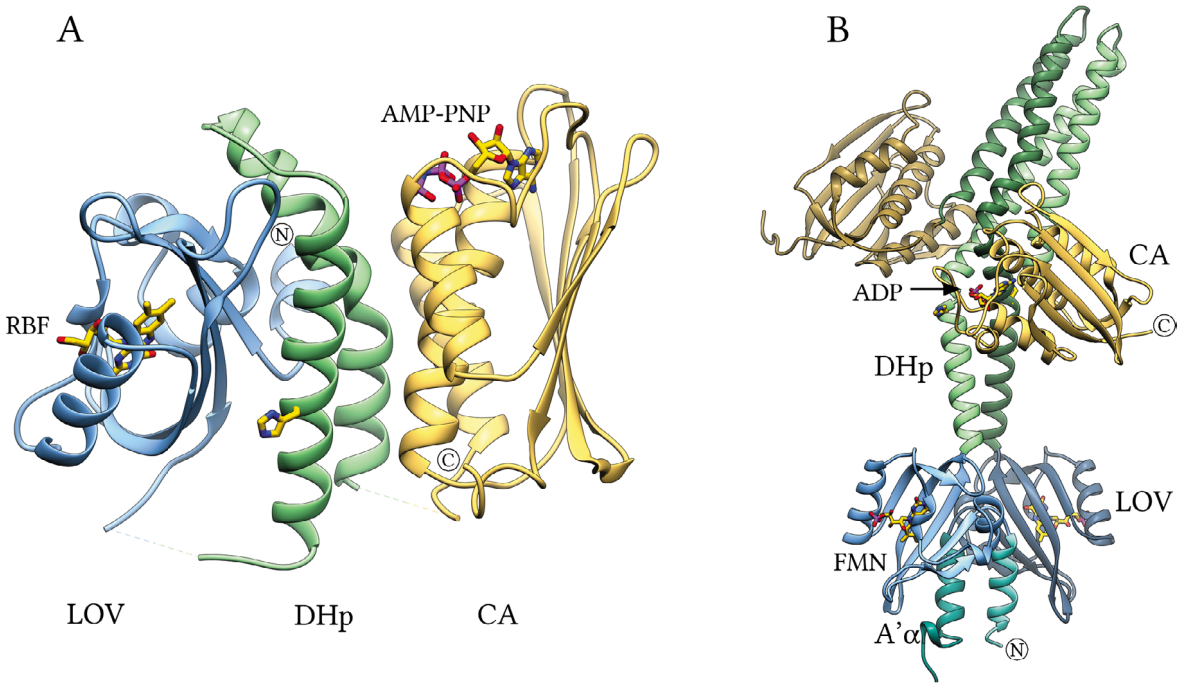


Figure 7. Crystal structures of natural EL346-LOV-HK and engineered LOV-HK, YF1. (A) The monomeric crystal structure of EL346-LOV-HK (PDB ID: 4R3A) is shown as a ribbon diagram, where the LOV domain is colored in blue, the DHp domain in green and the CA domain in gold. Bound RBF, AMP-PNP and the phospho-accepting histidine are shown as stick models. (B) The dimeric crystal structure of YF1 (PDB ID: 4GCZ) is shown as a ribbon diagram, where the N-terminal extension A'α helix is colored in cyan, the LOV domain in blue, the DHp domain in green and the CA domain in gold. Chains A and B are colored in the dark and light shades. Bound FMN and ADP are shown as stick models.

LOV histidine kinase EL346-LOV-HK was identified in marine bacterium and shown to relay signal to conjugated RR, which seems to regulate the general stress response (Swartz *et al.*, 2007; Correa *et al.*, 2013). This first LOV-HK crystal structure showed a novel tertiary structure for SHK. So far, SHK have been reported to be dimeric, which is essential for their functionality. In contrast, EL346-LOV-HK is a monomer in both light and dark states in solution and in the crystal structure despite the fact that it contains the DHp domain (Rivera-Cancel *et al.*, 2014). Dimerization through the DHp domain is apparently inhibited by the LOV domain, where the DHp is sandwiched between the LOV and CA domains (Figure 7A). Light-induced changes in the LOV domain are proposed to propagate structural changes through the DHp domain, which alter the DHp and CA interactions. These alterations would in turn initiate the release of the CA domain from the inhibited conformation.

The second structurally characterized LOV-HK is an engineered YF1 protein (Möglich, Ayers and Moffat, 2009) (Figure 7B). The YF1 protein is a chimeric protein constructed by the substitution of the oxygen-sensitive PAS domain in *Bradyrhizobium japonicum* FixL histidine kinase (Gilles-Gonzalez, Ditta and Helinski, 1991) with the LOV domain from *Bacillus subtilis* YtvA (YtvA-LOV). Substitution of the structurally similar PAS domain of FixL to YtvA-LOV resulted in the formation of a functional light-dependent histidine kinase. YF1 showed similar autophosphorylation rates as parental FixL, whereas illumination decreases the net kinase activity by a ~1000-fold (Möglich, Ayers and Moffat, 2009a).

FixL has the conjugated RR protein, FixJ, which in the phosphorylated state promotes the expression of specific gene. YF1, like parental FixL, retained ability of phosphotransfer to FixJ, and was used to develop the optogenetic tools, pDusk and pDawn (Ohlendorf et al., 2012). The plasmids pDusk and pDawn enabled the expression of the target gene only under dark or under light conditions, respectively.

Although YF1 is an engineered protein, it is similar in architecture to natural LOV-HK and can give insight into their function. YF1 in the dark and light states consistently forms a dimer in solution such as other SHKs. Crystal structure of the YF1 dimer showed a domain organization similar to that of VicK SHK (Diensthuber *et al.*, 2013). LOV domains are connected to long coiled-coil DHp domain helices, which are flanked by two lateral CA domains, as shown in Figure 7B. The CA domains are connected by flexible loops to the DHp, which perhaps allows them to adapt different spatial positions depending on the state. YF1 was crystallized under dark conditions in the presence of ATP, yet only one CA domain has bound ADP while the other is free of nucleotide. The distance between ADP and phospho-accepting histidine was ~23 Å. However, such distance and orientation of the CA domains does not allow unambiguous assignment of the YF1 functional state, whether it is prior to the autophosphorylation or after that.

Although YF1 was only crystallized in the dark state, a model of light-induced changes was proposed using the light state homology model of its LOV domain dimer, which was based on the crystal structure of PpSB1-LOV in the light state (Circolone *et al.*, 2012; Diensthuber *et al.*, 2013). The model suggested that light-induced changes from the LOV domain in YF1 propagate through torque motions of the DHp helices and result in the super-twisting of DHp helices. Changes in the DHp domain, in turn, would influence the CA domains and thus the autophosphorylation activity.

SB1F1 and SB2F1 – model LOV-HKs investigated in this work

The SB1F1 and SB2F1 protein presented next in this study were engineered in similar fashion as YF1 (Krauss, unpublished), where instead of the YtvA-LOV domain (residues 1 – 127), PpSB1-LOV (residues 1 – 120) and PpSB2-LOV (residues 1 – 120) short LOV proteins were used, respectively. Beyond residue 120, SB1F1 and SB2F1 share with YF1 an identical sequence for the DHp and CA domains (Figure 8).

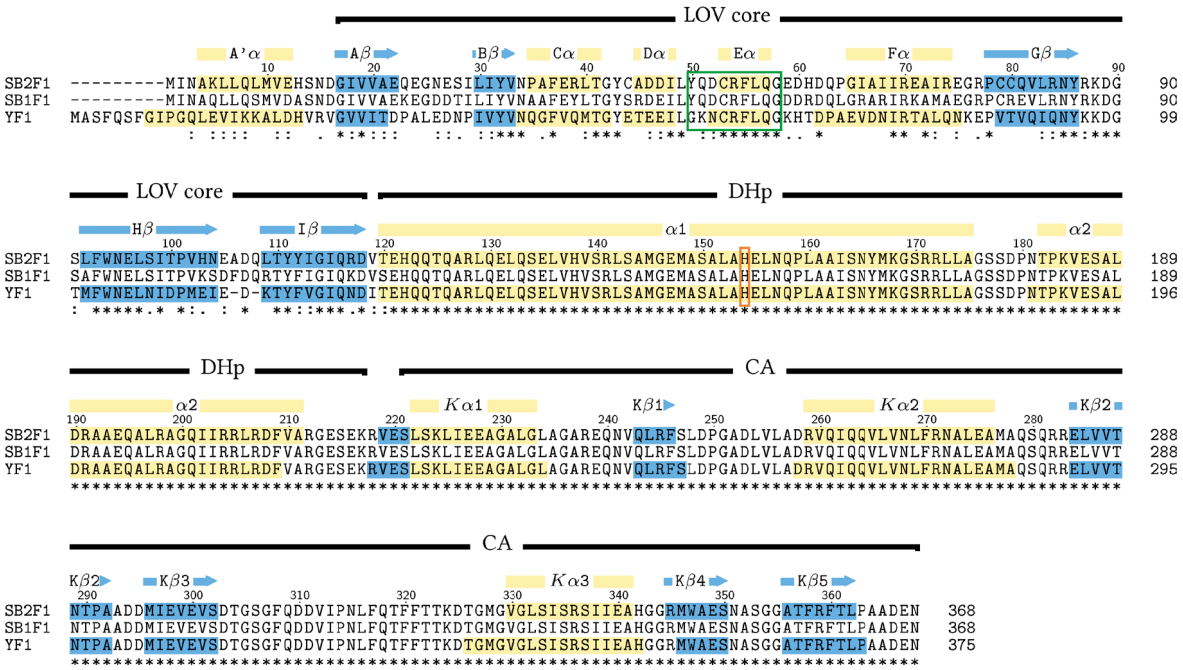


Figure 8. Sequence alignment of SB1F1, SB2F1 and YF1. The secondary structure was assigned with the DSSP program from crystal structures of SB2F1 (determined in this work) and YF1 (PDB ID: 4GCZ, Diensthuber *et al.*, 2013), where strands (blue) and helices (yellow) are marked in sequence. Secondary structure element labels are shown at the top. The asterisks at the bottom line of the alignment indicate identical residues in a given sequence position, while double and single dots refer to highly and moderately similar residues, respectively. Sequence numbering for SB2F1 is shown at the top. The conserved sequence identified in *Pseudomonas* LOV proteins is emphasized by the green box. The phospho-accepting histidine is emphasized by the red box.

The purpose of SB1F1 and SB2F1 engineering was not only to create a model system resembling natural LOV-HK but also to develop optogenetic tools. For this reason, two LOV proteins, PpSB1-LOV and PpSB2-LOV, were chosen with different adduct lifetimes (5.2 min vs 2471 min at 20 °C) (Jentzsch *et al.*, 2009). The different lifetimes would enable the signaling response in optogenetic applications to be modulated. Preliminary data on SB1F1

and SB2F1 showed light-dependent autophosphorylation activity with different specificity, such as that of YF1 (Krauss, unpublished). The structural and functional characterization of SB1F1 and SB2F1 would further enrich understanding of LOV-HK function as well as provide basis for design of new optogenetic tools.

1.3 Scope and outline of the thesis

In the last two decades, valuable data has been obtained on the structure, function and physiological role of LOV photoreceptor proteins. LOV proteins from different kingdoms share several key features, such as primary photochemistry, however, they vary to some extent in their signal propagation mechanisms within the LOV domain and their signal transduction mechanisms to the effector domains. Previous works have proposed the structural basis for dark recovery kinetics, which has features common to all LOV domains, however, they show differences in details between them. Additionally, as mentioned above, further improvement of LOV proteins need to be done for them to serve as optical tools.

Short LOV proteins constitute one of the prevalent functional cluster of LOV-containing proteins (Glantz *et al.*, 2016), and are the current focus in generation of optogenetic tools (Pudasaini, El-Arab and Zoltowski, 2015). In addition, due to their small size, inherent modularity and tunable photocycle, short LOV proteins provide an ideal model system for structure-function studies. In the first half of this thesis, several short LOV proteins have been characterized. Following questions have been addressed in this thesis:

- What is the structural basis for adduct lifetime variation among bacterial short LOV proteins from *P. putida* and *P. fluorescens*?
- How does a signal propagate upon blue-light absorption from the chromophore within the LOV domain towards the periphery and the potential effector domains?
- Can one exploit modified or uncommon flavin chromophores to improve the photophysical and photochemical properties of LOV proteins for their application as an optical tool?

Most of LOV domains are part of multi-domain proteins, which are commonly connected via an α-helical linker or via a signal transduction domain to the effector domains (Möglich, Ayers and Moffat, 2009a; Glantz et al., 2016). In the second half, the following aims have been pursued by studying the engineered LOV-HK proteins SB2F1 and SB1F1:

- Understanding the signal transduction mechanism from the LOV domain to the effector domain in engineered LOV-HK, that is expected to be similar to that of natural occurring LOV-HKs.
- Gain structural information on different states, which would facilitate further development of optogenetic tools to control complex signaling pathways.

Major biophysical methods used in the thesis are:

- X-ray protein crystallography to obtain structural information.
- Small-angle X-ray scattering to gain structural information in solution.
- UV-Vis absorbance spectroscopy and fluorescence spectroscopy to measure photophysical and photochemical properties.
- Molecular dynamics (MD) simulation to assess the dynamics of LOV proteins in different states.
- Quantum chemical calculations to optimize geometry of the FMN-cysteinyl adduct for MD simulations and the refinement of the light state crystal structure.

In this thesis, first the Material and Methods section is described, which also provide the basic background for the major techniques used here.

The results are presented for each protein separately, starting with the short LOV proteins and then the engineered LOV-HKs. The discussion section is organized in topics where the analysis of data and conclusions are based on data from this work supported by the relevant literature.

2 Materials and methods

2.1 Materials

Chemicals

The chemicals used for all experiments, if not stated separately, were of the highest purity (≥ 95%)

Table 4. Sparse matrix and additive crystallization screens.

Screen Name	Article Number	Company; Location
Additive Screen HT	HR2-138	Hampton Research; Aliso Viejo, USA
AmSO4 Suite	130705	Qiagen; Hilden, Germany
Anions Suite	130707	Qiagen; Hilden, Germany
Cations Suite	130708	Qiagen; Hilden, Germany
Crystal Screen	HR2-110	Hampton Research; Aliso Viejo, USA
Crystal Screen 2	HR2-112	Hampton Research; Aliso Viejo, USA
Ionic Liquid Screen	HR2-214	Hampton Research; Aliso Viejo, USA
JBScreen Plus HTS	CS-507L	Jena Bioscience; Jena, Germany
JBScreen Wizard 1	CS-311	Jena Bioscience; Jena, Germany
JBScreen Wizard 2	CS-312	Jena Bioscience; Jena, Germany
JCSG Core I Suite	130724	Qiagen; Hilden, Germany
JCSG Core II Suite	130725	Qiagen; Hilden, Germany
JCSG Core III Suite	130726	Qiagen; Hilden, Germany
JCSG Core IV Suite	130727	Qiagen; Hilden, Germany
Low Ionic kit	86684	Sigma-Aldrich; St. Louis, USA
MIDAS	MD1-59	Molecular Dimensions; Suffolk, UK
PACT Suite	130718	Qiagen; Hilden, Germany
PEG Grid Screening Kit	36436	Sigma-Aldrich; St. Louis, USA
PEG Suite II	130716	Qiagen; Hilden, Germany
PEG Suite	130704	Qiagen; Hilden, Germany
PGA Screen	MD1-50	Molecular Dimensions; Suffolk, UK
ProPlex Eco Screen	MD1-38-ECO	Molecular Dimensions; Suffolk, UK
Silver Bullets	HR2-096	Hampton Research; Aliso Viejo, USA

Table 5. Ligands and labeling substances.

Compound	CAS number	Purity	Article Number	Company; Location
7-Br-RF	–	–	T308447	Sigma-Aldrich; St. Louis, USA
8-Br-ATP	23567-97-7	≥ 95% (HPLC)	NU-114S	Jena Bioscience; Jena, Germany
8-Cl-RF	–	–	T308544	Sigma-Aldrich; St. Louis, USA
ADP	20398-34-9	≥ 95% (HPLC)	A2754	Sigma-Aldrich; St. Louis, USA
AMP-PNP	25612-73-1	≥ 93% (HPLC)	A2647	Sigma-Aldrich; St. Louis, USA
ATP	987-65-5	≥ 98%	HN35	Carl-Roth; Karlsruhe, Germany
ATP-γ-S	93839-89-5	≥ 90%	CAYM14957-1	Cayman Chemical; Michigan USA
FAD	84366-81-4	≥ 95% (HPLC)	F6625	Sigma-Aldrich; St. Louis, USA
FMN	130-40-5	73% – 79%	F6750	Sigma-Aldrich; St. Louis, USA
FMN	130-40-5	≥ 95% (HPLC)	F8399	Sigma-Aldrich; St. Louis, USA
L(+)-Selenomethionine	3211-76-5	≥ 99%	10553601	Fisher Scientific; Waltham, USA
Lumichrome	1086-80-2	≥ 90%	103217	Sigma-Aldrich; St. Louis, USA
RBF	83-88-5	≥ 98%	A6279	AppliChem; Glenview, USA

Bacterial strains, plasmids and primers

Table 6. Bacterial strains.

Strain	Genotype	Use	Source
E. coli BL21 (DE3)	*F- ompT hsdSB(rB- mB-) gal dcm (DE3)*	Protein overexpression	Novagen; Darmstadt, Germany
E. coli OmniMAX 2 T1R	F‘ {*pro*AB^{+} *lacI*^{q} *lacZ*ΔM15 *Tn*10(Tet^{R})Δ(*ccd*AB)} *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80(*lacZ*)ΔM15 Δ(*lacZYA-arg*F) U169 *end*A1 *rec*A1 *sup*E44 *thi*-1 *gyr*A96 *rel*A1 *ton*A *pan*D	Plasmid multiplication	Invitrogen; Carlsbad, USA
E. coli B834	*F- ompT hsdSB(rB- mB-) gal dcm met (DE3)*	SeMet labeled protein expression	Novagen; Darmstadt, Germany

Table 7. Vectors and plasmids.

Vector	Genotype	Reference
pET28a(+)	P_{T7}, His_6-Tag, MCS, *lac*I *bla*, Km_r pBR322	Novagen; Darmstadt, Germany
pET28a(+)PpSB2-LOV	Km_r, protein under control of the T7 promoter	(Jentzsch *et al.*, 2009)
pET28a(+)SB2F1	Km_r, protein under control of the T7 promoter	(Krauss, unpublished)
pET28a(+)SB2F1-I66R	Km_r, protein under control of the T7 promoter	This work
pET28a(+)SB1F1	Km_r, protein under control of the T7 promoter	(Krauss, unpublished)
pET28a(+)W619_1-LOV	Km_r, protein under control of the T7 promoter	(Rani *et al.*, 2013)
pET28a(+)SBW25-LOV	Km_r, protein under control of the T7 promoter	(Rani *et al.*, 2013)
pET28a(+)Pf-5-LOV	Km_r, protein under control of the T7 promoter	(Rani *et al.*, 2013)

Table 8. Primers.

Name	Primers	Reference
SB2F1-I66R	forward 5'- GGATCACGACCAGCCGGGCCGTGCAATTATCCG -3' reverse 5'- CGGATAATTGCACGGCCCGGCTGGTCGTGATCC -3'	(Jentzsch *et al.*, 2009)
SB2F1-H61R	forward 5'- GTTTTCTTCAGGGCGAGGATCGTGACCAGCCGG -3' reverse 5'- CCGGCTGGTCACGATCCTCGCCCTGAAGAAAAC -3'	(Jentzsch *et al.*, 2009)

Culture media and stock solutions

Table 9. Culture media and stock solutions.

Medium	Recipe for 1 L	Medium / Stock Solutions	Recipe for 1 L
AI medium	50.4 g TB medium powder 4 mL Glycerol 50 µg/mL Kanamycin 0.5 g/L Glucose 2 g/L Lactose 50 µM Riboflavin*	TFB1	30 mM Sodium acetate 50 mM $MgCl_2$ 100 mM NaCl 10 mM $CaCl_2$ 15% Glycerin pH 6.0
SOB medium	20 g Tryptone 5 g Yeast extract 0.6 g NaCl 0.2 g KCl pH 7.0	TFB2	10 mM MOPS 75 mM $CaCl_2$ 10 mM NaCl 15 % Glycerin pH 7.0
LB Agar plates	5 g Yeast extract 10 g Tryptone 10 g NaCl 15 g Agar-Agar	M9 salt solution	80 g Na_2HPO_4 30 g KH_2PO_4 5 g NaCl 5 g NH_4Cl pH 7.2
LB medium	5 g Yeast extract 10 g Tryptone 10 g NaCl	Vitamin mix	1 µg/mL Choline chloride, Folic acid, Nicotinamide, D-Pantothenic acid hemicalcium salt, Pyridoxal hydrochloride
SeMet medium	100 mL M9 salt solution 50 mg/mL Kanamycin 0.4% (w/v) Glucose 0.3 mM $CaCl_2$ 1 mM $MgSO_4$ 1 µg/mL Biotinin 1 µg/mL Thiamine 50 µM Riboflavin* 20 mL Amino acid mix 1 mL Vitamin mix 10 mL Trace elements mix 20 µg/mL Adenine and Uracil	Trace elements mix	30 µM $FeCl_3$ 6.2 µM $ZnSO_4$ 0.76 µM $CuCl_2$ 0.42 µM $CoCl_2$ 1.62 µM H_3BO_3 0.2 µM $NiCl_2$ 0.2 µM Na_2SeO_3 0.2 µM Na_2MoO_4 81 µM $MnCl_2$ EDTA (5 mg/mL) at pH 7.5
		Amino acid mix	40 µg/mL all L-amino acids except Cysteine, Tyrosine and Methionine

* Riboflavin was first dissolved in 10 mL of 0.1 M NaOH and added fresh to the medium.

Software and graphical representations

Table 10. Stand-alone and web-based software.

Software Name	Web-site/reference
3V	http://3vee.molmovdb.org (Voss and Gerstein, 2010)
Anisotropic B-factors server	http://services.mbi.ucla.edu/anisoscale/ (Strong *et al.*, 2006)
ATSAS	http://www.embl-hamburg.de/biosaxs/software.htmL (Petoukhov *et al.*, 2012)
CASTp	http://sts.bioe.uic.edu/castp/ (Dundas *et al.*, 2006)
CCP4 Suite	http://www.ccp4.ac.uk/ (Winn *et al.*, 2011)
Coot (CCP4 Suite)	http://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/ (Emsley *et al.*, 2010)
Clustal Omega	http://www.ebi.ac.uk/Tools/msa/clustalo/ (Sievers *et al.*, 2011)
DSSP	http://swift.cmbi.ru.nl/gv/dssp/DSSP_5.html (Kabsch and Sander, 1983)
Gnuplot	http://www.gnuplot.info/ (Williams and Kelley, 2016)
Phenix	http://www.phenix-online.org/ (Adams *et al.*, 2010)
ProtParam	http://web.expasy.org/protparam (Artimo *et al.*, 2012)
Quantity One	Bio-Rad (Hercules, California, USA)
Situs	http://situs.biomachina.org/ (Wriggers, 2010)
UCSF-Chimera and ChimeraX	http://www.cgl.ucsf.edu/chimera (Pettersen *et al.*, 2004)
XDS	http://xds.mpimf-heidelberg.mpg.de/ (Kabsch, 2010)

If not stated otherwise, all plots were made with Gnuplot, and protein representations were made with UCSF Chimera or UCSF ChimeraX. Standardized atom-type coloring used in figures: carbon, yellow; nitrogen, blue; oxygen, red; phosphorus, purple; sulfur, orange; hydrogen, white. Secondary structure assignment for ribbon representation and in sequence alignments was done with DSSP program. The sequence alignment was made with the Clustal Omega program.

Equipment

Table 11. Equipment used.

Equipment name	Company; Location
ÄKTA pure system	GE Healthcare; Buckinghamshire, UK
Blue-light laser (440 nm, 15 mW, cw mode)	Laser Components GmbH; Olching, Germany
Blue-light LED (450 nm)	Luxeon Lumileds, Phillips; Aachen, Germany
GelDoc 2000	Bio-Rad; Hercules, California, USA
DLS SpectroSize 300	Xtal Concepts; Hamburg, Germany
NanoPhotometer	IMPLEN; Munich, Germany
UV-1800	Shimadzu; Duisburg, Germany
Microfluidizer M-110P	Microfluidics Corporation; Westwood, USA
HPLC system	Agilent; Santa Clara, California, USA

Kits

Table 12. Ready-to-use kits for molecular biological and biochemical procedures.

Kit Name	Company; Location
innuPREP plasmid mini kit	Analytik Jena; Germany
Bio-rad protein assay	Bio-Rad; Hercules, California, USA
QuickChange II XL kit	Agilent; Santa Clara, California, USA

2.2 Microbiological methods

Preparation and transformation of chemically competent cells

$CaCl_2$ protocol

Pre-cultures of strains frozen at -80 °C were grown in LB-medium overnight at 37 °C and shaken at a speed of 130 rpm. Cultures were inoculated in 300 mL SOB-Medium at an OD_{600nm} of 0.1, and subsequently incubated at 37 °C with shaking at 130 rpm until the OD_{600nm} reached 0.3 – 0.4. Cell suspensions were transferred into sterile 50 mL centrifuge tubes and incubated on ice for 15 min. Cell suspensions were centrifuged at 3 000 g for 15 min at 4 °C.

The cell pellet was resuspended in 60 mL TFB1 buffer and incubated for 15 min on ice. Subsequently, cell suspension was centrifuged at 3 000 g for 15 min at 4 °C. The cell pellet was resuspended in 12 mL TFB2 buffer and incubated for 15 min on ice. Afterwards 100 µL aliquots of the suspension were frozen in liquid nitrogen and stored at -80 °C.

Heat shock transformation method

The heat shock method was used to transform chemically competent cells with the plasmids carrying the DNA of the target protein. In brief, frozen competent cells were thawed on ice for 10 min. Next, 1 µL of the plasmid, with a DNA concentration of at least 50 ng/µL, was added to the cell suspension and was subsequently mixed gently. After incubation for 20 min on ice, heat shock was applied by incubation at 42 °C for 90 s. Cells were placed on ice thereafter. 1 mL of LB-medium was added to the cells, which were then incubated for 3 h at 37 °C while being shaken at 130 rpm. The cells were centrifuged at 3 000 g and the excess medium was removed. About 100 µL of the cell suspension were added to LB-agar plates supplemented with 50 µg/mL kanamycin, and incubated for 24 h at 37 °C. A single colony from the LB plates was used to inoculate 10 mL of LB-medium supplemented with 50 µg/mL kanamycin, and subsequently incubated overnight at 37 °C, with shaking at 110 rpm.

Protein overexpression

E. coli BL21 (DE3) cells were transformed according to the above described protocol using the heat shock method. The transformed cultures were used to inoculate autoinduction media (AI) overnight at an OD_{600nm} of 0.05. Overexpression was carried out in 250 mL AI media cultures for 3 h at 37 °C. Afterwards, the incubation temperature was changed to 30 °C and the cells were incubated for 24 h at constant agitation of 110 rpm.

For W619_1-LOV overexpression with low chromophore content, the incubation temperature was changed from 30 °C to 15 °C and incubated for 48 h. Additionally, the AI medium did not contain riboflavin.

Selenomethionine-labeled protein overexpression

The selenomethionine protein labeling protocol was adapted from (Doublié, 2007). A methionine auxotrophs strain *E. coli* B834 (DE3) was made chemically competent according to $CaCl_2$ protocol. Cells were transformed according to the heat shock method described previously and grown in 100 mL LB-medium, supplemented with 0.5% glucose at 37 °C and 130 rpm until an OD_{600nm} of 1.0 was reached.

To enable the cells to adapt to the new medium, the cells were first spun down and resuspended at an OD_{600nm} of 0.1 in 1 L SeMet medium supplemented with L-Methionine at a concentration of 50 mg/L (500 mL of suspension in 2 L Erlenmeyer baffled flasks) and grown at 37 °C while shaking at 130 rpm until an OD_{600nm} of 1.0 – 1.5 was achieved.

The cells were subsequently centrifuged at 3 000 g for 15 min and resuspended at an OD_{600nm} of 0.35 in SeMet medium and incubated for 6 h at 37 °C and 130 rpm in order to deplete remaining L-Methionine. Then L-Selenomethionine was added to a final concentration of 75 µg/mL and incubated with a gradual temperature reduction to 30 °C until an OD_{600nm} of 0.6 was reached and subsequently induced with 1 mM IPTG. After incubation for 4 h, the cells were harvested by centrifugation at 5 000 g. The wet cell pellets were frozen at -20 °C until further use.

2.3 Biochemical and molecular biological methods

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

For purity control of the protein solution during purification and molecular weight determination, SDS-PAGE was used under denaturation conditions (Laemmli, 1970). The protein solution was mixed with 4x SDS loading dye in proportions of 3:1 and subsequently incubated at 90 °C for 10 min. Power supply were set to the constant voltage of 340 V and running time was 35 min – 40 min.

Table 13. Recipes for SDS-PAGE components.

Solution	Recipe	Solution	Recipe
5% stacking gel	16.8% (v/v) Acrylamide 30% 125 mM Tris-HCl pH 6.8 0.1% (w/v) SDS 0.1% (v/v) APS 10% 0.1% (v/v) TEMED	4x SDS loading dye	20 mM Tris-HCl pH 6.8 8% SDS 8% β-Mercaptoethanol 40% (v/v) Glycerol 0.3% (w/v) Bromphenolblue
15% separation gel	50% (v/v) Acrylamide 30% 375 mM Tris-HCl pH 8.8 0.1% (w/v) SDS 0.1% (v/v) APS 10% 0.04% (v/v) TEMED	Coomassie stain	25% Isopropanol 10% Acetic acid 0.05% (w/v) Coomasie Brilliant Blue R250
1x SDS buffer	50 mM Tris-HCl pH 8.3 385 mM Glycerol 0.1% SDS		

Finally, the gel was stained with warm Coomassie blue stain (~60 °C) for 30 min. The unbound dye was removed by washing several times with hot water (~90 °C). Gel imaging was done with Bio-Rad Geldoc 2000. Molecular mass estimation was done with Quantity One software, based on the relative migration distance of the sample and a molecular weight marker.

Isolation of plasmid DNA

Chemically competent *E. coli* OmniMAX 2 T1R cells were transformed with target plasmid by the heat shock procedure. The transformed cells were used to inoculate a 10 mL culture in the LB-medium overnight (37 °C, 110 rpm). The cells were subsequently centrifuged for 10 min at 16 000 g. Plasmid DNA was isolated using "innuPREP plasmid mini kit" (Analytik Jena; Jena, Germany) according to the manufacturer's protocol.

Single amino acid mutagenesis

Single amino acid mutagenesis of the protein plasmid DNA sequence was done with the help of the QuickChange II XL kit. The standard protocol of the manufacturer was used. The resulting plasmid DNA sequences were verified by SeqLab company (Göttingen, Germany).

Protein purification

The cell pellet (5 g cells, wet weight) was dissolved in 30 mL of lysis buffer [50 mM NaH_2PO_4/Na_2HPO_4, 300 mM NaCl and 10 mM imidazole (pH 8.0)]. The cells were then lysed by passing the cell suspension through a Microfluidizer three times at a constant pressure of 2 000 bar. The soluble fraction was separated from the pellet by centrifugation at 50 000 g for 30 min at 4 °C. The soluble supernatant was purified by immobilized metal affinity chromatography (IMAC) using a Superflow Ni-NTA resin (QIAGEN; Hilden, Germany) with the gravity flow protocol. The standard column volume (CV) of the Ni-NTA resin was 5 mL. The lysate was applied to the resin and washed with 10 CV of lysis buffer and 5 CV of wash buffer [50 mM NaH_2PO_4/Na_2HPO_4, 300 mM of NaCl and 40 mM imidazole (pH 8.0)]. Elution was performed with 5 CV of elution buffer [50 mM NaH_2PO_4/Na_2HPO_4, 300 mM NaCl and 250 mM imidazole (pH 8.0)]. The purity of the eluted fractions was evaluated by SDS-PAGE.

The pure protein fractions were pooled together, and the elution buffer was exchanged to the storage buffer (Table 14), using ÄKTA pure FPLC system (GE Healthcare; Buckinghamshire, UK) with an HiPrep 26/10 Desalting column according to a standard protocol. The protein fractions were pooled together, supplemented with 3 mM TCEP and concentrated by ultrafiltration using Vivaspin centrifugal concentrator units (molecular mass cut-off: 10 kDa) (Sigma-Aldrich; St. Louis, USA). All steps were performed either at 4 °C or at 10 °C, with only the buffer exchange procedure done at room temperature. If not used within

2 days, the excess protein was supplemented with glycerol to final concentration of 40% and flash-cooled in liquid nitrogen and stored at -80 °C.

The protein solution was then loaded with the chromophore as described in next section. The only exception was the W619_1-LOV expressed at low temperature, which was not loaded with the chromophore.

Storage buffers

Table 14. Storage buffer at purification stage.

Protein	Buffer
SB1F1	20 mM Tris-HCl (pH 8.0) and 40 mM NaCl, 2 mM TCEP
SB2F1	
SB2F1-I66R	
SB1F1 SeMet	20 mM Tris-HCl (pH 8.0) and 40 mM NaCl, 4 mM TCEP
SB2F1 SeMet	
Pf-5-LOV	10 mM Tris-HCl (pH 8.0) and 10 mM NaCl, 2 mM TCEP
PpSB2-LOV	
SBW25-LOV	
W619_1-LOV	

Protein concentration and chromophore loading

Bio-Rad Protein Assay

For routine protein concentration determination, the commercially available Bio-Rad Protein Assay (based on the Bradford Assay) was used. In brief, 800 µL of diluted protein sample was mixed with 200 µL of dye reagent and incubated for 2 minutes at room temperature. The absorbance at 595 nm were recorded and interpreted according to the calibration curve made with bovine serum albumin. Measurements were done in duplicates and triplicates for protein concentration and calibration, respectively.

UV-Vis based concentration determination and chromophore loading (denatured protein)

The 7,8-dimethylisoalloxazine of the bound flavin chromophore displays absorption at 280 nm and its molar attenuation coefficient changes, which depends on whether the protein is in the light or dark state (see e.g. Figure 10). Moreover, the determined molar attenuation coefficient of a free chromophore is not necessarily the same as that of bound protein. This makes UV-Vis spectroscopy an unreliable method to estimate the protein concentration and chromophore loading concurrently.

To overcome these limitations the protein was denatured by mixing it with 6 M guanidine hydrochloride (in buffer 20 mM NaH_2PO_4/Na_2HPO_4 at pH 6.5) in 5:1 ratio. Due to the simultaneous release of the chromophore upon protein denaturation, the following calculations could be done. First, concentration of the chromophore was derived from its

absorption value at 447 nm and the FMN/RBF molar attenuation coefficient of 11 825 $M^{-1}cm^{-1}$ (determined in this work, see Figure 9).

Second, the absorption contribution of the chromophore at 280 nm was subtracted from the total absorption by using the FMN molar attenuation coefficient of 20 670 $M^{-1}cm^{-1}$ (determined in this work, see Figure 9).

$$C^{chrom}\,[M] = \frac{Abs_{447}\,[AU]}{\varepsilon_{447}^{chrom}\,[M^{-1}]} \qquad Abs_{280}^{chrom}\,[AU] = C^{chrom}\,[M] \cdot \varepsilon_{280}^{chrom}\,[M^{-1}]$$

By using the corrected absorption at 280 nm, the molar concentration of the protein was determined using the molar attenuation coefficient for the tryptophan, tyrosine and cystine amino acids in 6 M guanidine hydrochloride, 20 mM NaH_2PO_4/Na_2HPO_4 at pH 6.5: tryptophan – 5 690 $M^{-1}cm^{-1}$, tyrosine – 1 280 $M^{-1}cm^{-1}$, cystine – 120 $M^{-1}cm^{-1}$ from (Gill and von Hippel, 1989).

$$C^{prot}\,[M] = \frac{Abs_{280}^{total}[AU] - Abs_{280}^{chrom}[AU]}{\varepsilon_{280}^{protein}\,[M^{-1}]}$$

Then, chromophore loading was expressed as a ratio of the concentration of protein and chromophore in percent.

$$Load\,[\%] = \frac{C^{prot}\,[M]}{C^{chrom}\,[M]} \cdot 100\,\%$$

This method is similar to the previously described methods based on heat, pH change and surfactant-assisted protein denaturation (Aliverti, Curti and Vanoni, 1999). However, these denaturation methods can lead to protein loss and thus result in underestimation of protein concentration and chromophore loading. The method proposed herein overcomes this problem due to the high solubility of both protein and chromophores in 6 M guanidine hydrochloride.

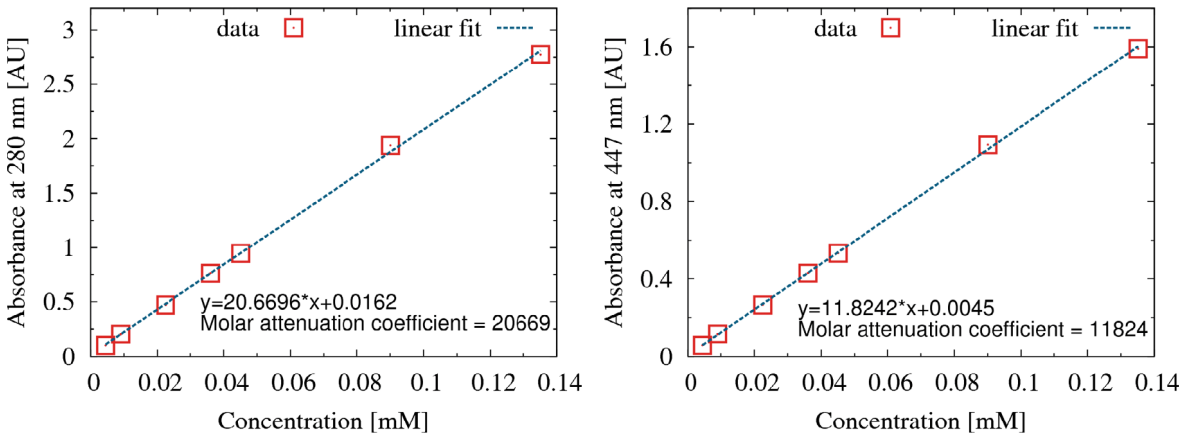


Figure 9. RBF attenuation coefficient calculation in 6 M guanidine hydrochloride, 20 mM NaH_2PO_4/Na_2HPO_4 at pH 6.5.

For the determination of the molar attenuation coefficients of the chromophore, RBF was used due to the similarity of absorption properties. RBF, FMN and FAD have slightly different absorbance peak maxima (447 nm, 446 nm, 450 nm respectively) which can contribute to error. However RBF and FMN are much closer to each other than FAD, and based on HPLC data of chromophore composition, FAD is present less frequently, and also in much smaller quantities (below 10%).

Loading proteins with natural and modified chromophores

In order to achieve higher chromophore loading *in vitro*, the protein sample was mixed with FMN to the final concentration of 10 mM and subsequently incubated overnight in the dark at 4 °C. Unbound chromophore was removed by size-exclusion chromatography (SEC) using a HiLoad 26/600 Superdex 200 pg column (GE Healthcare; Buckinghamshire, UK) on an ÄKTA pure FPLC system at room temperature. Eluted protein fractions were pooled, supplemented with 3 mM TCEP and concentrated to the desired protein concentration.

Apo W619_1-LOV protein was loaded with FMN, RBF, FAD, LC, 8-Cl-RF, 7-Br-RF and Roseoflavin in a similar manner.

Production of the Apo protein

In order to achieve the complete apo form of W619_1-LOV, the dark-adapted sample was denatured by dilution in 6 M guanidine hydrochloride, 20 mM NaH_2PO_4/Na_2HPO_4 pH 6.5 and washed a few times by subsequent dilution and concentration steps in Vivaspin centrifugal concentrator units (molecular mass cut-off: 3 kDa) to thoroughly remove the unbound chromophores. Refolding was done by rapid dilution with a storage buffer in two steps: up to ~1 M guanidine hydrochloride with 5 min incubation at room temperature, followed by dilution to ~0.1 M guanidine hydrochloride. The refolded protein was concentrated and the buffer was exchanged to 10 mM NaH_2PO_4/Na_2HPO_4 pH 8.0, 10 mM NaCl for spectroscopic studies.

2.4 Biophysical methods

UV-Vis and fluorescence spectroscopy

Static measurements

UV-Vis spectra of the protein solutions were measured with a Shimadzu UV-1800 spectrometer (Shimadzu; Kyoto, Japan). For generation of the light state protein, the sample was illuminated for at least 30 s using a blue-light (λmax = 450 nm) emitting LED with an approximate radiant flux of 50 mW (Luxeon Lumileds, Phillips; Aachen, Germany). All static measurements were done at 20 ± 2 °C in a 1 mL quartz cuvette with a 1 cm light path. The protein samples were in buffer with 10 mM NaH_2PO_4/Na_2HPO_4 pH 8.0, 10 mM NaCl.

Fluorescence measurements

Fluorescence spectra were measured with PerkinElmer LS 55 (PerkinElmer Inc.; Waltham, Massachusetts, USA). The excitation wavelength was set to 450 nm, with an excitation and emission slit width of 2.5 nm. For the W619_1-LOV protein loaded with LC, the excitation wavelength was set to absorption peak maximum at 420 nm. All fluorescence measurements were done at 20 ± 2 °C in 1.5 mL quartz cuvette with 1 cm light path and in buffer with 10 mM NaH_2PO_4/Na_2HPO_4 at pH 8.0, 10 mM NaCl. Samples were diluted to an approximate absorbance of 0.1 at the excitation wavelength in order to minimize inner filtering and re-absorption effects.

The relative fluorescence quantum yield was determined according to formula:

$$\Phi_F = \Phi_F^{ref} \frac{\eta^2 \, I \, A_{ref}}{\eta_{ref}^2 \, I_{ref} \, A}$$

where Φ_F^{ref} is the quantum yield of reference substance, η – the refraction index of the medium, A – the absorbance at the excitation wavelength and I – the integrated fluorescence intensity over the wavelength range. As a reference substance, RBF with a Φ_F of 0.27 was used, as reported (Drössler *et al.*, 2002, 2003).

Dark recovery kinetic measurements

All kinetic measurements were done with a Shimadzu UV-1800 spectrometer at 20 ± 2 °C. The protein samples were measured in a buffer containing 10 mM NaH_2PO_4/Na_2HPO_4 at pH 8.0, and 10 mM NaCl filled in 1 mL quartz cuvette with a 1 cm path length, if not stated otherwise. To generate the light state, the sample was illuminated with blue-light (λmax = 450 nm) emitting LED with an approximate radiant flux of 50 mW for at least 30 s (Luxeon Lumileds, Phillips; Aachen, Germany).

Dark recovery kinetics was then measured at 475 nm as a function of time. The raw data fitting was done with the Gnuplot program using the following exponential and biexponential decay functions:

$$Abs = a_0 + Ae^{-\frac{t}{\tau_{rec}}} \qquad Abs = a_0 + A_1 e^{-\frac{t}{\tau_{rec1}}} + A_2 e^{-\frac{t}{\tau_{rec2}}}$$

where t is the time since illumination stops, Abs – measured absorbance at 475 nm, τ_{rec} – is the time constant referred to as the adduct lifetime. The dark recovery data of the PpSB2-LOV and SB2F1 protein in the presence of ATP was a better fit with biexponential function, which indicated that the two processes occur with different time constants. It was previously reported that the biexponential function describes better experimental data in the case of some LOV proteins (Jentzsch *et al.*, 2009).

Crystal microspectrometry

Protein crystals grown in the dark were mounted in a loop under dim red-light and flash cooled with gaseous nitrogen at 100 K. Absorbance spectra in the wavelength range of 250 nm – 700 nm were recorded using an UV/Vis micro spectrometer at ID29S at ESRF (Grenoble, France) (von Stetten *et al.*, 2015). Fluorescence spectra were measured by excitation with a blue-light laser (440 nm, 15 mW, cw mode) while keeping crystals at 100 K. The same crystals were annealed to room temperature, illuminated for 10 s with the same laser and subsequently cooled to 100 K for spectra measurement.

Size-exclusion chromatography

Size-exclusion chromatography was used for molecular weight estimation. Superdex 200 10/300 GL and Superdex 200 Increase 10/300 GL columns (GE Healthcare; Buckinghamshire, UK) were used with an ÄKTA pure FPLC system. The standard running buffer was 20 mM Tris at pH 8.0, 150 mM NaCl. The flow rate was set to 0.5 mL/min.

HPLC-based determination of the chromophore composition

Protein chromophore composition and quantification before and after loading with FMN was done with HPLC (Agilent Technologies, Inc., Santa Clara, USA) according to a protocol described previously (Cao *et al.*, 2008).

In brief, the sample was incubated at 90 °C for 20 min to denature protein and release bound chromophores. Thereafter, the sample was centrifuged at 20 000 g, 40 °C for 15 min to remove denatured protein. The sample was next applied to Vivaspin concentrator units (molecular mass cut-off: 3 kDa) (Sigma-Aldrich; St. Louis, USA) to remove the rest of the denatured protein. The collected filtrate was directly applied to HPLC system equipped with 250/4 C18-RP column (Macherey and Nagel; Düren, Germany) to separate chromophores based on their hydrophobicity.

The following solvents were used: 50 mM ammonium acetate at pH 6.0 – eluant B and 70% acetonitrile in eluant B – eluant A. Following solvent gradient (t = 0 min, 5:95 A:B; t = 20 min, 40:60 A:B; t = 22 min, 40:60 A:B) was used. During gradient elution, absorption value at chromophore-specific wavelength at 447 nm was recorded. Calibration was done with commercially available pure RBF, FMN and FAD.

The chromophore composition was determined by integrating the area under each peak. Taking the total area of all peaks as 100%, the ratio of each single peak to the total area represents the chromophore content of the corresponding peak.

Small-angle X-ray scattering

Small-angle X-ray scattering (SAXS) is a technique based on the elastic scattering of a sample at small angles. The SAXS of protein solutions produce isotropic scattering, which can be averaged to the intensity I against the scalar value of the scattering vector q. The scattering

vector $\boldsymbol{q}$ is the vector difference between incoming X-rays and scattered X-ray vectors, which have the origin at the sample position and the end at the detector surface. If the scattering angle is defined as 2θ, then the $\boldsymbol{q}$ scalar value is defined as $4\pi sin\theta/\lambda$. Analysis of SAXS data can determine following properties: molecular weight, the oligomeric state, the distance distribution function, the radius of gyration as well as enable the calculation of *ab initio* envelope models.

Data collection and processing

SAXS data for all proteins were collected at beamline BM29 at the European Synchrotron Radiation Facility (ESRF) (Grenoble, France) (Pernot *et al.*, 2013) with the exception of SBW25-LOV, which was measured at beamline P12 at PETRA III, DESY (Hamburg, Germany) (Blanchet *et al.*, 2015). The collected data was evaluated by programs from the ATSAS software package (Petoukhov *et al.*, 2012)

Table 15. List of SAXS data collection on proteins in this study with experimental details. SAXS data was collected at 10 °C. All samples except SBW25-LOV were measured both under light and dark conditions.

Protein	Wavelength [Å]	Beamline	Buffer	Concentrations [mg/mL]
Pf5-LOV	0.992	BM29	10 mM NaH_2PO_4/Na_2HPO_4 pH 8.0, 10 mM NaCl	2.5; 5.0; 10.3
PpSB2-LOV			10 mM NaH_2PO_4/Na_2HPO_4 pH 8.0, 10 mM NaCl	2.2; 4.8; 10.3
SB1F1			20 mM Hepes pH 7.5, 150 mM NaCl, 2% Glycerol, 2 mM TCEP	0.9; 1.7; 3.2
SB2F1			20 mM Hepes pH 7.5, 150 mM NaCl, 2% Glycerol, 2 mM TCEP	1.0; 1.6; 2.9; 4.9
SB2F1			20 mM Hepes pH 7.5, 150 mM NaCl, 2% Glycerol, 2 mM TCEP, 1.5 mM ATP, 3 mM $MgCl_2$	1.0; 1.6; 2.9; 4.9
SBW25-LOV	1.240	P12	20 mM Hepes pH 7.5, 150 mM NaCl, 2% Glycerol, 1 mM TCEP	0.7; 1.2; 2.6; 3.9; 6.6
W619_1-LOV	0.992	BM29	10 mM NaH_2PO_4/Na_2HPO_4 pH 8.0, 10 mM NaCl	3.0; 5.1; 9.1

The reference measurement of the protein buffer was done before and after each protein sample was measured. During X-ray exposure, samples were continuously flowing through a quartz capillary to minimize radiation damage that can occur during the measurement. Probes with proteins in the dark state were kept in the dark prior to measurement at least three times longer than the adduct lifetime, and the whole SAXS setup was under dim red light. Light state samples were first illuminated with a blue-light (λmax = 450 nm) LED in the sample chamber. Additionally, the capillary containing the sample was continuously illuminated with blue-light LED (at BM29 beamline) or laser (λmax = 450 nm) (at P12 beamline) until right before X-ray exposure to maximize the light state population. The temperature for data collection was set to 10 °C to slow down dark recovery, thus

enabling more efficient population of the light state. For each sample, ten frames with an exposure time of 1 s each were recorded. Frames were checked by DATCMP (ATSAS package) for any radiation damage before being merged.

The measured intensity I was calibrated with measurement of pure water, verified by measurement of bovine serum albumin (BSA) protein molecular weight, and then normalized by the dividing to the concentration of the sample.

Ranges of q with higher signal-to-noise ratios from different concentrations were combined, if necessary. The AUTORG program, which is based on the Guinier approximation (Guinier, 1939) was used to estimate the radius of gyration Rg and to determine the I_0 value. Here R_g essentially describes the root mean square distance of the atoms in a protein relative to the protein center of mass. Whereas I_0 shows molecular weight of measured protein if it is calibrated to molecular weight determination as mentioned above.

The distance (or pair) distribution function $p(r)$, describing the statistical distribution of distances between each part of the protein molecule (in this case scattering atoms), was calculated with the AUTOGNOM program. The longest dimension of the molecule D_{max} can be determined for further analysis. Additionally, experimental SAXS data was compared to theoretical scattering curves of the crystal structures calculated by the program CRYSOL.

Ab initio models generation with GASBOR and DAMMIN

Ab initio bead modeling was done with the GASBOR and DAMMIN programs. In GASBOR, the initial model is randomly generated out of dummy beads connected to each other, at average protein residue-to-residue distance. The number of beads is equal to the number of protein residues, where each bead scatters as if it were an average residue. Additionally, the model is surrounded by solvent dummy beads and the system is constrained to a maximum size D_{max}. In DAMMIN, the initial model is randomly generated out of densely packed dummy beads for protein and solvent molecules in a sphere of size D_{max}. The modeling consists of simulated annealing as a global minimization method to find model which agree best with the measured scattering data or distance distribution function.

In both GASBOR and DAMMIN modeling, additional restraints are imposed to improve, for example, the connectivity and compactness of the model. In cases where the symmetry of the oligomer is known, it can be added as an additional restraint to the model.

In both cases, 50 *ab initio* models were generated, aligned and averaged to the most probable model. No significant difference was observed with and without two-fold symmetry restraint for short LOV proteins. The crystal structures were then aligned with the averaged model. Averaged bead models were transformed to envelope models for graphical representation with the Situs program package (Wriggers, 2010).

2.5 Protein crystallography

Crystallization methods

Protein crystallization can be achieved with various methods with the aim to initiate the phase transition of the protein solution into crystalline form. For this, one can exploit dependence of protein solubility on protein concentration, temperature, pH, ionic strength, specific solutes etc. Decreasing the protein solubility should take place in such a way to avoid nonspecific aggregation and to promote the ordered formation of crystals.

The most popular and easy to handle method of crystallization is based on vapor diffusion. The crystallization solution is composed of precipitating agent, buffering molecules to maintain specific pH and salt or solutes aimed to decrease protein solubility and promote crystallization. A drop of the protein solution is mixed with the crystallization solution in a closed system and connected through the vapor phase to a reservoir with a sufficiently larger volume of pure crystallization solution. The equilibration between them leads to an equal concentrations of crystallization solution components in both compartments, and a gradual increase in protein concentration. Increased concentration of protein and crystallization agent intrinsically decreases amount of free water molecules, leading to an effect similar to that as "salting out", which promotes crystallization.

The vapor diffusion crystallization setup can be done as a sitting drop or as a hanging drop. In the sitting drop setup, a drop is settled on top of the bridge which separates it from the reservoir containing the crystallization solution and sealed. In the hanging drop setup, a drop is settled on a glass slip pointing inward to the reservoir containing the crystallization solution, which simultaneously seals the apparatus. Each of the two setups has advantages and disadvantages depending on the application for which they are used.

Crystallization setups

For the initial crystallization screening, the protein samples were concentrated in their respective storage buffers (Table 14) until they reached concentrations in the range of 5 – 10 mg/mL or 20% below the concentration at which the protein starts to precipitate. The screening was done with commercial screens (Table 4) in 96-well sitting drop vapor diffusion plates (Greiner Bio, Kremsmünster, Austria). The protein solution was mixed in equal ratios with the crystallization solution, typically 0.7 µL and 0.7 µL, respectively, and equilibrated against the 70 µL crystallization solution contained in the reservoir. For crystallization under dark conditions, plates were wrapped in aluminum foil, whereas under light conditions the plates were constantly illuminated with blue-light LED arrays (λ_{max} = 450 nm, Luxeon Lumileds, Phillips, Aachen, Germany). The final optimized crystallization conditions for all studied proteins are listed in Table 16.

Screening plates were regularly examined under microscope to determine their outcome. The observation of crystallization plates kept under dark conditions was done under dim red light, and the microscope illumination was filtered to red light only. The conditions yielding protein crystals were refined by varying the concentration of the precipitant and other components such as the buffer pH, protein concentration or the temperature of the crystallization setup. Additionally, micro-seeding techniques were used to produce better crystals, where initial crystals were crushed with glass beads and added to crystallization drops as nucleants.

Table 16. Best crystallization conditions for proteins from this study.

Protein	Concentration [mg/mL]	T [°C]	Crystallization conditions	Cryo-protection	Remarks
Pf-5-LOV	15	14	1.0 M – 1.2 M $(NH_4)_2SO_4$, 0.1M MES (pH 6.0 – 6.3)	+ 25% Glycerol	–
PpSB2-LOV	8	19; 14	22% – 26% (w/v) PEG 200, 3% – 7% (w/v) PEG 3350, 0.1 M Sodium Acetate (pH 4.9 – 5.3)	–	Micro-seeding
SBW25-LOV	8	21	10% – 12% PEG 3350, 0.1 M MES (pH 6.0 – 6.3)	+ 25% PEG 3350	–
W619_1-LOV	32	14	1 M $(NH_4)_2SO_4$, 0.1 M MES (pH 6.0 – 6.3)	+ 35% Glycerol	–
SB1F1 dark	4.5	19	18% – 22% PEG 3350, 0.2 M $LiSO_4$, 0.1 M Bicine (pH 9.0 – 9.3)	+ 15% PEG 200	Micro-seeding
SB1F1 light	3.5	21	20% – 24% PEG 3350, 0.2 M $LiSO_4$, 0.1 M Tris (pH 7.3 – 7.7)	+ 10% PEG 200	Micro-seeding
SB2F1[1]	4	19	10% – 12% PEG 8000, 0.2 M NaCl, 0.1 M Sodium Citrate (pH 5.8 – 6.1)	+ 20% PEG 200	Micro-seeding
SB2F1-I66R[2]	4	14	9% – 12% PEG 8000, 0.2 M NaCl, 0.1 M Sodium Citrate (pH 5.7 – 5.9)	+ 20% PEG 200	Micro-seeding

[1] Co-crystallized with ligand combinations of ATP/$MgCl_2$, ADP/$MgCl_2$, AMP-PNP/$MgCl_2$, and ATP-γ-S/$MgCl_2$. [2] Co-crystallized with ATP/$MgCl_2$.

Cryo-protection and crystal mounting

Protein molecules inside crystals absorb X-ray photons during X-ray exposure, which can cause radiation damage. The radiation damage can be divided into primary and secondary radiation damage. The primary damage is a result of the ionizing power of X-rays, which produces electrons via the photoelectric effect. The secondary damage is a consequence of electrons interacting to produce free radicals and bond breakage. Altogether this can result e.g. in C–S or S–S bond breakage or the decarboxylation of aspartic or glutamic acid residues. More generally, radiation damage reduces resolution and data quality through destruction of lattice order.

In order to protect crystals from X-ray induced radiation damage during data collection, crystals can be cooled to 100 K, which slows the diffusion of the free radicals. The crystallization solution, however can become crystalline that could also diffract X-rays which would interfere with the protein crystal diffraction. To overcome this problem, crystallization

solution is supplemented with cryo-protectant, which prevents formation of the crystal phase and traps the solution in the disordered amorphous phase. If the crystallization solution contains glycerol, ethylene glycol or short chain PEGs (100 Da – 3 500 Da), their concentration can be increased up to 30% – 35% to act as a cryo-protectant. However, crystallization solutions containing long chain PEGs or salt-based precipitating agents (like ammonium sulfate) needs to be supplemented with either the above-mentioned compounds or with other cryo-protectants such as sucrose or special oils. Cryo-protection can be introduced in two separate ways: either the crystal is subjected to cryo-protection solutions with incrementally increasing concentrations or it is submerged into the final cryo-protection solution and flash-cooled to 100 K. Optimal compositions of cryo-protecting solution and protocols that demonstrate minimal damage to the protein crystal itself needs to be tested individually by trial and error.

X-ray diffraction

A crystal is a solid material which is formed by the periodic arrangement of atoms, ions, molecules or even macromolecules in a crystal lattice. Due to its periodicity, the crystal lattice can be described by its smallest group, called a unit-cell, which repeats itself through translation in all three dimensions. The unit-cell is defined by three length dimensions along the principal axes (**a**, **b**, **c**) and the angles between them (α, β, γ). Triclinic crystal systems are the most general unit-cell (Table 17). There are seven crystal systems in three-dimensional space, which are special cases of the triclinic cell. In addition to the seven crystal systems, centered non-primitive Bravais lattices can be formed. These include body-centered (I), face-centered on C (the plane formed by **a** and **b**) (C), face-centering of all three faces (F) and rhombohedral-centering (R).

Table 17. Crystal systems and their cell geometry.

Crystal systems	Cell geometry	
triclinic	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma \neq 90°$
monoclinic	$a \neq b \neq c$	$\alpha = \gamma = 90° \beta \neq 90°$
orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90°$
tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90°$
trigonal	$a = b \neq c$	$\alpha = \beta = 90° \gamma = 120°$
hexagonal	$a = b \neq c$	$\alpha = \beta = 90° \gamma = 120°$
cubic	$a = b = c$	$\alpha = \beta = \gamma = 90°$

Within the unit-cell, elements are related by symmetry operations. These symmetry operations are: identity, translation, reflection through a plane, inversion at a point, rotation around axis (360°/n) and combinations of thereof. This internal symmetry allows then to reduce the description of a unit-cell to a so-called asymmetric unit.

The 14 Bravais lattices combined with symmetry operations lead to 230 space groups which fully describes symmetry. However, due to asymmetric and the chiral nature of typical

protein amino acids, the reflection through a plane and inversion at point symmetry operations are not allowed. This leads to the reduction of the possible space groups for protein crystals from 230 to 65.

X-ray scattering from the electron clouds of atoms or molecules in crystal lattices result in constructive or destructive interference. The resulting interference can be explained by Bragg equation, which describes the scattering taking place on lattice planes instead of single atoms. The Bragg equation is:

$$n\lambda = 2dsin\theta$$

where n is an integer, λ – the wavelength, d – spacing between the planes, θ – diffraction angle between the scattered light vector and the plane, that describes conditions for constructive interference.

Diffraction patterns can be described with a mathematical construct called a reciprocal lattice, where each diffraction spot is the reciprocal of the point in real space. The reciprocal lattice is thus a Fourier transformation of the crystal lattice, where the lattice basis vectors are perpendicular to the crystal lattice. Each reflection in reciprocal space is indexed by the Miller indices: h, k, l.

The structure factor, $\boldsymbol{F}(hkl)$, for each point in reciprocal space mathematically describes the amplitude and phase of the diffracted X-rays. With the help of Fourier transformation, from known $\boldsymbol{F}(hkl)$ one can calculate the electron density at each given x, y, z grid point in real space by summation over all Miller indices h, k, l

$$\rho(x,y,z) = \frac{1}{V}\sum_{h=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\boldsymbol{F}(hkl)e^{-2\pi i(hx+ky+lz)}$$

$$\boldsymbol{F}(hkl) = F(hkl)e^{i\alpha(hkl)}$$

In an experiment, the intensity of X-rays $I(hkl) = F^2(hkl)$ are measured, which lack phase information α for each reflection, leading to the so-called phase problem in crystallography. To reconstruct an electron-density map one needs to derive the phase information by additional procedures or experiments. For instance, phasing with the help of molecular replacement procedure or anomalous scattering experimental phasing can reconstruct the missing phase information. These procedures will be described next, in the structure solution and refinement section.

Data collection

The collection of X-ray diffraction data was performed at synchrotron ESRF in Grenoble, France at various beamlines (ID23-1, ID23-2, ID29, ID30B, ID30A-3). Typically, the monochromatic X-ray wavelength was around 1 Å, where synchrotron X-ray sources have

maximum flux. For the collection of anomalous scattering from heavy-atom-labeled protein crystals, the wavelength of the heavy-atom-label X-ray absorption edge was determined with the help of X-ray fluorescence measurements at the beamline. The collection of X-ray diffraction data at this wavelength provide maximum anomalous scattering and is sufficient for single-wavelength anomalous diffraction (SAD) experiments. Multi-wavelength anomalous diffraction (MAD) experiments can be done when SAD experiments do not yield sufficient initial phases for structure solution. For this, at least two more datasets can be determined at wavelength where absorption edge has inflection point, and another one far after the absorption edge.

Single crystals frozen in a loop were mounted on a spindle under a stream of gaseous nitrogen at a temperature of 100 K. Initially, the crystal was centered with respect to the X-ray beam spot and if possible the beam size was adjusted to match crystal size. In order to sufficiently depict a three-dimensional reciprocal lattice onto a two dimensional detector, at least two single images separated by a $\Delta\varphi_{sep}$ = 90° rotation around the spindle axis were taken for calculation of data collection strategy and for diffraction quality estimation.

Images were automatically indexed and various parameters for the radiation dose-optimized data collection were determined by the BEST program which, routinely uses HKL, XDS, MOSFLM and Labelit programs (Otwinowski and Minor, 1997; Bourenkov and Popov, 2010; Kabsch, 2010; Battye *et al.*, 2011). The determined point group and crystal lattice orientation implies how much of the three-dimensional reciprocal space needs to be measured to reconstruct the whole diffraction pattern with help of the symmetry. The completeness shows how many unique reflections were observed in comparison to the theoretically possible reflections for a given point group and the size of the unit-cell at a particular resolution. Moreover, the intensity of diffracted X-rays is evaluated for the signal-to-noise ratio $I/\sigma(I)$, which is an intensity of the reflection to the standard deviation among several observations of this unique reflection.

Using these and additional parameters, the BEST program finds the optimal strategy to collect data without introducing significant radiation damage to the protein crystals. Additionally, if the crystal was larger than the X-ray beam size, it was translated along a defined path during X-ray exposure to collect data from unexposed parts of the crystal.

In the presence of anomalous scattering, Friedel's law becomes invalid and the Friedel reflection pairs $I(hkl)$ and $I(\text{-}h\text{-}k\text{-}l)$ are no longer equal. For this reason, the SAD or MAD data collection strategy produced by BEST takes into account measurement of all Friedel pairs. The intensity differences between the reflections of Friedel pair essentially carry information necessary for the initial phase calculation. As X-ray exposure at absorption edge of anomalous scatterer could intrinsically increase radiation damage, a lower radiation dose is used for the SAD or MAD data collection strategy.

Data processing

The collected X-ray diffraction images were processed with the XDS program (Kabsch, 2010). The beamline instruments parameters were used for data processing, like the detector to crystal distance, wavelength λ, starting angle φ, rotation angle step $\Delta\varphi$ and the various setup parameters for the correction of data distortion. After the subtraction of the average background signal, sets of images separated by a $\Delta\varphi_{sep}$ = 90° degrees were used to extract strong reflections with high signal-to-noise ratios $I/\sigma(I)$. These were used to determine possible unit-cell dimensions and space groups, alternative indexing origins and crystal orientation in respect to the X-ray beam. After successful determination of these, reflection intensities from all images were integrated. However, if no strong evidence was available for a specific unit-cell and space group, reflection intensities were integrated in the P1 space group. After integration of reflections, the unit-cell parameters and space group choice were evaluated and eventually refined based on the whole dataset.

Further on, the data was evaluated with the POINTLESS program (CCP4 suite) to find the final space group with the highest symmetry based on the integrated intensities. The POINTLESS program checks systematic absence of reflections that are specific for a certain space groups. Some space groups cannot be distinguished based on the systematic absence of reflections, for instance, the $P3_121$ and $P3_221$ space groups. Here, data evaluation and solution of the structure were attempted in both space groups, and the correct one was chosen after the phasing step.

After determining the final space group, Friedel pairs (not merged for anomalous data) and symmetry equivalent reflection were merged and scaled to average intensities with standard deviations, with the help of the AIMLESS program (CCP4 suite). Furthermore, based on the statistics calculated with AIMLESS, a resolution cut-off was chosen, which took into account: $I/\sigma(I)$, completeness and $CC_{1/2}$ values of the highest resolution shell. The $CC_{1/2}$ value is particularly helpful as it estimates non-randomness of observed intensities and is calculated as a correlation of reflections intensities observed several times between two parts of randomly split dataset.

In the case of anisotropic data, where reflections along the crystallographic axes don't have an equal $I/\sigma(I)$ distribution, a compromise between $I/\sigma(I)$, completeness, and $CC_{1/2}$ along these axes was chosen. For anomalous data, besides the above-mentioned values, the $CC_{1/2}$ correlation of anomalous signal was used to judge the resolution to which the anomalous signal extends.

Structure solution and refinement

To obtain the initial phases for structure solution either molecular replacement procedure or experimental anomalous data phasing was performed in this work. Molecular replacement procedure was sufficient to solve short LOV protein structures.

The crystal structure of a close homologue can serve as a search template for molecular replacement. For this, homology modeling is first performed, which includes sequence alignment, respective amino acids exchanges with subsequent energy minimization. If a homologue structure contained more than one protein chain in the asymmetric unit, a biologically relevant assembly (e.g. dimer) was used as a search model.

