# Development and application of new computer simulation techniques to describe the multi-scale relaxation dynamics of light-sensitive protein systems

Dissertation
zur Erlangung
des Doktorgrades der Naturwissenschaften
(Dr. rer. nat. )
der Naturwissenschaftlichen Fakultät IV
-Chemie und Pharmazie-
der Universität Regensburg


vorgelegt von
Emanuel Peter
aus Regensburg
März 2012

“Ich weiß, dass ich nichts weiß”

Sokrates

**Eidesstattliche Erklärung**

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe des Literaturzitats gekennzeichnet.

Emanuel Peter

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Die vorliegende Arbeit wurde in der Zeit von Oktober 2008 bis März 2012 unter Anleitung von Herrn PD. Dr. Stephan A. Bäurle am Institut für physikalische und theoretische Chemie der Universität Regensburg angefertigt.
Promotionsgesuch eingereicht am: 15.03.2012
Tag des Kolloquiums : 26.06.2012

**Prüfungsausschuss:**

- Prof. Dr. Robert Wolf (Vorsitzender)
- PD. Dr. Stephan Bäurle
- Prof. Dr. Bernhard Dick
- Prof. Dr. Hartmut Krienke

**Danksagung**

Im Folgenden danke ich denjenigen, welche mir beim Verfassen dieser Arbeit massgeblich geholfen haben. Hierbei ist PD Dr. Stephan Bäurle an erster Stelle zu nennen, der mich in allen Fragestellungen der Doktorarbeit stets hervorragend betreut hat. Er scheute keine Mühen um mir in angeregten Diskussionen die interessante Materie der theoretischen Chemie näherzubringen. An dieser Stelle danke ich Herrn Prof. Dr. Bernhard Dick, der als Lehrstuhlinhaber mit wertvollen Anregungen zu dieser Arbeit beitrug. Nicht zu vergessen sind meine Bürokollegen - Roger-Jan Kutta und Kathrin Magerl, welche für eine gute Büroatmosphäre und ein freundschaftliches Klima sorgten. Beiden danke ich für die gute Stimmung und die Youtube-Freitage, ohne die ich wahrscheinlich ein anderer (in Musik un-gebildeter) Mensch wäre. Den beiden Kollegen - Sergii Donets und Anton Pershin, danke ich für das freundschaftliche Klima und die Anregungen sowie Diskussionen in theoretischen Fragestellungen.

Zuallerletzt und im Besonderen danke ich meinen Eltern Gung-ok und Alexander Peter - den beiden grossen Unterstützern meiner Arbeit. Ohne deren Ermutigungen in oft schwierigen Lagen wäre es für mich oft nicht möglich gewesen weiterzumachen. Meinem Bruder Robert Peter danke ich für die Unterstützung aus der Ferne und die vielen lustigen Treffen, wenn er sich mal nicht in der Weltgeschichte herumtrieb.

# Contents

# Part I.

# Introduction

# 1. Scope of thesis

The scope of this thesis is the investigation of the multi-scale signal-transduction dynamics of protein-based photosensors. Based on the previous experimental findings of Kottke *et al.* [1], Lanzl *et al.* [2] and Kutta *et al.* [3] on the LOV1-domain from *Chlamydomonas reinhardtii* (CrLOV1), we first used the popular molecular dynamics (MD) technique to describe the short-time relaxation dynamics of the CrLOV1-system, as well as the LOV2-domain from *Avena sativa* and the Vivid LOV-domain from *Neurospora crassa*. In addition, we simulated the early stages of signal-transduction of the artificial photoactivable-Rac1 PA-Rac1 photo-enzyme. These LOV-photosensors, which are a subclass of the PAS-family, were of essential interest in the Graduate College for sensory photoreceptors in natural and artificial systems GRK640 as well as in the Graduate College for chemical photocatalysis GRK1626, supported by the German Research Council (DFG). Their signal-transduction pathway includes relevant structural changes, which occur on time-scales form ns up to several ms. In order to handle this range of timescales, we developed rate-based Kinetic-Monte-Carlo (KMC) simulation algorithms as well as mixed mesoscopic-atomistic coarse-grained (CG) models. With this new simulation-technique, we described the signaling pathways of LOV-based photosensors on multiple timescales from nanoseconds to seconds. As a result of these simulations, we observed the early signal transduction dynamics in the ns-timescale in the vicinity of the FMN-chromophore, which were complementary to our previous MD-results on the same natural and artificial photosensory systems. Additionally to these early timescales we obtained informations about the late signaling stages of the LOV-based photosensors within timescales from $\mu$s to many seconds.

We start with a general introduction on the photosensory systems and the theory on the methods, which we used in our work. In the next sections we present the algorithmic developments, which we designed for the simulation on multiple timescales for the desciption of short-time as well as long-time signal-transduction events. We continue with the description of our results on the simulations of wild-type protein photosensors and finally will present our findings on artificial LOV-based systems.

*1. Scope of thesis*

In the scope of this thesis, following works have been published :

- "Mechanism of signal transduction of the LOV2-J photosensor from Avena sativa"
  Emanuel Peter, Bernhard Dick and Stephan A. Baeurle *Nature Communications* (2010), **1** (8), 122. DOI: 10.1038/ncomms1121
- "Effect of computational methodology on the conformational dynamics of the protein photosensor LOV1 from Chlamydomonas reinhardtii"
  E. Peter, Bernhard Dick, Stephan A. Baeurle *J. Chem. Biol.* (2011), **4** (4), 167-184. DOI: 10.1007/s12154-011-0060-z
- "Signals of LOV1: a computer simulation study on the wildtype LOV1-domain of Chlamydomonas reinhardtii and its mutants"
  E. Peter, B. Dick, S. A. Baeurle *J. Mol. Mod.* (2011), DOI: 10.1007/s00894-011-1165-6
- "Illuminating the early signaling pathway of a fungal LOV-photoreceptor"
  Emanuel Peter, Bernhard Dick, Stephan A. Baeurle *Proteins* (2012), **80** (2), 471-481. DOI: 10.1002/prot.23213
- "Signaling pathway of a photoactivable Rac1-GTPase in the early stages"
  Emanuel Peter, Bernhard Dick, Stephan A. Baeurle *Proteins* **80** (5), (2012), 1350-1362 DOI: 10.1002/prot.24031
- "A novel computer simulation method for simulating the multiscale transduction dynamics of signal proteins"
  Emanuel Peter, Bernhard Dick, Stephan A. Baeurle *J. Chem. Phys.*, **136** (12), 124112 (2012).

# 2. Light-oxygen-voltage sensitive domains

## 2.1. PAS-domain

Light-oxygen-voltage sensitive (LOV)-domains control the cell metabolism in response to external stimuli in all kingdoms of life. They are a sub-class of the PAS-domains, which are named after the first proteins where these structural motifs were identified: Period (Per), aryl-hydrocarbon-nuclear-translocator (ARNT), and single-minded (Sim) [4, 5, 6]. The first structural investigation which determined the characteristics of folding of these PAS-domains was performed by Pellequer *et al.*, who investigated the structure of the photoactive-yellow protein (PYP) as a prototype of PAS-domains [7]. The generalization of the overall structure of PAS-domains was supported by several structural investigations on the FixL- and the human HERG-protein [8, 9]. Characteristically the N-terminal cap of the PAS-domains is folded $\alpha$-helical, whereas the hydrophilic PAS-core region faces the hydrophobic $\beta$-scaffold. Both motifs, which are connected by one helical connector, together form a binding pocket for co-factors [5, 10, 11]. This co-factor controls the sensitivity to environmental stimuli, such as light, oxygen or voltage. The usage of co-factors is a widely used concept for cell survival in all kingdoms of life, which are the kingdoms of *Archea, Bacteria* and *Eucarya*. These sensitive proteins include histidine- and ser/thr-kinases, chemoreceptors and photoreceptors for taxis and tropism, circadian clock proteins, voltage-activated ion channels, cyclic nucleotide phosphodiesterases, and regulators of responses to hypoxia and embryological development of the central nervous system [5].

Oxygen itself is a terminal acceptor for oxidative phosphorylation and a toxic agent that forms reactive free radicals when partially reduced. Several authors characterized FixL as a sensor for oxygen. Oxygen dissociation from the FixL input PAS-domain causes conformational changes in this protein [8, 12, 13, 14, 15, 16]. This induces increased autophosphorylation activity of the output-domain [13, 14, 8]. FixL catalyzes a His-Asp phosphoryl transfer to the receiver module of the response regulator FixJ. This latter phosphorylated FixJ protein acts as a transcriptional activator of the genes involved in the nitrogen fixation. Sensing of light intensity and wavelength governs such cellular responses as phototropism in plants and phototaxis in bacteria. One given example is the photoactive-yellow-protein (PYP), which is a bacterial photoreceptor postulated to govern a photophobic swimming response in *Ectothiorhodospira halophila* [17]. This photoreceptor is an isolated PAS-domain with a non-covalently attached 4-hydroxycinnamyl chromophore [17]. By contrast, plant phytochromes use a tetra-pyrrole chromophore. In this photosensitive domain a hinge region separates two PAS-domains [18]. A histidine kinase transmitter domain is located C-terminal to the PAS-domains [19, 20]. These PAS-domains transmit a signal onto the kinase domain after light illumination and lead to the plantal response after a sequence of secondary signals. PAS-domains also act as regulators and signature motifs of circadian clocks as shown by many examples in nature [21, 22]. In addition to being transcriptional regulators with DNA-binding motifs, many circadian-clock proteins contain PAS-domains. Moreover, mammals regulate their master clock by PAS-domains within the so-called suprachiasmatic nuclei in the hypothalamus [23]. So far, PAS-domains are the only motifs known to be conserved among widely diverse clock proteins [5]. Clocks are made up of transcription factors that feed back and inhibit their own transcription. There is also evidence that depletion of cellular energy levels is first sensed by cellular organisms in a decreased electron transport and proton motive force. Moni-

toring electron transport or proton motive force can quickly alert a cell to energy loss [5]. It was found by Bespalov *et al.*, that *E. coli* migrate to a microenvironment with preferred redox potential [25, 26]. The PAS-domain, which is responsible for this movement is the *E. coli Aer*-protein. Some authors hypothesized, that cells sense a change in proton motive force [27, 28]. In fact cells recognize proton-motive force related parameters by using sensory-modules as Aer, ArcB and HIF. In metazoans, oxygen concentration is monitored by the hypoxic response pathway, where hypoxia-inducible factors (HIFs) regulate numerous genes in response to oxygen [24, 29]. HIF-proteins are heterodimeric basic helix-loop-helix PAS-domains. These act as transcription factors in an oxygen-dependent manner [30]. In summary, PAS-domains regulate a multitude of cellular functions in all kingdoms of life - bacteria, fungi, plants and mammals. In the next section we will concentrate on the LOV-domains, which are one subclass of the PAS-domain superfamily.

## 2.2. LOV-domain of phototropin1

Phot proteins are blue light photoreceptors found in higher plants as well as in micro-algae. They regulate biological processes such as phototropic plant movement, chloroplast relocation, stomatal opening, rapid inhibition of stem growth and gametogenesis [31, 32, 33, 34]. They are composed of 2 light-oxygen-voltage (LOV) sensitive PAS-domains, each containing a non-covalently bound flavin-mononucleotide FMN chromophore, and a C-terminal kinase domain. Upon blue-light absorption, a covalent bond between the FMN chromophore and an adjacent reactive cysteine residue of the apo-protein is formed in the LOV-domains. This subsequently mediates the activation of the kinase domain, which induces a signal in the organism via photoreceptor auto-phosphorylation [35]. A common feature of phot proteins is that the photochemical reactivity of LOV2 is required for the activation of the attached kinase domain. By contrast, the role of LOV1 is still unclear, although size exclusion chromatography in addition to small angle x-ray scattering analysis suggests that this domain may be involved in photoreceptor dimerization [3, 36]. In several studies it has been proposed that the LOV1-domain might be responsible for regulating the lifetime and signal intensity of the LOV2-domain under low light conditions and in this way affect the phototropin activity [34, 37, 38, 39]. However, it was found by Kaiserli *et al.* that the LOV1-domain in *Arabidopsis thaliana* is not involved into phot1 dimerization and intermolecular phosphorylation [40]. Remarkably, even phot1 which is deficient from LOV1, shows the same dimerization-mediated mechanism as full-length phot1. In the past decade the primary photoreaction mechanism and the initial structural changes of the wild-type LOV-domains, as well as related mutants, have been investigated in detail through experimental as well as theoretical means. However, despite tremendous efforts still only little is known about their signal transduction pathway as well as dark-state activity at the molecular level.

## 2.3. Experimental investigations on LOV-domains

To elucidate the photocycle, Kottke *et al.* [1] investigated the LOV1-domain of phot1 from *Chlamydomonas reinhardtii* with time-resolved absorption spectroscopy. They found that photoexcitation of the dark-state form of LOV1, designated as LOV1-447, causes transient bleaching and the formation of two spectrally similar red-shifted intermediates that can both be assigned to triplet states of the FMN. These triplet states decay with time constants of 800 ns and 4 $\mu$s with an efficiency of $> 90$ % into a blue-shifted intermediate, designated as LOV1-390, that can be correlated with the formation of a thio-adduct between the reactive cysteine residue (Cys57) of the apo-protein and the C4a-atom on FMN (FMN-C4a). Then, the adduct state LOV1-390 reverts

to the dark-state form in hundreds of seconds with a time constant being dependent on pH and salt concentration. Fedorov *et al.* [11] performed accurate measurements of the crystal structure of the LOV1-domain from the same species in the dark and illuminated state (resolution: 1.9 Å) and discovered two different dark-state conformers, populated with a probability distribution of 70 % : 30 %. They concluded from their study that the conformers differ in the rotation around the $C_\alpha$-$C_\beta$-bond of the Cys57-residue, which may possess different reactivity and could explain the occurrence of the two triplet states observed in the spectroscopic experiments of Kottke *et al.* [1]. Through Fourier transform infrared (FTIR) spectroscopy, Bednarz *et al.* [41] further established that in both conformers the thiol group is exposed to environments of different H-bonding strength and correlated the structural characteristics of the dark-state conformers with the decay kinetics of the respective triplet states. Schleicher *et al.* [42] studied the photoinduced triplet state and the photoreactivity of the FMN-cofactor in wild-type LOV1 and LOV2-domains as well as in mutant LOV-domains from *Chlamydomonas reinhardtii*, *Avena sativa* and *Adiantum capillus-veneris* by time-resolved electron paramagnetic resonance and UV-visible spectroscopy at low temperatures ($T \leq 80$ K). They suggested that differences in the electronic structure of the FMN, as reflected by altered zero-field splitting parameters of the triplet state, can be correlated with changes in the amino acid composition of the binding pocket. In a multitude of subsequent experimental studies, it has been demonstrated that the photochemically induced signal of phototropin, initiated by adduct formation, is very different from that of other photoactive proteins, where the photoreaction is initiated by photoisomerization, like for example in case of the rhodopsins, PYP and phytochrome. While the photoreactions of the latter proteins show only little temperature sensitivity, the phototropins show a highly temperature-dependent photoreaction and signal transduction pathway [43]. To investigate this aspect, Iwata *et al.* [44] compared the light-induced structural changes of the LOV1 and LOV2-domains from *Adiantum phytochrome3* (phy3) by means of UV-visible and FTIR spectroscopy in the temperature range from 77 K to 295 K. They observed that the thiol group of the reactive cysteine residue forms an H-bond in phy3-LOV1, which is strengthened at low temperatures. They related this with the fact that no adduct formation takes place for phy3-LOV1 at 77 K, as revealed by UV-visible spectroscopy. A reduction of the occurrence of the adduct form was also observed for phy3-LOV2 at low temperature [45], i.e. at 77 K the yield of the adduct state of phy3-LOV2 decreased by an amount of 36 % with regard to the yield at room temperature. Based on these experiments, Iwata *et al.* [44, 45] suggested that the presence of a non-reactive fraction of dark-state conformers at low temperature may originate from a local structural heterogeneity between the FMN and the reactive cysteine residue, whose thiol group might form an H-bond with some surrounding amino acid sidechain, peptide carbonyl or internal water molecule. In such a situation the respective reaction centers may thus be too far from each other and could possess a too low mobility to form an adduct state. In a subsequent FTIR study Sato *et al.* [43] observed dark-state conformers with different S-H-stretching frequencies and suggested that the micro-environment of the thiol group of the reactive cysteine residue determines the reactivity at low temperatures. Kennis *et al.* [46] investigated the molecular mechanism of signaling of the LOV2-domain with an additional $\alpha$-helix, called the J$\alpha$-helix, of phot1 from *Avena sativa* (AsLOV2-J$\alpha$), using FTIR spectroscopy. In this study they characterized a newly isolated low-hydration intermediate that shows a downshift of high-frequency amide I signals and possibly corresponds to a loop tightening without large structural changes at the $\beta$-sheet or the J$\alpha$-helix. In addition, they reported a heterogeneity involving two different populations of FMN-C4=O conformers, coexisting in the dark-state and characterized by two different C4=O-carbonyl frequencies. They explained the occurrence of these two carbonyl frequencies by the presence of conformers with either a singly H-bonded or twice H-bonded FMN-C4=O with the surrounding amino acids. According to their analysis, these conformers display slightly shifted absorption spectra and cause a splitting of the

475-nm band in the UV-visible spectra of the LOV-domains at low temperature.

### 2.3.1. Theoretical investigations on LOV-domains

From the theoretical side, a major focus was put on the investigation of the primary steps of photoreaction and the initial structural changes after blue-light absorption of the photocycle. Fedorov *et al.* [11] performed first quantum chemical calculations with their crystal structure configurations, which provided information about possible pathways of the photoreaction in the active site. They concluded from their study that the light-induced structural changes of the protein domain in the crystal are minor and mainly take place around the chromophore. Neiss and Saalfrank [47] performed *ab initio* quantum mechanical calculations to determine the electronic properties of several FMN-related model compounds in different charge and spin states. They deduced from their calculation results that the reaction pathway proceeds via the triplet T1-state of the FMN and that the process of adduct formation involves more likely radical intermediates, instead of ionic species. Dittrich *et al.* [48] studied the early steps of the photocycle by performing combined quantum mechanical/molecular mechanical (QM/MM) simulations of a complete LOV1-domain from *Chlamydomonas reinhardtii*. They investigated the electronic properties and initial structural changes that follow blue-light absorption, and concluded from their calculation results that the pathway for cysteinyl-FMN adduct formation proceeds via a neutral radical state, generated by H-atom transfer from the reactive Cys57-residue to the FMN chromophore. In a later work Freddolino *et al.* [49] investigated the structural changes, taking place at longer times after adduct formation, by performing molecular dynamics (MD) simulations with the LOV1-domain from *Chlamydomonas reinhardtii* and the LOV2-domain from *Avena sativa*. To this end, they performed five independent 12 ns simulations with the light and dark-states of both protein domains, revealing significant differences in how the LOV1 and LOV2-domains respond to photoactivation. Based on their calculation results, they proposed that the LOV1 activation is caused by a change in the H-bond network between FMN chromophore and the surrounding apo-protein that destabilizes a highly conserved salt bridge, whereas LOV2 activation results from a change in flexibility of a set of protein loops. Neiss and Saalfrank [50] performed MD simulations with the dark and illuminated state of the LOV2-domain of *Adiantum capillus-veneris*. In accordance with the accurate x-ray diffraction measurements of Fedorov *et al.* [11] for the LOV1-domain from *Chlamydomonas reinhardtii*, they detected three dark-state conformers, which are populated with about equal probability and differ with respect to a rotation around the $C_\alpha$-$C_\beta$-bond of the reactive Cys57-residue. However, they found no clear changes in the structure or dynamics of the protein domain, which could be relevant for signal transduction after adduct formation. They suggested that the conformational behavior of the isolated LOV2-domain in the crystal should differ from the biologically active LOV2 in solution in its natural environment. Similar findings were also recently made for related photoactive proteins. Arai *et al.* [51] found that for the putative signaling state of PYP there are serious discrepancies between the crystal structure, obtained from time-resolved Laue diffraction analysis, and the solution structure, obtained from various spectroscopic studies. To explore the origin of such disagreements, they performed MD simulations with the PYP in the crystal and compared the results with those from MD simulations of the protein in solution. They found clear differences in the root-mean square deviation, when they relaxed the crystallographically determined structure of the protein in solution. Finally, Sato *et al.* [43] performed MD simulations in conjunction with a FTIR study on the LOV2-domain from *Adiantum neochrome 1* (neo1) and observed that multiple protein conformers can co-exist, differing in the heterogeneous environment of the reactive cysteine residue. They found that the conformers can thermally be converted into each other by conformational fluctuations at physiological temperatures.

### 2.3.2. Experiments on LOV2-J$\alpha$-system

In the past decade several experimental works have attempted to clarify the response mechanism of phototropins to blue-light photoexcitation. To this end, major efforts were concentrated on the investigation of the dynamical changes of isolated LOV-domains in the light-state [35, 52]. Studies of independent protein parts, however, can only provide limited informations on the complex interaction mechanism taking place between the different structural elements of the full phototropin during the signal transduction process. Harper *et al.* [53] first characterized the light-dependent changes of the AsLOV2-J$\alpha$-photosensor by solution NMR-spectroscopy and found that the J$\alpha$-helix plays a key role in the kinase activation. From their results, they deduced that the photoinduced adduct formation in the LOV2-domain results in a disruption of the J$\alpha$-helix from the G$\beta$-, H$\beta$- and I$\beta$-strands, which in turn may trigger a signal in the kinase. To examine how the structural changes upon illumination are coupled to the kinase activation, Harper *et al.* [54] in a subsequent work performed a series of point mutations along the J$\alpha$-helix to disrupt its interaction with the LOV-domain under dark-state conditions. Using NMR-spectroscopy and limited proteolysis, they demonstrated that several of these mutations are able to displace the J$\alpha$-helix from the LOV-domain even in the absence of light. When placed into the full-length phototropin, these point mutations displayed constitutive kinase activation without illumination of the sample. These results indicate that unfolding of the J$\alpha$-helix is the critical event in the regulation of the kinase signal. Recent crystallographic data from Halavaty and Moffat [55] on the AsLOV2-J$\alpha$-photosensor also attributed the J$\alpha$-helix a predominant role in the signal transduction. While these studies clearly identified the C-terminal J$\alpha$-helix to be responsible for communicating the signal to a downstream effector domain, it remained unclear how covalent adduct formation in the core triggers the $\alpha$-helical unfolding on the surface of the domain. In a series of experimental and theoretical studies with LOV2-domains a conserved glutamine residue, located on the I$\beta$-strand in the immediate vicinity of the FMN-binding site, was found to change its H-bonding pattern with the FMN upon photoexcitation and was suspected to be involved in the transmission of the local stress to the surface of the LOV-domain [10, 35, 52, 49]. By designing point mutations of this glutamine residue and monitoring the effects on the LOV2-domain using UV-visible-absorbance and circular-dichroism spectroscopy, Nash *et al.* [56] recently confirmed that it plays a central role in both spectral tuning and signal propagation from the LOV2-core to the peripheral J$\alpha$-helix. However, up to now a detailed understanding of the spatio-temporal dynamics of the molecular processes involved in this critical event is still lacking.

### 2.3.3. Fungal LOV-domain Vivid

In fungal organisms LOV-domains are employed for adjusting the circadian temporal organization of the cells to the daily and seasonal periods [57]. In case of the fungus *Neurospora crassa* the circadian clock is controlled by two light-sensitive proteins, known as the White-Collar-Complex (WCC) and the LOV-domain Vivid (VVD-LOV) [58, 59]. WCC is primarily responsible for the light-induced transcription on the control-gene Frequency (FRQ) under daylight conditions, which drives the expression of VVD-LOV and governs the negative feedback loop onto the circadian clock [59, 60]. By contrast, the role of VVD-LOV is mainly modulatory and does not directly affect FRQ [58, 61]. In several recent works it has been demonstrated that under low-light conditions VVD-LOV modulates the circadian clock through inhibiting WCC after transmission of the initial signal [61, 62, 63]. Finally, Hunt *et al.* [64] showed that VVD-LOV serves as a buffer against external temperature fluctuations, keeping the circadian clock mainly temperature-insensitive.

Similarly to the LOV1-domain, it has been demonstrated *in vitro* through size-exclusion chromatography of Zoltowski *et al.* [65, 66] and SAXS-measurements by Lamb *et al.* [67] that VVD-

LOV possesses a higher dimerization tendency upon blue-light irradiation than in the dark and that the dimerization process takes place through its $\beta$-strands. Moreover, Zoltowski *et al.* [65] showed through crystal structure determination and mutational experiments that the Cys71-residue plays a central role in the signal transduction pathway from the FAD-chromophore to the N-terminal cap (Ncap) region, influencing the aggregation behavior of VVD-LOV. Based on these findings, they proposed a mechanism in which Cys71 is primarily affecting the H-bond network in the coil region between Ncap and the protein core. Although the major changes in the secondary structure during the signaling process have been resolved up to now, the signal transduction at the molecular level and the functional role of VVD-LOV remain still only poorly understood.

## 2.4. Artificial LOV-photosensors

### 2.4.1. Artificial photoactivable GTPase PA-Rac1

Artificial photoenzymes based on the concept of fusing light-activable protein photoswitches to a target enzyme show promising perspectives in cell signaling [68]. This has first been demonstrated in the pioneering work of Lee *et al.* [69, 70], who created an artificial light-activable photoenzyme based on the blue-light-activable AsLOV2-J$\alpha$-photosensor combined with the enzyme dihydrofolate reductase (DHFR). Even if their fusion protein was structurally not tuned for providing optimal light-dependent catalytic activity, this work provided first indications of the power of such artificial protein constructs for allosteric regulation of enzymatic reactivity in biological systems. In a subsequent work Möglich *et al.* [71] replaced the oxygen-sensing PAS-domain in FixL from *Bradyrhizobium japonicum* with the blue-light photosensor YtvA-LOV from *Bacillus subtilis* and generated the light-inactivated histidine kinase YF1. They showed that the signals of this protein construct are transmitted from the LOV-domain to the histidine kinase through a rotational movement propagating within an $\alpha$-helical coiled-coil linker that reduced net kinase activity by 1000-fold *in vitro* and enabled the control of gene expression in a light-dependent manner in *Escherichia coli*. Further evidences for the potential of such fusion proteins have recently been provided by Wu *et al.* [72], who connected the AsLOV2-J$\alpha$-photosensor to a mutant of the GTPase Rac1. They demonstrated through affinity-assaying that the LOV2-domain in its closed conformation in the dark blocks the binding of the effector protein PAK1 to Rac1, and that light-induced unwinding of the J$\alpha$-helix releases steric inhibition, leading to Rac1-activation. The light-sensitive control of the Rac1-activity allowed Wu *et al.* to stimulate cell motility and control the direction of motion of HeLa-cells by light. However, up to now a detailed molecular based understanding of the signal transduction pathway of photoenzymes based on the AsLOV2-J$\alpha$-photosensor is still lacking.

### 2.4.2. Artificial DNA-photoswitch LOV-TAP

Artificial DNA-regulation through fusion of LOV-photosensors with DNA transcription factors delivers artificial tools for the *in vivo* regulation of DNA-expression by light. These DNA transcription factors act as DNA-switches which tune cell processes in response to environmental signals, as inter- and extra-cellular signals, at the DNA. Notably, every of these DNA transcription regulators is characterizable through its defined and specific DNA binding motif. An immense number of these natural DNA-binding motifs exists, due to the multitude of different nucleotide sequences in all kingdoms of life. The photosensoric White-Collar-Complex WC-1 from *Neurospora crassa* performs light-regulated gene expression [73] and uses the zinc-finger-motif, which exists in many bacteria, phagae [74] and fungi. The loop-motif, which is one essential

structural module of the so-called apoptosis-factor p53 in human cells, is able to sense mutations in nucleotide sequences with its protruding loops. Through this mechanism cell growth and proliferation are tightly regulated. Remarkably, p53 is one of the elementary tumor suppressors and its importance can be derived from the fact, that nearly half of all human cancers are related to mutations in the p53-gene [75, 76]. We note that DNA-switches also occur in bacteria, where they control the expression of proteins depending on the bacteria's environmental conditions. It was found that bacteria, such as *E. coli*, adjust their metabolism onto the conditions which are present in their growth medium. For example, it was observed that falling levels of glucose in *E. coli* cause an increase of cyclic AMP. This molecule binds to the CAP protein to turn on appropriate genes, which enables *E. coli* to use alternative carbon sources than glucose. Moreover, this organism possesses 5 genes that code for enzymes that manufacture the amino-acid tryptophan. If tryptophan is present in the growth medium and enters the cell, the need for these enzymes is no longer given. The short region of regulatory DNA, responsible for the production of tryptophan-producing enzymes, is recognized by the so-called tryptophan-repressor (TrpR)-protein, a member of the helix-turn-helix (HTH) family [77, 78]. If tryptophan is present in the cell, the so-called co-repressor binds at two sites onto one TrpR-protein. Through this binding, the HTH-motif tilts and the DNA-sequence is blocked against the RNA-polymerase. Without tryptophan, the motif changes its sterical arrangement and is unable to bind to DNA [79]. In contrast to chemically induced DNA transcription as in case of TrpR, a multitude of bacterial organisms use the connection of light-sensing PAS-folded domains and DNA switches for light-induced DNA transcription. For example, bacteria *Eryobacter litoralis* (EL222) uses a DNA-photoswitch consisting of one LOV-domain, which is convalently linked to a DNA-binding HTH-motif. Nash *et al.* used X-ray crystallography, solution NMR, UV-absorbance and limited proteolysis and observed after illumination of the sample that the LOV-domain cleaves from the HTH-motif and enables HTH to bind onto DNA [80]. Moreover, they found that after light-induced cleavage of the J$\alpha$-helix , the HTH-motif is released and the LOV-domains of EL222 dimerize. Strickland *et al.* used the same principle of allostery DNA light-regulation and created an artificial DNA-photoswitch through generation of a fusion between AsLOV2-J$\alpha$ and TrpR of *E. coli* [81]. The resulting fusion between these two proteins is the so-called LOV-TAP DNA-photoswitch. They used Rsal protection assaying, CD-spectroscopy, structural modelling and small angle X-ray scattering to observe light induced changes in DNA and LOV-TAP. Through their experiments, they showed that LOV-TAP is undergoing light-induced changes in its equilibrium between states, bound or cleaved to DNA. In the light-state, LOV-TAP binds to its specific nucleotide-sequence, whereas in the dark-state, LOV-TAP is mostly cleaved from the DNA. As it was demonstrated by these authors, this equilibrium is changed solely by light. However, a detailed understanding of this dynamical effect is still lacking at the molecular level.

# 3. Methods

### 3.0.3. Sampling of ensembles

Computer simulation generates information at the microscopic level, while statistical mechanics converts this information into macroscopic terms [82]. Quantities, such as the diffusion coefficient, shear viscosity or the structure factor are state functions. Although they clearly describe the microscopic structure and dynamics of the system, their values are influenced by few variables (e.g. number of particles $N$, pressure $P$, temperature $T$), which characterize the thermodynamic state [82]. By contrast, the instantaneous mechanical state is given by atomic positions and momenta, which can be thought of coordinates in a multidimensional space, the phase space. A system of $N$ atoms has for example a $6N$-dimensional phase space.

The ensemble-average of a quantity $A$ in the canonical ensemble is written as :

$$\langle A \rangle_{ensemble} = \frac{\int d\mathbf{r} A e^{-\beta\Phi}}{\int d\mathbf{r} e^{-\beta\Phi}} \ , \tag{3.1}$$

where $\Phi$ stands for the potential energy of the system and $\beta$ stands for $\frac{1}{k_B T}$, where $k_B$ is Boltzmann's constant.

The experimentally observable 'macroscopic' quantity is the time average, which is taken over a long time interval :

$$A_{obs} = \langle A \rangle_{time} = \lim_{t_{obs} \to \infty} \frac{1}{t_{obs}} \int_0^{t_{obs}} A(t) dt \ . \tag{3.2}$$

In case, if the propability density $\rho$ does not change as the system evolves during sampling and if the ergodicity requirement is fulfilled, the ensemble average will be equal to the time-average :

$$A_{obs} = \langle A \rangle_{ensemble} \ . \tag{3.3}$$

### 3.0.4. Ensembles of statistical mechanics

In computer simulation 4 ensembles are in common use. These are the microcanonical (constant-NVE), the canonical (constant-NVT), the isothermal-isobaric (constant-NPT) and the grand canonical (constant-$\mu$VT) ensemble [82]. For each of the ensembles the previously mentioned thermodynamic variables are fixed, whereas the other thermodynamic quantities must be determined through ensemble averaging [82]. In the quasi-classical expression, the state function of the microcanical ensemble is written in case of an atomic system, as :

$$Q_{NVE} = \frac{1}{N!} \frac{1}{h^{3N}} \int d\mathbf{r} d\mathbf{p} \delta \left( H(\mathbf{r}, \mathbf{p}) - E \right) \ ; \tag{3.4}$$

where the integral stands for integration over all $6N$ phase space coordinates, $h$ is Planck's constant, $\mathbf{r}$ and $\mathbf{p}$ are coordinates and momenta. The indistinguishability of the particles is handled by $N!$ and $\delta$ is the Dirac delta function, in case of continuous states [82]. $H(\mathbf{r}, \mathbf{p})$ stands in eqn. (3.4) for the Hamiltonian. The appropriate thermodynamic potential is the negative of the entropy, defined as:

$$S = k_B \ln Q_{NVE} \ ; \tag{3.5}$$

where $S$ stands for the entropy and $k_B$ is Boltzmann's constant.

The state function of the canonical ensemble in its quasi-classical form, is given for an atomic system, as follows :

$$Q_{NVT} = \frac{1}{N!}\frac{1}{h^{3N}} \int d\mathbf{r} d\mathbf{p} \exp\left(-H(\mathbf{r},\mathbf{p})/k_B T\right) \ . \tag{3.6}$$

The appropriate thermodynamic function is the Helmholtz free energy $A$ :

$$A = -k_B T \ln Q_{NVT} \ . \tag{3.7}$$

In the canonical ensemble, exchange of energy with the heat bath is allowed, which causes that energy fluctuations are non-zero [82].

Because the energy is expressed as the sum of potential and kinetic energy, the partition function factorizes into a product of kinetic $E_{kin}$(ideal gas) and potential $\Phi(\mathbf{r})$(excess) parts, which permits to reformulate the canonical partition function as :

$$Q_{NVT} = \frac{1}{N!}\frac{1}{h^{3N}} \int d\mathbf{p} \exp\left(\frac{-E_{kin}}{k_B T}\right) \int d\mathbf{r} \exp\left(-\frac{\Phi(\mathbf{r})}{k_B T}\right) \ . \tag{3.8}$$

This results in :

$$Q_{NVT} = Q^{id}_{NVT} Q^{ex}_{NVT} \ . \tag{3.9}$$

In the ideal case of an atomic system with $U = 0$, we obtain :

$$Q^{id}_{NVT} = \frac{V^N}{N! \Lambda^{3N}} \ , \tag{3.10}$$

where $\Lambda$ stands for the thermal de Broglie wavelength :

$$\Lambda = h^2 / \left(2\pi m k_B T\right)^{1/2} \ . \tag{3.11}$$

The excess part of the canonical state function is given by :

$$Q^{ex}_{NVT} = V^{-N} \int d\mathbf{r} \exp\left(-\Phi(\mathbf{r})/k_B T\right) = \frac{Z_{NVT}}{V^N} \ . \tag{3.12}$$

with the configuration integral :

$$Z_{NVT} = \int d\mathbf{r} \exp\left(-\Phi(\mathbf{r})/k_B T\right) \ . \tag{3.13}$$

The importance of the $NPT$-ensemble is derived from the fact, that averages obtained from this ensemble reproduce experimental conditions, e.g. *in vivo* and *in vitro* experiments. The partition function of the $NPT$-ensemble in its quasi-classical form, in case of an atomic system can be written as :

$$Q_{NPT} = \frac{1}{N!}\frac{1}{h^{3N}}\frac{1}{V_0} \int dV \int d\mathbf{r} d\mathbf{p} \exp\left(-\frac{(H + PV)}{k_B T}\right) \ . \tag{3.14}$$

The corresponding thermodynamic function is the Gibbs free energy $G$ :

$$G = -k_B T \ln Q_{NPT} \ . \tag{3.15}$$

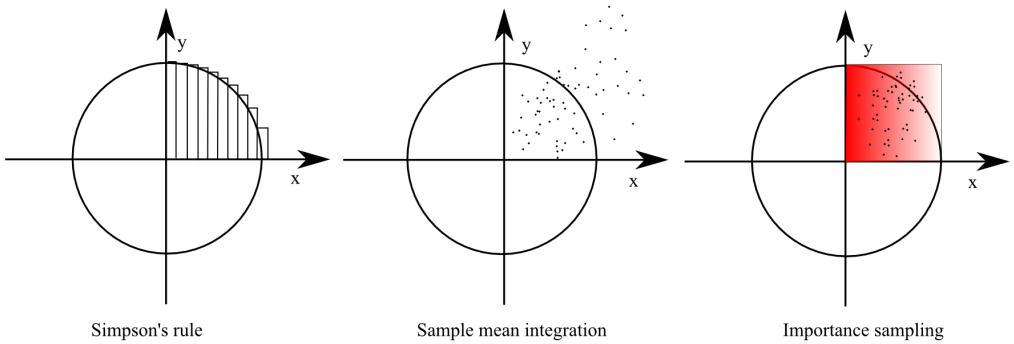


Figure 3.1.: Different schemes of numerical integration. From left to right : Simpson's rule, sample-mean integration, Importance sampling.

The configuration integral of this ensemble is :

$$Z_{NPT} = \int dV \exp\left(-\frac{PV}{k_B T}\right) \int d\mathbf{r} \exp\left(-\frac{\Phi(\mathbf{r})}{k_B T}\right) \ . \tag{3.16}$$

Finally, the grand canonical partition function is written as :

$$Q_{\mu VT} = \sum_N \frac{1}{N!} \frac{1}{h^{3N}} \exp\left(\frac{\mu N}{k_B T}\right) \int d\mathbf{r} d\mathbf{p} \exp\left(-\frac{H}{k_B T}\right) \ . \tag{3.17}$$

The appropriate thermodynamic function is in this case :

$$-\frac{PV}{k_B T} = -\ln Q_{\mu VT} \ . \tag{3.18}$$

## 3.1. Monte Carlo integration

In this section we consider the problem of numerical integration of multi-dimensional integrals. We note that the sample-mean integration technique is an alternative to the so-called Simpson's integral evaluation. This method simply measures the number of hits and misses of the area of a function within each trial, which is also shown in Fig. 3.1. Let us consider following 1-dimensional integral :

$$F = \int_{x_1}^{x_2} dx f(x) \ , \tag{3.19}$$

which we rewrite as :

$$F = \int_{x_1}^{x_2} dx \left(\frac{f(x)}{\rho(x)}\right) \rho(x) \ , \tag{3.20}$$

where $\rho(x)$ is an arbitrary probability density function [82]. To calculate this integral, we perform a number of trials $\tau$, each consisting a random-number $\xi_\tau$ from the distribution $\rho(x)$ in the range $(x_1, x_2)$. Then we obtain :

$$F = \langle \frac{f(\xi_\tau)}{\rho(\xi_\tau)} \rangle_{trials} \ . \tag{3.21}$$

In a simple case, we choose a uniform expression for $\rho(x)$ :

$$\rho(x) = \frac{1}{x_2 - x_1} \ , \tag{3.22}$$

with

$$x_1 \leq x \leq x_2 \ . \tag{3.23}$$

Then the integral F can be estimated as :

$$F \approx \frac{(x_2 - x_1)}{\tau_{max}} \sum_{\tau=1}^{\tau_{max}} f(\xi_\tau) \ . \tag{3.24}$$

For the simple one-dimensional integration, the Monte Carlo technique is not competitive with straightforward numerical methods such as Simpson's rule. However, for the multidimensional integrals of statistical mechanics, the sample mean method, with a suitable choice of $\rho(x)$, is the only useful solution. In order to calculate the configurational integral (3.13) for a system of 100 molecules in a cube of side L, we would have $10^{300}$ Simpson function evaluations to solve this integral. However, at realistic liquid densities the problem can be solved using a sample-mean integration as described in eqn. (3.24) where the random coordinates are chosen from a non-uniform distribution. For this purpose, the so-called importance-sampling techniques are used. These techniques choose random numbers from a distribution $\rho(x)$, which allows the function evaluation to be concentrated in the regions of space that make important contributions to the integral [82]. To explain this concept, let us consider a integral of the canonical ensemble :

$$\langle A \rangle_{NVT} = \int d\Gamma \rho_{NVT}(\Gamma) A(\Gamma) \ , \tag{3.25}$$

where $\rho_{NVT}$ is the normalized canonical distribution and $\Gamma$ is the related phase space vector, which defines one point in this space. The normalized canonical distribution $\rho_{NVT}$ is written as :

$$\rho_{NVT} = \frac{e^{-\beta\Phi(\mathbf{r})}}{\int d\mathbf{r} e^{-\beta\Phi(\mathbf{r})}} \tag{3.26}$$

By sampling configurations at random, from a chosen distribution $\rho$ we can estimate the integral as :

$$\langle A \rangle_{NVT} = \langle A \rho_{NVT} / \rho \rangle_{trials} \ . \tag{3.27}$$

We point out that $A(\Gamma)_{NVT}$ is significant where

$$\rho = \rho_{NVT} \ . \tag{3.28}$$

In this case

$$\langle A \rangle_{NVT} = \langle A \rangle_{trials} \ , \tag{3.29}$$

is valid.

## 3.2. Metropolis Monte-Carlo of Ising spin-systems

Let us consider a system consisting of 20 spins, where each spin can have two orientations in 1 dimension. This system has $2^{20}$ different possible configurations, while a 256 spin system would have $2^{256}$ different configurations. We point out that in a simple 2-dimensional case of $80 \times 80$ spins with $2^{6400}$ possible configurations, straight forward sampling over all different configurations and subsequent averaging is very inefficient and related to extensive computational cost. Techniques which apply stochastical averaging try to overcome this problem. In case of the Ising-spin system and particle-based systems as well as polymeric systems, the standard importance-sampling technique is the so-called Metropolis-algorithm [83]. This sampling technique can be used to compute ensemble-average quantities with low computational cost in case of Ising 2D systems. For the sake of simplicity we introduce this spin-systems, because its Hamiltonian is easily describable by nearest-neighbour interactions in form of boolean expressions. By contrast, straightforward averaging over all possible configurations would be impossible in case of larger systems. In case of 2D-Ising systems the Metropolis algorithm is formulated as follows:

1. Select one spin randomly and generate one new configuration by changing this spin.

2. Generate random-number $\xi \in [0, 1]$

3. Solve following equation and test if:

$$min\{1, \exp[-(H_{new} - H_{old})/(k_B T)]\} \geq \xi \ , \tag{3.30}$$

   where $H = -J \sum_{\langle ij \rangle} s_i s_j - B \sum_i s_i$, and B is the external magnetic field, $s_i$ is the spin at center i and each center is solely interacting with its direct 4 neighbors. J is the coupling parameter for spin-interaction;

4. if this eqn. (3.30) is fulfilled, then accept otherwise reject the new configuration

5. Return to (1).

## 3.3. Kinetic Monte Carlo of Ising spin systems

The standard method of efficiently sampling the Ising spin model is the Metropolis Monte-Carlo method. However, this method suffers from inefficient sampling of energy-states, where straightforward propagation in phase space is difficult. Such situations are given near equilibrium or within metastable states, where very low transition probabilities exist. One futher example is given by ferromagnetic couplings in ferromagnetic systems, when randomly selected pairs are different to their nearest neighbors with very low probability. Bortz *et al.* developed a new technique, which samples configurations of the Ising model much faster [84]. This algorithm is based on the fact, that there is a small number $n$ of classes of sites (neighbor pairs) classifying sites (pairs) by their probability of interchange [84]. This n-fold way chooses along all sites (pairs) with a probability which is weighted in such a way that the probability of choosing a given site (pair) is proportional to its probability of flipping (interchanging). Thus, once a site (pair) is selected, the flipping (interchanging) can be immidiately performed. Bortz *et al.* implemented the so-called ten-fold way and tested this algorithm on a lattice of $80 \times 80$-spin lattice. They divided the array of 6400 spins into 10 different classes with moveable partitions. For each spin flip in this algorithm ten numbers were calculated according to:

$$Q_i = \sum_{j=1}^{i} n_j P_j \ , \tag{3.31}$$

where $i \in [1, .., 10]$ and $P_j$ is the transition probability of spin j. In addition, the algorithm generates one random number $\xi$ and the class i of the selected spin will be defined by solving the equation:

$$Q_{i-1} \leq \xi < Q_i \ , \tag{3.32}$$

where $Q_0 = 0$.

The increment in time variable t is calculated as follows. Since $Q_{10}$ is the number of spins times the average probability that an attempt will produce a flip, the time t increment is proportional to $Q_{10}^{-1}$.

$$\Delta t = -\left(\frac{\ln \xi * \tau}{Q_{10}}\right) \ , \tag{3.33}$$

where $\xi$ is representing a random fraction, by which this event contributes to the time increment. This latter equation reflects approximately the real time of a physical system [84]. This realtime expression of eqn. (3.33) is derived as follows. Bortz *et al.* assumed that their lattice system is immersed in a bath of fixed temperature $T$. This lattice system randomly flips in space and time, while each attempted flip occurs per lattice site in time $\tau$. Moreover, they expected that this parameter $\tau$ depends on the temperature of the bath and the nature of the system but only weakly depends on the state of the system. The probability of flipping a spin on a given random attempt is $\frac{Q_{10}}{N}$ , where N is the number of lattice sites. Thus, they found that the probability of having a flip during the infinitesimal time interval $dt$ is

$$pdt = \frac{Q_{10}}{\tau} dt \ . \tag{3.34}$$

In the following, we take the different transition probablities into account which change according to changes in the time-axis. $P\Delta t$ is the probability, that no flip occurs during the time-span $\Delta t$. Moreover, the probability $P(\Delta t + dt)$ that no flip occurs at a given time $\Delta t + dt$ must be lower than the latter probability $P(\Delta t)$. However, if one flip occurs between given times $\Delta t$ and $\Delta t + dt$, then we can write for the given probability $P$ at time $\Delta t + dt$ :

$$P(\Delta t + dt) = P(\Delta t) - P(\Delta t)(Q_{10}/\tau)dt \ . \tag{3.35}$$

This equation, can be rewritten, as follows:

$$\frac{\partial}{\partial t} P(\Delta t) = (-Q_{10}/\tau) P(\Delta t) \ , \tag{3.36}$$

which has the solution:

$$P(\Delta t) = \exp(-Q_{10}\Delta t/\tau) \ , \tag{3.37}$$

with the boundary condition that $P(0) = 1$.