Molecular replacement essentially consists of a Patterson map calculation that does not require phase information. Patterson maps calculated from experimental data, which show peaks for each vector between X-ray scatterers, cannot be used to solve protein structure directly. Instead, the Patterson map from experimental data is evaluated for a correlation with the Patterson map of the search template. For this, the search template is translated and rotated within unit-cell until the correlation becomes significant between Patterson map of the search template and that of experimental data. Once this is successful, phases calculated from newly positioned search template are combined with experimental structure factors.

Molecular replacement phasing for short LOV proteins was done with the MOLREP program (Winn *et al.*, 2011) by using the already determined crystal structure of a close homologue PpSB1-LOV (PDB ID: 3SW1). The success of molecular replacement phasing is first evaluated by R factors, where R factors values lower than 60% indicate that a solution is not random (Wilson, 1950). R factors are generally used in crystallography to measure the agreement between the data and the model, which are calculated in the following way:

$$R = \frac{\sum_{hkl}|F_{obs}(hkl) - F_{calc}(hkl)|}{\sum_{hkl} F_{obs}(hkl)}$$

where F_{obs} and F_{calc} are observed structure factors and that calculated from model, respectively, which are summed over all reflections. In order to prevent over-fitting of the data, approximately 5% of randomly chosen reflections were excluded from any refinement and denoted as a free set, with respective R_{free}, whereas the rest as a working set with respective R_{work}.

After estimation of the initial phases based on the search model data, $2mF_o$-DF_c and mF_o-DF_c maps were calculated and inspected for plausibility. The $2F_o$-F_c map, is a difference electron-density map where, F_{calc} factors of the model are subtracted from twice F_{obs} and multiplied by the calculated model phases. Such a map not only shows the electron density for the model but also for the missing parts of the model. Whereas the F_o-F_c map only shows the negative and positive differences between maps based on the F_{obs} and F_{calc} factors. Several corrections are made to amplify the density of the missing parts of the model, to suppress noise in the data, and to reduce bias. First, calculated phases are weighted by the so-called figure of merit (denoted as "m" in the map), which express the statistical probability of phase error. Second, F_{calc} are weighted by the sigma A factor (denoted as "D" in map) which accounts for errors and missing parts of the model based on the statistical probability

distribution. Finally, the $2mF_o$-DF_c, map DF_c can be used instead of the unobserved F_{obs}, which leads to clearer maps. However, the substitution of F_{obs} with DF_c should be used with care due to the bias introduced by the model, especially when the data has a completeness lower than 70% – 80%.

In the case of the SB2F1 and SB1F1 proteins, the molecular replacement procedure has failed to find a structure solution. In order to do SAD or MAD experimental phasing, the SB2F1 and SB1F1 proteins were over-expressed with selenomethionine and subsequently crystallized. The collected SAD or MAD data was used to calculate the Patterson map. The Patterson map contains not only peaks corresponding to the vectors between protein atoms (N, C, O and S) but also peaks with higher intensity corresponding to the anomalous scatterers (i.e. selenium atoms (Se) of selenomethionines). The limited number of vectors corresponding to the anomalous scatterers allows then to find their position inside the unit-cell and obtain initial phase information.

The initial phases may be inaccurate because of experimental errors or low signal-to-noise ratios, therefore, it is usually necessary to improve the phases by using prior knowledge about the electron density distributions in the crystals. Such procedure can be a density modification aimed to enhance protein signals in maps, or phase extension procedure which statistically guesses the missing phase information of reflections at higher resolution that initial phases. All these steps starting with initial phases calculation, density modification, refinement and phase extension procedures were done with the SHELX programs (C / D / E) (Sheldrick, 2010).

Based on the $2mF_o$-DF_c and mF_o-DF_c maps, the model is further improved by making manual corrections and model building using the COOT program (CCP4 suite). The improved model was then subjected to refinement with the Phenix software package (Adams *et al.*, 2010).

In most cases, several cycles of manual building/rebuilding and refinement were done until the model could not be further improved. Depending on the resolution, only an appropriate amount of parameters were refined to avoid data over-fitting, which is further monitored during refinement as a difference between the R_{work} and R_{free} factors. For instance, for structures with the resolution lower than 3.0 Å, group B-factors (e.g. one B-factor per residue) were used, whereas for structures with resolution higher than 3.0 Å individual B-factors were used.

In the first refinements runs, regardless of the data resolution, the refinement strategy employed by the Phenix software was the rigid body fitting. The rigid body fitting refines the position of the whole model or its parts such as monomers of an oligomer. Additionally, simulated annealing steps were added to overcome phase bias introduced from the phases of the search template, and to increase the radius of refinement convergence. In all refinement

runs, data was fitted in both real and reciprocal space together with the B-factor refinement. In cases where multiple side chain conformations of amino acids were observed, relative occupancies were also refined.

Quality assessment of structure

During successful refinement the R_{work} and R_{free} factors decrease, however, difference between R_{work} and R_{free} should not be more than ~5% to avoid data over-fitting. In addition, the model quality was assessed with the Molprobity program (Chen *et al.*, 2010) implemented in the Phenix software package and by evaluation tools in the COOT program. This included but was not limited to the following parameters: bond and angle deviations from their respective ideal values, a Ramachandran plot, deviations from common sidechain rotamers and too close contacts of atoms. Additionally, these parameters were compared to the average values from the structures of published proteins at similar resolution.

2.6 Quantum chemical calculations

Covalent adduct formation between FMN and conserved cysteine residue in LOV protein was predicted and observed in several crystal structure of LOV proteins as a continuous electron density between the Sγ atom of Cys53 residue and the C4a atom of FMN (Crosson and Moffat, 2002; Circolone *et al.*, 2012). Such continuous electron density was also observed in the light state crystal structure of SBW25-LOV determined in this work.

However, an electron-density map at resolution of a typical protein structure (current mean resolution in PDB databank is 2.2 Å) does not allow the exact positions of atom to be determined without prior knowledge of the structure of proteins. In the absence of an accurate geometry of the ligand molecule, neither can the electron density of the ligand in protein crystal structures be well-interpreted nor can the molecular dynamics simulations be done. Unfortunately there is no experimental data with atomistic resolution (higher than ~1 Å) on the geometry of the FMN-cysteinyl adduct where the FMN-C4a atom becomes sp^3-hybridized. Sp^3-hybridization should also result in the distortion of the planarity of the FMN tricyclic ring system. To overcome lack of experimental data, similar structures can be used to construct the model by analogy or even be combined from different structures. This initial model can be then subjected to the quantum chemical geometry optimization procedures. In order to perform such geometry optimization, first the atoms must be described at the quantum mechanical (QM) level of theory. Generally, to calculate the quantum state of a system, the Schrödinger equation must be solved. The simple one-dimensional Schrödinger equation for a stationary state is:

$$\hat{\mathrm{H}}\Psi(\mathrm{x}) = \mathrm{E}\Psi(\mathrm{x})$$

where $\hat{\mathrm{H}}$ is the Hamiltonian operator, $\Psi(\mathrm{x})$ is the one-dimensional wave function and E is the total energy.

Once the Hamiltonian operator of the system is known, the wave function that satisfies Schrödinger equation can be found, and the energy of the system can be calculated. Analytical solution, however, is only possible for a single particle. In order to solve the equation for a simple hydrogen atom containing one proton and one electron, the Born-Oppenheimer approximation needs to be used. The approximation assumes that the positions of the nuclei with respect to the electrons are stationary, due to their much higher mass and enables them to be treated separately by using two separate Schrödinger equations.

To treat multi-electrons systems, orbital approximations can be made, where the multi-electron function is treated as a linear combination of single electron functions (basis function), which are called as atomic or molecular orbitals (AO or MO). One way of constructing MO functions is based on the assumption that MO can be represented as a linear combination of AO basis functions. Intrinsically, a higher number of basis functions (called a basis set) make the MO description better. The choice of a larger basis set also increases computational costs and thus its size needs to be appropriately chosen for a particular system. In this study the 6-31G* basis set was used, where the basis functions are comprised of Gaussians functions, which includes six basis function to describe the core orbitals, whereas valence p orbitals are split into three and one functions, and supplemented with one d orbital polarization function.

The implementation of the Hartree–Fock method of the wave function approximation allows solution to the Schrödinger equation for multi-electron systems to be approximated. The assumption in the Hartree–Fock method that electrons do not interact with each other, however, can lead to the significant deviation from experimental results and underestimation of the energy. To improve the energy estimation, the excluded electron correlation terms are approximated, by the density functional theory (DFT). The combination of the Hartree–Fock method with the approximation of the electron correlation term from DFT leads to the hybrid functional method. One such hybrid functional method is B3LYP, which approximates the electron correlation term, based on the empiric and *ab initio* data.

Given proper energy estimation, then the energy can be expressed as a function of atomic coordinates. The geometry optimization is then a general energy minimization problem. Iteratively the energy and the respective forces on each atom are calculated, then atoms are displaced in order to minimize the energy and the respective forces. The geometry optimization of the molecule is converged to the energy minimum when change of the energy after iteration is less than some threshold.

In order to obtain the optimized geometry of the FMN covalently bound to the cysteine residue, several cycles of model building and QM geometry optimization was done. First, the initial structure was constructed from a similar crystal structure of 4a-Isopropyl-3-methyl-4a,5-dihydrolumiflavin ($C_{17}H_{22}N_4O_2$) (CCDC ID: 1180634) (Bolognesi, Ghisla and

Incoccia, 1978) (Figure 24). In this molecule, the isopropyl group, which forms a covalent bond to the C4a atom, were exchanged to a methylsulfanyl group with an initial S-C4a distance of 1.82 Å. Additionally, a 3-methyl group was removed, as it is not present in FMN. Second, resulting lumiflavin C4a-S-CH_3 structure was then subjected to QM geometry optimization at the B3LYP/6-31G* level of theory in implicit water using the GAMESS software (Gordon and Schmidt, 2005).

Next, the ribityl chain and the phosphate group of FMN were added to the model in the same conformation as in the crystal structures of PpSB2-LOV and PpSB1-LOV. Thereafter, the cysteine residue was completed and the N, C, O, Cα, and Cβ atoms of two adjoining dummy amino acid residues at the N- and C-terminus were added. The addition of these two residues was necessary to exclude the influence of the free amide and carboxylic groups of cysteine from the calculations. Geometry optimization at the same level of the theory yielded the final model of the cysteine FMN adduct, which was used for crystal structure refinement and molecular dynamics simulations.

Additionally, for molecular dynamics simulations, partial charges were derived by restrained electrostatics potential fitting of the QM data with RESP program implemented in the AmberTools16 program package (Case *et al.*, 2016). In a similar manner, the FMN of the dark state structure was also first QM optimized and the partial charges were derived from QM calculation.

2.7 Molecular dynamics simulation

A wide array of LOV proteins were crystallized in the dark and light states and only few of them were characterized by NMR. The protein dynamics of crystal structures, in contrast to NMR, often remain elusive, as structural elements are sometimes stabilized by crystal contacts. Molecular dynamics (MD) simulation can supplement crystal structure with information about protein dynamics. The dark recovery process in LOV proteins are known to be thermally driven, thus MD simulation can give valuable dynamics information on the transition from the light to the dark state.

The exploration of the conformational space and dynamics of proteins can be performed with MD simulations. For systems as large as macromolecules, it is currently not possible to perform QM calculations. Instead the atomic and molecular interactions are parameterized based on the experimental data and *ab initio* calculations. This set of parameters is called a force field, which includes both bonded and non-bonded interactions. Bonded interactions are defined via bond lengths, bond angles, dihedral angles and partial charges with respective potentials. Non-bonded interactions are described as Coulomb-type electrostatics and as Lennard-Jones potentials, which include Pauli repulsion and van der Waals attraction forces. Generally, force fields for protein simulations have improved, although some seems to agree better, for example, with experimental NMR data (Beauchamp *et al.*, 2012; Lindorff-

Larsen *et al.*, 2012). To explore dynamics of molecules, the Newton's equation of motion must be solved for given set of forces and atomic coordinates to get the acceleration of each particle at given time points:

$$F_i = m_i a_i;\ F_i = m_i \frac{d^2 r_i}{dt^2}; F_i = -\frac{dE}{dr_i}$$

where acceleration a is the second derivative of coordinates with respect to time, and the force, F, is essentially the gradient of the potential energy, E. The potential energy function of a protein is complex and cannot be solved analytically, therefore a numerical approach is used instead. Numerical integration with fixed time step results in a trajectory, where at each time step the energy, coordinates and velocities (or acceleration) of all particles are known. The time step needs to be sufficiently small to appropriately approximate an otherwise continuous function.

Non-bonded interactions at large distances become small yet they do not disappear, which results in high computational costs. In practice, cut-offs for non-bonded interactions are introduced, such that the truncation effect becomes negligible, but significantly speeds up simulation. Moreover, when a protein molecule is simulated in explicit solvent, the amount of simulated water molecules is proportional to the computational time. Practically, a protein model is placed in a simulation box of the size big enough to prevent a protein from "seeing" itself. Because protein molecule can freely diffuse and would eventually reach the edge of the simulation box, the so-called periodic boundary conditions (PBC) are introduced. The PBC is a creation of infinite volume with a finite unit volume, such that if the molecule were to cross one border it would reappear through a conjugate border back to the unit volume.

To start a simulation, an initially static system is gradually heated and velocities for each atom are randomly generated based on a Maxwell-Boltzmann distribution at a given temperature. Next, the pressure in the simulation box is set to 1 bar for most biological systems. To avoid some nonphysical behavior, the system is equilibrated before the production MD simulation starts.

The initial models for the MD simulation in this study were the following crystal structures: the dimer structure (Chains B and D) of PpSB1-LOV in the dark state (PDB ID: 5J3W), the dimer structure of PpSB1-LOV in the light state (PDB ID: 3SW1); the dimer structures from this work: PpSB2-LOV in the dark state, SBW25-LOV in the light state and W619_1-LOV in the apo form. Water, ions and ligands except for FMN were removed from the protein crystal structures. Additionally, the homology model of PpSB2-LOV in the light state was created.

The initial model of the PpSB2-LOV in the light state was constructed based on the PpSB1-LOV crystal structure in the light state due to high similarity of their structure and

sequence. Essentially each chain of PpSB2-LOV crystal structure was divided into three parts: N-terminal (residues 1 – 16), core domain (residues 17 – 119) and C-terminal (residues 120 – 148) parts, and each superposed with the respective part of PpSB1-LOV.

The same residues at identical residue positions were preserved in conformations as seen in PpSB1-LOV, whereas in the case of sequence differences, conformations of residues from the superposed PpSB2-LOV structure was used as guidance. The C-terminal part of PpSB1-LOV (residues 120 – 134) is shorter than that of PpSB2-LOV (residues 120 – 148), therefore only the conformations of residues up to position 134 were changed. Residues beyond position 134 were left the same as they were in the crystal structure of PpSB2-LOV as they no longer participated in the dimer interface and were freely-exposed to the solvent.

The initial crystal structures and the initial model of PpSB2-LOV in the light state were protonated and parameterized by the pdb2gmx program as implemented in the GROMACS 5.1 software (Abraham *et al.*, 2015) using the AMBER-99SB-ILDN force field (Lindorff-Larsen *et al.*, 2010). The AMBER-99SB-ILDN force field was used as it has been reported to show good agreement with experimental NMR data (Beauchamp *et al.*, 2012; Lindorff-Larsen *et al.*, 2012).

The topology for the QM-optimized geometries of FMN-cysteinyl and FMN together with the QM-derived partial charges were created with the ACPYPE script (Sousa da Silva and Vranken, 2012). The models were solvated in a periodic box (with ≥ 12 Å distance from any protein atom to the edge of the box) with TIP3P water molecules, and neutralized at a 150 mM NaCl concentration. The systems were energy-minimized with the steepest descent method, and followed by the conjugated-gradient method, and afterwards equilibrated at temperature of 298 K and 1 bar pressure. Cut-off for the electrostatic interaction with Particle Mesh Ewald algorithm (Essmann *et al.*, 1995) was 10 Å and that for the short-range Van der Waals interactions was also 10 Å. The simulation time step was 2 fs and the trajectory was saved every 10 ps. The dark state simulations were run for 50 ns, whereas the light state simulations for PpSB1-LOV were run for 100 ns, and 200 ns for both SBW25-LOV and PpSB2-LOV. The simulation of W619_1-LOV in the apo form was run for 50 ns.

The root-mean-square deviation (RMSD) fluctuation after 170 ns of simulation time became small which indicates that the model of PpSB2-LOV in the light state was in apparent equilibrium. Trajectory snapshots from 170 ns to 200 ns were superposed and the average structure was taken. The structure was subsequently energy-minimized, first with steepest descent method and followed by the conjugated gradient to yield the final model of PpSB2-LOV in the light state. The root-mean-square fluctuation (RMSF) calculation for the simulation of PpSB2-LOV in the light state was derived from 170 ns to 200 ns trajectory. The RMSF calculation for the dark state and the apo form simulations were derived from the whole 50 ns trajectory.

3 Results

3.1 PpSB2-LOV

Biophysical and biochemical characterization

UV-Vis spectroscopy and single crystal microspectrometry

PpSB2-LOV undergoes a similar photocycle as the other LOV proteins, although it differs in the adduct lifetime. A typical UV-Vis absorption spectra of PpSB2-LOV in the dark and light states are depicted in Figure 10, left panel. The dark state spectrum has characteristic peaks, which originate from flavin chromophore bound to the protein: three peaks (422 nm, 447 nm and 475 nm) in the region 400 nm – 500 nm and broad unresolved peaks at around 365 nm. Upon blue-light illumination, an adduct is formed between the chromophore and the conserved cysteine residue leading to the formation of a broad peak with a maximum at around 390 nm.

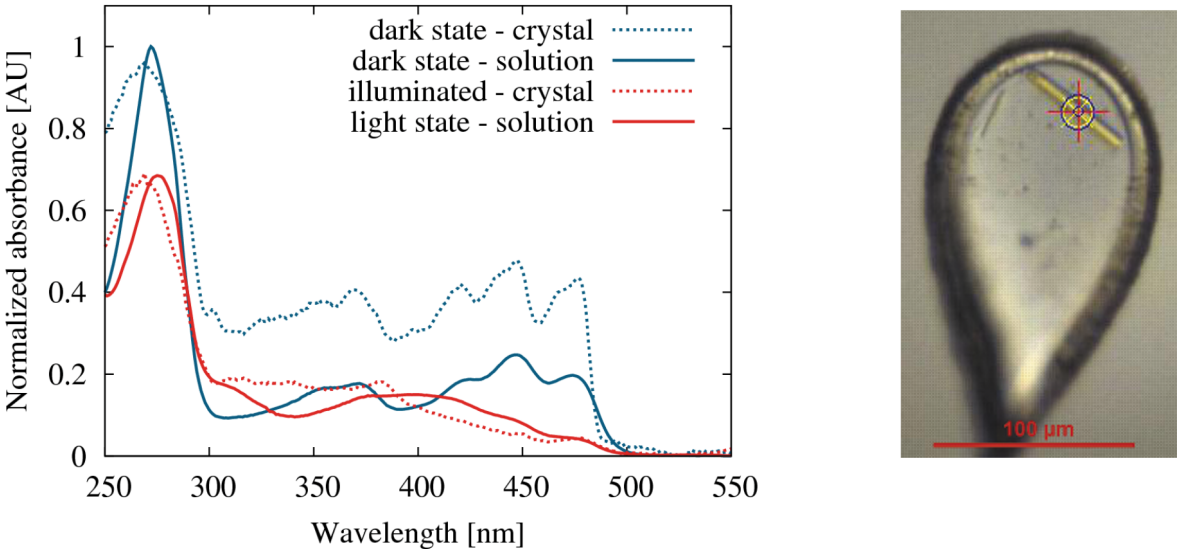


Figure 10. UV-Vis spectra of PpSB2-LOV in solution and in crystal form. (**left**) PpSB2-LOV UV-Vis spectra in the dark and light states, measured both in solution and as a crystal; (**right**) photo of a typical PpSB2-LOV crystal, mounted in a loop, used for structure determination and single crystal microspectrometry.

The PpSB2-LOV protein was crystallized under dark conditions. The UV-Vis spectrum of PpSB2-LOV crystals was measured by single crystal microspectrometry in order to verify that the protein was in the dark state. The measured spectrum of the PpSB2-LOV

crystal were indeed similar to the dark state spectrum in solution, with minor changes related to the fact that the spectra were obtained at 100 K temperature (Figure 10). Additionally, the crystals were exposed to a blue-light laser (λmax = 440 nm) to induce the light state. After the light exposure, the crystals were subsequently used for structure determination. This state will be called the illuminated state to distinguish it from the case when protein crystals grew under continuous blue-light illumination. Upon blue-light illumination of the crystal, its spectrum changed to the characteristic light state spectrum in solution (Figure 10).

Chromophore content

LOV domains can bind naturally-abundant cofactors like flavin mononucleotide (FMN), riboflavin (RBF) or flavin adenine dinucleotide (FAD) which function as a chromophore. All of these chromophores seem to undergo the same photochemistry, although it was reported that the dark recovery kinetics may be influenced (Mansurova *et al.*, 2011).

The chromophore content of recombinantly-expressed PpSB2-LOV in *E. coli* BL21 (DE3) was determined, before and after *in vitro* loading with FMN (Table 18). The previously reported (Jentzsch *et al.*, 2009; Circolone *et al.*, 2012) values: 70% FMN and 30% RBF are in a good agreement with data from this work, and the minor difference can be explained by some variations in recombinant overexpression. On the contrary, loading with FMN results in ~99% FMN content. This indicates a possibly higher affinity of PpSB2-LOV to FMN when compared to RBF and FAD, the same was also reported for the YtVA-LOV protein (Dorn *et al.*, 2013).

Table 18. HPLC analysis of the PpSB2-LOV chromophore content before and after *in vitro* loading with FMN.

PpSB2-LOV	FMN [%]	RBF [%]	FAD [%]
before	70	21.3	8.7
after	99.2	0.8	0

Dark recovery kinetics

The dark recovery kinetics of PpSB2-LOV was measured with protein loaded with FMN, at temperature of 20 °C (Figure 11). The adduct lifetime (τ_{rec}) was determined to be 330 ± 20 s by fitting the data with a single exponential function. For PpSB2-LOV it was reported, that a biexponential fit of the dark recovery kinetics showed better agreement, with one fast component, τ_{rec1} = 28 ± 2 s and one slow component with τ_{rec2} = 184 ± 15 s where each exponents had partial contribution of 29% and 71%, respectively (Jentzsch *et al.*, 2009). Indeed, the dark recovery data measured in this work could be also better fit by a biexponential function, which has one fast component, τ_{rec1} = 86 ± 4 s and one slow component with τ_{rec2} = 369 ± 11 s where each exponents has partial contribution of 19% and 81%. These values differ to published data, which might be related to different chromophore content, see above. The biexponential dark recovery kinetics can be due to the presence of two competitive dark recovery pathways, and further investigations must be done.

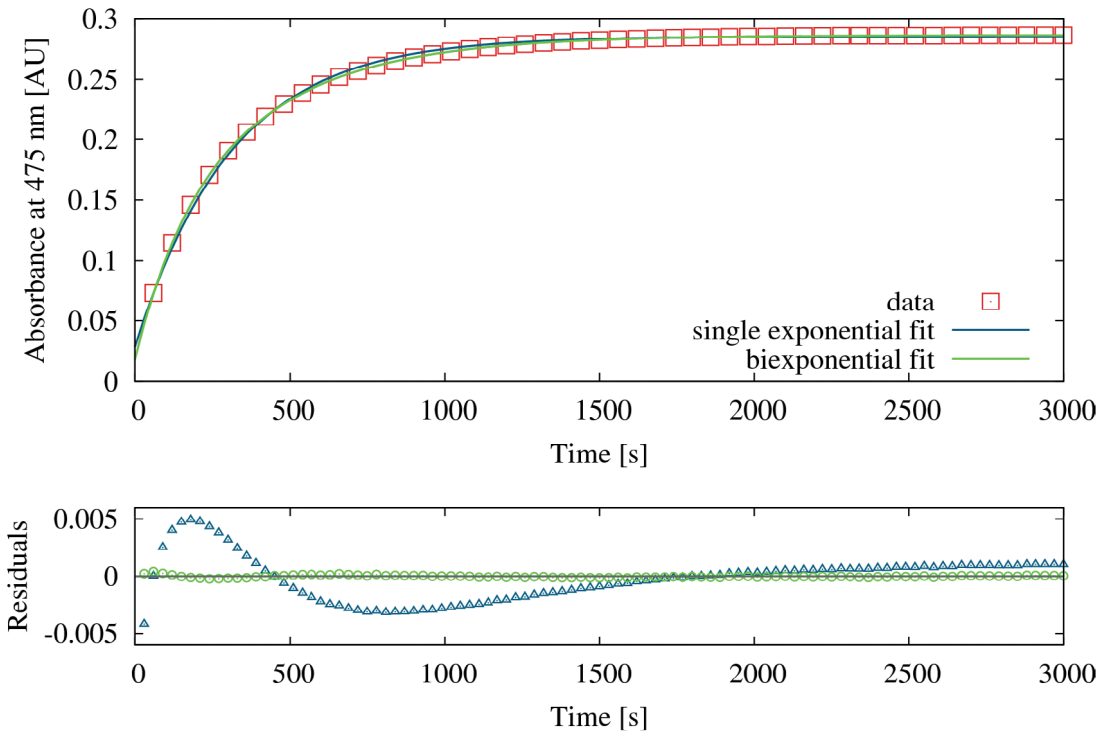


Figure 11. Dark recovery kinetics of PpSB2-LOV. (**top**) Dark recovery kinetics monitored as an absorbance increase at 475 nm over time at 20 °C in 10 mM Na_2HPO_4/NaH_2PO_4 pH 8.0, 10 mM NaCl. The dark recovery kinetics was fit by a single and by a biexponential function. (**bottom**) The respective residuals of data fit by a single and by a biexponential function.

The dark recovery kinetics of LOV proteins can be influenced for example, by imidazole, as reported by Alexandre *et al.*, 2007. It was proposed that, imidazole acts as a base and participates in the deprotonation of the FMN-N5 atom, when protein is in the light state. The dependency of the adduct lifetime on imidazole concentration was investigated for PpSB2-LOV and PpSB1-LOV (Figure 12). The dark recovery data of PpSB2-LOV was fitted not only with a single but also with a biexponential function. The fit with a biexponential function showed a better agreement with data compared to the fit with just single exponential function. The partial contribution of each exponential function in the biexponential fit, was varied from 54% – 81% for the slow component, and 19% – 46% for the fast component.

The adduct lifetime of both proteins decreased with an increase in imidazole concentration, although the x-fold change was higher in the case of PpSB1-LOV (Figure 12). Imidazole-dependent dark recovery at first sight appeared to follow an exponential function, where x-fold acceleration was probably linearly dependent. The response of PpSB2-LOV and PpSB1-LOV to the imidazole, however, was far less when compared to that of AsLOV2 protein (Alexandre *et al.*, 2007), where 10-fold increase was achieved at concentrations as low as 2 mM of imidazole. This might be related to a difference in the accessibility of imidazole to the chromophore pocket or the indirect effect of imidazole on deprotonation. It can also be speculated that the proximity of His61 to the phosphate group and the ribityl chain of FMN enables it to act as a base, thus the imidazole effect on the adduct lifetime of PpSB2-LOV is smaller than that on PpSB1-LOV, where the latter consists of an arginine residue.

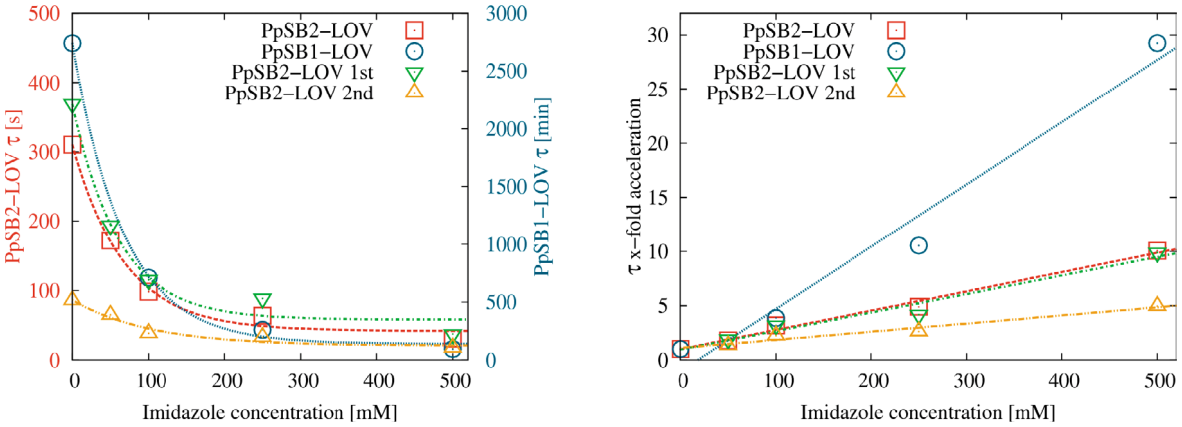


Figure 12. Imidazole-dependent acceleration of the dark recovery of PpSB2-LOV and PpSB1-LOV. The dark recovery kinetics of PpSB2-LOV was fit by a single exponential function (red squares), PpSB2-LOV fit by a biexponential function (green and orange triangles) and PpSB1-LOV (blue circles) by a single exponential function. The data was measured at 20 °C in 10 mM Na_2HPO_4/NaH_2PO_4 pH 8.0, 10 mM NaCl and imidazole at pH 8.0. (**left**) The τ_{rec} for PpSB2-LOV is shown on the left y-axis in seconds and for PpSB1-LOV on the right y-axis in minutes. (**right**) The x-fold acceleration was calculated relative to the imidazole-free adduct lifetime. The data points were fitted with a linear function.

Fluorescence spectra

The fluorescence spectra of PpSB2-LOV in solution and as crystals were acquired by excitation at the wavelengths 440 nm or 450 nm (Figure 13). The fluorescence spectrum in solution showed peak at ~500 nm and a shoulder at ~523 nm as observed for other LOV proteins (Drepper *et al.*, 2007). Whereas fluorescence spectra of protein crystals, measured at temperature of 100 K, showed two distinct peaks at 492 nm and 500 nm and a shoulder at around 555 nm. A better separation of peaks in crystal spectra is likely related to the decreased temperature of 100 K, at which molecular vibration becomes more reduced compared to that at room temperature.

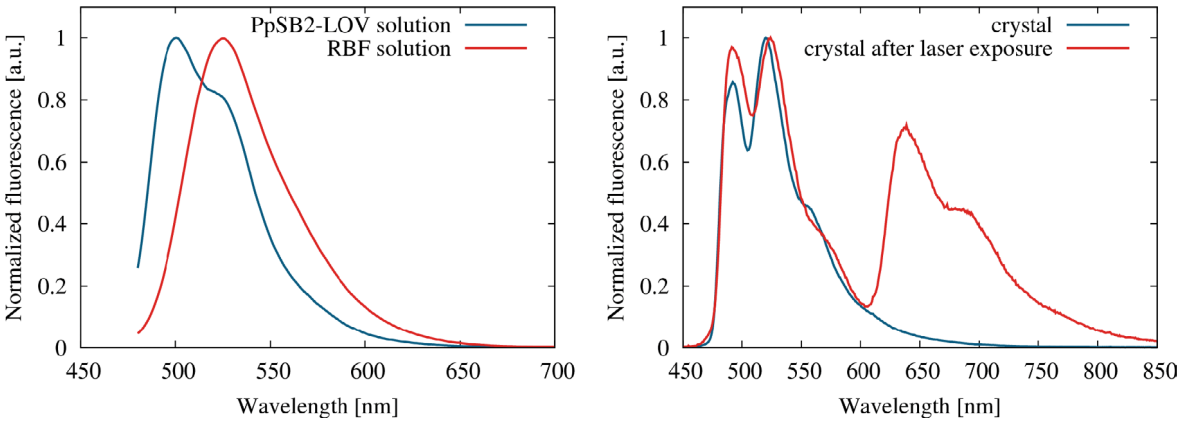


Figure 13. PpSB2-LOV fluorescence spectra both in solution and as a crystal. (**left**) Fluorescence spectra of PpSB2-LOV (blue) and RBF (red) in solution at pH 8.0. (**right**) The fluorescence spectrum of the PpSB2-LOV crystal (blue), which was used in structure determination, and the fluorescence spectrum of the PpSB2-LOV crystal (red), which was exposed with blue-light laser at room temperature.

Surprisingly, the exposure of crystals with blue-light at room temperature and subsequent freezing resulted in the appearance of two additional emission peaks in the far-red

range (Figure 13). It is plausible to consider these to be phosphorescence, which is rarely observed in solution at room temperature.

Crystal structure of PpSB2-LOV

The crystal structure of PpSB2-LOV in the dark state was determined at a resolution of 1.93 Å in $P2_1$ space group. The data collection and refinement statistics are listed in Table 19. The asymmetric unit contains a homodimer, formed by monomers with a non-crystallographic 2-fold symmetry (Figure 14A). All 148 residues of PpSB2-LOV could be observed in the electron-density map, and additionally the histidine residue from the N-terminal expression-tag was also observed and numbered as 0. Superposition of monomers showed their similarity, with the root-mean-square deviation (RMSD) of the Cα atom positions of only 0.8 Å (Figure 14B). The main differences between the structures were in the loop regions: Aβ-Bβ, Eα-Fα and Hβ-Iβ, which is probably induced by crystal packing.

Additionally, the crystals grown under dark conditions were illuminated with blue-light and the X-ray diffraction data was subsequently collected. Resulted illuminated structure was identical to that of the dark grown crystals, moreover, no covalent adduct between the chromophore and the conserved residue Cys53 was observed. Comparable results were previously reported for the PpSB1-LOV illuminated structure (Röllen *et al.*, 2016).

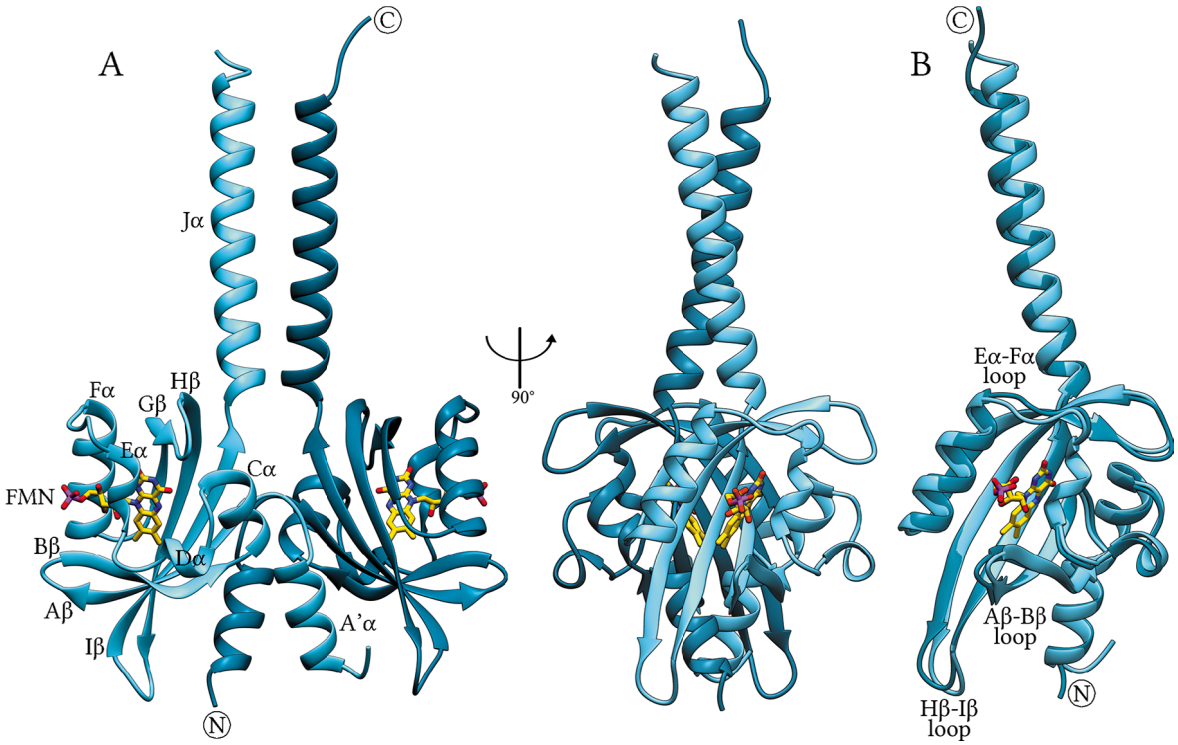


Figure 14. The crystal structure of the PpSB2-LOV in the dark state. (**A**) The structure is represented as ribbons, where chains A and B are colored in light and dark shades of blue. FMN is shown as a stick model. (**B**) Superposition of the monomers to each other showed differences in the Aβ-Bβ, Eα-Fα and Hβ-Iβ loops. After superposition, the RMSD of the Cα atom positions is 0.8 Å.

Table 19. PpSB2-LOV crystal structures: data collection and refinement statistics.

X-ray data	PpSB2-LOV dark state
Beamline	ID23-2, ESRF
Detector	PILATUS 2M
Wavelength [Å]	0.8726
Resolution range [Å]	44.88 – 1.93 (1.98 – 1.93)*
Space group	$P2_1$
Unit-cell a, b, c [Å]	37.14 88.47 54.18
α, β, γ [°]	90 106.03 90
Total reflections	107111 (6856)
Unique reflections	25353 (1705)
Multiplicity	4.2 (4.0)
Completeness [%]	100.0 (100.0)
Mean I/σ(I)	8.9 (1.4)
Wilson B-factor [Å^2]	29.14
R-merge	0.097 (0.890)
R-meas	0.111 (1.028)
$CC_{1/2}$	0.997 (0.450)
Refinement	
Resolution range [Å]	44.88 – 1.93 (2.00 – 1.93)
R-work	0.1769 (0.2711)
R-free	0.1944 (0.3319)
coordinate error (max.-likelihood based) [Å]	0.23
Number of non-hydrogen atoms	2530
macromolecules	2375
ligands	73
water	82
Protein residues	299
RMSD (bonds) [Å]	0.008
RMSD (angles) [°]	0.91
Ramachandran favored [%]	98.7
Ramachandran outliers [%]	0
Clashscore	2.93
Average B-factor [Å^2]	39.18
macromolecules [Å^2]	39.28
ligands [Å^2]	34.45
solvent [Å^2]	40.51

* Statistics for the highest-resolution shell are shown in parentheses

Dimer interface in crystal and solution

The PISA analysis (Krissinel and Henrick, 2007) of the dimer present in the crystal structure has a calculated total buried surface area of 3390 Å^2 with a -28 kcal/mol solvation energy gain upon dimer formation, indicative of a stable dimer in solution. The dimer interface is stabilized by hydrophobic interactions as well by hydrogen bonds and salt bridges (Table 20). Interfacial residues are predominantly situated in the A'α helix, the Jα helix and spread throughout the β-strands of core domain. The sequence of the Jα helix has

characteristic heptad repeats which lead to the coiled-coil like interactions that greatly contribute to the stabilization of the dimer interface.

Table 20. The interfacial residues of PpSB2-LOV monomers in the dimer within the hydrogen bond distance of ≤ 3.2 Å, which was observed in the crystal structure.

Secondary structure element	Residue of Chain A	Atom	Distance [Å]	Atom	Residue of Chain B	Secondary structure element
–	Met1	N	3.1	ND2	Asn104	Hβ
A'α	Asn15	ND2	3.1	OG	Ser98	Hβ
A'α	Asp16	OD1	2.8	NH2	Arg117	Iβ
Hβ	Ser98	OG	3.1	ND2	Asn15	A'α
Iβ	Arg117	NH2	2.9	OD1	Asp16	A'α
Jα	Arg140	NH2	3.1	OE2	Glu135	Jα

Analytical size-exclusion chromatography (SEC) gave a molecular weight (MW) estimation of PpSB2-LOV to be 55 kDa, which is higher than the MW calculated for the dimer based on its sequence (Table 21). An overestimation of the molecular weight could be due to the non-globular shape of the protein. Small-angle X-ray scattering (SAXS) data collection for PpSB2-LOV was performed under dark and light conditions. The SAXS data for both measurements were identical, probably due to the limitations of the method to detect the shape changes of a small protein such as PpSB2-LOV.

The estimation of the MW from the SAXS data based on the calculation of the Porod volume agreed well with the calculated MW of the dimer. The MW determined from the I_0 value, however underestimated it, probably because I_0 calculation depends on an exact protein concentration determination and can be therefore prone to errors. The estimations of the MW of PpSB2-LOV from the SEC and SAXS data together with the radius of gyration (R_g) consistently showed evidence of a dimer in solution (Table 21).

Table 21. An overview of determined molecular weight (MW) and SAXS data of PpSB2-LOV. The MW from Porod volume was calculated by dividing volume by the factor 1.7.

Calculated MW of monomer / dimer [kDa]	SEC MW [kDa]	SAXS MW from I_0 [kDa]	SAXS MW from Porod volume [kDa]	Crystal structure R_g [nm]	Guinier R_g from SAXS
17.3 / 34.6	55	26	31	2.4	2.6

Additionally, the theoretical scattering curves of the crystal structure (dimer and monomer) were calculated and fitted to the experimental scattering data (Figure 15). The scattering curve of the dimer fitted much better than that of the monomer, with the χ^2 value of 1.6 for the dimer. Moreover, an *ab initio* envelope model calculated from the SAXS data matched well the dimer structure present in the crystal structure, as shown in Figure 16.

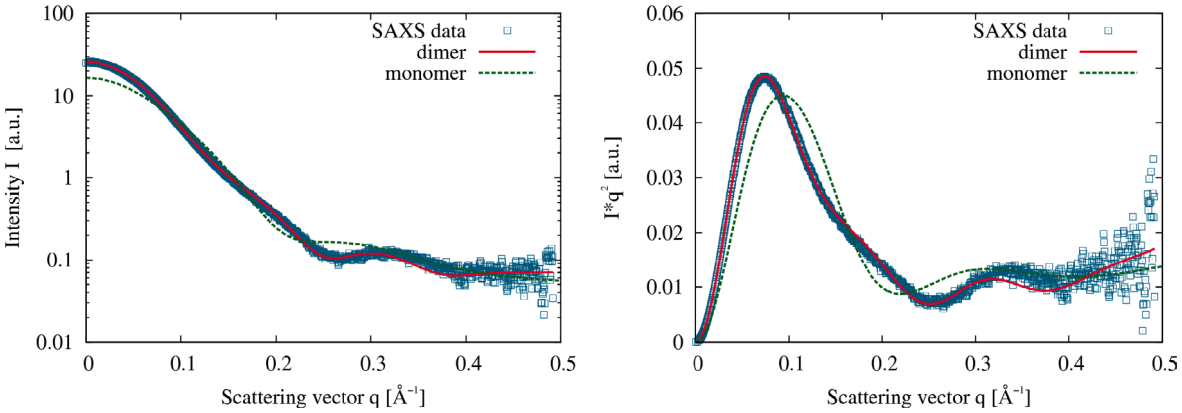


Figure 15. The PpSB2-LOV SAXS scattering curve fitted to the theoretical scattering curves of the dimer or monomer. (**left**) The theoretical scattering curves of the dimer (solid line) and monomer (dashed line) of the PpSB2-LOV crystal structure are shown as lines, and the experimental SAXS data are shown as squares. (**right**) The same data represented as a Kratky plot.

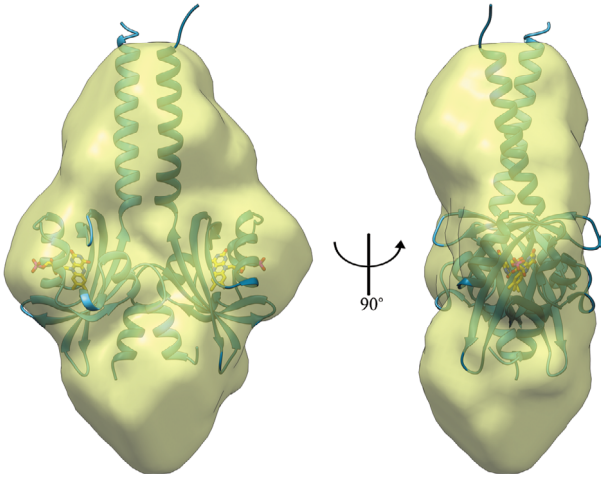


Figure 16. The SAXS envelope model of PpSB2-LOV aligned with the crystal structure. The envelope model is shown as a transparent yellow surface and PpSB2-LOV crystal structure as a ribbon diagram in blue. FMN is shown as a stick model.

Comparison with PpSB1-LOV in the dark state

PpSB2-LOV and PpSB1-LOV are homologous short LOV proteins from *Pseudomonas putida* KT2440. They share a sequence identity of about 67%, see the sequence alignment in Figure 17. Most of the sequence differences are located in the Fα and Jα helices.

The superposition of the PpSB2-LOV dimer to the PpSB1-LOV dimer (PDB ID: 5J3W, Röllen *et al.*, 2016) (Chains B and D), both in the dark states, showed their high similarity, with the RMSD of Cα atom positions of only 1.1 Å for residues 1 – 134 (Figure 18). PISA analysis (Krissinel and Henrick, 2007) of PpSB1-LOV dimer calculated total buried surface area of 3096 Å^2, which is smaller than 3390 Å^2 of PpSB2-LOV. Increased buried surface area can be accounted to the longer Jα helices in PpSB2-LOV.

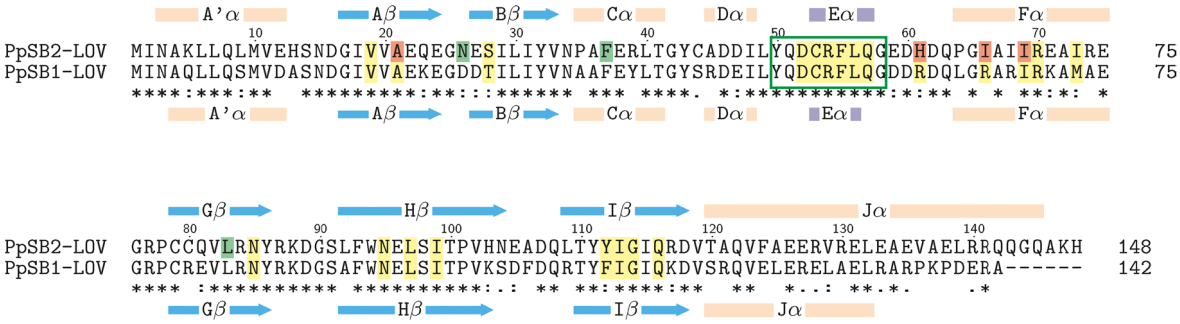


Figure 17. The sequence alignment of PpSB2-LOV with PpSB1-LOV. The secondary structure elements for PpSB2-LOV are shown on top, and for PpSB1-LOV on bottom, where: arrow – β-strand and cylinder – α-helix (apricot) and 3_{10}-helix (plum). The asterisks at the bottom line of the alignment indicate identical residues in a given sequence position, while single and double dots refer to highly and moderately similar residues, respectively. Residues within ≤ 4 Å distance to FMN in both PpSB2-LOV and PpSB1-LOV are marked in yellow. Residues within ≥ 4 Å distance to FMN in PpSB2-LOV but within ≤ 4 Å in PpSB1-LOV are marked in red, and vice versa, are marked in green. The conserved sequence identified in *Pseudomonas* LOV proteins is emphasized by the green box.

The C-terminal region of PpSB2-LOV is six residues longer in a sequence than that of PpSB1-LOV. However, their Jα helices differs by eleven residues in length (Figure 17). Not only the crystal structures but also the secondary structure propensities from the NMR data for the isolated C-terminal peptides (residues 120 – 148 of PpSB2-LOV and residues 120 – 146 for PpSB1-LOV) showed agreement (Rani *et al.*, 2013). The Jα helices of PpSB2-LOV formed crossing angle of 27.3° between them which is similar to 27.9° of PpSB1-LOV. However, longer Jα helices of PpSB2-LOV super-twist and bend more around each other, apparently to increase interfacial surface area.

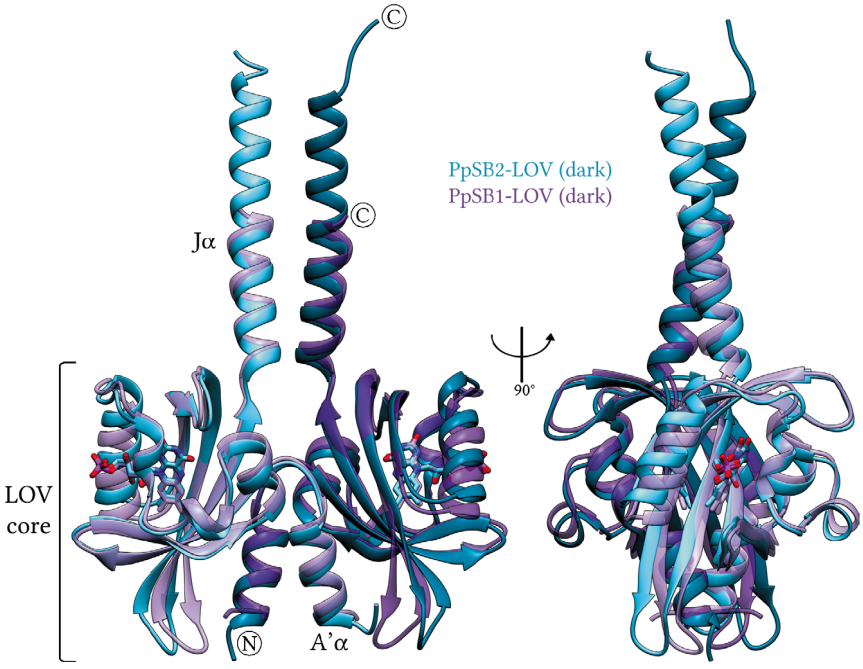


Figure 18. The superposition of PpSB2-LOV and PpSB1-LOV dimers in the dark state. The PpSB2-LOV (blue) and PpSB1-LOV (PDB ID: 5J3W, Röllen *et al.*, 2016) dimers are shown as ribbons. Chains A and B are colored in light and dark shades. FMN is shown as a stick model. After superposition, the RMSD of the Cα atom positions is 1.1 Å for residues 1 – 134.

Chromophore pocket

LOV proteins can generally accept RBF, FMN or FAD as a chromophore. The $2mF_o$-DF_c electron-density map clearly showed the electron density for an FMN molecule in both monomers and agree well with the chromophore composition determined by HPLC. FMN is tightly coordinated by amino acid residues with total buried surface area of ~860 Å^2 according to PISA analysis (Krissinel and Henrick, 2007). The residues are spread mainly over the β-strands Gβ, Hβ and Iβ, the 3_{10}-helix Eα and the α-helix Fα (Table 22). The isoalloxazine ring system of FMN is coordinated by hydrogen bonds with the conserved residues Gln57, Asn85, Asn95 and Gln116 (Figure 19A). Furthermore, residues Asp52 and Gln57 form hydrogen bonds with the ribityl chain of FMN. Additionally, the phosphate group of FMN is coordinated by two arginine residues, Arg54 and Arg70. Due to the crystal packing of PpSB2-LOV, Arg70 forms a salt bridge with the phosphate group of FMN of the symmetry-equivalent dimer. As previously observed for phot1 LOV1 (Fedorov *et al.*, 2003) and YtvA-LOV (Möglich and Moffat, 2007) the conserved Cys53 residue, which forms adduct to FMN in the light state, in one monomer of PpSB2-LOV adopts two conformations, with relative occupancies of 0.53 and 0.47.

Table 22. The residues of PpSB2-LOV, which are within the hydrogen bond distance ≤ 3.2 Å to FMN.

Chain A					Chain B				
Ligand	Atom	Distance [Å]	Atom	Residue	Ligand	Atom	Distance [Å]	Atom	Residue
FMN500	O3P	2.8	NH2	Arg54	FMN500	O3P	2.9	NE	Arg54
	O1P	2.6	NE	Arg54		–			
	O2P	3.1	NH2	Arg70		O2P	2.7	NH2	Arg54
	O4'	3.0	NE2	Gln57		O4'	2.9	NE2	Gln57
	O2'	2.7	OD1	Asp52		O2'	2.6	OD1	Asp52
	N5	3.3	OE1	Gln116		N5	3.3	OE1	Gln116
	O4	3.1	ND2	Asn95		O4	3.3	ND2	Asn95
	N3	2.8	OD1	Asn85		N3	2.9	OD1	Asn85
	O2	2.9	ND2	Asn85		O2	3.0	ND2	Asn85
	O2	2.9	NE2	Gln57		O2	3.0	NE2	Gln57

The chromophore pockets of PpSB2-LOV and PpSB1-LOV (Röllen *et al.*, 2016), both in the dark states, are shown in Figure 19. Overall, their secondary structure elements and their length remains the same, however, they differ in conformation of the Eα-Fα loop. The sequence differences of PpSB2-LOV and PpSB1-LOV among residues, which belong to the chromophore binding pocket, show altered interactions with the chromophore. In the Eα-Fα loop, the His61 residue instead of the Arg61 in PpSB1-LOV moved further away from FMN and lost strong interaction with the phosphate group of FMN. The presence of the Ile66 residue instead of the Arg66 in PpSB1-LOV resulted in loss of salt bridge to FMN. On the other side of the pocket, Ile73 and Tyr112 residues compared to Met73 and Phe112

residues in PpSB1-LOV occupied almost equal volume yet they differ in their properties. However, Met73 residues also may have additional π-Sulphur interaction (~5.5 Å distance from Sδ and center of aromatic ring (1,2-Dimethylbenzol) of FMN with edge-on configuration) with the aromatic ring of FMN. Additionally, Tyr112 instead of Phe112 in PpSB1-LOV can form bond to Gln23 (Lys23 in PpSB1-LOV).

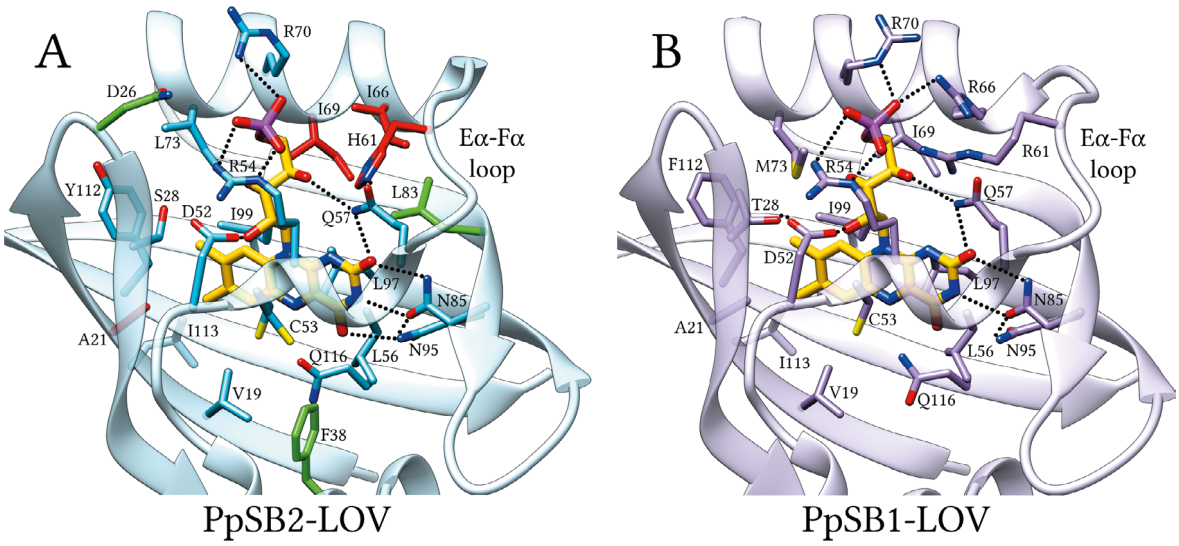


Figure 19. Comparison of the chromophore binding pockets of PpSB2-LOV and PpSB1-LOV in the dark states. (**A**) PpSB2-LOV (blue) and (**B**) PpSB1-LOV (PDB ID: 5J3W, Röllen *et al.*, 2016) (plum) structures are represented as ribbons. FMN and the interacting residues are shown as stick models. The carbon atoms of the residues within ≤ 4 Å distance to FMN are colored in model color. The carbon atoms of the residues within ≥ 4 Å distance to FMN in PpSB2-LOV but within ≤ 4 Å in PpSB1-LOV are marked in red, and vice versa, are marked in green. The hydrogen bonds and salt bridges within ≤ 3.2 Å distance to FMN are shown as dotted lines.

The conserved residue Gln116 was reported as one of the key residues for signaling (Crosson and Moffat, 2002; Zoltowski *et al.*, 2007; Vaidya *et al.*, 2011). Residue Gln116 in PpSB2-LOV can form weak hydrogen bonds with the FMN-N5 atom (3.3 Å distance) or with the backbone oxygen atom of Gly17 residue, which is slightly closer (3.2 Å distance). Most likely in solution and under dark conditions, the side chain of Gln116 residue flips with its protonated nitrogen to each side. Whereas upon formation of the light state with the subsequent protonation of the FMN-N5 atom, Gln116 residue would stay, probably, in the more favorable single conformation.

The Asp52 residue preceding the adduct-forming Cys53 forms hydrogen bonds: on one side to the FMN-O2' atom and to the backbone N atom of Cys53, on the other side to the backbone N atom of Ser28 and water molecule in the chromophore pocket. Moreover, the side chain of Ser28 forms a hydrogen bond with this water molecule. In PpSB1-LOV, Asp52 forms a hydrogen bond with the side chain of Thr28 instead of a water molecule.

The solvent-accessible cavities of the chromophore pocket for PpSB2-LOV and PpSB1-LOV were calculated with the 3V program (Figure 20A). The cavity volumes essentially reflected the chromophore shape together with its solvent accessibility. The PpSB2-

LOV cavity showed protrusions both in the direction of the Eα-Fα loop and in the direction of the C-terminal end of the Fα helix. This possibly indicates the higher solvent accessibility to FMN of PpSB2-LOV compared to that of PpSB1-LOV. It was previously shown that the rate-limiting step for the thermally driven dark recovery is the deprotonation of the FMN-N5 (Zoltowski, Vaccaro and Crane, 2009). An increased solvent accessibility in PpSB2-LOV may thus enable higher deprotonation rates of the FMN-N5 and accelerate dark recovery.

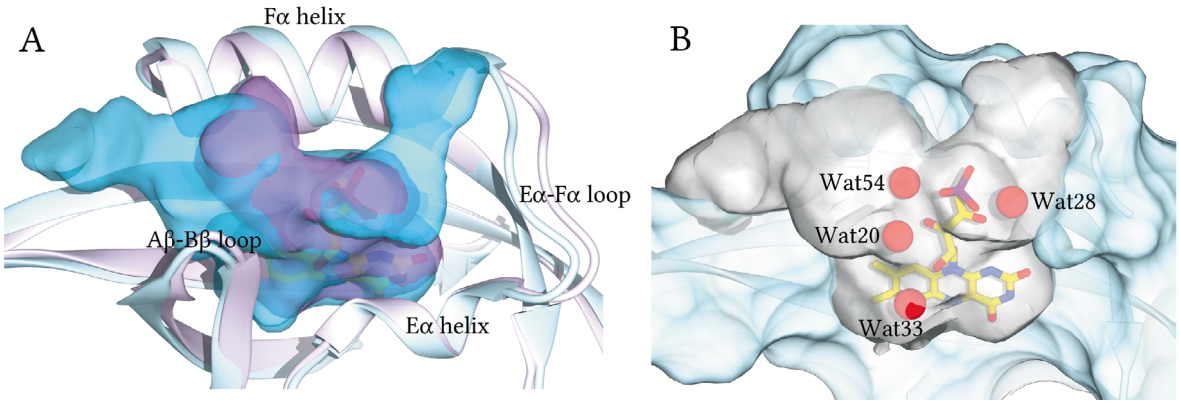


Figure 20. The superposition of PpSB2-LOV and PpSB1-LOV structures in the dark state together with their solvent-accessible cavities of the chromophore pocket. Water molecules in the vicinity of the chromophore in the PpSB2-LOV structure. (**A**) PpSB2-LOV (pale blue) and PpSB1-LOV (PDB ID: 5J3W, Röllen *et al.*, 2016) (pale plum) structures are represented as ribbons. The solvent-accessible cavities of PpSB2-LOV (blue) and PpSB1-LOV (plum) were created with a 1.4 Å probe radius and represented as transparent surfaces. The FMN from PpSB2-LOV is shown as a stick model. FMN molecules were not included in the calculation of the cavities but shown to indicate their positions in the respective structures. (**B**) Water molecules in the vicinity of the chromophore in the PpSB2-LOV structure. PpSB2LOV is represented as a ribbon diagram and as a blue transparent molecular surface. Residues 1 – 62 are not shown for clarity. The calculated solvent-accessible cavity of the chromophore pocket is shown as a gray transparent surface. Water molecules are represented as red spheres.

Table 23. The hydrogen bond distances between water molecules and FMN in chain A of PpSB2-LOV. Chain B contained the same water molecules except Wat 54, which was absent.

Water molecule	Distance [Å]	Atom	Residue
Wat20	2.9	O3'	FMN500
	2.4	OG	Ser28
	3.0	OD2	Asp52
Wat28	2.4	O1P	FMN500
	3.1	OE1	Gln57
Wat33	2.6	O	Ser28
	2.9	O2'	FMN500
	2.9	N	Cys53
Wat54	2.7	O5'	FMN500

In the crystal structure of PpSB2-LOV, four water molecules are present in the chromophore pocket, as depicted in Figure 20B. The water molecule in the proximity of conserved Cys53 seems to be observed frequently in other LOV structures (Möglich and Moffat, 2007; Conrad, Bilwes and Crane, 2013; Endres *et al.*, 2015). Some water molecules, such as water molecules 20, 28 and 33, act as a mediator for the hydrogen bonding between FMN and the residues that form the chromophore pocket, as shown in Table 23.

Comparison with PpSB1-LOV in the light state

The superposition of the PpSB2-LOV dimer in the dark state with the PpSB1-LOV dimer (PDB ID: 3SW1, Circolone *et al.*, 2012) in the light state showed a poor match, indicating strong differences between dark and light states. The superposition by the core domain residues 17 – 117 of the PpSB2-LOV monomer to the monomer from PpSB1-LOV light state structure showed that they have different dimer interface (Figure 21). Essentially, the monomers rotate around a virtual axis going through the FMN molecules of both monomers, as shown in a simplified cartoon in Figure 21C. This movement might suggest a mechanistic way of transducing the signal from the chromophore of the LOV domain to the fused effector domain via the C-terminal helices. The movement observed here is similar to that reported previously for PpSB1-LOV where the dark and light states were superposed in a likewise manner (Röllen *et al.*, 2016). Further on it will be shown that the light state model of PpSB2-LOV obtained from homology modeling and molecular dynamics simulations retain a similar crossing angle of the Jα helices as that in the PpSB1-LOV light state structure and similar dimer interface.

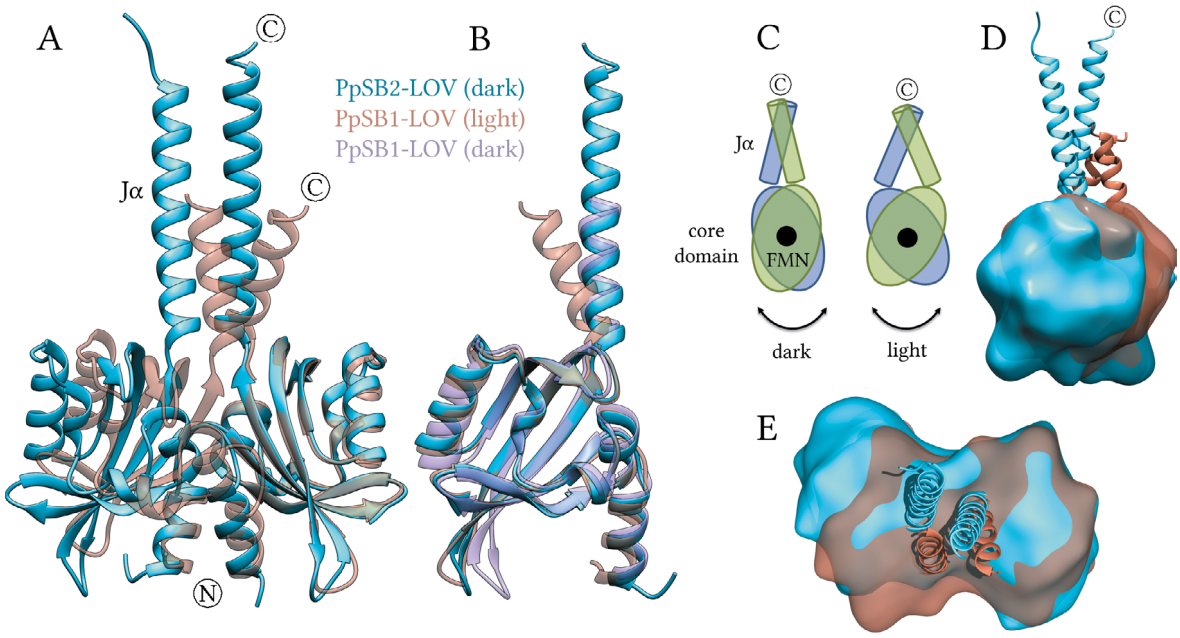


Figure 21. The superposition of the dark state structure of PpSB2-LOV with the dark and light structures of PpSB1-LOV. (**A**) The superposition of the PpSB2-LOV (blue) dark state structure with the PpSB1-LOV (PDB ID: 3SW1, Circolone *et al.*, 2012) (coral) light state structure by the core domain residues 17 – 117 of chain A is shown in ribbon representation. (**B**) The superposition of the PpSB2-LOV (blue) monomer with the PpSB1-LOV monomers of the light (coral) and dark (PDB ID: 5J3W, Röllen *et al.*, 2016) (plum) structures by the core domain residues 17 – 117 of chain A. (**C**) Cartoon representation of the dimer interface rotation observed between the dark and light structures when viewed from the side. The black circle indicates the FMN molecules aligned as an axis around which the rotation takes place. (**D,E**) The superposition of the PpSB2-LOV (blue) dark state structure with the PpSB1-LOV (coral) light state structure by the core domain residues 17 – 117 of chain A, which shows the dimer interface rotation. The residues 1 – 117 are shown as smoothed surfaces and the residues 118 – 148 as ribbons. (**D**) similar viewpoint as in cartoon (C). (**E**) viewpoint from the top.

Molecular dynamics

Molecular dynamics simulation of the dark state

Despite high similarities in structure, the adduct lifetime of PpSB2-LOV is much shorter than that of PpSB1-LOV, ~2 min vs 2741 min (Jentzsch *et al.*, 2009). Certainly, differences between their adduct lifetimes should be related to their sequence. Swapping of single residues between PpSB2-LOV and PpSB1-LOV, such as H61R and I66R mutations in PpSB2-LOV (adduct lifetimes increased by factors 4 and 7, respectively), did not have as dramatic effect as the R61H and R66I mutations in PpSB1-LOV (adduct lifetimes decreased by factors 107 and 3, respectively) (Jentzsch *et al.*, 2009). This suggests that a single residue is not responsible for a change but instead it is a collaborative effect of several residues, especially in the vicinity of the chromophore.

PpSB2-LOV failed to crystallize under light conditions and this is presumably due to its short adduct lifetime, which limits the size of the population of protein molecules in the light state. Molecular dynamics (MD) simulations were performed to address the lack of structural information on the light state and probe the dynamics of the protein in both the dark and light states.

The MD of PpSB2-LOV and PpSB1-LOV in the dark states were simulated for 50 ns. Their respective crystal structures were used as initial structures, as described in the Material and Methods section on page 44. The structures remained stable over the course of the simulation, see snapshots and RMSD of the trajectory in Figure 22. Major movements were observed for the C-terminal ends together with the Aβ-Bβ and Hβ-Iβ loops. The Jα helices of PpSB2-LOV partially unfold at the C-terminus compared to that in the crystal structure. However, NMR-derived secondary structure propensities of isolated C-terminal peptides showed similar findings, that last turns of the Jα helices had lower propensities (Rani *et al.*, 2013). Additionally, it was observed that FMN was fluctuating more in the chromophore pocket of PpSB2-LOV than in the chromophore pocket of PpSB1-LOV. This correlates with the previously observed difference in the volume of the chromophore pocket cavities, and thus indicates a higher solvent accessibility in PpSB2-LOV than in the PpSB1-LOV. While a higher solvent accessibility can in turn be one of the reasons of the short adduct lifetime.

The root-mean-square fluctuation (RMSF) of the core domain residues 17 – 119 of PpSB2-LOV and PpSB1-LOV was calculated and is presented in Figure 23. The overall RMSF of main chain atoms (N, C, O, and Cα) for most of the structure remained similar between PpSB2-LOV and PpSB1-LOV in the dark states. Whereas beside strong motions of the Aβ-Bβ and Hβ-Iβ loops, the Cα and Dα short helices were also more mobile in PpSB2-LOV.

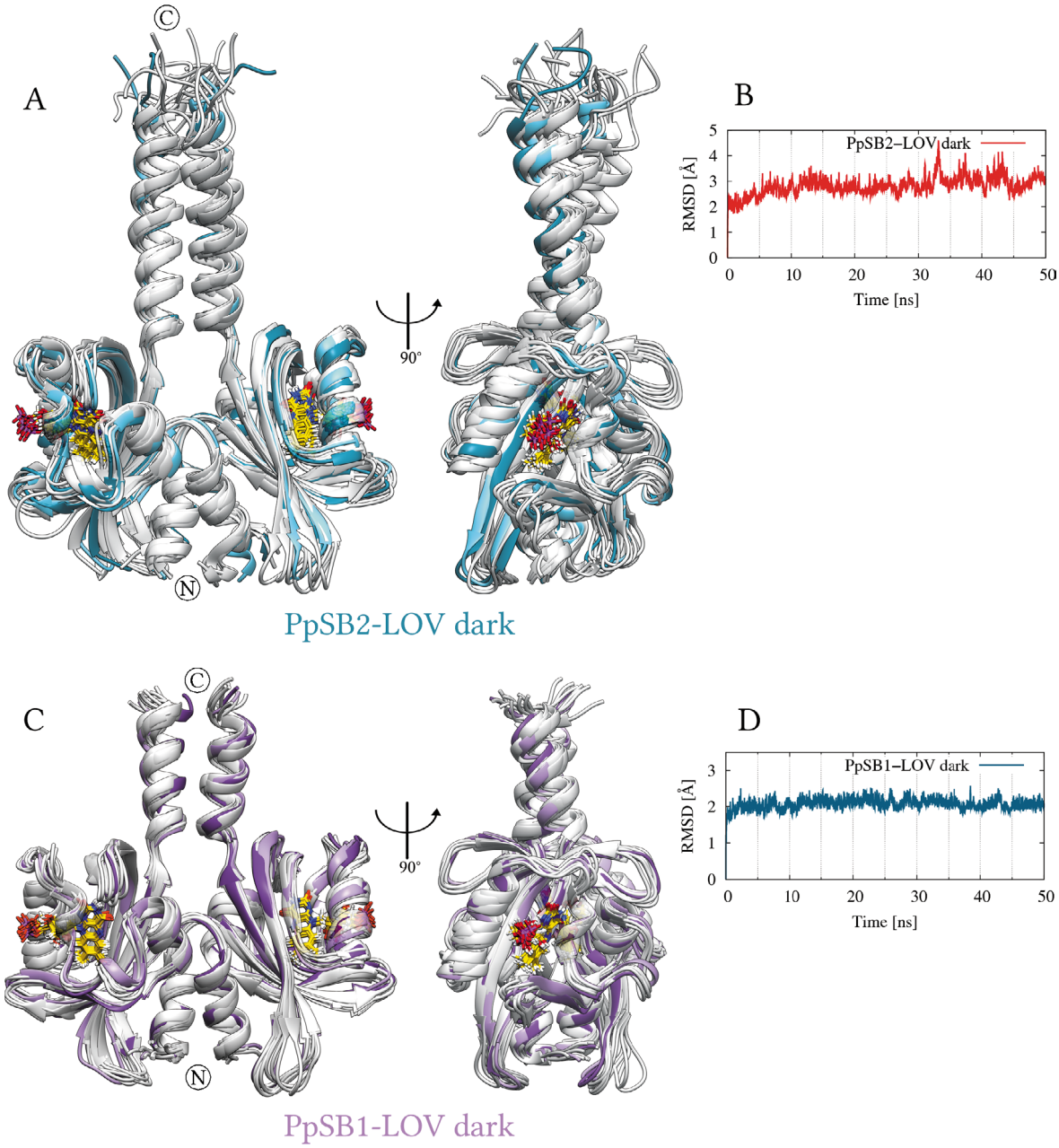


Figure 22. MD simulation snapshots of PpSB2-LOV and PpSB1-LOV in the dark states. (**A**) MD simulation snapshots of PpSB2-LOV in the dark state, with Δt = 5 ns, are shown as ribbons, where the first snapshot is colored in light blue, the last snapshots in dark blue and the snapshots in between in gray. FMN is shown as a stick model. (**B, D**) The RMSD calculated for the whole trajectory with respect to the starting structure. (**C**) MD simulation snapshots of PpSB1-LOV in the dark state, with Δt = 5 ns, shown as ribbons, where the first snapshot is colored in light plum, the last snapshots in dark plum and the snapshots in between in gray. FMN is shown as a stick model.

Comparison of side chains RMSF showed that some of the discrepancies are directly related to the differences in protein sequences where side chains have different conformational freedom. For example, the sequence difference of PpSB2-LOV to PpSB1-LOV: Ala45 to Arg45, Pro64 to Leu64, and Cys80 to Arg80 led to an increase in RMSF in PpSB1-LOV, whereas Arg74 to Ala74 and Leu92 to Ala92 led to decrease in RMSF. The side chain of Cys53 fluctuated more in PpSB2-LOV, which may correlate with increased FMN fluctuation

in the chromophore pocket. The side chain of His61 is more mobile compared to that of Arg61 in PpSB1-LOV, which may be due to the shorter side chain of His61, which cannot form a stable salt bridge to FMN as Arg61 in PpSB1-LOV. The RMSF of Ile66, on the contrary, stayed same as that of Arg66 in PpSB1-LOV. However, isoleucine residue has intrinsically less conformational freedom compared to arginine residue, thus equal RMSF of them may indicate that Ile66 has higher fluctuation than Arg61. Such findings agree well with the previously mentioned results from the H61R and I66R mutations of PpSB2-LOV, which accelerated the dark recovery.

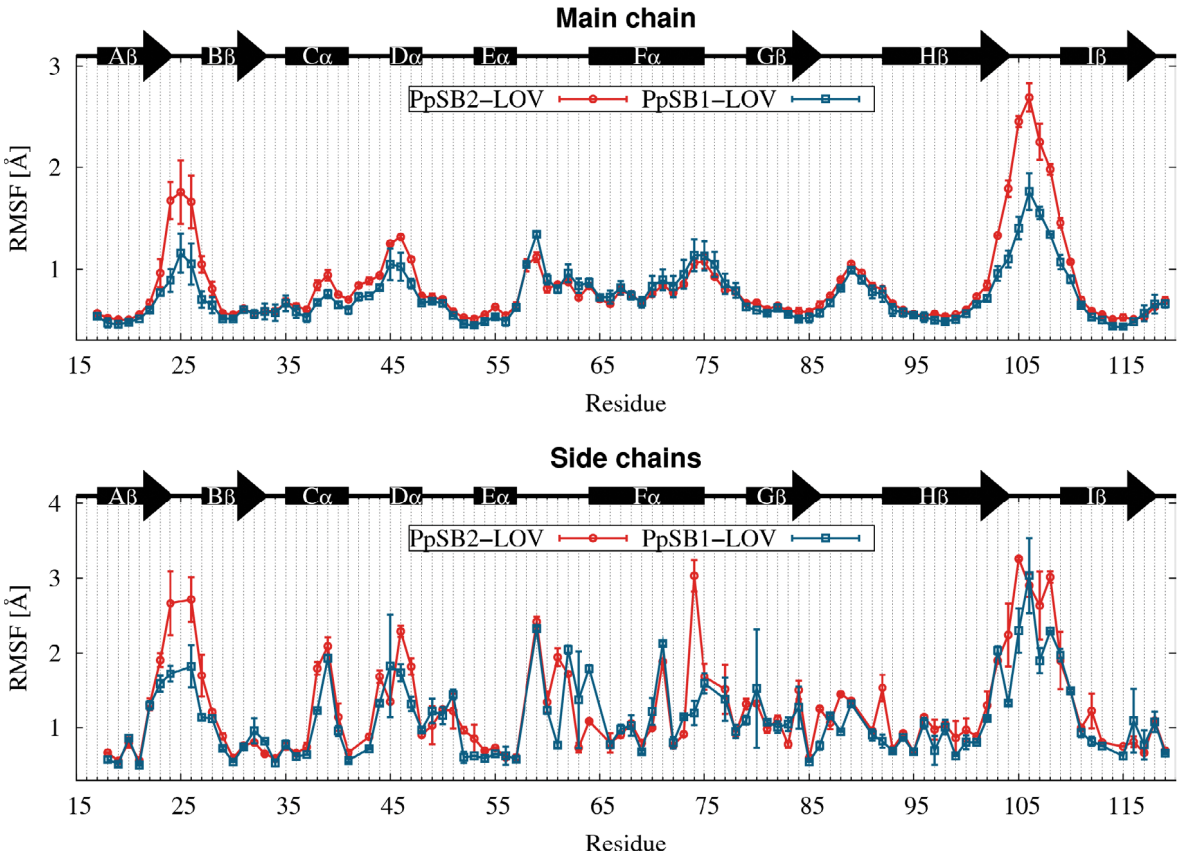


Figure 23. Residue-resolved RMSF of MD simulations of the PpSB2-LOV and PpSB1-LOV in the dark states. The bars indicate the values of two monomers, whereas the points indicate their average. The secondary structure elements of PpSB2-LOV are show as arrows for strands and rectangles for helices. Secondary structure assignment was done with the DSSP program for the PpSB2-LOV crystal structure.

Molecular dynamics simulation of the light state

In order to study molecular dynamics of both proteins in the light state, first the geometry of the FMN-cysteinyl adduct needs to be elucidated. As FMN-cysteinyl adduct is apparently unstable outside of the protein environment, no precise structural information such as the crystal structure at atomistic resolution (≤ 1 Å) is available. However, the crystal structure of lumiflavin(C4a)-isopropyl, which has covalent adduct similar to FMN-cysteinyl adduct, was determined previously (CCDC ID: 1180634) (Bolognesi, Ghisla and Incoccia, 1978). Based on this, the FMN-Cys structure was constructed, and then its geometry was optimized with quantum mechanical (QM) calculations using the B3LYP/6-31G* level of

theory, as described in the Material and Methods section on page 42. The starting and final geometrically-optimized structures in water are depicted in Figure 24. It was previously assumed that the bond length of the FMN-cysteinyl adduct should be ~1.82 Å, as it is in the C-S bond present in cysteine (Crosson and Moffat, 2002). Nonetheless, during geometry optimization, the bond length increased to 1.96 Å, which is not unusual as the C4a atom of the FMN-cysteinyl is sp^3-hybridized.

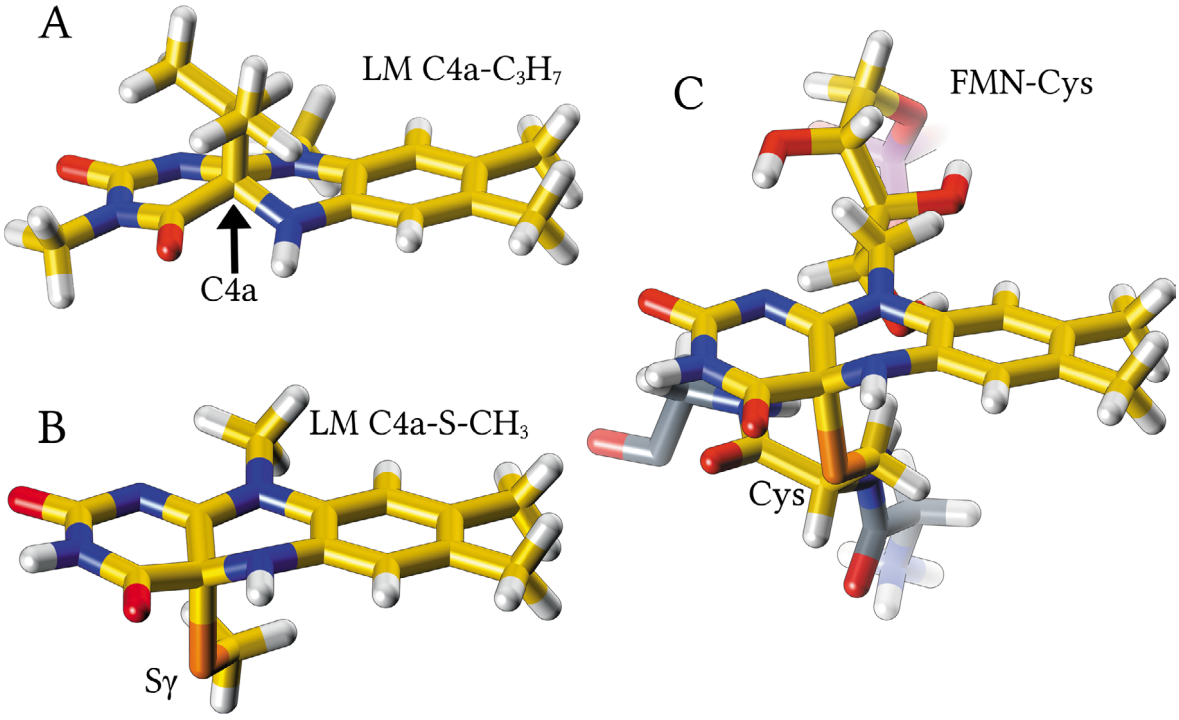


Figure 24. Initial structures and the final QM-optimized structures of FMN-Cys. The structures are shown in stick representation and are colored by elements: carbon, yellow; nitrogen, blue; oxygen, red; sulfur, orange; phosphorus, purple; hydrogen, white. (**A**) The crystal structure of lumiflavin(C4a)-isopropyl (CCDC ID: 1180634) (Bolognesi, Ghisla and Incoccia, 1978). (**B**) The initial structure, lumiflavin(C4a)-S-CH_3, was built and its geometry was optimized with QM, see details on page 42. (**C**) The final QM-optimized structure of FMN-Cys together with two glycine residues (gray), which were used during QM geometry optimization to exclude the influence of the free amine and carboxylic groups of cysteine.

Further on, the initial structure of PpSB2-LOV in the light state was constructed based on the PpSB1-LOV crystal structure in the light state due to their similarity, as described in the Material and Methods section. The QM-derived FMN-Cys structure was fitted to the PpSB1-LOV light state crystal structure and the PpSB2-LOV light state model. During the 100 ns MD simulation, the structure of PpSB1-LOV in the light state remained stable with a constant RMSD fluctuation (Figure 25D). In contrast, the PpSB2-LOV light state model was "artificially" constructed and certainly even after energy minimization the structure would represent nearest local energy minimum. This implies that in the beginning of the MD simulation, the structure explores the conformational space and would eventually find apparently much lower energy minimum. During the 200 ns simulation the structure seemed to be in equilibrium after 170 ns simulation based on its lower RMSD fluctuation compared to that at the beginning (Figure 25B).

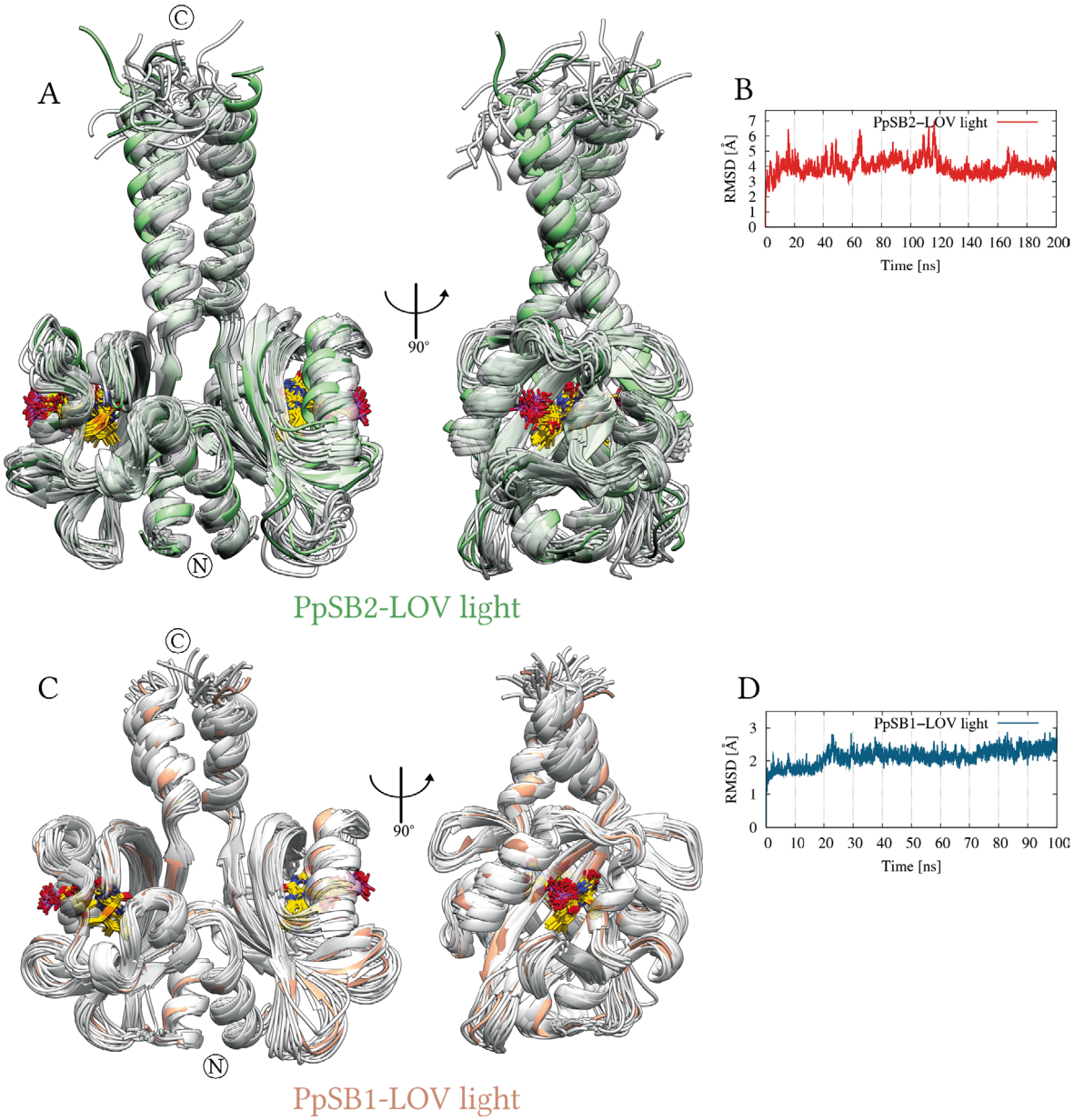


Figure 25. PpSB2-LOV and PpSB1-LOV light state MD simulation snapshots. (**A**) MD simulation snapshots of PpSB2-LOV in the light state, with Δt = 5 ns, are shown as ribbons, where the first snapshot is colored in light green, the last snapshots in dark green and the snapshots in between in gray. FMN and FMN-cysteinyl adduct are shown as a stick model. (**B, D**) The RMSD calculated for the whole trajectory with respect to the starting structure. (**C**) MD simulation snapshots of PpSB1-LOV in the light state, with Δt = 5 ns, shown as ribbons, where the first snapshot is colored in light coral, the last snapshots in dark coral and the snapshots in between in gray. FMN and the FMN-cysteinyl adduct are shown as a stick model.

The RMSF of the core domain residues 17 – 119 of PpSB2-LOV and PpSB1-LOV was calculated during their MD simulations in the light state and presented in Figure 26. The RMSF of main chain atoms (N, C, O, and Cα) for most of the structure remained similar between PpSB2-LOV and PpSB1-LOV in the light state. However, higher RMSF values were observed for residues 62 – 85 in the Eα-Fα loop, the Fα helix, the Gβ strand and the Hβ-Iβ loop. Taking into account that there is only minor difference in the side chain RMSF in residues 62 – 85, this can be interpreted as the concerted motion of the entire region. This

region contains many residues that directly interact with FMN in the crystal structure, such as Ile66, Ile69, Arg70, Leu73, Leu83 and Asn85. In addition, this region contains changes to eight amino acid residues compared to the same region in PpSB1-LOV. The increased dynamics of this residue range can result in the increased solvent access to the FMN. This finding seems to correlate with the above-mentioned observations, that PpSB2-LOV has a faster dark recovery compared to PpSB1-LOV, as a result of higher solvent accessibility to the chromophore in PpSB2-LOV than in PpSB1-LOV.

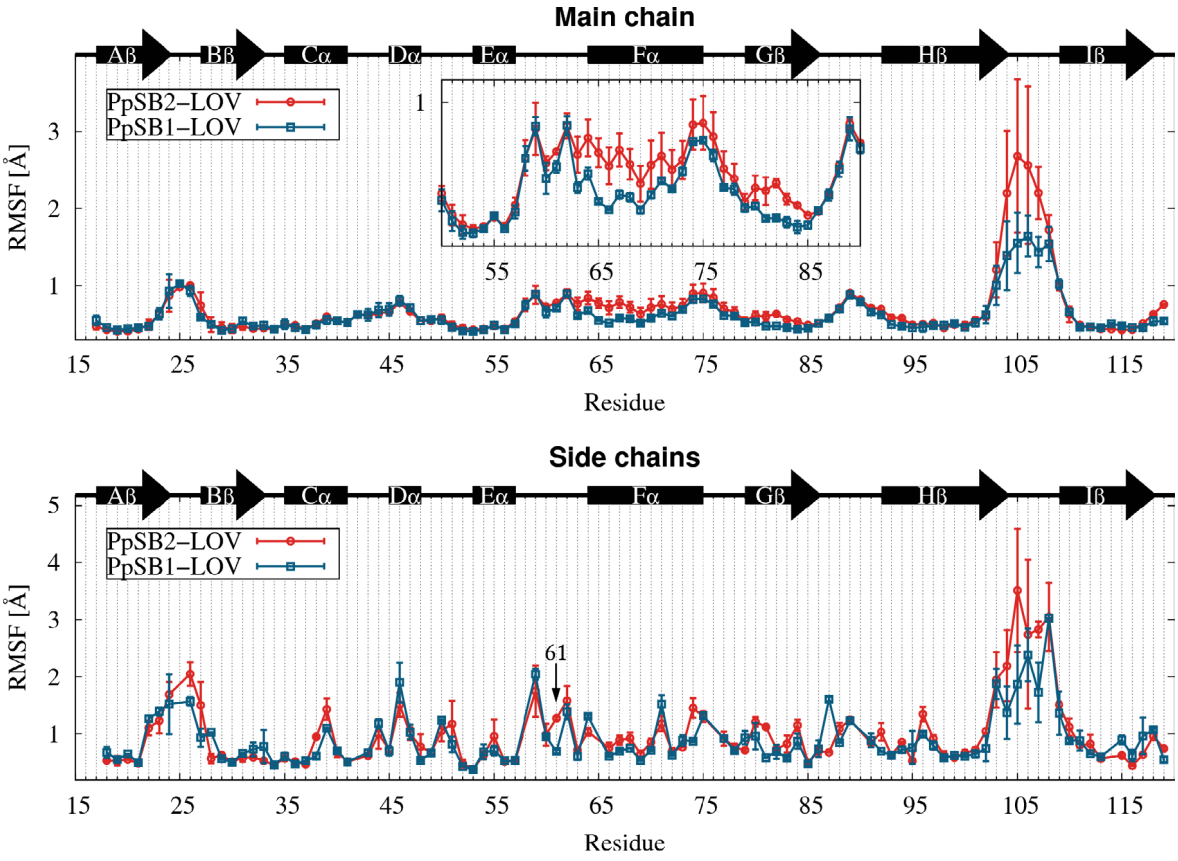


Figure 26. Residue-resolved RMSF of MD simulations of the PpSB2-LOV and PpSB1-LOV in the light states. The bars indicate the values of two monomers, whereas the points indicate their average. The secondary structure elements of PpSB2-LOV are show as arrows for strands and rectangles for helices. Secondary structure assignment was done with the DSSP program for the PpSB2-LOV crystal structure.

Regarding the RMSF of residue side chains, as previously observed, the side chain of His61 is more mobile than that of Arg61 in PpSB1-LOV. Altogether, compared to the dark state simulations, the dynamics of the side chains is decreased in the light state. A similar observation was made for the PpSB1-LOV side chain dynamics studied by neutron scattering and NMR spectroscopy (Röllen *et al.*, 2016; Stadler *et al.*, 2016)

To obtain a representative model of PpSB2-LOV in the light state, an averaged structure obtained during 170 ns to 200 ns of the trajectory was calculated and energy minimized with steepest descent method. Superposition of the averaged structure of PpSB1-LOV in the light state taken from MD with the crystal structure of PpSB1-LOV in the

light state demonstrated their similarities, as depicted in Figure 27. The dimer interface formed by the N-terminal A'α helix and the core region remained stable and similar to the starting structure at the beginning of the simulation. On the contrary, the C-terminal Jα helices in the model became super-twisted around each other. In the starting structure of PpSB2-LOV in the light state, the Jα helices were at the same crossing angle as they were in the crystal structure of PpSB1-LOV in the light state (Circolone *et al.*, 2012). However, the Jα helices in PpSB2-LOV are longer, and at such crossing angle many hydrophobic residues became solvent exposed. Hydrophobic forces drove them to avoid exposure to the solvent and arranged them into a coiled-coil-like interaction. Similar behavior will also be shown in the crystal structure of SBW25-LOV in the light state, described later in this thesis.

The obtained representative model of PpSB2-LOV in the light state brings insight into the light-induced activation of proteins which have the LOV domain N-terminally fused to effector domains, such as histidine kinases, regulation factors etc. (Diensthuber *et al.*, 2013; Engelhard *et al.*, 2013). Conformational changes of Jα helices are crucial and can help understand the mechanisms for signal transduction from the LOV domain to the effector domain.

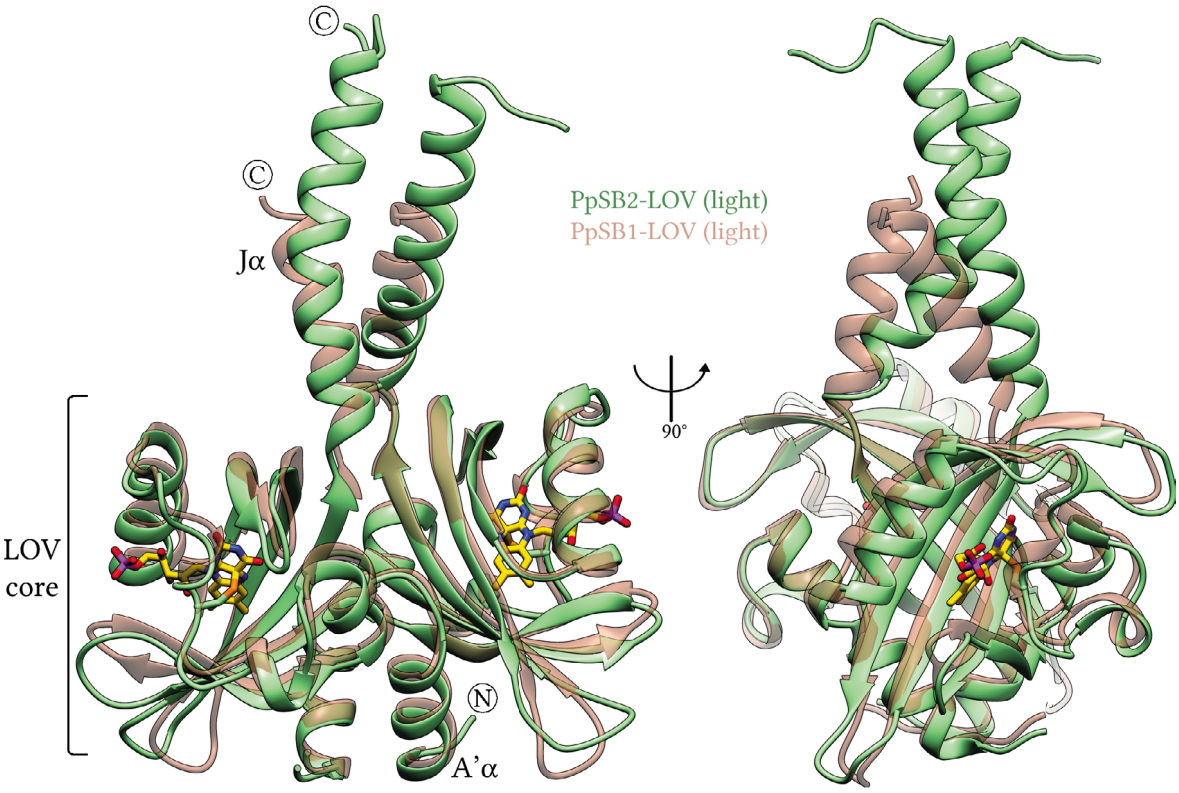


Figure 27. The superposition of the averaged structure of PpSB2-LOV in the light state with the crystal structure of PpSB1-LOV in the light state. The MD-derived averaged structure of PpSB2-LOV in the light state (green) and the crystal structure of PpSB1-LOV in the light state (coral) are shown as ribbons. FMN and the FMN-cysteinyl adduct are shown as stick models. After superposition, the RMSD of the Cα atom position is 1.46 Å over the residues 1 – 133.

3.2 SBW25-LOV and Pf5-LOV

Biophysical and biochemical characterization

UV-Vis spectroscopy, single crystal microspectrometry and dark recovery kinetics

SBW25-LOV is a short LOV protein from the SBW25 strain of *Pseudomonas fluorescens* that can be found on plants, in soil and water. *P. fluorescens* gained its name because they produce fluorescent siderophores, which can be used for instance to chelate iron for storage.

SBW25-LOV undergoes a similar photocycle as the other LOV proteins, although it differs in the recovery time from the dark to the light state. The adduct lifetime (τ_{rec}) of SBW25-LOV, however, is the slowest among the reported LOV proteins. For instance, the previously described protein PpSB1-LOV with the slow dark recovery has a τ_{rec} of 2471 ± 22 min at 20 °C and 109 ± 18 min at 37 °C (Jentzsch *et al.*, 2009). In contrast, SBW25-LOV showed no dark recovery after 16 hours of observations at 25 °C, and only at 37 °C was τ_{rec} measured to be 1470 ± 100 min (Rani *et al.*, 2013).

Typical UV-Vis absorption spectra of SBW25-LOV in the partially dark state and light state are depicted in Figure 28, left panel. The complete dark state spectrum could not be measured due to the technical reasons and the instability of the protein over long time. Nevertheless, the dark state spectrum showed characteristic peaks, which originate from flavin chromophore bound to the protein: three peaks in the region 400 nm – 500 nm and broad peak at 365 nm. Upon blue-light illumination, an adduct is formed between the chromophore and the conserved cysteine residue leading to the formation of a broad peak with a maximum at around 390 nm.

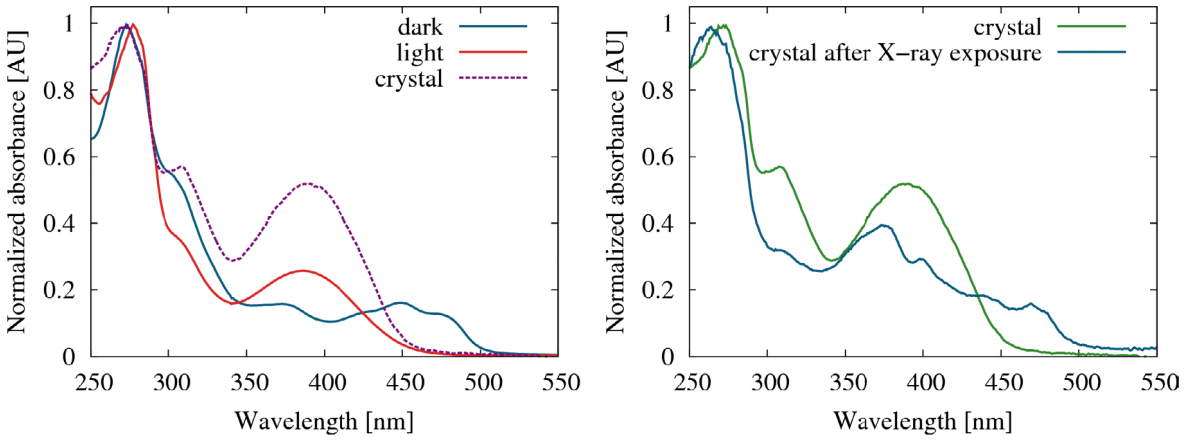


Figure 28. UV-Vis spectra of SBW25-LOV in solution and in crystal form. (**left**) The solution spectrum of SBW25-LOV in the light and partial dark states are shown as solid lines, the spectrum of the protein crystal used in the structure determination is shown as a dashed line. (**right**) The spectrum of protein crystal before and after X-ray exposure during data collection.

The SBW25-LOV protein was crystallized under permanent blue-light (λmax = 450 nm) illumination. The UV-Vis spectrum of protein crystals was measured in order to confirm that the protein is in the light state. Indeed, the spectrum was identical to the light state spectrum of SBW25-LOV in solution, except that the peak intensity was higher. One possibility can be that the higher intensity of the peak is due to the overlapping absorption of the photoproducts of MES buffer molecules, which was used in the crystallization and cryoprotection solution of the crystals.

An adduct formed in the light state between the conserved cysteine residue and the chromophore was previously found to be sensitive to radiation damage during the X-ray diffraction data collection (Fedorov *et al.*, 2003; Halavaty and Moffat, 2007; Zoltowski *et al.*, 2007). In order to reveal the impact of X-rays on the scission of adducts during data collection, the UV-Vis spectrum of the crystal after X-ray exposure was taken (Figure 28, right panel). The spectrum of the exposed crystal did not show prominent peak at 390 nm anymore, typical for the light state, instead some smaller peaks appeared. The peaks around 450 nm could indicate the recovery to the dark state upon disruption of the adduct. Moreover, the spectra of products generated by radiation damage can also contribute to the observed spectrum. The implication of these observations will be discussed in the next section, which describes the electron-density maps of the adduct observed in the crystal structure.

Pf5-LOV is a close homolog of SBW25-LOV and originates from the Pf-5 strain of *Pseudomonas fluorescens*. Pf5-LOV is a short LOV protein, which undergoes a similar photocycle as the other LOV proteins. The spectrum of Pf5-LOV in the dark and light states showed characteristic absorption peaks, which originate from the bound flavin chromophore (Figure 29, left panel).

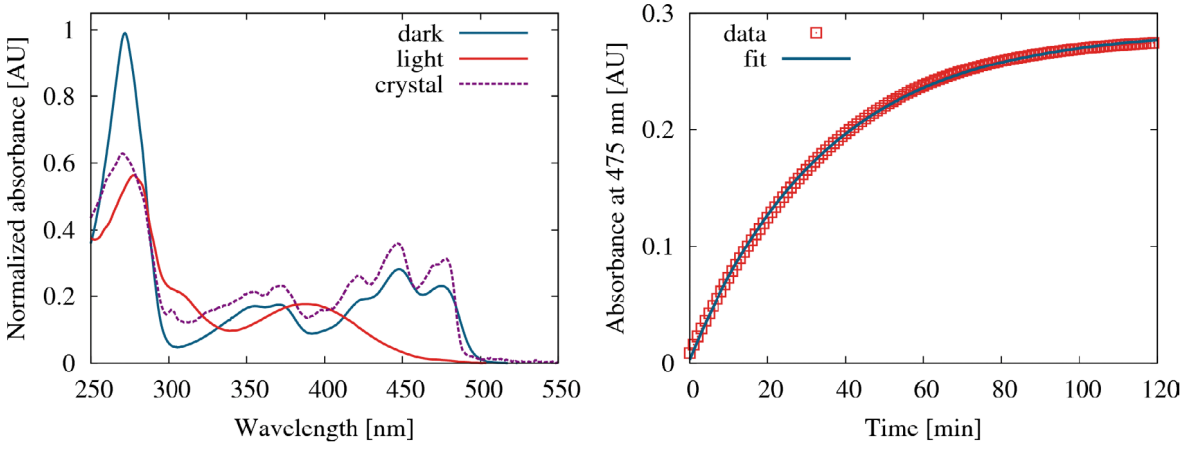


Figure 29. UV-Vis spectra of Pf5-LOV in solution and as a crystal. (**left**) The solution spectrum of Pf5-LOV in the light and dark states are shown as solid lines, the spectrum of the protein crystal used in the structure determination is shown as a dashed line. (**right**) Pf5-LOV dark recovery kinetics was monitored as an absorbance increase at 475 nm over time at 20 °C in 10 mM Na_2HPO_4/NaH_2PO_4 pH 8.0, 10 mM NaCl.

Pf5-LOV was crystallized under dark conditions, the UV-Vis spectrum of protein crystals confirmed that the protein is in the dark state (Figure 29, left panel). Indeed, the spectrum is similar to the dark state spectrum of Pf5-LOV in solution, except minor differences related to the fact that spectra were measured at temperature of 100 K instead of room temperature.

Pf5-LOV, in contrast to SBW25-LOV, has shorter adduct lifetime of 45.3 ±0.7 min at 20 °C (Figure 29, right panel). Previously, the adduct lifetime was only measured at 37 °C and was determined to be 3.6 ± 1.5 min (Rani *et al.*, 2013).

Chromophore content

The chromophore content of the recombinantly expressed SBW25-LOV and Pf5-LOV in *E. coli* BL21 (DE3) was determined and is listed in Table 24.

Table 24. HPLC analysis of the SBW25-LOV and Pf5-LOV chromophore content.

Protein	FMN [%]	RBF [%]	FAD [%]
SBW25-LOV	70.6	29.4	0
Pf5-LOV	89.8	8.1	2.1

Crystal Structure of SBW25-LOV

The crystal structure of SBW25-LOV in the light state was determined at a resolution of 1.6 Å in P1 space group. The data collection and refinement statistics are listed in Table 25. The asymmetric unit contains a homodimer, which is formed by monomers with a non-crystallographic 2-fold symmetry (Figure 30A). In the crystal structure, 148 of the 155 residues in Chain A (150 in Chain B) were observed, moreover, the N-terminal expression-tag of Chain A was also partially observed and its residues were numbered from 0 to -17. Superposition of monomers showed their similarity, with the root-mean-square deviation (RMSD) of the Cα atom positions of 0.89 Å over the residues 1 – 133 (Figure 30C). Whereas, the position differences from residue 134 onwards in the Jα helix becomes more than 4 Å and were excluded from the RMSD calculation. The main differences between the monomers are located in the Jα helix, the Gβ-Hβ loop and the A'α helix. The differences in the Gβ-Hβ loop and the A'α helix were probably induced by the crystal packing.

The Jα helix of chain B has a kink with an approximate angle of 26° when measured between the two vectors defined by the Cα atoms of residues, 125 to 128 and 128 to 131, with the HELANAL-Plus program (Kumar and Bansal, 2012). In comparison, chain A has maximum kink angle of only 9.6°. At these positions SBW25-LOV has neither proline nor known specific sequence which is prone to induce α-helix kink.

The backbone O atom of residue 125 in chain B formed a hydrogen bond with the i+3 residue instead of the classical i+4 for a α-helix, whereas backbone O atom of the next

residue 126, formed no hydrogen bond (Figure 30B). Additionally, backbone O atom of the residue 127 stabilized kink by i+2 hydrogen bonding. Gly130 residue clearly contributed to the stabilization of the helix with its intrinsic high conformational freedom. In fact, a kink in only one helix promotes the increased contact between Jα helices, which otherwise would be freely solvent-exposed at such crossing angle of the α-helices.

Table 25. SBW25-LOV crystal structures: data collection and refinement statistics.

X-ray data	**SBW25-LOV light state**
Beamline	ID23-1, ESRF
Detector	PILATUS 6M
Wavelength [Å]	0.87260
Resolution range [Å]	39.75 – 1.60 (1.63 – 1.60)*
Space group	P1
Unit-cell a, b, c [Å]	36.24 42.42 51.72
α, β, γ [°]	95.3 95.7 109.0
Total reflections	127729 (12553)
Unique reflections	36742 (3653)
Multiplicity	3.5 (3.4)
Completeness [%]	96.8 (95.4)
Mean I/σ(I)	11.50 (1.85)
Wilson B-factor [Å^2]	19.40
R-merge	0.0613 (0.6436)
R-meas	0.07254
$CC_{1/2}$	0.998 (0.73)
Refinement	
Resolution range [Å]	33.94 – 1.60 (1.66 – 1.60)
R-work	0.1827 (0.2359)
R-free	0.2341 (0.3019)
coordinate error (max.-likelihood based) [Å]	0.22
Number of non-hydrogen atoms	2626
macromolecules	2432
ligands	63
water	131
Protein residues	313
RMSD (bonds) [Å]	0.007
RMSD (angles) [°]	0.984
Ramachandran favored [%]	99
Ramachandran outliers [%]	0
Clashscore	2.88
Average B-factor [Å^2]	34.10
macromolecules [Å^2]	34.30
ligands [Å^2]	26.30
solvent [Å^2]	34.20

* Statistics for the highest-resolution shell are shown in parentheses

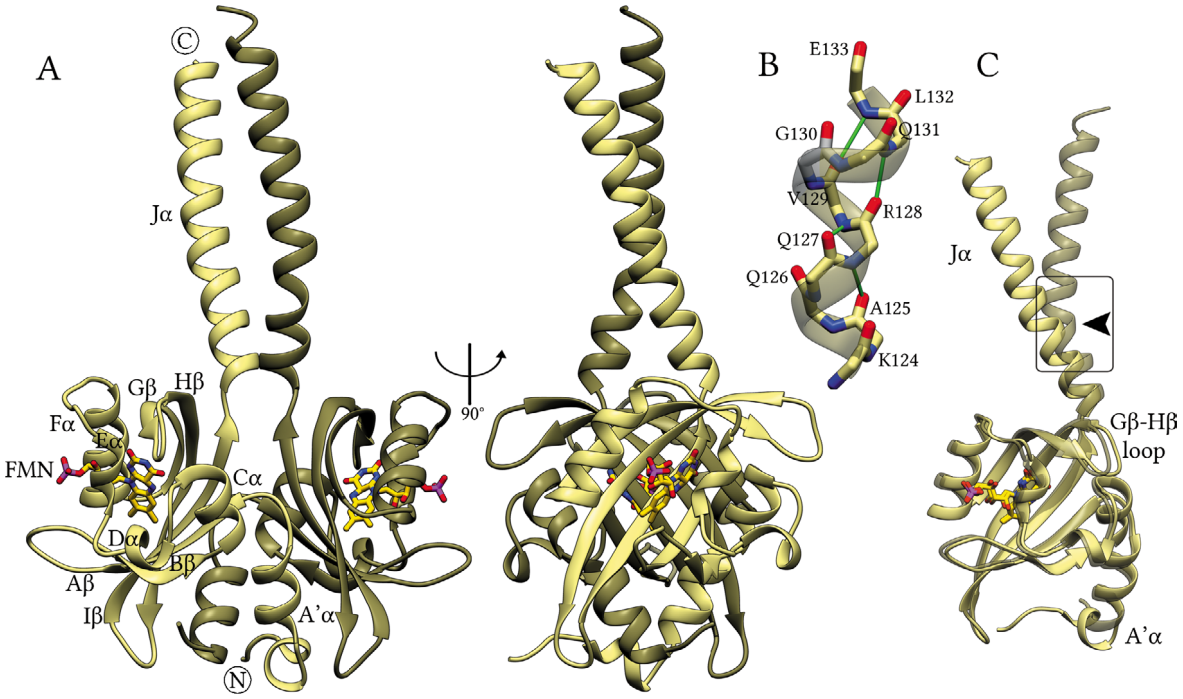


Figure 30. The crystal structure of the SBW25-LOV in the light state. The structure is represented as ribbons, where chains A and B are colored in light and dark shades of gold. FMN is shown as a stick model. (**A**) The dimer of SBW25-LOV that is observed in the crystal structure. (**B**) The residues at maximum kink of Chain B Jα helix, which is indicated by a box in (C) are shown as sticks. The hydrogen bonds between backbone N and O atoms are shown as green lines. (**C**) The superposition of the SBW25-LOV monomers by the core domain residues 17 – 117, which showed the differences in the Gβ-Hβ loop and the A'α, Jα helices. After superposition, the RMSD of the Cα atom positions is 0.89 Å over the residues 1 – 133.

SBW25-LOV in the light state

In the SBW25-LOV crystal structure, FMN was clearly identified in the $2mF_o$-DF_c electron-density map (Figure 31). This observation agrees with the HPLC finding that FMN dominates the chromophore content (~71%). The electron density of FMN atoms, Cys53 and a few residues in the chromophore pocket has an anisotropic distribution, while the others did not. Such anisotropic distribution can be the consequence, for example, of the collective motions of these atoms or can be the result of an averaging across the different conformations present in the crystal.

The continuous electron density between the Sγ atom of Cys53 residue and the C4a atom of FMN indicates the presence of covalent bond between them. In general, at the typical data resolution in protein crystallography, measurement errors contribute to the uncertainty in determination of atom positions and respective bond lengths. Often to improve the structure, restraints are used during refinement of crystal structures based on experimentally-derived data from protein structures with exceptional atomistic resolution (higher than ~1 Å) or from model peptides.

Till the date, there is no experimental data at atomistic resolution on the geometry of the FMN-cysteinyl adduct where the FMN-C4a atom is sp^3-hybridized. Sp^3-hybridization

should also result in the distortion of the planarity of the FMN tricyclic ring system. In this work, a model based on similar structures was created and its geometry was optimized with quantum chemical calculations (QM), additional details are in chapter 2.6 Quantum chemical calculations on page 42. The QM-optimized FMN was then used in the refinement of the SBW25-LOV crystal structure.

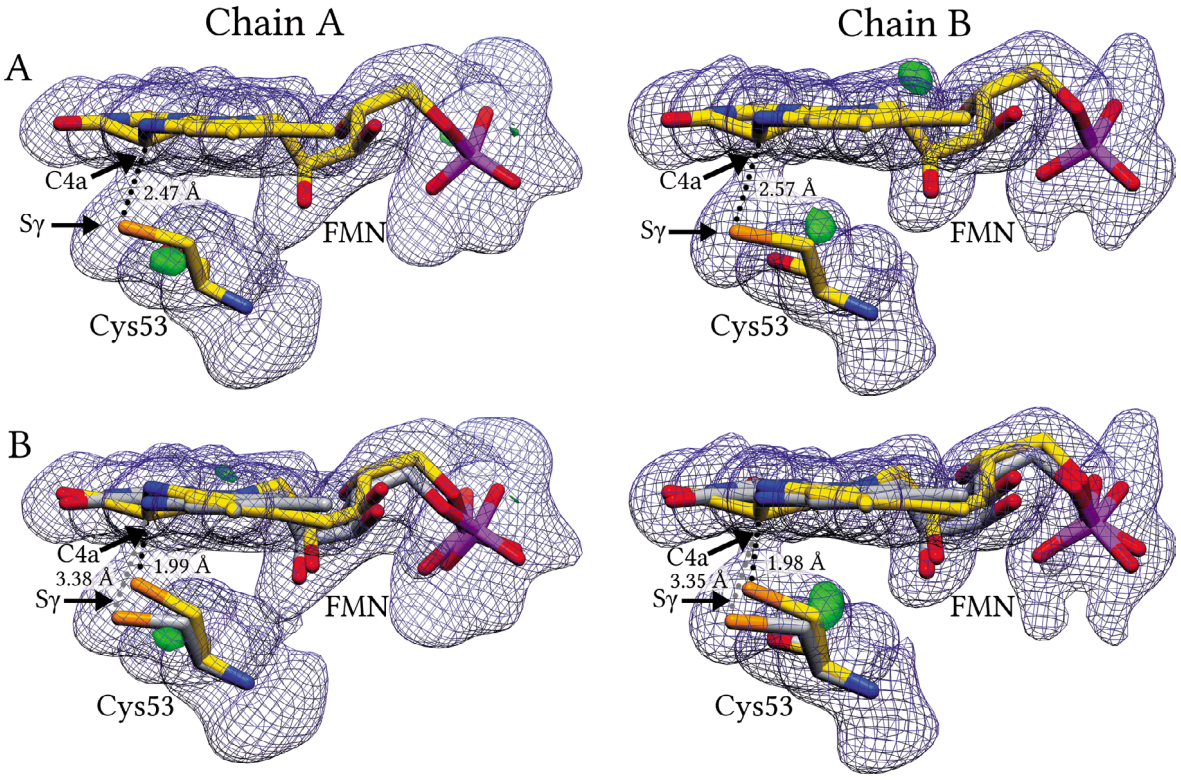


Figure 31. The FMN electron-density maps observed in the SBW25-LOV light state structure. $2mF_o$-DF_c electron-density maps contoured at 1 rmsd are shown as a blue mesh. mF_o-DF_c difference maps contoured at ± 3 rmsd are shown as a transparent surface, where difference maps with positive values are colored in green and with negative values are colored in red. FMN and the conserved Cys53 are shown as a stick model and colored by element: carbon, yellow; nitrogen, blue; oxygen, red; phosphorus, purple; sulfur, orange. (**A**) One conformation of the Cys53 and the "light state" FMN structure were used in the refinement. (**B**) Two conformations of the Cys53, and the "light state" and "dark state" FMN structures were used in the refinement, representing the dark and light states. The relative occupancies between the "light" and "dark" states were 0.49 / 0.51 (0.38 / 0.62 in chain B). The carbon atoms of "dark state" molecules are colored in grey and that of the "light state" in yellow.

In the absence of distance restraints between Cys53 and FMN, the distance between the Sγ atom of Cys53 and the C4a atom of FMN became 2.47 Å (2.57 Å in chain B) after structure refinement. The distance of ~2.5 Å is larger than the C-S covalent bond distance of 1.8 Å, which was previously assumed to be the bond distance for an FMN-cysteinyl adduct (Crosson and Moffat, 2002). The QM-optimized geometry of FMN-Cys structure, in contrast, predicted the distance to be 1.96 Å, which is closer to 2.5 Å but still shorter.

Solely from anisotropic distribution of the electron density of the Cys53-Sγ atom towards the FMN-C4a atom, it cannot be determined if the Sγ atom has strong anisotropic motion or the observed electron density is an average of two cysteine conformations. The UV-

Vis absorbance spectrum of the crystal showed that before X-ray exposure the chromophore was in the light state thus a covalent adduct was present. The UV-Vis spectrum of the crystal after X-ray exposure during data collection, in contrast, no longer showed spectrum identical to that of the light state. The observed spectrum instead showed apparent disruptions of the light state by scission of the adduct and also presence of the degradation products caused by radiation damage. This implies that observed electron-density maps is likely an average between the light state with the intact adduct and "disrupted state" with the disrupted adduct, which should be similar to the dark state.

Next, two conformations of FMN, in the “dark state” with planar ring system and in the “light state” with non-planar ring system, disrupted by the sp^3-hybridization of atoms involved in the FMN-cysteinyl adduct were modeled. Additionally, two conformations of the adduct-forming Cys53 residue were modeled. The relative occupancies of FMN and Cys53 conformations were coupled together. The distance between the “light state” FMN and one conformation of the Cys53 was constrained to 1.96 ± 0.04 Å, which was derived from the above-mentioned QM model. After refinement, no significant difference density was observed in the mF_o-DF_c map beyond the contouring at ± 3 rmsd levels, suggesting the validity of this interpretation.

The refined distance between the FMN-C4a and Cys53-Sγ atoms, which belong to the “light state” conformation, became 1.99 Å (1.98 Å in chain B), whereas that between the atoms, which belong to the “dark state”, was 3.38 Å (3.35 Å in chain B). The relative occupancies between the “light” and “dark” states were 0.49 / 0.51 (0.38 / 0.62 in chain B). Overall, the modeling of the two conformations seems to be valid from the crystallographic point of view and is supported by UV-Vis microspectrometry data of the protein crystals.

Crystal structure of Pf5-LOV

The crystal structure of Pf5-LOV in the dark state was determined at a resolution of 3.43 Å in $I4_122$ space group. The data collection and refinement statistics are listed in Table 26. The asymmetric unit contains a homodimer, which is formed by monomers with a non-crystallographic 2-fold symmetry (Figure 32). In the crystal structure, 149 of the 152 residues in Chain A (152 in Chain B) were observed. Superposition of monomers showed their similarity, with the root-mean-square deviation (RMSD) of the Cα atom positions of 0.89 Å over the residues 1 – 139. The Jα helix of chain B has a kink with an angle of 26°, identical as that in the SBW25-LOV structure, when measured between the two vectors made by the Cα atoms of residues, 125 to 128 and 128 to 131, with the HELANAL-Plus program (Kumar and Bansal, 2012). In comparison, chain A has maximum kink angle of only 15.8°, which is higher than 9.6° observed in the crystal structure of SBW25-LOV.

The diffraction data showed a relatively high Wilson B-factor of 169 $Å^2$, which indicates high flexibly in the structure. The estimated solvent content of the protein crystal

based on the Matthews probability was ~83%, which is unusually high compared to 40% to 60%, which is typically observed for protein crystals. Accordingly, the crystal had very loose packing with broad solvent channels. In this case, the high solvent exposure plus low data resolution with high B-factor resulted in a poor and often featureless $2mF_o$-DF_c electron-density map, such that usually only residues truncated to Cβ atom could be modeled, nevertheless almost all residues were observed in the electron-density map. The C-terminal Jα helices appeared as almost featureless tubes. Therefore, most of the analyses, which rely on the exact atom coordinates, were not performed due to the poor model of Pf5-LOV.

Table 26. Pf5-LOV crystal structures: data collection and refinement statistics.

X-ray data	**Pf5-LOV dark state**
Beamline	ID29, ESRF
Detector	PILATUS 6M
Wavelength [Å]	0.97625
Resolution range [Å]	48.12 – 3.43 (3.70 – 3.43)*
Space group	$I4_122$
Unit-cell a, b, c [Å]	148.70 148.70 200.79
α, β, γ [°]	90 90 90
Total reflections	104471 (10629)
Unique reflections	15387 (1516)
Multiplicity	6.8 (7.0)
Completeness [%]	99.6 (99.9)
Mean I/σ(I)	10.60 (0.72)
Wilson B-factor [Å^2]	169.32
R-merge	0.0872 (3.086)
R-meas	0.09467
$CC_{1/2}$	0.998 (0.354)
Refinement	
Resolution range [Å]	48.12 – 3.43 (3.55 – 3.43)
R-work	0.2658 (0.3806)
R-free	0.2986 (0.3778)
coordinate error (max.-likelihood based) [Å]	0.36
Number of non-hydrogen atoms	2226
macromolecules	2164
ligands	62
water	0
Protein residues	302
RMSD (bonds) [Å]	0.012
RMSD (angles) [°]	1.85
Ramachandran favored [%]	95
Ramachandran outliers [%]	0
Clashscore	19.01
Average B-factor [Å^2]	164.10
macromolecules [Å^2]	164.50
ligands [Å^2]	152.10
solvent [Å^2]	–

* Statistics for the highest-resolution shell are shown in parentheses

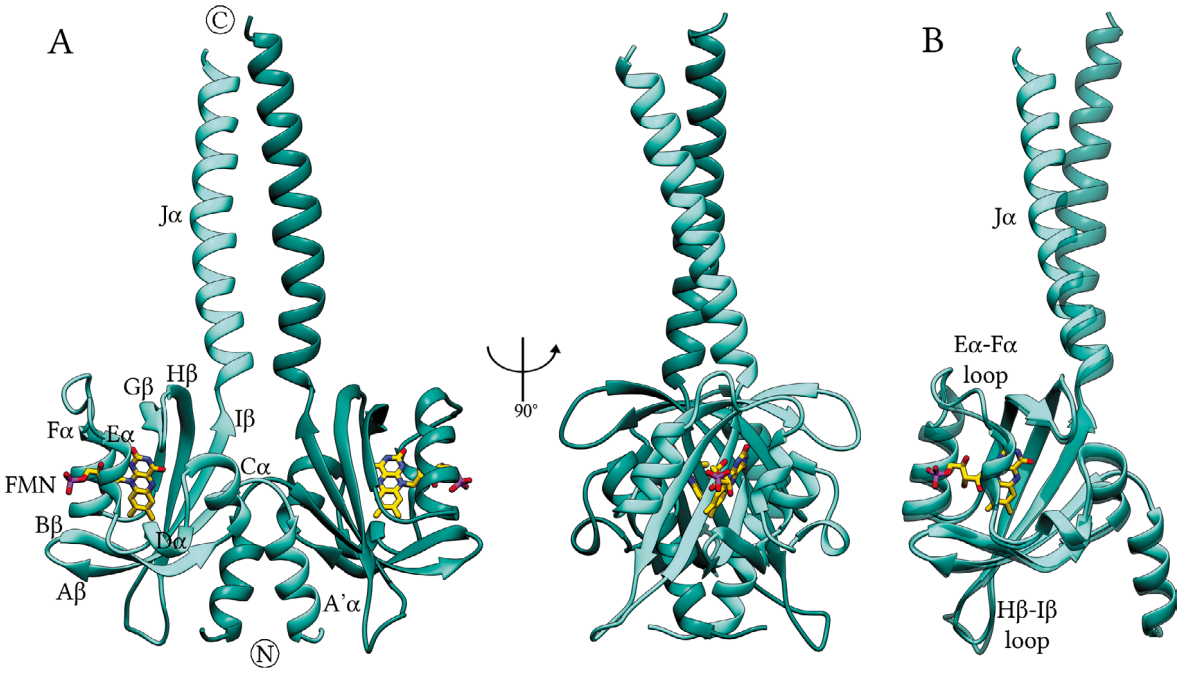


Figure 32. The crystal structure of Pf5-LOV in the dark state. The structure is represented as ribbons, where chains A and B are colored in light and dark shades of cyan. FMN is shown as a stick model. (**A**) The dimer of Pf5-LOV that is observed in the crystal structure. (**B**) The superposition of the Pf5-LOV monomers by the core domains residues 17 – 117, which showed the differences in the Hβ-Iβ and Eα-Fα loops and the Jα helices. After superposition, the RMSD of the Cα atom positions is 0.89 Å over the residues 1 – 139.

Dimer interface in crystal and solution

The PISA analysis (Krissinel and Henrick, 2007) of the SBW25-LOV dimer observed in the crystal structure has calculated total buried surface area of 4157 Å^2 with a -37.1 kcal/mol solvation energy gain upon dimer formation, indicative of a stable dimer in solution. The dimer interface is stabilized by hydrophobic interactions as well by hydrogen bonds and salt bridges (Table 27). Interfacial residues are predominantly situated in the A'α helix, Jα helix and spread throughout the β-strands. The sequence of the Jα helix has characteristic heptad repeats which lead to the coiled-coil like interactions that greatly contribute to the stabilization of the dimer interface.

Analytical size-exclusion chromatography (SEC) gave a molecular weight (MW) estimation of SBW25-LOV to be 53 kDa, which is higher than the MW calculated for the dimer based on its sequence (Table 28). The SEC gave a MW estimation of Pf5-LOV to be 64 kDa, which is also higher than the MW calculated for the dimer. An overestimation of the molecular weight could be due to the non-globular shape of the proteins. Small-angle X-ray scattering (SAXS) data collection for Pf5-LOV was performed under dark and light conditions, whereas for SBW25-LOV it was done only under light conditions due to the long adduct lifetime. The SAXS data for Pf5-LOV showed that both measurements were identical, probably due to the limitations of the method to detect the shape changes of such a small protein.

Table 27. The interfacial residues of SBW25-LOV monomers in the dimer within the hydrogen bond distance of ≤ 3.2 Å, which are observed in the crystal structure.

Secondary structure element	Residue of chain A	Atom	Distance [Å]	Atom	Residue of chain B	Secondary structure element
A'α	Gln8	NE2	3.1	O	Ile2	–
A'α	Ala13	O	2.7	OG	Ser100	Hβ
			2.8	OG	Ser100	Hβ
A'α-Aβ loop	Asn15	ND2	3.1	OG	Ser98	Hβ
A'α-Aβ loop	Asp16	OD1	3.0	NH1	Arg80	Gβ
			2.8	NH2	Arg80	Gβ
			2.9	OG	Ser98	Hβ
Gβ	Arg80	NH2	2.7	OD1	Asp16	A'α-Aβ loop
Hβ	Glu96	OE1	2.9	NZ	Lys117	Iβ
Hβ	Ser98	OG	2.8	OD2	Asp16	A'α-Aβ loop
			2.8	ND2	Asn15	A'α-Aβ loop
Hβ	Ser100	OG	2.8	O	Ala13	A'α
Jα	Arg128	NE	2.9	OE2	Glu133	Jα
Jα	Arg128	NH2	2.9	OE2	Glu133	Jα

The estimation of the MW from the SAXS data based on the calculation of the Porod volume and based on the I_0 value also agreed well with the calculated MW of the dimer. Altogether, the MW and radius of gyration (R_g) estimations consistently showed evidence of a dimer for both proteins.

Table 28. An overview of determined molecular weight (MW) and SAXS data of SBW25-LOV and Pf5-LOV. The MW from Porod volume was calculated by dividing volume by the factor 1.7.

Protein	Calculated MW of monomer / dimer [kDa]	SEC MW [kDa]	SAXS MW from I_0 [kDa]	SAXS MW from Porod volume [kDa]	Crystal structure R_g [nm]	Guinier R_g SAXS [nm]
SBW25-LOV	17.5 / 35.0	53	33.9	37.5	2.4	2.6
Pf5-LOV	17.3 / 34.6	64	26.7	42.4	2.5	3.0

Additionally, the theoretical scattering curves of the crystal structures (dimer and monomer) were calculated and fitted to experimental scattering data (Figure 33). The dimer scattering curves fitted much better than that of the monomers with the χ^2 values of 1.4 and 5.2 for the dimers of SBW25-LOV and Pf5-LOV, respectively. The higher χ^2 value of Pf5-LOV might be due to the fact that ~25% amino acid residues in the crystal structure were only modeled upto the Cβ atom due to the poor quality of the data, thus their contribution to the scattering curve calculation was missing. Moreover, the *ab initio* envelope models calculated from the SAXS data matched well to both dimers seen in the crystal structures (Figure 34).

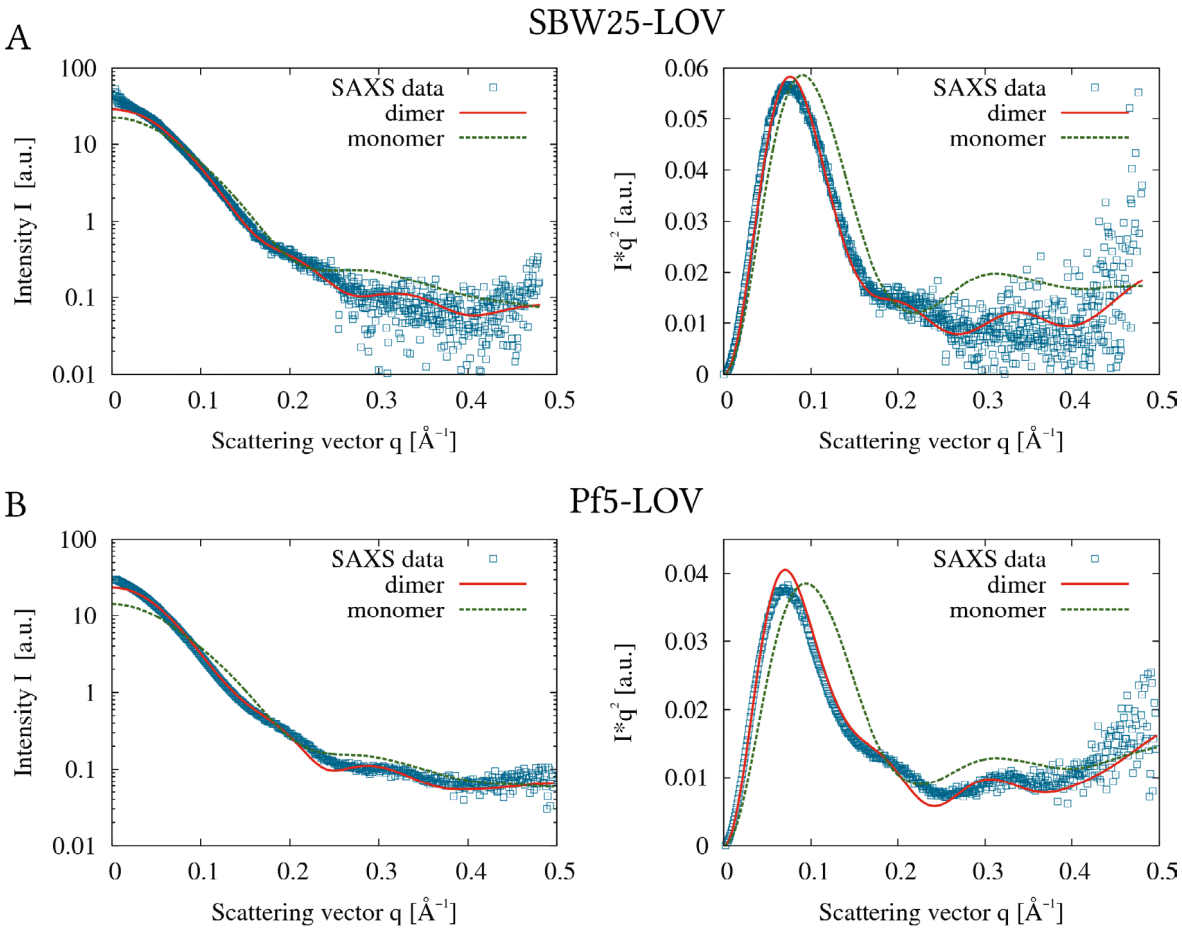


Figure 33. The SBW25-LOV and Pf5-LOV SAXS scattering curves fitted to theoretical scattering curves of their dimers and monomers. (**A, left**) The theoretical scattering curves of the dimer (solid line) and monomer (dashed line) of the SBW25-LOV crystal structure are shown as lines, and the experimental SAXS data are shown as squares. (**A, right**) the same data as in (A, left) represented as a Kratky plot. (**B**) The same data representation for Pf5-LOV as in (A).