Since the events within the time-intervals occur in a Poisson-process, the probability $P(\Delta t)$ is equal to a uniformly distributed random number. To this end Bortz *et al.* chose a random fraction $\xi$ over the interval $[0, 1]$. With $P(\Delta t) = \xi$ they obtained following result :

$$\Delta t = -(\tau/Q_{10}) \ln \xi \ . \tag{3.38}$$

## 3.4. Molecular dynamics

The molecular dynamics simulation method samples on a deterministic trajectory through the phase space of a system through the numerical solution of classical equations of motion [85]. Thus, in this simulation method statistical correlation always exists between new and old configurations by contrast to the previously mentioned monte-carlo techniques. The equations of motions written in their most fundamental form in the so-called Lagrangian equation of motion [82] are

$$\frac{d}{dt}(\partial L/\partial \dot{q_k}) - (\partial L/\partial q_k) = 0 \ , \tag{3.39}$$

where the Lagrangian function $L(q, \dot{q})$ is defined in terms of kinetic and potential energies :

$$L = K - V \ , \tag{3.40}$$

and is considered to be a function of the generalized coordinates $q_k$ and their time-derivatives $\dot{q_k}$. If we consider a system of atoms, with Cartesian coordinates $r_i$ and the usual definitions of $K$ and $V$ then eqn. (3.39) becomes :

$$m_i \frac{\partial^2}{\partial t^2} r_i = f_i \ ; \tag{3.41}$$

where $m_i$ is the mass of atom $i$ and

$$f_i = \nabla_{r_i} L = -\nabla_{r_i} V \ , \tag{3.42}$$

is the force on that atom.
In its strict definition, the so-called Hamiltonian is defined by :

$$H(p, q) = \sum_k \dot{q_k} p_k - L(q, \dot{q}) \ , \tag{3.43}$$

where it is assumed that we can write $\dot{q_k}$ on the right as some function of the momenta $p$. For Carthesian coordinates, Hamilton‘s equations become

$$\dot{r_i} = p_i / m_i \ , \tag{3.44}$$

and

$$\dot{p_i} = -\nabla_{r_i} V = f_i \ . \tag{3.45}$$

The standard method for solution of eqns. (3.41,3.44) is the finite difference approach. This approach follows the general idea, that if molecular positions, velocities and other dynamic information at time $t$ is given, we are able to obtain information about molecular positions, velocities and related information at the time $t + \delta t$. These equations using this finite difference approach are successively solved on a step-by-step basis [82, 85]. Principally, all of these finite difference based propagators are based on a Taylor expansion of $r(t)$ about the time $t$ :

$$r(t + \delta t) = r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t) + \frac{1}{6}\delta t^3 b(t) + \ \ldots \ , \tag{3.46}$$

where $a(t)$ is denoted as the second time derivative of $r(t)$ and $b(t)$ as the third time-derivative of $r(t)$. Velocities and accelerations are estimated within the same scheme of the Taylor expansion. Principally, $r(t + \delta t)$ is an estimate and strongly depends on the value of $\delta t$. Following desirable properties shall be fulfilled by a finite-difference algorithm :

1. Fast and low-memory requirement during computation.
2. Permission of long time step $\delta t$.
3. Duplication of classical trajectory as closely as possible.
4. Conservation of energy and momentum in the microcanonical ensemble.
5. Time-reversibility.

Algorithms, which are commonly used for numerical integration and which are equivalent due to the previously mentioned properties are the Gear predictor-corrector- , the Leapfrog-, the Verlet- and the Velocity-Verlet-algorithm. For standard atomistic simulations commonly a timestep of 1 fs is used, while the usage of constraints and other methods of reduction of degrees of freedom permits even larger timesteps.

As an example, we show the Leapfrog-integrator, whose two-step equation can be expressed as follows :

$$x_{i+1} = x_i + v_{i+1/2}\Delta t \ , \tag{3.47}$$

$$v_{i+3/2} = v_{i+1/2} + a(x_{i+1})\Delta t \ , \tag{3.48}$$

where $x_i$ are the coordinates of configuration $i$, $v_i$ are the velocities of configuration $i$ and $a(x_i)$ is the acceleration related to coordinate $x_i$ of configuration $i$.

Additionally, we mention that also other integration-techniques exist which allow the use of multiple-time-stepping [139, 140]. Commonly computational restrictions on the number of digits within the numerical propagation lead to numerical noise, which further reduces time-reversibility in long-time simulations. We point out that domain-decomposition in parallel calculations and fast-fourier transformation (FFT) in parallel particle-mesh-ewald cycles do also lead to restrictions of the previously mentioned desirable characteristics of finite difference methods.

Generally, the interactions are calculated during each step-by-step integration on the basis of a reduction to pairwise-interactions, while 3-body and higher interaction-terms are neglegted. These two-body interactions are split up into the so-called bonded and non-bonded interactions :

$$U = U_{bonded} + U_{non-bonded} \ , \tag{3.49}$$

while $U_{non-bonded}$ is split up into van-der-Waals and electrostatic interactions

$$U_{non-bonded} = U_{vdW} + U_{electrostatic} \ . \tag{3.50}$$

In their original implementation, bonded interactions have been represented by constraints, which define the internal coordinates throughout the simulation. The first approach, which handles bonded interactions, uses so-called internal coordinates, e.g. the dihedral angles of a macromolecule represent an independent set of coordinates, that specify the positions of each atom. However, difficulties arise from the resulting complexity of the equations of motion and cyclic systems might be difficult to describe with this approach, e.g. disulfide bonds within proteins [86]. We point out that most molecular dynamics algorithms rely on the method of Lagrangian multipliers within the equations of motion to describe bonded interactions, e.g. the unconstrained particle positions $\vec{x_i}(t + \Delta t)$ are updated using:

$$\vec{x}_i(t + \Delta t) \longleftarrow \vec{x}_i(t + \Delta t) + \sum_{k=1}^{n} \lambda_k^{(l)} \frac{\partial \sigma_k}{\partial \vec{x}_i} \ . \tag{3.51}$$

In addition to the Lagrange-multipliers method, a multitude of constraint algorithms have been developed to describe the bonded interactions of a molecule. The so-called Settle algorithm solves the system of non-linear equations analytically for 3 constraints in constant time, which is very often used for constraining rigid water molecules, e.g. SPC/E or TIP3P water models [87, 88]. Common to all popular MD-implementations is the SHAKE algorithm, which satisfies bond geometry constraints during MD-simulations [89]. In principle this method also relies on the methodology based on Lagrange multipliers. However, Ciccotti *et al.* introduced a cost-reducing method, which is called the M-SHAKE method [90]. Another method, which scales better in computation of the constraint iteration procedure is the LINCS-method, developed by Hess *et al.* [91]. LINCS applies Lagrange multipliers to the constraint forces and solves for the multipliers by using a series expansion, to approximate the inverse of the Jacobian matrix. Finally, we point out that these previously mentioned constraint-algorithms lead to numerically unstable situations during simulation, when subjected to problems in the constraint-iteration process. Bonded harmonic-oscillator potentials are more stable during integration. These are available in nearly all standard all-atom forcefields, such as CHARMM27 or GROMOS96 [92]. In the following section, we will focus on the calculation of long-range electrostatic interaction, which represents one of the major problems in the treatment of many-body interaction. We mention, that vdW-interactions in the non-bonded term, usually are described by the so-called Lennard-Jones potential. In standard implementations, this short-ranged interaction is treated by the so-called shift function. However, electrostatics are long-ranged by nature and often lead to a manifold of artifacts and errors in simulation. Rather than the vdW-description, electrostatic descriptions by forcefields are the major subject of evaluations and discussions in present science.

## 3.5. Electrostatics

Simulation of protein-solvent systems generally involves the accurate handling of a huge amount of bonded and non-bonded many-body interactions. A key issue in such simulations is the efficient treatment of the non-bonded long-range electrostatic interactions. There are currently three prevalent methods to treat electrostatic interactions in simulations under periodic boundary conditions [82, 93]: (1) straight truncation of the Coulomb interactions at a convenient cutoff distance [82]; (2) smooth truncation of the Coulomb interactions by using a switching or shifting function [94, 95, 96, 97, 98, 99] or by including a reaction-field (RF) correction [100, 101, 102]; (3) application of lattice-sum (LS) techniques, like the Ewald summation [103] or related, computationally less-expensive methods, such as the particle-mesh Ewald (PME) method [104, 105] or the particle-particle-particle-mesh (P3M) method [106]. Cutoff truncation techniques have the huge advantage of lowering the computational costs and the effect of artificial periodicity in simulations. These properties are important requirements for performing long-time calculations of large protein systems. However, it is well established that straight truncation of the long-range electrostatics represents a severe approximation, leading to numerous artifacts, such as e.g. heating, instabilities in the numerical integration procedure as well as inaccuracies in the simulated properties [93]. To alleviate the problem, electrostatic calculation techniques based on switching and shifting functions have been developed, which go to zero smoothly and reduce in this way some of the difficulties caused by abrupt truncation. In many cases, however, they lead to an amplification of other undesirable effects [93]. For example, it has been observed that simulations with atom-based switching functions show artifacts when a too short switching range is chosen [107]. Instead of that, simulations with a shifted potential provide a too large root-mean square deviation for cutoff criteria less than 14 Å, but show reasonable fluctuations with larger cutoffs [97]. However, in a recent study Yonetani [108] found that in computer sim-

ulations of TIP3P water an unphysical layer formation can occur when a long cutoff is used. In the twin-range cutoff (TRC) technique [109] the entire electrostatic potential is shifted, to reduce truncation effects, and the interaction range is divided into two parts. The short-ranged part of the electrostatics within a neighbor-list cutoff is calculated every step and the long-ranged part is only determined with a certain neighbor-list update frequency. In a system with two well-separated length- and time-scales, the TRC technique may be adequate and can significantly accelerate the calculation in comparison to other electrostatic techniques. As a costly alternative to treat long-range electrostatics, lattice-sum LS methods can be used, which rely on the Ewald summation or related techniques [103, 104, 105, 93]. Nowadays, the Ewald summation is the most widely employed technique for calculating electrostatic interactions in a periodic or pseudo-periodic system. The basic model relies on a system of point charges, mutually interacting through the Coulomb potential. The Ewald method introduces two amendments to this simple model system, which facilitates its calculation significantly [82]. First of all, each ion is effectively neutralized at long range by the superposition of a spherical Gaussian cloud of opposite charges, centered on the ion. The Gaussian cloud acts like an ionic atmosphere, which screens the interactions between neighboring charges and ensures that they become short-ranged. The total screened potential is then calculated by summing over all the ions in the central cube as well as its periodic images on the real-space lattice. Secondly, a second set of Gaussian charges is superposed, however this time possessing the same sign as the original point charges and again located at the center of the ions. Their function is to cancel the effect of the first set of Gaussian charges and to reduce the overall potential to that resulting from the original point charges. The potential, due to these Gaussians, is obtained by solving Poisson's equation and summing the contributions as a Fourier series in reciprocal space. Finally, the complete Ewald sum requires an additional correction term known as the self-energy correction, arising from the self-interaction of the canceling distribution with its own site. This term must be subtracted from the total sum. As a result, the Ewald summation technique replaces a potentially infinite only conditionally convergent sum in real space by two finite sums, i.e. a short-ranged part that sums quickly in real space and a long-ranged part that sums quickly in Fourier space, in addition to a constant self-energy correction contribution. The particle mesh-based approaches [118] all attempt to accelerate the solution of Poisson's equation under periodic boundary conditions, using the advantages of the fast Fourier transform (FFT) for calculating discrete Fourier transforms. However, they differ in how they transform the continuous charge density, due to the sum of compensating Gaussians, onto a regular three-dimensional grid and in how they compensate for the loss of accuracy introduced in this process. In case of the PME technique the Ewald sum is taken in its original form and the complex exponentials, appearing in the reciprocal sum, are approximated by local polynomial interpolation. Although LS and RF methods rely on more or less reasonable approximations for dealing with the long-range component of the electrostatic interactions, some dependence of the calculated properties on the cutoff distance or system size has also been evidenced for these methods [110, 111, 112, 113, 114, 117, 116]. Moreover, it has been stated in several works that the long-range periodicity of the Ewald boundary conditions can artificially stabilize local equilibrium structures of peptides and small proteins by inhibiting conformational fluctuations [118]. For example, Lins and Röthlisberger [115] performed a series of MD simulations, to investigate the influence of the long-range electrostatics on the folding of the N-terminal H4-histone tail peptide. In their study they employed three different long-range electrostatic treatments, i.e. spherical cutoff, RF as well as PME technique, and analyzed their effect on the dynamical behavior of the peptide system. They concluded that, among the three tested methods, the PME technique yields the least conformational variation and tends to a structural stabilization around the initial topology of the peptide. In another study Hünenberger and McCammon [117] used an approach based on continuum electrostatics, to investigate

the nature and magnitude of periodicity-induced artifacts on the dynamics and conformational equilibria of biomolecules. In their work they considered three model applications, i.a. the unfolding of polyalanine oligopeptides, the separation of the strands of a DNA tetranucleotide, as well as the conformational fluctuations of the small hyper-thermophilic protein Sac7d from *Sulfolobus acidocaldaricus*. From these calculations, they deduced that the artificial periodicity can significantly affect the conformational equilibrium of the biomolecules, in each case stabilizing the most compact conformation. Moreover, they found that periodicity-induced artifacts can be enhanced by different system-specific factors, like e.g. a solvent of low dielectric permittivity, a non-negligible solute size compared to the size of the unit cell or a non-neutral solute. In a further study Weber *et al.* [116] investigated the consequences of artificial periodicity on the conformational equilibrium of a polyalanine octapeptide in water. To this end, they performed explicit-solvent MD simulations at different system sizes, using the P3M method for handling electrostatic interactions, and compared the results to continuum electrostatics calculations. From their investigation, they concluded that the R-helical conformation of the polyalanine octapeptide is stabilized by artificial periodicity relative to other configurations, sampled during the simulation runs, and that the stabilizing tendency increases with decreasing system sizes. In summary, all these studies demonstrate that the artificial periodicity imposed by the use of the infinite periodic (Ewald) boundary conditions in the treatment of the long-range electrostatics may significantly perturb the dynamics as well as the conformational fluctuations of peptides and small proteins in explicit-solvent simulations. However, their influence on the electrostatics of larger protein systems still needs to be assessed.

## 3.6. Molecular-dynamics simulation parameters

We discuss in this section the molecular-dynamics related parameters, which we have used to generate the results of this thesis, unless specified otherwise. We generated the all-atom MD-trajectories making use of the GROMACS molecular dynamics simulation packages 3.3.1 and 4.0.3. The interactions were described by the GROMOS96-43A1 united-atom forcefield [119, 120] in all of the simulations carried out in this work. We point out that this widely used forcefield has been tested against NMR-spectroscopic data in case of the hen egg white globular protein lysozyme in water by Soares *et al.* [121] and has been found to reproduce its solution structure and conformational behavior very well. In a recent work Todorova *et al.* [122] performed extensive MD-simulations on the 51-amino-acid protein insulin and subjected the GROMOS96-43A1-forcefield to a systematic comparison against other popular biomolecular forcefields, including the CHARMM27-, AMBER03-, OPLS- and GROMOS96-53A6-forcefields. They analyzed in detail the effect of each forcefield on the conformational evolution and structural properties of the protein and compared the results with the available experimental data. They observed that each forcefield favors different structural trends. Moreover, they found that the united-atom forcefield GROMOS96-43A1, together with the CHARMM27-forcefield, delivered the best description of the experimentally observed dynamic behavior of the chain B of insulin. In all of our simulations we employed full particle-mesh-Ewald electrostatics with a a real-space Coulomb cutoff of 1.4 nm and computed the vdW-interactions, using a Lennard-Jones-shift function with a cutoff of 1.4 nm. We updated the neighborlist with a list cutoff of 1.4 nm on a grid every 10th-timestep of simulation. We placed the proteins into cubic boxes with respect to a minimum protein-box distance of 0.7 nm. After centering the proteins in such a box, we minimized the systems using the steepest-descent algorithm or the l-BFGS algorithm. The minimization of the protein was performed with the use of shift functions describing vdW- and Coulombic- interactions under non-periodic boundary conditions. We note, that the stepwidth of the minimization

algorithm as well as the cutoffs of the Coulombic-interactions were fitted onto the minimization problem under consideration. Unless specified otherwise, interactions during minimization were treated as mentioned previously in the MD-case. We then filled the remaining free space with pre-equilibrated SPC-water molecules as well as supplementary sodium ions, if needed to render the system electrically neutral. For the simulation-systems we chose the protonation states at the physiological relevant pH-value of 8, which means fully deprotonated phosphate groups, while e. g. lysine remains fully protonated. To generate an isothermal-isobaric ensemble with a temperature of 300 K and a pressure of 1 atm, the system was equilibrated for 1 ns, unless specified otherwise. During this equilibration-phase every part of the system was coupled to a Nosé-Hoover-thermostat and a Parrinello-Rahman barostat [123]. During each MD-production run in the ns-timescale, the protein and respective co-factors, such as the FMN-chromophore, were decoupled from the thermostat, whereas the solvent and ions remained coupled. This technique is known as the *non-invasive thermostating technique*, which allows the protein to sample configurations far from equilibrium and to follow its natural dynamics under solvent-mediated thermostating control [246, 247] and will explicitly explained in chapter 9. For the numerical integration of the equations of motion, we employed the leapfrog integrator with a timestep of 1 fs and isotropic periodic-boundary conditions. In case of the signaling state simulations of the LOV-based systems, which are described in some of the following chapters, we employed the parameters of Neiss and Saalfrank [47, 50], which were determined from B3LYP-6-31G*-calculation results and through comparison with similar groups in the forcefield to reach consistency. The reliability of the parametrization of the signaling state as well as the dark-state of the LOV-based systems was tested and confirmed on related FMN-derivates by these authors on the LOV2-domain from *Adiantum capillus-veneris* without the J$\alpha$-helix [47, 50], as well as by us in various studies involving LOV1- and LOV2-domains [246, 247].

To create the initial structures of our adduct-state simulations, we employed prior to all of these simulations the respective dark-state structures and generated the cysteinyl-FMN (CFN) adduct by forming a covalent bond between the reactive cysteine-sulfur and atom FMN-C4a. We point out that the covalent linkage between the reactive cysteine and FMN is required to transmit the stress from the reaction center to the protein and trigger the protein signal, as we will demonstrate in the following through comparison between dark- and light-state simulations.

# Part II.

# Multi-scale modelling of photosensory proteins through application of Kinetic-Monte-Carlo

# 4. Introduction

## 4.1. Signal-protein complexes

Signal-protein complexes act as regulators of the signaling pathways of metazoan and bacterial cells, responding to a multitude of environmental stimuli such as light, temperature change and/or mechanical stress [124]. A family of signal proteins of major importance in mammalian cells are the Rac-proteins, which are small GTPases that are involved in cell growth, cell-membrane adhesion and cell survival [125]. Their activation is triggered by binding specific effector proteins at characteristic activation sites in the cell cycle. For example, the GTPase Rac1 is known to form aggregates with the Ser-Thr-protein kinase PAK1 at the activation site switchII, whose mutation is suspected to lead to deregulations inducing several malign tumors [126].

## 4.2. Recent large-scale simulation approaches

To describe the structural-dynamics of such protein complexes on an atomistic length scale, several computational methodologies were developed starting from the late Seventies [127]. A prominent example among those is the molecular-dynamics (MD) technique, which describes the time-evolution of many-particle systems through phase space by numerically integrating Newton's equations of motion [128]. However, since its application range for small proteins typically spans from nanoseconds up to submicroseconds, its usefulness to study signaling processes on experimental timescales is only limited [129]. In order to reach longer timescales with the MD technique, several approaches have been proposed in the past decades. One of those is the coarse-graining (CG) approach, in which the system's degrees of freedom and, thus, the number of interactions are reduced allowing the use of larger timesteps at lower computational costs. A successful CG method for protein systems is the united-atom approach, implemented e.g. in the GROMOS96-forcefield [130]. It involves the representation of all hydrogens on aliphatic carbons as single effective atomic units. Approaches undertaking a more severe coarse-graining, such as the MARTINI method [131, 132], reliably reproduce protein structures but fail in providing the correct dynamics of complex protein systems, due to their strong heterogenity at the atomistic level of description. To cope with the multiple length-scale problem of complex protein systems, several multiscale-modeling methods have been proposed, which partition a complex system in different space regions with varying degree of chemical resolution defined in an *ad hoc* fashion prior to the simulation. They rely on the idea of coupling theoretical methods with different levels of coarse-graining, i.e. quantum, atomistic, mesoscopic or continuum-scale approaches [133, 134, 135, 136, 137], within one simulation method. However, such techniques generally lack transferability, because they are specifically adapted to the nature of the physical problem under consideration, and, thus, are not suitable in reproducing the multiscale relaxation dynamics of complex protein systems far from equilibrium. To extend the scope of computer simulation techniques to such non-equilibrium situations multiple-timestepping approaches, like e.g. the reversible reference-system propagator algorithms (RESPA's) [138] or the Langevin dynamics integrator technique [139, 140], have been devised that greatly accelerate simulations of systems

based on a separation of timescales and/or potential ranges. They generally rely on a decomposition of the dynamical range of the macromolecule into slow and fast modes, which allows an efficient propagation of the slower dynamical components through the use of larger timesteps in the numerical integration procedure. With these approaches, an acceleration from 10 up to 100 times with regard to the conventional MD technique could be achieved, extending the scope of application of MD-based techniques to the microsecond or even sometimes to the submillisecond timerange for peptides and small proteins [141, 142]. Even if the gains in timescale are substantial, these techniques are still computationally very demanding, generally requiring several months of calculation on large parallel computer clusters [142, 143] or supercomputers [141, 144], and, therefore, will not be useful for the calculation of the signaling pathways of complex protein systems in the forseeable future. Other techniques, such as the shadow hybrid Monte Carlo (SHMC) method [145] or the meta-dynamics method [146, 147], enable a faster sampling of the free-energy surface of peptides and medium-sized proteins, using an effective Hamiltonian or modified potentials respectively. These techniques allow a significant computational saving with regard to MD-based approaches, however by construction are not able to reproduce the real dynamics of complex protein systems.

## 4.3. Rate-based approaches

An alternative to the dynamical methods introduced previously are rate-based approaches, which mimic the time evolution of the system's trajectory through performing uncorrelated jumps from one state to another by circumventing activation barriers with a certain probability. One effective implementation of this conception is the kinetic Monte-Carlo (KMC) method, relying on the Bortz-Kalos-Lebowitz (BKL) algorithm [84], which was described in the previous section 3.3 with the application onto the Ising-spin system. This technique permits to generate a dynamical pathway, in which the most probable events are selected with higher probability from a list of events, whose rates have been previously determined from a simplified potential model. Several authors have applied this or related KMC techniques to increase the accessible timescales in simulation studies of important protein folding [148, 149, 150] and unfolding problems [151, 152]. For example, Makarov *et al.* [151, 152] introduced a KMC-based algorithm to simulate the force-induced unfolding process of the muscle protein titin, using double-well potentials to sample the breakage and formation of H-bonds between different secondary-structure elements. In their algorithm the list of events is updated dynamically as the simulation proceeds, including all possible events associated with H-bond breakage or H-bond formation available at a particular instant of time during the simulation. The rates are calculated through an Eyring-type formula, where the energy barriers for H-bond breakage and formation are dependent on the pulling force acting on these processes. A KMC method, relying on a modified BKL algorithm, was used to select the events with probabilities proportional to the rates. With this approach, Makarov *et al.* were able to reproduce time-dependent pulling experiments with titin immunoglobulin domains based on single-molecule atomic force microscopy [153]. A major drawback of the KMC implementations discussed previously is that the rates are determined from simplified potential models, which do not reflect the full interaction spectrum of complex protein systems. In this part of the thesis we introduce a novel algorithm, based on the combination of a BKL-type KMC- and MD-algorithm. Our technique permits to calculate rates on the fly from an existing standard atomistic forcefield, optimized for protein simulations, and allows to extend the overall simulation time of conventional forcefield-based computer simulation techniques to experimental timescales.

## 4.4. Methodology of algorithm

To develop our simulation algorithm, we consider that the protein complex under consideration behaves as an infrequent-event system, in which the dynamics is characterized by occasional transitions from one state to another with long periods of relative inactivity between the transitions [154]. In this picture each state corresponds to a single energy basin and the long time between the transitions arises because the system must surmount an energy barrier to get from one state to another. This so-called relaxation phase ensures that the excess energy of the system can fully dissipate, allowing its thermal equilibration, and that the configuration becomes statistically uncorrelated to the previous one. It usually takes a few picoseconds or a few tens of picoseconds, i.e. several vibrational periods of a complex protein system and can easily be carried out using a MD-type algorithm. The rates of all possible events out of each state $i$ are calculated on the fly during each relaxation phase from a biomolecular forcefield and a list is created with them. The events are chosen according to the BKL algorithm [84], which selects the events with a certain probability and rate corresponding to the most probable pathway. In our approach we assume that the multiscale transduction dynamics of complex protein systems is dominated by rate-determining steps, which according to Makarov *et al.* [151, 148, 149] are mainly given by the breakage and formation of H-bonds in the protein system. We note that the stress concentration in these H-bonds is able to affect the energy barriers significantly, generally rendering the transitions more likely. Moreover, in our algorithm we assume that these events follow an Eyring-type mechanism, in which local fluctuations induce transitions of local events driving the subsequent global conformational change [155]. If all the previous requirements are fulfilled, the state-to-state dynamics of our protein system performs a Markov walk through state pace, which means that the transition probabilities for exiting state $i$ are independent of the history prior to entering state $i$. In this regard it is also worth emphasizing that, if the rates are calculated exactly for each of the energy states visited, the state-to-state trajectory will be similar to the one calculated from a pure MD simulation at significantly lower computational costs [154]. A detailed description of our new algorithm will be given in the following sections.

# 5. KMC-MD algorithm

To treat the different length- and time-scales of the complex protein system, we assume that the protein can be split into a network of dynamically heterogeneous entities [156], each having their individual internal relaxation processes characterized by specific relaxation times. Moreover, we consider that, by going over to the high temperature limit, the protein unfolds and undergoes a transition from inhomogeneous to homogeneous kinetics by passing through the folding-unfolding transition temperature. This is caused by the fact that in this regime the landscape of the potential energy surface flattens and the relaxation behavior of all entities become dynamically homogeneous [157]. In case of our LOV-based protein complexes, we take into account in the first stage that the dynamics in the vicinity of the FMN chromophore, triggering the initial signal, is essentially dominated by the dynamics of single residues, which relaxes on very fast timescales in the range from picoseconds up to subnanoseconds. These residues have been identified in several experimental works [49, 1], to be involved in the signaling pathway. The algorithm, describing this scale, will be designated in the following as the residue-based KMC (RB-KMC) procedure. In the second stage we consider the dynamical process between different protein domains, each relaxing on timescales in the range from nanoseconds up to seconds. These can typically be secondary structure elements or large protein entities, which can easily be identified through a fluctuation analysis made from a short MD run prior to the KMC simulation. We call the KMC algorithm, treating this phase, the inter-domain KMC (ID-KMC) procedure. Both KMC algorithms generate a dynamical trajectory through state space by sampling the path of the highest probability, guaranteed by the application of the BKL technique for the rate selection. To ensure that the trajectory satisfies the requirements of a Markov chain, the KMC walk from state to state is complemented by a MD-relaxation phase in each state, allowing the thermal equilibration and relaxation of the system within a few picoseconds. We call the overall algorithm the KMC-MD approach, whose implementation and parameters will be discussed in more detail in the following sections.

## 5.1. Residue-based KMC

This phase of the KMC algorithm treats the reorganization of the amino acids in the vicinity of the FMN chromophore, taking place in the timerange from picoseconds up to subnanoseconds. From several experimental and theoretical investigations [11, 158, 246, 249, 248, 250], it is well-known that the three amino acids Gln513, Asn492 and Asn482 are essential in triggering the signaling pathway of the AsLOV2-J$\alpha$- as well as the PA-Rac1-system. We assume that these amino acids can carry out the following rate-dependent transformation processes, i.e. (1) sidechain rotation, (2) H-bond formation and (3) H-bond breakage, which are included in the list of events of the RB-KMC algorithm. Our approach relies on the BKL algorithm and is composed of the following steps:

1. MD phase;

2. H-bond analysis;

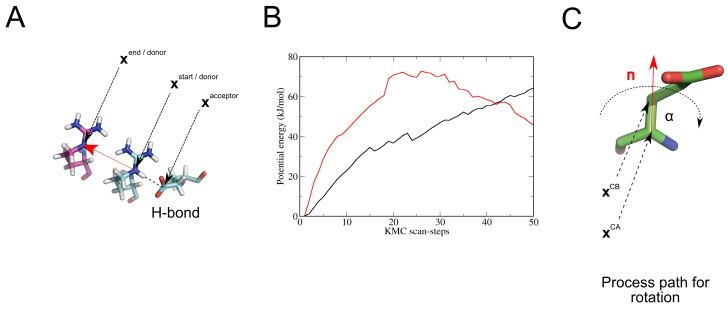


Figure 5.1.: Process path of H-bond breakage event with (B) representative corresponding potential energy profiles for dark state (black curve) and light state (red curve) of the AsLOV2-J$\alpha$ system. (C) Process path of event of sidechain rotation about an angle $\alpha$ with the axis of rotation in the direction of the CA-CB axis.

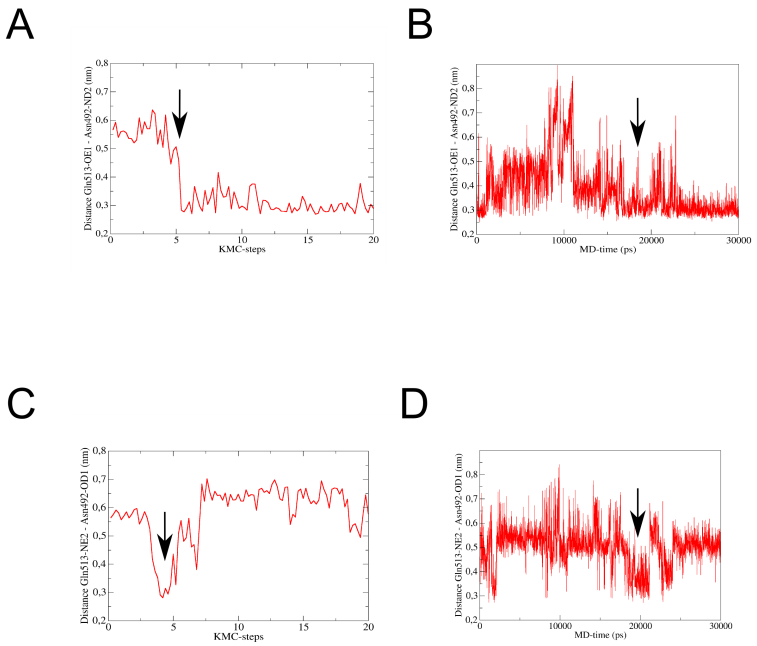


Figure 5.2.: Determination of pre-exponential factor $\bar{r}^{T\longrightarrow\infty}$ by comparing KMC-MD- to MD-simulation results for the light state of the PA-Rac1 system. Inter-atomic distance between Gln513-OE1 and Asn492-ND2 from (A) KMC-MD simulation in the RB-KMC phase as a function of KMC steps and from (B) MD simulation as a function of MD time. Inter-atomic distance between Gln513-NE2 and Asn492-OD1 from (C) KMC-MD simulation in the RB-KMC phase as a function of KMC steps and from (D) MD simulation as a function of MD time.

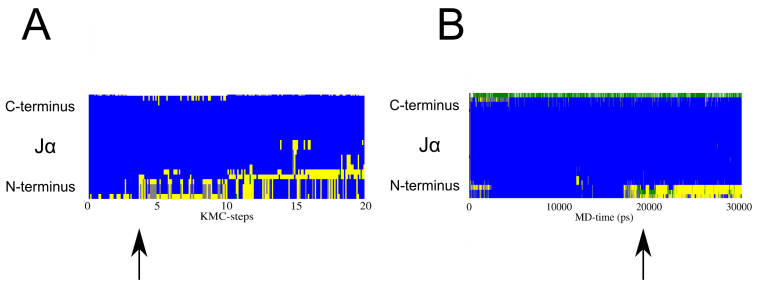


Figure 5.3.: (E) Secondary structure analysis of the Jα-helix of PA-Rac1 system from (A) KMC-MD simulation in the RB-KMC phase as a function of KMC steps and from (B) MD simulation as a function of MD time.

3. Scan of potential energy difference $\Delta\Phi^{\dagger/protein}$ along process path for each rate-determining event, which might take place in the simulation phase;

4. Calculation of rate of each event according to

$$r = \nu^{\dagger} e^{\left(\frac{-\Delta G^{\dagger}}{RT}\right)} \approx \bar{r}^{T\longrightarrow\infty} e^{\left(\frac{-\Delta\Phi^{\dagger/protein}}{RT}\right)} \quad , \tag{5.1}$$

where $\nu^{\dagger}$ and $R$ represent the pre-exponential factor and universal gas constant respectively, whereas $\bar{r}^{T\longrightarrow\infty}$ designates the rate in the infinite temperature limit $T \longrightarrow \infty$;

5. Select event $i$ from list of events by generating a uniform random number $\xi \in (0, 1]$ and by solving:

$$R_{i-1}/R_N < \xi \leq R_i/R_N \quad , \tag{5.2}$$

where $R_i = \sum_{j=1}^{i} r_j$ denotes a cumulative function with $i = 1, \ldots, N$ and $N$ the total number of events;

6. Execute selected event $i$ and update time through $t = t + \Delta t$ by generating a uniform random number $\xi \in (0, 1]$ and calculating:

$$\Delta t = -\frac{\ln \xi}{R_N} \quad ; \tag{5.3}$$

7. Return to step (1).

In order to demonstrate that the approximation in Eq. (5.1) can be used to obtain a reliable estimate of the Gibbs energy of activation $\Delta G^{\dagger}$ in our algorithm, we assume that under isothermal conditions it adopts the following form:

$$\Delta G^{\dagger} = \Delta H^{\dagger} - T\Delta S^{\dagger}, \tag{5.4}$$

where $\Delta H^{\dagger}$ and $\Delta S^{\dagger}$ denote the enthalpy and entropy of activation of the process, respectively. This permits to reformulate the rate expression in Eq. (5.1) as [159]

$$r = \bar{r}(T) e^{\left(-\frac{\Delta H^{\dagger}}{RT}\right)} \tag{5.5}$$

with

$$\bar{r}(T) = \nu^{\dagger} e^{\left(\frac{\Delta S^{\dagger}}{R}\right)}. \tag{5.6}$$

In Eq. (5.5) the temperature dependence of the pre-exponential term is generally negligible in comparison to the much stronger temperature dependence of the exponential term [161]. Using this assumption, we can re-write the rate as follows:

$$r \approx \bar{r}^{T \longrightarrow \infty} e^{\left(-\frac{\Delta H^{\dagger}}{RT}\right)} \tag{5.7}$$

with the enthalpy of activation given by [160]

$$\Delta H^{\dagger} = \Delta U^{\dagger} + p\Delta V^{\dagger} + \Delta p V^{\dagger}. \tag{5.8}$$

Here, we note that $\Delta p$ and $\Delta V^{\dagger}$ represent, respectively, the changes of the external pressure and volume of activation, whereas $\Delta U^{\dagger}$ is the change of the internal energy of activation defined as

$$\Delta U^{\dagger} = \Delta K^{\dagger} + \Delta \Phi^{\dagger}, \tag{5.9}$$

where $\Delta K^{\dagger}$ and $\Delta \Phi^{\dagger}$ designate the changes in the kinetic and potential energy of activation, respectively. The change in the potential energy $\Delta \Phi^{\dagger}$, associated with each KMC event, can further be decomposed in the following contributions:

$$\Delta \Phi^{\dagger} = \Delta \Phi^{\dagger/protein} + \Delta \Phi^{\dagger/solvent} + \Delta \Phi^{\dagger/interface}, \tag{5.10}$$

where $\Delta \Phi^{\dagger/protein}$, $\Delta \Phi^{\dagger/solvent}$ and $\Delta \Phi^{\dagger/interface}$ designate the changes in the potential of the protein, solvent and protein-solvent interface of activation, respectively. To obtain a reliable approximation of $\Delta H^{\dagger}$, we executed each event under solvent-free conditions at zero temperature. Then, we re-inserted the solvent and equilibrated it in a short canonical MD phase by freezing the degrees of freedom of the protein and adjusting the temperature related average kinetic energy of the solvent to its external value. This implementation ensured that for each executed event $\Delta K^{\dagger} \approx 0$, $\Delta p \approx 0$ and $\Delta V^{\dagger} \approx 0$, leading to

$$\Delta H^{\dagger} \approx \Delta \Phi^{\dagger}. \tag{5.11}$$

Moreover, by considering that in protein-solvent systems the potential energy contributions related to the bonded and non-bonded interactions of the protein during the event execution dominate in magnitude with regard to the potential energy changes of the solvent and protein-solvent-interface, we can finally approximate the enthalpy of activation as

$$\Delta H^{\dagger} \approx \Delta \Phi^{\dagger/protein}. \tag{5.12}$$

By computing the signaling behavior of the AsLOV2-J$\alpha$- and PA-Rac1-systems together with their mutants and comparing the results to experiments, we will show in the following that this implementation permits to reliably approximate $\Delta H^{\dagger}$ as the finite difference between the absolute maximum, i.e. the so-called transition state, and the absolute minimum prior to the transition state of the 1-dimensional potential energy profile. To evaluate the profiles of all possible events during a KMC cycle, we accomplished the scan stepwise within a scanning range $d_{scan}$ and completed each step by subsequent minimization using the Broyden-Fletcher-Goldfarb-Shannon (BFGS) algorithm [162], while constraining the donor-acceptor distance of the amino acids under consideration. The hopping range for the execution of each event was determined in an adaptive way as the absolute minimum in the potential energy after the transition state and, subsequently, the selected event was executed. Then, the algorithmic cycle was re-started with a MD phase within the NPT ensemble by imposing periodic boundary conditions.

## 5.2. Inter-domain KMC

The second KMC phase considers rate-dependent processes between different protein domains, occuring on timescales in the range from nanoseconds up to several seconds. These domains can either be secondary structure elements or larger protein entities, which can be identified as dynamically homogeneous regions of the protein by means of a fluctuational analysis prior to the KMC-MD simulation. The ID-KMC phase is accomplished by carrying out the same algorithmic steps as in the RB-KMC algorithm, introduced in section 5.1. In case of the AsLOV2-J$\alpha$ system we consider the LOV2-core and the J$\alpha$-helix as the dynamically homogeneous entities, supplemented in case of the PA-Rac1 system with the LOV2-GTPase interfacial region. Furthermore, we assume that the corresponding inter-domain relaxation is dominated by the rate-dependent processes of H-bond breakage. This latter simplification takes into account that the H-bond formation and sidechain rotation processes are much faster compared to the H-bond breakage processes and, therefore, the former processes can be neglected on the larger timescales, addressed in the ID-KMC phase. This allowed us to reduce the number of KMC cycles and, thus, the costs of the overall calculation significantly.

## 5.3. Simulation parameters and system preparation

### 5.3.1. KMC-MD parameters

We performed 5 independent simulations for each state of the AsLOV2-J$\alpha$- and PA-Rac1-system to demonstrate reproducibility of the simulations. The length of the MD phase between KMC steps of each KMC-MD simulation was determined through measurement of the statistical inefficiency of the MD-simulation algorithm in case of the AsLOV2-J$\alpha$- and the PA-Rac1-system. This latter parameter represents the length of the MD trajectory, which is required to get statistical independent configurations. To compute this quantity, we used the method of Fincham *et al.* [163] and evaluated a length of the MD phase of 5 ps. Another important parameter, which needs to be determined prior to the calculation, is the maximum scanning range $d_{scan}$ for the processes of H-bond formation and H-bond breakage as well as the maximum angle of rotation $\alpha_{scan}$ for the sidechain rotation processes. To this end, we performed independent MD simulations on the AsLOV2-J$\alpha$ system and evaluated for both H-bonding processes an optimal value of 0.35 nm from the average translational fluctuations, whereas for the process of sidechain rotation we estimated a value of 20 degrees. As illustrated in the Fig. 5.1, we determined the process path of each possible H-bond breakage and H-bond formation event $i$, which could take place in the protein during KMC-MD run, through the following equation:

$$\mathbf{x}_i^{end/donor} = \mathbf{x}_i^{start/donor} + d_{scan} \frac{\left(\mathbf{x}_i^{start/donor} - \mathbf{x}_i^{acceptor}\right)}{\left|\mathbf{x}_i^{start/donor} - \mathbf{x}_i^{acceptor}\right|}, \tag{5.13}$$

where $\mathbf{x}_i^{start/donor}$ and $\mathbf{x}_i^{end/donor}$ denote the starting and final coordinates of the donor, whereas $\mathbf{x}_i^{acceptor}$ is the coordinate of the acceptor kept at a fixed position in space during the execution of the event. In Fig. 5.1B we show corresponding representative potential energy profiles for a H-bond breakage process in case of the dark and light state of the AsLOV2-J$\alpha$-system. Moreover, the process path of the sidechain rotation in 3-dimensional space was defined through the following operation of rotation by an angle $\alpha$ about an axis in the direction of the unit vector $\mathbf{n} = (n_1, n_2, n_3)^T$:

$$R_{\mathbf{n}}(\alpha)\mathbf{x} = \mathbf{n}(\mathbf{n} \cdot \mathbf{x}) + \cos\alpha(\mathbf{n} \times \mathbf{x}) \times \mathbf{n} + \sin\alpha(\mathbf{n} \times \mathbf{x}), \tag{5.14}$$

where the rotation matrix is given by

$$R_{\mathbf{n}}(\alpha) = \begin{pmatrix} \cos\alpha + n_1^2(1-\cos\alpha) & n_1 n_2(1-\cos\alpha) - n_3\sin\alpha & n_1 n_3(1-\cos\alpha) + n_2\sin\alpha \\ n_2 n_1(1-\cos\alpha) + n_3\sin\alpha & \cos\alpha + n_2^2(1-\cos\alpha) & n_2 n_3(1-\cos\alpha) - n_1\sin\alpha \\ n_3 n_1(1-\cos\alpha) - n_2\sin\alpha & n_3 n_2(1-\cos\alpha) + n_1\sin\alpha & \cos\alpha + n_3^2(1-\cos\alpha) \end{pmatrix}. \tag{5.15}$$

More specifically, in our algorithm we chose in case of the amino acids Gln513, Asn492 and Asn482 in vicinity to the FMN the axis of rotation in the direction of the CA-CB axis (see Fig. 5.1C) and performed a stepwise rotation up to the maximum angle of rotation $\alpha_{scan}$. We note that each execution of an event along a process path was complemented by a subsequent MD-relaxation phase, which reduces the dependence of our algorithm on the process path realization. As we will demonstrate in the following, this implementation was found to provide reliable results for the AsLOV2-J$\alpha$- and PA-Rac1-systems under consideration and to be the most effective procedure with regard to computational expense. To determine the pre-exponential factor $\bar{r}^{T\longrightarrow\infty}$ in Eq. (5.1), we considered that, as described in section 5, the protein unfolds and undergoes a transition from inhomogeneous to homogeneous kinetics by passing through the folding-unfolding transition temperature with increasing temperature, which causes that in the infinite temperature limit all events possess identical rates designated by $\bar{r}^{T\longrightarrow\infty}$. We point out that the latter assumption has been confirmed by several theoretical and experimental works, which will be discussed in the following. For example, Leite *et al.* [157] have studied fluctuations of protein folding kinetics using a lattice-based dynamic Monte Carlo algorithm by examining the ratios of the moments of the first-passage time. They found that at high temperatures, due to the large thermal fluctuations of the dynamical entities, local details are smeared out and the different kinetic paths sense roughly similar barriers, resulting in single-exponential kinetics. Moreover, Zhang *et al.* [164] observed by performing neutron scattering spectroscopy experiments and MD simulations on the hen egg white protein lysozyme that in the high temperature regime the lifetime of the H-bonds between the protein and the hydration water decreases substantially, which causes an increased mobility of the protein in the water environment and leads to its subsequent unfolding. Ultimately, in the high temperature limit the protein looses the remaining dynamical heterogeneity related to its interaction with the hydration water and adopts a homogeneous kinetic behavior, which is reflected by an Arrhenius-type temperature dependence of the inverse diffusion constant [164]. Finally, we emphasize that the assumption of homogeneous kinetics in the high-temperature limit has also been successfully employed by Makarov *et al.* [151, 148, 149], to reproduce pulling experiments on the muscle protein titin [153], using a KMC method based on a modified BKL algorithm. Despite the ineffective infrequent-event sampling of the pure MD technique compared to the KMC-MD technique, we consider that the high local stress concentration in the binding pocket resulting from the formation of the cysteinyl-FMN adduct allows to reduce the activation barriers for the events, involving the residues near to the FMN chromophore in the AsLOV2 core. This ensures that the signaling pathway of both simulation methods can be assumed to be similar in the early stages. As a consequence, we can estimate the pre-exponential factor $\bar{r}^{T\longrightarrow\infty}$ by identifying similar characteristic events from the inter-atomic distances of the amino acids adjacent to the FMN chromophore, calculated with the KMC-MD and MD techniques, and adjusting the KMC-MD- onto the MD-time. To demonstrate this, we consider in the Figs. 5.2A and 5.2B the inter-atomic distances between Gln513-OE1 and Asn492-ND2 of the PA-Rac1 system, obtained with the MD as well as KMC-MD techniques respectively. We compare these results with the corresponding ones for the distance between Gln513-NE2 and Asn492-OD1, visualized in the Figs. 5.2 C and 5.2D. From the graphs, we deduce that at the simulation stage, marked with a black arrow, a permanent coupling through H-bond formation between Asn492 and Gln513 does occur, which triggers the N-terminal disruption of the J$\alpha$-helix of the PA-Rac1

system. The latter finding is confirmed by comparing the secondary structure analysis of the Jα-helix of the PA-Rac1 system, obtained with both simulation techniques and shown in the Figs. 5.3A and 5.3B. By considering that the early signaling pathways of both simulation techniques are similar, we conclude that these events must be identical and that, therefore, the $5^{th}$ KMC step must correspond to a real time of about 20 ns. Using this procedure, we obtained in case of the PA-Rac1 system a value for the pre-exponential factor of $3.94 * 10^7$ $s^{-1}$. In case of the AsLOV2-Jα system, we used the same pre-exponential factor as for the PA-Rac1 system by taking into account that both signaling pathways are triggered through cysteinyl-FMN-adduct formation in the AsLOV2 core and, thus, they should be similar in the early stages. The MD-parameters were taken according to the data, which is described in the previous section 3.6.

## 5.3.2. System preparation

As starting structures for the dark-state simulations, we used the dark-state crystal structure of the PA-Rac1- (PDB-code: 2WKP) [72] and AsLOV2-Jα system (PDB-code: 2V0U) [55], visualized in Fig. 5.4 and Fig. 5.5. To create the initial structures of our adduct-state simulations, we followed the scheme described in section 3.6.

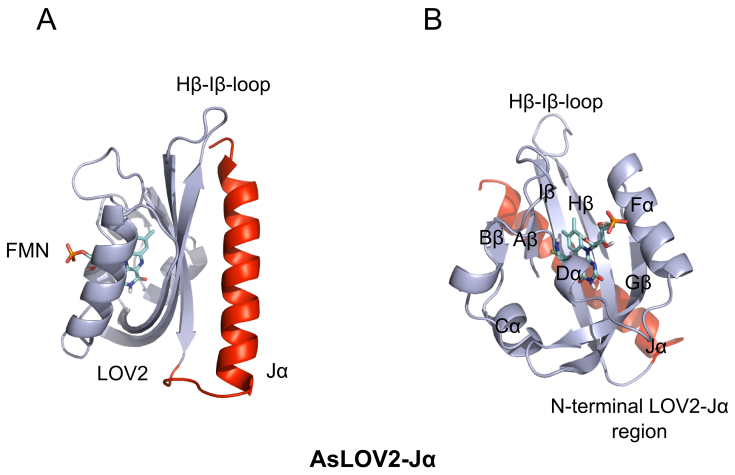


Figure 5.4.: Crystal structure of AsLOV2-Jα system (PDB-code: 2V0U), determined through X-ray diffraction by Halavaty *et al.*. (A) Side-view of AsLOV2-Jα system with FMN chromophore. (B) Front-view of AsLOV2-Jα system with different secondary-structure elements.

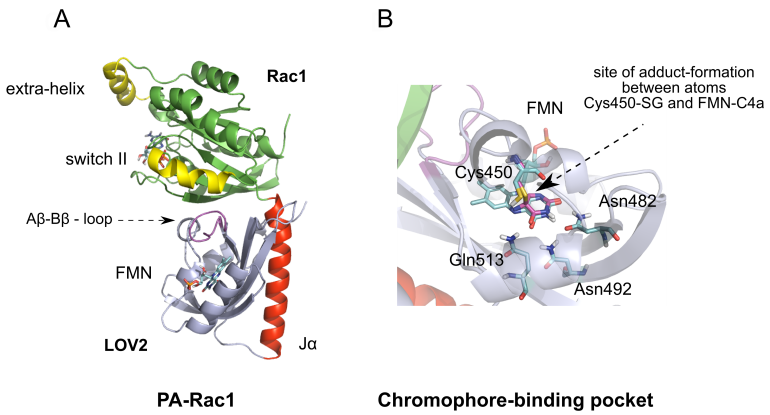


Figure 5.5.: Crystal structure of PA-Rac1 system (PDB-code: 2WKP), determined through X-ray diffraction by Wu *et al.*. (A) PA-Rac1 with activation site switchII on Rac1 enzyme connected through LOV2-core via J$\alpha$-helix. (B) Amino-acid environment of FMN chromophore in AsLOV2-core [Gln513 (I$\beta$), Asn492 (H$\beta$), Asn482 (G$\beta$)].

# 6. Results and discussion

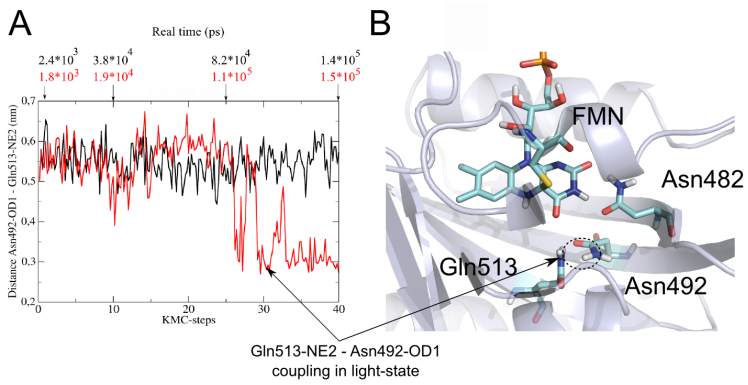


Figure 6.1.: Inter-atomic distances and representative structures of the AsLOV2-J$\alpha$ system, obtained by using KMC-MD and non-invasive MD methods. (A) Inter-atomic distance between Gln513-NE2 and Asn492-OD1 from KMC-MD simulation in the RB-KMC phase of the light-state (red curve) and dark-state (black curve) as a function of KMC steps. (B) Amino-acid configuration around FMN chromophore of the light-state from KMC-MD simulation after 40 RB-KMC steps.

To demonstrate the reliability and potential of our KMC-MD approach, we start by considering the structural-dynamical changes of the AsLOV2-J$\alpha$ system in the dark and light-state on the nanosecond timescale. In Figs. 6.1, 6.2 and 6.3 we display the results for the amino-acid configuration around the FMN chromophore and its consequences on the peripheral J$\alpha$-helix, obtained with our KMC-MD method, and compare them to the results, obtained from a 20 ns MD-simulation run on the same system. To begin, we show in Fig. 6.1A the distances between Gln513-NE2 and Asn492-OD1 for the dark and the light-state in the RB-KMC-MD phase. In case of the light-state curve, we see a clear drop in the distance at the $26^{th}$-KMC step indicating H-bond formation, whereas the dark-state curve remains at a constant distance of 0.55 nm up to the end of this phase. To further illustrate the coupling between Gln513 and Asn492 in the KMC simulation, we show a representative configuration after 40 RB-KMC steps in Fig. 6.1B. We observe a clear H-bonding between Gln513-NE2 and Asn492-OD1, which is a result of the rotation of the sidechain of Gln513. This finding is confirmed by our recent theoretical investigation [246] and various experimental works [10, 158, 56, 46], which provided evidence for the important role of Gln513 and its coupling with Asn492 in triggering the signaling pathway of the AsLOV2-J$\alpha$ domain. In particular, Nash *et al.* prove through mutational analysis that the Gln513 switch is directly implicated in the subsequent detachement of the J$\alpha$-helix from the LOV-core [56], which is supported by our recent MD simulations on the AsLOV2-J$\alpha$ system [246]. By contrast, we see in Fig. 6.2A in the corresponding MD-simulation results of the AsLOV2-J$\alpha$-light-state that no coupling occurs between Gln513-NE2 and Asn492-OD1. This is due to the limited MD time of 20 ns, in which the Gln513 switch and H-bond formation does not take place. In Fig.

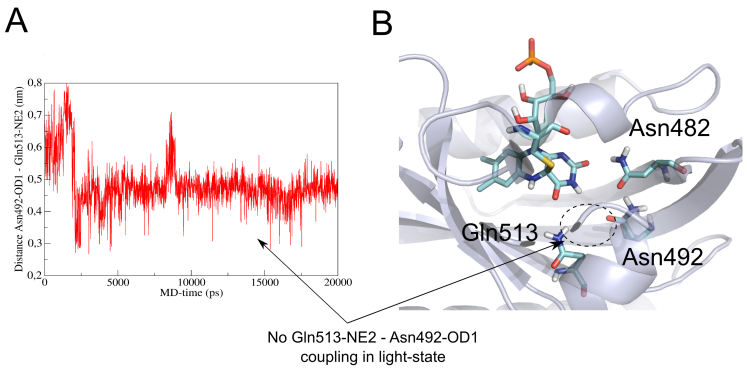


Figure 6.2.: Inter-atomic distances and representative structures of the AsLOV2-J$\alpha$ system, obtained by using KMC-MD and non-invasive MD methods. (A) Inter-atomic distance Gln513-NE2 and Asn492-OD1 of the light-state from MD simulation as a function of MD time. (B) Amino-acid configuration around FMN chromophore of the light-state after 20 ns of MD simulation.

6.2B we display a representative configuration of the chromophore environment after 20 ns of MD simulation, which shows that the amino acids Gln513 and Asn492 remain disconnected. We explain the discrepancy between the KMC-MD- and MD-simulation results by the difference in timescales accessed by both simulation approaches, which amounts to 130 ns at the end of the RB-KMC-MD phase. We point out that due to the higher computational expense per timestep the MD-simulation approach significantly depends on the quality of the starting structure, as can be deduced by comparing the current MD-simulation results based on a crystal starting structure with our previous MD simulation using a refined NMR-solution structure [246, 51]. Next, we display in Fig. 6.3A the consequence of the H-bond coupling between Gln513 and Asn492 onto the tertiary structures of the AsLOV2-J$\alpha$ system at the last KMC step of the RB-KMC-MD phase. By comparing the structures of the dark- and light-state, we see that the N-terminal region of the J$\alpha$-helix in the light-state (bright red) is disrupted and bended, whereas in the dark-state (dark red) it is similarly arranged as in case of the starting structure. The disruption of the J$\alpha$-helix is caused by the coupling between the I$\beta$- and H$\beta$-strands, mediated through the H-bond formation between the residues Gln513 and Asn492, as shown in the Figs. 6.1A and B. By contrast, we note that the tertiary structure of the light-state after 20 ns of MD simulation in Fig. 6.3B resembles the starting crystal or average dark-state structures. From this fact, we deduce that suitable coupling between Asn492 and Gln513 is a crucial step for N-terminal disruption of the J$\alpha$-helix, because no coupling occured in case of the MD simulation scheme, as demonstrated in the Figs. 6.2A and B. This is confirmed by Nash *et al.* [56], who performed CD-spectroscopical- and HSQC-NMR-experiments on the Gln513Asn-AsLOV2 mutant. Through their experiments they demonstrated, that this mutant does not exhibit light-state signaling behavior, which might be explained by the smaller sidechain of Asn513 in comparison to Gln513. To study this aspect, we performed two independent KMC-simulations with the Gln513Asn-AsLOV2 mutant and found that already after a few simulation steps H-bond coupling occurs between Asn513 and Asn492. However, we observed that this coupling does not lead to a disruption of the J$\alpha$-helix from the LOV2-core due to the shorter sidechain of Asn513, as shown in Fig. 6.4 and Fig. 6.5, which is consistent with the experiment of Nash *et al.*. Next, we continue with the analysis of the results from the subsequent ID-KMC-MD phase, which follows the RB-KMC-MD phase. In Fig. 6.6A,

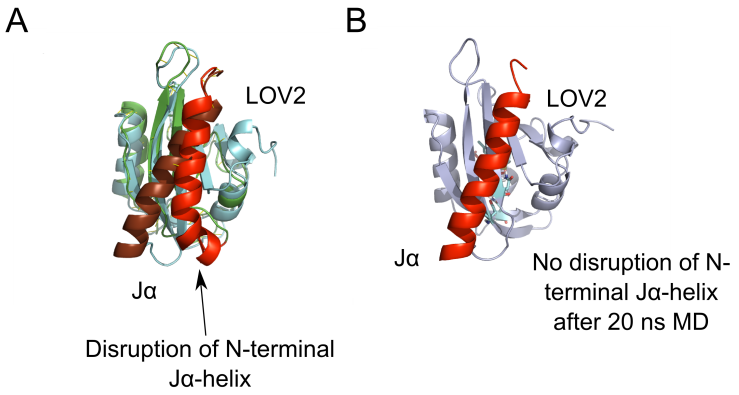


Figure 6.3.: Inter-atomic distances and representative structures of the AsLOV2-J$\alpha$ system, obtained by using KMC-MD and non-invasive MD methods.(A) Overlaid structures of dark and light-state of AsLOV2-J$\alpha$ system after 40 RB-KMC steps. (B) Final structure after 20 ns of MD simulation.

we show the final structures from two independent KMC-simulation runs of the dark-state. Both structures have been overlaid using the smallest root-mean-square deviation between the atoms. We observe that the J$\alpha$-helix remains attached to the LOV-core until the end of the ID-KMC-MD phase, in which a total simulation time of 149 $\mu$s and 624 s is reached respectively. In contrast to this, we see that the J$\alpha$-helix in the structures of the two light-state simulations in Fig. 6.6B is disrupted, which correspond to a final simulation time of 330 $\mu$s and 127 s respectively. From these observations, we conclude that in the light-state the signal propagates from the amino acids near to the FMN chromophore to the peripheral AsLOV2-J$\alpha$-interface region resulting in the detachment of the J$\alpha$-helix from the LOV-core, whereas in the inactive dark-state the J$\alpha$-helix remains attached to the LOV-core. This behavior is confirmed by several experimental works using X-ray-, 2D-HSQC-NMR- and FTIR-techniques [165, 55, 53, 54, 158]. For example, Harper *et al.* demonstrated through 2D-HSQC-NMR experiments that CFN-adduct formation induces the disruption of the J$\alpha$-helix from the LOV2-core [53]. Nakasone *et al.* performed time-resolved transient-grating and transient-lens experiments to determine diffusive changes of the LOV2-J$\alpha$ system of *Arabidopsis* upon light excitation [129]. In their paper, they speculated that complete unfolding of the J$\alpha$-helix takes place after cleavage from the LOV2-core. To investigate the detachment and subsequent unfolding processes of the J$\alpha$-helix in more detail, we visualize in the Figs. 6.6C and D the secondary structure analysis of the J$\alpha$-helix from the AsLOV2-J$\alpha$ system in the dark and light-state over the entire simulation run, obtained with the KMC-MD method using the crystal structure of Halavaty *et al.* [55] as starting structure. We compare these results with the ones obtained with the MD-simulation technique, using the refined NMR-solution structure of Harper *et al.* [53] in the Figs. 6.7A and B as well as with the crystal structure of Halavaty *et al.* [55] in Fig. 6.7D. While in the dark state the J$\alpha$-helix remains folded with both simulation strategies, we see that the length of the folded sequence of the J$\alpha$-helix decreases in the light-state KMC-MD simulation as well as MD simulation with the NMR-solution structure, whereas in case of the MD simulation with the crystal starting structure this behavior is not observed. By further comparing these results with the secondary structure analysis from the MD-simulation run of the isolated J$\alpha$-helix in Fig. 6.7C, we infer that after the coupling of the H$\beta$- and I$\beta$-strands the J$\alpha$-helix partially unfolds in the N-terminal region and after its complete cleavage

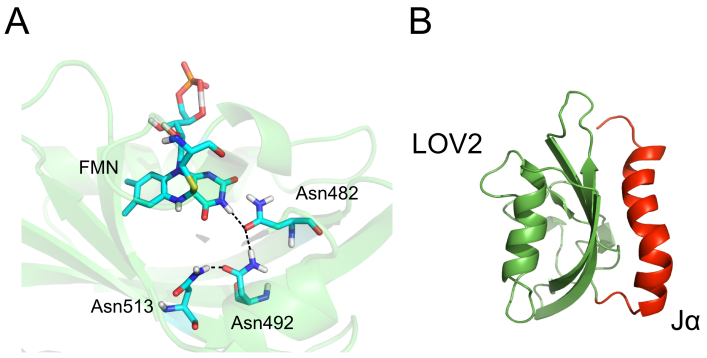


Figure 6.4.: (A) Final amino-acid configuration in vicinity of the FMN chromophore, (B) final overall structure from Gln513Asn-AsLOV2 mutant in the light-state obtained with KMC-MD method.

from the LOV-core the unfolding continues to a minor extent at the C-terminus. We explain the partial unfolding of the J$\alpha$-helix by the fact that no force is excerted by the C-terminal region of the LOV-core, which could completely unfold the J$\alpha$-helix after full cleavage. This causes that the J$\alpha$-helix of the AsLOV2-J$\alpha$ system in the disrupted signaling state behaves similarly as the isolated J$\alpha$-helix in solution. This finding is confirmed by the FTIR measurements of Alexandre *et al.* [46], who demonstrated that the J$\alpha$-helix unfolds only partially upon illumination. We point out that the partial unfolding at the C-terminal end of the J$\alpha$-helix might be a consequence of the high mobility of the J$\alpha$-helix at this free chain end, caused by the missing kinase enzyme. This is confirmed by the fact that this behavior is not observed in the PA-Rac1 system discussed in the following, where such kinetic effects are excluded. The importance of the force at the C-terminus for the complete unfolding of the J$\alpha$-helix is demonstrated by the mutational experiments of Harper *et al.* [54], who performed the point mutation Ile539Glu at the C-terminus of the J$\alpha$-helix. To study the consequence of this mutation on the secondary structure of the J$\alpha$-helix, we consider next in Figs. 6.8 and 6.9 the results of the KMC-MD simulation of the Ile539Glu-AsLOV2-J$\alpha$ mutant in the dark-state form. From the secondary structure analysis of the sequential range of the J$\alpha$-helix, we deduce that at the $46^{th}$-KMC step the J$\alpha$-helix looses its secondary structure, which is due to the formation of the salt bridge between Glu539-OE2 and the N-terminal Lys416-NZ (see snapshot in Fig. 6.8). This is in agreement with the observations of Harper *et al.* [54], who found that complete unfolding of the mutated J$\alpha$-helix takes place even in the dark-state form. We conclude from these observations that, additionally to the detachment process of the J$\alpha$-helix, our KMC-MD method reproduces the experimental findings about its subsequent unfolding in a reliable way. To study and understand in more detail the different disruption times, we compare in Fig. 6.10 the tertiary structures and APBS-electrostatic surfaces in Fig. 6.11 of the AsLOV2-J$\alpha$-interfacial region and the J$\alpha$-helix from the crystal structure of Halavaty *et al.* [55] as well as the NMR-solution structure of Harper *et al.* [53]. We conclude from the tertiary structures of the AsLOV2-J$\alpha$ surfaces that the J$\alpha$-helix in case of the crystal structure of Halavatay *et al.* [55] is longer and has a different orientation on the $\beta$-sheet of the AsLOV2-core than in case of the solution structure of Harper *et al.* [53]. Moreover, we deduce from the APBS surfaces that the AsLOV2-J$\alpha$-interfacial region and the J$\alpha$-helix in the former case are less polar and therefore more hydrophilic, which leads to a reduced attractivity for water to enter the AsLOV2-J$\alpha$ interface and to a lower tendency to dissociate both entities. This causes

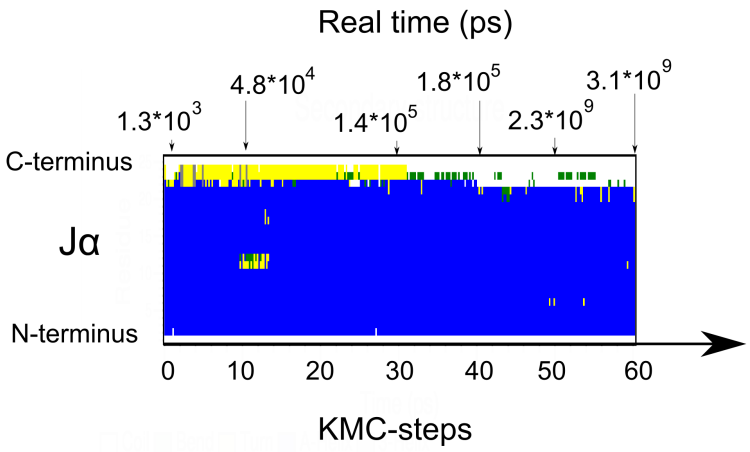


Figure 6.5.: Secondary-structure analysis of the J$\alpha$-helix as a function of KMC steps and real time from Gln513Asn-AsLOV2 mutant in the light-state, obtained with KMC-MD method.

that the AsLOV2-J$\alpha$ configuration from the crystal is thermodynamically more stable and thus has a higher disruption time than the one from solution. Finally, from the Fig. 6.6D we can estimate the time for the dissociation of the J$\alpha$-helix from the AsLOV2-core and its subsequent partial unfolding to take place in the range from 100 ns up to 300 $\mu$s. This can be correlated with the experimental timerange of 300 $\mu$s - 1 ms for these events, determined by Nakasone *et al.* [129] for the LOV2-domain from phototropin1 of *Arabidopsis*, using transient-grating methods. Thus, we conclude from the previous investigation that the KMC-MD method reproduces the multiscale relaxation behavior of the AsLOV2-J$\alpha$ system after CFN-adduct formation well and permits to extend the computationally accessible simulation times to experimental timescales.

In the next part we study the usefulness of the KMC-MD method, to describe the multiscale relaxation dynamics of complex photoenzymes. As an example, we consider the signaling behavior of the photoactivable Rac1 PA-Rac1 system, where the Rac1 enzyme has been fused with the AsLOV2-J$\alpha$ photoswitch. We start by analyzing the structural-dynamical changes in the RB-KMC phase of the amino acids in vicinity of the FMN chromophore, suspected to be involded in the primary steps of the signaling pathway of PA-Rac1 from prior MD simulations on the same system. In Fig. 6.12A we plot the distances between Gln513-OE1 and Asn492-ND2 from the dark and light-state, obtained with the KMC-MD method, as a function of KMC steps and real time. We see that after the $6^{th}$-KMC step the light-state curve performs a severe drop from 0.6 nm to H-bond distance at around 0.3 nm, whereas the dark-state curve undergoes a stable fluctuation around this value throughout the whole RB-KMC phase. Similar time behavior is observed for the quantity after an initial relaxation phase in MD simulations on PA-Rac1, as shown in Fig. 6.12B, as well as in KMC-MD simulations with the isolated AsLOV2-J$\alpha$ system in Fig. 6.1A. In particular, we note that the H-bond formation process in the KMC-MD simulation of the light-state of PA-Rac1 takes place at a simulation time of 20 ns, which is in agreement with the time range of 11 ns - 21 ns found in the corresponding MD simulation. Next, we compare in the Figs. 6.12C and D the distance between Gln513-NE2 and the carbonyl-oxygen FMN-C4=O of the dark and light-state, obtained with the KMC-MD- and MD-method respectively. We notice

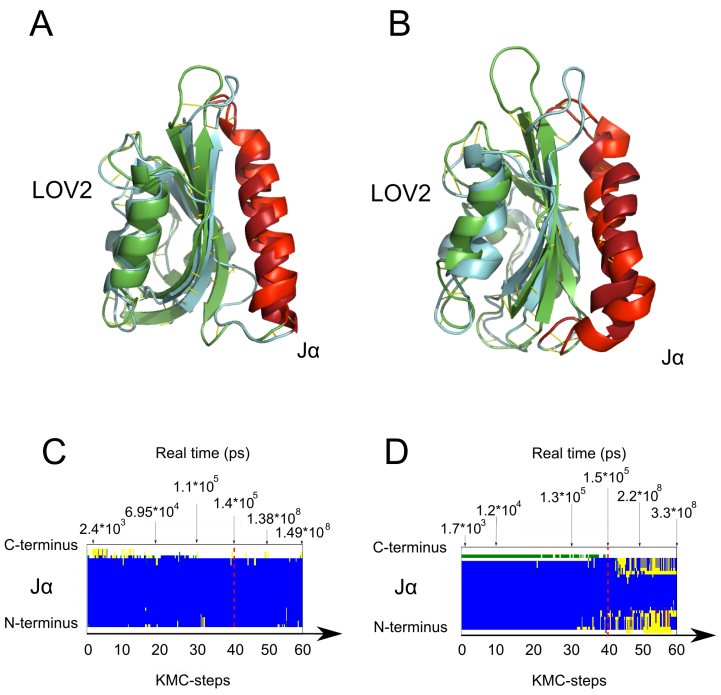


Figure 6.6.: Final structures of the AsLOV2-J$\alpha$ system in the dark-state (A) and light-state (B) after the ID-KMC phase at the end of KMC-MD simulation. Secondary-structure analysis of the J$\alpha$-helix of the AsLOV2-J$\alpha$ system in the dark-state (C) and light-state (D) as a function of KMC steps and real time, obtained by using KMC-MD method.

that, while in case of the dark-state the curves remain stable at H-bond distance, the light-state curves fluctuate at increased distance with large magnitude. To further illustrate the changes in the amino-acid environment of the FMN chromophore, we show in the Figs. 6.13A and B the respective representative snapshots from the final configurations of PA-Rac1 in its dark- and light-state form after the RB-KMC phase in comparison to the final configuration of PA-Rac1 in the light-state from MD simulation in Fig. 6.13C. From these graphs, we deduce that upon CFN-adduct formation the Gln513, located on the I$\beta$-strand, detaches from the carbonyl-oxygen FMN-C4=O and forms a H-bond with the adjacent Asn492 on the H$\beta$-strand, which results in the coupling between the I$\beta$- and H$\beta$-strands. We point out that this process has been found by us to be crucial in triggering the signaling pathway of the AsLOV2-J$\alpha$- and Vivid-LOV-photoswitches, as demonstrated in the previous section and in the MD-simulation studies of Refs. [246, 250]. We conclude from our discussion that it is also well-reproduced by our RB-KMC-simulation algorithm, introduced in section 5.1. To assess the consequences of the initial process near the FMN chromophore on the tertiary structure of the PA-Rac1 system, we further visualize in the Figs. 6.14A and B the final configurations of the LOV2-J$\alpha$ domain from PA-Rac1 in its dark- and light-state form, obtained after the ID-KMC phase and MD simulation respectively. We deduce from these graphs that the coupling between the I$\beta$- and H$\beta$-strands causes a N-terminal

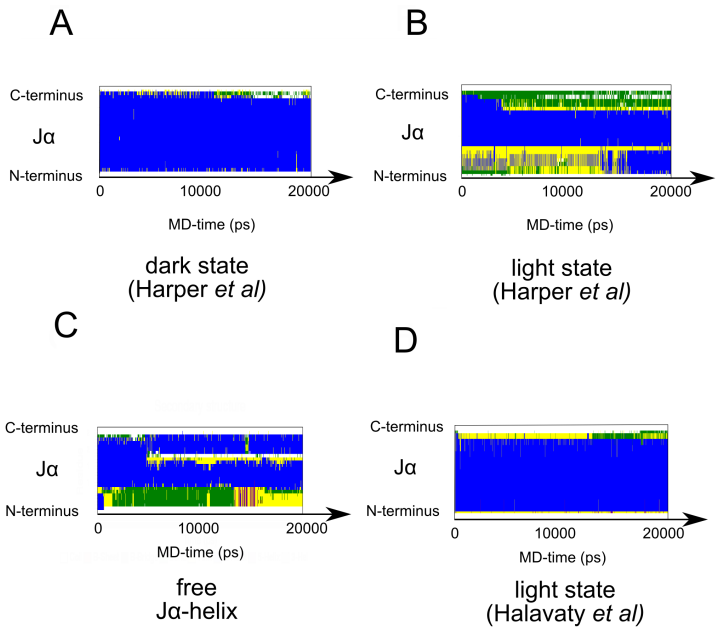


Figure 6.7.: Secondary-structure analysis of the J$\alpha$-helix of the AsLOV2-J$\alpha$ system in the dark-state (A) and light-state (B) as a function of MD steps, obtained by using MD method in conjunction with the NMR-solution structure of Harper *et al.* [54] as initial structure. Secondary-structure analysis of the J$\alpha$-helix of the AsLOV2-J$\alpha$ system in the free state (C) and light-state (D) as a function of MD steps, obtained by using MD method in conjunction with the X-ray crystal structure of Halavaty *et al.* [55] as initial structure.

disruption of the J$\alpha$-helix in the light-state of the PA-Rac1 system, which is not observed in the dark-state. Next, we analyze the impact of this disruption on the functional activity of the Rac1 enzyme. To this end, we display in the Figs. 6.15A and B the final structures from the ID-KMC phase of PA-Rac1 in the dark and light-state. From Fig. 6.15B, we deduce that in the light state the A$\beta$-B$\beta$-loop detaches from the activation site switchII on the GTPase, which is triggered by the N-terminal disruption of the J$\alpha$-helix. This causes that the steric inhibition from the LOV2-core is relieved and the activation site of Rac1 becomes accessible to interactions with its effector domains, like e.g. PAK1 [166, 167, 168]. By contrast, we infer from Fig. 6.15A that in case of the dark-state the switchII region is blocked by the A$\beta$-B$\beta$-loop, which causes that the Rac1 enzyme is inactive under dark-state conditions. To elucidate the reasons for this behavior, we consider in the Figs. 6.16A and B the APBS-electrostatic surfaces from the final dark- and light-state structures of PA-Rac1. In case of the dark state we see that the switchII region as well as the A$\beta$-B$\beta$-loop possess a high negative elecrostatic surface charge, which causes that both structural elements repell each other and are subjected to a high tension. This tension is released upon illumination through the N-terminal disruption of the J$\alpha$-helix from the LOV2-core, as can be concluded from the reduced charge concentrations at the same sites observed in Fig. 6.16B. Ultimately, this leads to the detachment of the A$\beta$-B$\beta$-loop from the switchII

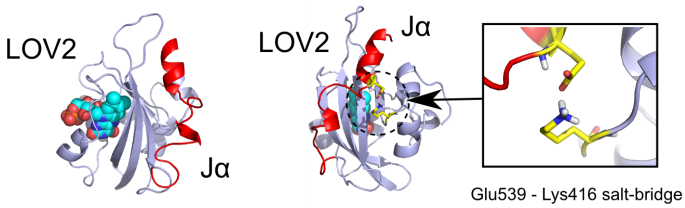


Figure 6.8.: (A) Side-view of the final overall structure, (B) alternative side-view of the final overall structure with snapshot of Glu539-Lys416 salt bridge from Ile539Glu-AsLOV2 mutant in the dark-state, obtained with KMC-MD method.

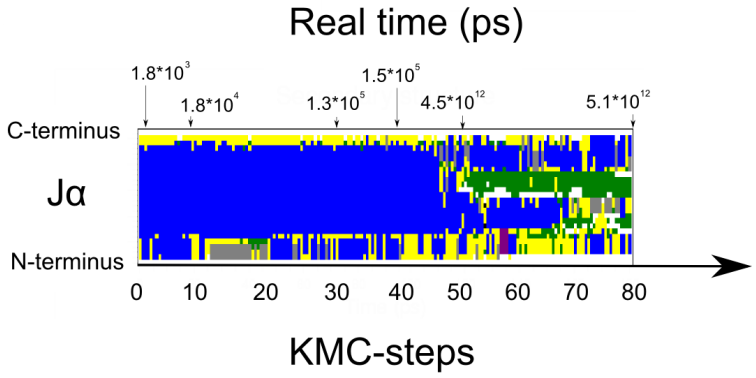


Figure 6.9.: Secondary-structure analysis of the J$\alpha$-helix as a function of KMC steps and real time from Ile539Glu-AsLOV2 mutant in the light-state, obtained with KMC-MD method.

region, which is confirmed by the increase in the distance of the respective center of masses of the light-state visualized in Fig. 6.17. From the same figure, we deduce that in the final stage of the KMC-MD simulation the difference in the distance between these two structural elements in the dark and light-state amounts to 0.4 nm, which remains stable over several microseconds. We emphasize that these observations are in agreement with the X-ray measurements of Wu *et al.* [72], who demonstrated that the A$\beta$-B$\beta$-loop from the AsLOV2-domain in PA-Rac1 acts as a steric inhibitor enabling the light-sensitive regulation of the functional activity of the Rac1 enzyme.

To conclude, we evaluate the computational expense required by the KMC-MD method in comparison to the MD-simulation technique, to access similar timescales. To this end, let us consider the light state of the PA-Rac1 system, for which we reached a total simulation time of 300 $\mu$s using the KMC-MD method. In this calculation a total number of 40 KMC steps were accomplished, requiring a total number of 200000 MD-relaxation steps. Now, if we would use the conventional MD-simulation technique with a timestep of 1 fs, we would need a total number of $3 * 10^{11}$ MD steps, to reach the same time of 300 $\mu$s as obtained with the KMC-MD approach. By considering that the costs to execute the KMC events are on the order of a MD-relaxation phase, we conclude that the KMC-MD-simulation method reduces the computational expense

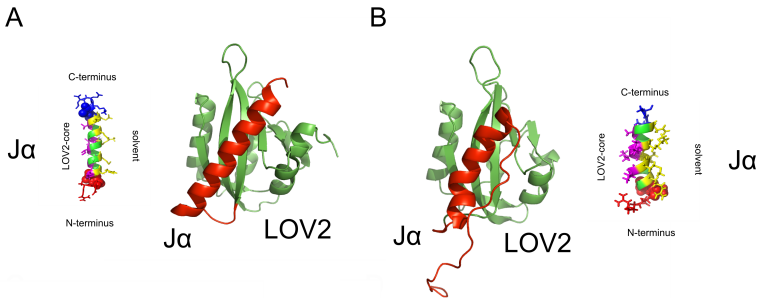


Figure 6.10.: (A) Front-view of the AsLOV2-core with the Jα-helix from NMR-solution structure of Harper *et al.* [54]. (B) Front-view of the AsLOV2-core with the Jα-helix from the X-ray crystal structure of Halavaty *et al.* [55].

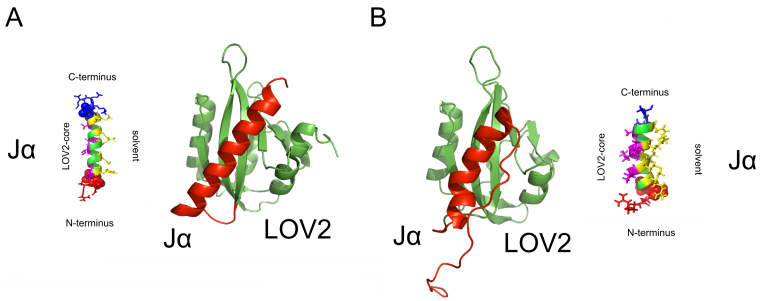


Figure 6.11.: (A) APBS surface of the AsLOV2-core with the Jα-helix from NMR-solution structure of Harper *et al.* [54]. (B) APBS surface of the AsLOV2-core with the Jα-helix from the X-ray crystal structure of Halavaty *et al.* [55].

by a factor of $7.5 * 10^5$ for the PA-Rac1 system in the light state.

## 6.1. Conclusions

In summary we introduced a novel multiscale-modeling method in this work, based on a combination of the kinetic Monte Carlo- and molecular-dynamics-technique, and demonstrated its usefulness for the simulation of the multiscale signaling behavior of the AsLOV2-Jα-photoswitch as well as the AsLOV2-Jα-based photoactivable Rac1-GTPase. After photoexcitation with blue light, these systems typically exhibit a complex signaling pathway by carrying out multiscale relaxation dynamics. This implicates local structural changes at the residue level near to a light-sensitive reaction center on a nanosecond timescale, inducing extensive structural rearrangements between peripheral protein domains on timescales ranging from several microseconds up to seconds. More specifically, with our new approach we find that the early stages of the signaling pathway are characterized by the coupling between the Iβ- and Hβ-strands through H-bond formation between Gln513 and Asn492. This event leads to a N-terminal unfolding of the Jα-helix and its subsequent disruption from the LOV-core. In case of the photoactivable Rac1 system, we show that this latter disruption process results in the release of a functional α-helix located at the switchII region of the Rac1 enzyme, which is known to be a primary docking site for Rac1

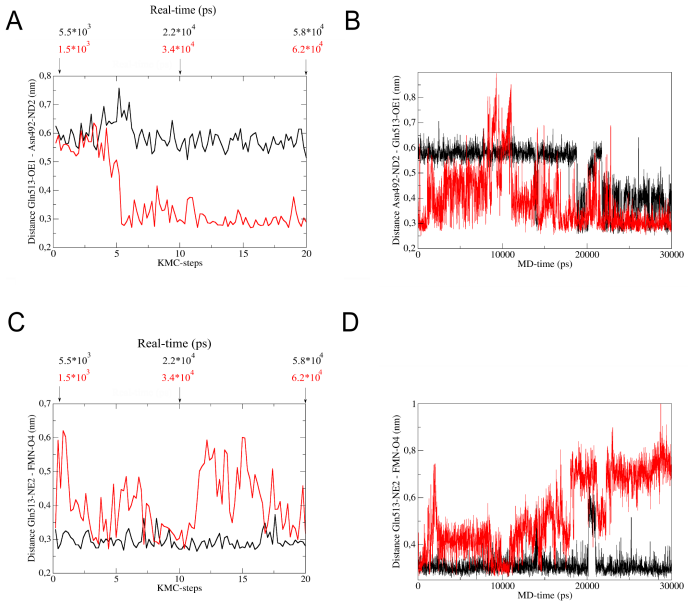


Figure 6.12.: Inter-atomic distances of the PA-Rac1 system, obtained by using KMC-MD and non-invasive MD methods. (A) Inter-atomic distance between Gln513-OE1 and Asn492-ND2 from KMC-MD simulation in the RB-KMC phase as a function of KMC steps. (B) Inter-atomic distance between Gln513-OE1 and Asn492-ND2 from MD simulation as a function of MD time. (C) Inter-atomic distance between Gln513-NE2 and FMN-C4=O from KMC-MD simulation in the RB-KMC phase. (D) Inter-atomic distance between Gln513-NE2 and FMN-C4=O from MD simulation as a function of MD time [red: light-state; black: dark-state].

with its effector domains, like e.g. PAK1. Through comparing these findings with the available experimental results, we conclude that our method is a powerful new multiscale-modeling tool for simulating the signaling behavior of large protein complexes from ultrafast up to experimental timescales. Further gains in efficiency can be expected by using multiple-timestepping approaches, like e.g. RESPA's, to carry out the MD-rexalation phase. Finally, we emphasize that its underlying algorithm is transferable to a large number of alternative protein-solvent systems, and will open new avenues for investigating the light-induced regulation of enzymatic reactivity [169] as well as gene function [170].

## 6.2. Determination of KMC-parameters

For suitable and appropriate simulation of previously described systems, parameters have to be found preliminary to the KMC-MD simulation of our systems. Primarily parameters have to be estimated, which include the scan-radius for the breakage of H-bonds as well as the regions of interest for ID-KMC and RB-KMC. We start with the description of the determination of the regions of interest for RB-KMC. This region is determined by an H-bond analysis between co-factor and its surrounding protein in the starting structure. The subsequent list of H-binding

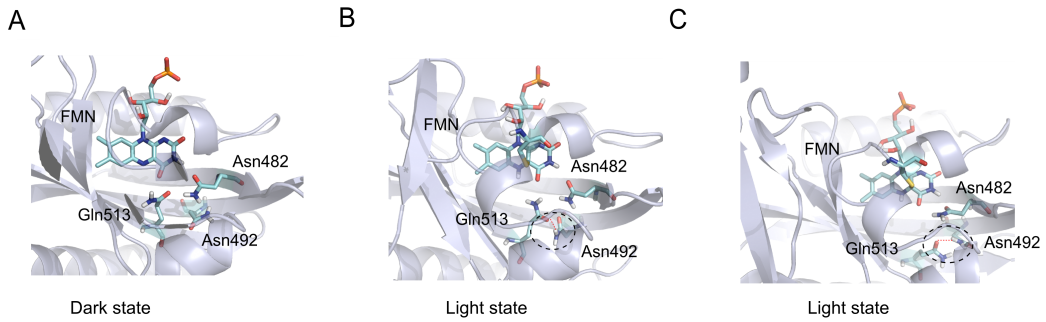


Figure 6.13.: Final representative amino-acid configurations near to the FMN chromophore, obtained with (A) KMC-MD simulation of the dark-state, (B) KMC-MD simulation of the light-state as well as (C) MD simulation of the light-state [cyan: C-atom; red: O-atom; blue: N-atom; white: H-atom; orange: P-atom; yellow: S-atom].

amino-acids comprises the region of interest for this RB-phase. If necessary, essential amino-acids known from literature have to be added to this RB-list. We point out that the domains for the ID-KMC phase have to be chosen according to the flexibility and secondary structure of the system. These domains can be identified by secondary structure analysis and subsequent root-mean-square fluctuation analysis, i.e. domains are enbordered by largely fluctuating regions and at the same time characterizable through their uniform secondary structure. Additionally, the break-radius and the radius of formation of H-bonds is estimated according to the root-mean-square fluctuation of H-bonds in between the ID- and RB-regions during a long-time AA-MD simulation. A good estimate of the break-radius is the value of 100%-150% of the maximum value of the root-mean-square fluctuation of the H-bonding distance in between donors and acceptors of the H-bond.

## 6.3. Process path of H-bond breakage

In the recent KMC-MD algorithm, the breakage of H-bonds is performed according to following equation:

$$\vec{r}_{new} = \vec{r}_{old} + \frac{r_{max}}{|\vec{r}_{donor} - \vec{r}_{acceptor}|} * (\vec{r}_{donor} - \vec{r}_{acceptor}) \ , \tag{6.1}$$

where $r_{max}$ stands for the scan-radius, $r_{donor}$ is the initial H-bond donor coordinate and $r_{acceptor}$ the H-bond acceptor coordinate, respectively. The propagation of coordinates is performed along the given process-path in eqn. 6.1 along a stepwise pathway. Each step is followed by subsequent minimization, as mentioned in the previous section 5.1.

Due to heterogeneties within the system, this type of breakage cannot perfectly fit the process-path onto the present situation in the system. An improvement of this breakage, is the manifold way, where simply different translational paths of breakage are sampled independently from each other and the energetically most favourable path is chosen, according to:

$$\vec{r}_{new} = \vec{r}_{old} + \frac{r_{max}}{|\vec{r}_{donor} + \Delta x - \vec{r}_{acceptor}|} * (\vec{r}_{donor} + \Delta x - \vec{r}_{acceptor}) \ , \tag{6.2}$$

where $\Delta x$ could have following values: $\{+x, 0, 0\}$, $\{-x, 0, 0\}$, $\{0, +y, 0\}$, $\{0, -y, 0\}$, $\{0, 0, +z\}$, $\{0, 0, -z\}$ and $\{0, 0, 0\}$.

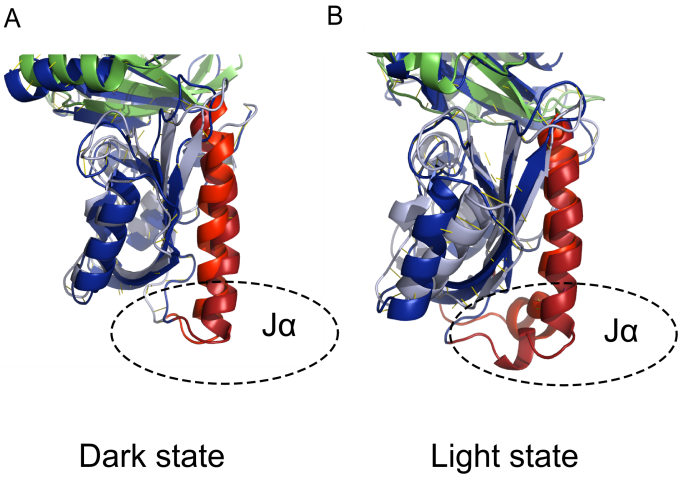


Figure 6.14.: Overlaid final structures of dark- (A) and light-state (B) of PA-Rac1 after 30 ns of MD- (dark colors) and 60 KMC-MD-steps (bright colors).

Finally, after a manifold scan over all 7 possible translational breakages we could select the energetically most favourable pathway of breakage. Another approach, which might be an alternative to the previous approach, would be the determination of a Metropolis-Monte-Carlo sampled path towards the broken H-bonded state. We illustrate this scheme in Fig. 6.18, where we show how the breakage of the H-bond is performed on a Metropolis sampled process path. For each scanstep a range of points on a radius with equal distance to the acceptor is given, e.g. 3 equally distributed points with $\alpha = 120$ degrees (see Fig. 6.18). Each of these points is tested according to its potential energy value and the actual step is selected according to the Metropolis criteria. With this new approach we gain a propability-weighted path of H-bond breakage, which allows also configurations with an increase in energy with some probability.

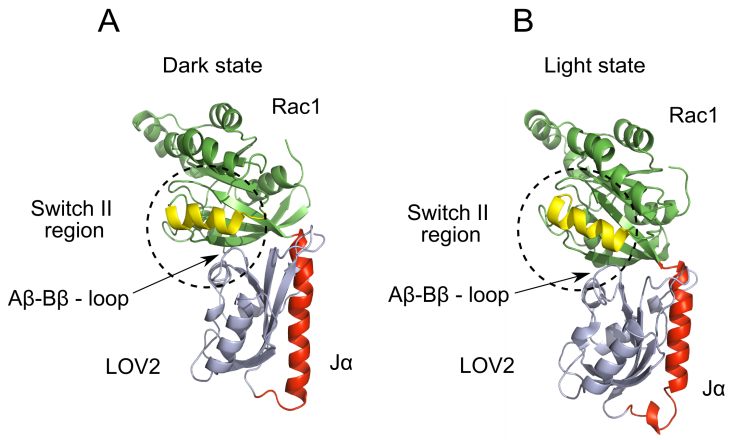


Figure 6.15.: Final structures of PA-Rac1 system in the dark-state (A) and light-state (B) from KMC-MD simulation after 60 KMC-MD steps.