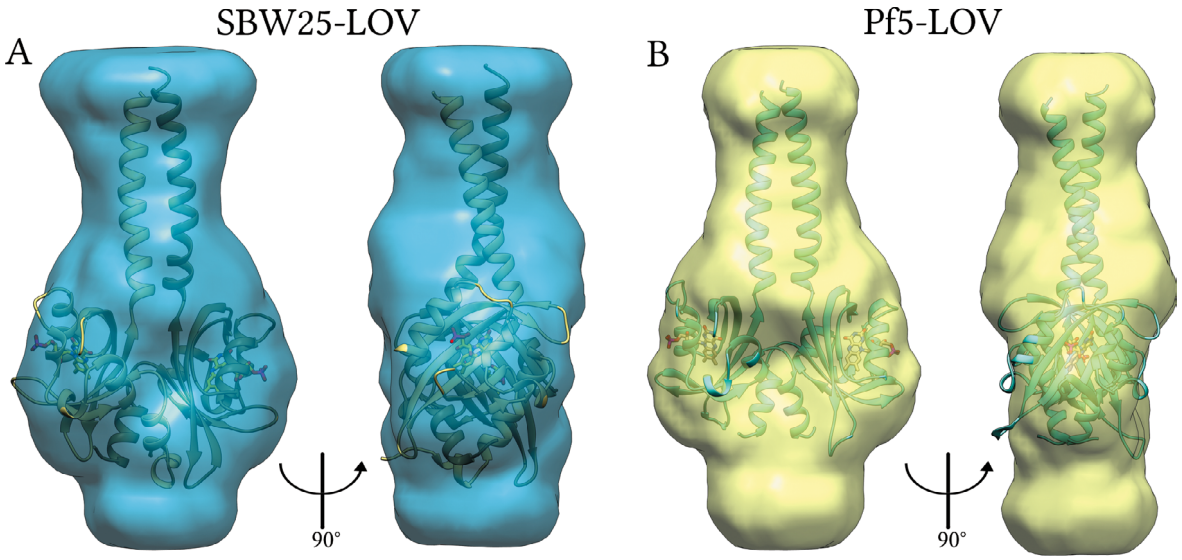


Figure 34. The SAXS envelope models of SBW25-LOV and Pf5-LOV aligned with their respective crystal structures. FMN is shown as a stick model. (**A**) The SAXS envelope model of SBW25-LOV is shown as a blue transparent surface and its crystal structure as a ribbon diagram in gold. (**B**) The SAXS envelope model of Pf5-LOV is shown as a yellow transparent surface and its crystal structure as a ribbon diagram in cyan.

Comparison between SBW25-LOV and Pf5-LOV

SBW25-LOV and Pf5-LOV are homologous short LOV proteins from strains SBW25 and Pf-5 of *Pseudomonas fluorescens*, respectively. They share a sequence identity of about 73%, see sequence alignment in Figure 35. Nevertheless, they differ dramatically in their adduct lifetimes, 1470 min vs 3.6 min at 37 °C, respectively. Most of the sequence differences are located in the A'α, Fα, Jα helices, the Aβ-Bβ, Hβ-Iβ loops and the Hβ strand.

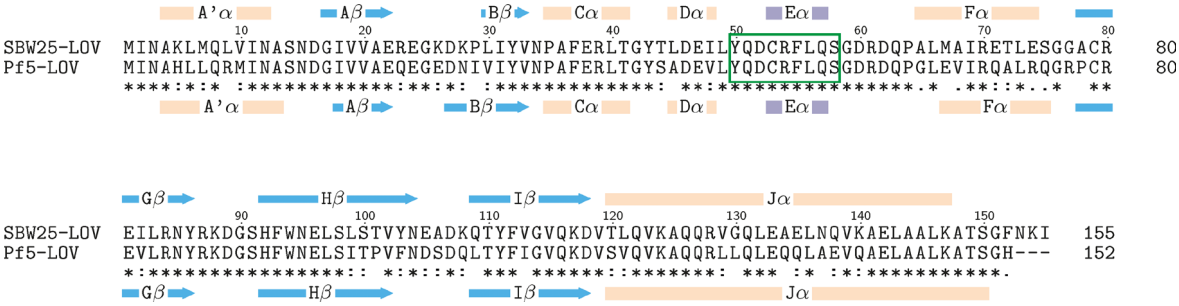


Figure 35. The sequence alignment of SBW25-LOV with Pf5-LOV. The secondary structure elements for SBW25-LOV are shown on top, and for Pf5-LOV on bottom, where: arrow – β-strand and cylinder – α-helix (apricot) and 3_{10}-helix (plum). The asterisks at the bottom line of the alignment indicate identical residues in a given sequence position, while single and double dots refer to highly and moderately similar residues, respectively.

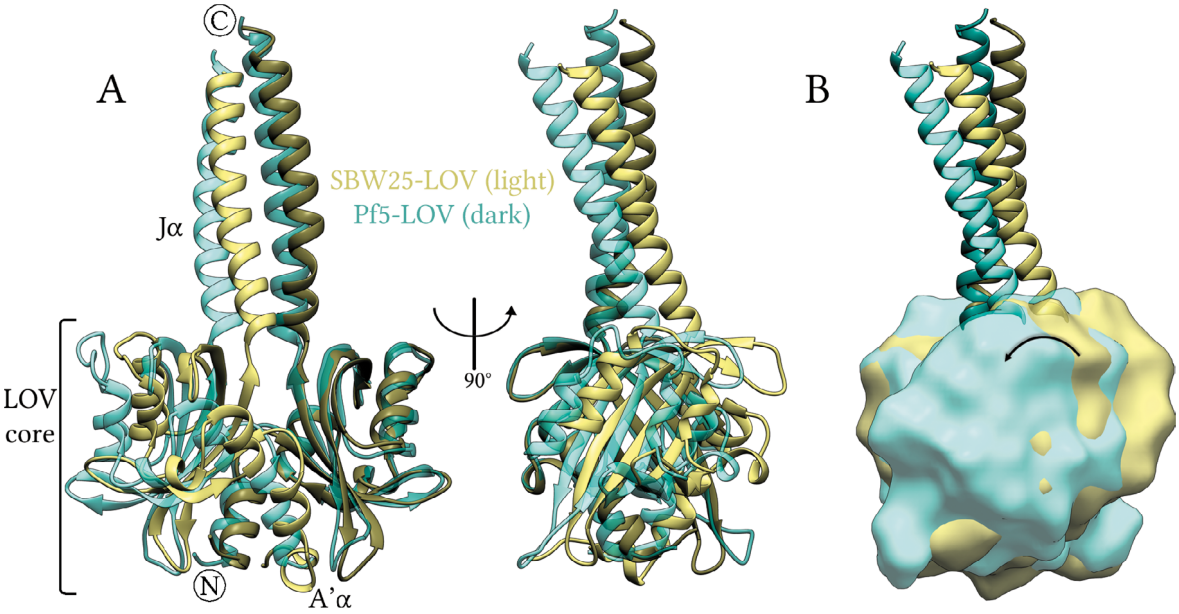


Figure 36. The superposition of the light state structure of SBW25-LOV with the dark state structure of Pf5-LOV. (**A**) The structure of SBW25-LOV dimer in the light state (gold) and the structure of Pf5-LOV dimer in the dark state (cyan) are shown as ribbons, where chains A and B are colored in light and dark shades. Their superposition by the core domain residues 17 – 117 of chain B shows the dimer interface rotation. The RMSD between the core domain residues 17 – 117 of SBW25-LOV and Pf5-LOV is 1.1 Å and 0.9 Å for their A chains and B chains, respectively (**B**) The same superposition as in (A) where LOV core domains are represented as surfaces, and Jα helices as ribbons to emphasize the dimer interface rotation.

The superposition of the SBW25-LOV dimer in the light state with the Pf5-LOV dimer, due to their high sequence and structural similarity, might demonstrate the changes between dark and light states of them. The superposition of dimers results in a poor match,

due to the difference in the dimer interface, the superposition by the core domain residues 17 – 117 of one monomer to another one can better depict this change (Figure 36A). Essentially, the monomers rotate around a virtual axis going through the FMN molecules of both monomers (Figure 36B). This is similar to the differences observed between the dimer interfaces of PpSB2-LOV in the dark state and PpSB1-LOV in the light state, as shown in a simplified cartoon in Figure 21C. However, the longer Jα helices in SBW25-LOV and Pf5-LOV, in contrast to those in PpSB1-LOV, has a kink in one α-helix. The motion of the helices can be then described as the screw-like motion of a kinked α-helix, where this helix moves towards the C-terminus following the rotation of core domain. This movement might describe a way of mechanistic signal transduction from the chromophore of the LOV domain to the fused effector domain via the C-terminal helices.

Comparison between SBW25-LOV and PpSB1-LOV in the light state

The superposition of the SBW25-LOV dimer with the PpSB1-LOV dimer (PDB ID: 3SW1, Circolone *et al.*, 2012) (Chain B and D) both in the light states, due to their sequence and structural similarity, might demonstrate features common for *Pseudomonas* short LOV proteins. Indeed, the superposition of dimers resulted in an overall good match (Figure 37A). The core domains overlaid well with some differences observed in the following loops: Aβ-Bβ, Eα-Fα and Hβ-Iβ. The different arrangement of the SBW25-LOV A'α helix (chain A) is solely due to the crystal contacts induced by the expression-tag. Superposition of the SBW25-LOV monomers with the PpSB1-LOV monomers showed that the crossing angle of the Jα helices is similar up to residue 125, whereas beyond that the longer Jα helices of SBW25-LOV are kinked.

A similar kinking and super-twisting of the helices was observed in MD simulation of the PpSB2-LOV model in the light state shown in this work, as mentioned previously on the page 62. Indeed, the superposition of the averaged model of PpSB2-LOV in the light state, which was derived from MD simulation, with the structure of SBW25-LOV showed that their structures are similar, as depicted in Figure 38.

FMN in the binding pocket of LOV proteins is tightly coordinated by hydrogen bonds, salt bridges and hydrophobic interactions. Most of the residues are highly conserved and are indispensable for the optimal positioning of the chromophore. The SBW25-LOV residues within ≤ 4 Å distance of FMN are shown in Figure 39A, and the residues within ≤ 3.4 Å distance are shown in Table 29. The Asp52 residue, preceding the adduct-forming Cys53, forms hydrogen bond with the OH groups of the FMN ribityl chain. The Asn85 residue forms hydrogen bonds to the N3 and O2 atoms of the FMN ring system. Whereas Gln57 residue interacts with Arg61, FMN-O2 of the ring system and FMN-O4' of the ribityl chain. Two conserved arginine residues, Arg54 and Arg70, form salt bridges to the phosphate group of FMN. Additionally, the non-conserved Arg61 forms a salt bridge to the FMN phosphate group, and at the same it is also coordinated by the Gln57 and Arg70 residues.

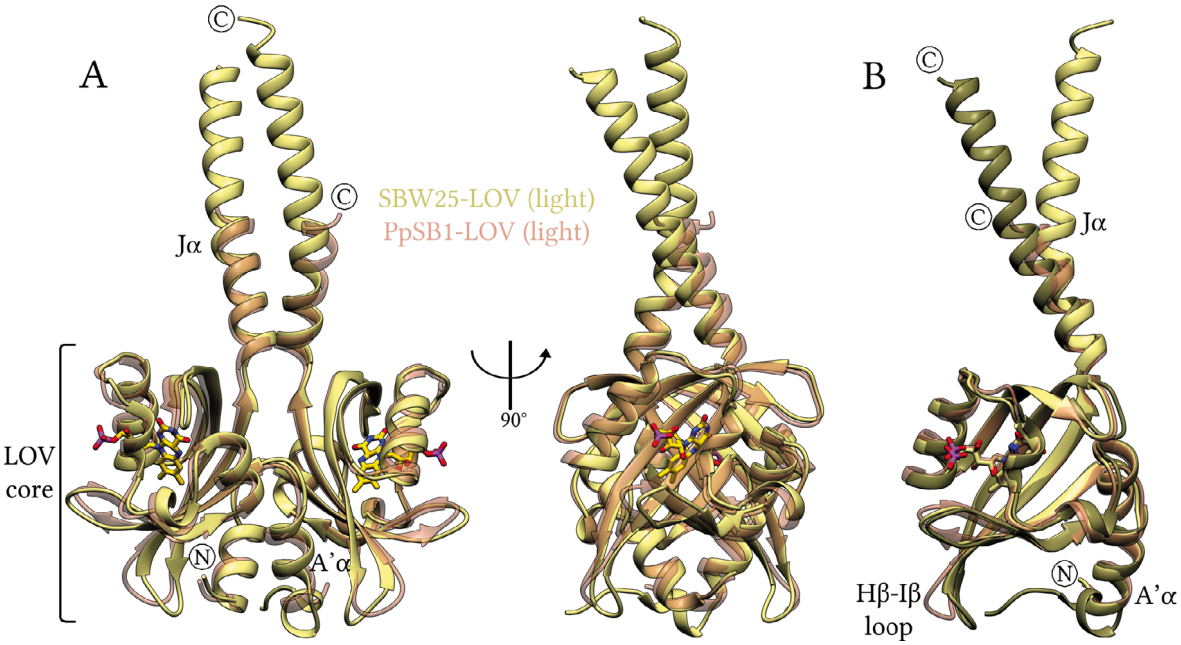


Figure 37. The superposition of SBW25-LOV and PpSB1-LOV structures in the light state. The structures of SBW25-LOV (gold) and PpSB1-LOV (PDB ID: 3SW1, Circolone *et al.*, 2012) (coral) are shown as ribbons. FMN is shown as a stick model. (**A**) The superposition of SBW25-LOV and PpSB1-LOV dimers in the light state, which were observed in the crystal structures. The RMSD of the Cα atom positions between the SBW25-LOV and PpSB1-LOV dimers was 0.98 Å over the residues 1 – 133. (**B**) The superposition of the two SBW25-LOV monomers (dark and light gold) with the PpSB1-LOV monomer (chain A) by the core domain residues 17 – 117. After superposition, the RMSD of the Cα atom positions of the PpSB1-LOV monomer (chain A) to the SBW25-LOV monomers is 0.84 Å and 0.94 Å, respectively.

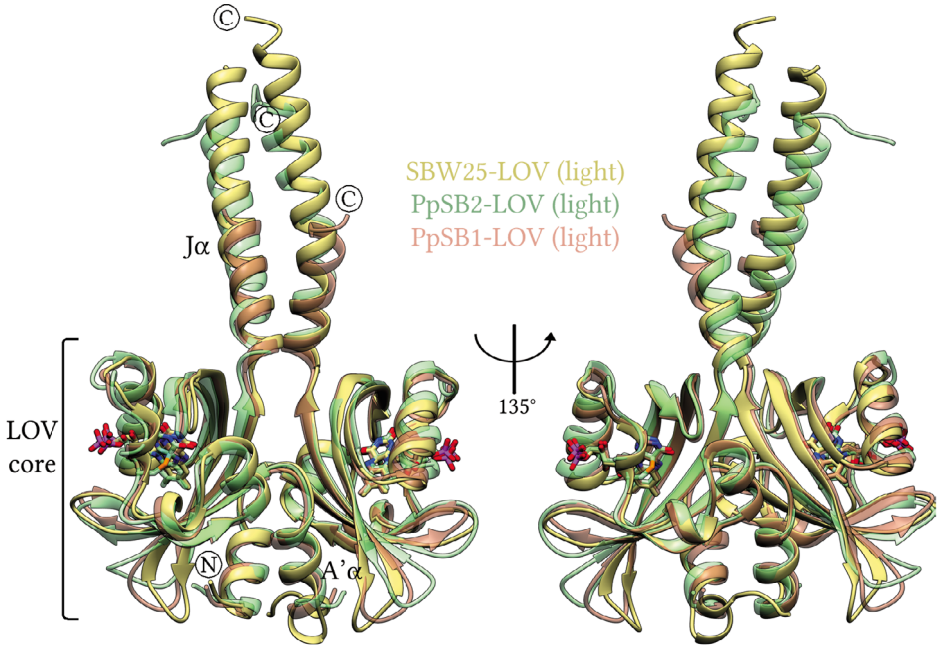


Figure 38. The superposition of SBW25-LOV and PpSB1-LOV structures in the light state with the averaged structure of PpSB2-LOV in the light state, which was derived from the MD simulation (see page 62). The structures of SBW25-LOV (gold), PpSB1-LOV (PDB ID: 3SW1, Circolone *et al.*, 2012) (coral) and PpSB2-LOV (green) model are shown as ribbons. FMN and the FMN-cysteinyl adduct are shown as stick models. After superposition, the RMSD of the Cα atom positions between the SBW25-LOV and PpSB2-LOV model is 1.58 Å over the residues 1 – 143, and between the SBW25-LOV and PpSB1-LOV is 0.98 Å over the residues 1 – 133.

As a comparison, the residues within the ≤ 4 Å distance of PpSB1-LOV in the light state structure (Circolone *et al.*, 2012) are shown in Figure 39B. PpSB1-LOV, in contrast to SBW25-LOV, has an additional non-conserved Arg66 residue. Together, the four arginine residues (Arg54, Arg61, Arg66 and Arg70) were previously described as being responsible for the slow dark recovery via their tight coordination of FMN (Circolone *et al.*, 2012). The four arginine residues have similar arrangement to the three arginine residues (Arg54, Arg61 and Arg70) of SBW25-LOV, yet with some differences. The three arginine residues of SBW25-LOV seems to be arranged more effectively around the phosphate group of FMN, having the three pairs of phosphate group's oxygens equally distributed between the arginine residues for better salt bridging. Whereas in PpSB1-LOV, the four arginine residues compete, what apparently led to rotation of the Arg70 side chain away, leaving only the NH group for bonding. Similarly, in PpSB1-LOV, the guanidinium group of Arg61 uses only one NH_2 group for coordination of the phosphate group instead of coordination with both NH_2 groups.

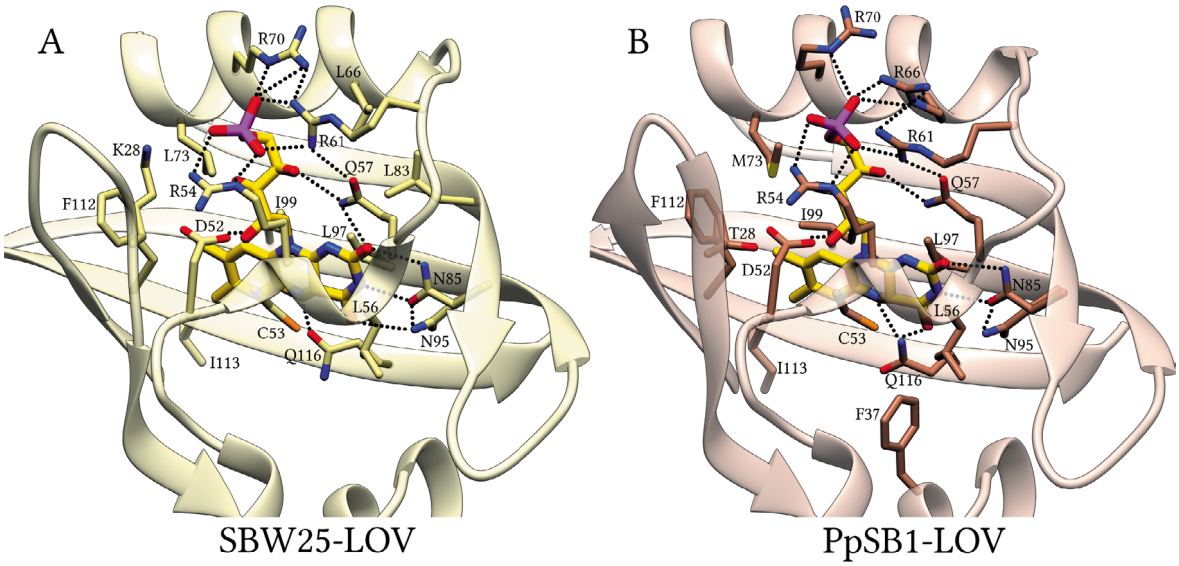


Figure 39. Comparison of the chromophore binding pockets of SBW25-LOV and PpSB1-LOV in the light states. (**A**) the SBW25-LOV (gold) and (**B**) the PpSB1-LOV (PDB ID: 3SW1, Circolone *et al.*, 2012) (coral) structures are represented as ribbons. FMN and the interacting residues, which are within ≤ 4 Å distance are shown as stick. The carbon atoms are colored in model color. The hydrogen bonds and salt bridges within ≤ 3.2 Å distance to FMN are shown as dotted lines.

The conserved residue Gln116 was reported as one of the key residues for signaling (Crosson and Moffat, 2002; Zoltowski *et al.*, 2007; Vaidya *et al.*, 2011). Residue Gln116 in SBW25-LOV is only within hydrogen bonding distance to the FMN-N5 atom, whereas in PpSB1-LOV, Gln116 residue can apparently participate in two hydrogen bonds, with the N5 and O4 atoms of FMN.

Sequence alignment of SBW25-LOV with the other LOV proteins discussed in this study showed that it has a sequence identity with PpSB1-LOV and PpSB2-LOV of ~60% (Table 2). The differences between their adduct lifetimes should be intrinsically related to their primary structure. To address this, all three crystals structures were superposed by their

core domain residues 17 – 117 and shown in Figure 40A, where the SBW25-LOV and PpSB1-LOV are the light state structures and the PpSB2-LOV is the dark state structure.

Table 29. The residues of SBW25-LOV, which are within the hydrogen bond distance ≤ 3.2 Å to FMN.

Chain A					Chain B				
Ligand	Atom	Distance [Å]	Atom	Residue	Ligand	Atom	Distance [Å]	Atom	Residue
FMN500	O2	3.1	NE2	Gln57	FMN500	O2	3.0	NE2	Gln57
		3.1	ND2	Asn85			3.0	ND2	Asn85
	N3	2.9	OD1	Asn85		N3	2.9	OD1	Asn85
	O4	3.1	ND2	Asn95		O4	3.2	ND2	Asn95
	N5	3.0	OE1	Gln116		N5	3.2	OE1	Gln116
	O2'	2.6	OD1	Asp52		O2'	2.6	OD1	Asp52
	O4'	3.0	NE2	Gln57		O4'	3.0	NE2	Gln57
	O1P	2.6	NE	Arg54		O1P	2.8	NE	Arg54
		3.3	NH1	Arg61			2.9	NH1	Arg61
		3.3	NH2	Arg61		O2P	2.8	NH2	Arg54
	O2P	2.8	NH2	Arg54		O3P	2.4	NE	Arg70
	O3P	2.2	NE	Arg70			3.0	NH2	Arg61
		3.0	NH2	Arg70			3.0	NH2	Arg70

Superposition not only showed different conformations of the Bβ strand and the Aβ-Bβ loop but also that these changes could be related to differences in their primary structure. PpSB1-LOV and PpSB2-LOV have the Thr28 and Ser28 residues respectively, which form hydrogen bonds to the water molecule, which is coordinated by the OH groups of the FMN ribityl chain. SBW25-LOV, in contrast, has the Lys28 residue, which can form hydrogen bond not only to a water molecule but also to Glu74 residue, thus closing the pocket from the solvent on that side of the protein. In the sequence alignment with the related LOV proteins, only the slowest protein SBW25-LOV has this combination of Lys28 and Glu74 residues, whereas others have either Arg74 or Ala74 residues.

Additionally, Arg23 residue in SBW25-LOV is also within the hydrogen bond distance to Glu74 residue, thus enabling better closure of the pocket from the solvent on that side of the protein. Such interactions cannot take place in PpSB1-LOV, as the amine group of the Lys23 side chain cannot interact with the Ala74 side chain. Residue Arg74 in PpSB2-LOV prefer interaction with Arg70 rather than Gln23 residue, moreover, Gln23 already form a hydrogen bond to the Tyr112 side chain. A consequence of such tight interactions in SBW25-LOV is readily seen in the solvent-accessible cavity produced by rolling a probe with 1.4 Å radius, as depicted in Figure 40B. The solvent-accessible cavity of SBW25-LOV is compacter than that of PpSB1-LOV, and even more than that of PpSB2-LOV, which has the fastest dark recovery among them (Figure 20B). This indicates that solvent accessibility to the ring system of FMN correlates with the adduct lifetime.

The Pf5-LOV crystal structure has poor resolution, yet the above-mentioned hypothesis seem to be valid as well. It was previously emphasized that a unique four arginine residues cluster in PpSB1-LOV is responsible for its slow dark recovery (Circolone *et al.*, 2012). However, Pf5-LOV has also three arginine residues like SBW25-LOV does, which seem to retain a similar coordination of the FMN phosphate group. Nevertheless, its adduct lifetime is much shorter than that of SBW25-LOV. Pf5-LOV has an identical combination of the Arg74 and Gln23 residues as in PpSB2-LOV, which are arranged in a similar way. Phe112 residue instead of Tyr112 residue in PpSB2-LOV, prevents the hydrogen bonding to Gln23 residue, as seen in PpSB2-LOV. Such sequence differences of Pf5-LOV to SBW25-LOV can be responsible for the faster dark recovery of Pf5-LOV, by analogy with PpSB2-LOV.

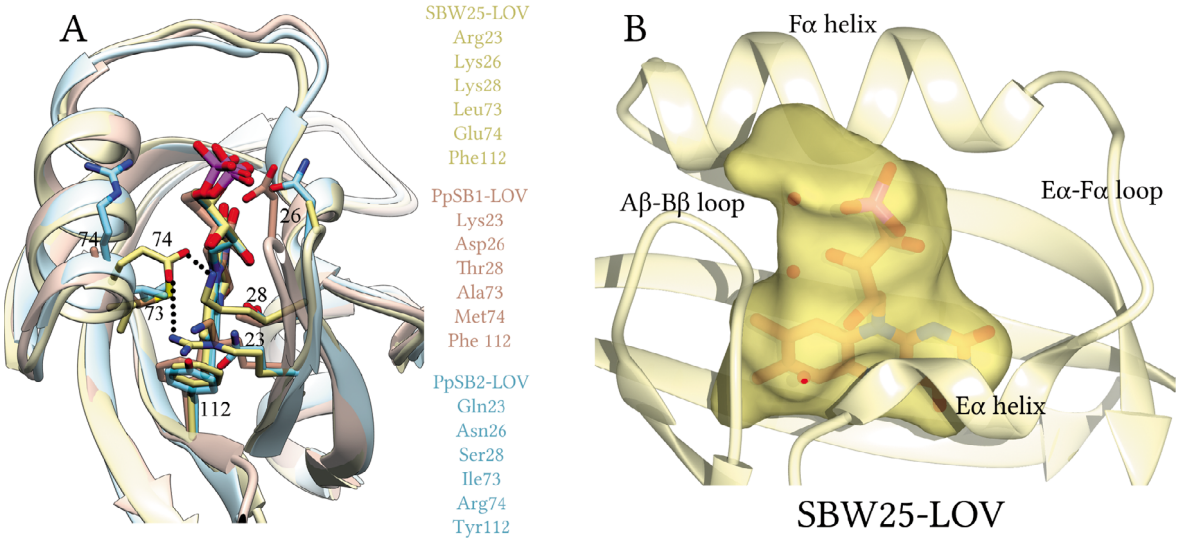


Figure 40. The superposition of the chromophore binding pockets of SBW25-LOV, PpSB1-LOV and PpSB2-LOV with each other and the solvent-accessible cavity of the chromophore pocket of SBW25-LOV. (**A**) Structures of SBW25-LOV (gold), PpSB1-LOV (coral) and PpSB2-LOV (blue) are shown as ribbons. After superposition by the core domain residues (17 – 117), the RMSD of the Cα atom positions between SBW25-LOV and PpSB1-LOV is 0.8 Å and between SBW25-LOV and PpSB2-LOV is 0.7 Å. FMN and selected residues are shown as stick models. The carbon atoms are colored in model color. The hydrogen bonds and salt bridges within ≤ 3.2 Å distance to FMN are shown as dotted lines. (**B**) The solvent-accessible cavity of SBW25-LOV (gold) was created with a 1.4 Å probe radius and represented as a transparent surface. The structure of SBW25-LOV (gold) is shown as a ribbon. FMN is shown as a stick model and water molecules are shown as red spheres. FMN and water molecules were not included in the calculation of the cavities but shown to indicate their positions in the respective structures.

Molecular dynamics

Molecular dynamics (MD) simulation of SBW25-LOV in the light state was performed to address the lack of information about its dynamics in the light state. The aforementioned MD simulations of PpSB2-LOV and PpSB1-LOV in the light state have shown that the increased mobility of the residues interacting with FMN might be related to their differences in the dark recovery kinetics. In the similar way as for PpSB2-LOV and PpSB1-LOV, a 200 ns MD simulation of SBW25-LOV in the light state was performed.

The crystal structure of SBW25-LOV in the light state was used as the initial structure, see details in the Material and Methods section. Overall, the structure remained stable over the course of the simulation, where higher fluctuations of the RMSD can be accounted for the higher dynamics of the Jα helices (Figure 41C). In the crystal structure of SBW25-LOV, the Jα helices are arranged into a coiled-coil-like interaction, where one helix has a kink. The Jα helices adopted various conformations during MD simulation, see snapshots of the trajectory in Figure 41. These conformations are essentially variation of super-twisting of the helices, concerted movement of the helices as a whole relative to the core domain and kinking of the helix, which was happening only for one helix at a time.

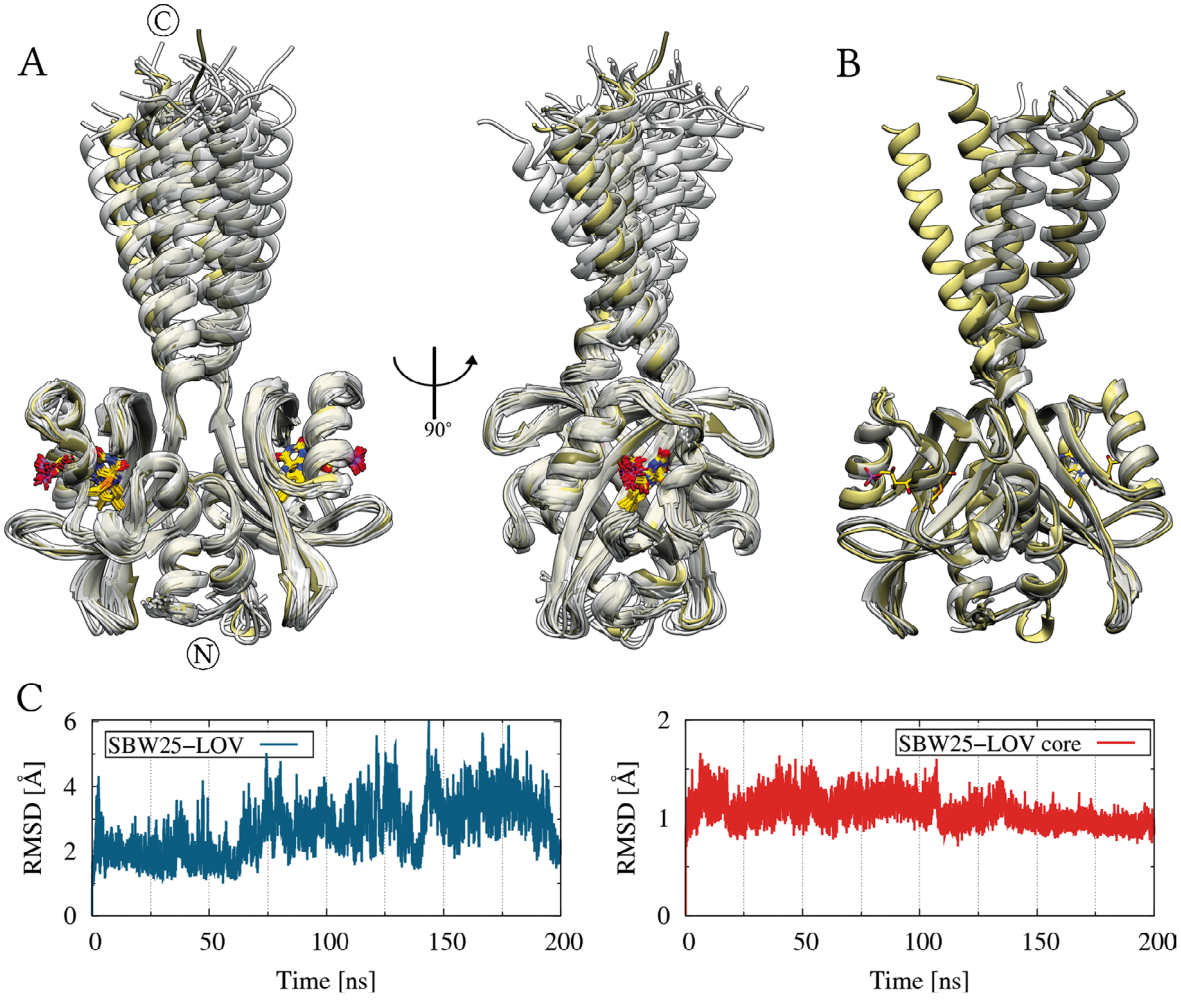


Figure 41. MD simulation snapshots of SBW25-LOV in the light state and the RMSD. (**A**) MD simulation snapshots, with Δt = 5 ns, are shown as ribbons, where the first snapshot is colored in light gold, the last snapshots in dark gold and the snapshots in between in gray. FMN and the FMN-cysteinyl adduct are shown as stick models. (**B**) Selected MD simulation snapshots of SBW25-LOV in the light state, which showed the strongest differences in the orientations of the Jα helices. (**C**) The RMSD calculated for the whole trajectory of SBW25-LOV MD simulation with respect to the starting structure. (**C, left**) The RMSD calculated for the dimer (398 residues). (**C, right**) The RMSD calculated only for the core domain residues 17 – 117 of the dimer.

The kink of one helix, observed in the crystal structures of SBW25-LOV and Pf5-LOV, might not be a coincidence, however, no specific kink-inducing sequence motif could be identified.

Similar kinking of the helix was also observed in previously described MD simulation of PpSB2-LOV in the light state (Figure 27).

Residue-resolved root-mean-square fluctuation (RMSF) of the core domain residues 17 – 119 of SBW25-LOV was calculated during its MD simulation in the light state and compared to that of PpSB2-LOV and PpSB1-LOV MD simulations in the light states (Figure 42). The overall RMSF of main chain atoms (N, C, O, and Cα) for most of the structure remained similar to the respective RMSF of PpSB2-LOV and PpSB1-LOV. However, lower RMSF values were observed in some regions of SBW25-LOV in comparison to those in PpSB2-LOV and PpSB1-LOV. The Aβ-Bβ loop in SBW25-LOV became more rigid, probably due to the presence of the Pro29 residue, which reduced the conformational freedom of the loop. Additionally, the presence of the above-mentioned interactions of the Lys28 and Arg23 residues with the Glu74 residue also stabilized the Aβ-Bβ loop.

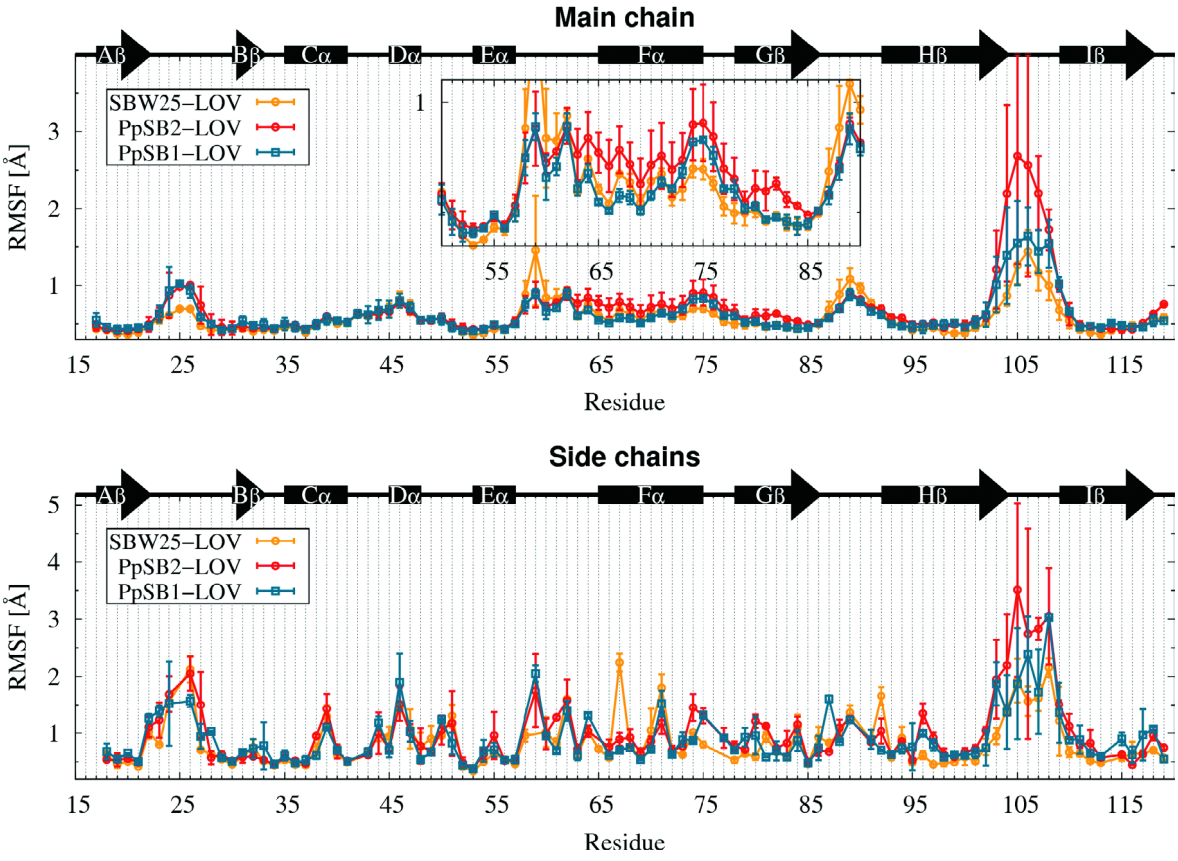


Figure 42. Residue-resolved RMSF of MD simulation of the SBW25-LOV, PpSB2-LOV and PpSB1-LOV in the light states. The bars indicate the values of two monomers, whereas the points indicate their average. The secondary structure elements of SBW25-LOV are show as arrows for strands and rectangles for helices. Secondary structure assignment was done with the DSSP program for the SBW25-LOV crystal structure.

Similarly, the Hβ-Iβ loop became even more rigid, which may be related to the sequence divergence of this loop even among homologous short LOV proteins in this study, see sequence alignment in Figure 5. The primary structure of the Gβ-Hβ loop, on the contrary, is more conserved, yet these residues showed higher dynamics of the main chain

atoms than those in PpSB2-LOV and PpSB1-LOV. Additionally, the Gly59 residue in one of the protein chains showed high RMSF value of the main chain atoms, which is probably related to its peptide-plane flipping.

In the crystal structure of SBW25-LOV, the Lys28, Glu74 and Arg23 residues interact with each other through hydrogen bonding, as mentioned earlier (Figure 40), and thus stabilize the chromophore pocket from that side of the protein. In fact, RMSF values are lower for the side chains of these residues in the Aβ-Bβ loop, the C-terminal part of the Fα helix and the regions around them when compared to those of PpSB2-LOV and PpSB1-LOV. Furthermore, the Fα helix and the Gβ strand of SBW25-LOV, have two glycine residues, Gly76 and Gly77, which probably contribute to the conformational freedom of the Fα helix, thus enabling it to adapt a more stable conformation. As a result of this, the main chain and side chain atoms of the residues around them showed reduced RMSF values.

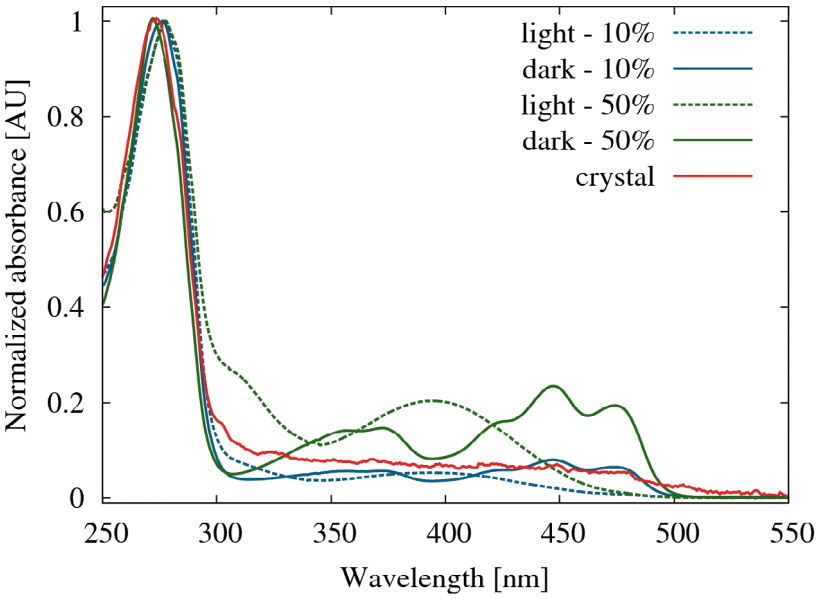
In the crystal structure of SBW25-LOV, three arginine residues (Arg54, Arg61 and Arg70) located in the Eα and Fα helices and the Eα-Fα loop, coordinate the phosphate group of FMN, perhaps more effectively than four arginine residues in PpSB1-LOV. This seems to correlate with the observed reduction in dynamics of the Eα-Fα loop and the Fα helix. The observed dynamics reduction of SBW25-LOV compared to that of PpSB1-LOV and PpSB2-LOV can be related to the decreased solvent accessibility, which is believed to stabilize the FMN-cysteinyl adduct formed in the light state and slow the recovery to the dark state.

3.3 W619_1-LOV

Biophysical and biochemical characterization

UV-Vis spectroscopy and single crystal microspectrometry

During heterologous overexpression under different conditions, W619_1-LOV protein samples with different chromophore loading were obtained. Protein overexpression at 14 °C resulted in ~10% chromophore loading, whereas at 30 °C and after additional *in vitro* loading with FMN resulted in ~50% chromophore loading. W619_1-LOV, independent of the chromophore loading, has typical UV-Vis absorption spectra as the other LOV proteins (Figure 43). The dark state spectrum showed characteristic peaks, which originate from flavin chromophore bound to the protein: three peaks (422 nm, 447 nm and 475 nm) in the region 400 nm – 500 nm and broad unresolved peaks at around 365 nm. Upon blue-light illumination, an adduct is formed between the chromophore and the conserved cysteine residue leading to the formation of a broad peak with a maximum at around 390 nm



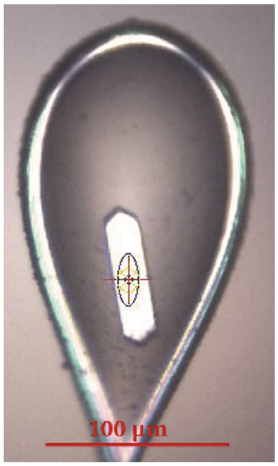


Figure 43. UV-Vis spectra of W619_1-LOV protein with different chromophore load in dark and light states, as well as spectrum of protein crystal (**left**). (**right**) Photo of a typical W619_1-LOV crystal, which is mounted in a loop and was used for structure determination and microspectrometry.

Samples of the W619_1-LOV protein with 10% and 50% chromophore loads were crystallized, resulting in colorless, rod-shaped crystals (Figure 43). Surprisingly, in the electron density maps of determined crystal structures, no electron density could be assigned to the possible flavin chromophores FMN or RBF. The absence of any chromophore in the crystal was also verified by single crystal microspectrometry measurements on W619_1-LOV crystals, which showed no characteristic absorption in the wavelength range between 300 nm and 700 nm (Figure 43). Even the protein samples having 50% chromophore loading resulted in crystallization of the apo form.

These results suggested that the apo form of W619_1-LOV is stable, in contrast to the previous report on the apo form of YtvA-LOV (Dorn, 2012; Dorn *et al.*, 2013). Next, the apo form of W619_1-LOV in solution was produced. This was done by removal of the remaining bound chromophores with guanidine hydrochloride, and subsequent renaturation yielded the stable apo form of W619_1-LOV. Stability of the apo form offered possibility to study biophysical and biochemical properties of W619_1-LOV with bound natural and modified flavin chromophores.

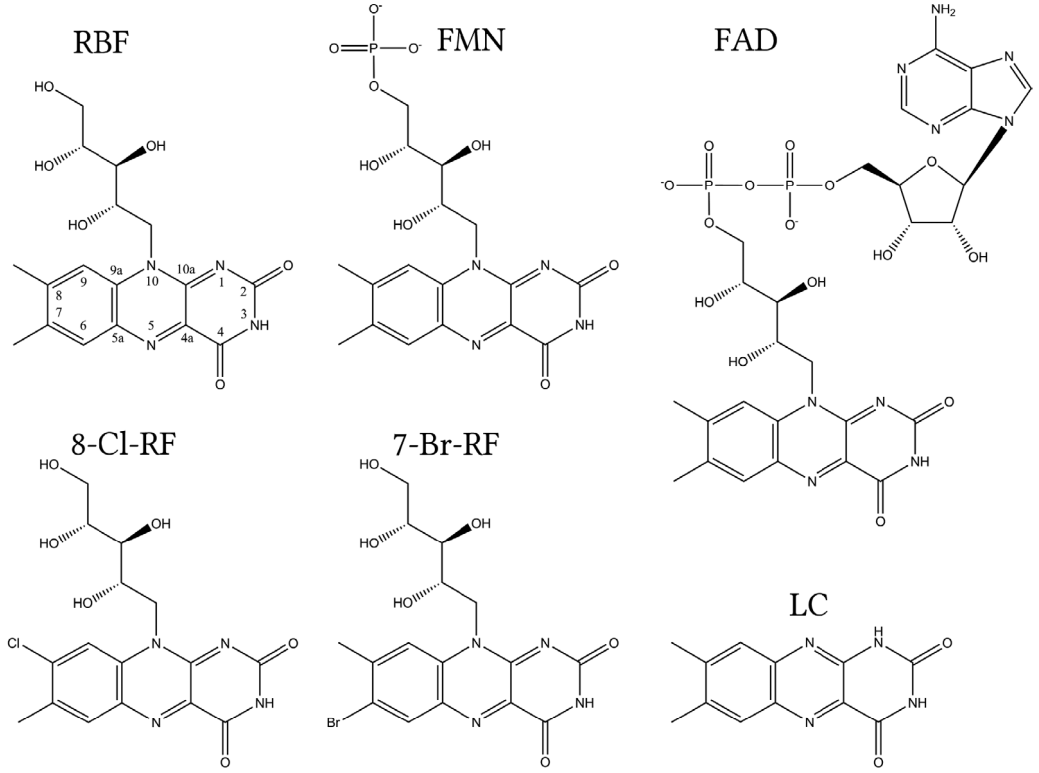


Figure 44. Natural and modified flavin chromophores. Where RBF – riboflavin, FMN – flavin mononucleotide, FAD – flavin adenine dinucleotide, 8-Cl-RF – 7-methyl-8-chloro-riboflavin, 7-Br-RF – 7-bromo-8-methyl-riboflavin, LC – lumichrome.

Table 30. The UV-Vis absorption maxima, λmax, for W619_1-LOV with different bound chromophores and that of other LOV proteins.

Protein		λmax, bound [nm]	λmax, free chromophore [nm]	Reference
W619_1-LOV	FMN	447	445	This work
	RBF	445	445	
	FAD	447	450	
	7-Br-RF	449	444	
	8-Cl-RF	456, 474	442	
	LC	337, 421	352, 387	
Pp2FbFP Q116V[1,2]		439	445	(Wingen *et al.*, 2014)
iLOV-Q489K[2]		440	445	(Davari *et al.*, 2016)

[1] Pp2FbFp is C53A mutant of PpSB2-LOV. [2] Mutation of the equivalent glutamine residue.

First, loading was done with naturally abundant pure flavins (RBF, FMN and FAD) previously detected in LOV proteins (Figure 44). After successful loading, the subsequent UV-Vis spectra showed similarities, however with slightly different peak maxima, λmax, as shown in Figure 45 and Table 30. Shifts in the peak maxima between of these chromophores were also previously observed in the spectra of flavin binding enzymes (Kozioł, 1971).

Second, the apo protein was loaded with natural flavins: lumichrome (LC) and roseoflavin (ROF) and modified flavins: 8-chloro-riboflavin (8-Cl-RF) and 7-bromo-riboflavin (7-Br-RF) (Figure 45). All chromophores except ROF could successfully bind to the W619_1-LOV in the apo form. UV-Vis spectra of free and protein-bound chromophores are depicted in Figure 45.

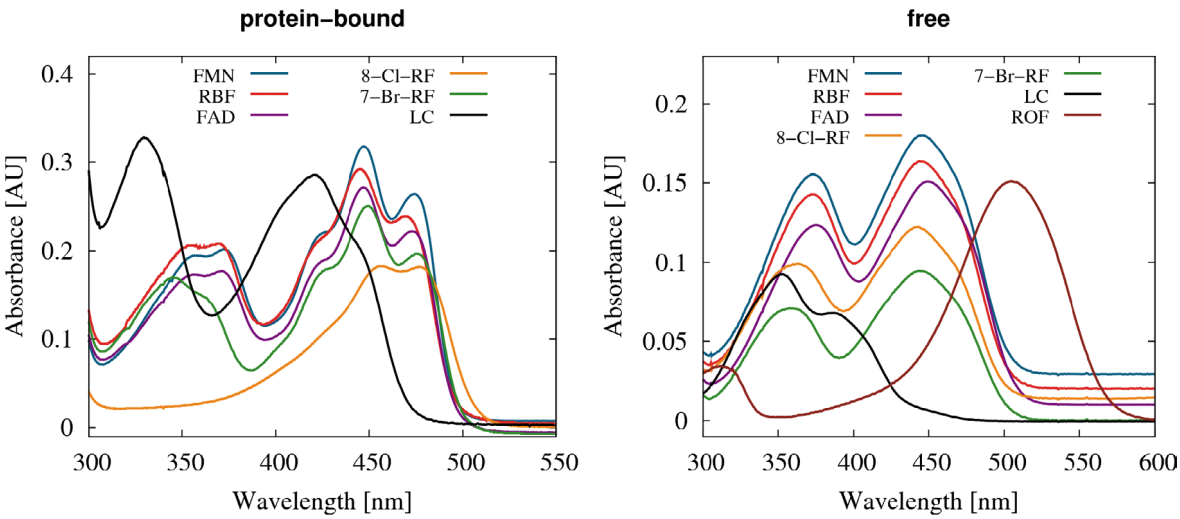


Figure 45. The UV-Vis absorption spectra of the W619_1-LOV bound to chromophores and those of the free chromophores.

ROF was particularly interesting chromophore due to its absorption maxima in the free state, which is shifted to the red region. Presumably, the bulky dimethylamino group, instead of the methyl group at position eight in the flavin ring system, prevented ROF from binding in the rather tight chromophore pocket. 7-Br-RF bound to protein showed the UV-Vis spectrum similar to that of W619_1-LOV with bound FMN, with absorption maximum shift of only 2 nm.

LC, on the contrary, lacks a ribityl side chain compared to RBF and only has a tricyclic ring system. Consequently, the UV-Vis spectrum of free LC was quite different, with the two peaks being close to each other and blue-shifted compared to the peaks in the spectrum of RBF. Upon binding to the W619_1-LOV in the apo form, its UV-Vis spectrum changed dramatically, leading to a single peak at 337 nm and one broad unresolved peak with a maximum at 421 nm. Thus the absorption peak maximum of bound LC became blue-shifted by 26 nm compared to that of the bound FMN. Until now, no more than a 10 nm shift was

achieved by using site-directed mutagenesis on LOV proteins with FMN as a chromophore (Wingen *et al.*, 2014; Davari *et al.*, 2016).

The UV-Vis absorption spectrum of free 8-Cl-RF was similar to that of W619_1-LOV with bound FMN, RBF, FAD and 7-Br-RF, whereas in the bound state it became distinct. The spectrum of bound 8-Cl-RF is unusual for LOV proteins and was seen only once before for the similar 8-bromo substituted RBF bound to the YtvA-LOV protein (Mansurova *et al.*, 2013). The absorption peaks at around 360 nm disappeared and only two partially unresolved peaks were present, one at 456 nm and one at 474 nm. An asymmetric form of the peak at ~456 nm probably indicates presence of an additional unresolved broad peak in this region. Careful analysis by calculation of the second derivative, indeed, revealed a small shoulder at around 423 nm. These three peaks, at ~423 nm, 456 nm and 474 nm, correspond to the peaks at 422 nm, 447 nm and 475 nm, of FMN bound to W619_1-LOV, where the absorption peak maximum is red-shifted by 9 nm, from 447 nm to 456 nm.

Surprisingly, it was found that upon denaturation in 6 M guanidine hydrochloride, the spectrum of 8-Cl-RF bound to W619_1-LOV did not return to the spectrum of the free chromophore (Figure 46). Denaturation in sodium dodecyl sulfate or heating to 90 °C for 1 hour also failed to release the chromophore. Additionally, this denatured sample was filtered through a centrifugal concentrator unit with 3 kDa cut-off and no free 8-Cl-RBF was detected spectroscopically, suggesting that 8-Cl-RF forms a stable covalent bond to W619_1-LOV.

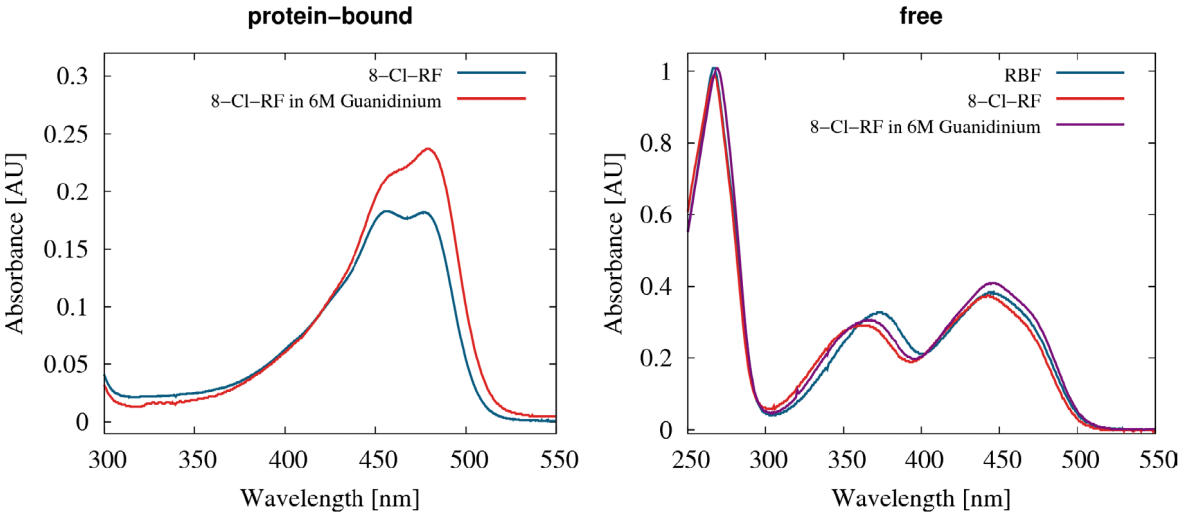


Figure 46. The UV-Vis spectra of the W619_1-LOV with bound 8-Cl-RF in phosphate buffer and denatured in guanidinium chloride. (**left**) The UV-Vis spectrum of the W619_1-LOV with bound 8-Cl-RF in 20 mM NaH_2PO_4/Na_2HPO_4 pH 8.0, 10 mM NaCl is shown as a blue curve. The UV-Vis spectrum of the W619_1-LOV with bound 8-Cl-RF denatured in 6 M guanidine hydrochloride, 20 mM NaH_2PO_4/Na_2HPO_4 pH 6.5 is shown as a red curve. (**right**) The spectra of RBF (blue) and 8-Cl-RF (red) in 20 mM NaH_2PO_4/Na_2HPO_4 pH 8.0, 10 mM NaCl and 8-Cl-RF in 6 M guanidine hydrochloride, 20 mM NaH_2PO_4/Na_2HPO_4 pH 6.5 (purple) are shown as solid lines.

Next, the incubation of W619_1-LOV protein in the apo form with 8-Cl-RF was monitored with UV-Vis spectroscopy, by measuring the spectra over time (Figure 47, left

panel). Right after the addition of 8-Cl-RF, the typical spectrum of LOV protein was observed, such as that of the W619_1-LOV with bound FMN. Over time it was transformed to the previously observed spectrum of the W619_1-LOV with bound 8-Cl-RF, probably as a result of covalent adduct formation.

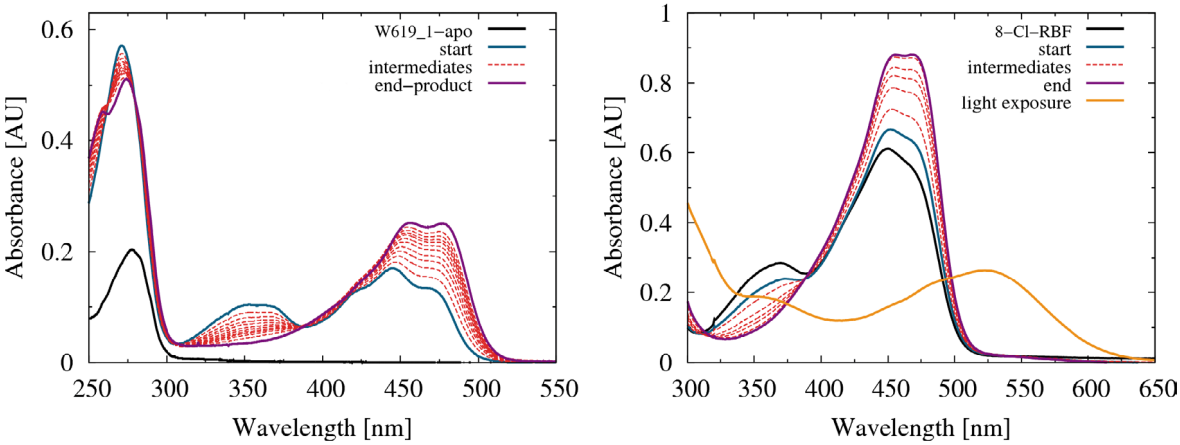


Figure 47. Binding of apo W619_1-LOV or L-Cysteine to 8-Cl-RF monitored by UV-Vis spectroscopy. (**left**) black curve – apo W619_1-LOV and blue – W619_1-LOV protein right after addition of 8-Cl-RF. Red dashed curves show intermediates and purple – the end product, which was filtered through 3 kDa cut-off membrane to remove unbound 8-Cl-RF. (**right**) black – 8-Cl-RF in 10 mM NaH_2PO_4/Na_2HPO_4 pH 8.0, 10 mM NaCl, blue – spectrum upon addition of L-Cysteine solution at pH 8.0, red dashed line – intermediates observed overtime, purple – end product and orange – end product after illumination with blue-light (450 nm)

In order to identify the covalent bonding partner of 8-Cl-RF, first free amino acids were incubated with 8-Cl-RF in solution. Among them, almost identical spectral changes were taking place over time after the incubation of L-Cysteine with 8-Cl-RF in a molar ratio of 400:1 in solution at pH 8.0 (Figure 47, right panel). The same spectral changes were observed after incubation with β-Mercaptoethanol, which also has a thiol group. However, compared to the W619_1-LOV binding, this process was much slower and required a higher molar excess of L-Cysteine. After illumination with blue-light the sample turned rose, which is reflected in the spectrum as the appearance of a broad peak at around 525 nm. It can be speculated that the illumination changed the oxidation state of the chromophore in the presence of ambient oxygen.

W619_1-LOV has two cysteine residues, the conserved Cys53 in the chromophore pocket and the solvent-exposed Cys79. In order to find which of these two residues is involved in the formation of a covalent adduct to 8-Cl-RF, additional experiments were done. First, to make sure that guanidine does not interfere with covalent binding, 8-Cl-RF and L-Cysteine were also incubated in 6 M guanidine hydrochloride, 20 mM NaH_2PO_4/Na_2HPO_4 at pH 6.5, which led to similar spectral changes. Second, the W619_1-LOV protein in the apo form was denatured and incubated with 8-Cl-RF in the 6 M guanidine hydrochloride, 20 mM NaH_2PO_4/Na_2HPO_4 at pH 6.5. This time, no spectral changes were observed, suggesting that only Cys53 in the chromophore pocket has high propensity to form the covalent adduct to 8-Cl-RF. Third, the purchased protein, Aprotinin, with surface exposed cysteines, under

reducing conditions with TCEP, also did not produce, either spectral changes or covalent binding. These facts led to the conclusion that the formation of the covalent adduct takes place exclusively at Cys53 in the pocket of the folded protein.

Literature data revealed that a similar chromophore, 8-chloro-FAD, was also forming a covalent bond to the flavoenzyme, dihydrolipoamide dehydrogenase (Moore, Cardemil and Massey, 1978; Moore, Ghisla and Massey, 1979). It was also found that covalent bonding takes place between the chlorine atom and a cysteine residue in the active center.

Dark recovery kinetics of W619_1-LOV bound to natural and modified flavin chromophores

The dark recovery kinetics of W619_1-LOV was measured at 20 °C and fitted by a single exponential function with the adduct lifetime (τ_{rec}) of 460 ± 10 min. Previously, the τ_{rec} of W619_1-LOV was measured at 37 °C and determined to be 53 ± 4 min (Rani *et al.*, 2013). For comparison, PpSB1-LOV is a close homologue of W619_1-LOV with an 88% sequence identity and has the τ_{rec} of 2471 ± 22 min at 20 °C (Jentzsch *et al.*, 2009).

The W619_1-LOV with bound RBF, FAD or 7-Br-RF showed a typical LOV photocycle property, such as the disappearance of absorption in the region 400 nm – 500 nm upon illumination with blue-light. Next, the dark recovery kinetics were measured for W619_1-LOV with these chromophores and fitted by a single exponential function (Table 31). The W619_1-LOV bound to these three chromophores, RBF, FAD or 7-Br-RF, showed a faster dark recovery compared to that of FMN, where RBF accelerated the photocycle by a factor of ~307. For YtvA-LOV it was previously reported that RBF accelerated the photocycle only by a factor of ~1.3 – 2, whereas FAD slowed it down by a factor of ~1.2 (Mansurova *et al.*, 2011; Dorn *et al.*, 2013).

Table 31. The adduct lifetimes of W619_1-LOV bound to different chromophores and of the homologues PpSB1-LOV and PpSB2-LOV proteins for a comparison.

Protein		Dark recover time τ_{rec} [min]	x-fold change of τ_{rec}	Reference
W619_1-LOV	FMN	460 ± 10	1 (reference)	This work
	RBF	1.5 ± 0.1	307	
	FAD	54.0 ± 0.5	8.5	
	7-Br-RBF	28.5 ± 0.5	16	
	8-Cl-RBF	–		
	LC	–		
PpSB1-LOV		2471 ± 22		(Jentzsch *et al.*, 2009)
PpSB2-LOV		2.3 ± 0.2		This work

Illumination with blue-light of the W619_1-LOV with bound LC or with bound 8-Cl-RBF did not change their absorption spectra. Apparently, the photocycle cannot take place for some reasons. In the case of LC, it can be speculated that the photocycle may take place on a time scale that is faster than that the regular UV-Vis absorption spectroscopy can

access. In the case of 8-Cl-RF, the conserved Cys53 is likely involved in the covalent bonding to the chromophore, hence no canonical photocycle can take place.

Fluorescence properties

The fluorescence spectra of W619_1-LOV with bound natural and modified flavin chromophores were measured by excitation at 450 nm except for LC, which was done at 420 nm and is depicted in Figure 48. Protein-bound FMN, RBF and FAD showed a fluorescence peak at around 500 nm with unresolved shoulder at around 520 nm, however in the case of RBF, the shoulder peak was almost as high as the main peak. In contrast, free chromophores has only one single peak at around 524 nm. The wavelengths of the fluorescence maximum (λmax) are summarized in Table 32.

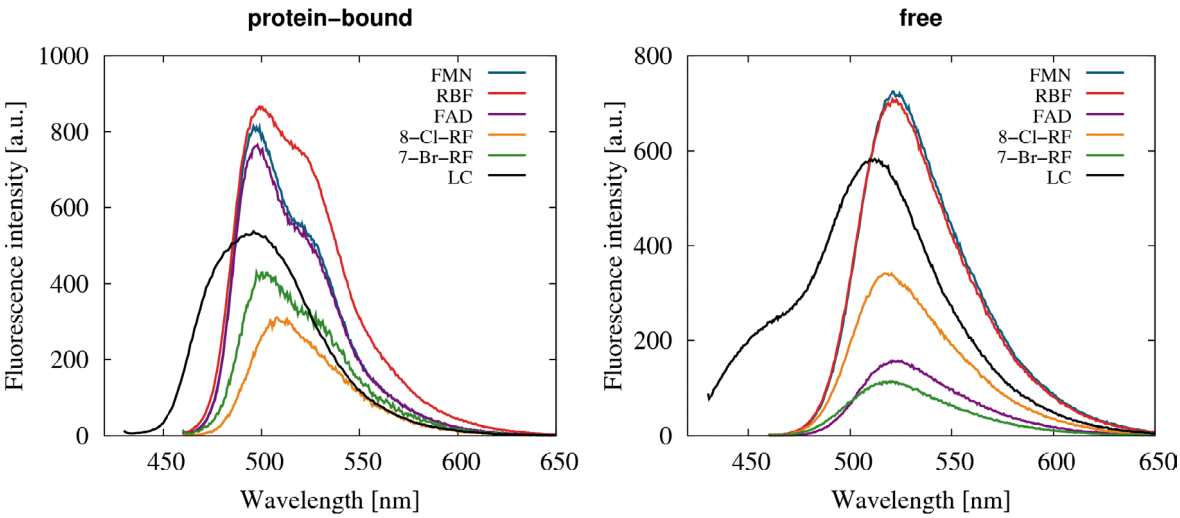


Figure 48. The fluorescence spectra of the W619_1-LOV with bound chromophores and those of the free chromophores. Fluorescence measurements were done by excitation at a wavelength of 450 nm for all samples, except for LC where it was 420 nm.