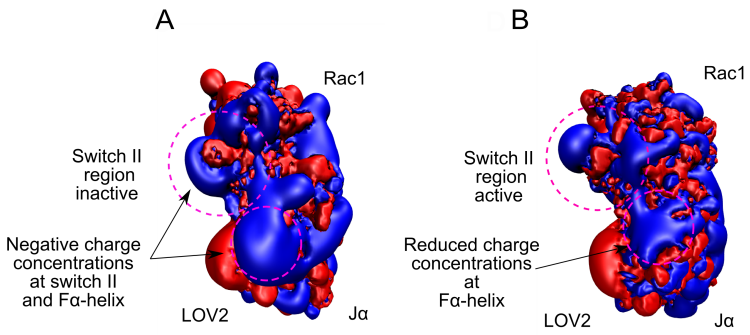


Figure 6.16.: APBS surface of PA-Rac1 system in the dark-state (C) and light-state (D).

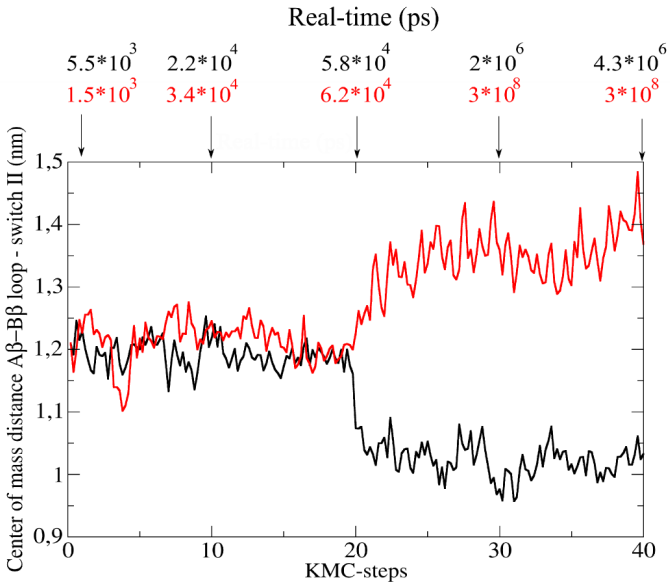


Figure 6.17.: Center-of-mass distance between A$\beta$-B$\beta$-loop and switchII from KMC-MD simulation of the light-state (red) and dark-state (black) as a function of KMC steps.

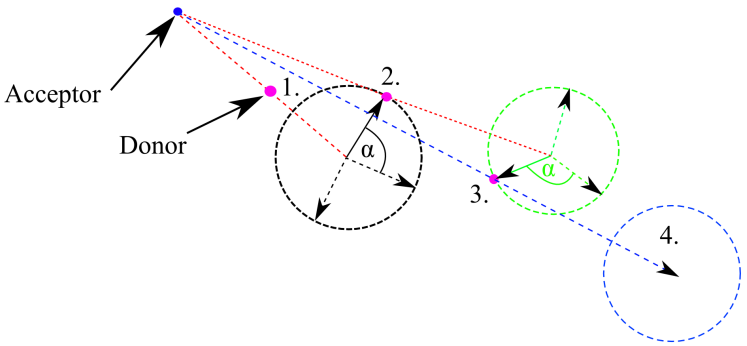


Figure 6.18.: Scheme for Metropolis-sampled alternation of process-path throughout breakage of an H-bond.

# 7. Biased Kinetic-Monte-Carlo for co-factor binding

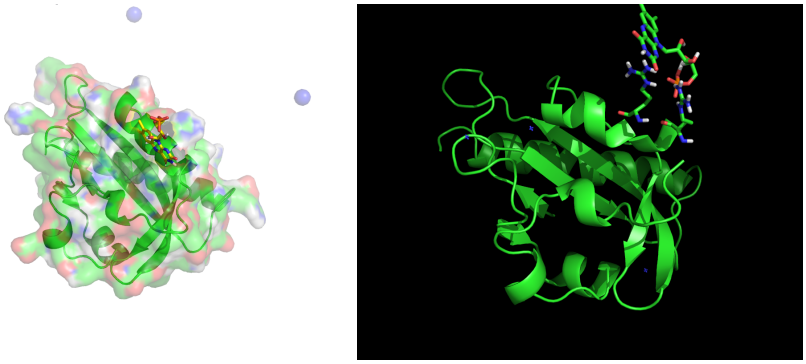

Figure 7.1.: Left: Starting geometry of the FMN-chromophore at its reaction coordinate in vicinity to Arg451 and Arg467. Right: First step of approach of ligand into binding pocket, where it is attached to Arg451 and Arg467.

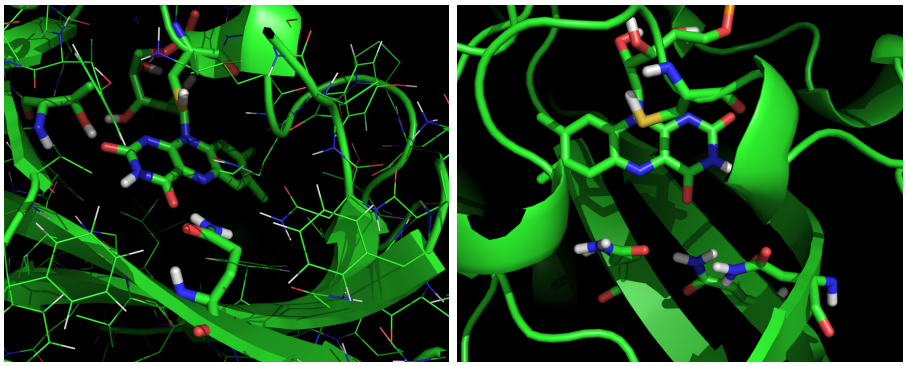

Figure 7.2.: Final configurations of FMN-chromophore after 4 independent KMC-MD simulations. Left: Final configuration of FMN-chromophore in binding pocket of LOV2-domain after real-time of 3.4 s. Right: Final configuration of FMN-chromophore in binding pocket of LOV2-domain after real-time of 3.84 s

## 7.1. Introduction

In all kingdoms of life, co-factors act as regulators for enzymatic activity in cell signaling processes. The Flavin co-factor Flavin-Mono-Nucleotide regulates light-sensitivity in plants, bacte-

ria, green algae and fungi [33]. In LOV-domains, this co-factor is non-covalently attached in the so-called binding pocket within the PAS-fold. Another example is signaling molecule Guanine-Tri-phospate (GTP), which is responsible for signaling in GTPases in mammals [171]. The dynamics of co-factor binding as well as dissociation from its target enzyme are presently poorly understood [29]. Even in case of known co-factor binding sites, the pathway of the co-factor into its binding pocket at the target enzyme is still unknown [172].

Several approaches for the description of co-factor association, i.e. ligand-host binding exist [173, 174]. Strasser *et al.* introduced a new algorithm, called *LigPath*, which describes the co-factor translation along a defined path into the protein and minimizes the ligand with respect to a defined number of internal degrees of freedom [173]. Meiler *et al.* introduced a popular technique, which initially localizes the ligand at the proposed association site and minimizes this configuration with the use of Metropolis Monte-Carlo [174]. Although these approaches deliver insights into ligand binding and can evaluate the stability in the binding pocket, a simulation technique, which delivers insight into the dynamical details of ligand binding is still lacking. In our work, we elucidate this pathway and the mechanism of complex-formation through a new technique, which is based on KMC-MD. In the next sections, we start with an introduction of our new algorithm and then go over to present our results of the simulations on AsLOV2 and the Hypoxia-Inducible-Factor (HIF-2$\alpha$).

## 7.2. Algorithm

The algorithm uses the same RB-KMC technique as the KMC-MD algorithm, described in the previous chapter 4. We consider the co-factor as one single residue, which undergoes the major RB-transitions for each KMC-step. By contrast to the RB-KMC algorithm, the co-factor undergoes a stepwise reduction of distance to its given target prior to each RB-KMC step. We note, that the biased reduction in the distance between target and co-factor does not contribute to the actual real-time and is considered as a non-sampled event in this algorithm. Moreover, the target is one of the needed input-parameters, known from literature or analysis of the interior binding cavities of the host-system. Notably, each of the enzymatic or sensory host-proteins has at least one significant binding cavity, e.g. all PAS-folded systems, kinases or GTPases. In the following, we describe the conception of our new technique :

1. MD-relaxation
2. Stepwise reduction of distance between ligand and binding-pocket with subsequent minimization without constraints.
3. H-bond analysis.
4. Scan of potential energy difference $\Delta\Phi$ along process path and determination of corresponding Gibbs free energy $\Delta$G.
5. Calculation of rate $r_j$ of each event j according to:

$$r_j = \nu e^{\left(\frac{-\Delta G}{RT}\right)} \; ; \tag{7.1}$$

   where $\nu$ represents the pre-exponential factor and R the universal gas constant.

6. Select event j from list of events by generating a uniform random number $\xi \in [0, 1]$ and by solving:

$$R_{j-1}/R_N < \xi \leq R_j/R_N \; ; \tag{7.2}$$

   where $R_N = \sum_{j=1}^{N} r_j$

7. Execute selected event j, and update time by generating a uniform random number $\xi \in [0, 1]$ and calculating:

$$\Delta t = -\frac{\ln \xi}{R_N} \ . \tag{7.3}$$

8. Return to (1)

We mention, that for sake of efficiency of the simulation, the co-factor has to be located at one of the optimized diffusion paths of the target enzyme in the starting configuration. Otherwise, the stepwise reduction in step 2 of the previous algorithm leads to a surfacial scanning of the host, until the co-factor possibly enters the protein. We point out that on one hand this leads to inefficiencies in computational cost, while on the other hand this leads to first insights into previously unknown systems and their unknown paths of ligand host binding.

## 7.3. FMN co-factor binding

The non-covalently bound FMN co-factor controls light-sensitivity in LOV-domains, which is located within the hydrophilic and the hydrophobic LOV-core regions and forms 3 H-bonds with one adjacent glutamine and two asparagines. In all LOV-domains, one reactive cysteine is located in vicinity to the FMN-C4a atom, which performs covalent adduct-formation after blue-light irradiation (see chapter 2.3). For simulation and testing of our algorithm on FMN co-factor binding, we chose the X-ray structure of AsLOV2-J$\alpha$ of Halavaty *et al.* [55]. We generated the starting configuration prior to our simulation through locating the chromophore approximately onto the most probable diffusion path within a 1 nm distance to the binding cavity. This reaction coordinate was chosen to be located in vicinity to two arginine residues, Arg451 and Arg467, as shown in Fig. 7.1. In additon we filled the cubic box with 8500 SPC-waters and neutralized the system by adding 5 sodium-ions. The bias for the per-step-reduction of the distance between Gln513-NE2 and FMN-C4=O was set to a value of 0.5 nm. The MD-relaxation had a length of 1 ps with a timestep of 1 fs. The MD-related parameters were chosen as described in the previous section 3.6.

### 7.3.1. Results

We start by analysing the pathway of cofactor binding in the first steps of our KMC-MD simulation. At this stage of simulation we observe, that the FMN-chromophore starts to approach the surface of the AsLOV2-domain. In this configuration after 3 KMC-MD steps we see, that the phosphate group of the FMN co-factor binds through H-bonds onto 2 arginine residues Arg451 and Arg467, which are located in vicinty to the binding pocket, as shown in Fig. 7.1. Through X-ray structure determination on LOV-domains by Federov *et al.*, Halavaty *et al.* and Crosson *et al.* observed, that these two arginine residues bind to the phosphate group of the FMN-chromophore [11, 55, 10]. We propose, that the initial binding of these two arginine residues is one of the major steps, which determines the diffusion pathway of the FMN-chromophore into the binding pocket in the early stages. In the final configurations of 4 independent KMC-MD-simulations with a simulation time of 20 KMC-MD steps, we observed two different types of configuration of the chromophore within the binding pocket (see Fig. 7.2). In the first configuration we see, that the FMN-chromophore is bound by the N-terminal residue Ser418 and Gln513. We observe, that the atom Cys450-SG is located close to FMN-C4a and could also perform the same adduct formation with FMN-C4a, due to the vicinity of the reactive cysteine [11, 55, 10]. Moreover, the carbonyl-oxygen FMN-C4=O is singly bound, which was also found by Alexandre

*et al.* through FTIR-measurements [46]. In the second configuration we observe, that the FMN-chromophore is bound analogously as in the crystal structure, as we show in Fig. 7.2. Finally, we conclude that the fully resolved pathway of the FMN-chromophore into the binding pocket of the AsLOV2-J$\alpha$-system ends in two different final FMN-configurations after 20 KMC-MD steps in 4 independent simulations. We point out that both configurations are in vicinity to the reactive cysteine and could undergo formation of the covalent adduct.

## 7.4. Artificial co-factors binding to HIF-2$\alpha$

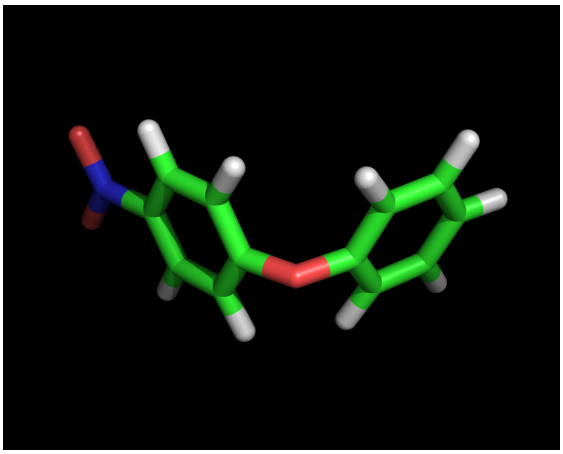

Figure 7.3.: Ligand KG-721 (*1-nitro-4-phenoxy-benzene*) [29].

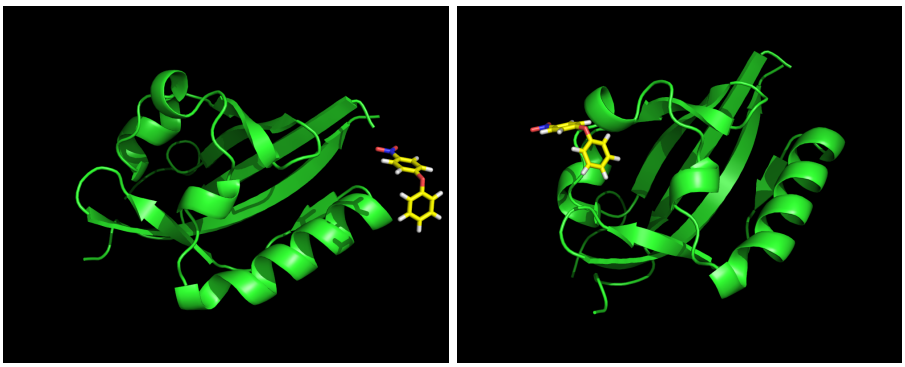

Figure 7.4.: Left: First starting configuration of co-factor KG-721 outside of HIF2$\alpha$-domain, Right: Second starting configuration of co-factor KG-721 outside of HIF2$\alpha$-domain.

The so-called Hypoxia-Inducible-Factors (HIF) play an important role in hypoxic-pathway regulation and, thus, are major factors influencing cancer-cell resistance against chemotherapy in several cancers [39]. In response to oxygen, they act onto numerous hypoxic-pathway related genes through their heterodimeric basic helix-turn-helix (HTH) motif [24]. Many PAS-domains are internally regulated by ligands or so-called cofactors, which regulate inter-protein interactions,

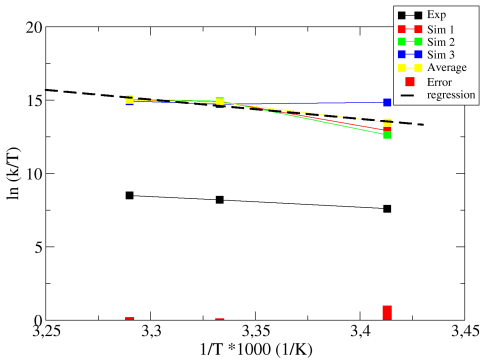


Figure 7.5.: Plot for the determination of activation free energy of KG-721 co-factor binding on HIF2$\alpha$-domain. $ln(k/T)$ plotted against inverse of temperature $1/T$.

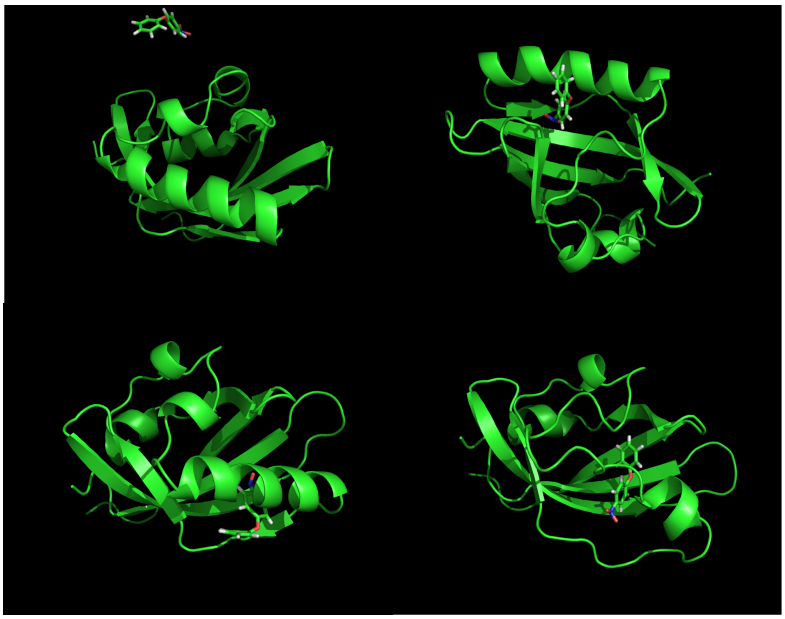

Figure 7.6.: Different snapshots of intrusion of KG721 into the HIF-2$\alpha$ protein during biased KMC-MD simulation.

such as in case of LOV-domains, BLUF or Cryptochrome [175]. The question arose in the last four years to which extent HIF-domains are possibly regulated by co-factors. Though, no recent study could demonstrate the existance of relevant co-factors, which act onto HIF-domains. However, an internal water-filled cavity of 290 $Å^3$ has recently been found, which suggests that this would be an appropriate binding site for co-factors [176]. For example, Key *et al.* performed X-ray diffraction, titration calorimetry, NMR-spectroscopy and molecular-dynamics simulations for the determination of binding affinities of several compounds with the binding pocket of HIF2$\alpha$ [29]. These compounds are composed of benzylic rings, which are substituted by hetero-atomic side-groups. In order to test the affinities of the ligands and to simulate the pathway of binding, we chose the compound KG-721 for simulation, as shown in Fig. 7.3 [29]. We simulated the pathway of ligand binding, starting with two structures, shown in Fig. 7.4. We varied the simulation temperatures from 283 K up to 303.95 K, to validate our data with the experiment, and performed 3 simulations for each temperature. For simulation we chose the same parameters as described in section 7.3. For comparison of our data with the data of Key *et al.*, we considered, that for each temperature, a final KMC-MD time $\tau$ of ligand binding was obtained, determined through eqn. (7.3). When we assume the ligand-binding of KG-721 on HIF2$\alpha$ to be a first-order reaction, then

$$k_{on} = 1/\tau \ , \tag{7.4}$$

where $k_{on}$ is the first-order rate-constant for co-factor binding. By suitably plotting $ln(k_{on}) = \frac{\Delta G}{RT} + \ln(A)$, we obtain $\Delta G$, where R is the universal gas-constant and A the pre-exponential factor. From Fig. 7.5, we observed that the values for $ln(\frac{k_{on}}{T})$ were shifted by a constant value of about 5, resulting from a shift in the pre-exponential factor $\ln(A)/T$. By averaging over the result of 9 independent simulations we obtained a value for $\Delta G = -30.743 \pm 1.5 \frac{kJ}{mol}$, whereas the experimental value determined by Key *et al.* is at $\Delta G_{exp} = -38.9 \pm 2.51 \frac{kJ}{mol}$, as shown in Fig. 7.5. Finally, we deduce from our results that this method is suitable for the investigation of binding affinities of newly designed co-factors.

The curve of $\ln(\frac{k_{on}}{T})$ differs by a value of 5 from the simulated curve. Thus, our pre-exponential factor has to be shifted by $\exp(-5)$. In the following, we propose a technique for the estimation of Eyring's pre-factor without experimental input. For this purpose, we make use of a molecular dynamics run prior to the KMC-MD simulation. We set up the system in a discrete number of configurations by which the co-factor might intrude into the binding pocket of the protein, as shown in Fig. 7.6. Through measurement of the co-factor's root-mean-square-fluctuation of each trajectory, we could calculate the co-factor's time-dependent tendency for the intrusion into the binding pocket. From linear interpolation of this average intrusion into the binding pocket and the known simulation time, we then can estimate the pre-exponential factor, through linear fitting of the KMC-real time.

# 8. Multi-scale modelling using mesoscopic-atomistic Coarse-Graining

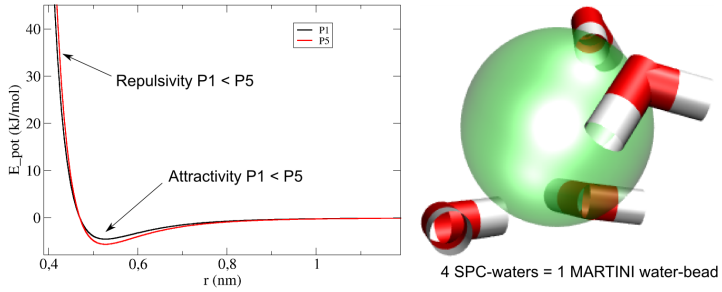


Figure 8.1.: Left: 12-6 Lennard-Jones potential of P1-P1 (black) compared with P5-P5 (red). Right: MARTINI coarse-grained model of 4 waters

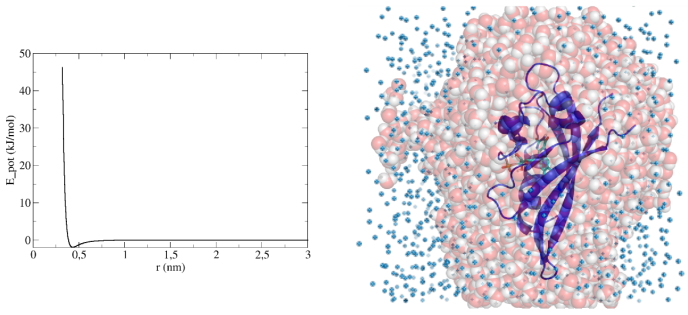


Figure 8.2.: Left: 12-6 Lennard-Jones potential AA-oxygen - P5-water. Right: Atomistic-mesoscopic model of protein-solvent system.

## 8.1. Introduction

Coarse-graining (CG) reduces the degrees of freedom of an atomistic system and thus reduces the system's description to a mesoscopic CG-level. The computational cost for describing such a CG-system is significantly lowered by the degree of coarse-graining, but at the same time reduces

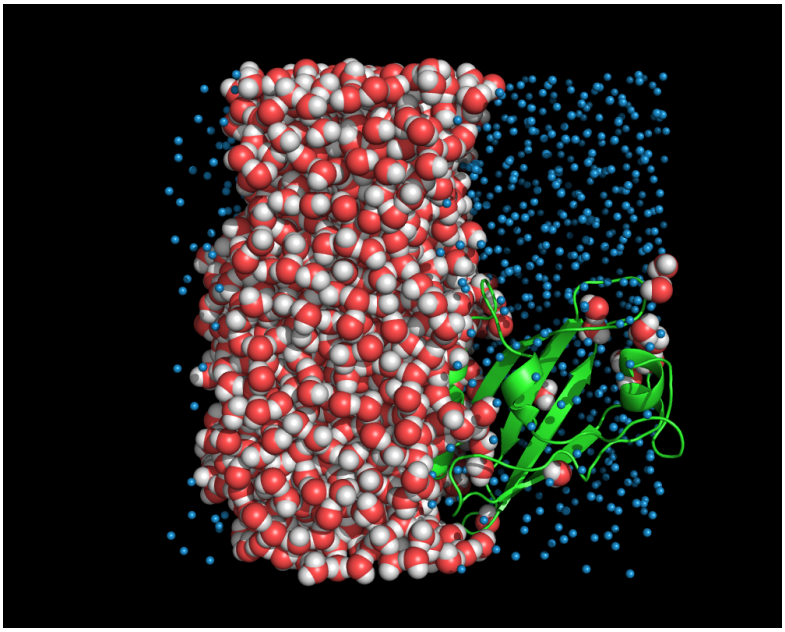

Figure 8.3.: Atomic-mesoscopic model of protein-solvent system after 20 ns of MD-simulation without application of solvent-exchange algorithm.

the chemical information, which we can obtain from the trajectory. Several approaches have been introduced, which coarse-grain especially biological systems without a loosening of system-shape and -size as well as all-atom (AA) quantities, i.e. the macromolecule's self-diffusion or solvation free-energies. Baron *et al.* as well as Marrink *et al.* introduced a CG-scheme, which uses fitting onto experimental phase-separation-, hydration- and partitioning free-energies. Moreover, they fitted their CG potential parameters onto AA phase-interaction quantities, such as the overall potential of mean force (PMF) between different phases. For the validation of their approach they used a multitude of data, which is based on the experiments on several liquid mixtures, i.e. water-alkane and water-alkene mixtures [177, 178]. It was included in the MARTINI-forcefield, which is one of the most widespread biomolecular CG-forcefields in case of lipid-membrane simulations and has the advantage of transferability. Shelley *et al.* and Izvekov *et al.* developed the so-called force-matching procedure, using a comparative process of CG- and AA-Hessian matrices. During this procedure the CG-Hessians are iteratively fitted onto the reference forces of the AA-simulation. They demonstrated that their CG-structure is reproducing the atomistically defined structure with high accuracy [179, 180, 181]. Moreover, Prapotnik and Shi *et al.* developed a technique, which describes the system on multiple scales [182, 183]. Prapotnik *et al.* used an adaptive algorithm, which alters the description of water from CG to AA, in case of an approximation close to a defined sphere of interest in which the AA-description is necessary. In their case, this defined sphere is e.g. the AA-solvation shell of a solvated AA polymer-chain. However, Shi *et al.* used an additive potential-description of AA- as well as CG-pair-interactions, and demonstrated that their mesoscopic-atomistic (MA)-model reproduced the fluctuations of the full AA-simulation.

## 8.2. Coarse-grained water

Marrink *et al.* generated one CG-bead representing 4 AA-waters in the scope of the transferable MARTINI-forcefield [178]. This water model performs phase-separation with other aliphatic CG-particles and reproduces solvation free-energies of alkanes. Marrink *et al.* defined 5 different MARTINI-waters using the LJ 12-6 potential description. In Fig. 8.1 we show how hydophobic interactions of the MARTINI-P water type were expressed by an increased slope in the repulsive LJ 12-6 part, while hydrophilic interactions were implemented by a lower minimum in the attractive part of the potential. Moreover, Baron *et al.* used soft-core potentials describing CG-water interaction with the advantage of a better reproducibility of thermodynamic data of partitioning free-energies of phase-separating liquids [177]. Marrink *et al.* performed CG-simulations on membrane-water systems with the use of equally partitioned mixtures of all 5 MARTINI P-waters, in order to mimic thermodynamic properties of water as close as possible [178], i.e. liquid state at room-temperature. In our simulations, which we present in the following sections, we solely used one single species of CG-water and varied the water types in several independent simulations. The water type, named P5, has the largest hydrophilicity, whereas P1 is the most hydrophobic water, with respect to its Lennard-Jones parameters (see Fig. 8.1). In the next section, we introduce the atomistic-mesoscopic approach for protein-solvent systems and discuss the influences of the different water-types on the dynamics of the protein.

## 8.3. Atomistic-mesoscopic approach of protein-water systems

Protein-dynamics is primarily dependent from the H-bond network of the first hydration shell around the protein [184]. This H-bond network is ordered in dependence of the specific hydrophilicity and hydrophobicity of the protein surface and has an imminent influence on the stability of the protein [185]. We conclude from these facts, that an explicit-water shell with a radius of 1 nm can effectively stabilize the protein in its dynamical behavior. In our approach we assume that the influence of the bulk solvent on protein dynamics is negligible and can be replaced by CG-waters. We point out that this approach significantly reduces the number of atoms in the system as well as the computational cost of the simulation, while it may retain the main characteristics of the AA-trajectory. In Fig. 8.2 we show the LJ-interaction between AA-water oxygen and CG-P5 MARTINI water, which was performed according to the Lorentz-Berthelot combination rules. In the same figure we show the LOV1-domain solvated by MA-solvent, which is composed from CG- as well as AA-water. It is important to note, that AA-water permanently has the tendency to diffuse into regions with lower density, and AA-water and CG-water have a large tendency to de-mix. This effect is thermodynamically driven by entropy, which always gains for its maximum in the bulk, i.e. the solvent shell is thermodynamically unstable in its mixture with CG-solvent. The reason for this instability is the potential describing the different CG- and AA-solvent particles. Un-charged CG-MARTINI water is largely hydrophobic, while charged AA-water (SPC-water) is hydrophilic. In Fig. 8.3, we visualize the effect of the de-mixing tendency of the two different water types on the protein-solvent system. We see, that the AA-water as well as the CG-water form separated phases, while the protein becomes dis-solvated from the AA-water and is largely opposed to CG-water. We conclude, that the atomistic-mesoscopic (AM) approach for protein-solvent systems significantly reduces the computational cost. However, this leads to de-mixing and dis-solvation of the protein, as shown in Fig. 8.3. We solved this problem of de-mixing and phase-separation through developing the so-called solvent-exchange algorithm, which consists of following steps :

1. MD-phase ;

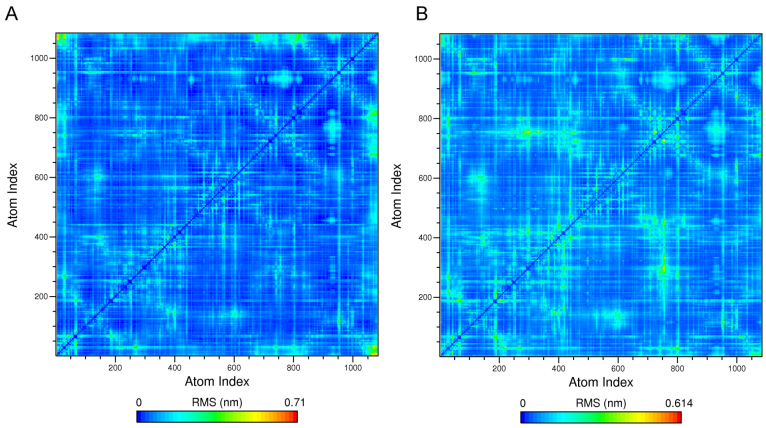


Figure 8.4.: RMSD as a function of atomic-number, averaged over 20 ns of AA-simulation (A) as well as over 20 ns of AM-simulation (B).

2. Test if AA- or CG-waters have diffused out of their regions ;
3. Set AA- and CG-waters randomly back into their regions ;
4. Minimize the sterical interaction between waters ;
5. Equilibrate the solvent by freezing the protein at fixed position ;
6. Return to (1).

## 8.4. Atomistic-mesoscopic approach with LOV1-domain of Chlamydomonas reinhardtii

### 8.4.1. System preparation

As starting structure we chose the X-ray structure from Fedorov *et al.* [11] (PDB-code: 1N9L). We centered this protein in a cubic box with box-length 6.09 nm. Then, we solvated the protein in a shell of SPC-water with a radius of 1 nm and neutralized the system by adding one sodium ion. Subsequently, the free space in the box was filled with P5-MARTINI water with a vdW-distance of 0.23 nm between explicit and coarse-grained water (see also Fig. 8.2). For simulation we used the GROMACS-3.3.1. simulation package and a combination of the GROMOS96-43a1 and the MARTINI-forcefield, to describe the interactions. Each MD-phase was carried out over a duration of 20 ps, using a timestep of 1 fs and a total MD-production run of 20 ns, after an equilibration of 1 ns. Unless specified otherwise, the constant NPT-ensemble was imposed using the parameters described in section 3.6.

### 8.4.2. Results

We start our analysis of the AM-trajectory with the average residue-based RMSD matrix, which we show in Fig. 8.4. Here, we compare the average RMSD distribution of the AM-simulation with the AA-simulation on the LOV1-domain. We see in Fig. 8.4A, that the RMSD distribution in the

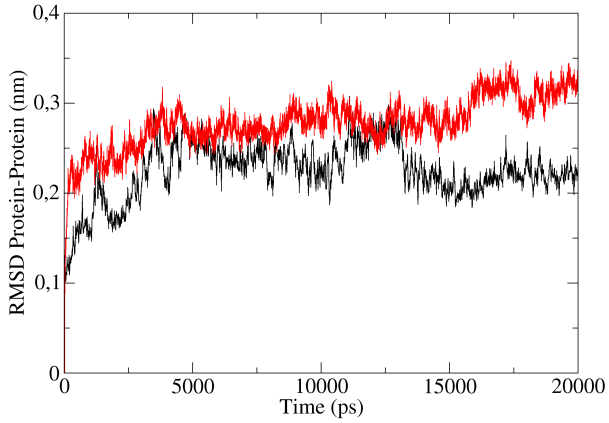


Figure 8.5.: Root-mean-square deviation from starting configuration as a function of simulation time. AA-case (black); AM-case (red).

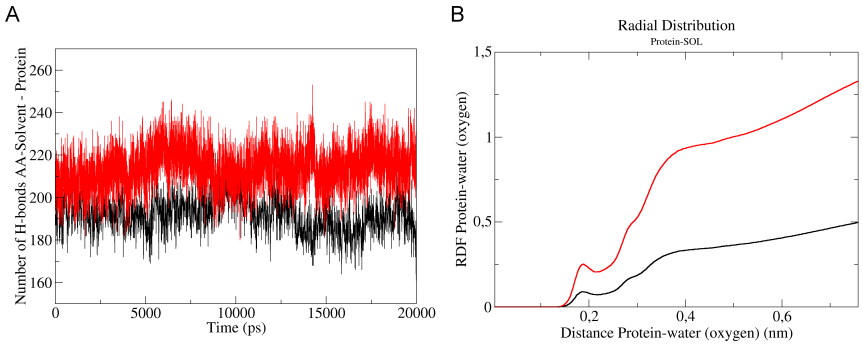


Figure 8.6.: A: Number of H-bonds solvent - protein; AA-case (black); AM-case (red). B: Radial-distribution function protein-water; AA-case (black); AM-case (red).

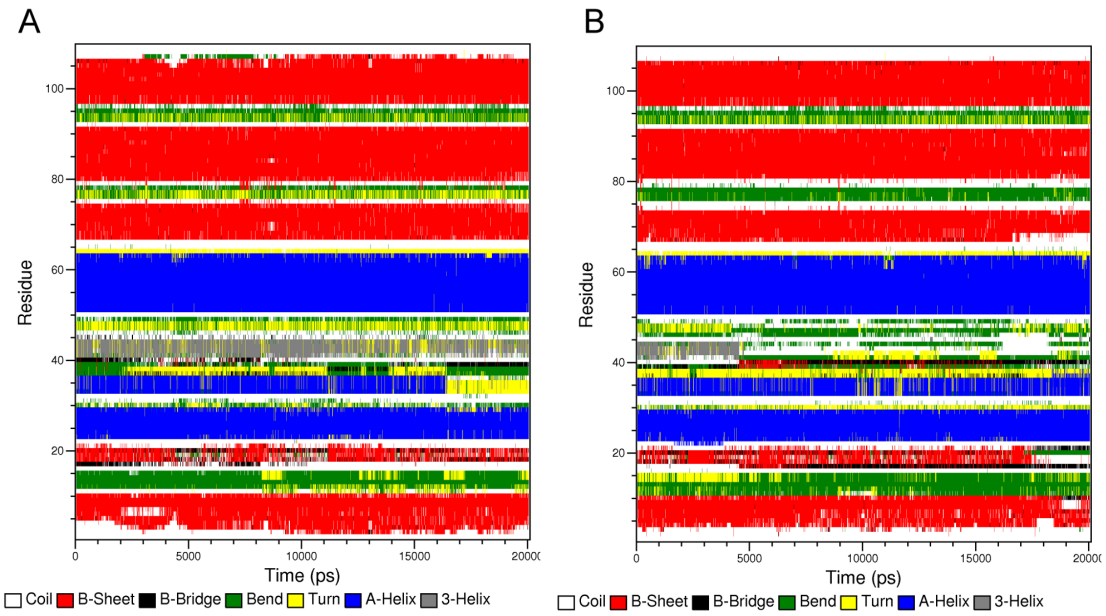


Figure 8.7.: Secondary structure analysis as a function of MD-time. A: AA-simulation; B: AM-simulation.

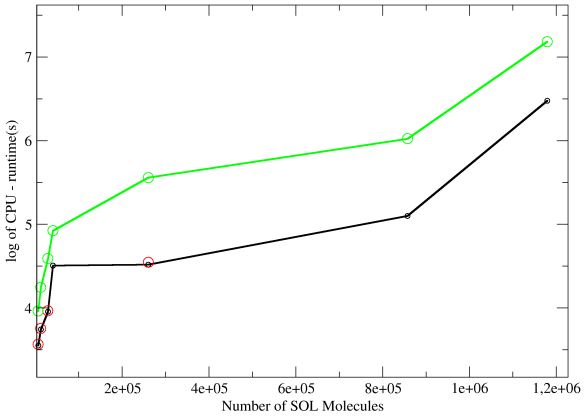


Figure 8.8.: CPU-time using the AM-approach compared with AA-approach as a function of number of solvent molecules

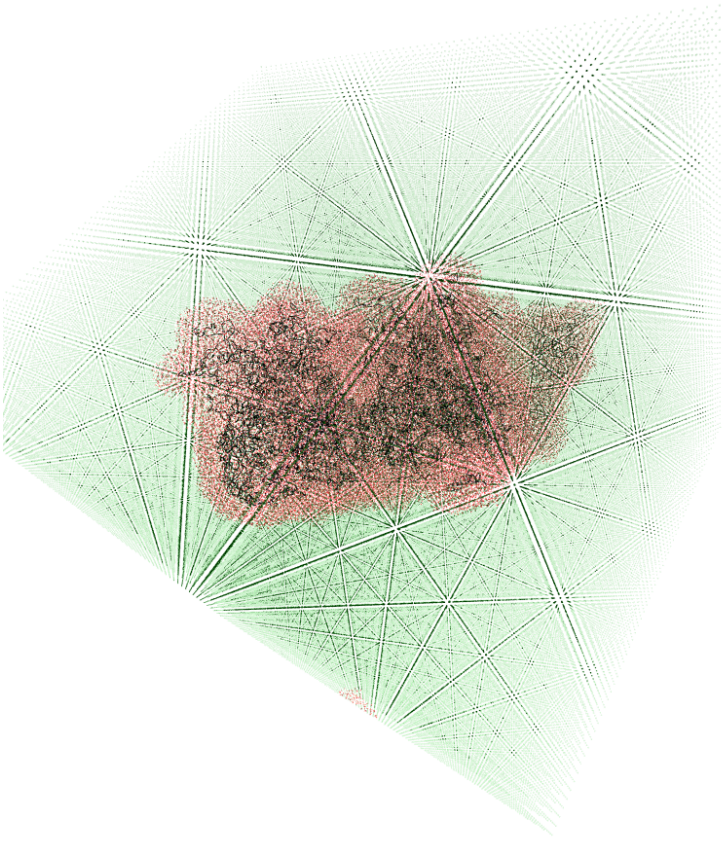

Figure 8.9.: AM approach for Hexon-protein of human adenovirus (PDB: 1VSZ).

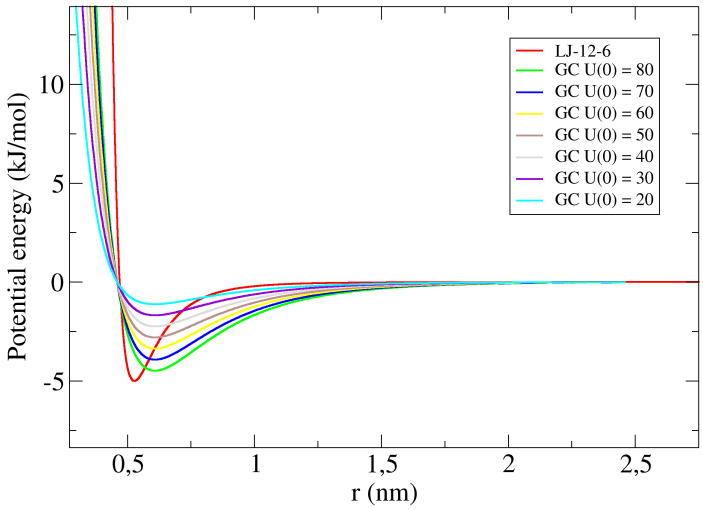


Figure 8.10.: Lennard Jones 12-6 potential of P5-P5(MARTINI) interaction, compared with SW-potential of Stillinger and Weber [186], with variation of $U_0$ from 10 to 80 $\frac{kJ}{mol}$.

AA-case has a sharp distribution of fluctuating regions, while in the N-terminal regions (atom-number 0-50) as well as in the C-terminal regions (atom-number 1050-1100) large fluctuations occur. By contrast, the AM-case in Fig. 8.4B has overall broadened fluctuating regions as well as an overall dampening of the magnitude of fluctuation in comparison to the AA-case, which is reduced from 0.71 nm to 0.61 nm. Moreover, largely fluctuating regions around the atom-numbers 750 in the AM-RMSD matrix (see Fig. 8.4B) do not appear in the AA-case. We conclude, that although the overall shape of the RMSD-distributions in Fig. 8.4B can be approximately compared with the AA-case in Fig. 8.4A, the overall magnitude of the RMSD value is dampened in comparison to the AA-case and the RMSD-distribution is broadened in contrast to the sharp distribution in the AA-case. In Fig. 8.5, we show how the protein deviates from its starting configuration. In this figure we see that the AA-simulation has overall a lower RMSD from its starting structure than the AM-simulation. In the first steps of the AM-simulation we observe a huge jump in the RMSD, i.e. a huge deviation from the starting structure in the early steps of the simulation. We conclude from this behavior, that the first hydration-shell is on one hand primarily stabilizing the protein, but on the other hand other solvent-effects contribute to effects on the protein-dynamics, which we do not observe in the AA-case (see Figs. 8.5 and 8.4). Next, we investigate the dynamical and structural behavior of the solvent in order to understand the AM-solvent effects in more detail. We analyze the self-diffusion coefficients of AA- and CG-water in the AM-system and compare them to the self-diffusion coefficients of water in the AA-system. Surprisingly, the values of the AM-system hugely deviated from the values of the AA-system. The water in the AA-system has an average self-diffusion of $3.98 * 10^{-5} cm^2/s$, while in the AM-system the self-diffusion coefficient of AA-water adopts a value of $21.2 * 10^{-5} cm^2/s$ and a self-diffusion coefficient of CG-water of $22.0 * 10^{-5} cm^2/s$. We note, that there is a huge difference between the AM-case and the AA-case and the huge diffusion-coefficients of AA-water and CG-water in the AM-case might contribute to the broader RMSD-distribution of the AM-protein,

as shown in Fig. 8.4. Moreover, the larger RMSD-value as a function of simulation time in the AM-case might be caused by the drastically increased self-diffusion of the AA- and CG-waters, which as a consequence might lead to instabilities in the trajectory of the system. In Fig. 8.6, we visualize the number of H-bonds between protein and water as well as the radial distribution functions between protein and water, to analyze the protein solvent interaction. We deduce from Fig. 8.6A, that in the AM-simulation an average number of 220 H-bonds exists between protein and solvent, while in the AA-simulation the average number of H-bonds is 185 between the same parts of the system. We conclude, that the larger mean-square displacement of the AA-waters in the AM-case leads to a larger number of H-bonds between protein and solvent. Moreover, this larger number of H-bonds might transmit the large solvent-fluctuation onto the protein and as a consequence cause a larger distribution of RMSD over the protein, as visualized in Fig. 8.4. However, the water structure at the protein-solvent interface in the AM-case is comparable to the AA-case, as shown in the solvent-protein radial-distribution function (RDF) in Fig. 8.6B. In this context it is worth mentioning, that the AA- and the AM-RDF between water and protein have different magnitudes due to the different AA-water densities in the AM-case compared to the AA-simulation. We point out that the stability of the protein in the AM-case is also comparable to the AA-case, as we infer from Fig. 8.7B. From this figure we conclude, that all secondary structure elements remain conserved and show the same behavior as in case of the AA-simulation, shown in Fig. 8.7A. Finally, we emphasize that the AM-solvent system does not resemble AA-solvent with respect to self-diffusion and protein-solvent H-bond statistics. Moreover, the fluctuations of the protein in the AM-case are larger and broadened in its distribution over the protein in comparison to the AA-case. However, secondary structure analysis of the AM-simulation shows a stable secondary-structure over the trajectory, which is comparable with the AA-simulation. We stress that our scheme can also be useful for long-time aggregation simulations of huge aggregates, where the molecular details are not of interest. Notably, our new technique saves about 50% of computation time in comparison with the AA-simulation technique, which means a 2-times speedup of the calculation (see Fig. 8.8). In very large systems, which need an immense amount of bulk-solvent, the computational speed-up grows with the reduction of the number of AA-waters in the system. To give a further impression about the benefit of the AM-approach let us consider the Hexon-protein of the virus-capsid of the human adenovirus with a total number of $1.1885 * 10^5$ atoms, as shown in Fig. 8.9. This system, immersed into a box with a length of 33.35 nm, would need a total number of $1,179278 * 10^6$ waters in the full AA-description. We point out that in the AM-case, this number of AA-waters is reduced to $8,3655 * 10^4$ waters, which means a reduction of $1,095623 * 10^6$ waters in comparison to the AA-case. This amount of $1,095623 * 10^6$ waters is replaced by $1,92830 * 10^5$ CG-waters in the AM-case, which leads to a reduction of $3,094039 * 10^6$ particles, i.e. sites in the system.