The fluorescence peaks of the W619_1-LOV bound to the 7-Br-RF and 8-Cl-RF chromophores have a similar shape as that of the protein-bound RBF, however, the peak maxima were shifted by 2 nm and 10 nm compared that of RBF, respectively. Fluorescence spectrum of LC, in contrast, has only single broad peak with the same peak maximum as FMN.

The efficiency of fluorescence emission, i.e. the quantum yield (Φ_F), was measured for all chromophores in the free and bound states and are presented in Table 32. The common LOV chromophores, FMN, RBF and FAD, showed a similar quantum yield in the protein-bound state. The fact that the fluorescence of free FAD was quenched in the solution, but not in the bound state is in agreement with the measurements of YtvA-LOV with bound FAD (Mansurova *et al.*, 2011). The binding of 8-Cl-RF to the protein increased the rather low quantum yield of 0.02 to 0.10. The quantum yield of the protein-bound LC of 0.40, however, is a very promising result. LOV proteins, when used as fluorescent reporters, demand the quantum yield to be as high as possible. The Φ_F of 0.40 is significantly high among LOV-

based fluorescent proteins with the highest reported values of 0.51 in the CreiLOV protein, which was achieved by the site-directed mutagenesis while using FMN as a chromophore (Mukherjee *et al.*, 2015; Davari *et al.*, 2016).

Table 32. The fluorescence properties of W619_1-LOV with different chromophores and those of other LOV proteins for comparison. Fluorescence spectra and quantum yield measurements were done by excitation at a wavelength of 450 nm for all samples, except for LC where it was 420 nm. The relative quantum yield, Φ_F, was calculated with RBF as a reference (Φ_F = 0.27), for additional details see page 30.

Protein		Fluorescence λmax (bound)	Fluorescence λmax (free)	Φ_F (bound)	Φ_F (free)	Reference
W619_1-LOV	FMN	496	524	0.28	0.27	This work
	RBF	500	524	0.29	0.27	
	FAD	499	523	0.25	0.03	
	7-Br-RBF	502	520	0.03	0.04	
	8-Cl-RBF	510	518	0.10	0.02	
	LC	496	511	0.40	n.d.**	
Pp2FbFp*		496	-	0.27	-	(Wingen *et al.*, 2014)
Pp1FbFp*		495	-	0.22	-	(Wingen *et al.*, 2014)
YtvA 8-Br-RF		-	520	0.07	0.07	(Mansurova *et al.*, 2013)

* measured for the so-called Pp2FbFp and Pp1FbFp proteins, which are C53A mutants of PpSB2-LOV and PpSB1-LOV, respectively. ** Not determined, as riboflavin is not a valid reference anymore.

Crystal Structure of W619_1-LOV

The crystal structure of W619_1-LOV was determined at a resolution of 2.5 Å in $I4_1$ space group. The data collection and refinement statistics are listed in Table 33. The asymmetric unit contains a homodimer, formed by monomers with a non-crystallographic 2-fold symmetry (Figure 49A).In the crystal structure, 134 of the 146 residues in Chain A (137 in Chain B) were observed, moreover, the N-terminal expression-tag of Chain A was also partially observed and its residues were numbered from 0 to -9. Superposition of monomers showed their similarity, with the root-mean-square deviation (RMSD) of the Cα atom positions of only 0.82 Å (Figure 49B).

The dimer interface of W619_1-LOV is similar to that observed in the dark state structure of PpSB1-LOV (PDB ID: 5J3W, Röllen *et al.*, 2016). The PISA analysis (Krissinel and Henrick, 2007) of the dimer present in the crystal structure has a calculated total buried surface area of 3272 Å^2, which is close to 3096 Å^2 of the PpSB1-LOV dimer. Such large total buried surface area is indicative of a stable dimer in solution.

In the $2mF_o$-DF_c electron-density map no electron density could be assigned to the possible flavin chromophores FMN or RBF, suggesting that the W619_1-LOV protein was crystallized in the apo form, which was confirmed by the UV-Vis microspectrometry.

Table 33. W619_1-LOV crystal structures: data collection and refinement statistics.

X-ray data	**W619_1-LOV apo (PDB ID: 5LUV)**
Beamline	ID29, ESRF
Detector	PILATUS3 6M
Wavelength [Å]	0.97625
Resolution range [Å]	47.60 – 2.50 (2.60 – 2.50)*
Space group	$I4_1$
Unit-cell a, b, c [Å]	109.54 109.54 95.21
α, β, γ [°]	90 90 90
Total reflections	110498
Unique reflections	19498 (2202)
Multiplicity	5.7 (5.9)
Completeness [%]	99.9 (100.0)
Mean I/σ(I)	12.3 (1.1)
Wilson B-factor [Å^2]	57.04
R-merge	0.101 (1.740)
R-meas	0.111 (1.908)
$CC_{1/2}$	0.998 (0.411)
Refinement	
Resolution range [Å]	40.56 – 2.50 (2.59 – 2.50)
R-work	0.2108 (0.2944)
R-free	0.2528 (0.3077)
coordinate error (max.-likelihood based) [Å]	0.32
Number of non-hydrogen atoms	2214
macromolecules	2179
ligands	17
water	18
Protein residues	281
RMSD (bonds) [Å]	0.009
RMSD (angles) [°]	1.02
Ramachandran favored [%]	100
Ramachandran outliers [%]	0
Clashscore	7.50
Average B-factor [Å^2]	76.86
macromolecules [Å^2]	76.80
ligands [Å^2]	100.26
solvent [Å^2]	61.75

* Statistics for the highest-resolution shell are shown in parentheses

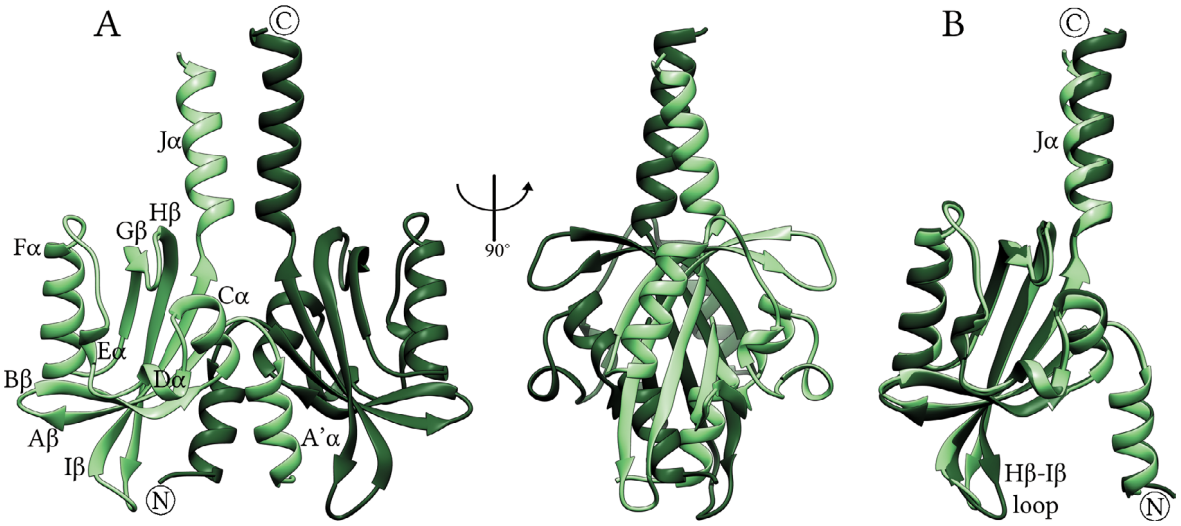


Figure 49. The crystal structure of W619_1-LOV in the apo form. The structure is represented as ribbons, where chains A and B are colored in light and dark shades of green. (**A**) The dimer of W619_1-LOV, which is observed in the crystal structure. (**B**) The superposition of the W619_1-LOV monomers by the core domain residues 17 – 117, which showed the differences in the Hβ-Iβ loop and the Jα helix. After superposition, the RMSD of the Cα atom positions is 0.77 Å over the residues 2 – 134.

Comparison of W619_1-LOV to PpSB1-LOV with bound FMN

W619_1-LOV from the W619_1 strain of *Pseudomonas putida* is a close homologue of PpSB1-LOV from the KT2440 strain of *P. putida* with a sequence identity of 88%. Alignment of W619_1-LOV and PpSB1-LOV sequences showed only small differences between them, as shown in Figure 50. In the core domain region, these are (W619_1-LOV to PpSB1-LOV): Gln23 to Lys23, Pro35 to Ala35, Arg39 to Tyr39, Ala65 to Gly65, Leu73 to Met73, Cys104 to Ser104, Ala106 to Phe106 and His108 to Gln108. The rest of the differences lie in the C-terminal part of the Jα helix beyond residue 132 and are probably unstructured, as these residues were not observed in the crystal structures of W619_1-LOV and PpSB1-LOV (Circolone *et al.*, 2012; Röllen *et al.*, 2016).

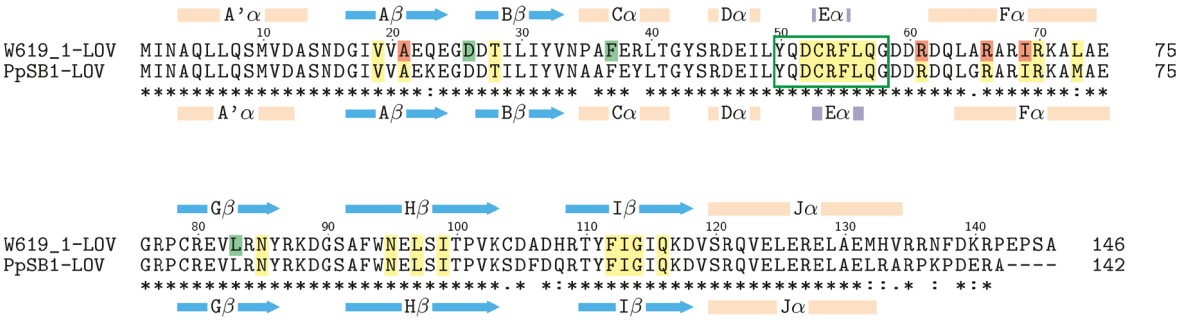


Figure 50. The sequence alignment of W619_1-LOV with PpSB1-LOV. The secondary structure elements for W619_1-LOV are shown on top, and for PpSB1-LOV (PDB ID: 5J3W, Röllen *et al.*, 2016) on bottom, where: arrow – β-strand and cylinder – α-helix (apricot) and 3_{10}-helix (plum). Residues lining the chromophore pocket at cut-off distances < 4 Å and ≥ 4 Å from FMN in the crystal structure of PpSB1-LOV in the dark state are highlighted in yellow and green, respectively. The asterisks at the bottom line of the alignment indicate identical residues in a given sequence position, while single and double dots refer to highly and moderately similar residues, respectively.

The superposition of the W619_1-LOV structure in the apo form with the PpSB1-LOV structure (PDB ID: 5J3W, Röllen *et al.*, 2016) in the dark state with bound FMN showed major differences around the chromophore binding pocket, as shown in Figure 51. The most dramatic changes were in the Eα helix (residues 53 – 55), the Eα-Fα loop (residues 56 – 65) and the Hβ-Iβ loop (residues 104 – 109), with the RMSD of the Cα atom positions of 2.65 Å (residues 53 – 65) and 0.60 Å (residues 104 – 109).

The Eα helix contains the conserved Cys53 residue which is involved in the formation of an adduct to the chromophore. This 3_{10}-helix Eα partially unfolds in the W619_1-LOV structure (residues 53 – 55), shortening by two residues in comparison to the PpSB1-LOV structure (residues 53 – 57). Moreover, the succeeding Eα-Fα loop adopts different conformation, probably due to the absence of any chromophore. The subsequent Fα helix was extended by two residues, thus starting at position 62 instead of 65 as in PpSB1-LOV, resulting in the shortening of the Eα-Fα loop. Detailed comparison of the residues lining the chromophore binding pockets in the crystal structures of the W619_1-LOV in the apo form to those of the PpSB1-LOV with bound FMN is shown in Figure 52.

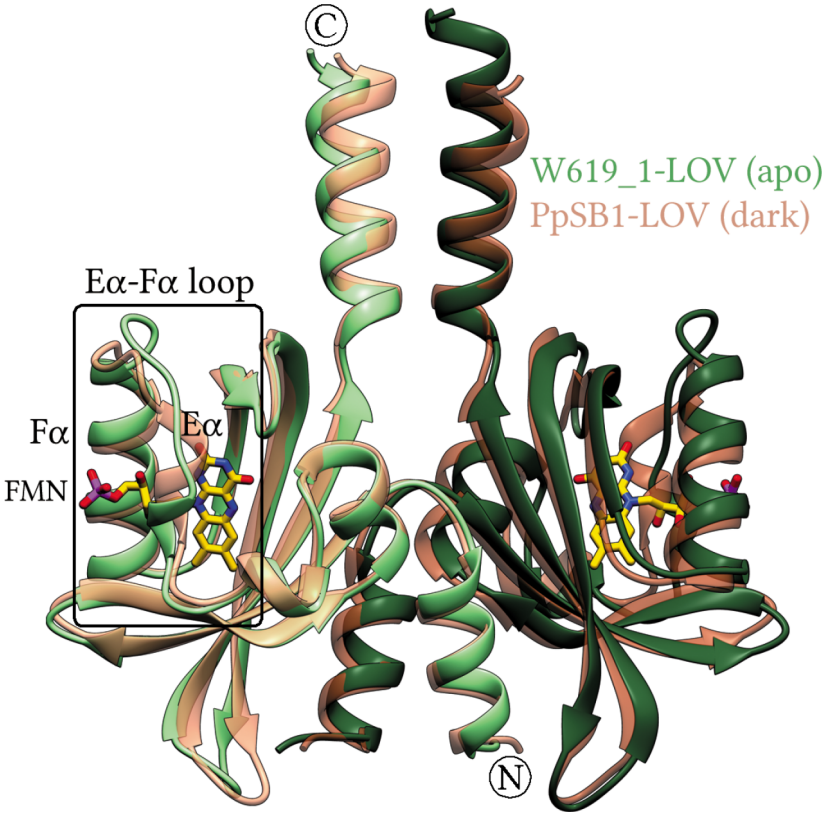


Figure 51. The superposition of the W619_1-LOV in the apo form with the PpSB1-LOV in the dark state with bound FMN. The structures of W619_1-LOV (green) and PpSB1-LOV (PDB ID: 5J3W, Röllen *et al.*, 2016) (coral) are represented as ribbons, where chains A and B are colored in light and dark shades. FMN in PpSB1-LOV is shown as a stick model. After superposition of W619_1-LOV dimer (residues 1 – 134) with the PpSB1-LOV dimer (residues 1 – 134), the RMSD of the Cα atom positions is 1.35 Å. The most dramatic changes were in the Eα helix, the Eα-Fα loop and Fα helix.

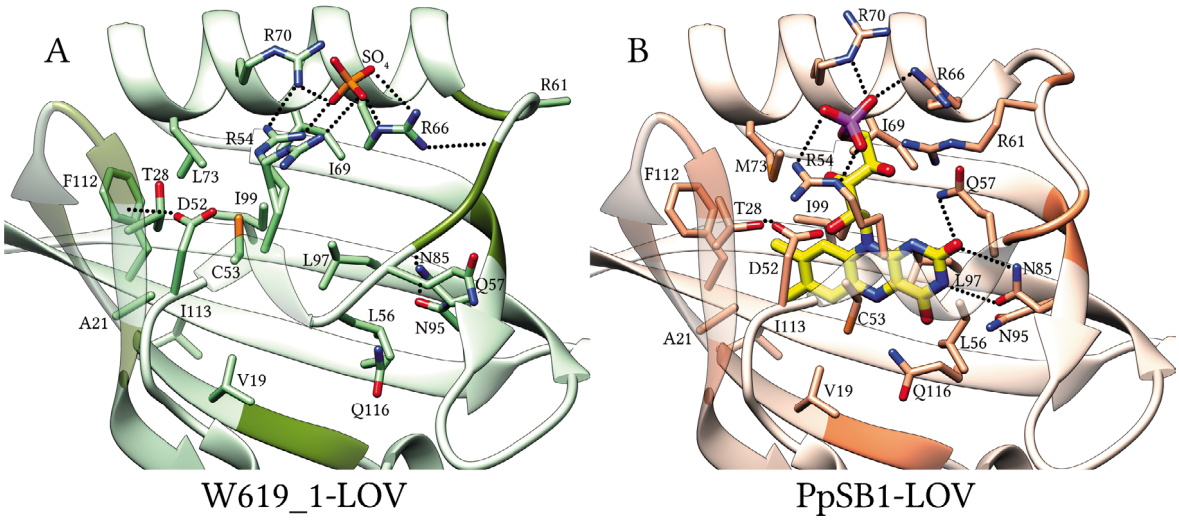


Figure 52. Comparison of chromophore binding pockets of W619_1-LOV in the apo form and PpSB1-LOV in the dark state. Residues lining the chromophore pocket within ≤ 4 Å cut-off distance are shown as stick models and > 4 Å cut-off distance are shown as dark colored ribbon sections. FMN, sulfate ion and amino acid residues are shown as stick models colored by elements: carbon, yellow; nitrogen, blue; oxygen, red; sulfur, orange; phosphorus, purple. Hydrogen bonds and salt bridges within ≤ 3.2 Å distance are shown as dotted lines.

In the apo form, the residues usually involved in hydrogen bonding with the chromophore now point away from it or form intramolecular interactions, as seen for the conserved residues Gln116, Asn95 and Asn85. Likewise, the residues Leu56, Gln57 and Arg61 of the Eα-Fα loop, which usually interacts with the ribityl chain and the phosphate group of FMN in the structure of FMN-bound LOV proteins, now point away from the pocket. The adduct-forming Cys53 also showed a different conformation, which would clash against FMN if FMN would be in the same position as in the PpSB1-LOV structure with bound FMN.

The positions of most hydrophobic residues remained similar (Val19, Ala21, Ile69, Leu97, Ile99, Phe112, Ile113 and Gly114) in the chromophore pocket. Several residues (Gln23, Asp26, Gln27 and Thr28) in the Aβ-Bβ loop have different rotamers of their side chains. Whereas the backbone N atom of Thr28 in W619_1-LOV forms a hydrogen bond to Asp52, which precedes the Eα helix. For comparison, Asp52 in PpSB1-LOV instead makes a hydrogen bond with the side chain OH group of Thr28.

The absence of FMN with its phosphate group led to the loss of interactions with the four arginine residues, Arg54, Arg61, Arg66 and Arg70, which coordinates the phosphate group of FMN in the PpSB1-LOV structure. The side chain of Arg61 is apparently flexible in W619_1-LOV structure as an electron density was observed only for the Cβ atom, additionally, it is pointing in the opposite direction as that of PpSB1-LOV. Two buffer molecules, a sulfate and a chloride ions, fitted well in the traces of electron density, which was detected in the vicinity in chain A (interacting with Arg54, Arg66, Arg70) and chain B (interacting with Arg70), respectively. These ions mimicked the interaction between the phosphate group of FMN with the arginine residues as seen in FMN-bound LOV structures.

Difference between the sequences of W619_1-LOV and PpSB1-LOV, Pro35 to Ala35, and Arg39 to Tyr39, introduced no local changes in the structure. The following sequence differences, Cys104 to Ser104, Ala106 to Phe106, His108 to Gln108, are part of the solvent exposed Hβ-Iβ loop, which adopted different conformations in comparison to those of PpSB1-LOV. However, these difference between the sequences may not be responsible for different conformation of the loop, as this loop adopts different conformations in the monomers of W619_1-LOV and also in those of PpSB1-LOV. The variation of the Hβ-Iβ loop might vary solely due to its participation in the crystal contacts.

In the chromophore pocket, there are only three residues that are different in W619_1-LOV and PpSB1-LOV: Ala65 to Gly65 and Leu73 to Met73 in the Fα helix and Gln23 to Lys23 in the Aβ strand, respectively. The Leu73 side chain in W619_1-LOV is well superimposable on its equivalent Met73 in PpSB1-LOV, where the latter is involved in hydrophobic interactions with the FMN chromophore. Located on the N-terminal part of the Fα helix, Ala65 forms an additional hydrogen bond with Asp62, causing elongation and stabilization of the Fα helix at its N-terminal end. An alanine residue was previously shown to stabilize the helical structure more than glycine residue (Serrano *et al.*, 1992). This could trigger the conformational changes seen in the preceding Eα helix and Eα-Fα loop in W619_1-LOV apo structure.

The Glu23 residue forms a hydrogen bond to the side chain of Thr28 and the Thr28 backbone N atom forms a hydrogen bond with Asp52, which precedes the adduct-forming Cys53. Whether the sequence differences between W619_1-LOV and PpSB1-LOV, Ala65 to Gly65, Leu73 to Met73 and Gln23 to Lys23, are related to the faster dark recovery of W619_1-LOV compared to PpSB1-LOV remain elusive. However, as mentioned above for SBW25-LOV, residues 23 and 74 are involved in the stabilization of the chromophore pocket from the side of Aβ-Bβ loop and the C-terminal end of the Fα helix.

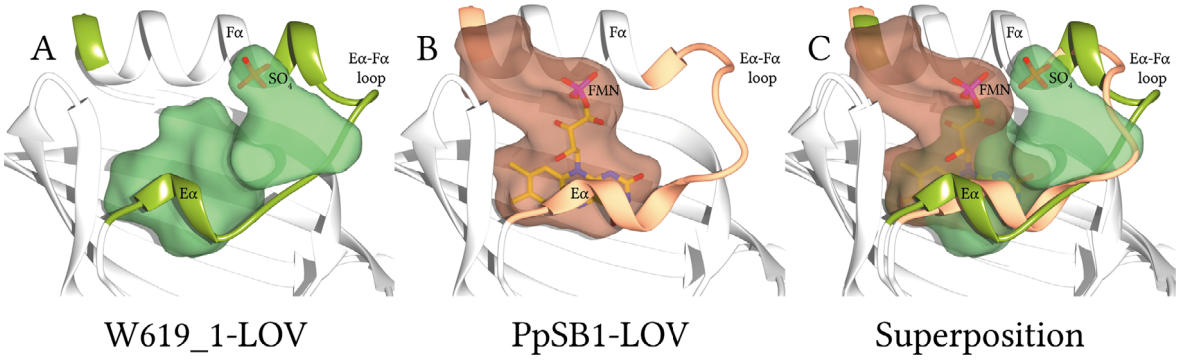


Figure 53. Comparison of the chromophore pocket solvent-accessible cavities of W619_1-LOV and PpSB1-LOV. The structures of W619_1-LOV and PpSB1-LOV (PDB ID: 5J3W, Röllen *et al.*, 2016) represented as ribbons. The solvent-accessible cavities of W619_1-LOV (**A**, green) and PpSB1-LOV (**B**, salmon) were created with a 1.4 Å probe radius and represented as transparent surfaces. The sulfate ion and FMN were not included in the calculations of the cavities but shown as sticks to indicate their positions in the respective structures. (**C**) The superposition of W619_1-LOV and PpSB1-LOV showing the alignment of their solvent-accessible cavities.

Next, the solvent-accessible cavities were determined for W619_1-LOV and PpSB1-LOV by rolling a sphere with a 1.4 Å radius inside binding pocket as described in Methods section. The three-dimensional cavity showed a van der Waals "shape" of an isoalloxazine ring system and most of the ribityl side chain of FMN (Figure 53). In PpSB1-LOV, the cavity extends towards the Aβ-Bβ loop and the C-terminal end of the Fα helix. In contrast, in W619_1-LOV it extends towards the Eα-Fα loop and the N-terminus of the Fα helix. These differences are likely related to the conformational changes caused by absence of the chromophore. First, the position of the phosphate group of FMN (as in PpSB1-LOV) is occupied by residues Arg54 and Arg70 in W619_1-LOV, which induced the closure of the pocket. Second, the aforementioned elongation of the Fα helix at its N-terminal end and the shift in the Eα-Fα loop creates more cavity space in the vicinity.

Molecular dynamics

Qualitative NMR data of YtvA-LOV in the apo from showed that the protein undergoes conformational exchange or was even partially disordered (Dorn, 2012; Dorn *et al.*, 2013). Molecular dynamics (MD) simulation of W619_1-LOV in the apo form, taking into account its high similarity to PpSB1-LOV, would allow drawing conclusion about the changes in dynamics caused by absence of the chromophore.

The MD of W619_1-LOV in the apo form was simulated for 50 ns. The respective crystal structure was used as initial structure, as described in the Material and Methods section. The structures remained stable over the course of the simulation, see snapshots and RMSD of the trajectory in Figure 54.

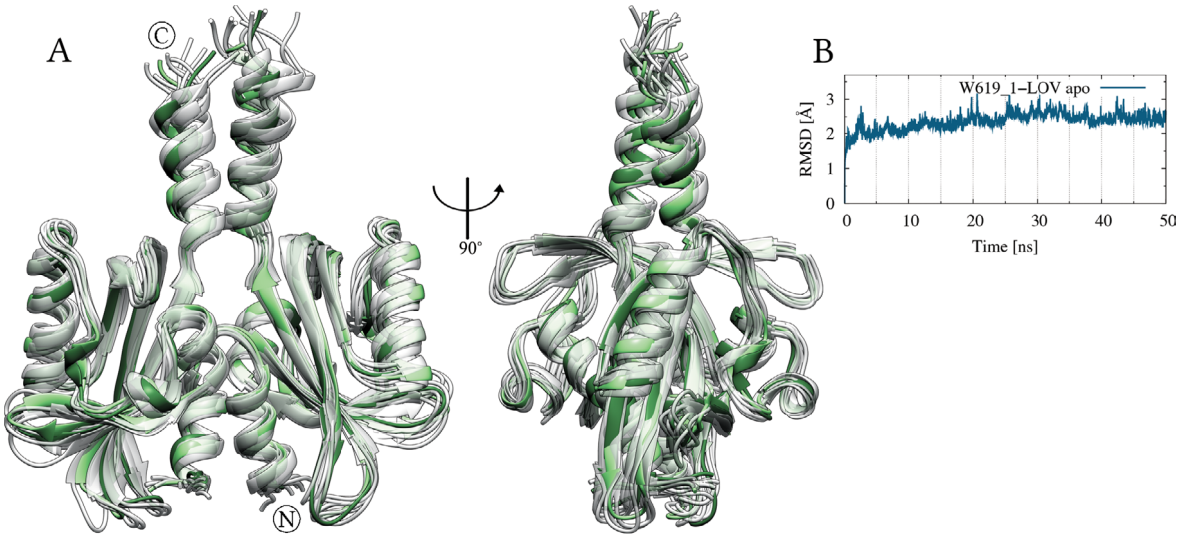


Figure 54. MD simulation snapshots of W619_1-LOV in apo form. (**A**) MD simulation snapshots, with Δt = 5 ns, are shown as ribbons, where the first snapshot is colored in light green, the last snapshots in dark green and the snapshots in between in gray. (**B**) The RMSD calculated for the whole trajectory with respect to the starting structure.

The root-mean-square fluctuation (RMSF) of the core domain residues 17 – 119 was calculated and presented in Figure 55. The RMSF of main chain atoms (N, C, O and Cα) showed increased dynamics of the loops: Aβ-Bβ, Gβ-Hβ and Hβ-Iβ, in the apo form. Similar differences in dynamics were also observed between the FMN-bound PpSB2-LOV and PpSB1-LOV in both the light and dark states, moreover, these loops vary also strongly among all LOV proteins. The regions of the Dα-Eα loop, the Eα helix and the Eα-Fα loop became significantly mobile, as emphasized in Figure 56. This region contains not only the adduct-forming Cys53 residue, but also residues that would otherwise interact with FMN, such as Asp52, Arg54, Gln57 and Arg61. The increased mobility indicates that chromophore binding significantly reduces dynamics. Moreover, the high dynamics of the Eα helix and surrounding residues, are necessary to bind chromophore and to enable the proper positioning of the adduct-forming Cys53 residue in this helix.

Comparison of the side chains RMSF showed an almost identical increase in dynamics of the Dα-Eα loop, the Eα helix and the Eα-Fα loop. Additionally, residues Arg66, Arg68 and Arg70 in the Fα helix have increased RMSF. Two residues Arg66 and Arg70, together with the above-mentioned Arg54 and Arg61, form the arginine residues cluster in PpSB1-LOV which coordinates the phosphate group of FMN (Circolone *et al.*, 2012).

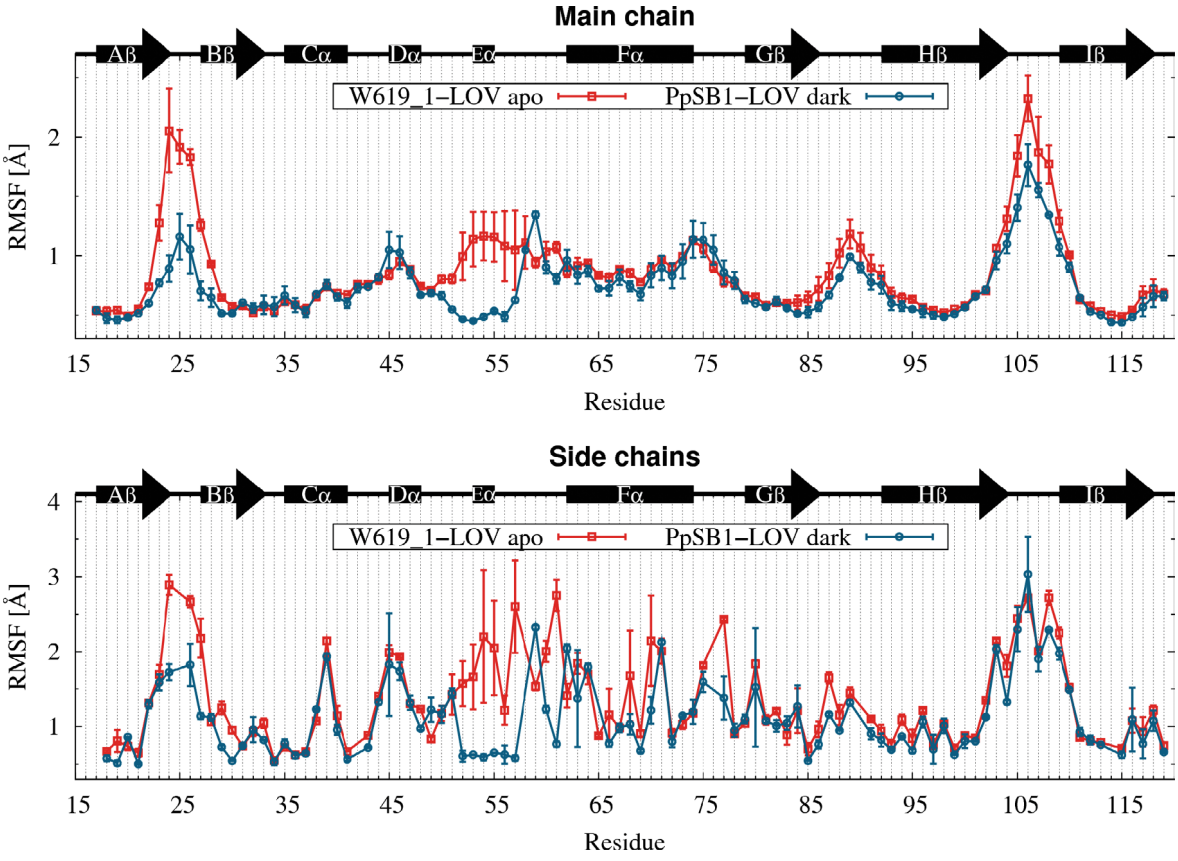


Figure 55. Residue-resolved RMSF of MD simulations of the W619_1-LOV in the apo and PpSB1-LOV in the dark state. The bars indicate the values of two monomers, whereas the points indicate their average. The secondary structure elements of W619_1-LOV are show as arrows for strands and rectangles for helices. Secondary structure assignment was done with the DSSP program for the W619_1-LOV crystal structure.

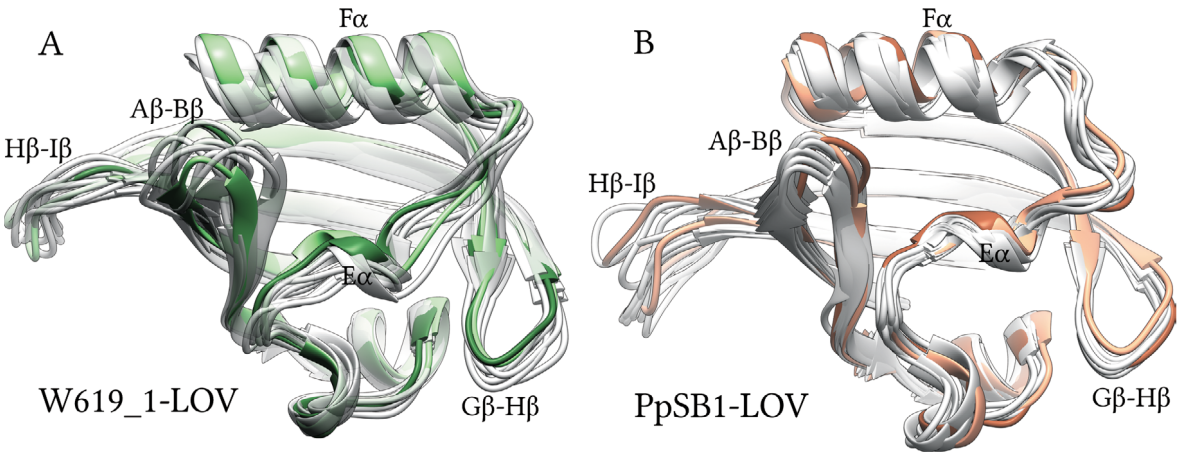


Figure 56. MD simulation snapshots of W619_1-LOV in the apo form and of PpSB1-LOV in the dark state with bound FMN. MD simulation snapshots of W619_1-LOV (**A**, green) and PpSB1-LOV (**B**, coral), with Δt = 5 ns, are shown as ribbons, where the first snapshot is colored in light shade, the last snapshots in dark shade and the snapshots in between in gray.

3.4 SB2F1

Biophysical and biochemical characterization

The SB2F1 is a chimeric sensory histidine kinase, which was engineered in similar fashion as YF1, see additional details on page 17 (Krauss, unpublished). For this, the oxygen-sensitive PAS domain in the histidine kinase FixL from *Bradyrhizobium japonicum* was replaced by the LOV domain from PpSB2-LOV protein (residues 1 – 120) from *Pseudomonas putida*. Beyond residue 120, SB2F1 share with YF1 an identical sequence for the DHp (dimerization and histidine phosphotransfer) and CA (the catalytic and ATP binding) domains (Figure 8).

YF1 was designed as the light-dependent histidine kinase by analogy to the natural sensory histidine kinases (Möglich, Ayers and Moffat, 2009a). Indeed, YF1 showed the dependency of the phosphorylation rates on the LOV domain photocycle (Möglich, Ayers and Moffat, 2009a; Diensthuber *et al.*, 2013). The similar dependency was shown by SB2F1, being more active in the dark state than in the light state (Krauss, unpublished).

UV-Vis spectroscopy, single crystal microspectrometry and dark recovery kinetics

SB2F1 protein showed photocycle properties such as the bleaching upon illumination with blue-light and the thermally-driven dark recovery. The dark state spectrum of SB2F1, similar to that of the parental PpSB2-LOV, showed three characteristic peaks around 450 nm (422 nm, 447 nm and 474 nm) and a double peak at around 360 nm (Figure 57). After the blue-light illumination, only a single broad peak at around 390 nm was observed, which was identical to that of PpSB2-LOV.

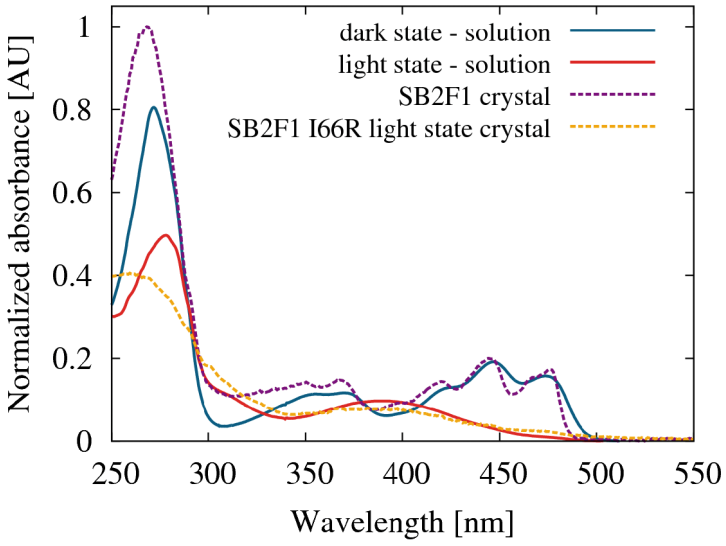


Figure 57. The UV-Vis spectra of SB2F1 in the dark and light states, measured both in solution and as a crystal together with the spectrum of a SB2F1 I66R crystal, which was grown under light conditions.

The single PpSB2-LOV domain has the adduct lifetime (τ_{rec}) of 5.2 ± 0.3 min at 20 °C. The presence of other protein domains in SB2F1, beside the LOV domain, might influence the dark recovery kinetics. Thus, the τ_{rec} was measured in the nucleotide-bound (ATP or ADP) and in the nucleotide-free states (Table 34).

The resulting τ_{rec} value of the SB2F1 in the nucleotide-free state was 7.1 ± 0.2 min at 20 °C, which is similar to the 5.2 ± 0.3 min of PpSB2-LOV, at identical conditions (Figure 58). The addition of ADP (1.5 mM) and $MgCl_2$ (2 mM) changed the τ_{rec} to 5.2 ± 0.4 min, which is identical to that of PpSB2-LOV. On the contrary, after incubation for 5 minutes with ATP (1.5 mM) and $MgCl_2$ (2 mM), the τ_{rec} was determined to be 17.3 ± 0.6 min, which is 2.4 times higher than τ_{rec} in the nucleotide-free state.

Thereafter, the same sample with ATP was further incubated for 4 h and 22 h, under dark conditions at 20 °C. The dark recovery data, so far, could be fitted well by just a single exponential function. In contrast, after incubation for 4 h and 22 h, the data could only be well-fitted by a biexponential function. After incubation for 4 h, the τ_{rec1} and τ_{rec2} constants were determined to be 6.3 ± 0.6 min and 16.8 ± 0.6 min, whereas after incubation for 22 h, they were determined to be 6.7 ± 0.5 min and 22.8 ± 0.6 min, with an almost equal relative contribution of each exponent to the fit, both after 4 h and 24 h of incubation time.

Table 34. Overview of the SB2F1dark recovery kinetics measurements. The abbreviation, nf, stands for the nucleotide-free state. See details about incubation with ATP and ADP in the Figure 58.

Protein	Adduct lifetime τ_{rec} [min]	Protein	Adduct lifetime τ_{rec} [min]
SB2F1 nf	7.1 ± 0.2	SB2F1 ADP	5.2 ± 0.4
PpSB2-LOV	5.2 ± 0.3	SB2F1 ATP	17.3 ± 0.6
SB2F1 I66R nf	16.8 ± 0.5	SB2F1 ATP 4h	6.3 ± 0.6; 16.8 ± 0.6
PpSB2-LOV I66R*	16.3 ± 6.1	SB2F1 ATP 22h	6.7 ± 0.5; 22.8 ± 0.6

* (Jentzsch *et al.*, 2009)

The change in the adduct lifetime after incubation with ATP can be related either to the activation of the CA domain upon ATP binding or to the influence of the resulting phosphorylated state, which affects the dark recovery kinetics. It can be expected that, after a short incubation with ATP, the activation of CA domain, which performs catalytic reaction, would be the primary factor that influences the adduct lifetime. Indeed, after incubation of SB2F1 with ATP for 5 min, the dark recovery kinetics data was well-fitted just by a single exponential function.

After the incubation for 4 h and 22 h, a biexponential function fitted better, where the first adduct lifetime (τ_{rec1}) was close to that of the SB2F1 in the nucleotide-free or ADP-bound states, whereas the second adduct lifetime (τ_{rec2}) was close to τ_{rec} after incubation for 5 min with ATP. This similarity of the τ_{rec2} suggests that they are related to the protein molecules in the same state. However, it remains elusive whether the τ_{rec2} is related to the binding of ATP or to the resulting phosphorylated state. Furthermore, the relative contribution of each

exponential function with τ_{rec1} and τ_{rec2} to the fit stayed almost equal. This prohibits drawing the model with the simple scheme, where the amount of one component decays while the another one increases.

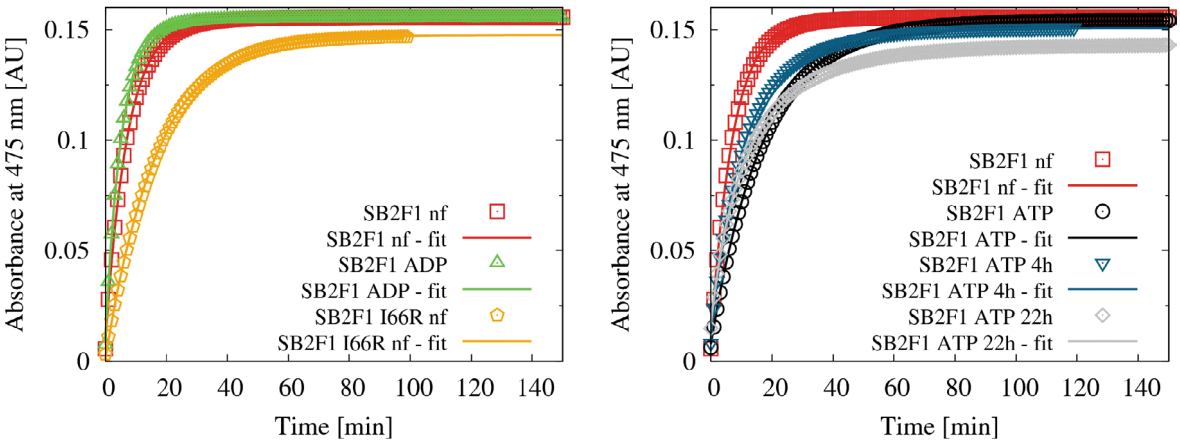


Figure 58. Dark recovery kinetics of SB2F1 and SB2F1 I66R, measured as an increase in absorbance at 475 nm over-time at 20 °C in 10 mM Na_2HPO_4/NaH_2PO_4 pH 8.0, 10 mM NaCl. (**left**) The dark recovery kinetics of the SB2F1 in the nucleotide-free state ("nf") and in the presence of 1.5 mM ADP, 2 mM $MgCl_2$ as well as that of the SB2F1 I66R in the nucleotide-free state. (**right**) The dark recovery kinetics of the SB2F1 in the nucleotide-free state and in the presence of 1.5 mM ATP, 2 mM $MgCl_2$ with the incubation times: 5 min – black, 4 h – blue and 22 h – gray.

The I66R mutant of SB2F1 was designed to enhance the crystallization of the protein in the light state, by increasing the adduct lifetime. Indeed, the adduct lifetime in the nucleotide-free state was increased to 16.8 ± 0.5 min at 20 °C, which is close to the adduct lifetime of 16.3 ± 6.1 min measured for I66R mutant of PpSB2-LOV (Figure 58) (Jentzsch *et al.*, 2009).

The protein crystals of SB2F1 used for the structure determination were grown under dark conditions in the presence of either ATP, ADP or non-hydrolysable/slowly-hydrolysable analogs of ATP (AMP-PNP, ATP-γ-S) supplemented with $MgCl_2$. Additionally, prior to X-ray exposure, some crystals grown in the presence of ATP and under dark conditions were illuminated with blue-light to induce the light state. Next, the UV-Vis microspectrometry of single crystals was performed in order to verify the photocycle state of the LOV domain in SB2F1. The spectra of dark grown and illuminated crystals were similar to that of the dark and light states in solution, respectively (Figure 57).

The UV-Vis spectra of the dark grown crystals of the SB2F1 I66R mutant showed the spectrum similar to that of the SB2F1 in the dark state in solution. The crystals of SB2F1 I66R, which were grown under light conditions, consistently showed the spectrum similar to that of SB2F1 in the light state in solution (Figure 57).

Chromophore content

All of the SB2F1 protein samples were additionally supplemented *in vitro* with FMN to increase the chromophore load. On average, the chromophore load of the protein samples

was greater than 85%. The analysis of the chromophore composition by HPLC yielded the following chromophore distribution: 88.8% of FMN, 11.2% of RBF and no FAD.

Determination of various crystal structures of SB2F1

The diffraction data of SB2F1, which was crystallized under dark conditions and in the presence of either ATP, ADP, AMP-PNP or ATP-γ-S nucleotides and supplemented with $MgCl_2$ was collected. The crystallization of SB2F1 under continuous blue-light illumination yielded crystals, which diffracted only to a very low resolution (lower than 10 Å). It was assumed that the fast photocycle disturbs the proper crystal packing because of the constant transition of protein molecules from the light state to the dark state and vice versa. First, the illumination of the crystals grown under dark conditions and in the presence of ATP with blue-light LED (λmax = 450 nm) was attempted in order to assess the light-induced structural changes. Such data will be referred to as the illuminated structure in order to distinguish it from the structures obtained from the crystals grown under continuous blue-light illumination.

In fact, the structures determined from the illuminated crystals may not reproduce the true light state structure due to the steric hindrance imposed by the crystal lattice. To clarify this fact, the SB2F1 I66R mutant was latter produced by analogy with respect to the I66R mutation of PpSB2-LOV, which was shown to increase the adduct lifetime. The increased adduct lifetime should in turn facilitate better crystallization under continuous blue-light illumination. Indeed, the SB2F1 I66R was successfully crystallized both under dark and light conditions, although resulting in a lower resolution than that of SB2F1. The short overview of the collected diffraction data is show in the Table 37 and the complete data collection and refinement statistics are shown in Table 35 and Table 36.

Initially, the molecular replacement phasing procedure was attempted to solve the crystal structure of SB2F1 by using the YF1 protein structure (PDB ID: 4GCZ, Diensthuber *et al.*, 2013) as a search template. However, this procedure failed, despite their similarity in the sequence and expected similar state (crystallization under dark conditions in presence of ATP/$MgCl_2$). In order to obtain the initial phase information experimentally, the selenomethionine-labeled protein was produced. Next, the SAD or MAD diffraction data were collected for the crystals with the selenomethionine-labeled protein. The determination of the initial phase information from this data enabled the structure to be solved, see additional details in the Materials and Methods section.

Moreover, the crystals of the selenomethionine-labeled protein diffracted to a higher resolution. A comparison of native and selenomethionine-labeled protein structures showed that the incorporation of the selenomethionine in place of methionine did not induce any structural changes in the protein. Therefore, all presented crystal structures of SB2F1, were determined from the crystals obtained from the selenomethionine-labeled protein.

Table 35. SB2F1 crystal structures: data collection and refinement statistics.

X-ray data	SB2F1 asymmetric dimer with ATP	SB2F1 symmetric dimer with ADP**	SB2F1 symmetric dimer with ADP
Beamline Detector	ID29, ESRF PILATUS 6M	ID29, ESRF PILATUS 6M	ID23-1, ESRF PILATUS 6M
Wavelength [Å]	0.97916	0.97916	0.97770
Resolution range [Å]	44.610 – 2.484 (2.573 – 2.484)*	45.658 – 2.505 (2.594 – 2.505)*	45.815 – 2.434 (2.521 – 2.434)*
Space group	$P3_121$	$P3_121$***	$P3_221$***
Unit-cell a, b, c [Å] α, β, γ [°]	138.8 138.8 96.1 90 90 120	138.7 138.7 98.7 90 90 120	138.7 138.7 49.6 90 90 120
Total reflections	761255 (78875)	773223 (78958)	160429 (4099)
Unique reflections	37952 (717)	37971 (643)	15487 (404)
Multiplicity	20.1 (21.1)	20.4 (21.1)	10.4 (10.1)
Completeness [%]	73 (100)	68 (100)	74 (20)
Mean I/σ(I)	12.50 (0.69)	8.00 (0.54)	21.43 (1.90)
Wilson B-factor [Å^2]	69.52	64.43	73.21
R-merge	0.1363 (4.867)	0.1967 (6.313)	0.0674 (1.484)
R-meas	0.1399 (4.987)	0.2018 (6.468)	0.0710 (1.564)
$CC_{1/2}$	0.999 (0.319)	0.999 (0.378)	0.999 (0.509)
Refinement			
Resolution range [Å]	45.428 – 2.484 (2.529 – 2.484)	45.658 – 2.505 (2.554 – 2.505)	45.815 – 2.434 (2.585 – 2.434)
R-work	0.2107 (0.5060)	0.2543 (0.4751)	0.2328 (0.4315)
R-free	0.2713 (0.5203)	0.2841 (0.5264)	0.2662 (0.3863)
coordinate error (max.-likelihood based) [Å]	0.39	0.28	0.37
Number of non-hydrogen atoms	5620	5783	2859
macromolecules	5494	5660	2801
ligands	126	116	58
water	–	–	–
Protein residues	728	730	364
RMSD (bonds) [Å]	0.009	0.010	0.011
RMSD (angles) [°]	1.26	1.34	1.40
Ramachandran favored [%]	98	98	98
Clashscore	17.99	17.91	17.31
Ramachandran outliers [%]	0	0	0
Average B-factor [Å^2]	123.90	129.73	150.89
macromolecules [Å^2]	124.16	130.22	151.24
ligands [Å^2]	112.39	109.35	133.85
solvent [Å^2]	–	–	–

* Statistics for the highest-resolution shell are shown in parentheses. ** Crystallization in the presence of ATP.
*** Datasets could be evaluated in both space groups $P3_121$ and $P3_221$, with long and short c-axis unit-cells.

Table 36. SB2F1 and SB2F1 I66R crystal structures: data collection and refinement statistics.

X-ray data	SB2F1 illuminated with ATP	SB2F1 I66R dark asymmetric dimer with ATP	SB2F1 I66R light asymmetric dimer with ATP
Beamline Detector	ID30B, ESRF PILATUS3 6M	ID29, ESRF PILATUS 6M	ID29, ESRF PILATUS 6M
Wavelength [Å]	0.97901	0.97625	0.97625
Resolution range [Å]	44.13 – 2.505 (2.595 – 2.505)*	41.09 – 3.004 (3.111 – 3.004)*	42.150 – 3.450 (3.573 – 3.450)*
Space group	$P3_121$	$P3_121$	$P3_121$
Unit-cell a, b, c [Å] α, β, γ [°]	138.2 138.2 94.9 90 90 120	138.6 138.6 96.9 90 90 120	139.4 139.4 89.9 90 90 120
Total reflections	607998 (26066)	121065 (11235)	267724 (27634)
Unique reflections	35459 (595)	21318 (2090)	13557 (1330)
Multiplicity	17.1 (8.5)	5.7 (5.4)	19.7 (20.8)
Completeness [%]	76 (86)	98 (98)	100 (100)
Mean I/σ(I)	20.85 (0.55)	12.06 (1.28)	9.12 (0.88)
Wilson B-factor [Å^2]	76.35	85.68	112.34
R-merge	0.1027 (3.357)	0.0798 (1.244)	0.2405 (3.986)
R-meas	0.1056 (3.575)	0.0878 (1.386)	0.2470 (4.086)
$CC_{1/2}$	1 (0.279)	0.999 (0.46)	0.998 (0.301)
Refinement			
Resolution range [Å]	45.239 – 2.505 (2.550 – 2.505)	45.371 – 3.004 (3.140 – 3.004)	45.622 – 3.450 (3.589 – 3.450)
R-work	0.2188 (0.3595)	0.2472 (0.4580)	0.3192 (0.4194)
R-free	0.2741 (0.3626)	0.3106 (0.4957)	0.4141 (0.4404)
coordinate error (max.-likelihood based) [Å]	0.38	0.64	0.84
Number of non-hydrogen atoms	5786	5578	5786
macromolecules	5660	5462	5660
ligands	126	126	126
water	–	–	–
Protein residues	728	728	728
RMSD (bonds) [Å]	0.009	0.010	0.015
RMSD (angles) [°]	1.30	1.42	1.90
Ramachandran favored [%]	98	94	95
Ramachandran outliers [%]	0	0	0
Clashscore	18.33	25.05	36.22
Average B-factor [Å^2]	114.10	154.35	164.35
macromolecules [Å^2]	114.37	154.78	165.08
ligands [Å^2]	101.86	134.28	131.29
solvent [Å^2]	–	–	–

* Statistics for the highest-resolution shell are shown in parentheses

Most of the crystals were not equally well diffracting along the crystallographic axes, where along one axis, c*, the resolution was lower (Figure 59A). For instance, in the dataset with asymmetric-dimer the significant diffraction spots were only observed up to the resolution along c* of 3.7 Å, while along a* and b* up to 2.5 Å. Moreover, the diffraction spots often had a streaky form, which led to the errors in the estimation of total intensity of the diffraction spot (Figure 59A). Both these facts led to the significant variation of the electron-density maps quality and to the high B-factors. This was largely evident in the extremely sparse electron-density maps of some parts of the CA domain, particularly the β-strands. On the contrary, the structural elements that were close to the DHp domain have well-defined electron-density maps. This can be attributed to the packing of the molecules within the crystal lattice. The crystal packing was such that the CA domains were exposed to the solvent channels of the crystal, what allowed them to be more flexible (Figure 59B).

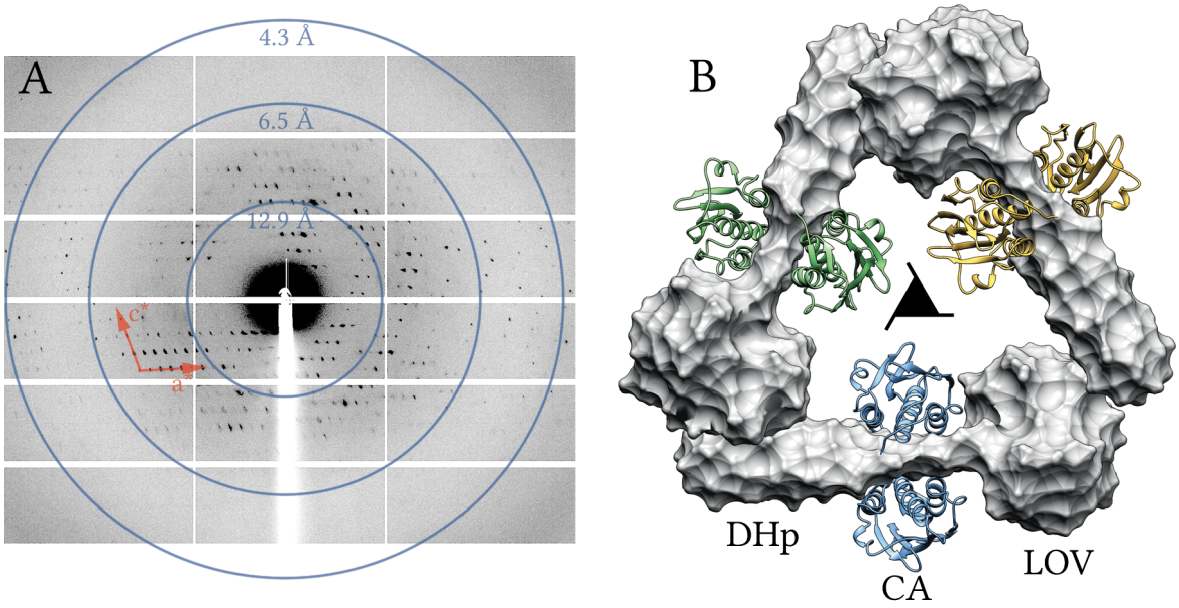


Figure 59. A typical diffraction image and representation of the crystal packing of the SB2F1crystal structures. (**A**) The diffraction image resulted from data collection by rotation with a $\Delta\varphi$ of 1° at ID30B beamline with PILATUS3 6M detector. (**B**) The crystal packing of SB2F1 molecules, when viewed through the 3_1 screw axis (▲). The LOV and DHp domains are shown as gray colored surfaces. The CA domains of symmetry-equivalent dimer molecules are shown as a ribbons, which are colored in blue, green and gold.

To guide and facilitate the refinement of the CA domains of the SB2F1 structures, the structure of the CA domain from the YF1 protein (PDB ID: 4GCZ, Diensthuber *et al.*, 2013) was used, which has an identical sequence. Additionally, the structure of the CA domain from YF1 was used as a geometrical restraint during structure refinement. The $2mF_o$-DF_c electron-density map showed the presence of the bound nucleotide, however, the low quality of the map of the CA domain did not enable it to be placed properly without prior knowledge. Instead, the structure of the ATP/Mg ion complex was used from the structurally homologous CA domain (PDB ID: 3SL2, Celikel *et al.*, 2012) to place it into the electron-density map. In

a similar way, the ADP structure from the YF1 structure (PDB ID: 4GCZ, Diensthuber *et al.*, 2013) was used for the refinement of the SB2F1 crystal structure with bound ADP.

Evaluation of the SB2F1 datasets showed that the crystals belonged to two related space groups: the $P3_121$ space group, with the approximate unit-cell dimensions a = b ≈ 138 Å, c ≈ 97 Å, and the $P3_221$ space group, with the approximate unit-cell dimensions of a = b ≈ 138 Å, c ≈ 48 Å. These combinations of the space group and the unit-cell will be denoted as the long and short c-axis, respectively (Table 37).

Table 37. A short overview of the SB2F1 and SB2F1 I66R crystal structures. All of the datasets from SB2F1 were obtained from crystals which were grown under dark conditions, except that the illuminated dataset was obtained from dark grown crystals exposed to blue-light.

Protein	Crystallization condition	Resolution [Å]	Short c-axis c ≈ 97 Å	Long c-axis c ≈ 48 Å
SB2F1	ATP	2.5	yes	yes
	ATP illuminated	2.5	no	yes
	ADP	2.4	yes	no
	AMP-PNP	2.5	yes	no
	ATP-γ-S	2.5	yes	no
SB2F1 I66R	ATP dark	3.0	yes	yes
	ATP light	3.5	no	yes

Among the collected diffraction data from the SB2F1 and SB2F1 I66R protein crystals, which were both crystallized in the presence of ATP and under dark conditions, there were some datasets which has the short c-axis as well as some which has the long c-axis. The crystals with different cells originated from the same crystallization drop, thus excluding the influence of differences in the crystallization conditions. Moreover, some datasets were first identified to have the short c-axis but later could be also evaluated with the long c-axis, but not vice versa. The reason for such discrepancy in the unit-cells was found after all the datasets were refined and will be presented next.

Crystals structures of SB2F1 with bound ATP or ADP

Structural domains of SB2F1

Next presented SB2F1 structures can be generally divided into three domains: LOV, DHp and CA, see sequence alignment in Figure 8. The LOV domain (residues 1 – 120), which originates from PpSB2LOV (residues 1 – 120) protein, has the non-canonical N-terminal helix A'α (residues 1 – 17) and the canonical LOV core domain (residues 17 – 117). All of the LOV domains of the SB2F1 structures were found to have a bound FMN, which agrees well with the chromophore composition determined by HPLC.

The DHp domain is comprised of two long helices, α1 (residues 121 – 175) and α2 (residues 182 – 205) that are connected by a loop (residues 176 – 181). Upon dimerization, the DHp domains are known to form an anti-parallel four-helix bundle (Bhate *et al.*, 2015),

which was also observed in the SB2F1 structures. Beside the facilitation of dimerization, the DHp domain carries the phospho-accepting histidine residue, from which the phosphoryl group is further transferred to the aspartate residues of the receiver domain. In the SB2F1 structures, the DHp domains seems to be arranged in the cis-autophosphorylation position where the CA domain would phosphorylate the His154 residue of the same monomer, such as in YF1, VicK, HK853 and other histidine kinases (Marina, Waldburger and Hendrickson, 2005; Wang *et al.*, 2013; Bhate *et al.*, 2015).

The CA domain has a highly conserved α/β sandwich fold with the three α-helices packed against the five anti-parallel β-strands (Bilwes *et al.*, 1999). The CA domain not only binds nucleotides such as ATP or ADP, but also catalyzes the autophosphorylation of the phospho-accepting histidine residue. In the SB2F1 structures, the CA domains (residues 218 – 365) were situated laterally to the DHp domains and connected by flexible loops to them. This connection may allow the CA domains to adopt the position close to the phospho-accepting histidine residue in either the cis- or trans-phosphorylation configuration. The flexibility of these loops made the electron-density map difficult to interpret unambiguously, therefore only the backbone atoms of residues were placed.

The asymmetric-dimer structure of SB2F1 with bound ATP

The crystal structure of the SB2F1 with bound ATP, which has the long c-axis, comprises two protein molecules in the asymmetric unit. Two protein molecules form long and rod-shaped homodimer (Figure 60). In both monomers, residues 1 – 365 out of 368 were observed in the $2mF_o$-DF_c electron-density map. However, as mentioned above, the CA domain structure partially has the $2mF_o$-DF_c electron-density map with poor resolution, thus the structure refinement was restrained to the CA domain from the crystal structure of YF1, which sequence is identical to that of SB2F1.

The superposition of one monomer with another one results in a mismatch as a consequence of the asymmetric dimer organization. For this reason, this dimer structure of SB2F1 will be referred to as the asymmetric-dimer. The superposition of monomers by residues 17 – 138 (The LOV core domain plus the DHp domain up to His138) showed the differences between their tertiary structures (Figure 60B). Essentially, the difference between the monomers is a result of DHp helix bending, which starts approximately beyond the His138 residue.

The symmetric-dimer structures of SB2F1 with bound ADP

The crystal structure of SB2F1, which has the short c-axis, was obtained under the same crystallization condition with ATP as it was in the above-mentioned asymmetric-dimer. Residues 1 – 365 out of 368 were observed in the $2mF_o$-DF_c electron-density map. This structure, as a consequence of the smaller volume of the short c-axis unit-cell, contains only one protein molecule in the asymmetric unit. However, the dimer can be formed by

application of the symmetry operation using the crystallographic 2-fold axis, which will be referred to as the symmetric-dimer (Figure 61A).

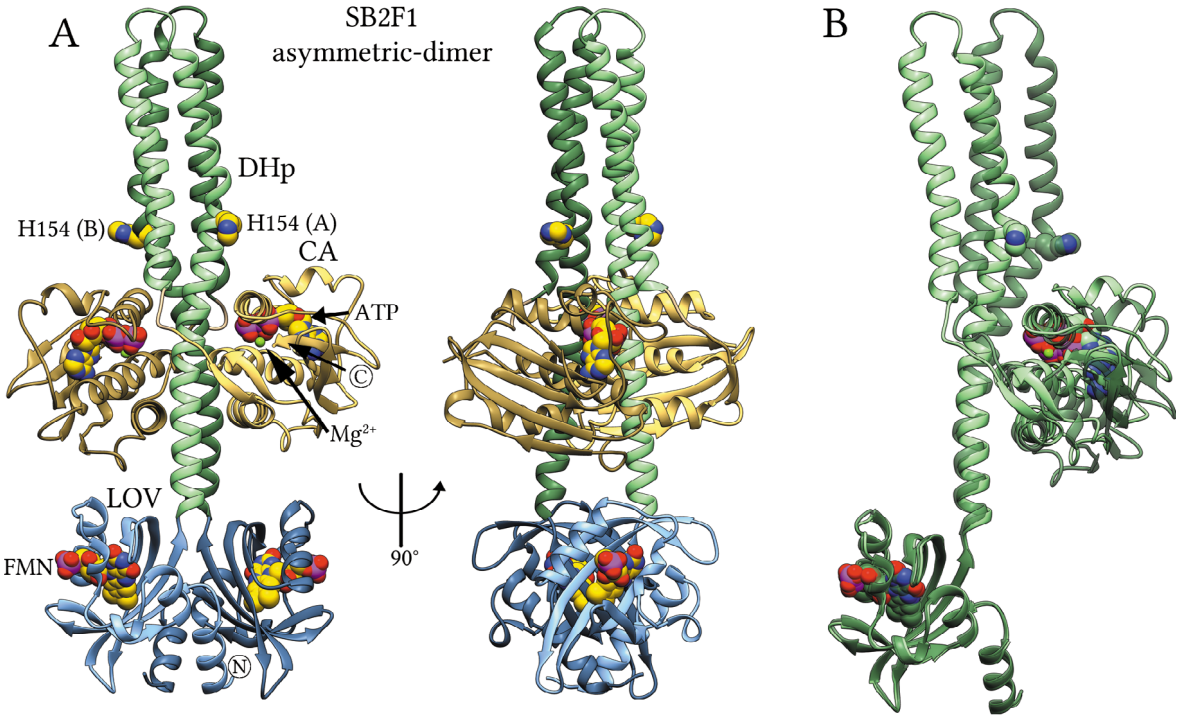


Figure 60. The crystal structure of the SB2F1 asymmetric-dimer with bound ATP. (**A**) The structure is represented as a ribbon diagram where the domains are colored as the following: LOV – blue, DHp – green and CA – gold. Chains A and B are colored in light and dark shades. FMN, ATP and the phospho-accepting His154 are shown as space-filling models. The magnesium ion is shown as a green sphere. (**B**) The monomers of the SB2F1 asymmetric-dimer are colored in light and dark shades of green and superposed by the residues 17 – 138 (the LOV core domain plus the DHp domain up to His138).

Superposition of this symmetric-dimer obtained from the short c-axis data with the asymmetric-dimer obtained from the long c-axis data by chain A, showed that the dimers differ in their tertiary structures. While chain A of the two structures roughly overlap, the CA domain from chain B adopts a different position. Such superposition with the focus on the whole chain, however, masks the degree of the differences between them. Thus, superposition by residues 17 – 138 of Chain A was done, with RMSD of the Cα atom positions of 1.6 Å (Figure 61B). It showed that there is not only a dramatic change in the position of the CA domain with respect to Chain B, but also the helix bundle of the DHp domain bends beyond the residue His138. For a comparison, monomers from the asymmetric-dimer were superposed with the monomer of the symmetric-dimer (Figure 61C). The superposition showed that the overall bending of the DHp domain helices in the monomer of the symmetric-dimer is somewhere between that of the asymmetric-dimer monomers.