One of the previous problems of too high fluctuations in the protein arise from the previously mentioned large diffusion and the inappropriate inter-phase behavior of the AM-solvent system. This might be caused by the inappropriateness of the LJ-6-12 potential, which is too repulsive. For future applications, we propose the implementation of the softer Stillinger Weber (SW) potential, which might lead to better performance of the AM-solvent at the interphase. This potential is given by :

$$U_{SW}(r) = U_0 \left[ \frac{B}{r^p} - \frac{1}{r^q} \right] * \exp\left(d/(r-b)\right) , \tag{8.1}$$

where $U_0$ is the pre-factor and $r$ is the inter-particle-distance. We point out that the SW-pair potential is zero, if $r \geq b$. The other parameters $B$, $p$, $q$, $d$ and $b$ are system-dependent [186] and have to be fit flexibly onto the interfacial problem in the AM-phase. In Fig. 8.10, we see a preliminary fit of the double-Gauss core potential on the LJ 12-6 potential of the MARTINI-

forcefield. We mention, that the repulsivity of the potential decreases, with an increase in attractivity at the same time.

# 9. Development of MD simulation technique for proteins

## 9.1. Protein states

Proteins are state-dependent systems, which undergo primarily transitions between partially disordered and ordered structures. These conformations of the protein-system essentially determine the functionalilty of the protein. We point out that these states are associated with a certain free-energy of the system, which determines their thermodynamic stability. For example, environmental factors, such as temperature or pressure as well as co-factors shift the equilibria between a protein's conformations. In Fig. 9.1A and Fig. 9.1B, we show two possiblities of the thermodynamic stability of different states. In Fig. 9.1A, we see that the unfolded conformation of the system is more stable in free energy than the ordered structure. Such systems are so-called chaperone assisted folded proteins [187], whose folding pathway is sterically controlled by chaperone-proteins. In this case, these systems remain partially disordered, but chaperone-proteins assist folding into a meta-stable protein-conformation [188]. In Fig. 9.1B, we see that the unfolded protein is unfavoured in contrast to the folded state. This thermodynamically favoured folding pathway is subject of large-scale MD-folding simulations, in which systems, such as the bovine-pancreatic trypsin inhibitor BPTI, fold into their native X-ray structure [141]. As demonstrated through these folding simulations, the unfolded protein folds along a direct pathway into its native structure in bulk-solvent, which is solely regulated by the sequence. By contrast, signaling proteins, such as LOV-domains undergo the so-called photocycle upon blue-light irradiation, which alters the stability of the different states, as shown in Fig. 9.1C. This photocycle contains at least one meta-stable intermediate, which decays back into the so-called dark-state within tens of seconds [1]. In case if this system undergoes this light-driven change out of the dark-state, a so-called stressed light-state is generated, which undergoes a further change into the un-stressed conformation of the light-state. This un-stressed state resembles the final light-state, which in Fig. 9.1C is lower in energy than the stressed light-state conformation. We point out that this stessed light-state is a conformation far from equilibrium. Moreover, experimental evidence is given that the dark-state is also a so-called non-equilibrium system. Mathes *et al.* performed an *in vivo* exchange of modified flavin-co-factors with the native FMN-chromophore [189], which suggests, that the dark-state is a meta-stable state when a FMN co-factor is bound. Moreover, Holzer *et al.* found through absorption- and fluorescence-spectroscopy, that in LOV-domains from *Chlamydomonas reinhardtii* a small amount of the FMN-chromophore is not bound onto the LOV-domain in the dark [190]. This also might indicate that LOV-domains with a bound FMN-cofactor are in non-equilibrium. In the next sections, we will describe our development of a simulation technique for the description of non-equilibrium properties of photosensory proteins. We used this technique throughout all of our simulations, which we present in this work.

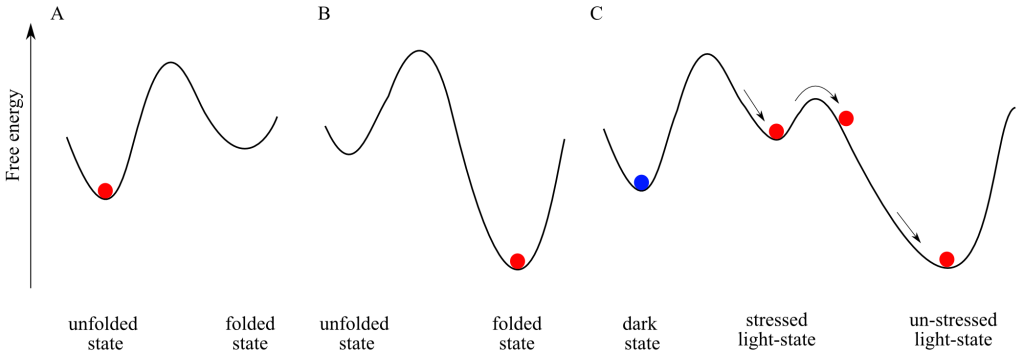


Figure 9.1.: Different states of a protein-system as a function of free-energy. A: Meta-stable folded state. B: Stable folded state. C: Distribution of states in dependence of dark- and light-state of a LOV-domain.

## 9.2. Non-invasive thermostating strategy

The influence of thermostating on the structure and dynamics of proteins is well-known and has been studied extensively in recent years [191]. However, no ideal thermostating strategy has been developed so far, to reproduce protein relaxation under experimental conditions. A standard procedure for controlling the temperature of large biomolecular systems is to make use of the Berendsen thermostat [192]. To maintain the desired temperature, the system is coupled to an external heat bath with fixed temperature $T_0$ by re-scaling the velocities at each timestep, so that the rate of change of the temperature is proportional to the difference in temperature $dT(t)/dt = (T_0 - T(t))/\tau$. In this scheme $\tau$ is an empirical heat bath parameter, which determines the strength of the coupling between the physical system and the heat bath. The procedure provides an exponential decay of the kinetic temperature towards the desired temperature. In practice, the parameter $\tau$ has to be chosen with care, since in the limit $\tau \longrightarrow \infty$ the Berendsen thermostat is inactive and the system recovers temperature fluctuations of magnitude of the microcanonical ensemble, while in the case of a too small $\tau$ it generates unrealistically small temperature fluctuations [193]. The Berendsen thermostat is widely used to equilibrate biomolecular simulations, because of its stability and efficiency in relaxing a system to the target temperature [194]. However, it is well-established that the ensemble generated by this approach is not consistent with the canonical phase-space distribution [194, 192]. Therefore, once the system has reached equilibrium it is more appropriate to sample the correct canonical ensemble, using the Nosé-Hoover thermostat [195, 196, 123]. This method relies on the extended system approach of Nosé [195], which mimics the heat bath by introducing an artificial variable $s$, associated with a fictive mass $Q > 0$, as well as a corresponding velocity $\dot{s}$. The advantage of the Nosé-Hoover thermostat is that it has been proven analytically to generate configurations belonging to the canonical ensemble. Moreover, it can easily be extended to reproduce the exact isothermal-isobaric ensemble [195], generally reflecting the external conditions of standard experimental investigations. However, similarly as in case of the Berendsen thermostat the size of the heat bath parameter $Q$ in this procedure affects the rate of exchange of the thermal energy between the physical system and the heat bath by determining the magnitude of the temperature fluctuations. A standard way to solve this difficulty is to couple the solute and solvent degrees of freedom to individual thermostats and optimize their respective $Q$-parameters in such a way, to ensure a

rapid thermal equilibration of the overall system. However, even if such ideal parameters could be found, a problem still remains when simulating systems composed of large protein domains containing covalently or non-covalently bound prosthetic groups, like the flavoproteins considered in this work. Such systems are inherently heterogeneous, involving many different subsets of degrees of freedom with characteristic relaxation times giving rise to regions of different heat diffusion rates. To ensure ergodicity in the corresponding simulations, it must be guaranteed that all the subsets of degrees of freedom can exchange energy among their various forms within the physical system and with the heat bath. To address this problem, Lingenheil *et al.* [191] recently introduced a non-invasive thermostating strategy, which relies on the idea that an explicit solvent environment represents an ideal thermostat with regard to the magnitude and time correlation of the temperature fluctuations of the solute molecule. Their approach consists in decoupling the solute from the heat bath, which remains solely thermostated by the surrounding solvent by exchanging kinetic energy with the solvent molecules through particle collisions. From test simulations with peptides, they concluded that this temperature control strategy permits to obtain a homogeneous temperature distribution in equilibrated systems, while allowing the solute molecule to sample conformations from the correct statistical ensemble and ensuring a minimal perturbation of its dynamics. Very recently, we have applied the non-invasive thermostating strategy to elucidate the signal transduction pathway of the LOV2-Jα-photosensor from *Avena Sativa* and could demonstrate that this technique allows to sample protein configurations far from equilibrium and to follow its natural dynamics under solvent-mediated thermostating control [246]. In the following we will compare this technique to the more conventional invasive thermostating strategy and analyze its consequences on the solution structure as well as the conformational fluctuations of the heterogeneous protein domain considered herein.

## 9.3. Influence of thermostating strategy on signal-transduction of AsLOV2-Jα

In the following, we compare the influence of the thermostating strategy on the signal transduction pathway of the AsLOV2-Jα photosensor. As we describe in the later chapter 12, the non-invasive thermostating strategy reliably describes the signal-transduction pathway of the photosensory system. By contrast, the invasive strategy cannot reproduce experimental data as we show in the later sections.

### 9.3.1. System-preparation

As a starting configuration, we employed the solution NMR structure of the wild-type LOV2-α-system of *Avena sativa*, which was determined in the 2D-HSQC NMR experiments of Harper *et al.* [53]. After the protein was centered in a cubic box with a boxlength of 7.5 nm the box was filled with 12735 SPC-216 water molecules and 7 supplementary sodium ions were added to render the system neutral. The protonation states of the amino acids and the FMN-chromophore were chosen for a physiological relevant pH-value of 8, as described in section 3.6. In Fig. 9.2 we visualize the secondary structure of the LOV2-Jα-domain with the non-covalently attached FMN-chromophore as well as the Jα-helix at the LOV2-core with relevant amino-acids, which will be discussed later. We point out that we generated the light-state and chose the MD-simulation parameters as described in section 3.6, while we used different thermostating strategies. In order to investigate the signal-transduction dynamics of AsLOV2-Jα under different thermostating techniques, we simulated the system under invasive thermostating (IT) as well as under non-invasive thermostating (NIT). In the IT-strategy we coupled each part of the system to the

Nosé Hoover thermostat with a coupling constant $t_c = 0.1$, while in the NIT-strategy the FMN-chromophore as well as the protein were decoupled from the thermostat and solely the solvent remained coupled after an IT-equilibration for 1 ns. Each system was simulated for a total amount of 20 ns using a timestep of 1 fs. In the following we describe the impact of the different thermostating strategies on the dynamics of AsLOV2-J$\alpha$ in its signaling state.

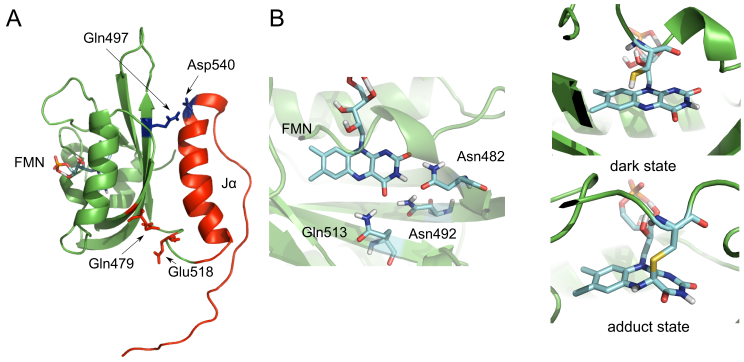


Figure 9.2.: Starting structure of AsLOV2-system from *Avena sativa*. A: Secondary structure of LOV2-domain with J$\alpha$-helix attached onto protein. B: FMN-chromophore and amino-acid environment. Dark- and adduct-state between Cys450 and FMN.

## 9.4. Results

We note that we will discuss the signaling of the AsLOV2-J$\alpha$-system using the NIT-strategy in the later chapter 12, where we demonstrate that the dark-state is in-active and stable throughout the simulation with the use of the NIT-strategy. We start with the analysis of the amino-acids in vicinity of the FMN-chromophore, and compare the IT- with the NIT-simulation (see Fig. 9.3). We observe approximately the same behavior of the NIT and the IT strategy in case of the dihedral angle between the Gln513-sidechain atoms Gln513-CD-OE1 and FMN-C4-N5, as shown in Fig. 9.3A. This dihedral angle performs the same transition at 2.4 ns from 20 to 220 degrees in the NIT- as well as in the IT-simulation. We observe the same switch-like behavior at a simulation time of 7.5 ns, where the dihedral angles alter their value from 220 to 300 degrees in the IT- and the NIT-case. However, after a simulation time of 10 ns, the dihedral angle of the IT-light-state decreases by a value of 30 degrees to a dihedral angle of 260 degrees and remains at this value until the end of the simulation. This is not the case for the NIT-simulation, where the dihedral angle does not alter after the first dihedral-switch at 7.5 ns and remains at a constant average value of 300 degrees. We interprete this behavior of the Gln513-CD-OE1 - FMN-C4-N5 dihedral angle in the IT-case as a damping effect, which impedes the dihedral switch. In Fig. 9.3B we see that the Asn492-OD1 - Gln513-NE2 distance also has approximately the same curvature in the NIT- and the IT-simulation in the time-range from 0 to 7.5 ns. However, this distance increases in the IT-case after a simulation time of 10 ns from a value of 0.3 nm to a value of 0.45 nm, while the distance in the NIT-case remains at 0.3 nm in a timespan from 7.5 ns to 16 ns. This increase in the Asn492-OD1 - Gln513-NE2 distance in the IT-case at a simulation time of 10 ns corresponds to the down-shift in the Gln513-CD-OE1 FMN-C4-N5 dihedral angle at the same time. We conclude, that permanent coupling between Gln513 and Asn492 does not

occur in the IT-case, while in the NIT-case this coupling persists over a simulation-period from 7.5 ns to 16 ns. Moreover, we see in Fig. 9.3C that the cleavage between atoms FMN-O4 and Gln513-NE2 does not reach the same distance in the IT-case than in the NIT-simulation. We see, that the FMN-O4 - Gln513-NE2 distance performs a huge jump from 0.45 to 0.9 nm after a simulation time of 7.5 ns in the NIT-case. This jump does not occur in the IT-case, where the distance increases in a drift from 7.5 ns to 8.5 ns from 0.4 to 0.78 nm. We observe that the FMN-O4 - Gln513-NE2 distance remains after 7.5 ns at an average distance of 0.6 nm in the IT-case, while the same distance reaches an average value of 0.9 nm in the NIT-simulation. We see in the distance between atoms Asn482-ND2 - FMN-O4, shown in Fig. 9.3D, that Asn482-ND2 binds onto FMN-O4 in both cases approximately at the same simulation time of 7.5 ns. However, this distance of 0.3 nm is stabilized in the IT-case, while in the NIT-case a second up-shift to a value of 0.5 nm occurs at 16 ns. We conclude, that the light-state signal is solely triggered in the NIT-case, which we see in the change in the dihedral angle and the coupling between Asn492 and Gln513. As we see later in chapter 12, these effects trigger both $\beta$-sheet tightening between the H$\beta$ and the I$\beta$-strands, which leads to an N-terminal disruption of the J$\alpha$-helix in the signaling state. However, we observe in the IT-case that the behavior of the amino-acids Gln513 and Asn492 is dampened by the thermostat, although the same dynamics in the beginning of the simulation from 0 to 7.5 ns are obtained with both strategies. We continue with the analysis of the inter-atomic distance between Asp540-OD1-OD2 and Gln497-NE2 at the interfacial region between the AsLOV2-core and the J$\alpha$-helix, as visualized in Fig. 9.4. In this distance we observe that cleavage between both amino-acids occurs in the NIT-case at a simulation time of 3 ns, where the distance increases from a value of 0.4 nm up to 1 nm. By contrast, this distance does not alter in the IT-strategy and remains at a constant value of 0.35 nm. We conclude, that no cleavage between J$\alpha$ and AsLOV2-core occurs in the IT-case, while we observe an immense increase in the Asp540-OD1-OD2 - Gln497-NE2 distance using the NIT-strategy. We point out that during this whole simulation phase of 20 ns, where the NIT strategy was applied, the instantaneous temperature was at an average temperature of 300.067 $K$, mediated by the fully thermostated solvent, as shown in Fig. 9.5. However, we observe that the averaged residue-based temperature of Asp540 in the NIT-strategy has a value of 301.32941 $K$, while the IT-case is at an average value of 300.42396 $K$. We point out that this residue based temperature in the IT-case is 0.9 $K$ lower than the temperature of the NIT-strategy. We conclude, that the IT-case does not allow higher temperatures and dampens the natural fluctuations of the protein, which we obtain through application of the NIT-strategy.

We continue with the description of the final structures after 20 ns of MD-simulation in Fig. 9.6. In Fig. 9.6A we show the final structure of the IT-strategy and in Fig. 9.6B the final configuration of the NIT-strategy, respectively. We observe in the IT-case, that the amino-acids located in the N- and in the C-terminal region of the J$\alpha$-helix remain H-bonded, whereas this helix is mainly cleaved in the NIT-case. In these final structures we see the overall effect of the thermostating strategy. In the IT-case the J$\alpha$-helix is conserved, whereas in the NIT-case disruption of the same helix occurs. We point out that the NIT-simulation of the dark-state remains stable, which we further explain in chapter 12. As demonstrated by Harper *et al.*, the J$\alpha$-helix is disrupted in the light-state of AsLOV2-J$\alpha$, which is well reproduced by the NIT-simulation [53]. We continue with the description of the secondary structure evolution of both simulations in Fig. 9.7. We observe in the IT-case shown in Fig. 9.7A, that the J$\alpha$-helix remains conserved, and changes in the amino-acid environment of the FMN-chromophore do not lead to a disruptive alternation in the secondary structure of this helix. By contrast, the J$\alpha$-helix alters its secondary structure at 4.75 ns and at 14.5 ns in the NIT-case, as shown in Fig. 9.7B. These disruptive events occur at the same time as the chromphore-environment re-arranges according to adduct formation, as desribed in the later chapter 12. We point out that the triggering events

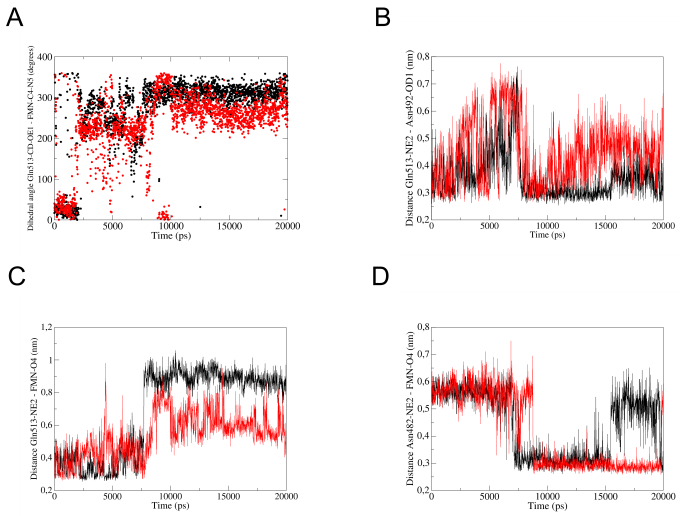


Figure 9.3.: Time-dependent inter-atomic distances and dihedral-angles in vicinity of FMN-chromophore. IT (red), NIT (black). A: Dihedral angle Gln513-CD-OE1 - FMN-C4-N5. B: Distance Asn492-OD1 - Gln513-NE2. C: Distance Gln513-NE2 - FMN-O4. D: Asn482-ND2 - FMN-O4. (red) IT-case; (black) NIT-case.

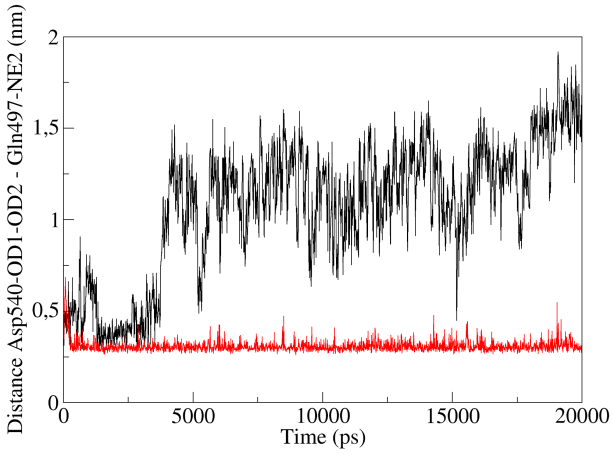


Figure 9.4.: Time-dependent inter-atomic distances between Asp540-OD1-OD2 - Gln497-NE2 at the protein-J$\alpha$-interphase.

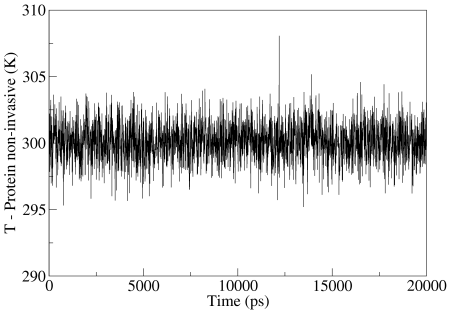


Figure 9.5.: Instantaneous temperature of the AsLOV2-J$\alpha$-protein during 20 ns of simulation using the NIT technique.

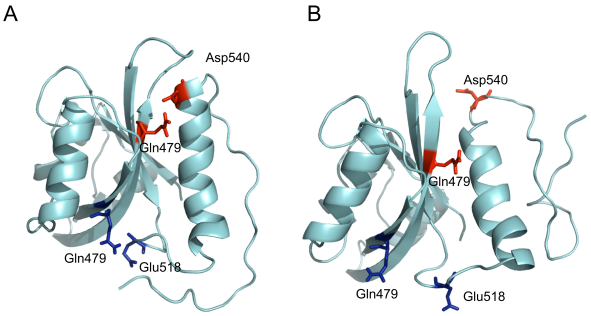


Figure 9.6.: Final structures after 20 ns in secondary structure representation of IT strategy (A) and NIT strategy (B).

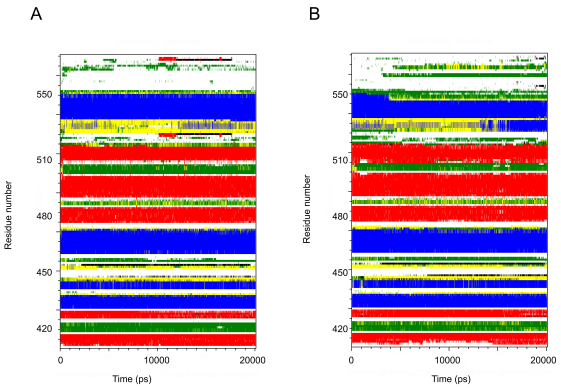


Figure 9.7.: Secondary structure analysis of 20 ns of IT simulation (A) and NIT simulation (B).

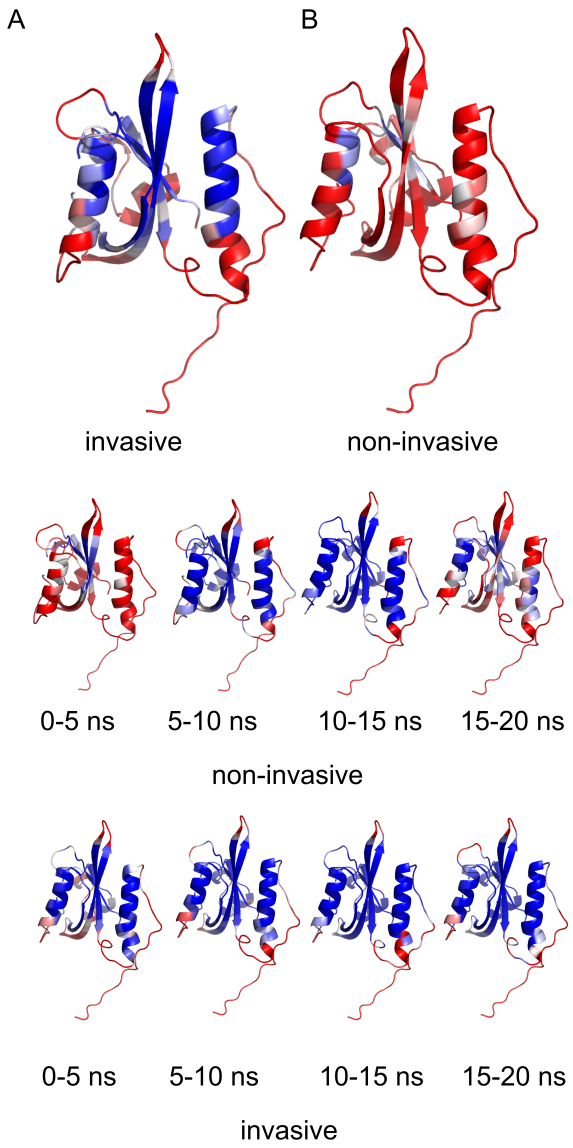


Figure 9.8.: B-factor distribution over secondary structure of LOV2-domain of *Avena sativa* of IT case (A) and NIT case (B).

in the IT-case do not lead to any disruptive event between LOV2-core and J$\alpha$-helix. In Fig. 9.8 we describe the differences in temperature evolution of the protein system in both cases through a residue-based B-factor colouring of the protein's secondary structure. Here we see the clear differences between both strategies. Whereas the temperature-pattern during the IT-strategy does not alter at all and nearly all regions have the same temperature, the NIT-strategy leads to a tolerance in the distribution of temperature over the protein-core and the J$\alpha$-helix. We point out that when the signal is triggered in the NIT case by the coupling process between Gln513 and Asn492 in the range from 0 to 5 ns, the overall protein-temperature is larger compared with the temperature-distribution in the time-range between 5 to 15 ns. Moreover, we see that when final cleavage in the NIT-case occurs in the timerange from 15 to 20 ns, the temperature increases again in the C- and N-terminal regions of the J$\alpha$-helix. We point out that these differences in the temperature distributions do not occur in the IT case. Moreover, in the IT-case the temperature remains equally distributed over the whole time-range of 20 ns. In the following chapters, we will demonstrate the applicability of our new NIT-approach on different protein photosensors in comparison with the experiment.

# Part III.

# Application of natural LOV-based photosensors

# 10. MD-study on dark-state of LOV1 from Chlamydomonas reinhardtii

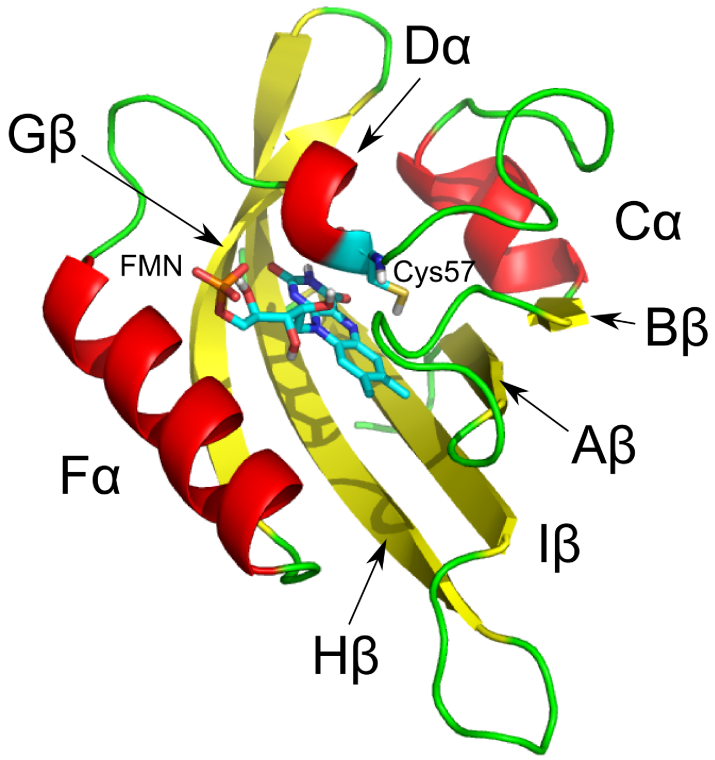


Figure 10.1.: Protein secondary structure of the wild-type LOV1-domain from *Chlamydomonas reinhardtii*, obtained by Fedorov *et al.* using x-ray diffraction [11] (MD starting structure).

## 10.1. Introduction

In the following sections, we examine the effect of different electrostatics techniques and different thermostating schemes onto the dark-state structure of the LOV1-domain of *Chlamydomonas reinhardtii*. We test the IT as well as the NIT-strategy, which we have previously introduced in chapter 9, where we demonstrated the experimentally verified dynamics on the LOV2-J$\alpha$-system from *Avena sativa* through using the NIT-strategy. Moreover, we apply different electrostatics onto this LOV1-system, in order to test if the very cost efficient twin-range cutoff technique

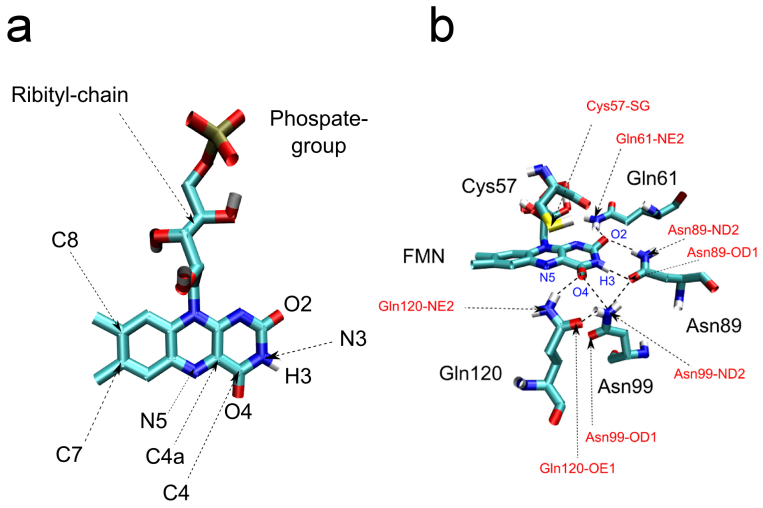


Figure 10.2.: (a) FMN and (b) crystal conformation of FMN with surrounding amino acids, obtained by Fedorov *et al.* using x-ray diffraction [11] (MD starting structure).

might reproduce the results of the more costly particle-mesh-ewald technique. Finally, we verify our simulation results with the experiment of several authors and subsequently the adequate method for the dynamical description of these photosensor proteins. As a starting configuration, we employed the protein crystal structure of the wild-type LOV1-domain of *Chlamydomonas reinhardtii*, which was determined in the x-ray diffraction experiments of Fedorov *et al.* [11] (PDB-code: 1N9L). The corresponding secondary structure as well as the local environment of amino acids surrounding the FMN chromophore are visualized in the Figs. 10.1 and 10.2, respectively. The protein was centered in a cubic box with a boxlength of 6.09 nm, and the box was filled with 7005 SPC-216 water molecules. The total charge of the system was neutralized by adding one supplementary sodium ion. The protonation states of the amino acids as well as the one of the phosphate group at the terminal end of the oxetyl chain of FMN were chosen for the physiological relevant pH-value of 8, which under these conditions means a fully deprotonated phosphate group. MD simulation parameters were chosen as described in section 3.6.

### 10.1.1. Results and discussion

We start our investigations of the influence of the long-range electrostatics on the conformational heterogeneity of dark-state conformers by performing MD simulations with the PME as well as TRC technique in conjunction with the NIT strategy introduced previously. In Fig. 10.3 we present results for the time-dependent behavior of characteristic inter-atomic distances of amino acids, either involved in the adduct formation process or forming H-bonds with the FMN chromophore. In Fig. 10.3a we show the distance between the adduct forming reaction centers, i.e. the sulfur atom Cys57-S on the apo-protein and the FMN-C4a atom on the chromophore, as a function of simulation time. We see that the curve, generated with the TRC technique, jumps from a distance of about 4.5 Å to 7 Å at around 2.5 ns after the start of the production phase, while the curve, obtained with the PME technique, provides a similar jump after 17 ns. In the

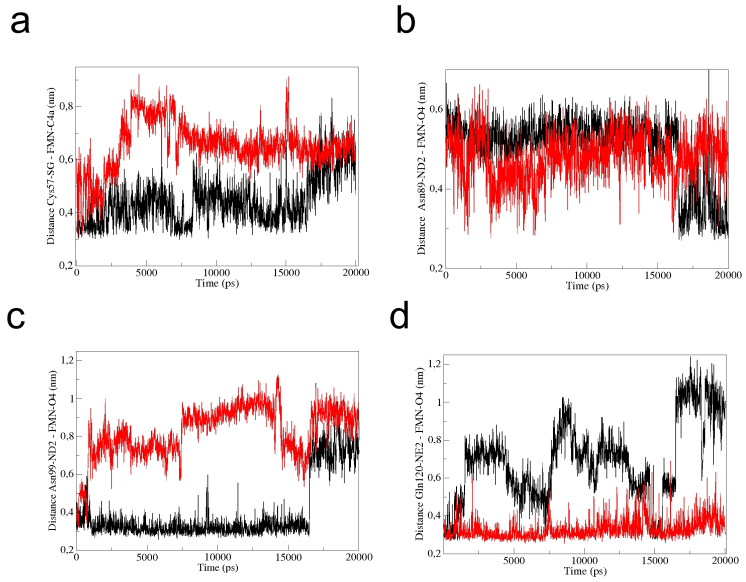


Figure 10.3.: Distance of characteristic atoms on the FMN chromophore to atoms of the surrounding amino acids, calculated with different electrostatic calculation techniques (black: NIT PME; red: NIT TRC) in conjunction with the NIT scheme [(a) Cys57-S - FMN-C4a. (b) Asn89-ND2 - FMN-O4. (c) Asn99-ND2 - FMN-O4. (d) Gln120-NE2 - FMN-O4].

next three plots of Fig. 10.3 we visualize the distances of the H-bond forming acceptor atom FMN-O4 with the donor atoms of the surrounding amino acids Asn89-ND2, Asn99-ND2 as well as Gln120-NE2. In case of the TRC technique we observe that the distances between the atoms Asn89-ND2 and FMN-O4, as well as the atoms Gln120-NE2 and FMN-O4, remain stable over the whole simulation time, while the distance between the atoms Asn99-ND2 and FMN-O4, after a sharp increase at around 1 ns, fluctuates about an average value of 8 Å with a large magnitude. By contrast, in case of the PME technique the distances between the atoms Asn89-ND2 and FMN-O4, as well as the atoms Asn99-ND2 and FMN-O4, fluctuate with a moderate magnitude around their average values up to 17 ns, followed by a sharp transition and stabilization around new average values. This sudden structural re-arrangement of the H-bond network concords well with the change in distance between the reaction centers Cys57-S and FMN-C4a at the same simulation time. Moreover, we observe by comparing the results from the TRC and PME technique that after the simulation time of 17 ns nearly the same distances are reached for the reaction centers Cys57-S and FMN-C4a, as well as the H-bond centers Asn99-ND2 and FMN-O4. To analyze these local structural changes on a molecular basis, we visualize in Fig. 10.4 the representative structures before and after the re-arrangement of amino acids, generated with the TRC and PME technique. If we first compare the two conformations in the Figs. 10.4a and 10.4b obtained with the TRC simulations before and after the structural re-arrangement at 2 ns, we notice that the arrangement of the amino acids with respect to FMN remains essentially the same, but the distances of Cys57 and Asn99 with respect to FMN become larger in agreement with the observations made in Fig. 10.3. By next comparing the conformation obtained with the PME technique before the structural re-arrangement, i.e. Fig. 10.4c, with the crystal structure

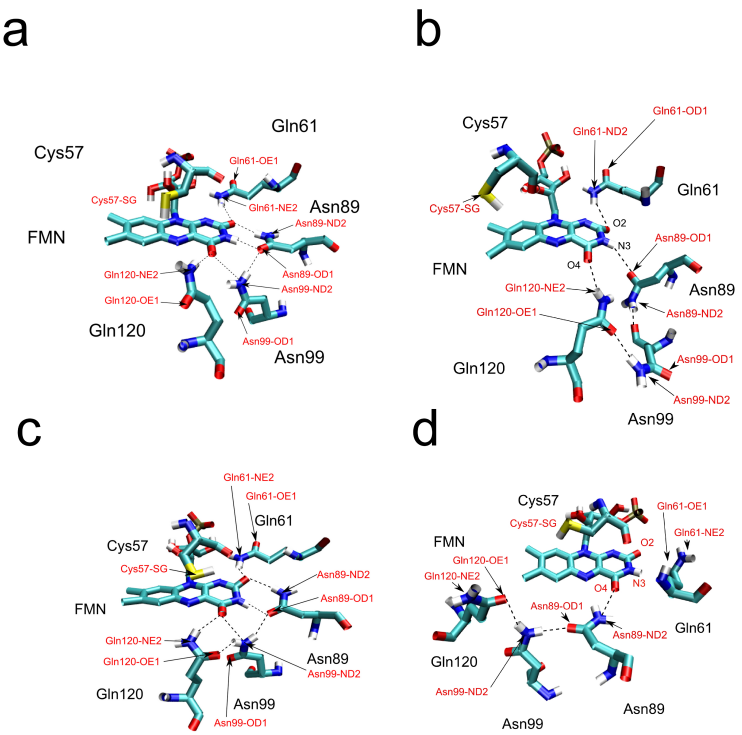


Figure 10.4.: Representative local configurations of amino acids surrounding the FMN chromophore, obtained with the NIT TRC and PME technique [(a) configuration at start of production phase with TRC NIT. (b) configuration at 20 ns with TRC NIT. (c) configuration at start of production phase with PME NIT. (d) configuration at 20 ns with PME NIT].

of Fedorov *et al.* in Fig. 10.2b, we conclude that in the initial part of the production phase the amino acids surrounding the FMN remain stable and fluctuate around the average positions, preset by the starting crystal configuration. In particular, we note that in this time range the carbonyl-oxygen atoms O2 and O4 on FMN are twice H-bonded by the hydrogen atoms of the amino acid pairs Gln120 and Asn99 as well as Asn89 and Gln61, respectively. By subsequently comparing the conformations obtained with the PME technique before and after the structural re-arrangement in the Figs. 10.4c and 10.4d, we see that the amino acids Gln120, Asn99, Asn89 and Gln61 in the latter case have been displaced with respect to the FMN plane, resulting in the formation of a new H-bond network and leaving the orientations of the amino acids with respect to each other unchanged. Moreover, we note that the carbonyl-oxygen atom O4 on FMN, which was originally twice H-bonded with the surrounding amino acid pair Asn99 and Gln120, becomes now singly H-bonded with the amino acid Asn89 in the time range after 17 ns. This observation is sustained by the drop in the distance at 17 ns of the H-bond centers Asn89-ND2 and FMN-O4 in Fig. 10.3b. Moreover, we further deduce from the Figs. 10.4c and 10.4d that the thiol group of the Cys57-residue, which was originally located close to the reactive C4a-atom, has in the new configuration re-oriented towards the aromatic benzene ring of the FMN. To analyze this issue in more detail, we present in Fig. 10.5 the distance between the reactive

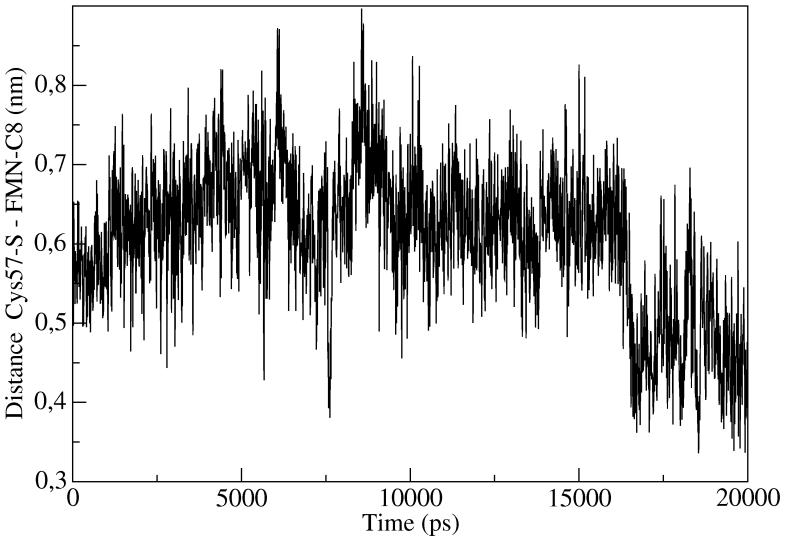


Figure 10.5.: Distance between the atoms FMN-C8 and Cys57-S, obtained with the NIT PME technique.

C8-position on the benzene ring of FMN and the reaction center Cys57-S, calculated with the NIT PME technique. We deduce from the figure that in the time range after 17 ns, the thiol group gets about 4 Å close to the FMN-C8 position, which enables orbital overlap between the new reaction centers and hinders at the same time the process of adduct formation as we will see in the further development. Therefore, the abrupt re-arrangement of the amino acids Cys57, Asn89, Asn99 and Gln120 in vicinity to the FMN indicates the existence of an additional dark-state configuration. This configuration is only temporary stable and can switch reversibly back to the original configuration. The latter conclusion is supported by our simulation results for the dimers, composed either from LOV1-domains in their dark or light-state form. In the Figs. 10.11a and 10.11b we present the distances between Asn89-ND2 and FMN-O4 in both monomers of the dark-state (black) and light-state (red) dimers, respectively. From both dark-state curves, we deduce that the distances reversibly jump from about 5.5 Å to 3.5 Å in both monomers of the dark-state dimer, which agrees well with the behavior of the same quantity in Fig. 10.3b of the dark-state monomer. In the time ranges, where the distance between Asn89-ND2 and FMN-O4 in the dimer is at 3.5 Å, the carbonyl-oxygen atom O4 on FMN becomes singly H-bonded with the amino acid Asn89, which is characteristic for the second dark-state conformer observed in the monomeric LOV1-domain at 17 ns in Fig. 10.3b. By contrast, in the time ranges, where the distance between Asn89-ND2 and FMN-O4 is at 5.5 Å, Asn89-ND2 forms an H-bond with the carbonyl-oxygen atom O2, which is characteristic for the first dark-state conformer. The second conformer can be suppressed upon illumination through adduct formation, as we can deduce from the corresponding results for the light-state dimer in the Figs. 10.11a and 10.11b. From these observations, we conclude that both dark-state configurations must be part of an equilibrium distribution of dark-state conformers, which can be influenced through light activation. These findings concord well with the results from the experiments of Kennis *et al.* [46], in which the molecular mechanism of signal transduction of the phot1 LOV2-domain from *Avena sativa* AsLOV2-J$\alpha$ has been investigated using FTIR spectroscopy. In their study a structural heterogeneity for AsLOV2-J$\alpha$ has been reported, involving two different equilibrium populations of C4=O conformers coexisting in the dark. These two conformers were characterized

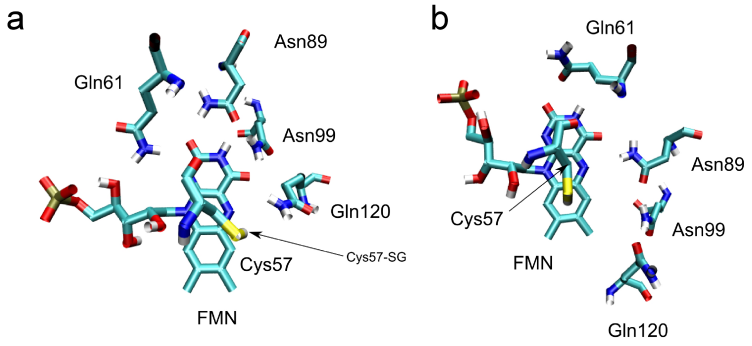


Figure 10.6.: Representative configurations showing the orientations of the Cys57-thiol group with regard to the FMN ring (a) before and (b) after the structural re-arrangement, obtained with the NIT PME technique.

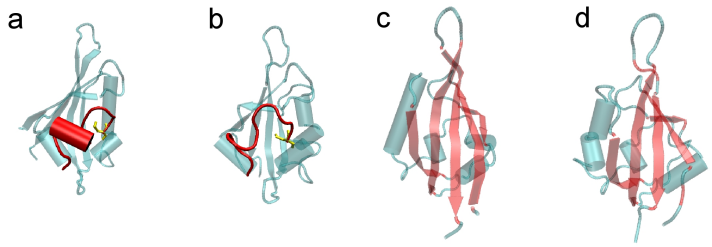


Figure 10.7.: Protein structure of the wild-type LOV1-domain at different times using the NIT PME simulation technique [(a) hydrophilic side at start of production phase. (b) hydrophilic side at 20 ns. (c) hydrophobic side at start of production phase. (d) hydrophobic side at 20 ns].

by different C4=O-carbonyl stretching frequencies, which could be attributed either to a singly or twice H-bonded C4=O. Moreover, they have been found to display slightly shifted UV-Vis absorption spectra, resulting in a structured absorption band at 475 nm at room temperature and a splitting of this band at low temperatures. Kennis *et al.* proposed that in case of the singly H-bonded conformer an H-bond is formed between the hydrogen of the amino group of Gln120 and FMN-O4, while in our study we find in contrast that an H-bond is formed between the amino acid Asn89 and FMN-O4. In the following we will analyze in more detail the structural changes involving the amino acid Cys57, because of its role as a reactive residue in the adduct formation process. To this end, we show in Fig. 10.6 representative structures corresponding to different thiol orientations before and after the structural re-arrangement. By comparing both structures, we recognize that through the structural re-arrangement the Cys57-thiol group rotates about an angle of 50° with regard to the FMN longitudinal axis and tends towards the C8-position of the FMN. It is worth noting in this context that such configurations have been proposed in various experimental works. In their x-ray diffraction experiments Fedorov *et al.* [11] detected two Cys57-rotamers with an occupancy ratio of about 70 % : 30 %, where in the more occupied conformation the thiol group of Cys57 is located about 3.8 Å from FMN-C5a and 4.4 Å from FMN-C4a, while in the less occupied conformation the thiol group is located about 3.5 Å from FMN-C4a. These findings have been confirmed by the FTIR difference spectroscopy

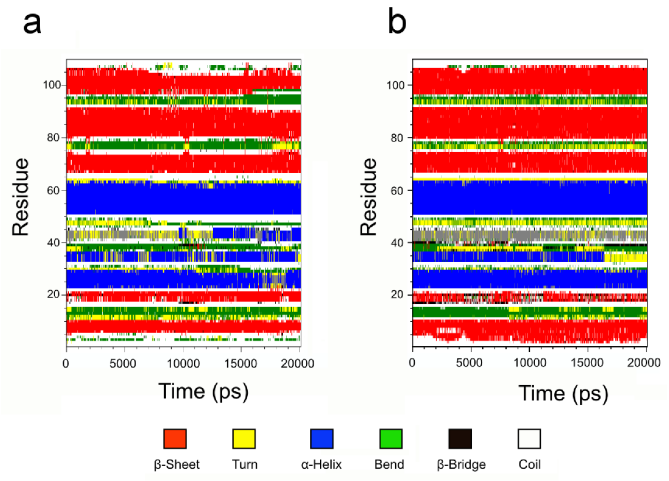


Figure 10.8.: Secondary structure analysis of NIT simulation trajectories of the wild-type LOV1-domain, using different electrostatic calculation techniques [(a) TRC technique. (b) PME technique].

experiments of Bednarz *et al.* [41]. They found that the two conformations are characterized by two different S-H vibrational modes of Cys57, which can be assigned to an H-bond-free state and an H-bonded state. In our simulation study we find that the FMN-N5-atom stays H-bonded with the reactive Cys57-residue in the time range before the structural re-arrangement, while the H-bond gets disrupted in the time range after the re-arrangement. Kennis *et al.* [197] suggested in case of LOV2 that the presence of the thiol group of the reactive cysteine residue in vicinity of the aromatic ring of FMN shifts the resonance equilibrium away from the oxidized form of FMN to one of its quinoid forms, which is depicted in Fig. 10.12. They concluded this by comparing vibrational spectra from fluorescence line narrowing experiments of the wild-type LOV2 species AsLOV2 and Phy3LOV2 with a cysteine-less mutant AsLOV2-C450A [197]. They found that the mutant shows small but significant frequency shifts of the FMN-ring vibrations as well as of the C2=O mode with respect to the wild-type species, similar to those found in FMN-model systems with an electron-donating group substituted at the C8-position of the FMN. Such substituents are known to induce a quinoid character to the electronic structure of oxidized flavin [198, 199, 200, 201], although with smaller absolute frequency shifts than in case of the non-covalently bound cysteine residue. Kennis *et al.* concluded that the induction of the quinoid resonance structure by the Cys57-residue may consequently strengthen the H-bond between the amino group of Asn89 and FMN-C2=O and stabilize in this way the H-bond network configuration that enables adduct formation. Moreover, they proposed that it increases the intersystem crossing (ISC) rate by rapid generation of the reactive FMN-triplet state, leading to a shortening of the FMN-singlet excited-state lifetime. These conclusions are consistent with several experimental studies, which state that the presence of the reactive cysteine at about 4 Å of the FMN [205, 11, 55] is responsible for a shortening of the singlet excited state lifetime and increase of the ISC rate by a factor of 2-3 as compared to FMN in solution [202, 197] or cysteine-less mutants [203, 204]. In this context it is worth pointing out that in LOV-domains an efficient ISC rate is crucial. This relates to the fact that the biologically active adduct state is generated from the first excited triplet state of FMN, which is itself formed through ISC from its

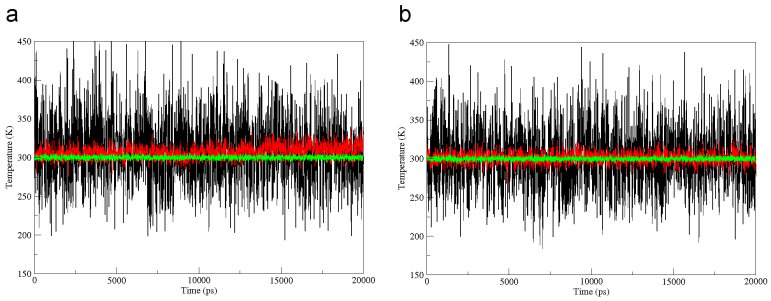


Figure 10.9.: Kinetic temperatures from the FMN chromophore (black), the apo-protein (red) as well as the solvent (green) of the wild-type LOV1-domain-solvent system, obtained by using the NIT strategy in conjunction with different electrostatic calculation techniques [(a) TRC technique. (b) PME technique].

first excited singlet state with a high quantum yield of about 60-80 % [206, 202, 1, 203, 207, 208]. In several experimental studies [202, 203, 11] it has been speculated that the enhanced ISC rate in LOV-domains is caused by a heavy-atom effect, induced by an interaction between the cysteine sulfur and the flavin isoalloxazine ring. This issue has subsequently been investigated in more detail in a series of theoretical studies. Dittrich *et al.* [48] performed quantum chemical calculations for the first excited triplet state of the FMN-related model compound lumiflavin at the ROHF/6-31G(2d,2p) level of theory and concluded from analyzing resulting Mulliken spin populations that a significant fraction of unpaired spin density is localized on the C4a- as well as the C8-position of lumiflavin. Salzmann *et al.* [209, 210] demonstrated that not only the ordering of low-lying potential energy hypersurfaces, but also the mechanism of triplet generation in case of FMN-related compounds significantly depends on the environment. In a recent work Salzmann *et al.* [211] explored the ground and low-lying excited states of FMN in the LOV-domain of the blue-light photosensor YtvA of *Bacillus subtilis* by means of combined QM/MM methods. They found increased spin-orbit coupling between the initially populated S1-state and the T1- and T2-states in YtvA-LOV, compared to free lumiflavin in water. In addition, an external heavy-atom effect was observed when the sulfur atom of the nearby cysteine residue was included in the QM region, which is in line with experimental findings. The electronic interaction between the $\pi_{H-1}$-molecular orbital of the isoalloxazine ring and the $p_S$-lone pair on sulfur was particularly pronounced for the conformation, where the thiol group of the reactive cysteine residue C62 tends towards the aromatic ring away from the C4a-position of FMN. Similar structural heterogeneities of the reactive cysteine residue have also been found in MD investigations with LOV2-domains [43, 50]. However, no structural re-arrangement of the surrounding amino acids have been detected in these cases. In summary, based on our calculation results and the conclusions from the comparative works presented previously, we retain that the LOV1-domain from *Chlamydomonas reinhardtii* is able to alter its ISC rate from the first excited singlet state to the first excited triplet state by adjusting the orientational distribution of the thiol group of the Cys57-residue towards the C4a- and C8-position. The regulation takes place by stabilizing the unpaired spin fractions at the sites FMN-C4a as well as FMN-C8 through electronic interaction between the $p$-lone pair sulfur orbitals of Cys57 with the $\pi$-molecular orbitals of the FMN.

To understand the effect of the change of the local H-bond network surrounding the FMN on the overall secondary structure of the LOV1-domain, we study next the overall protein conformations for different local arrangements of amino acids, obtained in different time ranges of our NIT

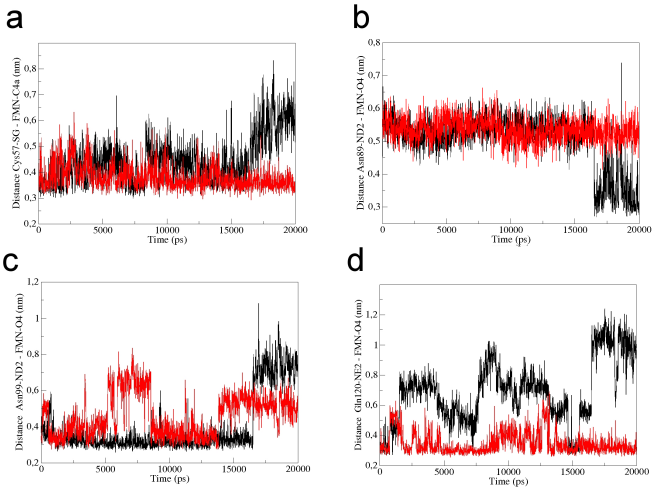


Figure 10.10.: Distance of characteristic atoms on the FMN chromophore to atoms of the surrounding amino acids, calculated with the PME technique using the NIT (black) and IT (red) thermostating strategy [(a) Cys57-S - FMN-C4a. (b) Asn89-ND2 - FMN-O4. (c) Asn99-ND2 - FMN-O4. (d) Gln120-NE2 - FMN-O4].

PME simulations. In the following we first examine the hydrophilic side of the protein domain by visualizing in the Figs. 10.7a and 10.7b the protein conformations, obtained before and after the structural re-arrangement of amino acids respectively. From the figures, we observe that a quite large $\alpha$-helical element, comprising the sequence of amino acids from 32 up to 38 and denoted as D$\alpha$ in Fig. 10.1, disappears in the second dark-state conformer after the structural re-arrangement at 17 ns and is converted into an extended unstructured protein chain. This process goes along with the re-orientation of the thiol group of the Cys57-residue towards the FMN-C8 position on the benzene ring of the FMN as well as the re-arrangement of the amino acids Asn89, Asn99 and Gln120 surrounding the FMN. Next, we consider in the Figs. 10.7c and 10.7d the corresponding hydrophobic side of the LOV1-domain, essentially characterized by the alignment of the $\beta$-strands. We observe that, due to the loss of structure of the D$\alpha$-helix, the tightening of the $\beta$-sheet is reduced and the flexibility of the $\beta$-strands is augmented in the second dark-state conformer, which increases the ability of the LOV1-domain for dimerization through this site with other LOV-domains or the kinase. As demonstrated in the Figs. 10.11a and 10.11b on the example of dimers formed from the dark-state LOV1-domains, this conformer is only temporary stable both in its monomeric and dimeric form, and can reversibly switch back to the original configuration. This indicates that the LOV1-domain is able to modulate its dimerization tendency in the dark through this $\alpha$-helical element by increasing its capacity to adjust onto the hydrophobic surface of its respective partner domain. These conclusions are confirmed through our simulation results for various $\alpha$-helical elements of the LOV1 dimer presented in Fig. 10.13, where we considered both monomers either in their dark or light-state form. In these graphs we show the center-of-mass distances as a function of simulation time between $\alpha$-helices located on different LOV1-domains. From the Figs. 10.13a and 10.13b, we deduce that after an initial adjustment phase the C$\alpha$- and F$\alpha$-helices, flanking the $\beta$-sheets in the monomers, come more close to each other by about 3 Å at 10 ns in the dark-state dimer compared to the light-state dimer. This enhances the affinity of the LOV1-domains for docking via their accommodated hydrophobic surfaces on the $\beta$-sheets. This docking process goes along with an increase of the

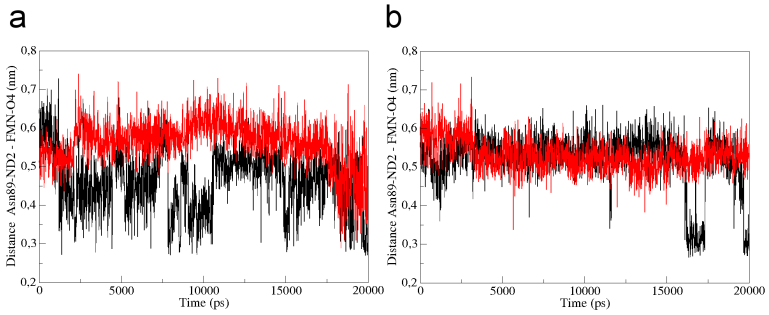


Figure 10.11.: Distance between Asn89-ND2 and FMN-O4 in the dimer formed from LOV1-domains in their dark-state (black) or light-state (red) form, obtained with the NIT PME simulation technique [(a) monomer 1. (b) monomer 2].

distance of the D$\alpha$-helices by about 5 Å in Fig. 10.13c, which together with the increased distance between the E$\alpha$-helices in Fig. 10.13d, results in an overall enlargement of the volume of the dark-state dimer with respect to the light-state dimer. The previous analysis is confirmed in Fig. 10.14, where we show the dimeric structures of the LOV1-domains in their dark and light-state form after a simulation time of 20 ns. Moreover, from the fact that the structural re-arrangement of amino acids surrounding the FMN chromophore can be suppressed through light activation, we infer that the distribution of dark-state conformers might play a role in controlling the docking affinity. In this context it is worth pointing out that our conclusions concord well with several recent experimental works [3, 212], where it has been suggested that the alignment of the $\beta$-sheets is responsible for controlling the dimerization process of the LOV1-domains and might regulate in this way the signal intensity of the phototropin under different light conditions. Since the equilibrium of dark-state conformers of the LOV1-domain can be influenced by light activation or a change of external conditions like e.g. a change in temperature, shifting their population distribution might permit to the LOV1-domain from *Chlamydomonas reinhardtii* to act as a modulator, which may allow a fine-tuning of its signal response to a change in its environment.

In order to assess the influence of different electrostatics on the conformational fluctuations and the effect of the structural re-arrangement of amino acids near the chromophore on the secondary structure of the LOV1-domain, we consider next in Fig. 10.8 the results from the secondary structure analysis of the MD trajectories, obtained with the NIT TRC and PME simulation technique. By comparing the parts of the sequence forming $\beta$-strands, we observe that, although the overall secondary structure is preserved with both simulation strategies, the TRC technique provides larger fluctuations in comparison to the PME technique, leading to a reduction in size of the $\beta$-strands. Similar reduction in size can be observed for the $\alpha$-helical elements known as the F$\alpha$- and C$\alpha$-helices, comprised between the residue numbers 50 and 64 as well as 22 and 30 respectively. Another clear difference between both techniques appears in the secondary structure in vicinity of the chromophore. While the TRC technique favors the formation of the E$\alpha$-helix after 12 ns within the range of residue numbers 41 to 46, the PME technique favors the disappearance of the D$\alpha$-helix after 17 ns, which directly correlates with the structural re-arrangement of the amino acids around the chromophore depicted in the Figs. 10.4c and 10.4d. The direct response of the secondary structure to local re-arrangements of the H-bond network in conjunction with the preservation of the remaining secondary structure of

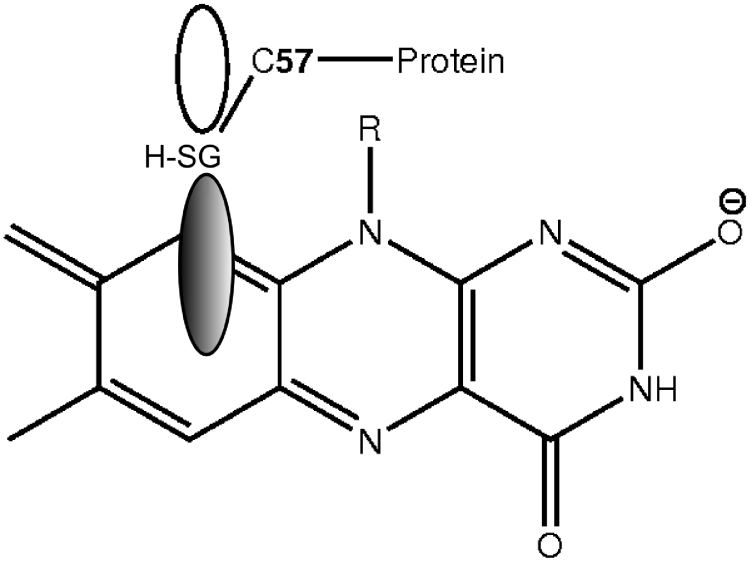


Figure 10.12.: Electronic structure of the quinoid form of FMN, induced by the influence of Cys57-S on the benzene ring of FMN.

the protein domain indicates that the NIT PME technique has generated an additional dark-state conformer of the wild-type LOV1-domain in the time range after 17 ns. By contrast, the formation of the E$\alpha$-helix in case of the TRC technique cannot be correlated with a local re-arrangement of the amino acids near the chromophore, as can be deduced from the Figs. 10.4a and 10.4b. The generation of this new helical element appears to take place at the expense of a weakening of the binding interaction with the chromophore and a strengthening of low-mode fluctuations, leading to a partial destabilization of the remaining secondary structure. This tendency can further be confirmed by regarding the kinetic temperature across the LOV1-domain-solvent system, which is an important indicator for assessing the stability of the simulation and efficiency of the thermostating scheme. To this end, we show in Fig. 10.9 the time-evolution of the kinetic temperatures from the FMN chromophore (black), the apo-protein (red) as well as the solvent (green), obtained by using the NIT strategy in conjunction with the TRC as well as PME technique. In case of the TRC technique we observe a slow drift of the temperature of the FMN chromophore and protein, which relates to the fact that the system is unstable due to the electrostatic calculation technique. This behavior is also reflected by considering the average kinetic temperatures produced by the TRC technique: $< \mathrm{T}^{\mathrm{FMN}}_{\mathrm{TRC}} >= 308.262 \pm 42.5179$ K, $< \mathrm{T}^{\mathrm{protein}}_{\mathrm{TRC}} >= 307.428 \pm 8.66738$ K and $< \mathrm{T}^{\mathrm{solv}}_{\mathrm{TRC}} >= 300 \pm 2.12618$ K. By contrast, in case of the PME technique the kinetic temperatures are perfectly stable over the entire simulation run and fluctuate around the externally imposed temperature of 300 K, differing solely in the magnitude of the fluctuations. The PME technique provided the following average temperatures: $< \mathrm{T}^{\mathrm{FMN}}_{\mathrm{PME}} >= 300.871 \pm 41.2056$ K, $< \mathrm{T}^{\mathrm{protein}}_{\mathrm{PME}} >= 301.082 \pm 7.52788$ K and $< \mathrm{T}^{\mathrm{solv}}_{\mathrm{PME}} >= 300 \pm 2.12519$ K. The larger fluctuations of the FMN chromophore can easily be explained by the fact that in the dark-state of the LOV1-domain the chromophore is unbounded in the binding pocket, which results in stronger fluctuations. In conclusion, these observations provide further evidence that the NIT PME technique is more reliable compared to the TRC technique in reproducing the structures and conformational behavior of the LOV1-domain-solvent system.

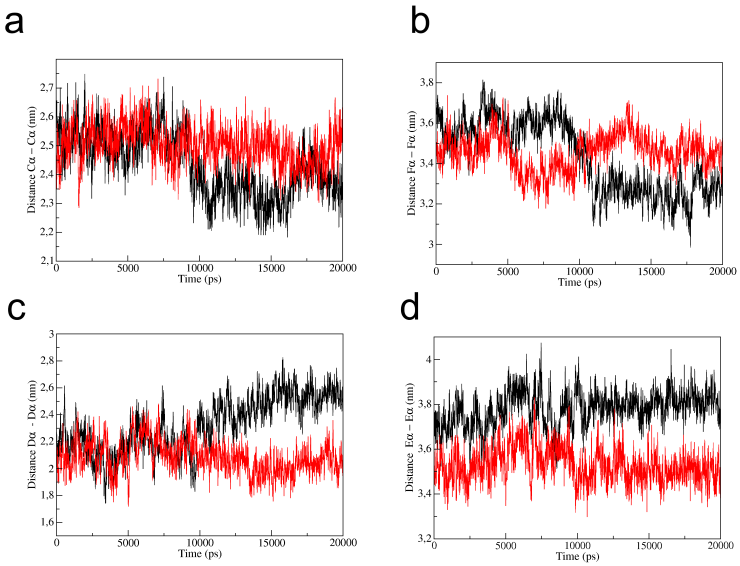


Figure 10.13.: Center-of-mass distance between $\alpha$-helices on different monomers from the dark-state (black) and light-state (red) dimer [(a) C$\alpha$-helix - C$\alpha$-helix. (b) F$\alpha$-helix - F$\alpha$-helix. (c) D$\alpha$-helix - D$\alpha$-helix. (d) E$\alpha$-helix - E$\alpha$-helix].

As we discussed in Section 9.4, the protein dynamics, besides the electrostatics, can also be tremendously affected by the thermostating procedure. To investigate this aspect, we performed simulations with the PME technique, using the IT and NIT strategy introduced in the same section. The effects on the dynamics are shown in Fig. 10.10, where we plot the distances of characteristic atoms on the FMN chromophore to atoms of the surrounding amino acids. From the figures, we deduce that both thermostating strategies provide a different dynamical behavior of the amino acids near to the chromophore. The IT strategy favors equilibrium fluctuations around the starting crystal configuration by enforcing the same temperature in each part of the system at any time of the simulation. This causes that configurations far from this local equilibrium configuration are suppressed and the system cannot drift towards a new equilibrium configuration. The NIT strategy, by contrast, avoids this problem by only thermostating the solvent, which permits the system also to sample configurations far from equilibrium. This is reflected by the sudden local re-arrangement of the amino acids at 17 ns and the emergence of a new local equilibrium configuration. By considering that both configurations are in agreement with the experimental observations discussed previously, we conclude that the NIT strategy provides a better thermostating strategy for simulating the protein dynamics and sampling the distribution of dark-state conformers realistically.

The importance of the distribution of dark-state conformers for the functionality of the LOV1-domain still needs to be assessed through theoretical as well as experimental means. However, it is plausible that, through a shift in the population distribution of inter-converting states, the LOV1-domain is able to adapt to a change of the external conditions, like e.g. a temperature change, and/or the solvation environment. This might permit the LOV1-domain to act as a modulating sensor with adjustable dark-state activity.

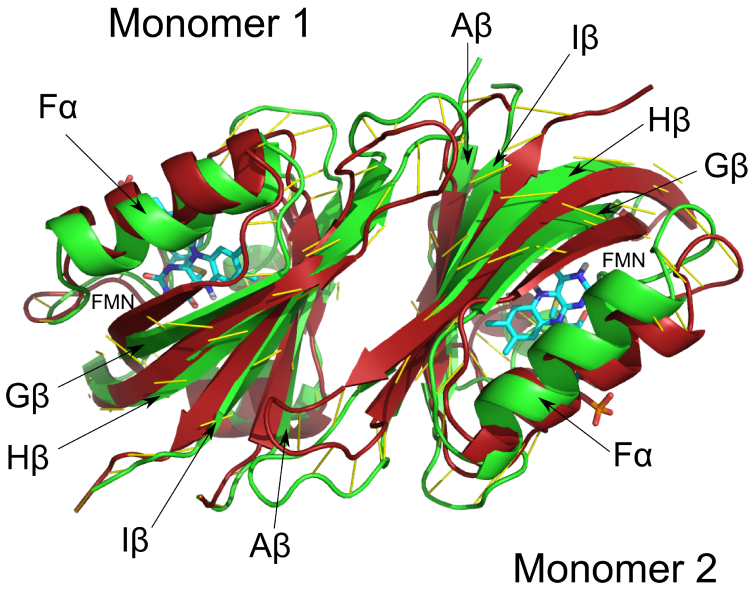


Figure 10.14.: Protein structure of the dark-state (green) and light-state (red) dimer at a simulation time of 20 ns, obtained with the NIT PME simulation technique.

### 10.1.2. Conclusions

In summary, in the present study we investigated the influence of the calculation procedure on the dark-state conformers of the wild-type LOV1-domain from the green algae *Chlamydomonas reinhardtii* in its monomeric and dimeric form. In particular, we considered the effect of different long-range electrostatic calculation techniques and thermostating strategies on their solution structures and conformational fluctuations. To this end, we first compared MD simulation results generated with the particle-mesh Ewald and the computationally less expensive twin-range-cutoff technique. The former technique has been suspected in previous works to suppress conformational fluctuations of peptides and small proteins because of artificial periodicity, while the second technique has been found to lead to an unrealistic destabilization of local configurations. In the particle-mesh Ewald case we find two local configurations of the amino acids surrounding the FMN chromophore, which can reversibly be inter-converted into each other. In the crystal-like dark-state conformer the carbonyl-oxygen atom O4 on FMN is twice H-bonded by the surrounding amino acid pair Asn99 and Gln120. In the second dark-state conformer the carbonyl-oxygen atom O4 on FMN becomes singly H-bonded with the amino acid Asn89, in agreement with experimental observations. Moreover, the thiol group of the reactive Cys57-residue, which in the crystal-like conformer is located close to the reactive center FMN-C4a, points in the second dark-state conformer away from this site in direction to the benzene ring of the FMN. Through electronic interaction with the C8-position on this ring, the thiol group can cause a stabilization of the quinoid resonance structure of the FMN and lead to an enhancement of the ISC rate, which promotes the population of the reactive FMN-triplet state from which the biological active adduct state of the protein is generated. Our calculations and the comparative experimental results indicate that the LOV1-domain from *Chlamydomonas reinhardtii* is able to alter the ISC rate and rate of adduct formation by adjusting the population distribution of dark-state conform-

ers. To understand the effect of the local arrangement of amino acids around the chromophore on the overall solution structure of the LOV1-domain, we also compared the protein secondary structures of the two dark conformers. In the second dark-state conformer we observed the disappearance of an $\alpha$-helical element, denoted as the D$\alpha$-helix, in conjunction with an augmentation of the flexibility of the $\beta$-strands, which might increase the capacity of the LOV1-domain to adjust on the hydrophobic surfaces of other LOV-domains or the kinase. Therefore, shifting the population distribution towards one or the other conformer via a change of external conditions, like e.g. a temperature change, might permit to the green algae *Chlamydomonas reinhardtii* to regulate the affinity for dimerization of the LOV1-domains in the dark and provide a mechanism for modulating the response under low-light conditions. In case of the twin-range-cutoff technique a weakening of the H-bond interactions between the chromophore and the apo-protein as well as a strengthening of low-mode fluctuations is observed, leading to a partial destabilization of the protein secondary structure. Overall, we can say that the twin-range-cutoff technique leads to larger conformational fluctuations resulting in a destabilization of the secondary structure of the protein domain, whereas the PME technique leads to conformational fluctuations in consistency with experimental observations.

Another major goal of this work was to study the effect of the thermostating strategy on the local arrangement of amino acids around the FMN and the conformational fluctuations of the dark-state conformers. To analyze this aspect, we investigated the usefulness of the NIT technique, where the protein solute is only indirectly coupled to the thermostat via the surrounding solvent. We compared this strategy to the standard IT thermostating strategy, where individual thermostats are coupled to both the solute and the surrounding solvent. From our investigation, we deduce that both thermostating strategies provide a different dynamical behavior of the amino acids near to the chromophore. The IT strategy favors equilibrium fluctuations around the starting crystal structure by enforcing the same temperature in each part of the system at any time of the simulation. This causes that configurations far from this local equilibrium configuration are suppressed and the system cannot drift towards a new local equilibrium. The NIT strategy, by contrast, avoids this problem by only thermostating the solvent, which enables the protein to sample configurations far from equilibrium. This permits the sudden local re-arrangement of the amino acids to take place, leading to the emergence of a new local equilibrium configuration. Since the resulting solution structure is in agreement with experimental observations, we conclude that the NIT thermostating strategy provides a better thermostating strategy for simulating the protein dynamics and sampling the distribution of dark-state conformers realistically.

Finally, our investigation indicates that, through a shift in the population distribution of dark-state conformers, the LOV1-domain is able to adapt its reactivity to a change of external conditions and act in this way as modulating sensor. However, the role of the different dark-state conformers for the functionality of the LOV1-domain still needs to be assessed in its natural environment. This requires the development of novel reliable cost-effective computational methods, to treat multiple length scales [128] and long-range interactions [213] effectively.

# 11. Molecular dynamics study on signaling state of Chlamydomonas reinhardtii and its mutants

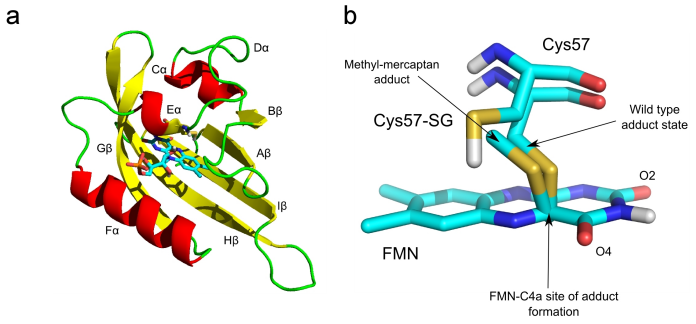


Figure 11.1.: Structural characteristics of the LOV1-domain. (a) Protein tertiary structure of the wild-type LOV1-domain, obtained by Fedorov *et al.* using x-ray diffraction measurements [11] (MD-starting structure). (b) FMN-chromophore with surrounding Cys57-residue for the CFN-adduct state as well as MM-adduct state (C-atom: cyan, S-atom: yellow, H-atom: white, O-atom: red, N-atom: dark blue).

## 11.1. Introduction

In the following sections we elucidate the early processes of the signal transduction pathway of the wildtype LOV1-domain from *Chlamydomonas reinhardtii* on the molecular level, using the MD-simulation method. A particular focus will be put on the understanding of the effect of the covalent-bond formation as well as the charge redistribution at the FMN-chromophore on the conformational changes of the CFN-adduct state. To this end, we compare the calculation results of the wildtype LOV1-domain in its dark and CFN-adduct state form to the MM-adduct state of the Cys57Gly-mutant (LOV1-Cys57Gly). The latter system is characterized by the same charge distribution on the FMN-chromophore as the CFN-adduct state, however it lacks the covalent bond between the FMN-chromophore and the apo-protein. We compare the calculation results with the available experimental data and demonstrate that the covalent-bond formation upon blue-light activation has a major influence on the response of the LOV1-domain. Moreover, our study will provide us new insights about the molecular mechanism and the functional role of the LOV1-domain.

### 11.1.1. Generation of starting structures

As starting structure for our dark-state simulation, we used the dark-state crystal structure of the wild-type LOV1-domain from *Chlamydomonas reinhardtii*, which was determined through x-ray diffraction measurements by Fedorov *et al.* [11] (PDB-code: 1N9L) and whose tertiary structure is visualized in Fig. 11.1a. To create the initial structure for our simulation of the cysteinyl-FMN CFN adduct of the wildtype LOV1-domain, we employed the dark-state crystal structure previously mentioned and generated a CFN-adduct by forming a covalent bond between the Cys57-S and FMN-C4a, as explicitly described in section 3.6. Similarly, we created the MM-adduct of the LOV1-Cys57Gly-mutant by using the same dark-state crystal structure and forming a covalent bond between the MM and FMN-C4a. This latter state solely differs from the CFN-adduct state through the lack of the covalent linkage between the reactive cysteine residue Cys57 and the FMN-chromophore (see Fig. 11.1b).

## 11.2. Results and discussion

### 11.2.1. Structural fluctuations during signaling process

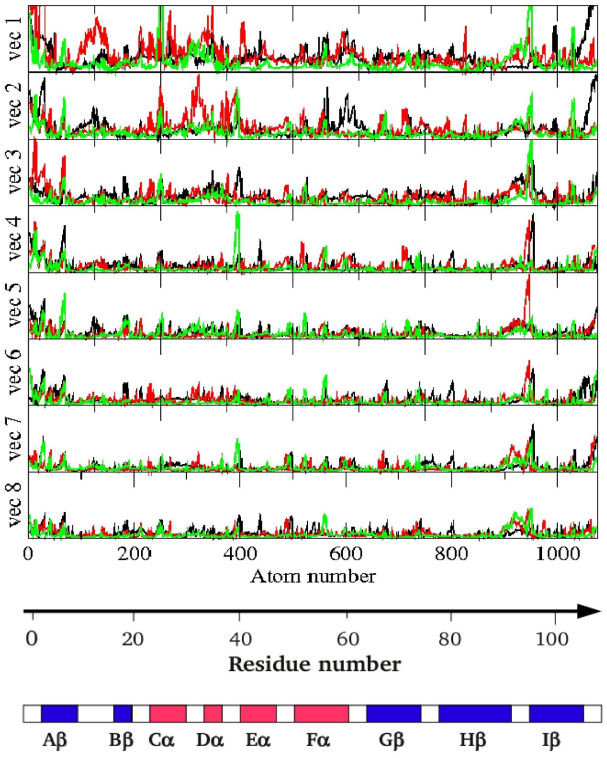


Figure 11.2.: Fluctuations of amino acids in the LOV1-domain. Root-mean square fluctuation of the dark-state (black) and CFN-adduct (red) state of the wildtype LOV1-domain, compared to the MM-adduct state of the LOV1-Cys57Gly-mutant (green), as a function of atomic number, residue number as well as secondary-structure element.

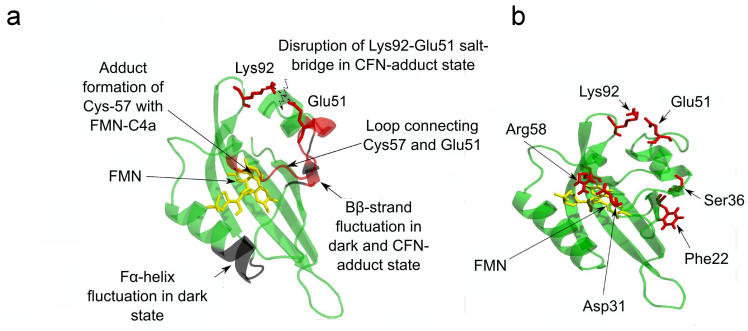


Figure 11.3.: Fluctuations of amino acids in the LOV1-domain. (a) Structural elements with largest change in mobility of the secondary structure of the dark-state (black) and CFN-adduct (red) state of the wildtype LOV1-domain. (b) Tertiary structure of the LOV1-domain with characteristic residues, identified in the RMSF-analysis as to be involved in the signal transduction pathway.

We start the analysis of our simulation results by identifying the region of the LOV1-domain of high and low mobility. To this end, we determine the root-mean-square fluctuation (RMSF), which is an indicator of the flexibility or rigidity of the protein structure [214]. The eigenvectors with the largest eigenvalues represent the collective motions that characterize the flexible degrees of freedom, whereas the smallest values represent the most constrained degrees of freedom. To clarify the role of the covalent-bond formation and change of the electrostatic distribution on the FMN-chromophore upon adduct formation, we compare in Fig. 11.2 the RMSF's for the dark-state and CFN-adduct state of the wildtype LOV1-domain to the one of the MM-adduct state of the LOV1-Cys57Gly-mutant. By analyzing the first eigenvector, we observe that in the range of residue numbers from 20 to 30 significant differences occur between the RMSF of the CFN-adduct state and the one of the MM-adduct state. A similar deviation from the RMSF of the MM-adduct state can be observed in case of the wildtype dark-state in the second eigenvector. This region corresponds to the linker region between the A$\beta$- and B$\beta$-strands. A second major deviation occurs in the second eigenvector in the range of residue numbers from 36 to 60, enclosing the reactive cysteine residue Cys57, between the RMSF of the CFN-adduct state and the MM-adduct state of the mutant. The increased mobility in this region in the CFN-adduct state relates to the propagation of the stress to its surrounding amino-acid environment, triggered by the covalent-bond formation between Cys57 and FMN-C4a. To provide an overview of the secondary structural elements primarily affected by the adduct formation process, we have visualized the regions with the largest RMSF in Fig. 11.3a. We recognize that the generation of the CFN-adduct induces an increased mobility in the $C\alpha$-, $D\alpha$- and $E\alpha$-helices on the hydrophilic side of the LOV1-domain, extending to a salt bridge between Glu51 and Lys92 at the surface of the protein. This observation is in agreement with the conclusions made by Iwata *et al.* [44] from the FTIR-spectroscopic measurements on the LOV1-domain of *Adiantum* Phytochrome3, who found that the changes in the secondary structure mainly take place in the region around Cys57, corresponding to the turn structure. An additional difference in the RMSF can be detected in the range of residue numbers from 70 to 82, located on the hydrophobic side of the LOV1-domain, between the dark-state and MM-adduct state of the mutant. We conclude from this observation that the change of the charge distribution upon adduct formation, but without the covalent-bond constrain, in the MM-case induces a reduction in mobility of the F$\alpha$-helix and the linker region

between the F$\alpha$-helix and the G$\beta$-strand.

### 11.2.2. Signal propagation near to reactive cysteine Cys57

To study the molecular details of the stress propagation following the covalent-bond formation and charge redistribution on the FMN-chromophore, we consider next in Fig. 11.4 the distances between characteristic atoms, identified in the previous RMSF-analysis as to be involved in the signal transduction pathway, as a function of simulation time. In the graphs we compare the results of the dark-state (black) and CFN-adduct state (red) to the MM-adduct state of the LOV1-Cys57Gly-mutant (green). The amino acids under consideration are represented in Fig. 11.3b. In Fig. 11.4a we show the distance between Asn56-N and FMN-O4 versus simulation time for all three species. We recognize that the curve of the CFN-adduct state shows a sudden drop of about 2 Å at a simulation time of 5000 ps, while the curves of the dark state and MM-adduct state remain stable over the whole simulation time. This shift in the distance relates to the stress release, originating from the covalent-bond formation between Cys57 and FMN-C4a. The stress then continues to evolve along the backbone chain, as can be deduced from the jumps in the distance between Glu51-N and FMN-O4 in Fig. 11.4b as well as Arg58-CZ and Asp31-OD2 in Fig. 11.4c of the CFN-adduct state curves. We emphasize that Arg58 is located in the center of a H-bond network and forms H-bonds with the FMN-phosphate as well as Asp31. It is apparent that both curves start to deviate at around 5000 ps, but return to their initial values at 15000 ps after full stress relaxation. In this time range the temporary increase in the distance between Arg58-CZ and Asp31-OD2 leads to a weakening of the H-bond between both residues. Since, by contrast, these distances in case of the MM-adduct state do not show a similar response, we conclude that the stress propagation along the backbone chain must be a consequence of the covalent-bond formation between Cys57 and FMN-C4a. We point out that the implication of the residues Arg58 and Asp31 in the signal transduction pathway of the LOV1-domain of *Chlamydomonas reinhardtii* was investigated by Losi *et al.* [208] through introducing point mutations at both sites. By then performing photoacoustic measurements with these mutants, they found that substitution of Arg58 significantly influence the H-bonds formed with Asp31 and the FMN-phosphate group, ultimately affecting the light-induced volume changes of the LOV1-domain. Besides Arg58, they recognized that substitution of Asp31 also exhibits well-detectable but smaller effects. Moreover, they concluded from their experiments that the covalent-bond formation between Cys57 and FMN-C4a causes a larger displacement of Asn56, which follows the Cys57-movement toward the FMN-chromophore, and brings it closer to Arg58. This goes along with the breakage of the H-bond between Asn99-ND2, which is attached to the H$\beta$-strand, and FMN-O4, as can be observed in Fig. 11.4d. The rearrangement of amino acids helps stabilizing the photoadduct and is reversed during the dark-state recovery reaction. This conclusion agrees well with the shifts and the subsequent recovery of the distances between Asn56-N and FMN-O4 as well as between Glu51-N and FMN-O4, observed in the Figs. 11.4a and 11.4b respectively. Because the H-bond between Arg58 and Asp31 connects the peripheral A$\beta$- and B$\beta$-strands to the coil region between the E$\alpha$ and D$\alpha$-helices, we conclude that its disruption in the adduct state can lead to a large change of the volume of the protein domain, which results from the covalent-bond formation between Cys57 and FMN-C4a. By contrast in case of the MM-adduct state of the LOV1-Cys57Gly-mutant no abrupt changes are observable for the amino acids previously mentioned over the entire simulation run. This indicates that, if the covalent bond between Cys57 and FMN-C4a is lacking, no stress can be built up and propagate along the backbone chain, which would permit to transmit the signal to the peripheral part of the protein. Moreover, it demonstrates that the light-induced spatio-dynamical changes are only weakly affected by a change of the charge distribution on the FMN-chromophore. Similar

observations can be made in case of the dark-state. The jump in the distance between Glu51-N and FMN-O4, observed in Fig. 11.4b, around 17000 ps is caused by a displacement of the chromophore in the binding pocket and therefore cannot be attributed to the displacement of the backbone chain, as in case of the adduct state. This is confirmed by a similar jump appearing at the same simulation time in the dark-state curve for the distance between Asn99-ND2 and FMN-O4, shown in Fig. 11.4d. In conclusion, we retain that the formation of a covalent bond between Cys57 and FMN-C4a induces a characteristic signal, transmitted via stress propagation through the outer peripheral of the protein domain, causing a change of its surface properties.

### 11.2.3. Salt bridge between Glu51 and Lys92

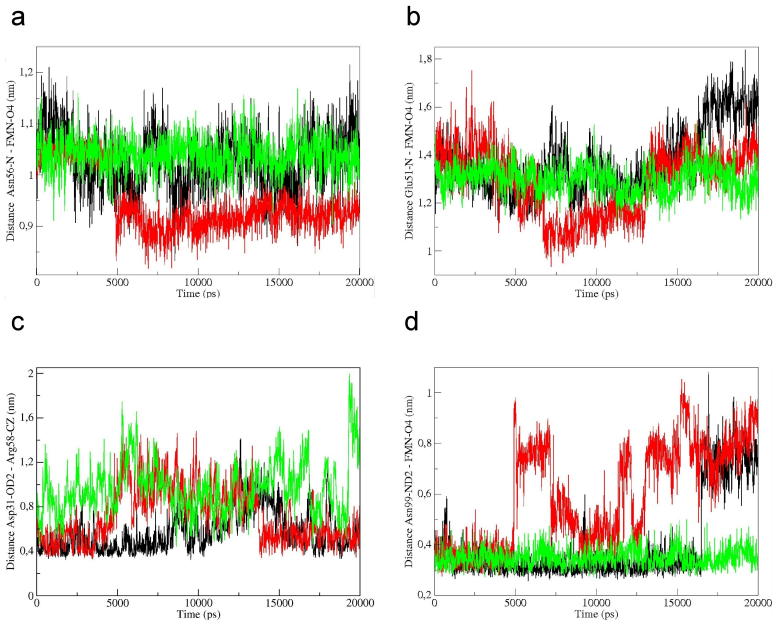


Figure 11.4.: Inter-atomic distances from characteristic residues surrounding the FMN-chromophore and/or located on the hydrophilic side of the LOV1-domain as a function of simulation time. (a) Asn56-N - FMN-O4. (b) Glu51-N - FMN-O4. (c) Asp31-OD2 - Arg58-CZ (d) Asn99-ND2 - FMN-O4. In the graphs we compare the curves of the dark-state (black) and CFN-adduct (red) state to the MM-adduct state of the LOV1-Cys57Gly-mutant (green). All these residues have been identified in RMSF-analysis as to be involved in the signal transduction pathway.

In previous works [49, 215] it was speculated that a salt bridge between Glu51 and Lys92 is involved in the signaling process of the LOV1-domain. To investigate this issue in more detail, we plot in Fig. 11.5 histograms of the distance between Glu51 and Lys92, performed over different time ranges. In Fig. 11.5a we first compare the histograms, taken over the entire simulation run, of the dark-state (black) and CFN-adduct state (red) to the MM-adduct state of the LOV1-Cys57Gly-mutant. We see that the dark-state curve possesses one major peak with a maximum centered around a distance of 0.23 nm, as well as a broad tail at larger distances with much smaller magnitude. By contrast, the corresponding histograms of the CFN-adduct and MM-

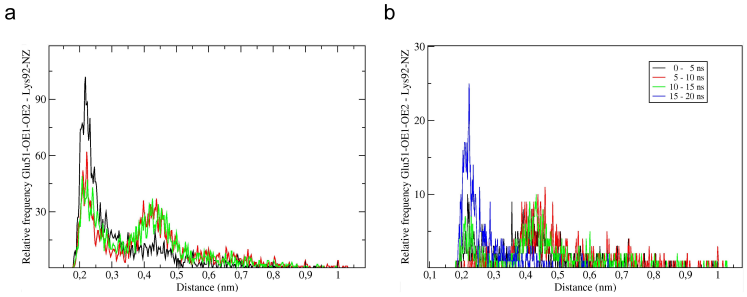


Figure 11.5.: Histogram of the distance between Glu51 and Lys91, characterizing a salt bridge on the hydrophilic side of the LOV1-domain, in different time ranges. (a) 20 ns-trajectory for the dark-state (black), CFN-adduct state (red) and MM-adduct state of the LOV1-Cys57Gly-mutant. (b) CFN-adduct state in the time range between 0-5 ns (black), 5-10 ns (red), 10-15 ns (green) and 15-20 ns (blue).

adduct states show two main peaks of almost similar height centered around 0.23 nm and 0.43 nm, which indicates that the salt bridge is partly dissociated in these cases. The agreement of the histograms of both species is rather surprising, since in case of the mutant no stress is propagated towards the surface, which would permit to dissociate the salt bridge. To analyze this issue in more detail, we consider in Fig. 11.4c the distance characterizing the H-bond between Arg58-CZ and Asp31-OD2 of the CFN-adduct and MM-adduct states. By comparing both curves, we recognize that in case of the mutant the distance fluctuates with large magnitude, which relates to the fact that Arg58 is rather flexible in the binding pocket and leads to a permanent rupture of the H-bond between Arg58-CZ and Asp31-OD2. In case of the dark-state curve we observe a moderate jump in the distance starting at around 8000 ps, which is much smaller than the jump of the adduct state curve, taking place at 5000 ps. The increase in the former case correlates with a temporary shortening of the distance between Cys57 and FMN-C4a. This confirms the conclusions of Gardner *et al.* [216, 217], made from solution NMR-measurements on the LOV2-J$\alpha$-system, that the conformations dominant in the lit state also exist in the dark-state, but with lower statistical weight. In our case the pseudo-lit state conformations can be considered as high-energy dark-state conformations, where the Cys57 is temporary attracted by the FMN-C4a through H-bond formation between the H-atom of the thiol group and FMN-N5. Let us next study the time-behavior of the Glu51-Lys92-salt bridge during the signaling process of the LOV1-domain. To this end, we analyze in Fig. 11.5b the histogram of the distance between Glu51 and Lys92 of the CFN-adduct state in the time ranges 0-5 ns (black), 5-10 ns (red), 10-15 ns (green) as well as 15-20 ns (blue). Confronting these curves to the histograms of the dark and adduct states in Fig. 11.5a carried out over the entire simulation run, we conclude that in the initial phase of the stress propagation, i.e. 0-15 ns, two peaks are generated with an emphasis on the larger distance. This shows that in this time range the salt-bridge is essentially broken. Moreover, we observe that in the subsequent time range, i.e. 15-20 ns, the histogram essentially adopts the pattern of the dark-state curve in Fig. 11.5a with a main maximum at around 0.23 nm. This demonstrates that after 15 ns the stress has fully relaxed and the salt-bridge readopts its dark-state conformation. By further comparing these results to the dark and adduct state curves in Fig. 11.4, we conclude that the stress, originating from the CFN-adduct formation in the binding pocket, propagates along the residues on the backbone chain upward and downward in sequence relative to Cys57. This causes the dissociation of the Glu51-Lys92-salt bridge at the surface of the LOV1-domain, confirming its implication in the signal transduction pathway, as

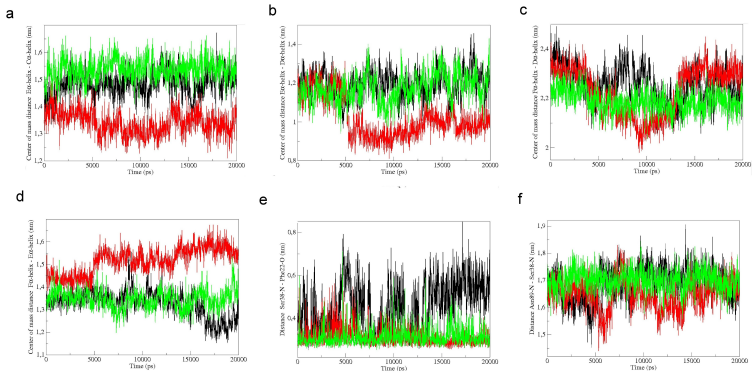


Figure 11.6.: Center of mass distance between $\alpha$-helices on the hydrophilic side and inter-atomic distances from characteristic residues on $\beta$-strands of the dark-state (black) and CFN-adduct (red) state of the wildtype LOV1-domain, compared to the MM-adduct state of the LOV1-Cys57Gly-mutant (green). (a) E$\alpha$-helix - C$\alpha$-helix. (b) E$\alpha$-helix - D$\alpha$-helix. (c) F$\alpha$-helix - D$\alpha$-helix. (d) F$\alpha$-helix - E$\alpha$-helix. (e) Ser38-N - Phe22-O. (f) Asn89-N - Ser38-N.