The SB2F1 crystals, obtained by crystallization in the presence of ADP, had only a short c-axis. Similar to the symmetric-dimer, the asymmetric unit contained only one protein

molecule. Residues 1 – 365 out of 368 could be observed in the $2mF_o$-DF_c electron-density map. Application of the symmetry operation to the monomer with the crystallographic 2-fold axis leads to the structure of the dimer (Figure 62). Superposition of this dimer with the above-mentioned symmetric-dimer, showed that they are virtually identical, with the RMSD of the Cα atom positions of 0.61 Å over the residue range of 1 to 365.

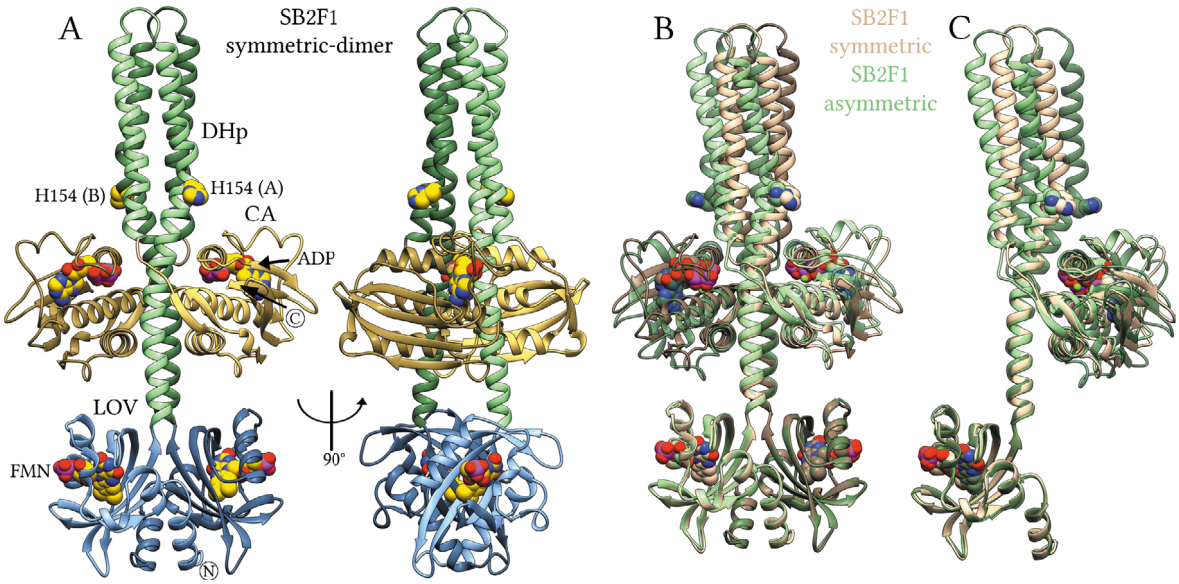


Figure 61. The crystal structure of the SB2F1 symmetric-dimer with bound ADP, which was crystallized in the presence of ATP. (**A**) The structure is represented as a ribbon diagram where the domains are colored as the following: LOV – blue, DHp – green and CA – gold. Chains A and B are colored in light and dark shades. FMN, ADP and the phospho-accepting His154 are shown as space-filling models. (**B**) Superposition of the SB2F1 asymmetric-dimer (green) with the SB2F1 symmetric-dimer (khaki) is shown, which were superposed by the residues range 17 – 138 of chain A (the LOV core domain plus the DHp domain up to His138). (**C**) The monomers of the SB2F1 asymmetric-dimer are colored in light and dark shades of green and the monomer of the SB2F1 symmetric-dimer in khaki, which were superposed by the residues range 17 – 138 (the LOV core domain plus the DHp domain up to His138).

The reason for the similarity in the structure of this dimer to that of the symmetric-dimer, and its dissimilarity to the asymmetric-dimer was found by comparison of their $2mF_o$-DF_c electron-density maps of the CA domains. In the datasets with the short c-axis, both CA domains have electron-density maps in which only the two phosphate groups of the bound nucleotide could be fitted. Whereas the datasets, which could only be evaluated by using the long c-axis, have electron-density maps in which three phosphate groups of the bound nucleotide and possibly the magnesium ion could be better fit.

This implies that during the crystallization of SB2F1, both in the presence of ATP and under dark conditions, at least two states were crystallized. The first state, represented by the asymmetric-dimer, is that where both CA domains have bound ATP and the second state, represented by the symmetric-dimer, is that where both CA domains have bound ADP. The second state is probably the result of the hydrolysis of ATP to ADP, which explains its similarity to the dimer structure obtained by crystallization in the presence of only ADP.

The AMP-PNP and ATP-γ-S nucleotides, are often used as non-hydrolysable analogs of ATP, yet some proteins are able to hydrolyze them as well (Olesen *et al.*, 2007; Miller and Lucius, 2014). The structures of SB2F1 obtained from crystals grown in the presence of AMP-PNP or ATP-γ-S nucleotides were also identical to the symmetric-dimer structure with bound ADP. It is possible that both AMP-PNP and ATP-γ-S were hydrolyzed either in the crystallization solution or by the CA domain, as only two phosphate groups could be fit into the electron-density map. The known instability of AMP-PNP in slightly acidic conditions might explain its hydrolysis in the crystallization buffer of SB2F1 at pH of 6.0 (Ferguson *et al.*, 2011).

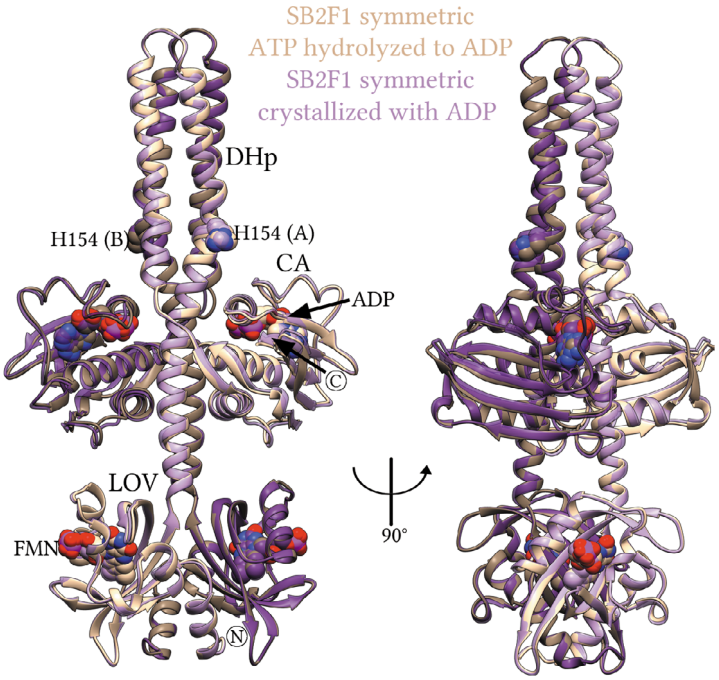


Figure 62. Superposition of the SB2F1 symmetric-dimers with bound ADP, which were crystallized in the presence of ATP and ADP. The structure obtained in the presence of ADP (plum) and the structure obtained in the presence of ATP but with bound ADP (khaki) are represented as ribbons. Chains A and B are colored in light and dark shades. FMN, ADP and the phosphor-accepting His154 are shown as space-filling models. After superposition, the RMSD of the Cα atom positions is 0.61 Å.

The illuminated asymmetric-dimer structure with bound ATP

The crystals of SB2F1 grown under continuous blue-light illumination, were not diffracting sufficiently enough to determine its structure. For this reason, the crystals, which were grown under dark conditions, were illuminated in order to see the structural changes induced by light. To accomplish this, the crystals grown in the presence of ATP were randomly selected, then illuminated with blue-light LED (λmax = 450 nm) at room temperature and subsequently flash-cooled to the temperature of 100 K.

The diffraction data collected from these crystals showed that they had only long c-axis. The resulting crystal structure contained two molecules in the asymmetric unit that forms a homodimer (Figure 63). The superposition of the illuminated crystal structure with

the asymmetric-dimer structure demonstrated their high similarity, with the RMSD of the Cα atom positions of 0.77 Å. Their superposition by chain A showed that the major differences were observed in the positions of their respective CA and LOV domains (Figure 63B). Consistent with the similarity of their structures, three phosphate groups of nucleotide could be fit in the electron-density map.

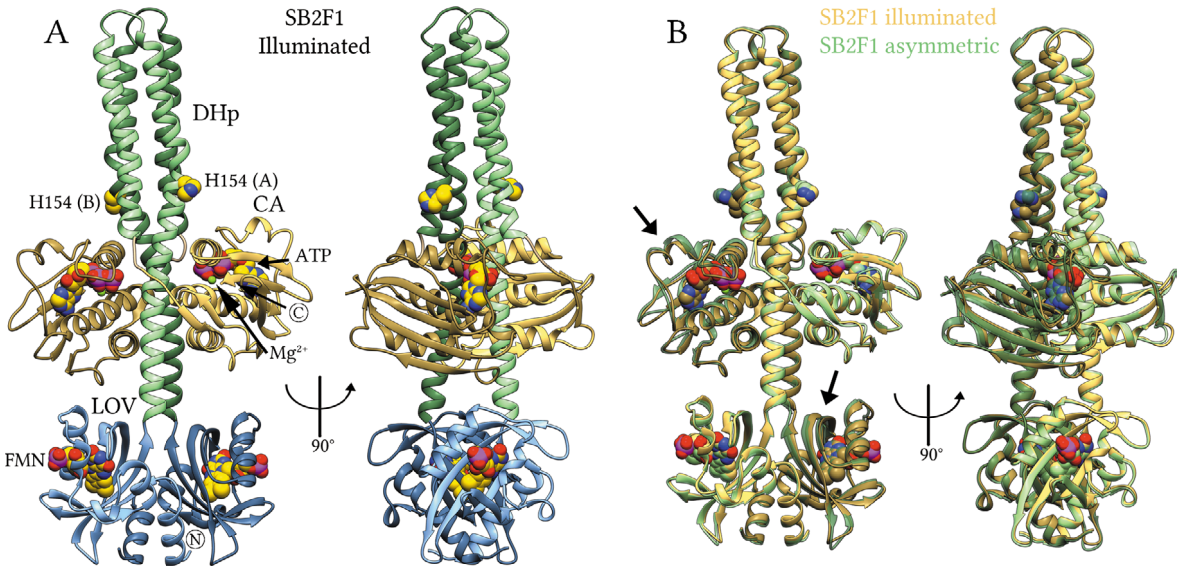


Figure 63. The illuminated crystal structure of the SB2F1 with bound ATP. (**A**) The structure is represented as a ribbon diagram where the domains are colored as the following: LOV – blue, DHp – green and CA – gold. Chains A and B are colored in light and dark shades. FMN, ATP and the phospho-accepting His154 are shown as space-filling models. The magnesium ion is shown as a green sphere. (**B**) Superposition of the SB2F1 asymmetric-dimer (green) and the illuminated structure (gold) is shown, which were superposed by the residues range 17 – 138 (the LOV core domain plus the DHp domain up to His138).

Microspectrometry showed that the spectrum of the illuminated crystals matched the spectrum of the SB2F1 in the light state measured in solution (Figure 57). Nevertheless, no continuous electron density was observed between the Cys53-Sγ and FMN-C4a atoms, which would have indicated the formation of an adduct between them. However, previous reports on LOV proteins mentioned that the adduct, typical for the light state, may not be observed in the crystal structure. Several reasons were mentioned, such as the sensitivity of the adduct to radiation damage during X-ray exposure or the fast dark recovery kinetics, which prohibits the stabilization of the adduct (Zoltowski *et al.*, 2007; Endres *et al.*, 2015).

From previous results on short LOV protein and as general consequence of crystal packing, protein molecules in crystals are often impaired to undergo large-scale motions upon external stimuli (Vaidya *et al.*, 2011; Röllen *et al.*, 2016). For instance, illumination of the PpSB1-LOV crystals resulted in a structure that was very close to the dark state structure and could not reproduce the light state structure, which were obtained under permanent blue-light illumination during the crystals growth (Röllen *et al.*, 2016). A similar observation was made in this work in the case of the illuminated PpSB2-LOV crystals.

To proceed with the crystallization of SB2F1 under permanent blue-light illumination, the I66R mutant of SB2F1 was designed to increase the adduct lifetime. This may improve the crystallization of the protein in the light state.

The structure of SB2F1 I66R in the light state

First, the SB2F1 I66R mutant was crystallized under dark conditions in the presence of ATP to verify that mutation did not induce changes in the structure. The collected diffraction data contained datasets with both short and long c-axis, as it was observed in the case of SB2F1.

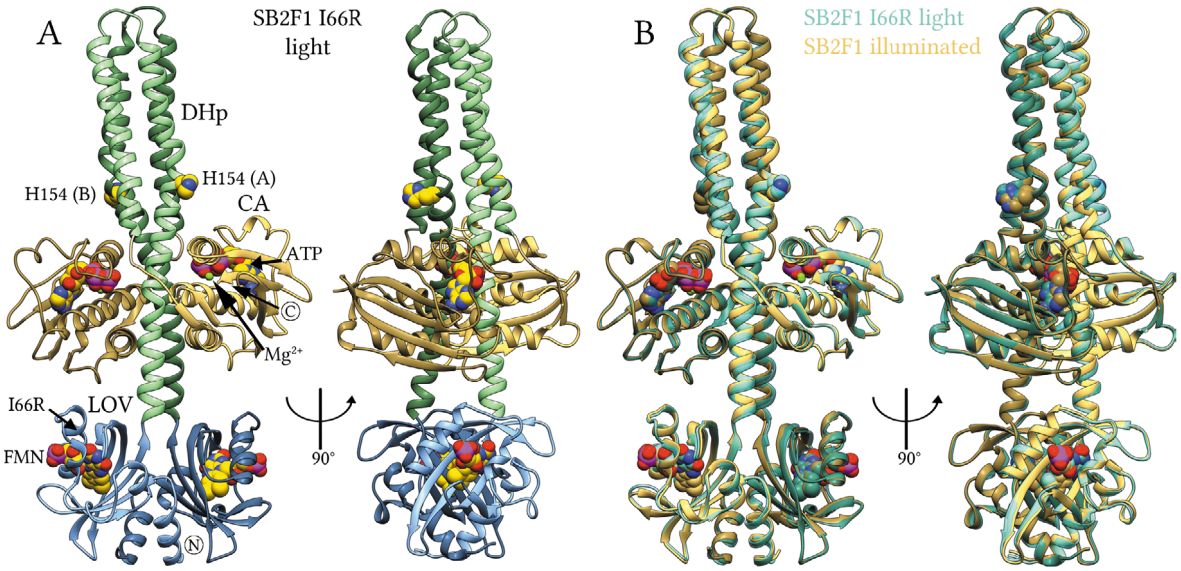


Figure 64. The light state crystal structure of the SB2F1 I66R with bound ATP. (**A**) The structure is represented as a ribbon diagram where the domains are colored as the following: LOV – blue, DHp – green and CA – gold. Chains A and B are colored in light and dark shades. FMN, ATP and the phospho-accepting His154 are shown as space-filling models. The magnesium ion is shown as a green sphere. (**B**) Superposition of the SB2F1 illuminated structure (gold) with the SB2F1 I66R light state structure (cyan) is shown, which were superposed by the residues range 17 – 138 (the LOV core domain plus the DHp domain up to His138).

The superposition of the SB2F1 I66R structure from the short c-axis dataset with the symmetric-dimer of SB2F1 showed the similarity of their structures. Furthermore, the CA domains contain a bound ADP, as it is in case of the symmetric-dimer of SB2F1. Superposition of the SB2F1 I66R structure from the long c-axis dataset with the asymmetric-dimer of SB2F1 also showed their structural similarity where both contained a bound ATP. Such results indicate that the I66R mutation did not alter the structure of SB2F1.

Next, the SB2F1 I66R mutant was crystallized under continuous blue-light illumination to facilitate the unrestricted crystallization of the protein in the light state. Microspectrometry showed that the spectrum of the crystals grown under light conditions matches the spectrum of the SB2F1 in the light state measured in solution (Figure 57).

However, no continuous electron density was observed between the Cys53-Sγ and FMN-C4a atoms, which would have indicated the formation of an adduct between them.

This observation can be attributed to the same reasons as those for the above-mentioned illuminated structure. However, the lower resolution of the SB2F1 I66R structure compared to that of the illuminated structure of SB2F1 (3.5 Å vs 2.5 Å) could also contribute to this finding. Additionally, it should be mentioned that the absence of the electron density for the adduct does not imply that protein was not in the conformation that is typical of the light state.

The superposition of the crystal structure of SB2F1 I66R in the light state with the structure of SB2F1 obtained from illuminated crystals demonstrated their high similarity, with the RMSD of the Cα atom positions of 0.61 Å over the residue range of 1 to 365 (Figure 64). The superposition of both structures by the residues 17 – 138 of Chain A showed small differences in the position of the CA and LOV domains in addition to small changes in the bending of the helix bundle from DHp. The fact that the illuminated and light state structures do not differ dramatically from the structure of the asymmetric-dimer will be discussed next in light of the SAXS data.

Catalytic and ATP binding domain

Analysis of all the SB2F1 and SB2F1 I66R structures revealed that asymmetric-dimer has the electron-density map of the CA domain that better fits ATP, whereas the electron-density map of symmetric-dimer better fits ADP (Figure 65). The DHp domain of SB2F1 contains two histidine residues, His138 and His154. According to the conserved sequence motif of the histidine kinases, **H**(D/E)(L/I)(K/R)(T/N)PL, the His154 should be the phospho-accepting histidine residue (Diensthuber *et al.*, 2013; Bhate *et al.*, 2015). The distance between the Nδ atom of His154 and the phosphate groups of bound ATP or ADP to the CA domains was approximately 20 Å in all structures. Such a long distance indicates that this might be not the active state.

However, the Nε atom of the His138 residues in both monomers of the asymmetric-dimer structures was in the approximate distance of 6 Å to the γ-phosphate group of ATP. This is surprising, nevertheless such a short distance suggest that, the His138 can be phosphorylated. The Nε atoms of the His138 in the symmetric-dimer structures were also in the approximate distance of 6 Å – 7.5 Å to the β-phosphate group of ADP.

Additionally, a somewhat smeared electron-density map was observed for the His138 in one of the monomers, yet it perhaps cannot be explained by its phosphorylation for a few reasons. First, phosphorylated histidine residue is known to be rather unstable and was not observed so far in any crystal structure (Klumpp and Krieglstein, 2002) and second, the "extra" density is smaller than the phosphoryl group would have been. Taking these considerations, two conformations of the histidine side chain were refined.

The smeared electron-density maps observed for the residues in the nucleotide binding pocket made it difficult to compare the structural changes induced by the binding of ATP or

ADP. Nevertheless, the Arg141 residue, which coordinates the phosphate group of the nucleotide and the His138 residue, seems to change its conformation depending on the binding of ATP or ADP. Apparently, only two instead of three phosphate groups of ADP enable Arg141 residue to adopt a more elongated conformation to facilitate the coordination of the β-phosphate group of ADP.

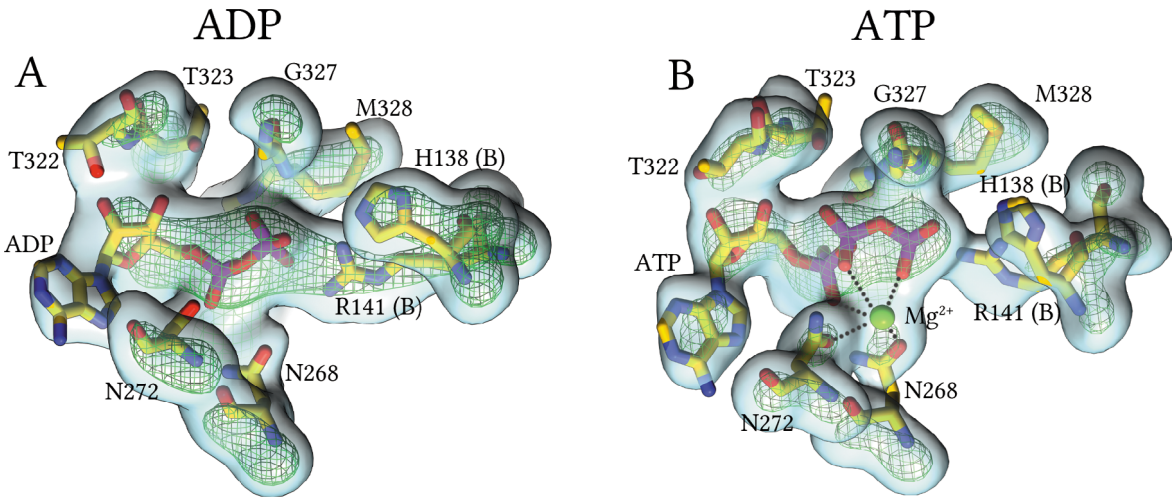


Figure 65. The $2mF_o$-DF_c electron-density maps for the ATP and ADP binding pockets in the SB2F1 structures. The $2mF_o$-DF_c electron-density map contoured at 1 rmsd is shown as a blue transparent surface, while the same map contoured at 4 rmsd is shown as a green mesh. ATP, ADP and selected residues are shown as stick models and colored by element: carbon, yellow; nitrogen, blue; oxygen, red; phosphorus, purple; sulfur, orange. The magnesium ion is shown as a green sphere. The His138 and Arg141 residues are from the chain B and all other residues from the chain A. (**A**) ADP was used in the refinement of the symmetric-dimer (**B**) ATP and the magnesium ion was used in the refinement of the asymmetric-dimer. Distances between magnesium ion and other oxygen atoms within ≤ 3.2 Å are shown as dotted lines.

In the ATP binding pocket of the asymmetric-dimer there is continuous electron density between Asn268, Asn272 and the phosphate groups of ATP, even at high rmsd contouring of the $2mF_o$-DF_c electron-density maps. On the contrary, this continuous density disappears at lower rmsd contouring in the case of ADP. Superposition with the structurally similar kinase domain, WalK (PDD ID: 3SL2), showed that WalK with bound ATP has a magnesium ion located approximately in this continuous density. Refinement including the magnesium ion with partial occupancy resulted in relative occupancies of 88% and 64%, in chains A and B, respectively. The mF_o-DF_c maps also showed no negative difference density beyond the 3rmsd contouring, which indicates that placement of the magnesium ion approximately in this continuous density may be plausible. However, due to the low resolution of the maps, this cannot be proved.

The CA domain of the SB2F1 protein is connected by a flexible loop to the C-terminus of the DHp domain and anchored to the middle part of the DHp domain of other monomer (Figure 66). The interaction between the CA and DHp domains is mostly mediated by hydrophobic interactions, hydrogen bonds and salt bridges. The attachment position of the CA domain is believed to depend on its state.

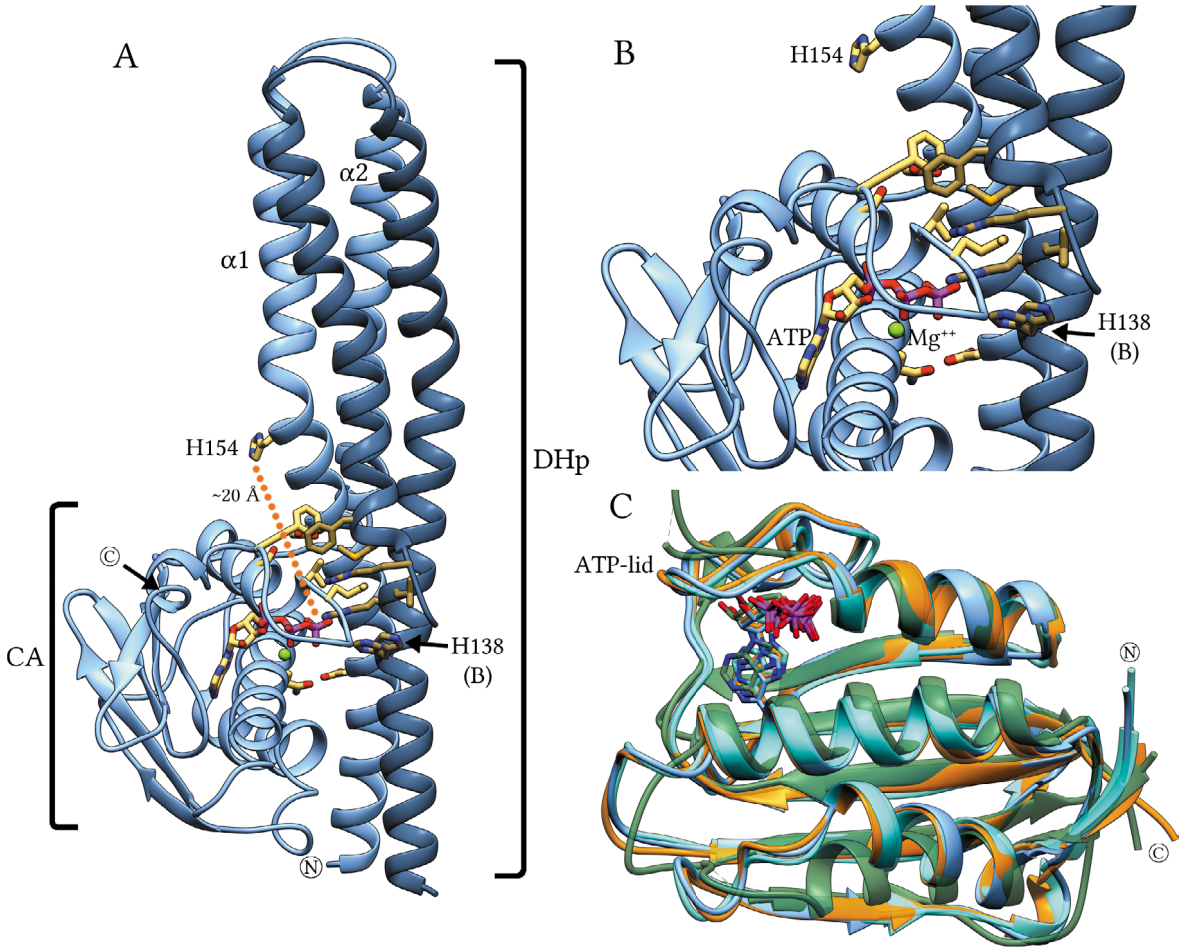


Figure 66. Interactions between the CA and DHp domains in the SB2F1 asymmetric-dimer; comparison of the CA domains from SB2F1, YF1 and WalK. (**A**) Interacting CA and DHp domains in the SB2F1 asymmetric-dimer are represented as ribbon diagrams, where chains A and B are colored in dark and light shades of blue. ATP and selected residues, which form contacts between the CA and DHp domains, are shown as stick models. The magnesium ion is shown as a green sphere. (**B**) Magnified section of panel A. (**C**) Superposition of the CA domains from SB2F1 (blue – asymmetric-dimer, cyan – symmetric-dimer), YF1 (PDB ID: 4GCZ) (orange) and WalK (PDB ID: 3SL2) (green). ATP from the Walk and the SB2F1 asymmetric-dimer, and ADP from the SB2F1 symmetric-dimer and the YF1 are shown as stick models with carbon atoms colored in the model color.

Attachment of the CA domain to the DHp domain in both the ADP- and ATP-bound structures are similar and differ only in their relative position and their orientation to the DHp domain. The main reason for the different positions and orientations is the change in the overall bending of the DHp domain helices. It can be speculated that the specific interactions, such as that of the Arg141 with the phosphate group may play a role in the change of the bending angles together with other still unknown factors. Indeed, as mentioned above, the binding of ADP or ATP resulted in the different conformations of Arg141, where Arg141 is located close to the position where the helices of the DHp domain starts to bend.

Superposition of the CA domains from the SB2F1 symmetric- and asymmetric-dimers with the CA domains of the YF1 (PDB ID: 4GCZ, Diensthuber *et al.*, 2013) and WalK (sequence identity of 30% to SB2F1) crystal structures (Celikel *et al.*, 2012) showed that the domain fold itself is well-preserved (Figure 66B). The CA domains of the WalK and SB2F1

asymmetric-dimer have bound ATP, while that of the YF1 and SB2F1 symmetric-dimer have bound ADP. It was previously reported for the CA domain of other histidine kinases, that the long loop, the so-called ATP-lid, changes conformation depending on the presence and type of the bound nucleotide (Bhate *et al.*, 2015). In the structures of SB2F1, on the contrary, the ATP-lid stays in the same conformation when bound to ADP or ATP.

LOV domain

The PpSB2-LOV protein (residues 1 – 120) was used in the construction of SB2F1 by analogy to the YF1 protein (Möglich, Ayers and Moffat, 2009a). Comparison of the LOV core domain from SB2F1 with that of PpSB2-LOV showed that their structure remained almost identical. Most of the side chains rotamers, besides a few exceptions, also stayed the same, which indicates that the fusion of the LOV domain did not alter its structure or function.

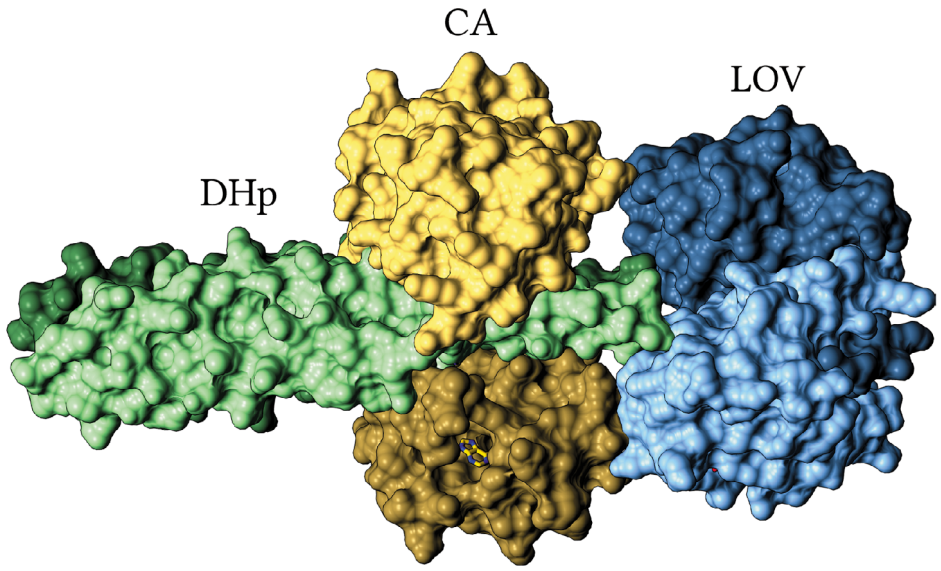


Figure 67. Domain contacts in the SB2F1 asymmetric-dimer structure represented as a molecular surface. Domains are colored as the following: LOV – blue, CA – gold, DHp – green, where chains A and B are colored as dark and light shades.

The two globular domains, LOV and CA, in the previously determined structure of YF1 are situated far from each other, which prevents their direct interaction (Diensthuber *et al.*, 2013). The LOV and CA domains in the SB2F1 structures, on the contrary, are situated such that the side chains of the CA domain residues can interact with the Fα helix of the LOV domain. The displacement of one CA domain in the illuminated and light structures toward one of the LOV domains increased this contact even more than it did in the symmetric-dimer (Figure 67).

Dimer interface

A dimerization and histidine phosphotransfer (DHp) domain in histidine kinases is responsible for the formation of a dimer and for the proper positioning of the phospho-

accepting histidine residue during the catalytic reaction. Moreover, the DHp domain is also responsible for the proper recognition of the conjugated response regulator, which is necessary for further signal propagation.

PISA analysis (Krissinel and Henrick, 2007) showed that the total surface area buried by the dimer interface was 8982 Å^2 for the asymmetric-dimer and 8548 Å^2 for the symmetric-dimer. Such a large area together with high solvation energy gain upon dimer formation (ΔG) of -76 kcal/mol and -69 kcal/mol is indicative of a stable dimer in solution. For a comparison, a similar engineered protein, YF1 (Diensthuber *et al.*, 2013), has a total surface area buried by the dimer interface of 8844 Å^2 with a ΔG of -83 kcal/mol.

The dimer interface of SB2F1 in all presented structures is formed not only by the DHp domains but also partially by the LOV domains. The influence triggered by the changes in the dimer interface of the LOV domains might be a key factor for the signal transduction from the LOV domains to the CA domains.

The arrangement of the dimer interface of the LOV domains in the SB2F1 structures is similar to that of PpSB2-LOV. In PpSB2-LOV, part of the interface is formed by the N-terminal A'α helices and partially by the β-strands, while the rest of it is formed by the C-terminal Jα helices. In the SB2F1 protein, the helices of the DHp domain instead of the Jα helices, contribute to the dimer interface in a similar way. However, the contribution of the DHp domain is significantly higher than that of the Jα helices, which is based on the coiled-coil interactions caused by heptadic residue repeats. Such coiled-coil like interactions between the DHp domains of monomers beyond residue 151 leads to the formation of the four-helix bundle.

Comparison of the asymmetric- and the symmetric-dimers of the SB2F1, which only differ in LOV domain to the previously determined structure of similar YF1 protein (Diensthuber *et al.*, 2013) might suggest how LOV domain influence the quaternary structure of the whole protein or give valuable information about changes related to nucleotide binding. Superposition of them demonstrated that the dimer interfaces differ dramatically (Figure 68). YF1, in contrast to the SB2F1, has a pronounced 35° kink within the N-terminal region of the DHp domain, which is believed to arise from the intermolecular packing within the crystal lattice (Diensthuber *et al.*, 2013). Such kinking of the DHp domain results in the dramatically different positions of the CA domains. While one of the CA domains has a similar anchoring to the DHp domain as the SB2F1 CA domains have, the second CA domain has a different anchoring and orientation with respect to the DHp domain. The different orientation and position of this CA domain, however, resulted in the closer distance (~13 Å) between the phospho-accepting His161 residue (YF1 sequence numbering, equal to the His154 residue in SB2F1) and the ADP.

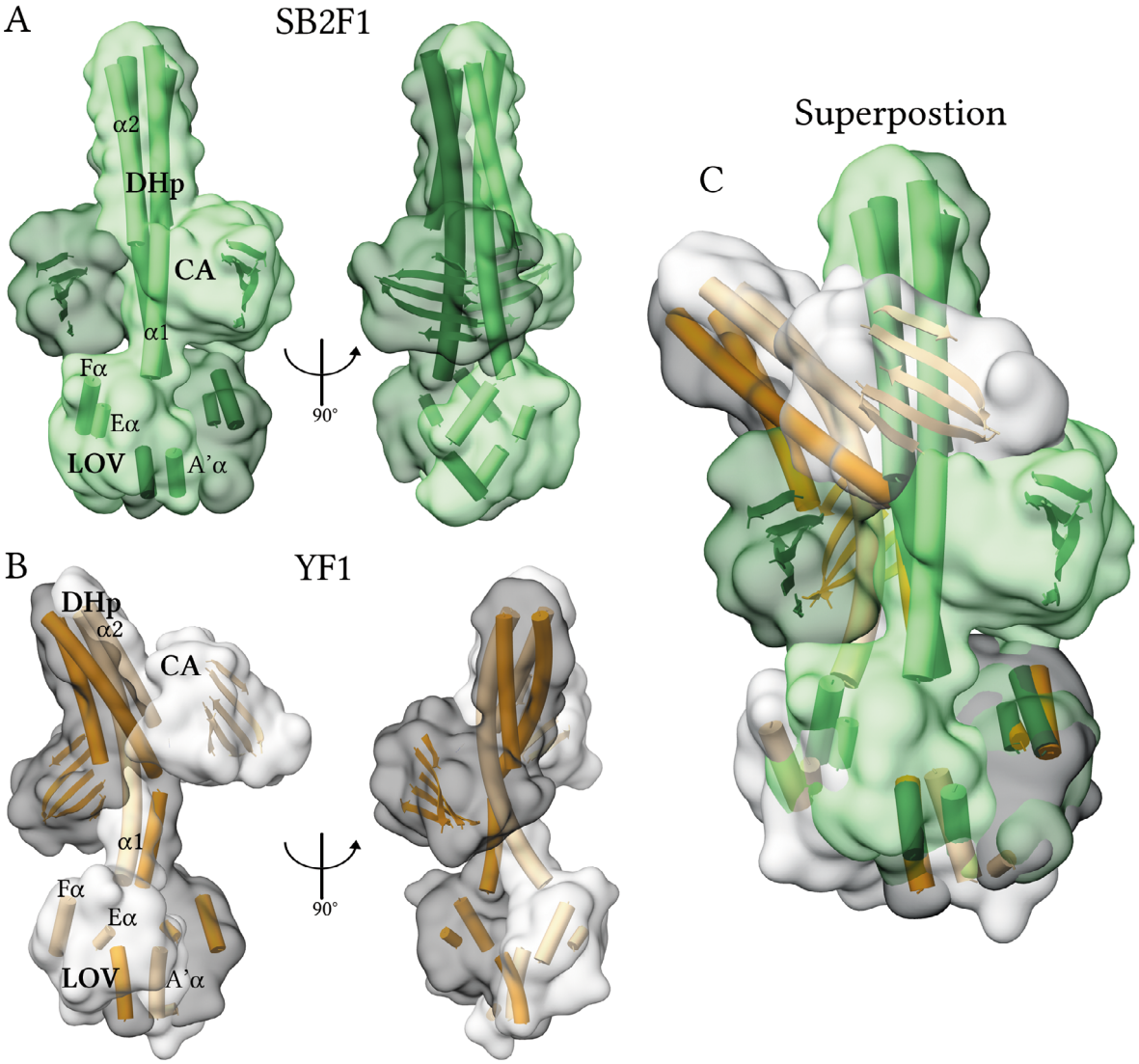


Figure 68. The superposition of the SB2F1 asymmetric-dimer with the crystal structure of the YF1 dimer. (**A**) The structure of SB2F1 (green) is represented as a transparent surface. Several structural elements of each domain are shown, where helices and strands are represented as cylinders and arrow ribbon, respectively. Chain A and B are colored as dark and light shades. (**B**) The structures of YF1 (PDB ID: 4GCZ, Diensthuber *et al.*, 2013) (orange) is represented as in (A). (**C**) The superposition of the SB2F1 with YF1 by the residues range 17 – 138 of chain A (the LOV core domain plus the DHp domain up to His138 in the SB2F1 residue numbering). Coloring and representation are as in (A, B).

SAXS data of SB2F1 in the dark and light state

The SAXS data of SB2F1 was measured both with and without ATP, and in the dark and light states, as summarized in Table 38. The molecular weight (MW) estimation based on the dynamic light scattering (DLS) and the size-exclusion chromatography (SEC) techniques showed that SB2F1 remains a dimer in each state. An overestimation of the MW, as determined by SEC, could be due to the non-globular shape of the SB2F1 protein. The estimation of the MW from SAXS data, based on the Porod volume calculation, also agrees well with the calculated MW of the dimer. However, the MW determined from the I_0 value is apparently underestimated, due to the fact that the calculation of I_0 depends on the exact determination of protein concentration, and thus can be prone to errors.

The radius of gyration (R_g) values calculated for all SB2F1 crystal structures were identical and equal to 3.4 nm. Estimation of the R_g from the SAXS data based on the Guinier approximation, in contrast, showed the variation between the different conditions. Such variation seems to correlate with the value of the longest dimension (D_{max}) of a molecule. According to the R_g value of 3.7 nm and the D_{max} value of 12.8 nm, the ATP-bound SB2F1 in the light state has the most compact conformation compared to the other states. Whereas, the ATP-bound SB2F1 in the dark state has an R_g value of 3.9 nm and a D_{max} value of 13.7 nm, which indicates a more elongated conformation. Such differences in values suggest that the structural changes induced by light should be higher than that observed for the SB2F1 structures obtained from crystals grown under dark or light conditions. Moreover, such structural changes are can be the reason of the different autophosphorylation rates of SB2F1 previously observed in the dark and light states (Krauss, unpublished).

Table 38. Overview of the molecular weight (MW) determination and SAXS data for the SB2F1 measured with and without ATP under dark and light conditions. The MW from Porod volume is calculated by dividing the volume by a factor of 1.7. The χ^2 value shows the agreement between the theoretical scattering based on crystal structures and the experimental data.

	SB2F1 dark		**SB2F1 light**	
	nucleotide-free	ATP-bound	nucleotide-free	ATP-bound
Calculated MW for monomer/dimer [kDa]	42.8 / 85.6			
SEC MW [kDa]	117			
DLS MW [kDa]	99			
SAXS MW from I_0 [kDa]	64.1	64.3	64.7	66.7
SAXS MW from Porod volume [kDa]	81.8	80.0	80.6	78.8
Crystal structure R_g [nm]	3.4			
Guinier R_g from SAXS [nm]	4.1	3.9	4.0	3.7
D_{max} from SAXS [nm]	14.4	13.7	14.0	12.8
χ^2 for the ADP-bound symmetric-dimer	5.5	5.4	5.2	2.6
χ^2 for the ATP-bound asymmetric-dimer	4.8	4.6	4.4	1.9
χ^2 for the ATP-bound illuminated	4.7	4.6	4.3	1.8
χ^2 for the ATP-bound SB2F1 I66R light	4.7	4.8	4.3	2.0
χ^2 for the YF1*	3.3	2.3	3.1	3.6

* PDB ID: 4GCZ (Diensthuber *et al.*, 2013)

Next, a comparison of the theoretical scattering of the SB2F1 crystal structures to the experimental data was done to understand how structures obtained from crystals differ to those observed in solution (Table 38 and Figure 69). The asymmetric-dimer, the illuminated structures of SB2F1 with bound ATP, and the structure of SB2F1 I66R in the light state demonstrated the best agreement with the SAXS data obtained from SB2F1 under light conditions in the presence of ATP, with the χ^2 values ranging from 1.8 to 2.0. Such agreement with the illuminated structure of SB2F1 and to the structure of SB2F1 I66R in the light state was expected, but not with the asymmetric-dimer structure of SB2F1, which was obtained from dark grown crystals. Moreover, the asymmetric-dimer structure of SB2F1 also matched to a lesser degree with the SAXS data obtained from SB2F1 under dark conditions in the presence of ATP, with a χ^2 value of 4.8.

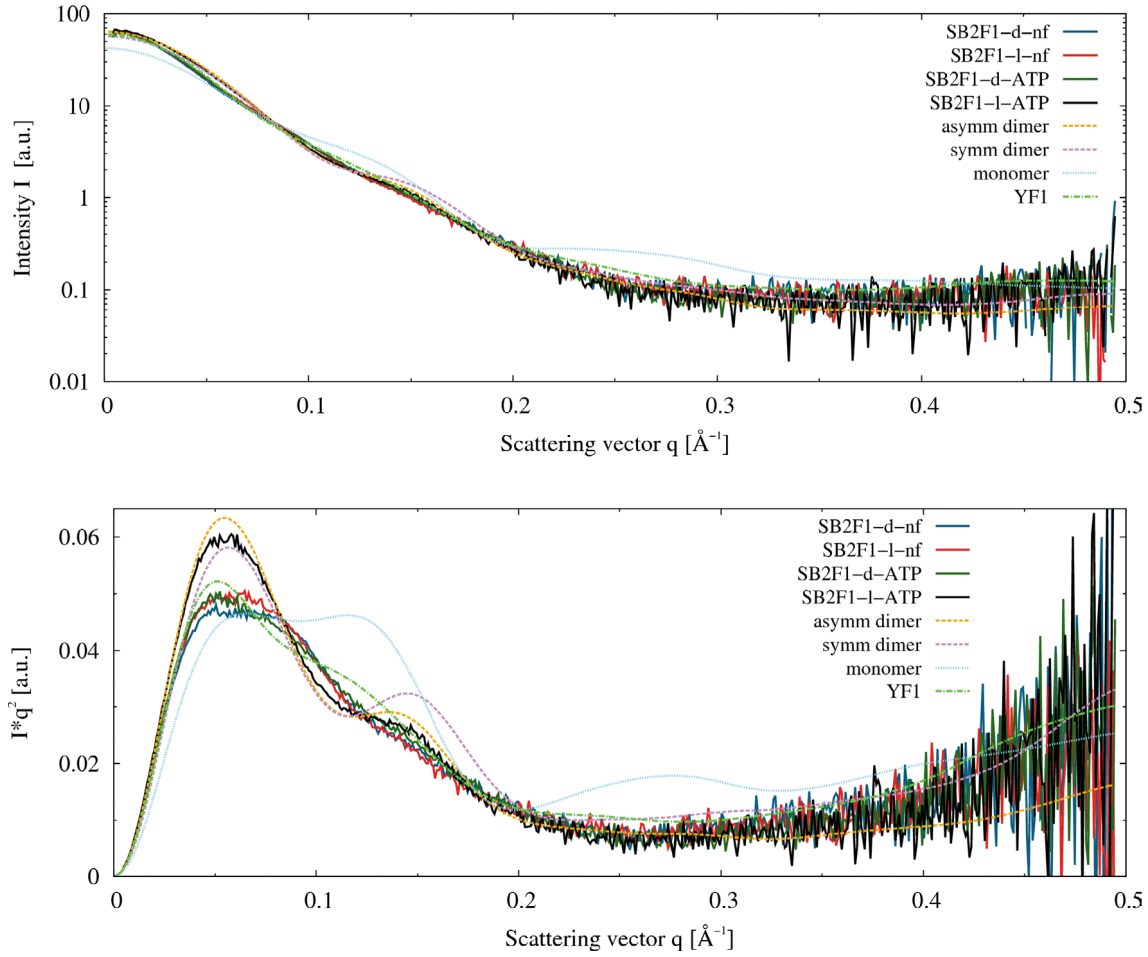


Figure 69. SAXS scattering curves of SB2F1 fitted to the theoretical scattering of the SB2F1 dimers, the SB2F1 monomer and the YF1 dimer. **(top)** The SAXS scattering data collected for the nucleotide-free (nf) SB2F1 in the dark (blue) and light (red) states are shown as solid lines. The SAXS scattering data collected for SB2F1 in the dark (green) and light (black) states incubated with 1.5 mM ATP, 2 mM $MgCl_2$ are shown as solid lines. The theoretical scattering curves for the asymmetric-dimer (orange), the symmetric-dimer (purple) and the monomer (cyan) of the SB2F1 crystal structure and the YF1 (PDB ID: 4GCZ, Diensthuber *et al.*, 2013) (green) structure in the dark state were fitted to the experimental SAXS data and shown as dashed lines. (**bottom**) The same data represented as Kratky plot.

The previously determined structure of similar protein YF1 (PDB ID: 4GCZ, Diensthuber *et al.*, 2013), obtained under dark conditions and in the presence of ATP, has ADP bound to one of the CA domains while the second CA domains has no bound nucleotide (Diensthuber *et al.*, 2013). Comparison of the theoretical scattering of the YF1 structure to the SAXS data obtained from SB2F1, under dark conditions in the presence of ATP, showed good agreement with a χ^2 value of 2.3. The good agreement indicates that the tertiary structure of YF1 is quite similar to that of SB2F1 in solution.

Based on the SAXS scattering data of SB2F1, *ab initio* envelope models were calculated and aligned with the crystals structures of SB2F1 and YF1 (Diensthuber *et al.*, 2013) to investigate the spatial distribution of the observed differences (Figure 70 and Figure 71). The illuminated structure of SB2F1 fitted well to the predicted envelope model of the

SB2F1 with ATP in the light state, which agrees with above-mentioned lower value of the χ^2 for the curve fitting of theoretical scattering to the SAXS data.

The symmetric-dimer of SB2F1 with bound ADP has a higher χ^2 value compared to the other structures of SB2F1, nevertheless, it also fitted just as well to the envelope model. This is not surprising as the intrinsically "low resolution" of calculated *ab initio* envelope models is not precise enough to identify small structural differences. However, the χ^2 value of 2.6 of the ADP-bound SB2F1 compared to the χ^2 value of 1.8 – 2.0 of the ATP-bound SB2F1, indicates that the observed difference between the ADP- and ATP-bound structures are likely to be valid for the solution structures as well.

The *ab initio* envelope models based on the SAXS scattering data of SB2F1 under dark condition and in the presence of ATP demonstrated a somewhat different shape compared to that obtained under light conditions (Figure 70). Alignment of this envelope with the illuminated structure of SB2F1 showed that the crystal structure could not cover the whole volume of the envelope model (Figure 71). On the contrary, the previously determined structure of YF1 (Diensthuber *et al.*, 2013) with its kinked shape, fitted better and covered the whole volume of the envelope model.

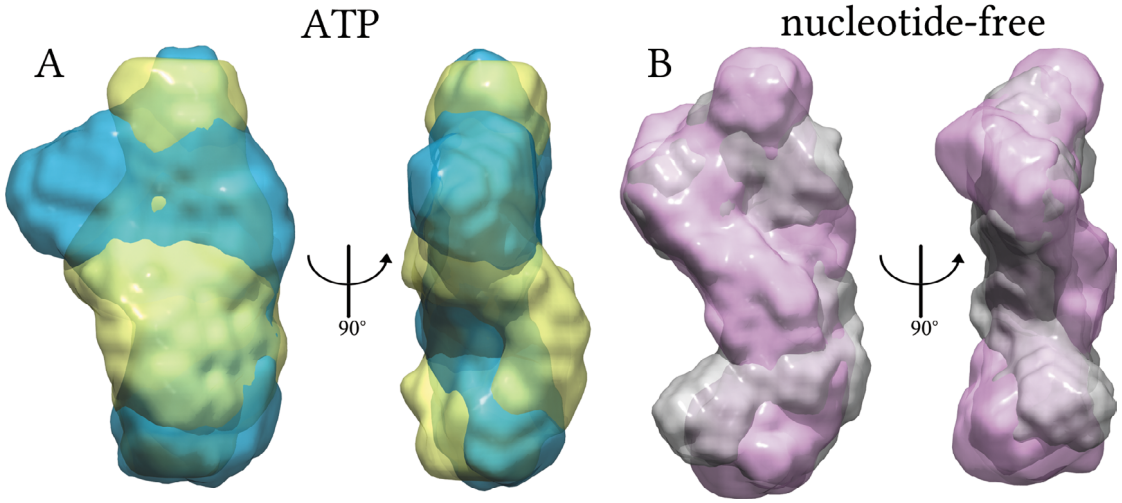


Figure 70. The SAXS envelope models of the dark and light states of SB2F1 in the ATP-bound and nucleotide-free states. (**A**) The aligned SAXS envelopes of the ATP-bound SB2F1 in the dark (blue) and light (yellow) states are shown as transparent surfaces. (**B**) The aligned SAXS envelopes of the nucleotide-free SB2F1 in the dark (gray) and light (plum) states are shown as transparent surfaces.

So far, the presented SAXS data suggest that all obtained structures of the SB2F1 and SB2F1 I66R are close to the light state structure of SB2F1 with bound ATP. The symmetric-dimer of SB2F1 with bound ADP is close to the state with bound ATP, although it should have some differences as indicated by a slightly higher χ^2 value. The structure of YF1 in the dark state, believed by their authors to be distorted by crystal packing (Diensthuber *et al.*, 2013), may indeed resemble the tertiary structure of the ATP-bound SB2F1 in the dark state.

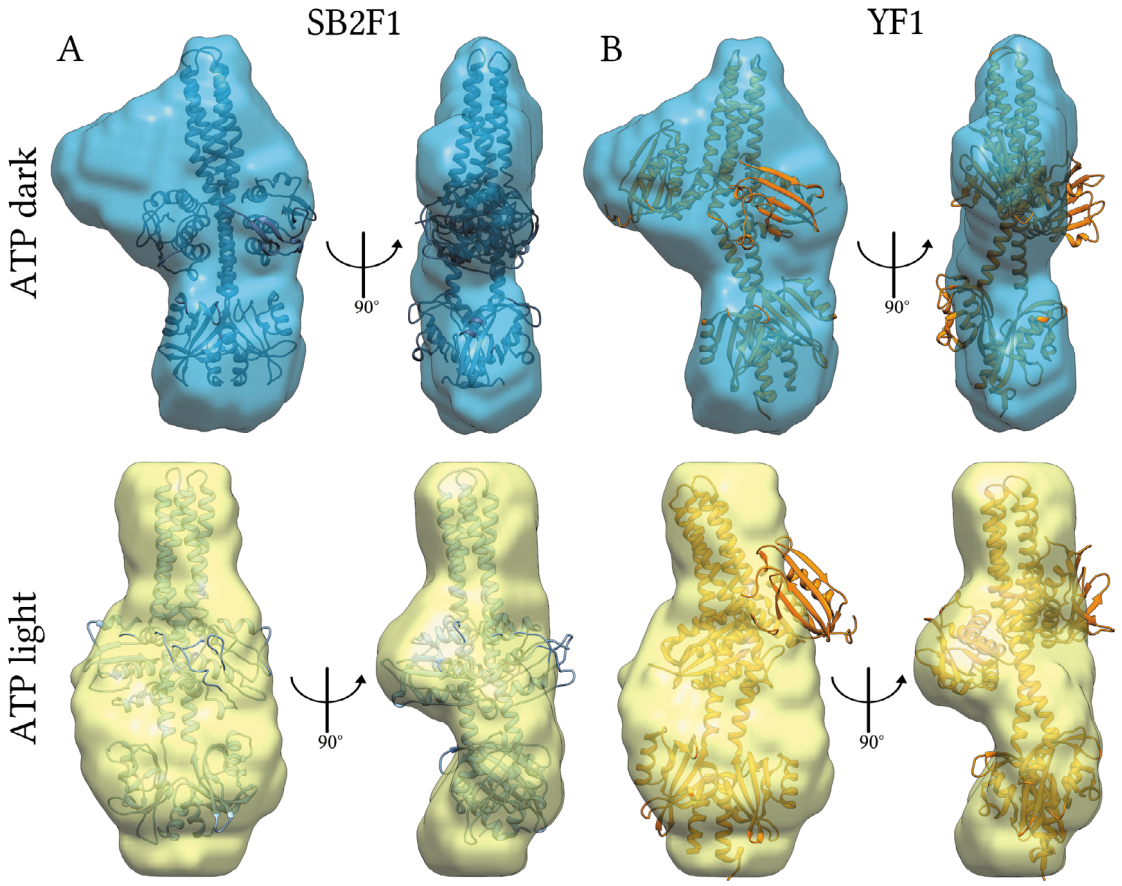


Figure 71. SAXS envelope models of the ATP-bound SB2F1 in the dark and light states aligned with the SB2F1 illuminated crystal structure and the structure of YF1 in the dark state. The illuminated crystal structure of SB2F1 (blue) and the dark state structure of YF1 (PDB ID: 4GCZ, Diensthuber *et al.*, 2013) (orange) are shown as ribbons and aligned with the SAXS envelope models of the dark (blue) and light (yellow) states of the ATP-bound SB2F1, which are shown as transparent surfaces. (**top row**) The SAXS envelope of the ATP-bound SB2F1 in the dark state aligned with the illuminated crystal structure of SB2F1 (**A**) and the dark state structure of YF1 (**B**). (**bottom row**) The SAXS envelope of the ATP-bound SB2F1 in the light state aligned with the illuminated crystal structure of SB2F1 (**A**) and the dark state structure YF1 (**B**).

The *ab initio* envelope models based on the SAXS data from SB2F1 in the nucleotide-free state, as expected from the theoretical scattering, did not fit well to any of the SB2F1 or YF1 (Diensthuber *et al.*, 2013) structures (Figure 72). This suggests that nucleotide binding causes significant changes to the tertiary structure. Moreover, there were also some differences between envelope models of the nucleotide-free SB2F1 in the dark and light states. However, the absence of any atomistic model makes it impossible to draw detailed information on the nucleotide-free state based only on the SAXS data. The data presented in the next section on the similar SB1F1 protein will show some additional details on the nucleotide-free state.

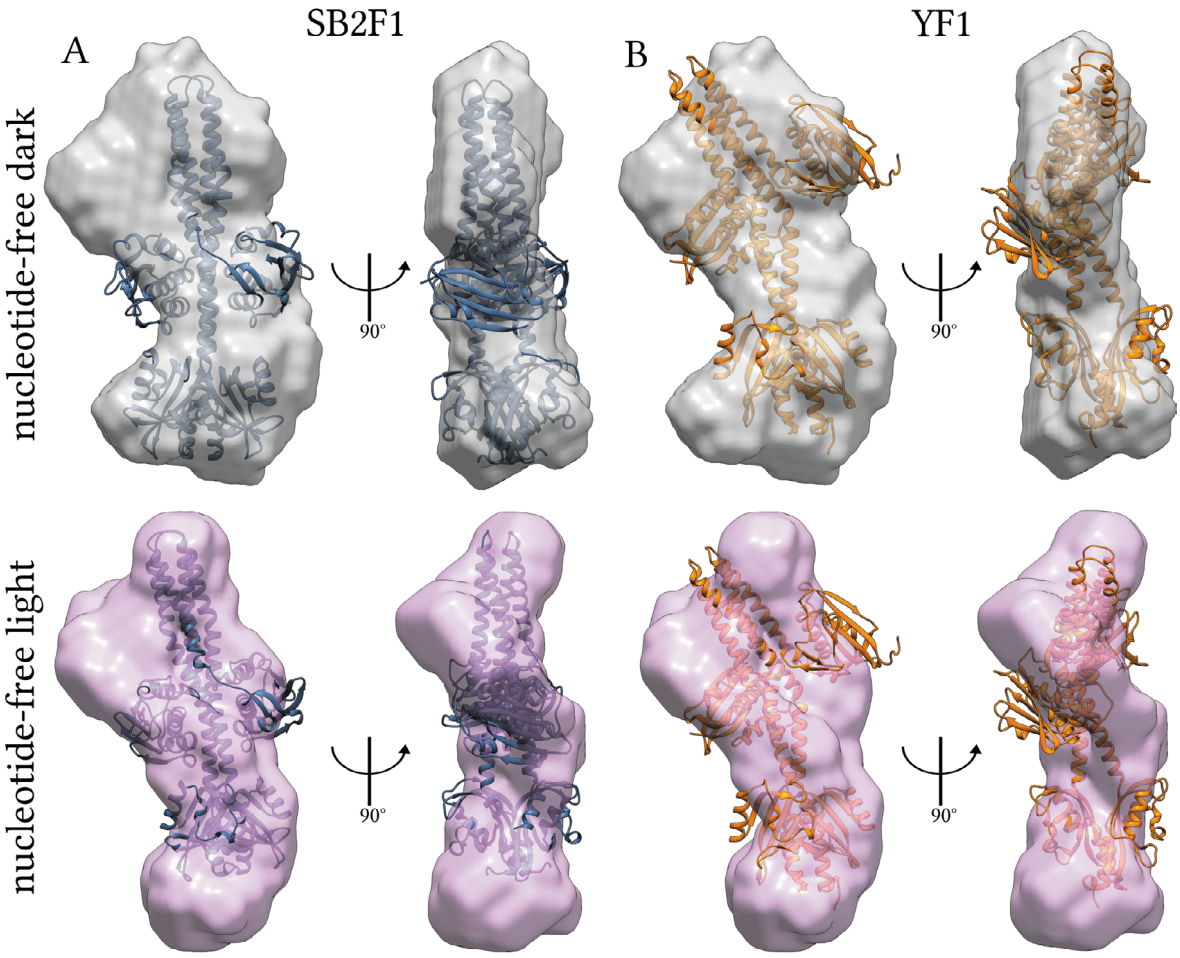


Figure 72. The SAXS envelope models of the nucleotide-free SB2F1 in the dark and light states aligned with the illuminated crystal structure of SB2F1 and the dark state structure of YF1. The SB2F1 illuminated crystal structure (blue) and the YF1 dark state structure (PDB ID: 4GCZ, Diensthuber *et al.*, 2013) (orange) are shown as ribbons and are aligned with the SAXS envelope models of the dark (gray) and light (plum) states of the nucleotide-free SB2F1 and shown as transparent surfaces (**A**) The SAXS envelope models of the nucleotide-free SB2F1 in the dark state aligned with the SB2F1 illuminated crystal structure and the YF1 dark state structure. (**B**) The SAXS envelope model of the nucleotide-free SB2F1 in the light state aligned with the same structures.

3.5 SB1F1

UV-Vis spectroscopy and dark recovery kinetics

The SB1F1 is a chimeric sensory histidine kinase, which was engineered in similar fashion, as SB2F1 and YF1 (Möglich, Ayers and Moffat, 2009a; Diensthuber *et al.*, 2013) were, see additional details on page 17 (Krauss, unpublished). For this, the oxygen-sensitive PAS domain in the histidine kinase FixL from *Bradyrhizobium japonicum* was replaced by the LOV domain from PpSB1-LOV protein (residues 1 – 120) from *Pseudomonas putida*. Beyond residue 120, SB1F1 and SB2F1 share with YF1 an identical sequence for the DHp (dimerization and histidine phosphotransfer) and CA (the catalytic and ATP binding) domains (Figure 8).

SB1F1 protein showed photocycle properties such as the bleaching upon illumination with blue-light and the thermally-driven dark recovery. The dark state spectrum of SB1F1, similar to that of the parental PpSB1-LOV, showed three characteristic peaks around 450 nm (423 nm, 448 nm and 475 nm) and a double peak at around 365 nm (Figure 73). After the blue-light illumination, only a single broad peak at around 390 nm was observed, which was identical to that of PpSB1-LOV.

The adduct lifetime (τ_{rec}) of SB1F1 in the absence of any nucleotide was measured to be 2850 min at 20 °C (Figure 73). The PpSB1-LOV has comparably long τ_{rec} of 2471 ± 22 min (Jentzsch *et al.*, 2009).

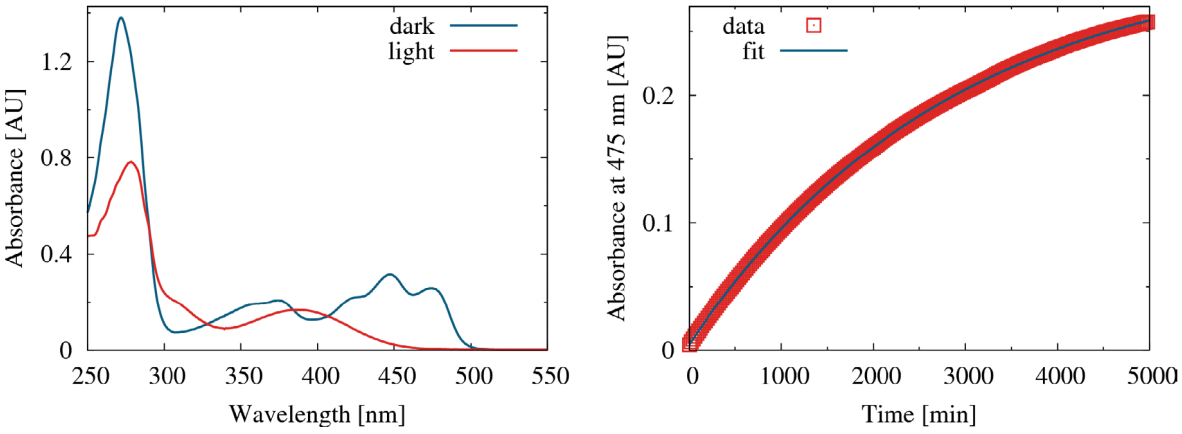


Figure 73. The UV-Vis spectra of the SB2F1 in the dark and light states in solution together with the dark recovery kinetics. (**left**) The dark (blue) and light (red) states spectra are shown as solid lines. (**right**) The dark recovery kinetics of the nucleotide-free SB1F1 was measured as an increase in the absorbance at 475 nm over-time at 20 °C in 10 mM Na_2HPO_4/NaH_2PO_4 pH 8.0, 10 mM NaCl.

Crystallization of SB1F1

SB1F1 was successfully crystallized in the nucleotide-free state under dark and light conditions. Crystallization trials of SB1F1 in the presence of nucleotide were unsuccessful. The X-ray diffraction data on nucleotide-free crystals was collected and evaluated (Table 39).

Unfortunately, despite the similarity of the SB1F1 sequence to that of the SB2F1 and YF1 proteins, the molecular replacement phasing procedure failed when these structures were used as search templates. The possible reason for this failure might be that the SB1F1 in the nucleotide-free state adopts a significantly different tertiary structure. This sounds plausible considering the fact that the envelope models of SB2F1 in the nucleotide-free state based on the SAXS data were not fitting well to the ADP- or ATP-bound structures of SB2F1.

In order to experimentally obtain the initial phase information, the selenomethionine-labeled protein was produced. Due to the small size of the crystals and low diffraction power, only the SAD anomalous scattering data was collected. The collected diffraction data has relatively low resolution of 3.8 Å for the crystals grown under dark conditions (Table 39). The values of anomalous signal correlation, CC_{anom}, indicated that anomalous signal was significant only until the resolution of 4.6 Å, which additionally decreased the chances for successful SAD phasing.

During the evaluation of the data, the positions of the four selenium atoms (out of 10 for each protein monomer) were found, however, with only low confidence. The initial phase information derived from these found atoms was apparently not enough to produce an interpretable $2mF_o$-DF_c electron-density map. No specific features of the electron-density map could be identified in order to enable the placement of protein residues.

Table 39. SB1F1 data collection statistics.

X-ray data	SB2F1 dark native	SB2F1 dark Selenomethionine	SB2F1 light native
Beamline	ID23-1, ESRF	ID23-1, ESRF	ID30A-3, ESRF
Detector	PILATUS 6M	PILATUS 6M	Eiger 4M
Wavelength [Å]	0.97296	0.97879	0.96770
Resolution range [Å]	47.09 – 3.28 (3.54 – 3.28)*	47.17 – 3.76 (4.20 – 3.76)*	47.15 – 3.70 (4.05 – 3.70)*
Space group	I422 or $I4_122$	I422 or $I4_122$	I422 or $I4_122$
Unit-cell a, b, c [Å]	127.25 127.25 140.03	129.37 129.37 137.86	127.57 127.57 140.02
α, β, γ [°]	90 90 90	90 90 90	90 90 90
Total reflections	171028 (35997)	152814 (43413)	138661 (33884)
Unique reflections	9123 (1837)	6226 (1721)	6434 (1495)
Multiplicity	18.7 (19.6)	24.5 (25.2)	21.6 (22.7)
Completeness [%]	100.0 (100.0)	99.8 (99.5)	99.8 (100.0)
Mean I/σ(I)	12.4 (1.0)	6.9 (3.3)	7.5 (1.3)
Wilson B-factor [$Å^2$]	88.7	68.4	162.5
R-merge	0.176 (3.444)	0.585 (2.348)	0.644 (5.545)
R-meas	0.186 (3.627)	0.608 (2.442)	0.674 (5.790)
$CC_{1/2}$	0.999 (0.331)	0.990 (0.769)	0.993 (0.415)

* Statistics for the highest-resolution shell are shown in parentheses.