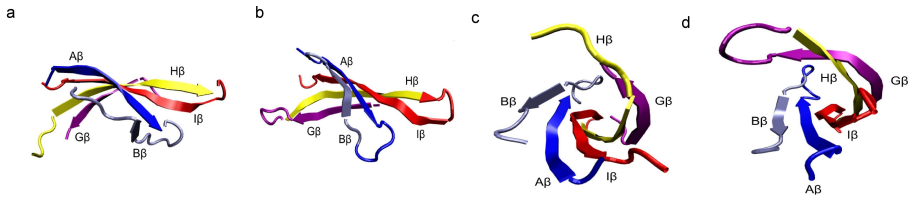


Figure 11.7.: Arrangement of $\beta$-strands of the LOV1-domain at the final simulation time of 20 ns. (a) dark-state - side view. (b) CFN-adduct state - side view. (c) dark-state - front view. (d) CFN-adduct state - front view.

suggested in various experimental and theoretical studies. Crosson *et al.* [215] speculated that the conservation of chemically-linked residues extending from the FMN to this surface salt bridge is an indication for the involvement of this protein part in the signaling process. Furthermore, they pointed out that, even if the salt bridge persists only 50 % of the time, it may still be useful to modulate the structure and dynamics of the B$\alpha$-A$\alpha$'-helices as well as the C$\beta$-D$\beta$-loop in the LOV-domain. Slight shifts in the stability of the bridge upon adduct formation could serve to increase or decrease the structural mobility of the B$\alpha$-A$\alpha$'-helices as well as the C$\beta$-D$\beta$-loop and, thus, affect the interaction with the LOV-partner domains. To substantiate their hypothesis, they proposed two distinct models for describing light-driven changes in the binding affinity between the LOV-domain and its partner domains, involving the conserved surface salt bridge previously mentioned [215]. In an enthalpically-driven interaction model, the process of adduct formation destabilizes the salt bridge and leads to the adoption of a distinct conformation, associated with a different affinity for the interacting LOV-partners. By contrast, in an entropically-driven interaction model, the same process also causes the destabilization of the salt bridge, but it does not change the average conformation of the LOV-domain. Instead of that, it increases its conformational flexibility, resulting in a different affinity for its interacting partners. Although

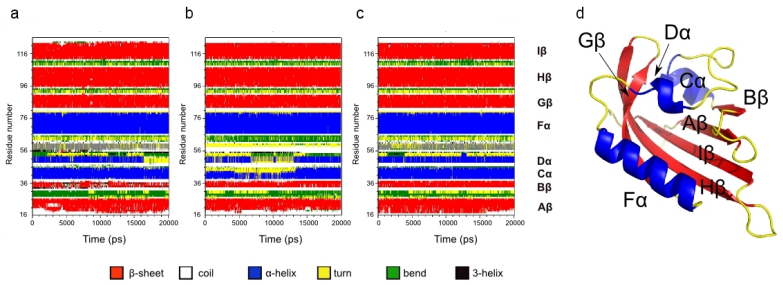


Figure 11.8.: Secondary structural analysis from MD-simulations. (a) dark-state. (b) CFN-adduct state of the wildtype LOV1-domain. (c) MM-adduct state of the LOV1-Cys57Gly-mutant (d) Secondary structure elements of the LOV1-domain.

the signaling process of the LOV-domains is likely to combine features of the two models, they favored a dynamic regulation model, where the dynamical state of the LOV-domain determines the interaction with its partner domains. In a subsequent work Freddolino *et al.* [49] investigated the structural changes, taking place after formation of the CFN-adduct, by performing MD-simulations with the LOV1-domain of *Chlamydomonas reinhardtii*. Based on their simulation results, they proposed that the LOV1-activation is induced by a change in the H-bond network between the FMN-chromophore and the surrounding apo-protein that causes the disruption of the highly conserved salt bridge between Glu51 and Lys91 at the surface of the LOV1-domain. By contrast our simulation results demonstrate that, although this H-bond network plays an important role, the major influence on the behavior of the Glu51-Lys91-salt bridge results from the covalent-bond formation between Cys57 and FMN-C4a in the CFN-adduct state. We find that the conformational stress, generated at the CFN-bond, mainly reaches the Glu51-Lys91-salt bridge through the direct route along the backbone chain down to the residue Glu51, as well as to a smaller extent through the route via the H-bond between Arg58 and Asp31.

### 11.2.4. Signal propagation on peripheral $\alpha$-helices

To assess how the adduct formation affects the secondary structural elements on the hydrophilic side of the LOV1-domain, we consider next in Fig. 11.6 the center of mass distances between different $\alpha$-helical elements. First of all, we display in Fig. 11.6a the distance between the E$\alpha$- and C$\alpha$-helices of the CFN-adduct state compared to the ones of the dark-state and MM-adduct state. We observe that the curve of the CFN-adduct state, after an initial slow decrease, stabilizes after 5000 ps around an average value of 1.35 nm. By contrast, in case of the dark-state and MM-adduct state the quantity remains stable over the whole simulation run and deviates in average from the CFN-adduct state curve by about 0.2 nm. A similar behavior is observed in Fig. 11.6b, in which we visualize the center of mass distance between the E$\alpha$- and D$\alpha$-helix. We point out that the D$\alpha$-helix, characterized by the range of residue numbers from 49 to 52, contains Glu51, which is the amino acid forming a salt bridge with Lys92. While the curves of the dark-state and MM-adduct state show a stable behavior over the whole simulation run, we see that the curve of the CFN-adduct state undergoes an abrupt drop in the distance at around 5000 ps, which is followed by a slow increase of the quantity indicating a systematic recovery up to the end of the simulation run. Similar perturbations at the same simulation time can also be detected in the distance curves between the F$\alpha$- and D$\alpha$-helix, shown in Fig. 11.6c, as well as between the E$\alpha$- and F$\alpha$-helix, shown in Fig. 11.6d. Comparing these observations with the ones from the

histograms over 5 ns-time intervals of the distance between Glu51 and Lys92 in Fig. 11.5b, we conclude that the rearrangement at 5000 ps and subsequent conformational relaxation of the $\alpha$-helical elements correlates with the breakage and subsequent recovery of the Glu51-Lys92-salt bridge, confirming that this salt bridge is indeed involved in the signal transduction process as suggested in the experimental and theoretical studies discussed previously. We conclude from these observations that the covalent-bond formation in the CFN-adduct state significantly affects the hydrophilic region of the LOV1-domain, causing the breakage of the Glu51-Lys92-salt bridge and leading to an increased mobility of the $\alpha$-helical elements. The larger fluctuations in the hydrophilic region implicate also a significant change in the surface charge distribution, causing an attenuation of its aptitude to function as an aggregation partner to other protein domains, such as the LOV2-domain or the kinase. The possibility of the LOV1-domain to act as an attenuator of the kinase activity under low light-conditions and reduce its inhibitory effect under high light-conditions has been confirmed by Matsuoka and Tokutomi [218] by performing *in vitro* phosphorylation experiments with various constructs of bacterially expressed phototropin 2 from *Arabidopsis*. In *Chlamydomonas reinhardtii* the hydrophilic side of the LOV1-domain might play a similar role in its sexual life cycle [219] by adapting its reproduction speed and growth to the lighting conditions. Finally, it is also worth pointing out that the importance of the peripheral $\alpha$-helices in the light-induced regulation of the aggregation tendency with partner domains have also recently been demonstrated in case of the BLUF-domain by Nemukhin *et al.* using molecular dynamics simulations [220].

### 11.2.5. Signal propagation on $\beta$-strands

However, the LOV1-domain is not only involved with regulating the signal intensity of phototropin by interacting with the LOV2-domain or the kinase through its hydrophilic side. It has recently also been found to act as a docking partner through face-to-face association with $\beta$-scaffolds of LOV1-domains from other phototropins [221]. *In vitro* experiments of Briggs *et al.* [222] with phototropins from *Arabidopsis* have indicated that cross-phosphorylation can occur between functional phot2 and inactivate phot1 by means of hetero-dimerization through hydrophobic interaction via the $\beta$-sheets. Based on these findings, a response mechanism has been proposed, to account for the phototropic adaptation of the plant to moderate light intensities. Recent *in vitro* experiments of Kutta *et al.* [3] on the LOV1-domain from *Chlamydomonas reinhardtii* established that adduct formation results in a decrease in the monomer/dimer-ratio and concluded that dimerization of phototropin might play a role in the signal transduction pathway. However, to date it is still unknown, if dimerization might also have a modulating effect on the phototropin function *in vivo* [222]. To assess how the adduct formation affects the dimerization tendency of the LOV1-domain, we put next a focus on the fluctuations of the $\beta$-strands on its hydrophobic side. From the RMSF-analysis in Fig. 11.2, we deduce that large deviations are observable in the first eigenvector of the CFN-adduct state and the second eigenvector of the dark-state, compared to the MM-adduct state, in the range of residues located between the A$\beta$- and B$\beta$-strands. To analyze this issue in more detail, we consider next distances of characteristic residues involved within the hydrophobic region of the protein. For this purpose, let us first regard the distance between Ser38-N and Phe22-O of the dark-state, CFN-adduct state as well as MM-adduct state in Fig. 11.6e. In the crystal structure both residues form a H-bond, which connects the A$\beta$-strand and B$\beta$-C$\alpha$-loop. From the figure, we deduce that the distance between both residues in the dark-state is subjected to large fluctuations, whereas the H-bond remains stable in the MM-adduct state. By contrast, in case of the MM-adduct state curve fluctuations with increased magnitude are only observed within the first 7500 ps after adduct formation, until stress has fully relaxed and the H-bond is stabilized. This indicates that the fluctuations

of the A$\beta$- and B$\beta$-strands are essentially influenced by the charge redistribution on the FMN-chromophore, but not by the formation of the covalent bond of the CFN. Next, we plot in Fig. 11.6f the distance between the backbone atoms Asn89-N and Ser38-N, which are part of the G$\beta$- and B$\beta$-strands respectively. The comparison of the results of all three species demonstrate that neither the covalent-bond formation nor the charge redistribution of the CFN-adduct do significantly affect the interaction between the G$\beta$- and B$\beta$-strands. In conclusion, we retain that the charge redistribution on the FMN-chromophore is primarily responsible for the damping of the fluctuations of the $\beta$-scaffold through tightening of the A$\beta$- and B$\beta$-strands. In Fig. 11.7 we show the arrangement of the $\beta$-strands after 20 ns in case of the dark-state (a-c) as well as CFN-adduct state (b-d). The largest differences between both states are observable in the region around the B$\beta$-strand. In the CFN-adduct state the B$\beta$-strand is significantly closer to the A$\beta$-strand compared to the dark-state, which is consistent with the light-induced $\beta$-sheet tightening process discussed previously. The latter process goes along with an increase of the number of van-der-Waals (vdW) contacts, which can be formed in a dimer between the side chains of residues located on the $\beta$-sheet surfaces of the LOV1-domains, allowing a stronger hydrophobic interaction and increased dimerization tendency. These findings concord well with the experimental results of Kutta *et al.* [3], who found that adduct formation results in an increase of the monomer interaction within dimers, indicating that dimerization might play a role in the signal transduction pathway. Similar conclusions have also been drawn by Nakasoka *et al.* [221] through performing small-angle x-ray scattering measurements on the LOV1-domain of phot1 and phot2 from *Arabidopsis*. They established that dimeric association is essentially dominated by vdW-contacts, in addition to some H-bonds, formed between side chains of residues on the $\beta$-scaffolds, stabilizing the dimer structures. Interestingly, similar findings have been made by Iwata *et al.* [45] on the LOV2-domain of *Adiantum* Phytochrome3 using FTIR- and UV-visible-spectroscopy. They concluded from their experiments that immediately upon formation of the CFN-adduct the system looses the local structure of amino acids surrounding the FMN-chromophore. In subsequent stages further changes take place to mainly tighten the $\beta$-sheets as well as to alter the arrangement of $\alpha$-helices on the hydrophilic side of the protein.

### 11.2.6. Structural evolution over long times

To assess how the covalent-bond formation influences the evolution and stability of the secondary structure over time, we consider next the secondary structure analysis of all three species in Fig. 11.8, i.e. dark-state (a), CFN-adduct state (b), MM-adduct state of the LOV1-Cys57Gly-mutant (c). Whereas the secondary structures of the MM-adduct state and dark-state remain nearly unaffected by fluctuations, the CFN-adduct state shows large fluctuations in the flexible $\alpha$-helical elements in the initial phase up to 15000 ps after adduct formation. In this time range the CFN-adduct state undergoes significant changes in the hydrophilic region of the protein, extending from the D$\alpha$-helix to the C$\alpha$-helix. We point out in this context that in case of the dark-state the D$\alpha$-helix also disappears in the time range from 17000 ps to 20000 ps, which relates to the formation of a second dark-state conformer characterized by a different orientation of the Cys57-residue with respect to the FMN-chromophore. Finally, it is also apparent that the CFN-adduct state shows stronger fluctuations in the A$\beta$- and B$\beta$-strands than the dark-state as well as MM-adduct state. These results demonstrate that the formation of the covalent bond upon photoexcitation mainly induces fluctuations in the residue range, including the D$\alpha$-helix and C$\alpha$-helix as well as the A$\beta$-strand. Moreover, we see that its neglect in case of the MM-adduct state leads to a structural stabilization, whereas the charge redistribution on the FMN-chromophore upon adduct formation does not have a significant effect on the structural changes. This indicates that the dynamics of signal transduction of the CFN-adduct state is

primarily affected by the covalent-bond formation and only to a minor extend by the charge redistribution on the FMN-chromophore.

### 11.2.7. Conclusions

In this study we have elucidated the early stages of the signal transduction pathway of the LOV1-domain of *Chlamydomonas reinhardtii* upon adduct formation on the molecular level using the molecular dynamics simulation method. To this end, we have compared the calculation results of the dark-state and cysteinyl-FMN state of the wildtype LOV1-domain to a methyl-mercaptan-adduct state of its Cys57Gly-mutant. This latter system has been generated in recent experiments [224, 2] by replacing the reactive cysteine residue Cys57 with glycine as well as photochemically producing an adduct through addition of methyl-mercaptan and irradiation with blue light. Our simulation results demonstrate that the covalent-bond formation between Cys57 and FMN-C4a in the adduct state primarily affects the hydrophilic side of the LOV1-domain. More specifically, we find that the conformational stress, generated at Cys57, mainly propagates to the Glu51-Lys91-salt bridge through the direct route along the backbone chain down to the residue Glu51, as well as to a smaller extent through the route via the H-bond between Arg58 and Asp31. This results in an increased mobility of the $C\alpha$-, $D\alpha$- and $E\alpha$-helices on the hydrophilic side of the LOV1-domain. The larger fluctuations in the hydrophilic region implicate also a significant change in the surface charge distribution, causing a decrease of its aptitude to function as an aggregation partner to other protein domains, such as the LOV2-domain or the kinase. By contrast, the hydrophobic side of the protein is essentially influenced by both the covalent-bond formation between Cys57 and FMN-C4a as well as a change of the charge distribution on the FMN-chromophore. This induces a $\beta$-sheet tightening process via coupling of the A$\beta$- and B$\beta$-strands. The latter process goes along with an increase of the number of van-der-Waals contacts, which can be formed within a dimer between side chains of residues located on the $\beta$-sheet surfaces of the LOV1-domains, allowing a stronger hydrophobic interaction and increase of the dimerization tendency. Therefore, our findings confirm the aptitude of the LOV1-domain to function as a dimerization site, to allow cross-phosphorylation of the phototropins, and permits to the green alga to adapt its reproduction and growth speed to the lighting conditions. We anticipate that the elucidation of the LOV1-signal transduction pathway on a molecular level will provide new perspectives for the development of novel photoactivable fusion proteins with adjustable binding affinity to substrates or membranes [71]. The fine-tuning could be done through the creation of mutants of the LOV1-domain with optimized light-induced structural response, which might open new avenues for allosteric control of protein activity in living cells [223, 225, 226].

# 12. MD study on the AsLOV2-J$\alpha$-photosensor

### 12.0.8. Introduction

In the following sections we elucidate the early stages of the signal transduction pathway of the AsLOV2-J$\alpha$-photosensor on the molecular level, using the molecular dynamics (MD) simulation method. Our study shows that the initial process after formation of the Cys966-FMN-adduct involve the breakage of a H-bond between the carbonyl-oxygen FMN-C4=O and the amino group of Gln1029, followed by a rotational re-orientation of its sidechain. This triggering event near the FMN-chromophore initiates the creation of a new H-bond between Gln1029 and Asn1008, coupling the I$\beta$- and H$\beta$-strands together, and causes a progressive extension of the I$\beta$-strand, transmitting the stress towards the LOV2-J$\alpha$-interface. The latter process then induces the breakage of H-bonds at opposite sides of the interface, involving a H-bond between Glu1034 on the J$\alpha$-I$\beta$-linker and Gln995 on the G$\beta$-strand as well as another between Asp1056 on the C-terminal end of the J$\alpha$-helix and Gln1013 on the H$\beta$-strand. Ultimately, this leads to a disruption and partial unfolding of the J$\alpha$-helix from the LOV-core. The details of the simulation as well as the generation of the starting structures of our dark and adduct state simulations (see Fig. 12.1c) are given in the method section appended below.

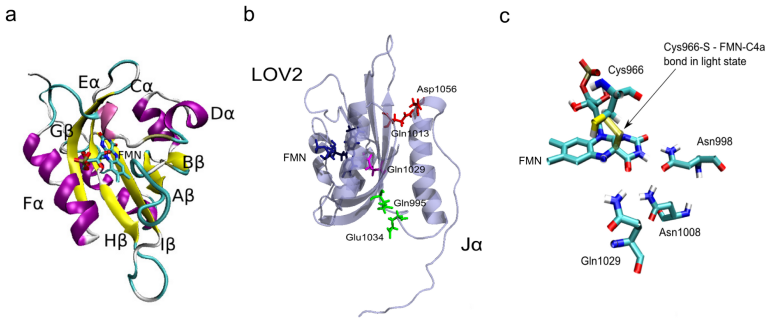


Figure 12.1.: Starting structures of the AsLOV2-J$\alpha$-photosensor. (a) dark-state solution structure of AsLOV2-J$\alpha$-system refined through NMR-measurements. (b) AsLOV2-protein core with adjacent J$\alpha$-helix and relevant amino acids [violet: Gln1029; red: Asp1056, Gln1013; green: Glu1034, Gln995]. (c) Difference in the starting structures between dark and adduct state form, structurally only differing with regard to the covalent bond of the CFN, and relevant amino acids surrounding the FMN-chromophore [cyan: C-atom; red: O-atom; blue: N-atom; white: H-atom; brown: P-atom; yellow: S-atom].

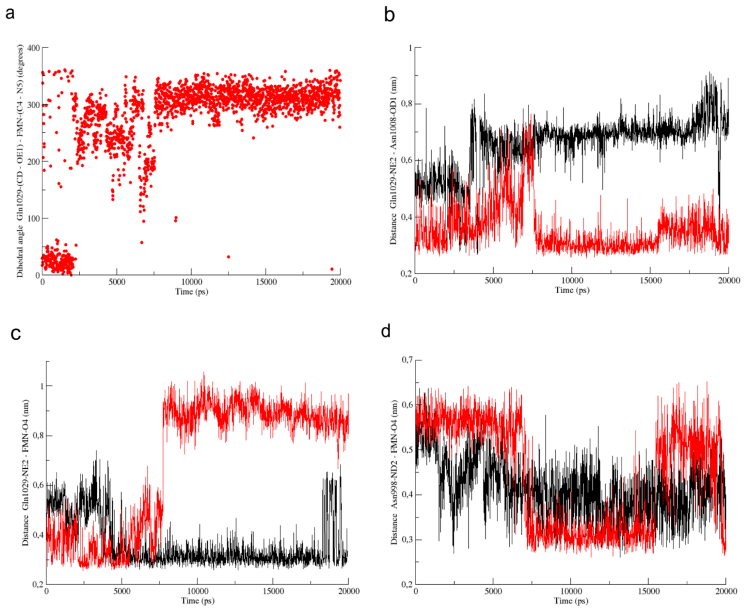


Figure 12.2.: Spatio-dynamical behavior of amino acids near the FMN-chromophore. Dihedral angle or distance of characteristic atoms on the FMN-chromophore or on surrounding amino acids to atoms of neighboring amino acid residues in the dark (black) and adduct (red) state [(a) dihedral angle Gln1029-CD - Gln1029-OE1 - FMN-C4 - FMN-N5. (b) distance Gln1029-NE2 - Asn1008-OD1. (c) distance Gln1029-NE2 - FMN-O4. (d) distance Asn998-ND2 - FMN-O4].

### 12.0.9. Generation of starting structures

As a starting structure for our dark-state simulation, we used the dark-state solution structure of the LOV2-Jα-protein domain from phototropin1 of *Avena sativa* AsLOV2-Jα of Harper *et al.*, which is a modeled structure refined with NOE-values originating from NMR-measurements [53]. To create the initial structures of our adduct state simulations for both the wildtype AsLOV2-Jα-system and its Gln1029Asn-mutant, we employed the dark-state solution structure previously mentioned and generated a cysteinyl-FMN-adduct (CFN) by forming a covalent bond between the Cys966-S and FMN-C4a, as described in section 3.6. In Fig. 12.1a we visualize the secondary structure of the LOV2-Jα-domain and FMN in its binding pocket, while in Fig. 12.1b we display the Jα-helix aggregated to the LOV2-core. The latter figure contains the amino acids relevant for signal transduction, discussed in the further development. In Fig. 12.1c we visualize the difference in the starting structures between the dark and adduct state form, which structurally only differ with regard to the covalent bond of the CFN.

### 12.0.10. Results

### 12.0.11. Wildtype AsLOV2-Jα-photosensor

We start the analysis of our simulation results by considering the spatio-dynamical behavior of relevant amino acids nearby the FMN-chromophore in the dark and adduct state. In Fig. 12.2a

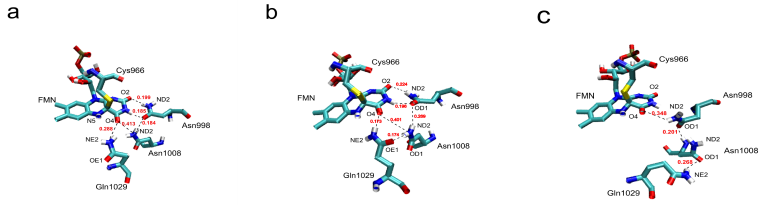


Figure 12.3.: Triggering event of the signaling pathway of the AsLOV2-J$\alpha$-photosensor. Representative local configurations of amino acids Gln1029, Asn1008 and Asn998 surrounding the FMN-chromophore at different simulation times, including H-bond distances in units of Å [(a) 0 ps. (b) 2500 ps. (c) 8000 ps].

we draw the dihedral angle of the adduct state, enclosing the atoms Gln1029-CD, Gln1029-OE1, FMN-C4 and FMN-N5, as a function of simulation time. We observe that in the initial stage the quantity fluctuates around an average value of 30 degrees, followed after a simulation time of 2500 ps by a sudden rotation of the angle of about 270 degrees to a new average value of 300 degrees. This rotational re-orientation of the glutamine residue is visualized in Fig. 12.3, where we compare the representative local configurations of the amino acids Gln1029, Asn1008 and Asn998, surrounding the FMN-chromophore, at different characteristic simulation times. In case of the configuration at 0 ps (Fig. 12.3a) the FMN-C4=O-carbonyl oxygen is twice H-bonded by the sidechains of Gln1029 and Asn1008, whereas the FMN-C2=O-carbonyl oxygen is twice H-bonded by the Asn998- and Gln970-sidechains. Comparing this configuration with the configuration at 2500 ps (Fig. 12.3b), we see that the rotation of the dihedral angle of about 270 degrees corresponds to a switch of the carbonyl-oxygen of Gln1029-OE1 about the axis formed by the sidechain atoms Gln1029-NE2 and Gln1029-CD of the glutamine residue. This process initiates the release of the amino acid Asn1008 through breakage of the H-bond between Gln1029 and Asn1008, as can be deduced from the increase in the distance between Gln1029-NE2 and Asn1008-OD1 in Fig. 12.2b, triggered at the same simulation time. From this graph, we can further conclude that this distance undergoes a sudden drop to a value of about 3 Å at around 7500 ps, which relates to the formation of a new H-bond between the amino group of Gln1029-NE2 and Asn1008-OD1 belonging to the I$\beta$- and H$\beta$-strand respectively. The corresponding configuration is visualized in Fig. 12.3c. We observe that the formation of a new H-bond causes a stable coupling between both $\beta$-strands, persisting over the remaining simulation time. We point out that this coupling process is consistent with the early stage of a $\beta$-sheet tightening process, observed in FTIR-experiments on LOV2- [45, 165] and LOV2-J$\alpha$-systems [46]. From the distances between Gln1029-NE2 and FMN-O4 in Fig. 12.2c as well as between Asn998-ND2 and FMN-O4 in Fig. 12.2d, we see that the H-bond formation process goes along with an abrupt detachment of the residue Gln1029 from the FMN-chromophore and a simultaneous stabilization of the local conformation through the creation of a new H-bond between Asn998-ND2 and FMN-O4, leaving the FMN-C4=O singly H-bonded with Asn998. This latter H-bond, however, is only transient and Asn998 can fluctuate back to its original position, as can be deduced from the subsequent jumps in Fig. 12.2d at 17000 ps and 19500 ps. In this context it is worth mentioning that the glutamine residue adjacent to the FMN-chromophore is highly conserved and has previously been suggested to be crucial for the signaling process of the LOV-domains. Several early experimental studies on the isolated LOV2-domain suggested that the primary process after adduct formation involves the breakage of a H-bond between the glutamine residue

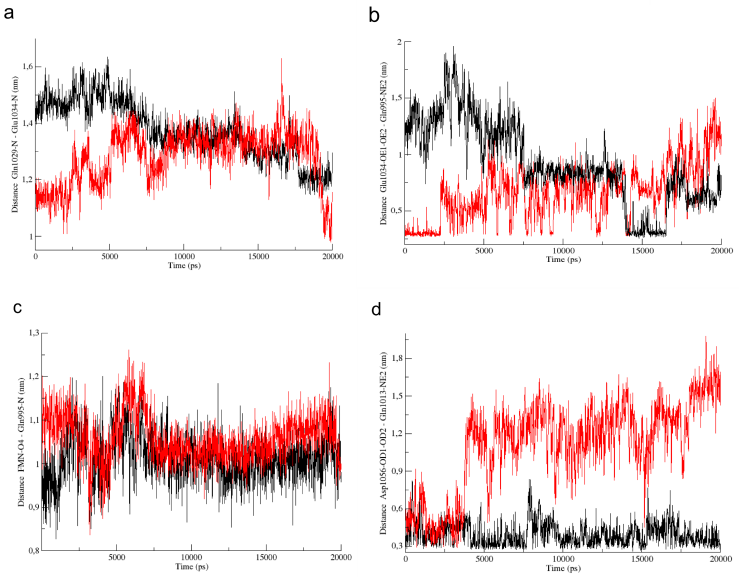


Figure 12.4.: Spatio-dynamical behavior of amino acids at the LOV2-Jα-interface. Inter-atomic distance from characteristic amino acids located on the interface between the Jα-helix and the LOV2-core in the dark (black) and adduct (red) state [(a) Gln1029-N - Glu1034-N. (b) Glu1034-OE1(OE2) - Gln995-NE2. (c) FMN-O4 - Gln995-N. (d) Asp1056-OD1(OD2) - Gln1013-NE2].

and the FMN-O4 [52], possibly followed via rotation of its sidechain by the formation of a new H-bond with the FMN-N5 [10, 11]. In a later crystallographic study on the AsLOV2-Jα-system Halavaty and Moffat [55] proposed a local re-orientation of the conserved glutamine, associated with a disruption of a H-bond between an asparagine and aspartic acid on the surface of the protein. In a series of spectroscopical investigations with point mutations of the LOV2-Jα-system [56, 227] as well as the full-length phototropin [35] it was demonstrated that this glutamine residue plays a central role in both spectral tuning and propagating the signal via the Iβ-strand from the LOV-core to the Jα-helix, while its lack attenuates light-induced autophosphorylation of the phototropin. In a recent theoretical work on the isolated AsLOV2-domain Freddolino *et al.* [49] deduced from MD simulations that adduct formation induces the breakage of a H-bond between the conserved glutamine and FMN-O4, followed by the creation of a new H-bond with FMN-N5. In addition, their simulation results suggested a second conformation, in which the glutamine interacted with neighboring residues on the Iβ-strand leading to an increased mobility in the LOV2-Jα-interface region. However, a clear molecular picture of the signaling process could not be established in this work, due to the neglect of the Jα-helix. From our simulation results, we conclude that the primary steps after adduct formation involve the breakage of a H-bond between Gln1029-NE2 and FMN-O4, followed by a rotational re-orientation of the glutamine residue Gln1029 and subsequent creation of a new H-bond between the amino group of Gln1029-NE2 and Asn1008-OD1. This latter process causes the coupling and tightening of the Iβ- and Hβ-strands. Moreover, it permits to propagate the stress to the LOV2-Jα-interface, which ultimately leads to the detachment of the Jα-helix from the LOV2-domain, as we will show in the following. In Fig. 12.4 we present data from our dark and adduct state simulations of characteristic inter-

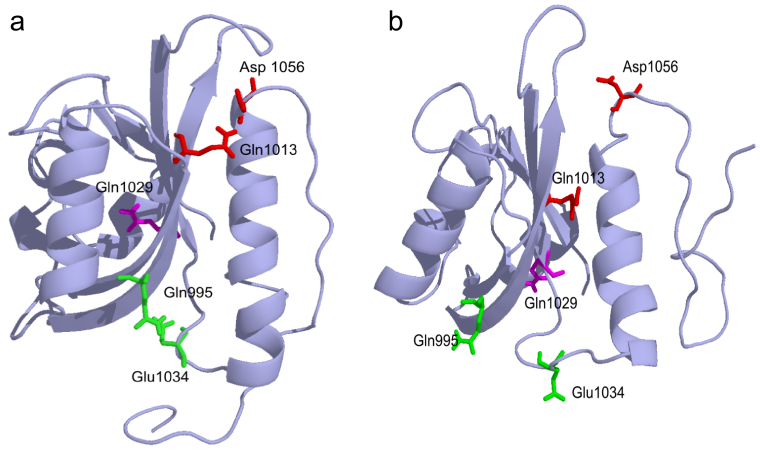


Figure 12.5.: Final secondary structures of the dark and adduct state of the AsLOV2-J$\alpha$-photosensor. Secondary structures at a simulation time of 20 ns from the dark (a) and adduct (b) state of the AsLOV2-J$\alpha$-system with relevant amino acids [violet: Gln1029; red: Asp1056, Gln1013; green: Glu1034, Gln995].

atomic distances of residues, located at the LOV2-J$\alpha$-interface, as a function of time. The respective residues are depicted in Fig. 12.1b. In Fig. 12.4a we first display the minimum distance between the backbone nitrogen atoms of the conserved glutamine residue Gln1029-N, which initially links the I$\beta$-strand via its sidechain to the FMN-O4, and glutamate Glu1034-N, located on the I$\beta$-strand in the linker region between the LOV2-core and the J$\alpha$-helix. In case of the adduct state we observe starting from 2500 ps a steady increase of the distance between both amino acids up to a maximum amount of 2 Å. This relates to the propagation of the stress along the I$\beta$-strand after the rotational re-orientation of the Gln1029-sidechain, leading to a progressive extension of the backbone chain. From Fig. 12.4b, we deduce that the latter process leads to the successive disruption of the H-bond between Glu1034-(OE1)OE2, located on the linker between the LOV2-core and the J$\alpha$-helix, and the amino group of Gln995-NE2 on the G$\beta$-strand. The release of the residue Glu1034 causes a perturbation of the G$\beta$-backbone in the time range from 2500 ps to 7500 ps, as can be deduced from the oscillating distance between the backbone nitrogen atom Gln995-N and FMN-O4 in Fig. 12.4c. However, afterwards the distance re-adopts fluctuations around a similar average value as before the perturbation, which shows that overall the G$\beta$-strand remains unaffected by this process. Next, we deduce from the strong jump of about 1 nm in the distance between Asp1056-OD1(OD2) and Gln1013-NE2, which can be observed at 3000 ps in Fig. 12.4d, that the cleavage of the H-bond between Glu1034 and Gln995 at 2500 ps causes a subsequent breakage of the H-bond between Asp1056-OD1(OD2) and the amino group of Gln1013-NE2, located at the opposite end of the J$\alpha$-helix and the H$\beta$-strand respectively. By further considering the distance between Gln1029-NE2 and Asn1008-OD1 in Fig. 12.2b, we clearly see that the rotational re-orientation of the glutamine Gln1029, which begins at 2500 ps, results in the subsequent disruption of the H-bond between Asp1056 and Gln1013. In this situation the J$\alpha$-helix starts successively to disrupt from one end of the LOV2-domain. At 17000 ps, we observe an accelerated increase of the distance between Asp1056-OD1(OD2) and Gln1013-NE2. This coincides with the full disruption of the H-bond between Glu1034 and Gln995, characterized through the rise in the distance between Glu1034-(OE1)OE2 and Gln995-NE2 in Fig. 12.4b at the same simulation time. The latter process

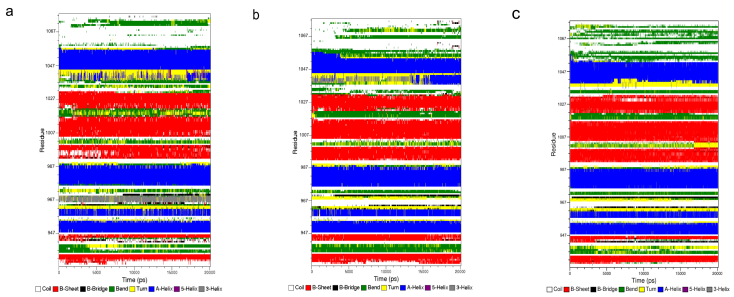


Figure 12.6.: Secondary structure analysis of the wildtype AsLOV2-Jα-photosensor and its Gln1029Asn-mutant. Secondary structure analysis from simulations of the wildtype AsLOV2-Jα-system in the dark (a) and adduct (b) state, as well as of its Gln1029Asn-mutant in the adduct state (c).

ultimately leads to the complete detachment of the Jα-helix from the LOV2-domain, which is followed by a relaxation of the β-strands. This conclusion can easily be drawn from the drops in the distance at 18000 ps between the backbone nitrogen atoms Gln1029-N and Glu1034-N on the Iβ-strand in Fig. 12.4a, as well as the distance between the backbone nitrogen Gln995-N on the Gβ-strand and FMN-O4 in Fig. 12.4c. In their seminal paper on phototropin switches Harper *et al.* [53] compared the dark and adduct state 3D-HNCO-spectra of the AsLOV2-Jα-system, using the minimum chemical-shift-difference method to quantify the spectral changes. They observed within the LOV2-core large differences for the residues close to the FMN-chromophore as well as for the ones on the Gβ-, Hβ- and Iβ-strands. Significant perturbations were also observed at sites on the Eα-helix and the Jα-helix over 15 Å away from the FMN-chromophore. They concluded from their work that the structural changes, induced by the formation of the cysteinyl-FMN-adduct, propagate through the LOV-domain, ultimately causing the disruption of the Jα-helix. To determine which of the protein segments at the LOV2-Jα-interface could be critical in the detachment process, they further created a truncated protein lacking the last 24 residues of the AsLOV2-Jα-system (AsLOV2ΔJα), starting from residue Glu1053 up to the end of the Jα-helix. This deletion removes a key portion of the C-terminal Jα-helix, including the H-bond forming residue Asp1056. The comparison of the dark-state $^{15}N/^{1}H$-HSQC-spectra of the AsLOV2-Jα- and AsLOV2ΔJα-system revealed significant chemical-shift differences for residues located in the core of the protein domains. Further removal of C-terminal residues, starting from Asp1038, leaving only the LOV2-core (AsLOV2Δ) produced no additional spectral changes in the dark-state spectra. These results indicate that the Jα-helix possesses a critical interaction with the LOV2-core in the protein part containing Asp1056 that is removed by truncation. In Fig. 12.5 we present secondary structures of the AsLOV2-Jα-system in the dark (a) and adduct (b) state at the final simulation time of 20 ns. In the figure we highlight the H-bonds between Gln995 and Glu1034 (green) as well as Asp1056 and Gln1013 (red), in addition to the residue Gln1029 (violet). By comparing both structures, we clearly see that the structural changes at the Hβ- and Iβ-strands upon adduct formation go along with the breakage of H-bonds at the LOV2-Jα-interface. Moreover, we recognize that the Jα-helix in case of the adduct state is partially unfolded at its C-terminal end. We note that in this segmental part of the Jα-helix, starting from Lys1049, the sequence is more conserved among phototropins than at the N-terminal end of the Jα-helix and the Iβ-Jα-loop [55], which indicates that this part of the protein might be important in the signal transduction pathway. This observation is further confirmed in Fig. 12.6,

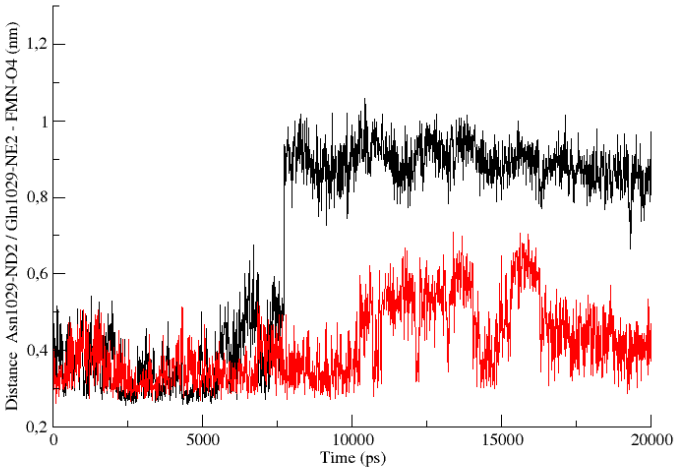


Figure 12.7.: Spatio-dynamical behavior of the conserved glutamine of the AsLOV2-J$\alpha$-photosensor versus the corresponding asparagine of its Gln1029Asn-mutant. Distance between Gln1029-NE2 and FMN-O4 in the wildtype AsLOV2-J$\alpha$-system (black) in comparison to the distance between Asn1029-ND2 and FMN-O4 in the AsLOV2-J$\alpha$-Gln1029Asn-mutant (red). Both photosensors are in their adduct state form.

where we present the secondary structure analysis of our simulations of the AsLOV2-J$\alpha$-system in the dark (a) and adduct (b) state. While in the dark-state simulation the J$\alpha$-helix remains stable over the entire production run, we see that the J$\alpha$-helix in the adduct state simulation partially looses its helical structure at a simulation time of 3000 ps, starting from Asn1054 up to the end of the J$\alpha$-helix. This unfolding process clearly correlates with the breakage of the H-bond at the same simulation time between Asp1056 on the J$\alpha$-helix and Gln1013 on the H$\beta$-strand, shown in Fig. 12.4d. In addition, in the secondary structure analysis we observe at around 17000 ps an increase of the length of the J$\alpha$-helix through the formation of an additional helical element between the residues Asp1038 and Glu1043. This event coincides with the full disruption of the H-bond between Glu1034 and Gln995 at the same simulation time, as can be deduced from Fig. 12.4b. Finally, we observe in case of the adduct state a stabilization of the smaller $\beta$-strands at the outer peripheral of the LOV2-domain, denoted as A$\beta$- and B$\beta$-strands, which may contribute to an increase of the hydrophobicity of the LOV2-domain at the interface and accelerate the detachment of the J$\alpha$-helix from the LOV2-domain. The rest of the secondary structure remains essentially unaffected through the adduct formation. It is worth emphasizing in this context that our conclusions concord well with the NMR-measurements of Harper *et al.* [53] on the AsLOV2-J$\alpha$-system in the dark and adduct state. In their study it was demonstrated that the photosensor undergoes a selective unfolding of the C-terminal J$\alpha$-helix upon adduct formation, implicating that the conformational changes in the LOV-domain must be sufficiently destabilizing, in order to displace the J$\alpha$-helix without globally unfolding the LOV-core. These findings were confirmed in a series of consecutive experimental works on the same system, using

various spectroscopic techniques [56, 46]. To investigate the role of the unfolding process of the Jα-helix after its detachment from the LOV2-core, Harper *et al.* [54] created point mutations that disrupted the Jα-helix in the absence of light. From these experiments, they concluded that the unfolding of the Jα-helix is the critical event for triggering the kinase activity.

### 12.0.12. Gln1029Asn-mutant of AsLOV2-Jα-photosensor

To demonstrate how the precise knowledge of the mechanism of signal transduction can be used to influence the transduction pathway and engineer the signal activity of the AsLOV2-Jα-system, we designed a point mutant, replacing Gln1029 with its methylene-less counterpart asparagine, and investigated the effect of this substitution on the mechanism of signal transduction. In Fig. 12.7 we compare the distance between Gln1029-NE2 and FMN-O4 in the wildtype AsLOV2-Jα-system to the corresponding distance in the mutant between Asn1029-ND2 and FMN-O4. We observe that the mutant does not show any major jump in the distance, which might indicate an abrupt detachment of the Asn1029-ND2 on the Iβ-strand from the FMN-chromophore, as observed in case of the wildtype for the distance between Gln1029-NE2 and FMN-O4 at 7500 ps. To assess the effect of the mutation on the light-induced structural changes and thus the signal intensity, we compare next the secondary structure analysis of the wildtype AsLOV2-Jα-system and its Gln1029Asn-mutant in the adduct state in Fig. 12.6b and 6c. We see that in case of the asparagine-mutant the light-induced structural changes, observed in the residue range of the Jα-helix, are reduced compared to the wildtype AsLOV2-Jα-system. This indicates that the signal activity is damped due to the substitution and that Gln1029 plays a central role in the signal propagation from the LOV2-core to the Jα-helix in the wildtype system, sensitively affecting the intensity of the signal. We note that these results agree well with recent experimental measurements of Nash *et al.* [56] on the same system, using a combination of UV-visible-absorbance-, circular-dichroism- and NMR-spectroscopy. These authors found that mutation at this site causes a significant reduction of the light-induced structural changes of the protein domain, while maintaining its photocycling capabilities. Moreover, they concluded, to explain the slightly accelerated dark-state recovery kinetics of the mutant with respect to the wildtype AsLOV2-Jα-system, that the mutant retains the H-bond between the amide group of Asn1029 and FMN-O4, observed in the dark-state, also in the adduct state form. Thus, we see that a precise knowledge of the mechanism of signal transduction allows to engineer the signal activity of the AsLOV2-Jα-system by altering the degree of light-induced structural changes via the introduction of mutations and demonstrates the predictive power of our simulation procedure.

### 12.0.13. Discussion

Our work demonstrates that the signal transduction pathway after adduct formation of the AsLOV2-Jα-photosensor begins with altering the H-bond network via a rotational re-orientation of a conserved glutamine residue adjacent to the FMN-chromophore, inducing the coupling of the Iβ- and Hβ-strands as well as the propagation of the stress via the Iβ-strand to the LOV2-Jα-interface. The stress then causes the breakage of a H-bond at the interface between residues on the C-terminal end of the Jα-helix and the Hβ-strand, as well as another H-bond between residues on the Jα-Iβ-linker and the Gβ-strand. Ultimately, this leads to a disruption of the Jα-helix from the LOV2-core and a partial unfolding of the Jα-helix. We anticipate that the elucidation of the AsLOV2-Jα-signal transduction pathway on a molecular level presented in this work will provide new perspectives for the creation of novel fusion proteins in combination with the AsLOV2-Jα-photosensor. As demonstrated on the example of its asparagine mutant AsLOV2-Jα-Gln1029Asn, this knowledge can be used to engineer novel fusion proteins with

optimized light-induced response, which might open new avenues for allosteric control of protein activity in living cells [223, 229, 225, 226].