In the future new phasing strategies need to be used to solve these structures. These can include a MAD experiment or preparation of heavy atom derivatives of protein crystals, which would produce higher anomalous signal than that of the selenium atoms. In addition,

certainly the search for new crystallization conditions or improvements on the current one needs to be further pursued, which would yield crystals with better diffraction.

SAXS data and *ab initio* models

The SAXS data was collected with the SB1F1 protein in the nucleotide-free state under both dark and light conditions. The results of the SAXS data together with the results from size-exclusion chromatography (SEC) are listed in Table 40. The molecular weight (MW) estimations based on the SEC suggested that SB1F1 is a dimer in the nucleotide-free state. An overestimation of the MW, as determined by SEC, could be due to the non-globular shape of the SB1F1 protein. The estimation of the MW from SAXS data, based on the Porod volume calculation, also agrees well with the calculated MW of the dimer. However, the MW determined from the I_0 value is apparently underestimated, due to the fact that the calculation of I_0 depends on the exact determination of protein concentration, and thus can be prone to errors.

Table 40. Overview of molecular weight determination (MW) and SAXS data for the SB1F1 measured in the nucleotide-free state both in the dark and light states. MW from Porod volume is calculated by dividing the volume by a factor of 1.7. χ^2 shows agreement between theoretical scattering based on crystal structures and experimental data.

	SB1F1 dark	SB1F1 light
Calculated Mw for monomer/dimer [kDa]	40.7 / 81.4	
SEC MW [kDa]	120	
SAXS MW from I_0 [kDa]	66	70
SAXS MW from Porod volume [kDa]	82	82
Guinier R_g from SAXS [nm]	4.1	4.0
Dmax from SAXS [nm]	14.2	14.2
χ^2 for the ADP-bound symmetric-dimer	4.5	4.4
χ^2 for the ATP-bound asymmetric-dimer	4.6	4.5
χ^2 for the YF1*	1.7	1.8

* PDB ID: 4GCZ (Diensthuber *et al.*, 2013)

A comparison of the agreement between the theoretical scattering based on the crystal structures of SB2F1 and YF1 (PDB ID: 4GCZ, Diensthuber *et al.*, 2013), and experimental data, expressed with χ^2 values, are shown in Table 40. The discrepancy between the theoretical scattering of the SB2F1 structures in the nucleotide-bound state and the experimental scattering of SB1F1 in the nucleotide-free state was somewhat similar to that seen in the SAXS data of SB2F1. In contrast, the theoretical scattering of YF1 (Diensthuber *et al.*, 2013) structure fitted better to the data of SB1F1 in the nucleotide-free state than to the scattering curve of SB2F1 with bound ATP, with the χ^2 value of 2.3 vs 1.7, respectively. The high similarity of the theoretical scattering of YF1 structure to the both scattering curves, however, does not imply that the differences between the states are the same. This can lead to the conclusion that the YF1 structure containing the large kink is somehow similar to that of the SB1F1 in the nucleotide-free state and SB2F1 in the dark state with bound ATP, but are not identical in detail.

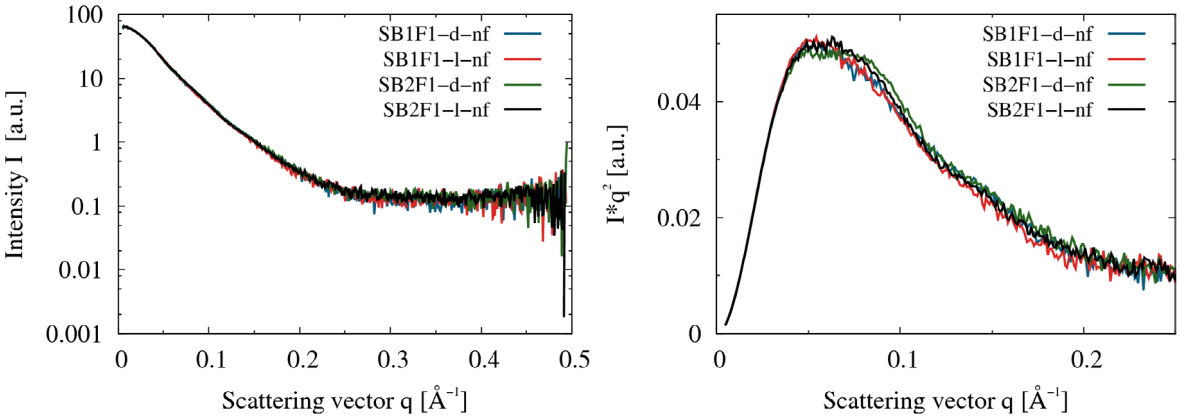


Figure 74. The SAXS scattering curves of the nucleotide-free SB1F1 and SB2F1 in the dark and light states (**left**). (**right**) The same data in the q range of 0 – 0.25 Å^{-1} represented as Kratky plot.

The estimated radius of gyration (R_g) from the SAXS data of SB1F1 in the nucleotide-free state was ~4 nm, which is equal to the estimated R_g of the SB2F1 in the nucleotide-free state from SAXS data. The value of the longest dimension, D_{max}, is also close to that of SB2F1, suggesting strong similarities between the proteins in their nucleotide-free states. Indeed, the scattering curves and envelope models of both proteins fitted well to each other, displaying only minor differences (Figure 74 and Figure 75).

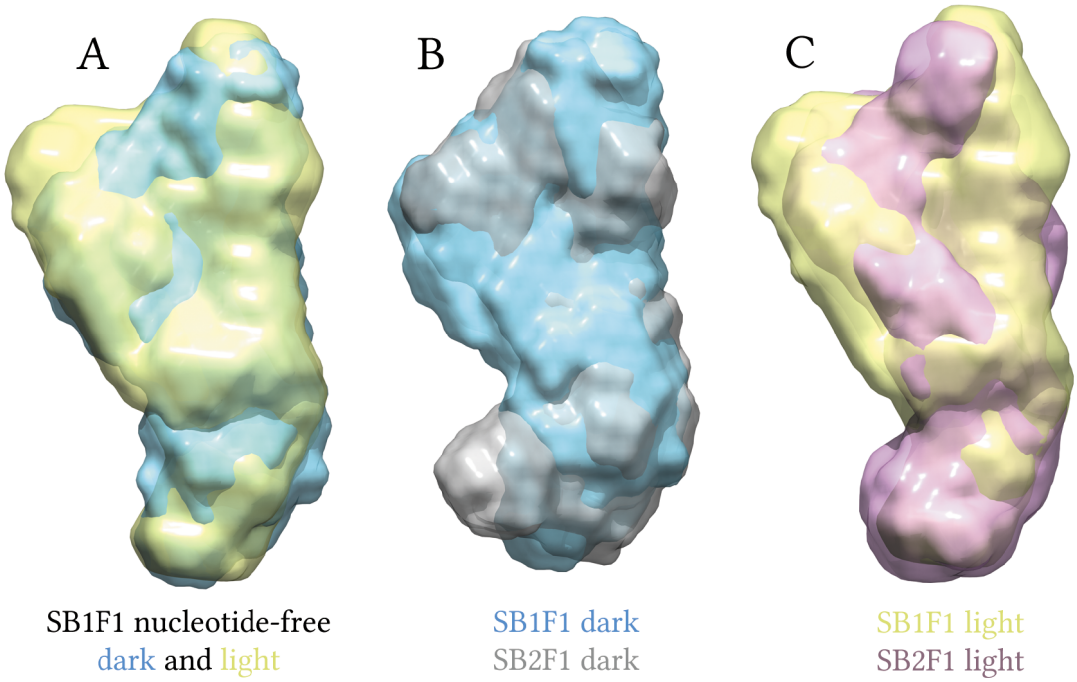


Figure 75. The SAXS envelope models of the dark and light states of SB1F1 in the nucleotide-free state. (**A**) The aligned SAXS envelopes of the nucleotide-free SB1F1 in the dark (blue) and light (yellow) states are shown as transparent surfaces (**B**) The aligned SAXS envelopes of the nucleotide-free SB1F1 (blue) and SB2F1 (gray) in the dark states. (**C**) The aligned SAXS envelopes of the nucleotide-free SB1F1 (yellow) and SB2F1 (plum) in the light states.

4 Discussion

4.1 Dark recovery kinetics

Different flavin chromophores influence dark recovery kinetics

LOV domains are known to bind naturally available flavins, like: FMN, RBF or FAD. Flavin chromophores have a high binding affinity to LOV proteins such that during heterologous expression in bacteria they bind ubiquitously available flavins. Depending on the expression rate and available amount of flavins it often results in mixed chromophore content. Relative amount of bound chromophores, however, is not proportional to their binding affinities due to a limit in the amount of available flavins in the cytosol during protein over-expression. The over-expression of the PpSB2-LOV in *E. coli* BL21 (DE3), in this study, resulted in the following chromophore composition: 70.0% of FMN, 21.3% of RBF and 8.7% of FAD. After *in vitro* loading of PpSB2-LOV with an excess of FMN, it resulted in a chromophore content of almost 99% FMN. Such exchange is direct evidence that FMN has higher affinity to the PpSB2-LOV than other flavins, whereas due to the limit of available flavins PpSB2-LOV also binds RBF and FAD during expression.

The dependence of the biophysical/biochemical properties of LOV proteins on different flavin chromophores was previously studied with the YtvA-LOV protein (Mansurova *et al.*, 2011; Dorn *et al.*, 2013). It was found that YtvA-LOV with bound RBF or FAD have different adduct lifetimes compared to the FMN-bound YtvA-LOV, with 85 min and 172 min compared to 152 min, respectively (Mansurova *et al.*, 2011). W619_1-LOV also showed chromophore-dependent changes of the adduct lifetimes, although with a higher magnitude. The RBF- and FAD-bound W619_1-LOV have adduct lifetimes of 1.5 min and 54 min at 20 °C, which is 307 and 8.5 times faster than that of the FMN-bound W619_1-LOV, which has an adduct lifetime of 460 min (Table 31).

Such a remarkably higher acceleration of dark recovery, compared to YtvA-LOV, can be related to the sequence feature of W619_1-LOV, such as the cluster of four arginine residues: Arg54, Arg61, Arg66 and Arg70. This cluster coordinates the phosphate group of FMN and was first identified in the PpSB1-LOV protein (close homologue of W619_1-LOV) (Jentzsch *et al.*, 2009; Circolone *et al.*, 2012). It was suggested that such tight coordination of

the phosphate group reduced the thermal fluctuations of FMN and thus slowed the dark recovery. Based on the results of this work, however, additional details will be presented next.

The study of solvent isotope effect on LOV proteins demonstrated that the rate limiting step in the dark recovery was the deprotonation of the FMN-N5 atom (Swartz *et al.*, 2001; Corchnoy *et al.*, 2003). The tight coordination formed by the salt bridges of the four arginine residues cluster might decreases the solvent accessibility of the adduct-forming cysteine residue and the FMN-N5 atom, and thus decreases the rate of the FMN-N5 atom deprotonation. The same considerations about the solvent accessibility and the deprotonation rates were also previously discussed and supported by the site-directed mutagenesis experiments (Christie *et al.*, 2007; Zoltowski, Vaccaro and Crane, 2009; Zoltowski, Nash and Gardner, 2011).

The aforementioned explanation is also well-supported by the molecular dynamics (MD) simulation data, where the absence of two arginine residues in PpSB2-LOV led to the increased mobility of the Eα helix, the Eα-Fα loop and the Fα helix, that surrounds the chromophore and form its binding pocket (Figure 26). Comparison of the solvent-accessible cavities of the PpSB2-LOV, PpSB1-LOV and SBW25-LOV chromophore pockets showed the most extensive volume of the cavity with lateral protrusions for the PpSB2-LOV, which has the shortest adduct lifetime among them (Figure 20 and Figure 40).

SBW25-LOV, in contrast to W619_1-LOV and PpSB1-LOV, has only three arginine residues to coordinate the phosphate group of FMN. Nevertheless, they seem to stabilize the residues surrounding the chromophore by forming salt bridges to the phosphate group of FMN in an effective manner without competition as in PpSB1-LOV, where four arginine residues coordinate three oxygens of the phosphate group (Figure 39). This tighter coordination is well-reflected in the decreased main chain dynamics of the Eα helix, the Eα-Fα loop and the Fα helix in the MD simulation of the SBW25-LOV, when compared to PpSB1-LOV (Figure 42).

Using the above-mentioned interpretation, one can explain the dramatic acceleration of photocycle of the W619_1-LOV with bound RBF. Only a single OH group of RBF instead of a phosphate group of FMN cannot establish salt bridges with all four arginine residues at the same time. The absence of these interactions leads to a higher dynamics for the four arginine residues, which in turn destabilizes the Eα helix, the Eα-Fα loop and the Fα helix. This destabilization can result in an increase in solvent access, and thus accelerate the deprotonation of the FMN-N5 atom.

The modest acceleration of the dark recovery that was observed in the W619_1-LOV with bound FAD can be likewise explained. Although FAD has the diphosphate group, it still possess less conformational freedom to adopt the optimal position to form salt bridges to the four arginine residues. Especially, the diphosphate group is not free but is instead connected to

the dinucleotide moiety of FAD. Altogether this might lead to a higher dynamics and conformational exchange of FAD and interacting residues, which in turn increases the solvent accessibility.

The less dramatic effect of the exchange of FMN to RBF or FAD in YtvA-LOV can be related to the fact that it has only two conserved arginine residues and thus the absence of the phosphate group would not have a strong influence on the dark recovery kinetics. Substitution of one of the conserved arginine residues, Arg63 ahead of the conserved cysteine Cys62 (Cys53 and Arg54 in the W619_1-LOV) in YtvA (full-length protein containing YtvA-LOV) to lysine residue accelerated the dark recovery eight-fold (Tang *et al.*, 2010). However, destabilization of the environment in the vicinity of the conserved cysteine residue can affect the dark recovery more directly.

Lateral solvent access to the chromophore influences dark recovery kinetics

Aforementioned arginine residues cluster, however, cannot solely explain variations in the adduct lifetimes. For instance, Pf5-LOV, a close homolog of SBW25-LOV, has three arginine residues like its homolog does, but Pf5-LOV still has an adduct lifetime of only 3.6 min vs 1470 min at 37 °C of SBW25-LOV (Rani *et al.*, 2013). The reason for such discrepancy seems to be that the lateral stabilization in the chromophore pocket of SBW25-LOV decreased the accessibility of solvent to it. Indeed, three arginine residues "close" the pocket by interaction with the phosphate group of FMN only from the Eα helix, the Eα-Fα loop and the Fα helix side, but not from the side of the C-terminal end of the Fα helix and the Aβ-Bβ loop. On that side, SBW25-LOV, in comparison to other faster LOV protein of this study, has a hydrogen bond network formed by the Glu74, Arg23 and Lys28 residues, as depicted Figure 40. In the MD simulation of SBW25-LOV, these residues also showed reduced dynamics, and the regions around them also became more rigid (Figure 42). Such interactions are missing in Pf5-LOV, which has a similar arrangement of the residues on this side like PpSB2-LOV, which also displays fast dark recovery. Therefore these residues are likely responsible for the difference in adduct lifetimes between these proteins.

Another bacterial short LOV protein, RsLOV, also has three arginine residues Arg56, Arg68 and Arg72 (equivalent to the Arg54, Arg66 and Arg70 of SBW25-LOV and Pf5-LOV), which coordinate the phosphate group of FMN (Conrad, Bilwes and Crane, 2013). The chromophore pocket from the side of the C-terminal end of the Fα helix and the Aβ-Bβ loop in RsLOV, similarly to the Pf5-LOV, is not well "closed" from solvent. This can probably explain why RsLOV, despite the tight coordination of the phosphate group, still has rather short adduct lifetime of ~40 min (no temperature was mentioned in the publication, assuming it was done at room temperature) (Conrad, Bilwes and Crane, 2013).

Imidazole effect on the dark recovery kinetics

Addition of imidazole can accelerate the kinetics of dark recovery in LOV proteins (Alexandre *et al.*, 2007; Zoltowski, Vaccaro and Crane, 2009; Herman *et al.*, 2013). It was proposed that in the light state, imidazole acts as a base and participates in the deprotonation of the FMN-N5 atom. The kinetics of the PpSB1-LOV and PpSB2-LOV dark recovery were also influenced by imidazole (Figure 12), although at comparably higher concentrations than that reported for AsLOV2 (Alexandre *et al.*, 2007). In the case of AsLOV2, a 10-fold increase was achieved already at 2 mM concentration of imidazole, whereas a similar acceleration was achieved in PpSB1-LOV and PpSB2-LOV at 250 mM and 500 mM concentrations, respectively (Figure 12). In the structure of AsLOV2, a solvent channel leading to the chromophore pocket was identified, which would be used by imidazole to approach the chromophore and deprotonate the FMN-N5 atom by acting as a general base. On the contrary, no such channels could be identified in the crystal structures of the PpSB1-LOV and PpSB2-LOV, which might explain their weaker response to imidazole.

Comparison of the x-fold acceleration of the adduct lifetime, of PpSB1-LOV and PpSB2-LOV at the same concentration of imidazole, showed that PpSB1-LOV has stronger x-fold acceleration than PpSB2-LOV (Figure 12). This result is unexpected as PpSB1-LOV should have a lower solvent accessibly to the chromophore pocket than PpSB2-LOV, according to the above-presented data on the pocket solvent-accessible cavities and the MD simulations. However, this discrepancy might be explained if imidazole participates in the deprotonation of the FMN-N5 atom indirectly in the both proteins, through the deprotonation of residues in the vicinity of the chromophore.

Indeed, it can be then speculated whether His61 of PpSB2-LOV, which is situated in proximity of the phosphate group and the ribityl chain of FMN, can influence deprotonation rates. First of all, a histidine residue itself can act as a base and thus trigger sequential deprotonation of residues or water molecules which are situated along the way to the FMN ring system (Chan and Bogomolni, 2012). The pKa value of His61 perhaps is lowered due to the strong electrostatic interactions of the phosphate group and arginine residues in the vicinity of His61 (Figure 19). In such case, the addition of imidazole would only moderately accelerate the deprotonation of histidine residues, and thus weaken the effect of imidazole, as it just complements with the deprotonation carried out by the histidine residue itself.

Indeed, when dark recovery kinetics of PpSB2-LOV was fit by a biexponential function, the x-fold acceleration of the adduct lifetime for one exponential function was almost twice as high as that of the second exponential function. Thus, it is possible that imidazole has stronger effect on one deprotonation pathway. On the contrary, PpSB1-LOV data was sufficiently fit just by a single exponential function. It can be related to the fact that PpSB1-LOV has an arginine residue at position 61, which might explain the stronger effect of imidazole, due to the intrinsically higher side chain pKa values of arginine residue.

4.2 First structure of LOV protein in the apo form and implications of use of natural and modified flavins as a chromophores

Fluorescent protein reporters are widely used for many *in vivo* and *in vitro* experiments (Chudakov, Lukyanov and Lukyanov, 2005). Most common among them are the green fluorescent protein (GFP) and its derivatives. GFP protein derivatives are extensively developed and their excitation wavelengths cover the full visible light spectrum (Matz, Lukyanov and Lukyanov, 2002). However, GFP proteins have certain limitations, for instance, the autocatalytic development of the chromophore requires molecular oxygen. During over-expression this can lead to the accumulation of the inactive, non-fluorescing reporter proteins as well as limit its applications in aerobic environments.

To overcome such limitations, novel fluorescent reporter proteins, based on LOV domains, were extensively developed in the last decade (Drepper *et al.*, 2007; Wingen *et al.*, 2014; Pudasaini, El-Arab and Zoltowski, 2015). In contrast to GFP proteins, LOV proteins bind and use ubiquitously available flavins as a chromophore, which have intrinsic fluorescence properties and do not require oxygen for the development of the chromophore. Additionally, LOV domains have a relatively small size of ~12 – 16 kDa as compared to the ~27 kDa of GFP proteins. Their smaller size makes applications of LOV proteins to be attractive where the size of fluorescent label is crucial. The short LOV proteins, PpSB2-LOV, PpSB1-LOV, YtvA-LOV and DsLOV, which were discussed in this work, were also used to develop so-called FMN-based fluorescent proteins (FbFPs) (Drepper *et al.*, 2007; Wingen *et al.*, 2014).

Currently, the wider application of LOV proteins as a fluorescent reporter is limited by the lack of broad choice of the excitation wavelengths. Until now, the excitation wavelength could be changed from the typical ~450 nm by only 10 nm through protein site-directed mutagenesis. For instance, the Q116V mutation of Pp2FbFp (PpSB2-LOV C53A mutant) or iLOV-Q489K (equivalent to the Q116 of Pp2FbFp) shifted the absorption maximum by 10 nm in blue region (Wingen *et al.*, 2014; Davari *et al.*, 2016). For a rational design and improvement of such LOV-based fluorescent reporters, the investigation of high-resolution structural information and photochemical/photophysical properties is necessary.

The crystal structure of W619_1-LOV reported here is the first structure of a LOV protein in the apo form, and opens the possibility to study the chromophore-protein interactions in detail. Two strategies have been generally applied to improve the optical properties of LOV-based fluorescent reporters: site-directed mutagenesis and the replacement of native chromophores with modified flavins (Mansurova *et al.*, 2011, 2013; Silva-Junior *et al.*, 2013; Wingen *et al.*, 2014; Davari *et al.*, 2016). The W619_1-LOV in the apo form was able to bind both the natural (FMN, FAD, RBF and LC) and modified (8-Cl-RF and 7-Br-RF) flavin chromophores in solution. Among them, 8-Cl-RF and LC showed remarkably

different photophysical and photochemical properties when bound to W619_1-LOV compared to the common flavins: FMN, RBF or FAD.

Lumichrome

Lumichrome comprises only the 7,8-Dimethylalloxazine ring system of common flavins (Figure 44). Absence of the ribityl chain results in a different distribution of electron clouds. This is directly reflected in the distinct UV-Vis spectrum of free LC, with two absorption peaks at 352 nm and 387 nm. The binding of LC to W619_1-LOV changed its spectrum, which then has one peak at 337 nm and one broad peak at 421 nm with two unresolved shoulders. Remarkably, the absorption maximum of bound LC shifted by 26 nm from 447 nm to 421 nm compared to the FMN-bound W619_1-LOV (Table 30). Moreover, the relative fluorescence quantum yield of bound LC was measured to be 0.4, which is rather high when compared to the typical value of 0.2 – 0.3 for other LOV proteins (Wingen *et al.*, 2014), yet lower than the fluorescence quantum yield of 0.51 measured in CreiLOV (Mukherjee *et al.*, 2015). The wavelength of maximum fluorescence of bound LC is equal to that of bound FMN, leading to a higher separation of the excitation and emission wavelengths that can be beneficial for application of LOV proteins as optical tool.

The lack of the ribityl chain in LC can be expected to result in a less tighter binding in the chromophore pocket when compared to RBF or FMN. This can be the reason for the observed broadening of the absorption and fluorescence peaks. Site-directed mutagenesis of the residues in the chromophore pocket, which would ensure the precise and tight positioning of LC, could improve its optical properties. The possible absence of the photocycle might be the consequence of lumichrome fluctuation inside the chromophore pocket, which prevents the formation of the covalent adduct. Alternatively, an intrinsically higher solvent access to the pocket, due to the absence of the ribityl chain in LC, facilitates scission of the formed adduct and results in the fast dark recovery on a time scale faster than the regular UV-Vis absorption spectroscopy can access. The use of time-resolved spectroscopy might clarify this question in the future.

8-Cl-RF

Free 8-Cl-RF in solution showed similar properties as its parental RBF, with a minor difference in the absorption maximum of 3 nm. Upon binding of 8-Cl-RF to W619_1-LOV, its spectrum first changed to the typical spectrum of the LOV proteins with bound FMN. However, overtime the spectra changed to a different one, with two unresolved peaks at 456 nm and 474 nm, of almost equal intensity and probably one extremely broad peak at ~420 nm which overlaps with them (Figure 47). A similar spectrum was observed for the YtvA-LOV protein bound to a modified flavin, with a bromine atom at the eighth position instead of a chlorine atom (8-Br-RF) (Mansurova *et al.*, 2013).

It was found that 8-Cl-RF upon binding to W619_1-LOV spontaneously forms a covalent adduct to the protein, which led to the observed spectral changes (Figure 47). The formation of the covalent bond most likely takes place between the conserved Cys53 residue and 8-Cl-RF. Experiments showed that 8-Cl-RF binds to free cysteine or β-mercaptoethanol molecules in solution (Figure 47). Previously, it was reported that the related flavin 8-Cl-FAD also spontaneously formed a covalent adduct to the flavoenzyme dihydrolipoamide dehydrogenase (DLD), isolated from pig heart (Moore, Cardemil and Massey, 1978). Although the crystal structures of pig heart DLD with both bound FAD or with 8-Cl-FAD are not available, it was confirmed from biochemical experiments that the covalent bond forms between the cysteine residues and the methyl group of FAD at the eighth position (C8a) (Moore, Cardemil and Massey, 1978).

For a long time, flavoenzymes were known to naturally bind FAD and in some cases FMN via covalent bonds (Singer *et al.*, 1956; Heuts *et al.*, 2009). They benefit from the covalent binding for several reasons: it increases the redox potential of flavin, prevents the dissociation of flavin from protein and increases protein stability. Natural covalent linkage to the 7,8-dimethylisoalloxazine ring system occurs at the eighth (C8a methyl group) and at the sixth (six member ring carbon C6) positions, refer to the atom numbering in Figure 44. Covalent bonds at C8a of flavin occurs with cysteine, histidine or tyrosine residues, whereas at C6 only with a cysteine residue (Heuts *et al.*, 2009). Covalent bond formation is believed to be auto-catalytic and so far no enzyme was identified that would catalyze such reaction.

The crystal structure of the human DLD protein with bound FAD, homologous to the above-mentioned DLD from pig hearth, has one cysteine in the 7 Å distance to the C8a methyl group of FAD and the active center with the disulfide bridge in the proximity of the FAD-C4a atom (Brautigam *et al.*, 2005). During catalytic reaction, this disulfide becomes reduced and one of the cysteine residues forms an adduct to the FAD-C4a atom, which resembles the formation of an adduct between the cysteine residues and flavin in LOV proteins. However, in order for 8-Cl-FAD to form a covalent adduct via Cl atom, the 8-Cl-FAD needs to be bound in a somewhat different position. Either cysteine, which is in the 7 Å distance, should come closer to the Cl atom of 8-Cl-FAD or the cysteine residues from the active center should come closer to the Cl atom. In the case of W619_1-LOV this also implies that 8-Cl-RF bind differently than, for instance, FMN, in order to have Cl atom of the 8-Cl-RF in the vicinity of Cys53 residue to enable the covalent adduct formation.

Further experiments must be done to elucidate the details of such adduct formation. For instance, mutation C53A, together with the substitution of the residues that are close to the C8a methyl group, to cysteine would help to clarify the binding configuration and may improve its fluorescence quantum yield. In fact, 8-Cl-RF has a low relative fluorescence quantum yield of 0.03 in the free state, which improves to 0.1 upon binding to W619_1-LOV. This suggest that the environment of the chromophore pocket can prevent the

quenching of the fluorescence even better, when 8-Cl-RF would bind in the similar position as natural flavins do, where almost the whole flavin molecule is buried within the chromophore pocket. This improvement of the fluorescence quantum yield together with the covalent attachment of the chromophore would make the W619_1-LOV with bound 8-Cl-RF very attractive for applications in harsh conditions.

The apparent absence of the photocycle with 8-Cl-RF bound to W619_1-LOV can be explained by a covalent binding of conserved cysteine to the flavin's C8a methyl group, which excludes the formation of an adduct to the C4a atom of flavin. Previously it was reported that the YtvA-LOV protein with bound 8-Br-RF, which has bromine at position eight instead of chlorine as in 8-Cl-RF, showed properties of the typical photocycle for LOV proteins with bound FMN (Mansurova et al., 2013). However, careful analysis of the absorption changes resulting from the illumination of YtvA-LOV with bound 8-Br-RF with blue-light does not seem to favor this hypothesis. First, the observed changes of absorption spectrum were small and did not resemble the typical light state spectrum of LOV domains. The measured recovery of the absorption might be related either to impurities, such as presence of bound RBF, or to some reversible/irreversible photoproducts. This is emphasized by the fact that the absorption spectrum did not revert back to the initial one. Indeed, it was reported that 8-Cl-FAD has a higher susceptibility to photodamage and the subsequent formation of photoproducts (Moore, Ghisla and Massey, 1979). Second, it can be argued that 8-Br-RF can also form covalent adduct to YtvA-LOV, taking into account the similarities between Br and Cl atoms and between the obtained bound-state spectra.

4.3 Primary light-induced changes

Several photochemical processes takes place upon the absorption of blue-light light by the chromophore bound to the LOV domain. Transitions between intermediate states often occur on a fast time scale, ranging from nanoseconds to seconds (Kottke *et al.*, 2003). Yet, an intermediate where an adduct between conserved cysteine residue and FMN was formed, which is referred as the light state, can be stabilized on a time scale ranging from seconds to several days (Jentzsch *et al.*, 2009; Tang *et al.*, 2010; Endres *et al.*, 2015). This allows LOV proteins with long adduct lifetimes to be trapped in the light state by crystallization under permanent blue-light illumination. These fully-adapted light state structures, in contrast to the illuminated structures obtained just by the exposure of dark grown crystals with light, can provide details on the light-induced changes without being restrained by the crystal lattice (Vaidya *et al.*, 2011; Circolone *et al.*, 2012).

Until recently, adduct formation in the light state has been believed to be the primary trigger of signal propagation in LOV proteins (Jones *et al.*, 2006). Adduct formation can either trigger signal propagation to the fused effector domain or trigger conformational changes, which results in binding to a partner protein. Mutants of some LOV-containing proteins that lack this conserved cysteine residue were shown to be inactive (Christie *et al.*,

2002; Matsuoka and Tokutomi, 2005) and it led to the assumption that all other LOV proteins are also deactivated. Nevertheless, it was found that the LOV-containing proteins, Vivid and YF1, which lack the conserved cysteine residue, were still able to generate a response upon blue-light illumination (Yee *et al.*, 2015). The signal propagation in these proteins originated from the flavin photo-reduction to the neutral semiquinone, which triggers changes in conformation of conserved glutamine residue. However, it should be mentioned that the generated response has a reduced magnitude. These results raises a question whether the formation of the adduct is the dominant factor, or if signal propagation can be triggered by other light-induced changes.

Three structures obtained in this work, two in the dark state and one in the light state can now show the details of primary changes induced by light. PpSB2-LOV, PpSB1-LOV, SBW25-LOV and Pf5-LOV proteins originate from bacteria of the same genus, *Pseudomonas*, and share about 60% of sequence identity (Table 2). Additionally, the YtvA-LOV protein from *Bacillus subtilis* was taken for comparison, because of its similarity in sequence and in the its dimer interface. Superposition of the PpSB2-LOV, PpSB1-LOV and Pf5-LOV dark state structures with the PpSB1-LOV and SBW25-LOV light state structures revealed that several structural changes take place around the chromophore and in the residues conserved between them (Figure 76).

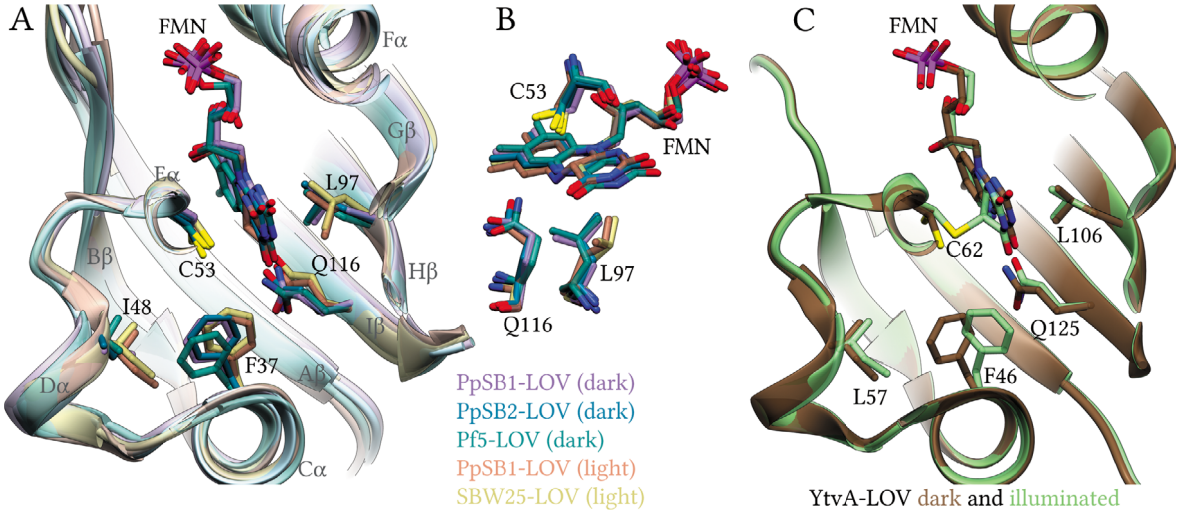


Figure 76. The primary and secondary light-driven changes in the short LOV. The structures are shown as ribbon diagrams. FMN and residues are shown as sticks. (**A**) Superposition of the light structures: SBW25-LOV (gold), PpSB1-LOV (coral) and the dark structures: PpSB1-LOV (plum), PpSB2-LOV (blue), Pf5-LOV (cyan) shows the primary light-induced changes in the vicinity of the chromophore. (**B**) The inset shows the two distinct conformations of Leu97 residue. (**C**) Superposition of the dark (brown) and illuminated structures (green) of YtvA-LOV.

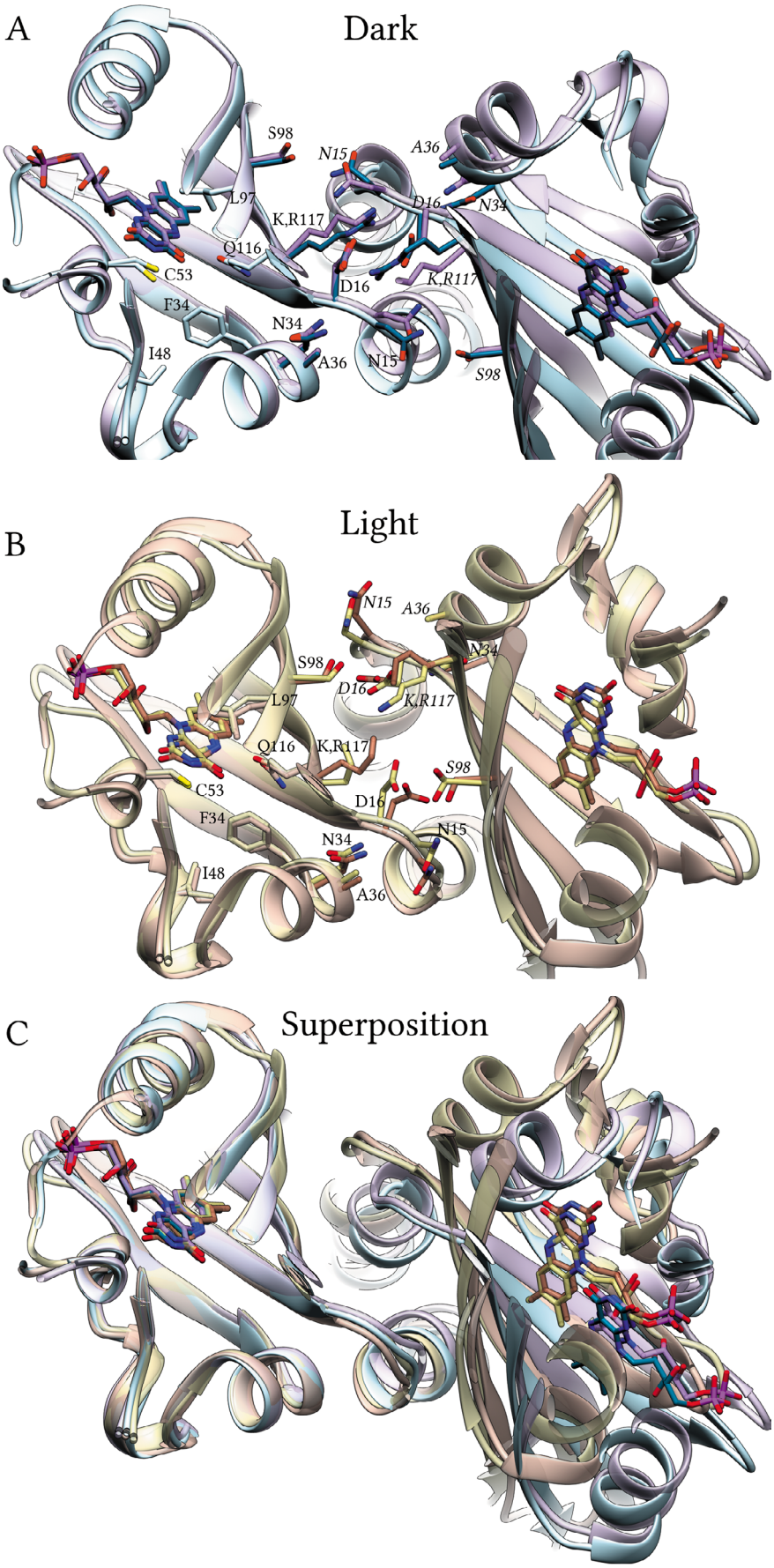


Figure 77. The primary and secondary light-driven changes in the short LOV proteins. The light structures: SBW25-LOV (gold), PpSB1-LOV (coral), and the dark structures: PpSB1-LOV (plum), PpSB2-LOV (blue) →

are shown as ribbon diagrams. FMN and residues are shown as stick models. Labels for the chain A residues are shown in regular font, whereas for the chain B residues in italic font. (**A, B**) Comparison of the residues affected by the primary light-induced changes shown in Figure 76A in the dark state (**A**) and the light state (**B**) structures. (**C**) Superposition of the dark and light state structures shows the rotation of the dimer interface. The structures were superposed by the core domain residues 17 – 117 of one chain.

The isoalloxazine tricyclic ring of the flavin chromophore has a planar structure in the dark state. The FMN-C4a atom position is stacked approximately between the conserved Cys53 and Leu97 residues. Upon blue-light illumination, a covalent adduct forms between the Cys53-Sγ and FMN-C4a atoms, where latter one becomes sp^3-hybridized. The sp^3-hybridization of FMN-C4a atom is reflected in the tetrahedral arrangement of the Cys53-Sγ atom and the C4, N5, C10a atoms of FMN around it. The Leu97 residue, positioned behind this tetrahedral arrangement, showed two distinct conformations in the dark and light state structures.

In the original structure of PpSB1-LOV in the light state, the conformation of Leu97 residue was slightly different than the one presented here. Due to the rather low resolution of the electron-density map, both conformations of Leu97 cannot be distinguished, where the dihedral angle χ_2 of the two Cγ methyl groups are rotated by 180°. Thus the conformation of the Leu97, resembling that in SBW25-LOV, could be well refined in the electron-density map of PpSB1-LOV, as shown in Figure 76B.

The distinct conformations of Leu97 residue can be related to the tetrahedral arrangement of adduct-forming atoms instead of planar one due to the presence of adduct, which leaves space for Leu97 residue to adopt a different side chain conformation. This change of conformation apparently leads to a change in the position of Ser98 and Ile99 residues through steric interactions. It can be speculated that this change enables rotation of the dimer interface as seen between dark and light states (Figure 77 and Figure 21). Indeed Ser98 residue participates in the dimer interface by forming a hydrogen bond to either the Asn15 or Asp16 residues of the other monomer.

Evidence of changes in the Leu97 conformation is well-supported by the solution NMR study on the PpSB1-LOV in the dark and light states (Röllen *et al.*, 2016). The chemical shift of Leu97 residue changed significantly upon illumination together with those of the adjacent Gln96 and Ser98 residues. Additionally, their S^2 values (refers to the square of the generalized model-free order parameter) also increased, which indicate a reduction of their dynamics in the light state.

The highly conserved Gln116 residue adopts two distinct side chain conformations between the dark and light states, as it was previously observed for other LOV protein (Crosson and Moffat, 2002; Zoltowski *et al.*, 2007; Vaidya *et al.*, 2011). This change in conformation is believed to be the result of FMN-N5 protonation in the light state. The protonation of FMN-N5 changes it from being a hydrogen bond acceptor to a donor. This in

turn results in the flipping of the Gln116 side chain atoms, Nε2 and Oε1, such that Oε1 atom becomes close to the FMN-N5 atom to form a hydrogen bond instead of the Nε2 amine group.

Such rearrangement results in the displacement of the Gln116 side chain that apparently triggers further structural changes through its steric interaction with Phe37 residue. As a consequence of that, the side chain of Phe37 adopts two distinct positions in the dark and light states. Movement of the Phe37, further on, results in the change of rotamer of Ile48 side chain, so that the Cδ methyl group moves away from the Phe37 residue. A solution NMR study on the PpSB1-LOV protein, also showed a large difference in the chemical shifts of both residues between the light and dark states (Röllen *et al.*, 2016).

A change in the Phe37 side chain position can propagate further structural changes such as the displacement of the Asn34 and Ala36 residues, which enables the Asn15 and Asp16 residues to accommodate a different position. The altered positions of the Asn15 and Asp16 residues, which participate in the dimer interface by interacting with the Ser98, Arg80 and Lys117 residues (Arg117 in PpSB2-LOV), might be directly involved in the observed dimer interface rotation between the dark and light states (Figure 77A,B).

The structural changes induced by Leu97 residue might be responsible for the signal propagation caused by adduct formation. Whereas a cascade of the above-mentioned changes, caused by Gln116 residue, could trigger signal propagation as seen in the Vivid and YF1 proteins that lacked a conserved cysteine residue (Yee *et al.*, 2015).

A comparison of the dark and illuminated structures of YtvA-LOV (LOV domain used in the YF1 protein) demonstrated similar changes in the Gln123, Phe46 and Ile57 residues (equivalent to above-mentioned Gln116, Phe37 and Ile48) (Figure 76C). The Leu106 residue (equivalent to Leu97) in the both structures did not adopt a distinct conformation, however, the mutation of leucine to phenylalanine (as in plant LOV domains) in the full-length YtvA protein (contains YtvA-LOV) impaired its function (Tang *et al.*, 2010).

The YtvA-LOV construct used for crystallization lacked the N-terminal A'α helix, which makes important contribution to the dimer interface. Its absence perhaps limited the light-induced changes that could be observed in the illuminated crystal structure of YtvA-LOV (Möglich and Moffat, 2007). However, the full-length protein YtvA and the chimeric protein YF1 contains A'α helix of the YtvA-LOV domain (Möglich, Ayers and Moffat, 2009a; Tang *et al.*, 2010) (Figure 4). The dark state structure of YF1 (Diensthuber *et al.*, 2013) showed that the A'α helix also participates in the dimer interface as that of other short LOV proteins studied here also do, thus a similar mechanism can be valid for the YtvA and YF1 proteins.

Since the LOV domains are a subset of the PAS superfamily, they may have inherited their general signaling mechanism. The PAS domains bind small ligands or use the heme

ligand to sense redox states (Taylor and Zhulin, 1999). The structural changes related to the binding of a ligand are mostly spread within a structurally conserved five antiparallel β-strands motif (Erbel *et al.*, 2003; Möglich, Ayers and Moffat, 2009b). These β-strands also participate in the dimerization of the PAS and LOV domains, or in the binding of the flanking helices such as the C-terminal Jα helices (Figure 1). In light of the structural data obtained in this study, such signaling mechanism seems to be valid for the short LOV proteins presented here. Moreover, there is also increasing experimental evidence that the β-sheet is a key transducer element of the signal propagation of the LOV proteins (Herrou and Crosson, 2011; Losi and Gärtner, 2011; Rinaldi *et al.*, 2012).

Other short LOV proteins do not have the combination of residues discussed here and might have different primary light-induced changes. The implication of the signal transduction from the LOV domain to the fused effector domain, via the Jα helices for instance, will be discussed next.

4.4 Light-induced structural changes and signal propagation in short LOV proteins

Many efforts have been put forth to understand how signals propagate from external stimuli, recognized by the LOV domain, to the fused effector domain. In this work, short LOV proteins were studied that lack the effector domain. In nature, most of the LOV domains are part of the multi-domain proteins, such as histidine or serine/threonine kinases, STAS sigma B factor, HTH DNA-binding domains etc. (Huala *et al.*, 1997; Losi *et al.*, 2002; Purcell *et al.*, 2010; Zoltowski, Motta-Mena and Gardner, 2013).

The LOV core domain is also frequently accompanied by N- or C-terminal helical extensions that are either packed against the core or extend from it (Harper, Neil and Gardner, 2003; Circolone *et al.*, 2012; Herman *et al.*, 2013; Endres *et al.*, 2015). Often the LOV domain is N-terminally fused to the effector via a helical linker, similarly to the Jα helix in short LOV proteins. Among LOV protein from different kingdoms, several signal transduction mechanisms that involve the Jα helix were identified. In AsLOV2, for instance, the Jα helix becomes displaced upon illumination thus allowing or preventing binding to the signaling partner protein (Harper, Neil and Gardner, 2003). In the LOV domain of the Aureochrome protein, the Jα helix unfolds upon illumination and facilitates the dimerization of LOV domains (Herman *et al.*, 2013; Hisatomi *et al.*, 2014).

Some bacterial LOV proteins, however, demonstrated different signaling mechanism, which is related to a different structural organization. For instance, the bacterial LOV domain YtvA-LOV, exists as a dimer both in the dark and light states. The Jα helices protrude away from the core domain and form the dimer interface (Figure 4). The short LOV proteins in this study, do this in a similar fashion, but the additional N-terminal A'α helix also participates in the dimer interface in both the dark and light states. From the crystal structure of PpSB1-

LOV in the dark and light states, a mechanism of signal transduction via the Jα helices was suggested, which involves the rotation of the dimer interface (Röllen *et al.*, 2016). The two additional structures in the dark states (PpSB2-LOV and Pf5-LOV) and one in the light state (SBW25-LOV) obtained in this work can now shed more light on the details of signal transduction.

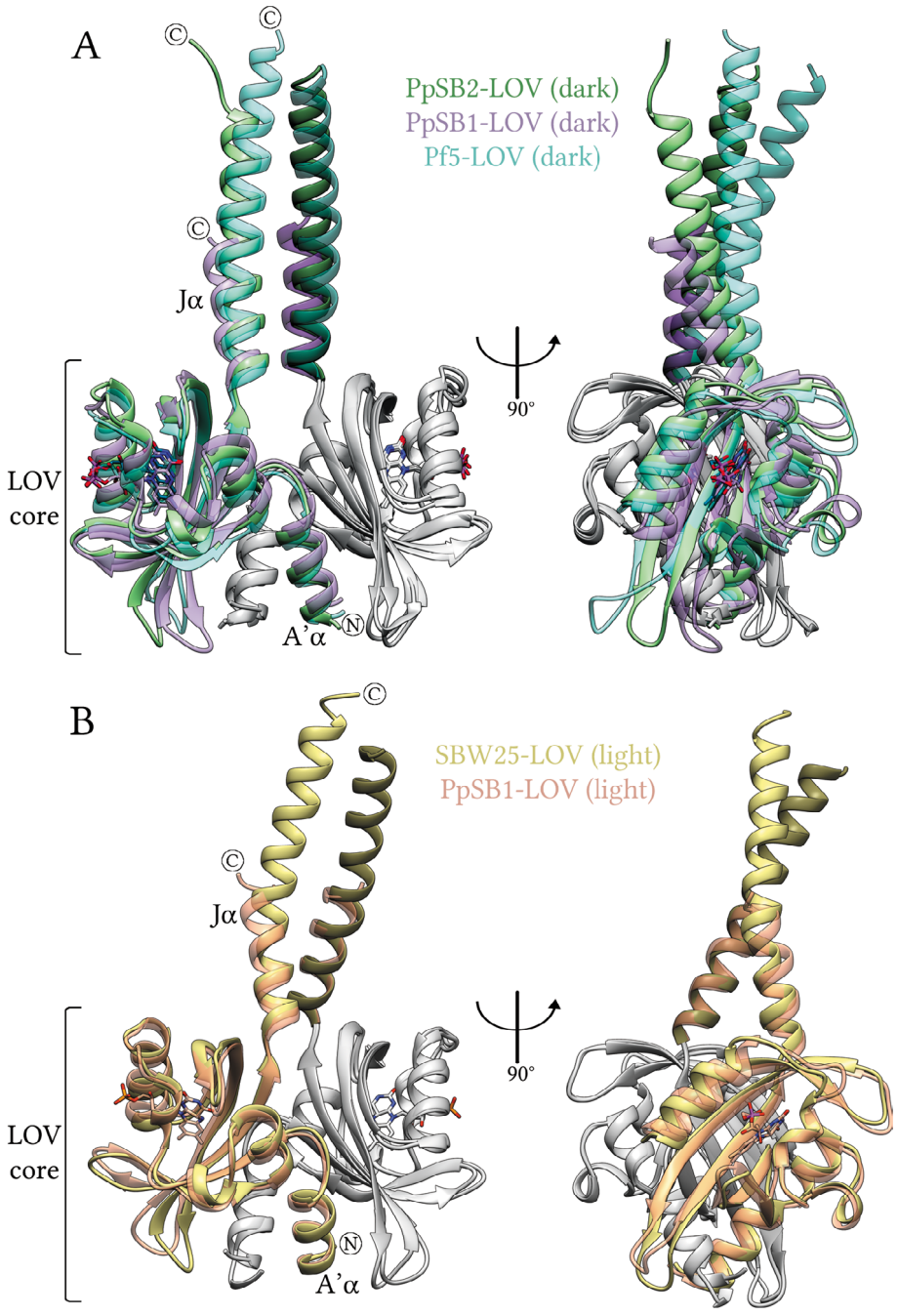


Figure 78. Superposition of the dark and light structures of the short LOV proteins discussed it this work. (**A**) The crystal structures in the dark state: PpSB1-LOV (plum), PpSB2-LOV (green), Pf5-LOV (cyan) and (**B**) those in the light state: SBW25-LOV (gold), PpSB1-LOV (coral) were superposed by the core domain residues 17 – 117 of chain B (gray for all) and are shown as ribbon diagrams. FMN is shown as a stick model.

A comparison between the dark and light structures of PpSB1-LOV and PpSB2-LOV revealed that the rotation of the dimer interface together with the change in the crossing angle of Jα helices is similar to that seen between the dark and light states of PpSB1-LOV (Figure 21). This resembles the motion of a piston connected by stiff rod to a crank, as shown in a simplified cartoon in Figure 21C. Superposition of the PpSB2-LOV structure in the dark state with the SBW25-LOV in the light state and Pf5-LOV in the dark state confirmed that such dimer interface rotation is common to each protein (Figure 78). However, the longer Jα helices of PpSB2-LOV, SBW25-LOV and Pf5-LOV in comparison to that of the PpSB1-LOV do not remain straight. While the first helical turns follow that of PpSB1-LOV, beyond that they bend and twist around each other, as a general consequence of their length and absence of any fused domain. Moreover, such an arrangement is perhaps necessary to cover newly-exposed hydrophobic residues of the Jα helices caused by the rotation of the core domains.

The Jα helices cross at a smaller angle in the dark state and at a larger angle in the light state. In general, the Jα helices in the both states do not have any strong interactions with the core domains, which would influence their crossing angle. Thus it can be concluded that observed change in the crossing angle of the Jα helices between the dark and light states is a secondary event resulting from the rotation of the core domains.

Moreover, a comparison of the Jα helices between the structures of Pf5-LOV in the dark state and SBW25-LOV in the light state, showed not only changes in their crossing angle but also a super-twisting of the helices. Such twisting leads to the displacement of one helix towards the C-terminus in the light state compared to the dark state. These changes might explain the observed structural changes in the LOV-containing proteins, YF1, SB2F1 and SB1F1, and will be discussed next.

4.5 LOV chimeric proteins as a model to study sensory histidine kinases

The SB2F1, SB1F1 and YF1 (Möglich, Ayers and Moffat, 2009a) proteins are engineered sensory histidine kinases (SHK) that were constructed by the substitution of the structurally homologous oxygen-sensitive PAS domain in FixL protein from *Bradyrhizobium japonicum* to the LOV domains from the PpSB2-LOV, PpSB1-LOV and YtvA-LOV proteins, respectively (Kraus, unpublished). Despite their engineered construction, they represent diverse, naturally occurring signal receptors proteins (Finn et al., 2006), including numerous PAS-linked (Möglich, Ayers and Moffat, 2009a) and LOV-linked proteins(Purcell et al., 2007; Swartz et al., 2007).

The crystal structures of SB2F1, YF1 and the entire effector domain of SHK HK853 (Marina, Waldburger and Hendrickson, 2005; Casino, Rubio and Marina, 2009) showed that the DHp and CA domains were connected by flexible loops. Such a flexible connection

enables them to adapt the different spatial orientations required for their function (Albanesi *et al.*, 2009). Arguably due to such flexibility, only a few full-length SHK proteins were structurally characterized, which contained both sensor and effector domains (Yamada *et al.*, 2009; Wang *et al.*, 2013; Rivera-Cancel *et al.*, 2014). A key feature among models for signaling in sensor histidine kinases is that they must form a dimer for their function, however, there is one report on SHK EL346 which is believed to function as a monomer (Rivera-Cancel *et al.*, 2014).

The Jα helices and helices of the DHp domain, function as so-called signaling helices, and enable mechanical propagation of the signal to the effector domains and represent a fairly versatile signal transduction mechanism (Anantharaman, Balaji and Aravind, 2006). The SB2F1 and YF1 structures, confirmed that the helices of the DHp domain that formed coiled coils were sufficient to transmit signal between the LOV sensor and CA effector domains, without need of their direct contact (Diensthuber *et al.*, 2013). However, the dark recovery kinetics of SB2F1 was affected by the presence of ATP. The influence of ATP seems to be related to the catalytic activity of the CA domains, which has few residues contact with the LOV domain in the proximity of FMN binding pocket.

Symmetric and asymmetric structures results from ADP and ATP binding

The crystal structures of SB2F1 and SB2F1 I66R mutant were obtained under dark and light conditions and in the presence of ATP, ADP and the non-hydrolysable/slowly-hydrolysable ATP analogs: AMP-PNP and ATP-γ-S. The overall arrangement of the domains in the structures of SB2F1 was found to be remarkably different to that of the previously determined YF1 structure (Diensthuber et al., 2013) (Figure 68). While SB2F1 also formed a dimer mostly with the long helices of the DHp domain, they adapted a more parallel arrangement in comparison to the strongly kinked DHp domain of YF1. Several crystal structures of the DHp domains alone or with the CA domain often showed an asymmetric arrangement; in contrast, NMR-based structures showed a parallel and elongated arrangement of the DHp domain helices (Bhate et al., 2015). Perhaps for a similar reason, the authors concluded that the kink in the helices of YF1 was induced by crystal packing. However, the SAXS data of SB2F1 in solution, discussed later, suggested that YF1 structure has a conformation that is close to the native one.

Two distinct dimers of SB2F1 were obtained by crystallization under dark conditions in the presence of ATP or ADP. The crystallization of SB2F1 in the presence of ATP produced crystals, which belonged to two related space groups $P3_121$ and $P3_221$, with respectively different unit-cells with long and short c-axis, while crystals obtained in the presence of ADP belonged to only one space group. The different volume of the unit-cells has the consequence that only one molecule per asymmetric unit is present in the short c-axis data, whereas in the long c-axis data there are two such molecules. The two molecules of the long c-axis structures form a homodimer with roughly two-fold symmetry (denoted as an

asymmetric-dimer), whereas in the short c-axis structures, a dimer was formed by application of the symmetry operation with the crystallographic two-fold axis (denoted as an symmetric-dimer).

The reason for the two distinct structures was found to be related to the bound nucleotides. The ATP-bound structures are all asymmetric-dimers whereas the ADP-bound structures are symmetric-dimers. The fact that the dimer becomes symmetrical in the ADP-bound state is perhaps related to the functional binding of the conjugated response regulator (RR). Indeed, in the crystal structure of histidine kinase (HK) complexed to the RR, the RR was found to bind symmetrically to both monomers (Casino, Rubio and Marina, 2009). This symmetrical arrangement could be necessary for proper recognition of SHK by the RR and the catalysis of the subsequent phosphotransfer reaction.

The asymmetric-dimer structure with bound ATP shows that the CA and DHp domains of one monomer accommodated different spatial arrangement than the other one. This can be physiologically relevant, as histidine kinases phosphorylate only one side at a time instead of both sides simultaneously (Bhate *et al.*, 2015).

Role of the LOV domain

The crystals of SB2F1 were grown under dark conditions and subsequently illuminated with blue-light to study the light-induced structural changes, as the preliminary data on the SB2F1 showed that it has different autophosphorylation activity depending on the dark or light state of the LOV domain (Kraus, unpublished). Additionally, the SB2F1 I66R mutant, which has a longer adduct lifetime, was also crystallized under the continuous blue-light illumination. Both of these structures were similar to the asymmetric-dimer, which was obtained under dark conditions, and showed further displacement of the CA and DHp domains and an increased asymmetry of the dimer.

The structures of short LOV proteins in this study, on the contrary, displayed a significant rotation of the dimer interface with a subsequent change of the Jα helices crossing angle, as a consequence of changes induced by light. However, comparing the relative orientation of the LOV core domains of the SB2F1 and YF1 (Diensthuber *et al.*, 2013) structures to that seen in the short LOV proteins in the dark and light states showed unexpected results (Figure 79 and Figure 80). The relative orientation of the LOV domains in the YF1 structure was close to that of the YtvA-LOV, which was used in the YF1 protein construction (Möglich and Moffat, 2007; Diensthuber *et al.*, 2013), as well as close to PpSB1-LOV, PpSB2-LOV and Pf5-LOV structures in the dark state (Figure 80). The long Jα helices of these proteins also followed the direction of the DHp helices in the YF1 structure. On the contrary, the relative orientation of the LOV domains in the SB2F1 asymmetric and symmetric structures did not match that of the PpSB2-LOV dimer structure in the dark state. The superposition by chain A showed that the chain A monomer, including the C-terminal Jα

helix of PpSB2-LOV, matches well the SB2F1 structures, while the chain B monomer is rotated and has a different position (Figure 79).

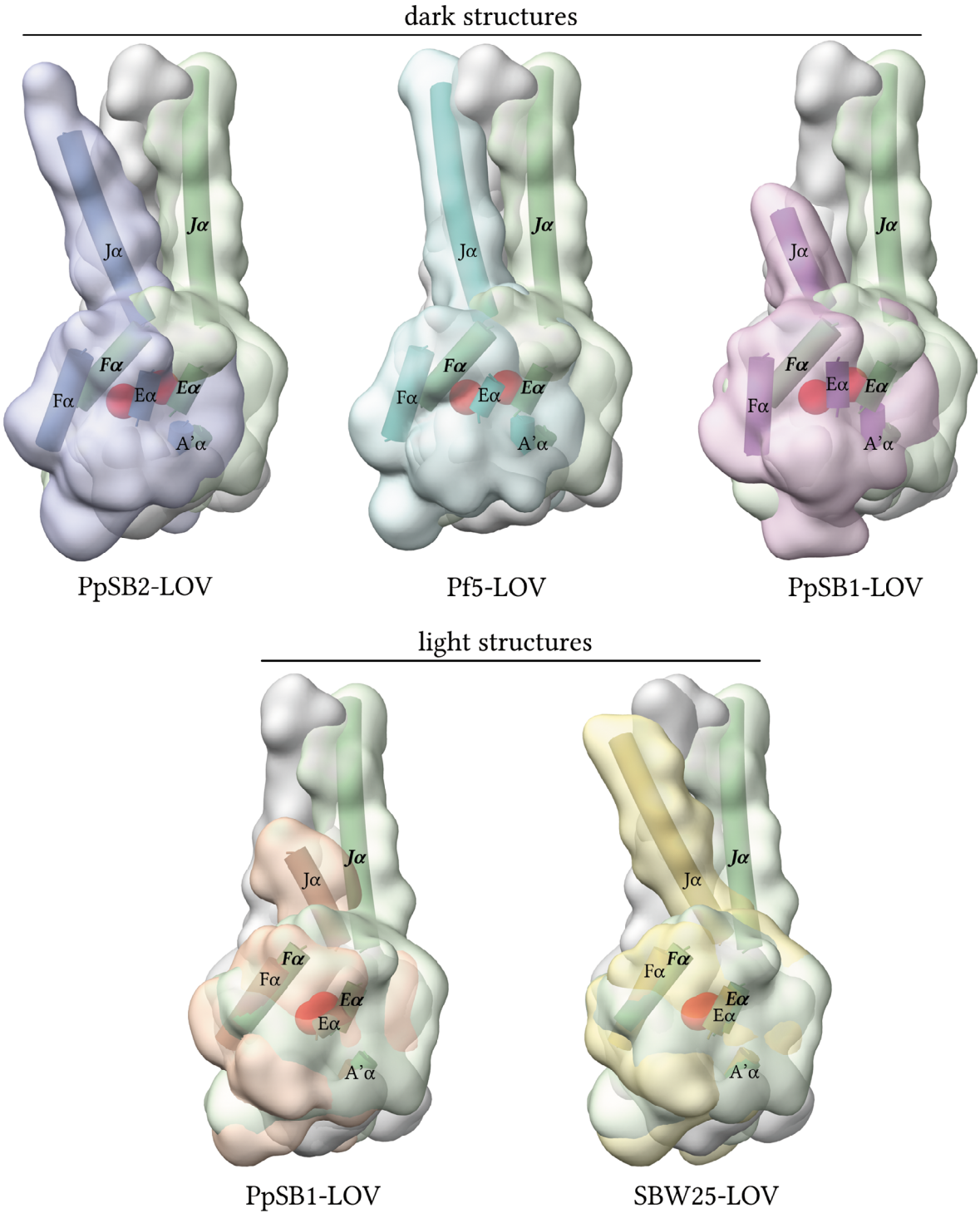


Figure 79. Superposition of SB2F1 with the short LOV protein structures. The superposition of the SB2F1 (residues 1 – 152) by residues range 17 – 138 of chain A (the LOV core domain plus the DHp domain up to His138) with LOV protein structures in both dark and light states. Chain B of the superposed structures: SB2F1 (green), PpSB2-LOV (blue), Pf5-LOV (cyan), PpSB1-LOV (dark state) (plum), PpSB1-LOV (light state) (coral) and SBW25-LOV (gold) are represented as transparent surfaces, whereas their chain A are represented as gray colored surfaces. Several structural elements of each domain are shown, where helices are represented as cylinders. The positions of FMN in each structure are represented by red spheres.

However, superposition of the SB2F1 with the light state structures of PpSB1-LOV and SBW25-LOV showed a good match (Figure 79). Why the dimer interface of the SB2F1 LOV domains is closer to that of the light state and not to that of the dark state is very controversial. The implication of this will be discussed further in light of the SAXS data obtained for SB2F1 in solution.

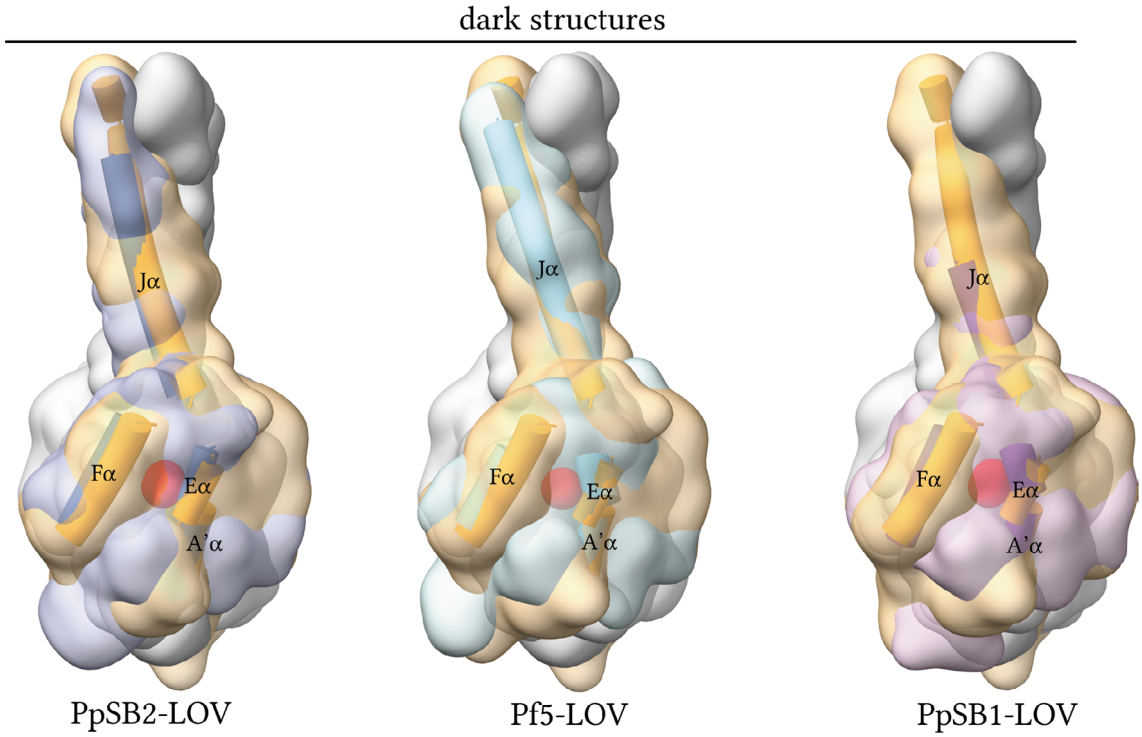


Figure 80. Superposition of YF1 with the short LOV protein structures. Superposition of the YF1 (residues 8 – 159) with the short LOV protein in the dark state. Chain B of the superposed structures: YF1 (PDB ID: 4GCZ, Diensthuber *et al.*, 2013) (orange), PpSB2-LOV (blue), Pf5-LOV (cyan), PpSB1-LOV (plum) are represented as transparent surfaces, whereas their chain A are represented as gray colored surfaces. Several structural elements of each domain are shown, where helices are represented as cylinders. The positions of FMN in each structure are represented by red spheres.

The SAXS data collected for the ATP-bound SB2F1 in the dark and light states, despite apparent inconsistency, showed the same agreement (Table 38). Comparison of the theoretical scattering calculated from the asymmetric-dimer of SB2F1 to the solution scattering curve of the ATP-bound SB2F1 in the light state showed better agreement than to the dark state. Whereas the theoretical scattering calculated from the YF1 (PDB ID: 4GCZ, Diensthuber *et al.*, 2013), on the contrary, showed better agreement to the scattering curve of the ATP-bound SB2F1 in the dark state (Table 38). Additionally, based on the SAXS scattering curves, *ab initio* envelope models were calculated and aligned with the crystal structures (Figure 70).

The envelope model of the ATP-bound SB2F1 in the light state agrees remarkably well in the shape and the spatial arrangement of the CA domains of the SB2F1 asymmetric-dimer. Whereas the structure of YF1 fits better to the envelope model that was calculated for

the ATP-bound SB2F1 in the dark state. This envelope model is even able to describe such a kinked structure of YF1 with asymmetric spatial positions of its CA domains.

In light of these findings, it can be argued that the SB2F1 structure, although obtained by crystallization under dark conditions, stayed trapped in the conformation typical for the light state structure, as it exists in solution. This also seems to be the reason for its high similarity to the SB2F1 illuminated structure and the structure of SB2F1 I66R in the light state. Several hypotheses can be proposed. First of all, despite the fact that they were crystallized under dark conditions, the protein samples were not handled in darkness prior to crystallization due to its short adduct lifetime. The measurement of the dark recovery kinetics, however, showed that the ATP-bound SB2F1 species has longer adduct lifetime (24 min vs 5 min) than the ADP-bound or nucleotide-free species. Moreover, the scission of the covalent adduct between the chromophore and the cysteine residue first of all results in the "dark state" of the chromophore, whereas the subsequent change of the protein conformation to the thermodynamically favorable one for the dark state may happen on a slower time scale. Such light state-like species might be then confined during the crystallization process by the crystal lattice.

The second explanation can be that the SB2F1 in solution undergoes a conformation exchange, where the light or dark states of the LOV domain thermodynamically favor one or the other conformation of the SB2F1. This may be the reason for the moderate sensitivity of SB2F1, which has the phosphorylation rates reduced approximately by half in the light state compared to the rate in the dark state (Krauss, unpublished). This implies that the light state of the LOV domain does not abolish the autophosphorylation but instead only reduces it.

Dark recovery kinetics of SB2F1 after incubation with ATP could be better fit by two exponential functions with τ_{rec1} and τ_{rec2}, where τ_{rec1} was similar to τ_{rec} of nucleotide-free state but τ_{rec2} was ~3 times longer than that (Table 34). The LOV and CA domains in the SB2F1 structures are situated such that the side chains of the CA domain residues can interact with the Fα helix of the LOV domain (Figure 67). The possibility that the observed dependence of the adduct lifetime of the LOV domain on the CA domain activity upon ATP binding is related to such interactions cannot be excluded. Especially keeping in mind that in the previously discussed short LOV proteins, the Fα helix was found to be important for the stabilization of the FMN binding pocket and thus mediates the dark recovery kinetics.

Spatial positions of the CA domains

The crystal structure of SB2F1 suggest that the CA domains are situated in the cis-autophosphorylation configuration with respect to the DHp domains, which enables the CA domain to autophosphorylate the histidine residue of the same monomer, such as in YF1, VicK, HK853 and other histidine kinases (Marina, Waldburger and Hendrickson, 2005; Wang *et al.*, 2013; Bhate *et al.*, 2015). One of the CA domains in the YF1 structure has

bound ADP, whereas the second one has no bound nucleotide. The phospho-accepting His161 residue in YF1 (His154 in SB2F1) has approximate distance to the bound ADP of 23 Å, while modeling of the ADP in the second CA domain, resulted in the distance of approximate 12.8 Å (Diensthuber *et al.*, 2013). The 12.8 Å distance is shorter than ~20 Å observed in the SB2F1. Assuming that the structures of SB2F1 represent the true light state structure as supported by the above-mentioned analysis, then the shorter distance in the YF1 structure might be related to its higher autophosphorylation rates. Whereas longer distance in the SB2F1 structure suggests that the autophosphorylation rates might be reduced in the light state. These hypotheses seem to agree with the SB2F1 preliminary data, which showed that SB2F1 has a higher autophosphorylation rate in the dark state than in the light state (Krauss, unpublished), however, still remain speculative.

Moreover, it can speculated that not only the longer distance to the conserved phosphate-accepting His154 residue in SB2F1 structures could explain the difference in the autophosphorylation rates, but this may also be due to the close proximity of His138 residue to the bound ATP/ADP in the SB2F1 structures. The His138 has only a ~6 Å distance to the ATP in the asymmetric-dimer, and indeed, such a short distance suggested that this residue might also be phosphorylated. However, phosphorylation of His138 cannot participate in a further relay of the phosphoryl group, as it would not be recognized by the conjugated response regulator protein, which usually binds in the range beyond this residue (Bhate *et al.*, 2015). Such unusual distortion of the phosphorelay signaling pathway can be the reason for the lower activity of the SB2F1 in the light state.

Signaling through change of the dimer interface

Following the assumption that the structures of SB2F1 represent the true light state structure, which was supported by the above-mentioned analysis, then the comparison of the YF1 (Diensthuber *et al.*, 2013) and SB2F1 structures may provide a model for the global structural changes induced by light. The rotation of the dimer interface observed between the dark and light states in the short LOV proteins, is similar to the differences in the orientation between the LOV domains in the YF1 and SB2F1 structures. Furthermore, the rotation of the dimer interface induced the displacement of the α1 helix of the DHp domain, which is connected to the LOV domain similarly as the Jα helix is connected in the short LOV proteins. The displacement of the α1 helix might cause the kinking of the DHp domain and results in reorientation of the CA domain.

It can be argued that although the details are not precise, such a model can generally explain the light-dependent phosphorylation rates in the YF1 and SB2F1 proteins. The model suggested here, however, is somewhat different in the details to the proposed one for YF1 (Möglich, Ayers and Moffat, 2009a; Diensthuber *et al.*, 2013). They proposed that the signal transduction from the LOV domain instead mainly propagates through torque motions of the helices and results in the super-twisting of the DHp domain helices. This model was

established by a comparison of the LOV domain dimer from the YF1 structure to the homology model of it in the light state, which was based on the structure of PpSB1-LOV in the light state (Circolone *et al.*, 2012). Nevertheless, their model also contains similar dimer interface rotation as seen in this study and could be the primary mechanism of the signal transduction.

Nucleotide-free state

The crystal structure of SB1F1 in the nucleotide-free state could not be solved yet, due to the insufficient experimental phasing. However, the SAXS data of the nucleotide-free SB2F1 and SB1F1 in the dark and light states was obtained. The scattering curves of SB2F1 and SB1F1 in the nucleotide-free states were virtually identical, suggesting a high structural similarity.

A comparison of the SAXS scattering curves of SB1F1 to the theoretical scattering of the YF1 and SB2F1 structures showed that they fitted worse to them than the SAXS scattering curves of the ATP-bound SB2F1 did. The calculated *ab initio* envelope models from the scattering curves of nucleotide-free SB1F1 showed consistently lower matching to the crystals structures of YF1 and SB2F1. Such different conformation of SB1F1 in the nucleotide-free state might explain the failure of the molecular replacement procedure to solve the SB1F1 structure by using the YF1 and SB2F1 structures as search templates.

4.6 Concluding remarks

The results presented in this thesis have contributed to understanding of several aspects of short LOV proteins as well as of engineered LOV histidine kinases. Determined crystal structure of short LOV proteins in the light and dark states enabled to identify the primary light induced changes, which propagate the signal from the chromophore to the periphery. These changes are likely to be the reason of observed dimer interface rotation between the dark and light states. Moreover, this rotation seems to be common for all studied short LOV protein from *P. putida* and *P. fluorescens*.

Analysis of the chromophore binding pockets together with the molecular dynamics simulations of short LOV proteins in the different states helped in the explanation of the strong variation of the adduct lifetimes among them. It was shown that the regions surrounding the chromophore are more flexible in the LOV proteins with short adduct lifetimes by enabling better solvent access to the chromophore, whereas LOV proteins with long adduct lifetimes showed reduced dynamics of these regions.

The first apo structure of the W619_1-LOV protein opened opportunities for a better understanding of the chromophore-to-protein interactions. Additionally, the apo form of W619_1-LOV could be successfully loaded with natural and modified flavin chromophores in solution, which showed remarkably different biophysical and biochemical properties. Among

them, two flavins showed the most intriguing properties: lumichrome, which absorption maximum became shifted by 26 nm in comparison to the FMN-bound one, together with fluorescence quantum yield of 0.4 and 8-chloro-riboflavin, which can auto-catalytically form a covalent bond to LOV protein.

The crystal structures of engineered LOV histidine kinases, SB2F1 and SB2F1 I66R, were determined in the dark and light states and with different nucleotides (ATP, ADP, AMP-PNP, ATP-γ-S). Additionally, small-angle X-ray scattering data provided structural information on the quaternary structure of the respective proteins in solution, in both the light and dark states. Analysis of all structural information led to the model of their light activation to be drawn. The proposed model is based on the observed structural changes induced by light in the short LOV proteins. This model can enrich understanding of light regulation of engineered LOV histidine kinases as well as be helpful in understanding of naturally-occurring sensory histidine kinases. However, analysis of the SB2F1 data presented here requires further support from additional experiments and deeper structural analysis. Furthermore, molecular dynamics simulations might provide valuable information on the possible tertiary structures which protein samples in the dark and light states.

5 Acknowledgment

I am especially thankful to my supervisors PD Dr. Renu Batra-Safferling and PD Dr. Joachim Granzin for giving me opportunity to work on interesting topics, their help and support during this PhD project.

I would like to thank Prof. Dr. Dieter Willbold for the possibility to work and accomplish my PhD project at the Institute of Complex Systems – Structural Biochemistry (ICS-6) at the Forschungszentrum Jülich.

I am also grateful to Prof. Dr. Karl-Erich Jäger and Dr. Ulrich Krauss from the Institute of Molecular Enzyme Technology in Heinrich-Heine University Düsseldorf and their collaborators who gave me an opportunity to work on exciting proteins presented in this thesis.

I am very thankful to Dr. Andreas Stadler who introduced me to the theory and practice of small-angle X-ray scattering.

I am deeply grateful to Dr. Michael Owen who greatly contributed in improvement of thesis text and supported me in the first steps of molecular dynamics simulations.

I would like to thank Dr. Max Michel, Anna Cousin and Dr. Katrin Röllen for introducing me to the molecular biological and biochemical aspect of laboratory work.

The X-ray diffraction experiments were performed on beamlines ID23-1, ID23-2, ID29, ID30B, ID30A-3 at the ESRF (Grenoble, France). I am grateful to local contact persons at the ESRF for providing assistance in using beamline. Additionally, I am thankful for support on ID29S beamline for microspectrometry measurements, especially Dr. David von Stetten.

Small-angle X-ray scattering experiments were performed on the beamline BM29 at the ESRF (Grenoble, France) and on the P12 beamline at PETRA III, DESY (Hamburg, Germany). I would like to specially thank Dr. Clément Blanchet and Dr. Martha Brennich for their great support at the beamlines and in the construction of blue-light laser or LED setup for measurements.

Jülich Supercomputing Centre at Forschungszentrum Jülich provided great access opportunity for molecular dynamics simulations and quantum chemical calculation. I want to thank all employees for their work and especially for their help in the software implementation in cluster environment.

6 References

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B. and Lindahl, E. (2015) 'GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers', *SoftwareX*, 1–2, pp. 19–25. doi: 10.1016/j.softx.2015.06.001.

Adams, P. D., Afonine, P. V, Bunkóczi, G., *et al.* (2010) 'PHENIX: a comprehensive Python-based system for macromolecular structure solution', *Acta Crystallographica Section D Biological Crystallography*. International Union of Crystallography, 66(2), pp. 213–221. doi: 10.1107/S0907444909052925.

Ahmad, M. and Cashmore, A. R. (1996) 'Seeing blue: the discovery of cryptochrome', *Plant molecular biology*, 30(5), pp. 851–861.

Albanesi, D., Martin, M., Trajtenberg, F., Mansilla, M. C., Haouz, A., Alzari, P. M., de Mendoza, D. and Buschiazzo, A. (2009) 'Structural plasticity and catalysis regulation of a thermosensor histidine kinase', *Proceedings of the National Academy of Sciences*, 106(38), pp. 16185–16190. doi: 10.1073/pnas.0906699106.

Alexandre, M. T. A., Arents, J. C., van Grondelle, R., Hellingwerf, K. J. and Kennis, J. T. M. (2007) 'A Base-Catalyzed Mechanism for Dark State Recovery in the Avena sativa Phototropin-1 LOV2 Domain', *Biochemistry*. American Chemical Society, 46(11), pp. 3129–3137. doi: 10.1021/bi062074e.

Alexandre, M. T. A., Domratcheva, T., Bonetti, C., van Wilderen, L. J. G. W., van Grondelle, R., Groot, M.-L., Hellingwerf, K. J. and Kennis, J. T. M. (2009) 'Primary Reactions of the LOV2 Domain of Phototropin Studied with Ultrafast Mid-Infrared Spectroscopy and Quantum Chemistry', *Biophysical Journal*, 97(1), pp. 227–237. doi: 10.1016/j.bpj.2009.01.066.

Aliverti, A., Curti, B. and Vanoni, M. A. (1999) 'Flavoprotein Protocols', in Chapman, S. K. and Reid, G. A. (eds). Totowa, NJ: Humana Press, pp. 9–23. doi: 10.1385/1-59259-266-X:9.

Anantharaman, V., Balaji, S. and Aravind, L. (2006) 'The signaling helix: a common functional theme in diverse signaling proteins', *Biology Direct*. BioMed Central, 1(1), p. 25. doi: 10.1186/1745-6150-1-25.

Artimo, P., Jonnalagedda, M., Arnold, K., *et al.* (2012) 'ExPASy: SIB bioinformatics resource portal', *Nucleic Acids Research*, 40(W1), pp. W597–W603. doi: 10.1093/nar/gks400.

Avila-Perez, M., Hellingwerf, K. J. and Kort, R. (2006) 'Blue Light Activates the B-Dependent Stress Response of Bacillus subtilis via YtvA', *Journal of Bacteriology*, 188(17), pp. 6411–6414. doi: 10.1128/JB.00716-06.

Ayers, R. A. and Moffat, K. (2008) 'Changes in Quaternary Structure in the Signaling Mechanisms of PAS Domains', *Biochemistry*. American Chemical Society, 47(46), pp. 12078–

12086. doi: 10.1021/bi801254c.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. and Leslie, A. G. W. (2011) 'iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM', *Acta Crystallographica Section D Biological Crystallography*. International Union of Crystallography, 67(4), pp. 271–281. doi: 10.1107/S0907444910048675.

Bauer, C., Rabl, C.-R., Heberle, J. and Kottke, T. (2011) 'Indication for a Radical Intermediate Preceding the Signaling State in the LOV Domain Photocycle', *Photochemistry and Photobiology*. Blackwell Publishing Ltd, 87(3), pp. 548–553. doi: 10.1111/j.1751-1097.2011.00901.x.

Beauchamp, K. A., Lin, Y.-S., Das, R. and Pande, V. S. (2012) 'Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements', *Journal of Chemical Theory and Computation*. American Chemical Society, 8(4), pp. 1409–1414. doi: 10.1021/ct2007814.

Bednarz, T., Losi, A., Gärtner, W., Hegemann, P. and Heberle, J. (2004) 'Functional variations among LOV domains as revealed by FT-IR difference spectroscopy', *Photochem. Photobiol. Sci.*, 3(6), pp. 575–579. doi: 10.1039/B400976B.

Berg, J., Hung, Y. P. and Yellen, G. (2009) 'A genetically encoded fluorescent reporter of ATP:ADP ratio', *Nature Methods*, 6(2), pp. 161–166. doi: 10.1038/nmeth.1288.

Bhate, M. P., Molnar, K. S., Goulian, M. and DeGrado, W. F. (2015) 'Signal Transduction in Histidine Kinases: Insights from New Structures', *Structure*, 23(6), pp. 981–994. doi: 10.1016/j.str.2015.04.002.

Bilwes, A. M., Alex, L. A., Crane, B. R. and Simon, M. I. (1999) 'Structure of CheA, a Signal-Transducing Histidine Kinase', *Cell*, 96(1), pp. 131–141. doi: 10.1016/S0092-8674(00)80966-6.

Blanchet, C. E., Spilotros, A., Schwemmer, F., *et al.* (2015) 'Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY)', *Journal of Applied Crystallography*, 48(2), pp. 431–443. doi: 10.1107/S160057671500254X.

Bolognesi, M., Ghisla, S. and Incoccia, L. (1978) 'The crystal and molecular structure of two models of catalytic flavo(co)enzyme intermediates', *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry*. International Union of Crystallography, 34(3), pp. 821–828. doi: 10.1107/S0567740878004124.

Bonomi, H. R., Posadas, D. M., Paris, G., Carrica, M. del C., Frederickson, M., Pietrasanta, L. I., Bogomolni, R. A., Zorreguieta, A. and Goldbaum, F. A. (2012) 'Light regulates attachment, exopolysaccharide production, and nodulation in Rhizobium leguminosarum through a LOV-histidine kinase photoreceptor', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 109(30), pp. 12135–12140. doi: 10.1073/pnas.1121292109.

Bourenkov, G. P. and Popov, A. N. (2010) 'Optimization of data collection taking radiation damage into account', *Acta Crystallographica Section D Biological Crystallography*, 66(4), pp. 409–419. doi: 10.1107/S0907444909054961.

Brautigam, C. A., Chuang, J. L., Tomchick, D. R., Machius, M. and Chuang, D. T. (2005) 'Crystal structure of human dihydrolipoamide dehydrogenase: NAD+/NADH binding and the structural basis of disease-causing mutations', *Journal of molecular biology*, 350(3), p. 543—552. doi: 10.1016/j.jmb.2005.05.014.

Briggs, W. R. (2007) 'The LOV domain: a chromophore module servicing multiple photoreceptors', *Journal of Biomedical Science*, 14(4), pp. 499–504. doi: 10.1007/s11373-007-9162-6.

Briggs, W. R. and Huala, E. (1999) 'Blue-Light Photoreceptors in Higher Plants', *Annual Review of Cell and Developmental Biology*, 15(1), pp. 33–62. doi: 10.1146/annurev.cellbio.15.1.33.

Buckley, A. M., Jukes, C., Candlish, D., *et al.* (2016) 'Lighting Up Clostridium Difficile: Reporting Gene Expression Using Fluorescent Lov Domains', *Scientific Reports*, 6, p. 23463. doi: 10.1038/srep23463.

Buckley, A. M., Petersen, J., Roe, A. J., Douce, G. R. and Christie, J. M. (2015) 'LOV-based reporters for fluorescence imaging', *Current opinion in chemical biology*, 27, pp. 39–45. doi: 10.1016/j.cbpa.2015.05.011.

Cao, Z., Buttani, V., Losi, A. and Gärtner, W. (2008) 'A Blue Light Inducible Two-Component Signal Transduction System in the Plant Pathogen Pseudomonas syringae pv. tomato', *Biophysical Journal*, 94(3), pp. 897–905. doi: 10.1529/biophysj.107.108977.

Case, D. A., Betz, R. M., Botello-Smith, W., *et al.* (2016) 'AMBER16'. University of California, San Francisco.

Casino, P., Rubio, V. and Marina, A. (2009) 'Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction', *Cell*, 139(2), pp. 325–36. doi: 10.1016/j.cell.2009.08.032.

Celikel, R., Veldore, V. H., Mathews, I., Devine, K. M. and Varughese, K. I. (2012) 'ATP forms a stable complex with the essential histidine kinase WalK (YycG) domain', *Acta Crystallographica Section D*, 68(7), pp. 839–845. doi: 10.1107/S090744491201373X.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. and Prasher, D. (1994) 'Green fluorescent protein as a marker for gene expression', *Science*, 263(5148), pp. 802–805. doi: 10.1126/science.8303295.

Chan, R. H. and Bogomolni, R. A. (2012) 'Structural Water Cluster As a Possible Proton Acceptor in the Adduct Decay Reaction of Oat Phototropin 1 LOV2 Domain', *The Journal of Physical Chemistry B*. American Chemical Society, 116(35), pp. 10609–10616. doi: 10.1021/jp304934t.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) 'MolProbity: all-atom structure validation for macromolecular crystallography', *Acta Crystallographica Section D Biological Crystallography*. International Union of Crystallography, 66(1), pp. 12–21. doi: 10.1107/S0907444909042073.

Christie, J. M. (2007) 'Phototropin Blue-Light Receptors', *Annual Review of Plant Biology*, 58(1), pp. 21–45. doi: 10.1146/annurev.arplant.58.032806.103951.

Christie, J. M., Corchnoy, S. B., Swartz, T. E., Hokenson, M., Han, I.-S., Briggs, W. R. and Bogomolni, R. A. (2007) 'Steric Interactions Stabilize the Signaling State of the LOV2 Domain of Phototropin 1', *Biochemistry*. American Chemical Society, 46(32), pp. 9310–9319. doi: 10.1021/bi700852w.

Christie, J. M., Swartz, T. E., Bogomolni, R. A. and Briggs, W. R. (2002) 'Phototropin LOV domains exhibit distinct roles in regulating photoreceptor function', *The Plant Journal*, 32(2), pp. 205–219. doi: 10.1046/j.1365-313X.2002.01415.x.

Chudakov, D. M., Lukyanov, S. and Lukyanov, K. A. (2005) 'Fluorescent proteins as a toolkit for in vivo imaging', *Trends in Biotechnology*. Elsevier, 23(12), pp. 605–613. doi: 10.1016/j.tibtech.2005.10.005.

Circolone, F., Granzin, J., Jentzsch, K., Drepper, T., Jaeger, K.-E., Willbold, D., Krauss, U. and Batra-Safferling, R. (2012) 'Structural basis for the slow dark recovery of a full-length LOV protein from Pseudomonas putida', *Journal of molecular biology*, 417(4), pp. 362–74. doi: 10.1016/j.jmb.2012.01.056.

Conrad, K. S., Bilwes, A. M. and Crane, B. R. (2013) 'Light-Induced Subunit Dissociation by a Light–Oxygen–Voltage Domain Photoreceptor from Rhodobacter sphaeroides', *Biochemistry*. American Chemical Society, 52(2), pp. 378–391. doi: 10.1021/bi3015373.

Corchnoy, S. B., Swartz, T. E., Lewis, J. W., Szundi, I., Briggs, W. R. and Bogomolni, R. A. (2003) 'Intramolecular proton transfers and structural changes during the photocycle of the LOV2 domain of phototropin 1', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 278(2), pp. 724–31. doi: 10.1074/jbc.M209119200.

Correa, F., Ko, W.-H., Ocasio, V., Bogomolni, R. A. and Gardner, K. H. (2013) 'Blue Light Regulated Two-Component Systems: Enzymatic and Functional Analyses of Light-Oxygen-Voltage (LOV)-Histidine Kinases and Downstream Response Regulators', *Biochemistry*, 52(27), pp. 4656–4666. doi: 10.1021/bi400617y.

Crosson, S. and Moffat, K. (2001) 'Structure of a flavin-binding plant photoreceptor domain: Insights into light-mediated signal transduction', *Proceedings of the National Academy of Sciences*, 98(6), pp. 2995–3000. doi: 10.1073/pnas.051520298.

Crosson, S. and Moffat, K. (2002) 'Photoexcited structure of a plant photoreceptor domain reveals a light-driven molecular switch', *The Plant cell*, 14(5), pp. 1067–1075. doi: 10.1105/tpc.010475.

Davari, M. D., Kopka, B., Wingen, M., Bocola, M., Drepper, T., Jaeger, K.-E., Schwaneberg, U. and Krauss, U. (2016) 'Photophysics of the LOV-Based Fluorescent Protein Variant iLOV-Q489K Determined by Simulation and Experiment', *The Journal of Physical Chemistry B*. American Chemical Society, 120(13), pp. 3344–3352. doi: 10.1021/acs.jpcb.6b01512.

Diensthuber, R. P., Bommer, M., Gleichmann, T. and Möglich, A. (2013) 'Full-Length Structure of a Sensor Histidine Kinase Pinpoints Coaxial Coiled Coils as Signal Transducers and Modulators', *Structure*, 21(7), pp. 1127–1136. doi: 10.1016/j.str.2013.04.024.

Djouani-Tahri, E.-B., Christie, J. M., Sanchez-Ferandin, S., Sanchez, F., Bouget, F.-Y. and Corellou, F. (2011) 'A eukaryotic LOV-histidine kinase with circadian clock function in the picoalga Ostreococcus', *The Plant Journal*, 65(4), pp. 578–588. doi: 10.1111/j.1365-313X.2010.04444.x.

Dorn, M. (2012) 'Structure, dynamics and interactions of domains of the bacterial photoreceptor YtvA from Bacillus subtilis', *Diss.Fu-Berlin.De*. Available at: http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000040363.

Dorn, M., Jurk, M., Wartenberg, A., Hahn, A. and Schmieder, P. (2013) 'LOV takes a pick: thermodynamic and structural aspects of the flavin-LOV-interaction of the blue-light sensitive photoreceptor YtvA from Bacillus subtilis', *PloS one*, 8(11), p. e81268. doi: 10.1371/journal.pone.0081268.

Doublié, S. (2007) 'Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems', *Methods in molecular biology*, 363, pp. 91–108. doi: 10.1007/978-1-59745-209-0_5.

Drepper, T., Eggert, T., Circolone, F., *et al.* (2007) 'Reporter proteins for in vivo fluorescence without oxygen', *Nature Biotechnology*, 25(4), pp. 443–445. doi: 10.1038/nbt1293.

Drössler, P., Holzer, W., Penzkofer, A. and Hegemann, P. (2002) 'pH dependence of the absorption and emission behaviour of riboflavin in aqueous solution', *Chemical Physics*, 282(3), pp. 429–439. doi: 10.1016/S0301-0104(02)00731-0.

Drössler, P., Holzer, W., Penzkofer, A. and Hegemann, P. (2003) 'Fluoresence quenching of riboflavin in aqueous solution by methionin and cystein', *Chemical Physics*, 286(2–3), pp. 409–420. doi: 10.1016/S0301-0104(02)00969-2.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. and Liang, J. (2006) 'CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues', *Nucleic Acids Research* , 34(suppl 2), pp. W116–W118. doi: 10.1093/nar/gkl282.

Emery, P., So, W. V., Kaneko, M., Hall, J. C. and Rosbash, M. (1998) 'CRY, a Drosophila Clock and Light-Regulated Cryptochrome, Is a Major Contributor to Circadian Rhythm Resetting and Photosensitivity', *Cell*. Elsevier, 95(5), pp. 669–679. doi: 10.1016/S0092-8674(00)81637-2.

Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010) 'Features and development of Coot', *Acta Crystallographica Section D Biological Crystallography*. International Union of Crystallography, 66(4), pp. 486–501. doi: 10.1107/S0907444910007493.

Endres, S., Granzin, J., Circolone, F., *et al.* (2015) 'Structure and function of a short LOV protein from the marine phototrophic bacterium Dinoroseobacter shibae', *BMC Microbiology*, 15(1), p. 30. doi: 10.1186/s12866-015-0365-0.

Engelhard, C., Raffelberg, S., Tang, Y., Diensthuber, R. P., Möglich, A., Losi, A., Gärtner, W. and Bittl, R. (2013) 'A structural model for the full-length blue light-sensing protein YtvA from Bacillus subtilis, based on EPR spectroscopy', *Photochemical & Photobiological Sciences*. The Royal Society of Chemistry, 12(10), p. 1855. doi:

10.1039/c3pp50128k.

Erbel, P. J. A., Card, P. B., Karakuzu, O., Bruick, R. K. and Gardner, K. H. (2003) 'Structural basis for PAS domain heterodimerization in the basic helix-loop-helix-PAS transcription factor hypoxia-inducible factor', *Proceedings of the National Academy of Sciences*, 100(26), pp. 15504–15509. doi: 10.1073/pnas.2533374100.

Espinosa-Urgel, M., Salido, A. and Ramos, J.-L. (2000) 'Genetic Analysis of Functions Involved in Adhesion of Pseudomonas putida to Seeds', *Journal of Bacteriology*, 182(9), pp. 2363–2369. doi: 10.1128/JB.182.9.2363-2369.2000.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. and Pedersen, L. G. (1995) 'A smooth particle mesh Ewald method', *The Journal of Chemical Physics*. AIP Publishing, 103(19), p. 8577. doi: 10.1063/1.470117.

Fedorov, R., Schlichting, I., Hartmann, E., Domratcheva, T., Fuhrmann, M. and Hegemann, P. (2003) 'Crystal Structures and Molecular Mechanism of a Light-Induced Signaling Switch: The Phot-LOV1 Domain from Chlamydomonas reinhardtii', *Biophysical Journal*, 84(4), pp. 2474–2482. doi: 10.1016/S0006-3495(03)75052-8.

Fenno, L., Yizhar, O. and Deisseroth, K. (2011) 'The Development and Application of Optogenetics', *Annual Review of Neuroscience*. Annual Reviews , 34(1), pp. 389–412. doi: 10.1146/annurev-neuro-061010-113817.

Ferguson, A. D., Sheth, P. R., Basso, A. D., Paliwal, S., Gray, K., Fischmann, T. O. and Le, H. V. (2011) 'Structural basis of CX-4945 binding to human protein kinase CK2', *FEBS Letters*, 585(1), pp. 104–110. doi: 10.1016/j.febslet.2010.11.019.

Galperin, M. Y., Nikolskaya, A. N. and Koonin, E. V. (2001) 'Novel domains of the prokaryotic two-component signal transduction systems', *FEMS Microbiology Letters*. Oxford University Press, 203(1), pp. 11–21. doi: 10.1111/j.1574-6968.2001.tb10814.x.

Gill, S. C. and von Hippel, P. H. (1989) 'Calculation of protein extinction coefficients from amino acid sequence data.', *Analytical biochemistry*, 182(2), pp. 319–26. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2610349.

Gilles-Gonzalez, M. A., Ditta, G. S. and Helinski, D. R. (1991) 'A haemoprotein with kinase activity encoded by the oxygen sensor of Rhizobium meliloti', *Nature*. Nature Publishing Group, 350(6314), pp. 170–172. doi: 10.1038/350170a0.

Glantz, S. T., Carpenter, E. J., Melkonian, M., Gardner, K. H., Boyden, E. S., Wong, G. K.-S. and Chow, B. Y. (2016) 'Functional and topological diversity of LOV domain photoreceptors', *Proceedings of the National Academy of Sciences*, 113(11), pp. E1442–E1451. doi: 10.1073/pnas.1509428113.

Gordon, M. S. and Schmidt, M. W. (2005) 'Chapter 41 – Advances in electronic structure theory: GAMESS a decade later', in *Theory and Applications of Computational Chemistry*, pp. 1167–1189. doi: 10.1016/B978-044451719-7/50084-6.

Grusch, M., Schelch, K., Riedler, R., Reichhart, E., Differ, C., Berger, W., Ingles-Prieto, A. and Janovjak, H. (2014) 'Spatio-temporally precise activation of engineered receptor tyrosine kinases by light', *The EMBO Journal*. EMBO Press, 33(15), pp. 1713–1726. doi: 10.15252/embj.201387695.

Guinier, A. (1939) 'La diffraction des rayons X aux tres petits angles; application a l'etude de phenomenes ultramicroscopiques', *Ann Phys (Paris)*, 12, pp. 161–237.

Halavaty, A. S. and Moffat, K. (2007) 'N- and C-Terminal Flanking Regions Modulate Light-Induced Signal Transduction in the LOV2 Domain of the Blue Light Sensor Phototropin 1 from Avena sativa', *Biochemistry*, 46(49), pp. 14001–14009. doi: 10.1021/bi701543e.

Harper, S. M., Neil, L. C. and Gardner, K. H. (2003) 'Structural basis of a phototropin light switch', *Science*, 301(5639), pp. 1541–1544. doi: 10.1126/science.1086810.

Heintzen, C., Loros, J. J. and Dunlap, J. C. (2001) 'The PAS Protein VIVID Defines a Clock-Associated Feedback Loop that Represses Light Input, Modulates Gating, and Regulates Clock Resetting', *Cell*, 104(3), pp. 453–464. doi: 10.1016/S0092-8674(01)00232-X.

Hendrischk, A.-K., Moldt, J., Frühwirth, S. W. and Klug, G. (2009) 'Characterization of an Unusual LOV Domain Protein in the α-Proteobacterium Rhodobacter sphaeroides', *Photochemistry and Photobiology*. Blackwell Publishing Ltd, 85(5), pp. 1254–1259. doi: 10.1111/j.1751-1097.2009.00554.x.

Herman, E., Sachse, M., Kroth, P. G. and Kottke, T. (2013) 'Blue-Light-Induced Unfolding of the Jα Helix Allows for the Dimerization of Aureochrome-LOV from the Diatom Phaeodactylum tricornutum', *Biochemistry*. American Chemical Society, 52(18), pp. 3094–3101. doi: 10.1021/bi400197u.

Herrou, J. and Crosson, S. (2011) 'Function, structure and mechanism of bacterial photosensory LOV proteins', *Nature Reviews Microbiology*, 9(10), pp. 713–723. doi: 10.1038/nrmicro2622.

Heuts, D. P. H. M., Scrutton, N. S., McIntire, W. S. and Fraaije, M. W. (2009) 'What's in a covalent bond?', *FEBS Journal*, 276(13), pp. 3405–3427. doi: 10.1111/j.1742-4658.2009.07053.x.

Hisatomi, O., Nakatani, Y., Takeuchi, K., Takahashi, F. and Kataoka, H. (2014) 'Blue Light-induced Dimerization of Monomeric Aureochrome-1 Enhances Its Affinity for the Target Sequence', *Journal of Biological Chemistry*, 289(25), pp. 17379–17391. doi: 10.1074/jbc.M114.554618.

Huala, E., Oeller, P. W., Liscum, E., Han, I. S., Larsen, E. and Briggs, W. R. (1997) 'Arabidopsis NPH1: a protein kinase with a putative redox-sensing domain', *Science*, 278(5346), pp. 2120–3. doi: 10.1126/science.278.5346.2120.

Huang, K., Merkle, T. and Beck, C. F. (2002) 'Isolation and characterization of a Chlamydomonas gene that encodes a putative blue-light photoreceptor of the phototropin family', *Physiologia plantarum*, 115(4), pp. 613–622.

Huang, Z. J., Edery, I. and Rosbash, M. (1993) 'PAS is a dimerization domain common to Drosophila period and several transcription factors', *Nature*, 364(6434), pp. 259–262. doi: 10.1038/364259a0.

Inouye, M. and Dutta, R. (2002) *Histidine kinases in signal transduction*. Edited by M. Inouye and R. Dutta. San Diego, USA: Academic Press.

Jentzsch, K., Wirtz, A., Circolone, F., Drepper, T., Losi, A., Gärtner, W., Jaeger, K.-E. and Krauss, U. (2009) 'Mutual exchange of kinetic properties by extended mutagenesis in two short LOV domain proteins from Pseudomonas putida', *Biochemistry*, 48(43), pp. 10321–33. doi: 10.1021/bi901115z.

Jones, M. A., Feeney, K. A., Kelly, S. M. and Christie, J. M. (2006) 'Mutational Analysis of Phototropin 1 Provides Insights into the Mechanism Underlying LOV2 Signal Transmission', *Journal of Biological Chemistry*, 282(9), pp. 6405–6414. doi: 10.1074/jbc.M605969200.

Jurk, M., Dorn, M., Kikhney, A., Svergun, D., Gärtner, W. and Schmieder, P. (2010) 'The Switch that Does Not Flip: The Blue-Light Receptor YtvA from Bacillus subtilis Adopts an Elongated Dimer Conformation Independent of the Activation State as Revealed by a Combined AUC and SAXS Study', *Journal of Molecular Biology*, 403(1), pp. 78–87. doi: 10.1016/j.jmb.2010.08.036.

Kabsch, W. (2010) 'XDS', *Acta Crystallographica Section D Biological Crystallography*. International Union of Crystallography, 66(2), pp. 125–132. doi: 10.1107/S0907444909047337.

Kabsch, W. and Sander, C. (1983) 'Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features', *Biopolymers*, 22(12), pp. 2577–637. doi: 10.1002/bip.360221211.

Kaserer, A. O. and West, A. H. (2010) 'Histidine Kinases in Two-Component Signaling Pathways', in *Handbook of Cell Signaling*. Elsevier, pp. 581–586. doi: 10.1016/B978-0-12-374145-5.00078-4.

Kay, C. W. M., Schleicher, E., Kuppig, A., *et al.* (2003) 'Blue light perception in plants. Detection and characterization of a light-induced neutral flavin radical in a C450A mutant of phototropin', *The Journal of biological chemistry*, 278(13), pp. 10973–82. doi: 10.1074/jbc.M205509200.

Kennis, J. T. M. and Alexandre, M. T. A. (2006) 'Chapter 12. Mechanisms of Light Activation in Flavin-Binding Photoreceptors', in *Flavins: Photochemistry and Photobiology*. The Royal Society of Chemistry, pp. 287–320. doi: 10.1039/9781847555397-00287.

Kennis, J. T. M., Crosson, S., Gauden, M., van Stokkum, I. H. M., Moffat, K. and van Grondelle, R. (2003) 'Primary Reactions of the LOV2 Domain of Phototropin, a Plant Blue-Light Photoreceptor', *Biochemistry*. American Chemical Society, 42(12), pp. 3385–3392. doi: 10.1021/bi034022k.

Klumpp, S. and Krieglstein, J. (2002) 'Phosphorylation and dephosphorylation of histidine residues in proteins', *European Journal of Biochemistry*, 269(4), pp. 1067–1071. doi: 10.1046/j.1432-1033.2002.02755.x.

Kottke, T., Heberle, J., Hehn, D., Dick, B. and Hegemann, P. (2003) 'Phot-LOV1: Photocycle of a Blue-Light Receptor Domain from the Green Alga Chlamydomonas reinhardtii', *Biophysical Journal*, 84(2), pp. 1192–1201. doi: 10.1016/S0006-3495(03)74933-9.

Kozioł, J. (1971) 'Fluorometric analyses of riboflavin and its coenzymes', *Methods in*

Enzymology, 18, pp. 253–285. doi: 10.1016/S0076-6879(71)18089-5.

Kraiselburd, I., Alet, A. I., Tondo, M. L., Petrocelli, S., Daurelio, L. D., Monzón, J., Ruiz, O. A., Losi, A. and Orellano, E. G. (2012) 'A LOV Protein Modulates the Physiological Attributes of Xanthomonas axonopodis pv. citri Relevant for Host Plant Colonization', *PLOS ONE*. Public Library of Science, 7(6), pp. 1–18. doi: 10.1371/journal.pone.0038226.

Krauss, U., Losi, A., Gärtner, W., Jaeger, K.-E. and Eggert, T. (2005) 'Initial characterization of a blue-light sensing, phototropin-related protein from Pseudomonas putida: a paradigm for an extended LOV construct', *Physical Chemistry Chemical Physics*. The Royal Society of Chemistry, 7(14), p. 2804. doi: 10.1039/b504554a.

Krissinel, E. and Henrick, K. (2007) 'Inference of Macromolecular Assemblies from Crystalline State', *Journal of Molecular Biology*, 372(3), pp. 774–797. doi: 10.1016/j.jmb.2007.05.022.

Kumar, P. and Bansal, M. (2012) 'HELANAL-Plus: a web server for analysis of helix geometry in protein structures', *Journal of Biomolecular Structure and Dynamics*. Taylor & Francis, 30(6), pp. 773–783. doi: 10.1080/07391102.2012.689705.

Laemmli, U. K. (1970) 'Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4', *Nature*, 227(5259), pp. 680–685. doi: 10.1038/227680a0.

Lindorff-Larsen, K., Maragakis, P., Piana, S., Eastwood, M. P., Dror, R. O. and Shaw, D. E. (2012) 'Systematic Validation of Protein Force Fields against Experimental Data', *PLoS ONE*. Edited by D. J. Muller. Public Library of Science, 7(2), p. e32131. doi: 10.1371/journal.pone.0032131.

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O. and Shaw, D. E. (2010) 'Improved side-chain torsion potentials for the Amber ff99SB protein force field', *Proteins*. Wiley-Blackwell, 78(8), pp. 1950–8. doi: 10.1002/prot.22711.

Losi, A. and Gärtner, W. (2008) 'Bacterial bilin- and flavin-binding photoreceptors', *Photochemical & Photobiological Sciences*. The Royal Society of Chemistry, 7(10), p. 1168. doi: 10.1039/b802472c.

Losi, A. and Gärtner, W. (2011) 'Old Chromophores, New Photoactivation Paradigms, Trendy Applications: Flavins in Blue Light-Sensing Photoreceptors', *Photochemistry and Photobiology*, 87(3), pp. 491–510. doi: 10.1111/j.1751-1097.2011.00913.x.

Losi, A., Polverini, E., Quest, B. and Gärtner, W. (2002) 'First Evidence for Phototropin-Related Blue-Light Receptors in Prokaryotes', *Biophysical Journal*. Elsevier, 82(5), pp. 2627–2634. doi: 10.1016/S0006-3495(02)75604-X.

Mansurova, M., Scheercousse, P., Simon, J., Kluth, M. and Gärtner, W. (2011) 'Chromophore Exchange in the Blue Light-Sensitive Photoreceptor YtvA from Bacillus subtilis', *ChemBioChem*, 12(4), pp. 641–646. doi: 10.1002/cbic.201000515.

Mansurova, M., Simon, J., Salzmann, S., Marian, C. M. and Gärtner, W. (2013) 'Spectroscopic and Theoretical Study on Electronically Modified Chromophores in LOV Domains: 8-Bromo- and 8-Trifluoromethyl-Substituted Flavins', *ChemBioChem*, 14(5), pp. 645–654. doi: 10.1002/cbic.201200670.

Marina, A., Waldburger, C. D. and Hendrickson, W. A. (2005) 'Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein', *The EMBO journal*, 24(24), pp. 4247–59. doi: 10.1038/sj.emboj.7600886.

Masuda, S. (2013) 'Light Detection and Signal Transduction in the BLUF Photoreceptors', *Plant and Cell Physiology*, 54(2), pp. 171–179. doi: 10.1093/pcp/pcs173.

Matsuoka, D. and Tokutomi, S. (2005) 'Blue light-regulated molecular switch of Ser/Thr kinase in phototropin', *Proceedings of the National Academy of Sciences of the United States of America*, 102(37), pp. 13337–13342. doi: 10.1073/pnas.0506402102.

Matz, M. V, Lukyanov, K. A. and Lukyanov, S. A. (2002) 'Family of the green fluorescent protein: Journey to the end of the rainbow', *BioEssays*. Wiley Subscription Services, Inc., A Wiley Company, 24(10), pp. 953–959. doi: 10.1002/bies.10154.

Mavrodi, D. V, Paulsen, I. T., Ren, Q. and Loper, J. E. (2007) 'Genomics of Pseudomonas fluorescens Pf-5', in Ramos, J.-L. and Filloux, A. (eds) *Pseudomonas: A Model System in Biology*. Dordrecht: Springer Netherlands, pp. 3–30. doi: 10.1007/978-1-4020-6097-7_1.

Metz, S., Jager, A. and Klug, G. (2012) 'Role of a short light, oxygen, voltage (LOV) domain protein in blue light- and singlet oxygen-dependent gene regulation in Rhodobacter sphaeroides', *Microbiology*, 158(2), pp. 368–379. doi: 10.1099/mic.0.054700-0.

Miller, J. M. and Lucius, A. L. (2014) 'ATP-γ-S competes with ATP for binding at Domain 1 but not Domain 2 during ClpA catalyzed polypeptide translocation', *Biophysical Chemistry*, 185, pp. 58–69. doi: 10.1016/j.bpc.2013.11.002.

Mishin, A. S., Belousov, V. V, Solntsev, K. M. and Lukyanov, K. A. (2015) 'Novel uses of fluorescent proteins', *Current Opinion in Chemical Biology*, 27, pp. 1–9. doi: 10.1016/j.cbpa.2015.05.002.

Möglich, A., Ayers, R. and Moffat, K. (2009a) 'Design and signaling mechanism of light-regulated histidine kinases', *Journal of molecular biology*, 385(5), pp. 1433–44. doi: 10.1016/j.jmb.2008.12.017.

Möglich, A., Ayers, R. and Moffat, K. (2009b) 'Structure and Signaling Mechanism of Per-ARNT-Sim Domains', *Structure*, 17(10), pp. 1282–1294. doi: 10.1016/j.str.2009.08.011.

Möglich, A. and Moffat, K. (2007) 'Structural basis for light-dependent signaling in the dimeric LOV domain of the photosensor YtvA', *Journal of molecular biology*, 373(1), pp. 112–26. doi: 10.1016/j.jmb.2007.07.039.

Möglich, A., Yang, X., Ayers, R. A. and Moffat, K. (2010) 'Structure and Function of Plant Photoreceptors', *Annual Review of Plant Biology*, 61(1), pp. 21–47. doi: 10.1146/annurev-arplant-042809-112259.

Moore, E. G., Cardemil, E. and Massey, V. (1978) 'Production of a covalent flavin linkage in lipoamide dehydrogenase. Reaction with 8-Cl-FAD', *J. Biol. Chem.*, 253(18), pp. 6413–6422. Available at: http://www.jbc.org/cgi/content/long/253/18/6413.

Moore, E. G., Ghisla, S. and Massey, V. (1979) 'Properties of flavins where the 8-methyl group is replaced by mercapto- residues', *Journal of Biological Chemistry*, 254(17), pp.

8173–8178. Available at: http://www.jbc.org/content/254/17/8173.abstract.

Moriconi, V., Sellaro, R., Ayub, N., Soto, G., Rugnone, M., Shah, R., P. Pathak, G., Gärtner, W. and Casal, J. J. (2013) 'LOV-domain photoreceptor, encoded in a genomic island, attenuates the virulence of Pseudomonas syringae in light-exposed Arabidopsis leaves', *The Plant Journal*, 76(2), pp. 322–331. doi: 10.1111/tpj.12289.

Mukherjee, A. and Schroeder, C. M. (2015) 'Flavin-based fluorescent proteins: emerging paradigms in biological imaging', *Current Opinion in Biotechnology*, 31, pp. 16–23. doi: 10.1016/j.copbio.2014.07.010.

Mukherjee, A., Weyant, K. B., Agrawal, U., Walker, J., Cann, I. K. O. and Schroeder, C. M. (2015) 'Engineering and Characterization of New LOV-Based Fluorescent Proteins from Chlamydomonas reinhardtii and Vaucheria frigida', *ACS Synthetic Biology*. American Chemical Society, 4(4), pp. 371–377. doi: 10.1021/sb500237x.

Nakatani, Y. and Hisatomi, O. (2015) 'Molecular Mechanism of Photozipper, a Light-Regulated Dimerizing Module Consisting of the bZIP and LOV Domains of Aureochrome-1', *Biochemistry*. American Chemical Society, 54(21), pp. 3302–3313. doi: 10.1021/acs.biochem.5b00320.

Ohlendorf, R., Vidavski, R. R., Eldar, A., Moffat, K. and Möglich, A. (2012) 'From Dusk till Dawn: One-Plasmid Systems for Light-Regulated Gene Expression', *Journal of Molecular Biology*, 416(4), pp. 534–542. doi: 10.1016/j.jmb.2012.01.001.

Olesen, C., Picard, M., Winther, A.-M. M. L., *et al.* (2007) 'The structural basis of calcium transport by the calcium pump', *Nature*, 450(7172), pp. 1036–1042. doi: 10.1038/nature06418.

Otwinowski, Z. and Minor, W. (1997) 'Processing of X-ray diffraction data collected in oscillation mode', in *Methods in Enzymology*, pp. 307–326. doi: 10.1016/S0076-6879(97)76066-X.

Parkinson, J. (1992) 'Communication Modules in Bacterial Signaling Proteins', *Annual Review of Genetics*, 26(1), pp. 71–112. doi: 10.1146/annurev.genet.26.1.71.

Parkinson, J. S. (2010) 'Signaling Mechanisms of HAMP Domains in Chemoreceptors and Sensor Kinases', *Annual Review of Microbiology*. Annual Reviews , 64(1), pp. 101–122. doi: 10.1146/annurev.micro.112408.134215.

Paulsen, I. T., Press, C. M., Ravel, J., *et al.* (2005) 'Complete genome sequence of the plant commensal Pseudomonas fluorescens Pf-5', *Nat Biotech*, 23(7), pp. 873–878. doi: 10.0.4.14/nbt1110.

Pernot, P., Round, A., Barrett, R., *et al.* (2013) 'Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution', *Journal of synchrotron radiation*. International Union of Crystallography, 20(Pt 4), pp. 660–4. doi: 10.1107/S0909049513010431.

Petoukhov, M. V., Franke, D., Shkumatov, A. V., *et al.* (2012) 'New developments in the ATSAS program package for small-angle scattering data analysis', *Journal of Applied Crystallography*. International Union of Crystallography, 45(2), pp. 342–350. doi: 10.1107/S0021889812007662.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. (2004) 'UCSF Chimera - a visualization system for exploratory research and analysis', *Journal of computational chemistry*, 25(13), pp. 1605–12. doi: 10.1002/jcc.20084.

Pfeifer, A., Majerus, T., Zikihara, K., Matsuoka, D., Tokutomi, S., Heberle, J. and Kottke, T. (2009) 'Time-Resolved Fourier Transform Infrared Study on Photoadduct Formation and Secondary Structural Changes within the Phototropin LOV Domain', *Biophysical Journal*, 96(4), pp. 1462–1470. doi: 10.1016/j.bpj.2008.11.016.

Pfeifer, A., Mathes, T., Lu, Y., Hegemann, P. and Kottke, T. (2010) 'Blue light induces global and localized conformational changes in the kinase domain of full-length phototropin', *Biochemistry*, 49(5), pp. 1024–1032. doi: 10.1021/bi9016044.

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', *Molecular and Cellular Biology* , 18(7), pp. 4079–4088. doi: 10.1128/MCB.18.7.4079.

Potzkei, J., Kunze, M., Drepper, T., Gensch, T., Jaeger, K.-E. and Buechs, J. (2012) 'Real-time determination of intracellular oxygen in bacteria using a genetically encoded FRET-based biosensor', *BMC Biology*, 10(1), p. 28. doi: 10.1186/1741-7007-10-28.

Pudasaini, A., El-Arab, K. K. and Zoltowski, B. D. (2015) 'LOV-based optogenetic devices: light-driven modules to impart photoregulated control of cellular signaling', *Frontiers in Molecular Biosciences*. Frontiers, 2, p. 18. doi: 10.3389/fmolb.2015.00018.

Purcell, E. B., McDonald, C. A., Palfey, B. A. and Crosson, S. (2010) 'An Analysis of the Solution Structure and Signaling Mechanism of LovK, a Sensor Histidine Kinase Integrating Light and Redox Signals', *Biochemistry*, 49(31), pp. 6761–6770. doi: 10.1021/bi1006404.

Purcell, E. B., Siegal-Gaskins, D., Rawling, D. C., Fiebig, A. and Crosson, S. (2007) 'A photosensory two-component system regulates bacterial cell attachment', *Proceedings of the National Academy of Sciences*, 104(46), pp. 18241–18246. doi: 10.1073/pnas.0705887104.

Rani, R., Jentzsch, K., Lecher, J., Hartmann, R., Willbold, D., Jaeger, K.-E. and Krauss, U. (2013) 'Conservation of Dark Recovery Kinetic Parameters and Structural Features in the Pseudomonadaceae "Short" Light, Oxygen, Voltage (LOV) Protein Family: Implications for the Design of LOV-Based Optogenetic Tools', *Biochemistry*, 52(26), pp. 4460–4473. doi: 10.1021/bi400311r.

Rinaldi, J., Gallo, M., Klinke, S., Paris, G., Bonomi, H. R., Bogomolni, R. A., Cicero, D. O. and Goldbaum, F. A. (2012) 'The β-Scaffold of the LOV Domain of the Brucella Light-Activated Histidine Kinase Is a Key Element for Signal Transduction', *Journal of Molecular Biology*, 420(1–2), pp. 112–127. doi: 10.1016/j.jmb.2012.04.006.

Rivera-Cancel, G., Ko, W.-H., Tomchick, D. R., Correa, F. and Gardner, K. H. (2014) 'Full-length structure of a monomeric histidine kinase reveals basis for sensory regulation', *Proceedings of the National Academy of Sciences*, 111(50), pp. 17839–17844. doi: 10.1073/pnas.1413983111.

Röllen, K., Granzin, J., Panwalkar, V., *et al.* (2016) 'Signaling States of a Short Blue-Light Photoreceptor Protein PpSB1-LOV Revealed from Crystal Structures and Solution NMR Spectroscopy', *Journal of Molecular Biology*, 428(19), pp. 3721–3736. doi: 10.1016/j.jmb.2016.05.027.

Russo, F. D. and Silhavy, T. J. (1993) 'The essential tension: opposed reactions in bacterial two-component regulatory systems', *Trends in Microbiology*, pp. 306–310. doi: 10.1016/0966-842X(93)90007-E.

Sample, V., Newman, R. H. and Zhang, J. (2009) 'The structure and function of fluorescent proteins', *Chemical Society Reviews*. The Royal Society of Chemistry, 38(10), p. 2852. doi: 10.1039/b913033k.

Sancar, A. (2004) 'Regulation of the Mammalian Circadian Clock by Cryptochrome', *Journal of Biological Chemistry*, 279(33), pp. 34079–34082. doi: 10.1074/jbc.R400016200.

Schleicher, E., Kowalczyk, R. M., Kay, C. W. M., Hegemann, P., Bacher, A., Fischer, M., Bittl, R., Richter, G. and Weber, S. (2004) 'On the Reaction Mechanism of Adduct Formation in LOV Domains of the Plant Blue-Light Receptor Phototropin', *Journal of the American Chemical Society*. American Chemical Society, 126(35), pp. 11067–11076. doi: 10.1021/ja049553q.

Schwerdtfeger, C. and Linden, H. (2003) 'VIVID is a flavoprotein and serves as a fungal blue light photoreceptor for photoadaptation', *The EMBO Journal*, 22(18), pp. 4846–4855. doi: 10.1093/emboj/cdg451.

Serrano, L., Neira, J.-L., Sancho, J. and Fersht, A. R. (1992) 'Effect of alanine versus glycine in α-helices on protein stability', *Nature*, 356(6368), pp. 453–455. doi: 10.1038/356453a0.

Shcherbakova, D. M., Shemetov, A. A., Kaberniuk, A. A. and Verkhusha, V. V (2015) 'Natural Photoreceptors as a Source of Fluorescent Proteins, Biosensors, and Optogenetic Tools', *Annual Review of Biochemistry*, 84(1), pp. 519–550. doi: 10.1146/annurev-biochem-060614-034411.

Sheldrick, G. M. (2010) 'Experimental phasing with SHELXC / D / E : combining chain tracing with density modification', *Acta Crystallographica Section D Biological Crystallography*. International Union of Crystallography, 66(4), pp. 479–485. doi: 10.1107/S0907444909038360.

Sievers, F., Wilm, A., Dineen, D., *et al.* (2011) 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Molecular systems biology*, 7(1), p. 539. doi: 10.1038/msb.2011.75.

Silva-Junior, M. R., Mansurova, M., Gärtner, W. and Thiel, W. (2013) 'Photophysics of Structurally Modified Flavin Derivatives in the Blue-Light Photoreceptor YtvA: A Combined Experimental and Theoretical Study', *ChemBioChem*, 14(13), pp. 1648–1661. doi: 10.1002/cbic.201300217.

Singer, T. P., Thimot, N. Z., Massey, V. and Kearney, E. B. (1956) 'Purification and properties of succinic dehydrogenase from yeast', *Archives of Biochemistry and Biophysics*. Academic Press, 62(2), pp. 497–499. doi: 10.1016/0003-9861(56)90148-5.

Sousa da Silva, A. W. and Vranken, W. F. (2012) 'ACPYPE - AnteChamber PYthon Parser interfacE', *BMC Research Notes*. BioMed Central, 5(1), p. 367. doi: 10.1186/1756-0500-5-367.

Stadler, A. M., Knieps-Grünhagen, E., Bocola, M., Lohstroh, W., Zamponi, M. and Krauss, U. (2016) 'Photoactivation Reduces Side-Chain Dynamics of a LOV Photoreceptor', *Biophysical Journal*, 110(5), pp. 1064–1074. doi: 10.1016/j.bpj.2016.01.021.

von Stetten, D., Giraud, T., Carpentier, P., *et al.* (2015) 'In crystallo optical spectroscopy (icOS) as a complementary tool on the macromolecular crystallography beamlines of the ESRF', *Acta Crystallographica Section D Biological Crystallography*. International Union of Crystallography, 71(1), pp. 15–26. doi: 10.1107/S139900471401517X.

Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D. and Eisenberg, D. (2006) 'Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis', *Proceedings of the National Academy of Sciences*, 103(21), pp. 8060–8065. doi: 10.1073/pnas.0602606103.

Swartz, T. E., Corchnoy, S. B., Christie, J. M., Lewis, J. W., Szundi, I., Briggs, W. R. and Bogomolni, R. A. (2001) 'The Photocycle of a Flavin-binding Domain of the Blue Light Photoreceptor Phototropin', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 276(39), pp. 36493–36500. doi: 10.1074/jbc.M103114200.

Swartz, T. E., Tseng, T.-S., Frederickson, M. A., *et al.* (2007) 'Blue-Light-Activated Histidine Kinases: Two-Component Sensors in Bacteria', *Science*, 317(5841), pp. 1090–1093. doi: 10.1126/science.1144306.

Tang, Y., Cao, Z., Livoti, E., Krauss, U., Jaeger, K.-E., Gärtner, W. and Losi, A. (2010) 'Interdomain signalling in the blue-light sensing and GTP-binding protein YtvA: A mutagenesis study uncovering the importance of specific protein sites', *Photochem. Photobiol. Sci.* The Royal Society of Chemistry, 9(1), pp. 47–56. doi: 10.1039/B9PP00075E.

Taylor, B. L. and Zhulin, I. B. (1999) 'PAS Domains: Internal Sensors of Oxygen, Redox Potential, and Light', *Microbiol. Mol. Biol. Rev.*, 63(2), pp. 479–506. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10357859.

Thöing, C., Pfeifer, A., Kakorin, S. and Kottke, T. (2013) 'Protonated triplet-excited flavin resolved by step-scan FTIR spectroscopy: implications for photosensory LOV domains', *Physical Chemistry Chemical Physics*, 15(16), p. 5916. doi: 10.1039/c3cp43881c.

Tokutomi, S., Matsuoka, D. and Zikihara, K. (2008) 'Molecular structure and regulation of phototropin kinase by blue light', *Biochimica et Biophysica Acta - Proteins and Proteomics*, pp. 133–142. doi: 10.1016/j.bbapap.2007.09.010.

Tomomori, C., Tanaka, T., Dutta, R., *et al.* (1999) 'Solution structure of the homodimeric core domain of Escherichia coli histidine kinase EnvZ', *Nature structural biology*. Nature Publishing Group, 6(8), pp. 729–734. doi: 10.1038/11495.

Vaidya, A. T., Chen, C.-H., Dunlap, J. C., Loros, J. J. and Crane, B. R. (2011) 'Structure of a light-activated LOV protein dimer that regulates transcription', *Science*

signaling, 4(184), p. ra50. doi: 10.1126/scisignal.2001945.

Voss, N. R. and Gerstein, M. (2010) '3V: cavity, channel and cleft volume calculator and extractor', *Nucleic Acids Research*. Oxford University Press, 38(Web Server), pp. W555–W562. doi: 10.1093/nar/gkq395.

Wang, C., Sang, J., Wang, J., *et al.* (2013) 'Mechanistic insights revealed by the crystal structure of a histidine kinase with signal transducer and sensor domains', *PLoS biology*. Public Library of Science, 11(2), p. e1001493. doi: 10.1371/journal.pbio.1001493.

Williams, T. and Kelley, C. (2016) *Gnuplot 5.0: an interactive plotting program*. Available at: http://gnuplot.info (Accessed: 17 April 2016).

Wilson, A. J. C. (1950) 'Largest likely values for the reliability index', *Acta Crystallographica*, 3(5), pp. 397–398. doi: 10.1107/S0365110X50001129.

Wingen, M., Potzkei, J., Endres, S., *et al.* (2014) 'The photophysics of LOV-based fluorescent proteins – new tools for cell biology', *Photochemical & Photobiological Sciences*. The Royal Society of Chemistry, 13(6), p. 875. doi: 10.1039/c3pp50414j.

Winn, M. D., Ballard, C. C., Cowtan, K. D., *et al.* (2011) 'Overview of the CCP 4 suite and current developments', *Acta Crystallographica Section D Biological Crystallography*. International Union of Crystallography, 67(4), pp. 235–242. doi: 10.1107/S0907444910045749.

Wriggers, W. (2010) 'Using Situs for the integration of multi-resolution structures', *Biophysical reviews*, 2(1), pp. 21–27. doi: 10.1007/s12551-009-0026-3.

Yamada, S., Sugimoto, H., Kobayashi, M., Ohno, A., Nakamura, H. and Shiro, Y. (2009) 'Structure of PAS-Linked Histidine Kinase and the Response Regulator Complex', *Structure*, 17(10), pp. 1333–1344. doi: 10.1016/j.str.2009.07.016.

Yee, E. F., Diensthuber, R. P., Vaidya, A. T., Borbat, P. P., Engelhard, C., Freed, J. H., Bittl, R., Möglich, A. and Crane, B. R. (2015) 'Signal transduction in light-oxygen-voltage receptors lacking the adduct-forming cysteine residue', *Nature communications*. Nature Publishing Group, 6, p. 10079. doi: 10.1038/ncomms10079.

Zoltowski, B. D. and Gardner, K. H. (2011) 'Tripping the Light Fantastic: Blue-Light Photoreceptors as Examples of Environmentally Modulated Protein–Protein Interactions', *Biochemistry*, 50(1), pp. 4–16. doi: 10.1021/bi101665s.

Zoltowski, B. D., Motta-Mena, L. B. and Gardner, K. H. (2013) 'Blue Light-Induced Dimerization of a Bacterial LOV–HTH DNA-Binding Protein', *Biochemistry*. American Chemical Society, 52(38), pp. 6653–6661. doi: 10.1021/bi401040m.

Zoltowski, B. D., Nash, A. I. and Gardner, K. H. (2011) 'Variations in Protein–Flavin Hydrogen Bonding in a Light, Oxygen, Voltage Domain Produce Non-Arrhenius Kinetics of Adduct Decay', *Biochemistry*. American Chemical Society, 50(41), pp. 8771–8779. doi: 10.1021/bi200976a.

Zoltowski, B. D., Schwerdtfeger, C., Widom, J., Loros, J. J., Bilwes, A. M., Dunlap, J. C. and Crane, B. R. (2007) 'Conformational switching in the fungal light sensor Vivid', *Science*, 316(5827), pp. 1054–7. doi: 10.1126/science.1137128.

Zoltowski, B. D., Vaccaro, B. and Crane, B. R. (2009) 'Mechanism-based tuning of a LOV domain photoreceptor', *Nature chemical biology*. Nature Publishing Group, 5(11), pp. 827–34. doi: 10.1038/nchembio.210.

7 Appendix

7.1 List of figures

7.2 List of tables

7.3 Protein sequences

PpSB2-LOV

MGSSHHHHHHSSGLVPRGSHMINAKLLQLMVEHSNDGIVVAEQEGNESILIYVNPAFE
RLTGYCADDILYQDCRFLQGEDHDQPGIAIIREAIREGRPCCQVLRNYRKDGSLFWNE
LSITPVHNEADQLTYYIGIQRDVTAQVFAEERVRELEAEVAELRRQQGQAKH

Pf5-LOV

MGSSHHHHHHSSGLVPRGSHMINAHLLQRMINASNDGIVVAEQEGEDNIVIYVNPAFE
RLTGYSADEVLYQDCRFLQSGDRDQPGLEVIRQALRQGRPCREVLRNYRKDGSHFWNE
LSITPVFNDSDQLTYFIGVQKDVSVQVKAQQRLLQLEQQLAEVQAELAALKATSGH

SB2F1

MGSSHHHHHHSSGLVPRGSHMINAKLLQLMVEHSNDGIVVAEQEGNESILIYVNPAFE
RLTGYCADDILYQDCRFLQGEDHDQPGIAIIREAIREGRPCCQVLRNYRKDGSLFWNE
LSITPVHNEADQLTYYIGIQRDVTEHQQTQARLQELQSELVHVSRLSAMGEMASALAH
ELNQPLAAISNYMKGSRRLLAGSSDPNTPKVESALDRAAEQALRAGQIIRRLRDFVAR
GESEKRVESLSKLIEEAGALGLAGAREQNVQLRFSLDPGADLVLADRVQIQQVLVNLF
RNALEAMAQSQRRELVVTNTPAADDMIEVEVSDTGSGFQDDVIPNLFQTFFTTKDTGM
GVGLSISRSIIEAHGGRMWAESNASGGATFRFTLPAADEN

SB1F1

MGSSHHHHHHSSGLVPRGSHMINAQLLQSMVDASNDGIVVAEKEGDDTILIYVNAAFE
YLTGYSRDEILYQDCRFLQGDDRDQLGRARIRKAMAEGRPCREVLRNYRKDGSAFWNE
LSITPVKSDFDQRTYFIGIQKDVSEHQQTQARLQELQSELVHVSRLSAMGEMASALAH
ELNQPLAAISNYMKGSRRLLAGSSDPNTPKVESALDRAAEQALRAGQIIRRLRDFVAR
GESEKRVESLSKLIEEAGALGLAGAREQNVQLRFSLDPGADLVLADRVQIQQVLVNLF
RNALEAMAQSQRRELVVTNTPAADDMIEVEVSDTGSGFQDDVIPNLFQTFFTTKDTGM
GVGLSISRSIIEAHGGRMWAESNASGGATFRFTLPAADEN

SBW25-LOV

MGSSHHHHHHSSGLVPRGSHMINAKLMQLVINASNDGIVVAEREGKDKPLIYVNPAFE
RLTGYTLDEILYQDCRFLQSGDRDQPALMAIRETLESGGACREILRNYRKDGSHFWNE
LSLSTVYNEADKQTYFVGVQKDVTLQVKAQQRVGQLEAELNQVKAELAALKATSGFNK
I

W619_1-LOV

MGSSHHHHHHSSGLVPRGSHMINAQLLQSMVDASNDGIVVAEQEGDDTILIYVNPAFE
RLTGYSRDEILYQDCRFLQGDDRDQLARARIRKALAEGRPCREVLRNYRKDGSAFWNE
LSITPVKCDADHRTYFIGIQKDVSRQVELERELAEMHVRRNFDKRPEPSA

* the His-tag residues are underlined.