# 13. Full-length Phototropin1

> LOV1 core
GLRHTFVVADATLPDCPLVYASEGFYAMTGYGPDEVLGHNCRFLQGEGTDPKEVQKIRD
AIKKGEACSVRLLNYRKDGTPFWNLLTVTPIKTPDGRVSKFVGVQVDVTS
> Jα
KTEGKALADNSGVPLLVK

Figure 13.1.: LOV1 sequence and part of L1-L2 used for modelling of new LOV1-linker model from *Chlamydomonas reinhardtii*

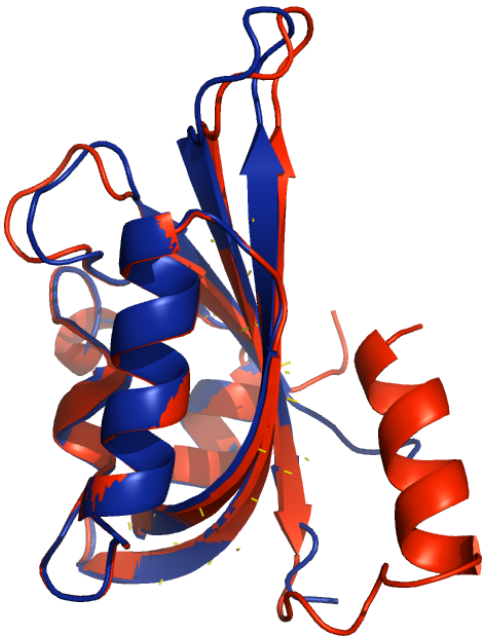

Figure 13.2.: RMS-alignment of LOV1-model (AsLOV2-template) (red) with LOV1-domain from *Chlamydomonas reinhardtii* (blue) of Fedorov *et al.* [11].

## 13.1. Introduction

Phototropin1 (phot1) is essential for the phototropic response in plants and green algae. It contains two LOV-domains and one kinase domain, which performs intra- and intermolecular phosphorylation of phot1 after blue-light irradiation *in vivo*. Several authors performed experiments on the functionality of phot1 *in vivo*. Wan *et al.* performed fluorescence microscopic

> LOV1-LOV2 linker
KTEGKALADNSGVPLLVKYDHRLRDNVARTIVDDVTIAVEKAEGVEPGQASAVAA
AAPLGAKGPRGTAPKSFP

Figure 13.3.: L1-L2 linker sequence from *Chlamydomonas reinhardtii*.

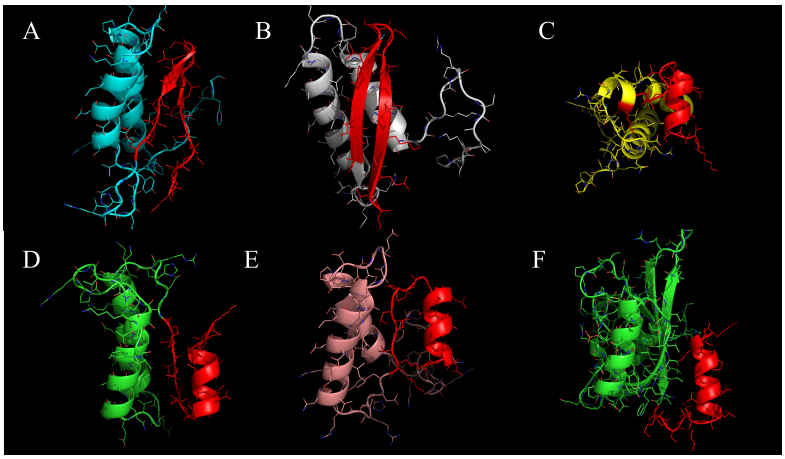


Figure 13.4.: Alignments of L1-L2 linker sequence (A-E); LOV1-J$\alpha$ represented in red colour. F: LOV1-J$\alpha$-system.

experiments on seedlings of *Arabidopsis thaliana* and found, that phototropin1 relocalizes within each stage of illumination. They showed, that phot1 translocates from the membrane into the cytoplasm, when exposed to blue light irradiation [230]. Folta *et al.* found through fluorescence microscopy and northern blotting, that phototropin1 is required for mRNA destabilization and causes chloroplast accumulation after blue light irradiation in *Arabidopsis thaliana* [231]. They showed that in case of the absence of phot1 in *Arabidopsis thaliana* the light-harvesting, clorophyll-binding (Lhcb) transcript is not destabilized. Folta *et al.* concluded from this fact, that phot1 is responsible for the destabilization of the Lhcb transcript in *Arabidopsis thaliana in vivo*. Through subsequent RNA-extraction and northern blotting after having grown the seedlings of *Arabidopsis thaliana* in the dark, Folta *et al.* prove that the fluence responce of the Lhcb transcript increases. This is not the case if phot1 is absent. This indicates that the WT-plant induces chloroplast accumulation, through direct interaction of phot1 with the RNA Lhcb-transcript. Inada *et al.* found that signal transduction is also performed through strong interaction with RPT2, which is another interacting protein, promoting stomatal opening in *Arabidopsis thaliana* [232]. We point out that all phot1-proteins in the different organisms have highly identical sequences and the mechanisms of signaling in all of these organisms must be similar. This was proven by Onodera *et al.*, who inserted the Cr-phototropin1 into the higher plant *Arabidopsis thaliana* [233]. For this purpose, they used a mutant, which is deficient of phot1 and phot2. Subsequently, they fused a Cr-phot1-vector with the promoter of the cauliflower mosaic virus. They observed, that the basic mechanism of phototropin activation was highly conserved, whereas the physiological functions remained diverse in comparison to WT-*Arabidopsis thaliana*. This latter fact might be caused by differences in the life-cycles of the different organisms. Huang *et al.* found through northern blotting, immunoblotting, nucleic acid transformation, gametogenesis and zygote germination that Cr-phot1 is responsible for the modulation of the sexual life

```
>lov2
RVALDLATTVERIQQNFCISDPTLPDCPIVFASDAFLELTGYSREEVLGRNCRFLQGAGT
DRGTVDQIRAAIKEGSELTVRILNYTKAGKAFWNMFTLAPMRDQDGHARFFVGVQVDVT
AQSTSPDKAPVWNKTPEEEVAKAKMGAEAASLISSALQGMAAPTTANP
```

Figure 13.5.: LOV2-J$\alpha$-sequence from *Chlamydomonas reinhardtii*

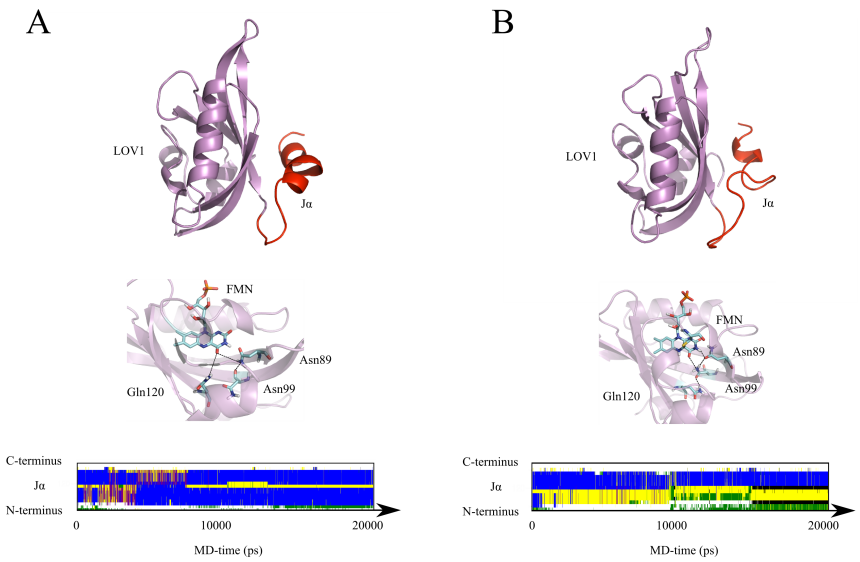


Figure 13.6.: Final structures and secondary structure analysis of LOV1-J$\alpha$. A: dark-state; B: light-state.

cycle of the green algae Cr [219]. They found that green algae *Chlamydomonas reinhardtii* (Cr), which were phot1-deficient, were partially impaired in the three steps of the life cycle of Cr. This indicates that Cr-phot1 is a molecule, which is primarily responsible for regulating the sexual life cycle of Cr. Especially in the pre-distinction between activated and in-activated gametes, Cr-phot1 plays an important role, because it was found, that gametogenesis does not take place in the dark. Moreover, the subsequent tetrade is also blue-light induced [219]. The overall activation mechanism of phototropin1 was at first investigated by Christie *et al.* [234]. They observed that the non-phototropic hypocotyl (NPH1)-gene in *Arabidopsis thaliana* is encoding a 120 kDa serine-threonine kinase. Through performing fluorescence excitation spectroscopy they found that phot1, which is encoded by NPH1, transduces its signal by auto-phosphorylation. Salomon *et al.* performed chromatographic experiments on the LOV1-LOV2 part of phot1 and hypothesized that LOV1 in phot1 of *Avena sativa* might be responsible for the dimerization of WT-phot1 and further might modulate trans-phosphorylation [36]. Kaiserli *et al.* performed absorbance spectroscopy, *in vitro* phosporylation analysis and confocal microscopy experiments onto phot1 in *Arabidopsis thaliana in vivo* and *in vitro* [40]. They found, that in phot1 solely the LOV2-domain is the primary dark-state repressor in phot1, while LOV1 is responsible for high-light chloroplast accumulation and does not act as a dark-state repressor. Independently of

LOV1, phot1 performs intermolecular phosphorylation, via light dependent dimerization. This transphosphorylation is sufficient to promote phot1-internalization triggered primarily by phosphorylation of Ser-851 in the kinase activation loop [40].

Experiments on Crphot1 *in vitro*, lead to futher insight into the functionality of phot1 on the level of secondary structure elements. Pfeifer *et al.* performed FTIR-spectroscopical investigations and mutational analysis experiments onto the full-length phototropin1 of *Chlamydomonas reinhardtii* [165]. At first, they investigated the dynamics of the J$\alpha$-helix within the LOV1-LOV2-J$\alpha$ system, with and without this $\alpha$-helix. They observed, that the LOV1-LOV2-J$\alpha$ construct closely resembles the spectrum LOV1-LOV2 construct upon illumination with a single difference in the band at 1646 $cm^{-1}$. They assigned this band to the perturbed J$\alpha$-helix, which consists of 20 amino-acids. They compared these spectra with the spectra of full-length phot1, the full-length phot1 without the J$\alpha$-helix and the full-length phot1 without the activation-loop (A-loop) extension. They concluded from these experiments, that upon light illumination the J$\alpha$-helix unfolds and the LOV2-domain cleaves from the kinase. This latter event causes further unfolding of the A-loop extension, which activates the kinase and leads to phosphorylation [165]. Through modelling, they finally proposed a structure in which the A-loop extension is located externally from the kinase. Moreover, through secondary structure prediction they found that this A-loop extension has $\alpha$-helical content. Finally, they concluded that unfolding occurs upon light irradiation in the J$\alpha$-helix and the A-loop extension of the kinase domain. Moreover, they proposed a model of the kinase and concluded that these latter unfolding events lead to an activation of the kinase domain and subsequent phosphorylation. However, a detailed mechanism, which involves a structural model of full-length phototropin1 on a molecular level is still lacking.

In this work, we introduce our model of phototropin1 of *Chlamydomonas reinhardtii* (Crphot1) and give insight into the mechanistic details upon light activation of this photo-sensor using the KMC-MD technique. At first we explain the preparation of the starting structure for the phototropin system. Then, we discuss the results of our simulations and compare them with experimental data. Based on this investigation, we derive a model of the mechanism of light-activation of Crphot1 *in vivo*.

## 13.2. Modelling of Phototropin1 of Chlamydomonas reinhardtii

To create the initial structure of phot1, we started with the modelling of the LOV1-domain and the LOV1-LOV2 linker (L1-L2). The CrLOV1-domain was determined through X-ray diffraction measurements of Fedorov *et al.* [11]. We point out that a multitude of structures based on the LOV2-domain from *Avena sativa* have been published by several authors [10, 55, 72]. However, the role of L1-L2 remains unclear and was not structurally resolved by experiment. Notably, all refined structures of LOV-domains have either N- or C-terminally located $\alpha$-helices. Most prominent examples for N-terminal helices are the PYP and VVD-LOV [7, 65], the so-called N-cap. Moreover, C-terminally located helices are found in all LOV2-J$\alpha$ systems of plant phototropins [10, 53, 54, 55], as well as in bacterial LOV-domains as Ytva-LOV from *Bacillus subtilis* and El-222 from *Erythrobacter litoralis* [236, 80]. One of the few examples of a LOV-structure, where no helix either N- and C-terminally has been found to date is the structure of CrLOV1. However, the sequence N- and C-terminally to this LOV1-domain is not resolved, while $\alpha$-helical content in the N- or C-terminal region of the LOV-core is highly probable. For generation of our LOV1 - L1-L2 model, we used the protein structure prediction program ROBETTA [235] and the SWISS-Modeller [237, 238, 239]. These tools have been conceived for the prediction of unknown structures. In particular ROBETTA uses an *ab initio* modelling approach, while SWISS-modeller performs homology modelling. We started with modelling of the LOV1-domain through apply-

ing the SWISS-Model algorithm on the sequence of LOV1- and the L1-L2 sequence, as shown in Fig. 13.1 [237, 238, 239]. As a template for modelling, we used the AsLOV2-structure of Harper *et al.* [53] as single template-structure. Through alignment of the resulting structure with the structure of Fedorov *et al.*, we found that both structures can be closely aligned by a value lower than 0.468 Å, which validates the high similarity between our model and the LOV1 X-ray structure of Fedorov *et al.*. We illustrate this further in Fig. 13.2. We note that our model fits very well with the experimentally refined structure of Fedorov *et al.* [11], while solely the regions of highly flexible linkers between the secondary structural elements are deviating by 0.468 Å in the alignment. Moreover we found an additional J$\alpha$-helix in the C-terminal region of CrLOV1, by *ab initio* modelling of L1-L2 by using the ROSETTA algorithm [235] with the L1-L2 sequence. We obtained 5 alignments, as shown in Fig. 13.4. We observe, that all alignments contain 2 large $\alpha$-helices, while 2 of the alignments contain a N-terminal $\beta$-strand alignment. However, 3 alignments show the same $\alpha$-helical structure in the same N-terminal region, as obtained from the SWISS comparative structural modelling, as described previously and shown in Fig. 13.4F. Moreover, most of the *ab initio* models are $\alpha$-helically structured in this region. It is worth mentioning, that $\beta$-sheet alignments also did not occur in all known C- or N-terminal regions of X-ray structures of LOV-domains. We conclude, that an $\alpha$-helical alignment, located C-terminally to the CrLOV1-domain is highly probabable. Finally we chose the structure, which is visualized in Fig. 13.4E, because its N-terminal L1-L2 structure in this model is closely overlapping with the SWISS-modelled LOV1-J$\alpha$ system. The CrLOV2-J$\alpha$-system was generated in the same procedure by using the comparative SWISS-modeller [237, 238, 239], with the input-sequence shown in Fig. 13.5.

## 13.2.1. MD simulation of LOV1-J$\alpha$-system

### System preparation

We centered the LOV1-J$\alpha$-system in a cubic box with boxlength 6.11 nm and solvated it with 7018 SPC-waters. No additional ions were needed to render the system electroneutral. We equilibrated the system under invasive thermostating control for 1 ns, while the production run for 20 ns was performed under non-invasive thermostating. We note, that the light-state and MD-related parameters were chosen as described in section 3.6.

### Results

We continue by describing the MD-simulation of the isolated CrLOV1-J$\alpha$-system of light and dark-state. We display the results in Fig. 13.12 A and B, where we show the final structures and the secondary structure analysis of the dark- and light-state MD-simulation. We observe, that the J$\alpha$-helix is highly conserved in the dark-state in contrast to the light-state simulation, where this helix is unfolded. We observe, that a slight tilting of the J$\alpha$-helix has occurred in both cases in comparison to the starting structure, which originally had a parallel alignment along the $\beta$-strand surface. In the dark-state as well as in the light-state this helix has tilted into an orthogonal orientation along the hydrophobic surface, as we display in the Figs. 13.6 A and B. We see that experimental findings of Harper *et al.* on LOV2-J$\alpha$ are in agreement with our LOV1-case, where the J$\alpha$-helix is disrupted in the light-state as a consequence of adduct formation [53]. Moreover, we find that this latter adduct formation leads to subsequent coupling between Gln120 and Asn99, as we previously have shown for other light-state LOV-domains [246] (see Figs. 13.6 A and B). Additionally, we deduce from secondary structure analysis of our simulations that the J$\alpha$-helix is disrupted N-terminally in the light-state, which is confirmed by our previous simulations, where we also observed N-terminal disruption of the J$\alpha$-helix.

>LOV2-Kinase linker
WAAISGVIMRRKPHKADDKAYQALLQLQERDGKMK
>Kinase
LMHFRRVKQLGAGDVGLVDLVQLQGSELKFAMKTLDKFEMQERNKVARVLTESAILAAVD
HPFLATLYCTIQTDTHLHFVMEYCEGGELYGLLNSQPKKRLKEEHVRFYASEVLTALQYL
HLLGYVYRDLKPENILLHHTGHVLLTDFDLSYSKGSTTPRIEKIGGAGAAGGSAPKSPKK
SSSKSGGSSSGSALQLENYLLLAEPSARANSFVGTEEYLAPEVINAAGHGPAAVDWWSLG
ILIFELLYGTTPFRGARRDETFENIIKSPLKFPSKPAVSEECRDLIEKLLVKDVGARLGS
RTGANEIKSHPWFKGINWALL

Figure 13.7.: LOV2-kinase linker- and kinase-sequence from *Chlamydomonas reinhardtii*.

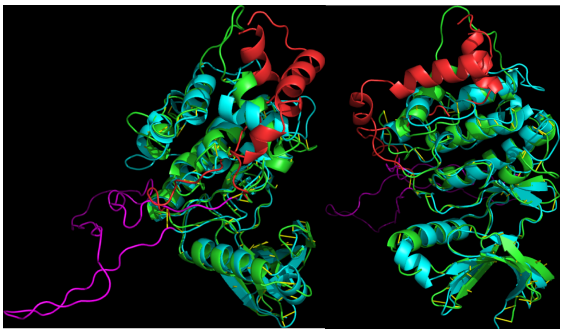

Figure 13.8.: Overlay of the structural models of the kinase-domain from *Chlamydomonas reinhardtii*. Green: Structure of Pfeifer *et al.*, Blue: Our model. Violet: A-loop extension of Pfeifer *et al.*. Red: A-loop extension of our model.

### 13.2.2. Model preparation of the Cr kinase-domain

In the following we describe the model of the Cr kinase-domain, where we used the sequence shown in Fig. 13.7. In the work of Pfeifer *et al.* this phot1-sequence was modelled using 30 kinase structures, which overlap in sequence with largest probability. From their model they inferred that the A-loop structure within the C- and N-terminal lobes of the kinase structure was totally unfolded and located externally from the kinase [165]. Moreover, they find that the A-loop extension is folded in the in-active dark-state of the kinase domain [165], whereas in the active state of the kinase, this A-loop extension unfolds. In order to obtain a dark-state model of the kinase, we modelled the kinase-domain, by taking one single template structure as an input for the comparative SWISS-modeller. In contrast to Pfeifer *et al.* we chose one kinase-structure as template (PDB-code: 3R63), which has a partially folded A-loop extension. As a result of our modelling, we obtained the structure of the Cr-kinase domain, as shown in Fig. 13.8. We observe, that in contrast to the model of Pfeifer *et al.*, this A-loop extension is folded, which mimics the in-active state [165]. By comparing our model with the model of Pfeifer *et al.*, we conclude that they only differ in the A-loop extension, as shown in Fig. 13.8. Moreover, we observe that in contrast to the model of Pfeifer *et al.* this A-loop extension is folded.

Finally, we connected each modelled part of Crphot1 through minimization using the l-BFGS minimizer, as we show in Fig. 13.9. Throughout the connection of each part we provided the orientations of each domain with lowest sterical repulsion. In the alignment of the LOV2-core onto the kinase-domain, we took into account that the LOV2-F$\alpha$-helix acts as an interaction site with the kinase domain, as demonstrated through 2D-solution NMR-experiments by Amezcua *et al.* [240]. Moreover, we considered the scheme of Pfeifer *et al.* [165], who demonstrated

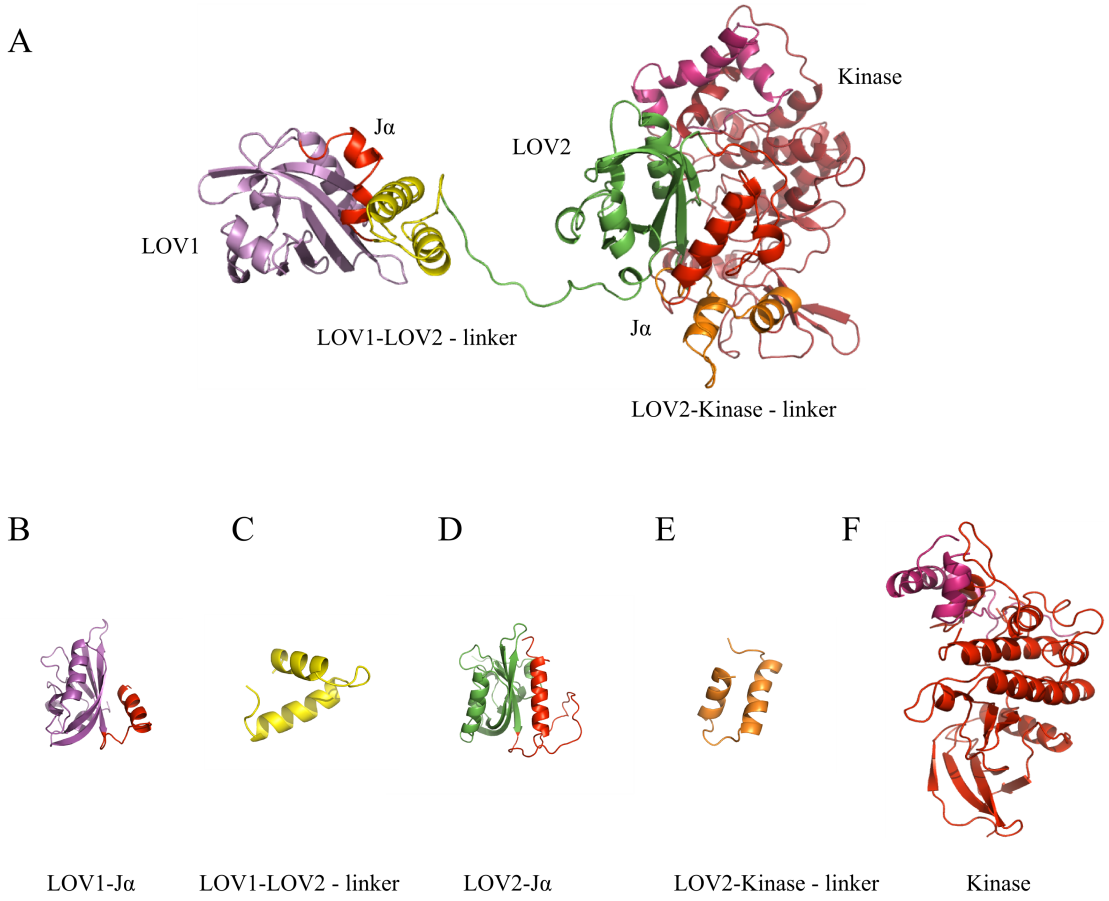


Figure 13.9.: Model of full-length-phototropin of *Chlamydomonas reinhardtii*. A: full-length-phototropin1, B: LOV1-Jα, C: LOV1-LOV2 - linker region, D: LOV2-Jα, E: LOV2-kinase - linker region, F: Kinase-domain

through their experiments, that the LOV2-domain deinhibits the kinase domain upon cleavage of the J$\alpha$-helix [165, 40]. Finally we point out, that in this model Ser611, which is located at the active site of the kinase, is inhibited by the LOV2-domain. This was demonstrated by Kaiserli *et al.*, who showed that primarily phosphorylation of Ser611 occurs upon illumination of the LOV-domain. In the next sections we discuss the results of our simulations on the light- and dark-state structures of phot1.

## 13.3. KMC-MD simulation of full-length Crphot1

We prepared the model of full-length phot1 as previously described and centered the protein into a cubic box with boxlength of 13.84 nm and filled the box with 82751 SPC-waters. The system itself was totally electroneutral and no additional ions were added. We subsequently performed a MD run for 1 ns with a timestep of 1 fs with the invasive thermostating procedure in the dark-state of Crphot1. We observed, that during this equilibration phase, no jumps in the potential energy and no jumps in the thermodynamic observables (temperature 300 K, pressure 1 atm) occurred. From this equilibrated structure, we generated the light-state, as described in section 3.6. Subsequently, we started 8 independent simulations with the dark and light-state using our newly implemented KMC-MD scheme, as described in chapter 4.

### 13.3.1. Simulation parameters

We used the same simulation parameters as the ones described in chapter 4, excepting the break-radius, which was set to a value of 0.3 nm to obtain better performance for H-bond breakage. Moreover, we changed the length of the MD-phase from 5 ps up to 20 ps, due to the large statistical correlations in the system.

## 13.4. Results

### 13.4.1. Vicinity of FMN-chromophore

We start with the analysis of the multiscale transduction dynamics of phot1 by visualizing the amino-acids in the vicinity of the FMN-chromophore for dark and light-state in Fig. 13.10. In case of the dark-state no coupling occurs between the amino-acids Asn99 and Gln120 in the LOV1-domain, as well as in the LOV2-domain, where also no coupling occurs between Asn292 and Gln313 (see Fig. 13.10 A and C). We point out that these latter couplings occur in the light-state of the LOV1- and LOV2-domain, as shown in Figs. 13.10 B and D respectively. In all of our previous works on LOV2-J$\alpha$-based systems and VVD-LOV, we observed the same coupling behavior between glutamine and asparagine in the light-state [246, 245]. As mentioned in the previous chapter 12 and 4, we showed that this H-bond formation process between these amino-acids leads to a coupling between H$\beta$ and I$\beta$, which triggers the disruption of the J$\alpha$-helix.

### 13.4.2. LOV1-LOV2 linker region

Next, we analyse the L1-L2 region and the secondary structure evolution of this part of the system, as shown in Fig. 13.11. In this figure we show the overlaid final structures of phot1 of two dark- and two light-state simulations. In Fig. 13.11A we observe that the two LOV-domains dimerize in the dark-state, in contrast to the light-state. We observe a loosening of $\alpha$-helical structure linker-region in the dark-state. Both segments of L1-L2 are disrupted and

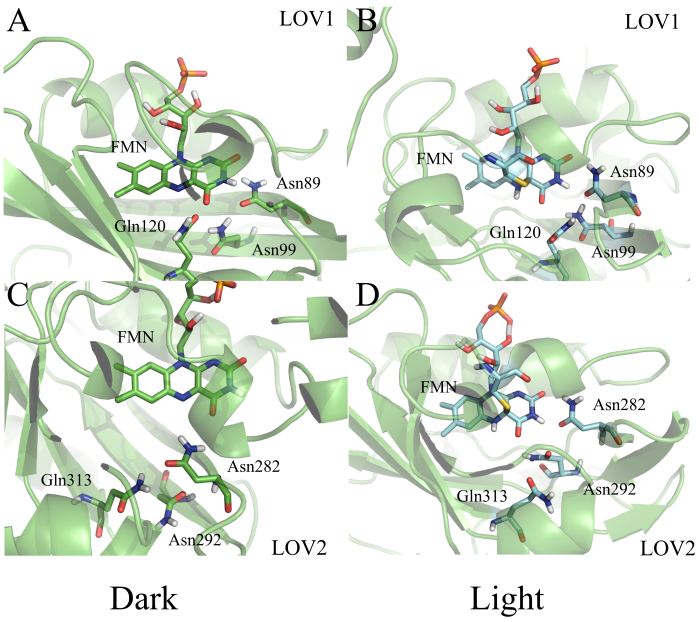


Figure 13.10.: Final structures in vicinity of chromophore FMN in LOV1 (A,B) - and LOV2 (C,D) -domain in dark (A,C) and light (B,D) -state.

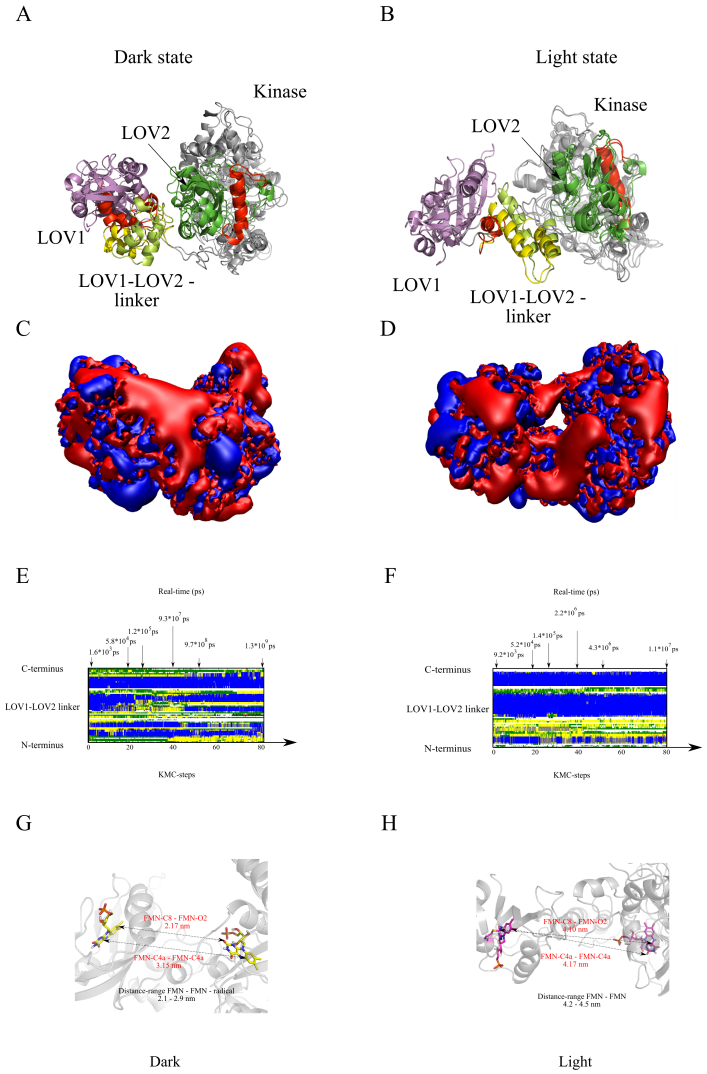


Figure 13.11.: Final structures of KMC-MD simulation of dark and light-state. A: dark-state, B: light-state. C: APBS-electrostatic surface of dark-state, D: APBS-electrostatic surface of light-state. E: Secondary structure analysis of LOV1-LOV2 linker region in dark-state. F: Secondary structure analysis of LOV1-LOV2 linker region in light-state. G: Final configuration of FMN-chromophores of LOV1 and LOV2 in dark-state. H: Final configuration of FMN-chromophores of LOV1 and LOV2 in light-state.

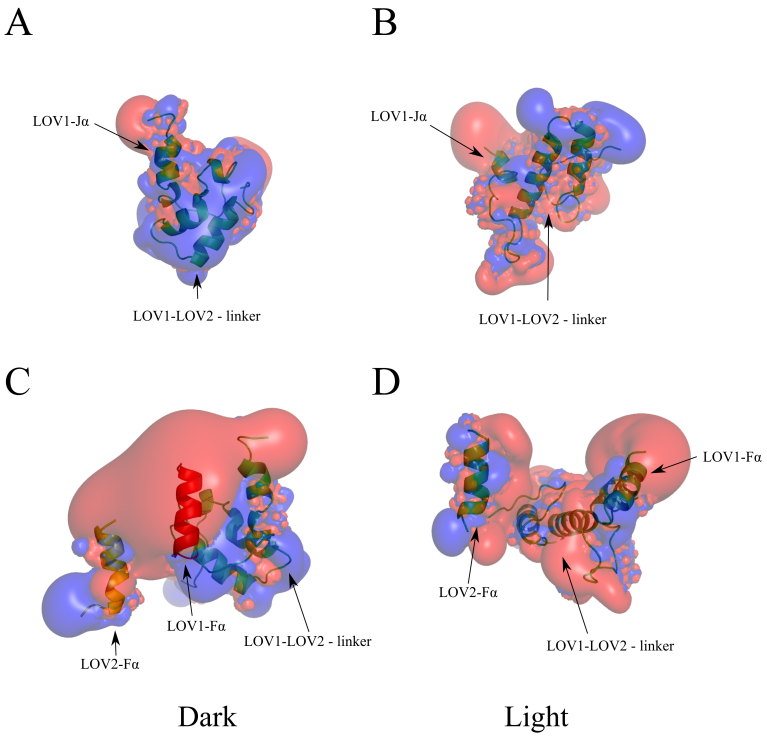


Figure 13.12.: Final structure and APBS analysis of L1-L2 part of phot1. APBS-analysis of L1-L2 region for dark (A,C) and light-state (B,D).

become unstructured between the LOV-domains in the dark-state. In the secondary structure analysis of L1-L2 in Fig. 13.11E, we observe the same behavior. By contrast, we see a stable inter-domain L1-L2 region with two $\alpha$-helical segments between the LOV-cores in the light-state, as shown in Fig. 13.11B and 13.11F. This L1-L2 linker segment forms a HTH-motif, in contrast to the dark-state. We point out that this HTH-motif might play an important role in the aggregation tendency of full-length Cr-phot1 towards cellular targets, such as DNA- or RNA-transcripts as well as the cell-membrane. Wan *et al.* [230] performed fluorescence microscopic experiments on plant seedlings of *Arabidopsis thaliana* (At). They found that in the light state, Atphot1 accumulates in the cytoplasm, whereas in its dark state it can be localized at the membrane [230]. We confirm this with the APBS-surface analysis of the final configurations of both states, which we display in the Figs. 13.11C and 13.11D. Doltchinkova *et al.* investigated the membrane potential of *Chlamydomonas reinhardtii* through electrokinetic and light-scattering measurements. They found that light exposure of the thylakoid substrate alters the surface electrostatic potential to a positive value, while in the dark this membrane is negatively charged [243]. We note that a positively charged surface is located between the LOV1- and the LOV2-domain in the dark-state structure, shown in Fig. 13.11 C. This indicates that this positively charged surface induces an attractive force towards the negatively charged membrane in the dark-state. Next, we investigate the electrostatic influence of L1-L2 onto the overall charge surface of phot1, as shown in Fig. 13.12. We observe that the positively charged surface is caused by the parallel alignment of the two F$\alpha$-helices of LOV1 and LOV2 in the dark-state, as shown in Fig. 13.12 E. Both latter helices induce a dipole along the parallel axis of the helices, in case of the dark-state. In the same figure we see, that the linker-region exerts a negative charge concentration (see Fig. 13.12C), which causes the dipolar distribution of the electrostatic potential. The light-state is positively charged at the same LOV1-LOV2-linker, due to the different folding situation. Moreover, we see in the Figs. 13.11 C and D, that the LOV1-J$\alpha$ helix in case of the dark-state is folded, whereas the same helix is unfolded in the light-state. We conclude that L1-L2 unfolds in case of a conserved LOV1-J$\alpha$-helix. By contrast, if LOV1-J$\alpha$ unfolds, L1-L2 forms a HTH-motif in the light-state. We further illustrate this in Fig. 13.11 D and E, where we show the secondary structure analysis of L1-L2 in the dark and the light-state of Crphot1. In these two analyses we see that the linker region is split into 5 different segments after a short simulation time in the dark-state, whereas in the light-state two large segments form a stable HTH-motif over the whole simulation time. Moreover, we observe that the $\alpha$-helical content of LOV1-J$\alpha$ is frequently disrupted in the light-state, in contrast to the dark-state simulation, where the number of residues with $\alpha$-helical content is twice as large as in the light-state.

To further analyze the previously discussed behavior at a molecular level, we additionally measured the inter-atomic distances in between the linker-region and the LOV1-core region, which we show in Fig. 13.13. In all simulations of the dark state we observe a condensation of the J$\alpha$-linker region onto the LOV1-core, as shown in the same figure. In case of the two distances between Asp145-OD1-OD2 (J$\alpha$) - Lys107-NZ (I$\beta$) and Arg147-NH1-NH2 (J$\alpha$) - Glu81-OE1-OE2 (H$\beta$) we observe a salt-bridge formation upon KMC-step 47, which means that after this step the J$\alpha$-helix effectively stabilizes along the $\beta$-strand surfaces (see Figs. 13.13 B and C). Moreover, the two amino-acids Glu168 and Asp150, which are located at the linker, form salt-bridges with the N-terminal region of CrLOV1, with the amino-acids Arg19 and Lys72 (A$\beta$-B$\beta$ - region). We point out that this formation of four different salt-bridges within the LOV1-linker region, has an overall electrostatic effect in the dark-state and changes the electrostatic pattern of this region in contrast to the light-state, where the salt-bridges remain cleaved.

Next, we study the influence of L1-L2 on the inter-domain distance between the LOV1- and LOV2-domain. Our focus is here put on the distance between the FMN-chromophores of LOV1

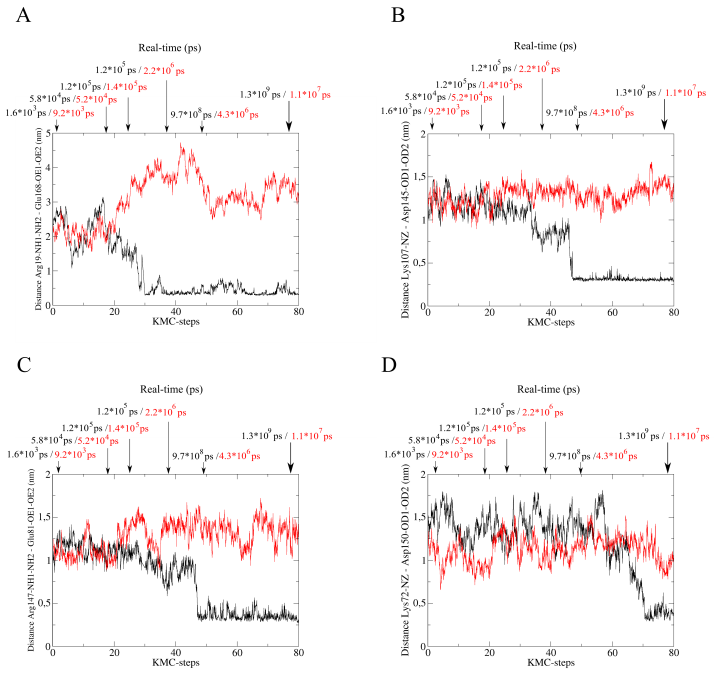


Figure 13.13.: Inter-atomic distances between LOV1-core region and LOV1-J$\alpha$ and LOV1-LOV2-linker versus KMC-simulation time in the dark-state (black) and light-state (red).[ A: Distance Arg19-NH1-NH2 (A$\beta$) - Glu168-OE1-OE2 (LOV1-LOV2-linker). B: Distance Lys107-NZ (I$\beta$) - Asp145-OD1-OD2 (LOV1-J$\alpha$). C: Arg147-NH1-NH2 (LOV1-J$\alpha$ - Glu81-OE1-OE2 (H$\beta$). D: Lys72-NZ (A$\beta$-B$\beta$)- Asp150-OD1-OD2 (LOV1-LOV2-linker).]

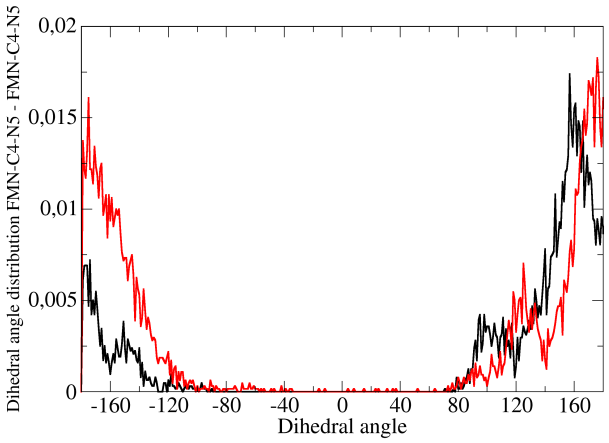


Figure 13.14.: Average dihedral angle FMN-C4-N5 (LOV1) - FMN-C4-N5 (LOV2) distribution of dark (black) and light-state (red).

and LOV2. We measured the distance between the FMN-chromophores of LOV1 and LOV2 and averaged over 4 dark-state as well as 4 light-state simulations. We observe, that the averaged FMN-FMN distances are within a range from 21.7±0.17 Å (FMN-C8 - FMN-O2) up to 31.5±0.21 Å (FMN-C4a - FMN-C4a) in case of the dark state. By contrast, we obtained an average range in the FMN-FMN distances from 41.0±0.24 Å (FMN-C8 - FMN-O2) to 41.7±0.25 Å (FMN-C4a - FMN-C4a) in the light state. In Fig. 13.11 G and H we display these distances for the dark and the light-state. By taking into account that the average electron-density in the dark-state is distributed over the range from the FMN-O2 and the FMN-C4a atom, we get an average-value of 26.6 Å which is in the range of the value of the experiment of Mathes *et al.* [241], who performed ESR-*ELDOR* experiments on the radical state of the Cr-phot1 Cys57-Ser- and the Cys57-Gly mutant. They found, that both FMN-chromophores of LOV1 and LOV2 are at a distance of 24.8 ± 0.7*Å* to each other [241]. We point out that our values fit very well to the experiment of Mathes *et al.*. We further analyzed the orientation of the two FMN-molecules and determined the average dihedral angle distribution over 4 independent KMC-MD simulations between the two FMN-isalloxazine rings. The curves are displayed in Fig. 13.14. We note that the average dihedral angle distribution in case of the light-state is broadened in comparison to the dark-state. In case of the light-state, we see that the angular distribution of the dihedral angle FMN-C4-N5 - FMN-C4-N5 is distributed from -90 degrees up to 80 degrees over a range of 170 degrees. However, the maximum of the light-state distribution (176 degrees) is close to that of the dark-state (156 degrees), while the dark state distribution ranges from -120 to 80 degrees, with a total-range of 140 degrees. We conclude that the light-state has overall a larger variance in the orientation of the two LOV-domains compared to the dark-state.

### 13.4.3. Light-induced phosporylation at kinase domain

In the following we continue with the description of the final structures of the LOV2-J$\alpha$-system of Crphot1 for the dark and light-states visualized in Fig. 13.15. From Fig. 13.15A we infer that in the final structure of the dark-state, the J$\alpha$-helix is conserved along the $\beta$-strand surface.

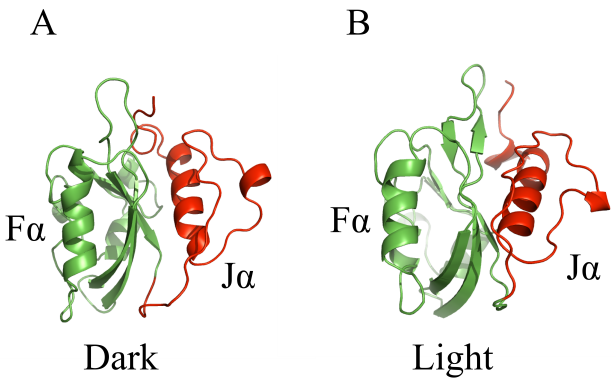


Figure 13.15.: LOV2-J$\alpha$ system of the dark (A) and light (B)-states.

Next, in Fig. 13.15B we display the final structure of the light-state of Crphot1. We see that primarily a shift of the J$\alpha$-helix along the $\beta$-strand surface occurs in the light-state. Moreover, the J$\alpha$-helix tilts its C-terminal end towards the N-terminal part of the LOV2-J$\alpha$ system in the light state. This cleavage of the J$\alpha$-helix in the light-state of LOV2-J$\alpha$ was also shown by us in MD- as well as KMC-MD-simulations, presented in chapters 4 and 12. We point out that the mutual orientation of the F$\alpha$-helix and the J$\alpha$-helix is different in both states, as depicted in Figs. 13.15A and B. Both helices have a parallel alignment in case of the dark-state, as shown in Fig. 13.15A. By contrast, the J$\alpha$-helix in case of the light-state is tilted with regard to the F$\alpha$-helix, as we conclude from Fig. 13.15B. The F$\alpha$- as well as the J$\alpha$-helix play a deceisive role in the interaction of the LOV2-domain with the kinase, as demonstrated in the experiments of Amezcua *et al.* and Pfeifer *et al.* [240, 165]. In Fig. 13.16 we display the final structures of the dark- and light-states after 80 KMC-MD steps with a focus on the influence of the F$\alpha$- and the J$\alpha$-helix on the A-loop of the kinase. From Fig. 13.16A we deduce that both helices are in contact with the kinase-domain in the dark-state. Moreover, we note that the location and mutual orientation of both $\alpha$-helices in the light-state differs drastically from the dark-state, (see Fig. 13.16 B). Kaiserli *et al.* as well as Pfeifer *et al.* demonstrated through their experiments, that primarily the phosphorylation at the active site occurs mainly between the amino-acids Asp547 and Ser611 [165], which are located at the A-loop. We point out that this site is inhibited through a tight binding between the LOV2-F$\alpha$-helix and the A-loop in the dark-state, as visualized in Fig. 13.16A. By contrast, the LOV2-F$\alpha$-helix is cleaved from the kinase in the light-state, which is shown in Fig. 13.16B. In Fig. 13.17 we further display the inter-atomic distances between Ser611-OG (A-loop) and Asp659-OD1-OD2 (C-terminal lobe of kinase) as well as the inter-atomic distances between Ser611-O (A-loop) and Gln265-O (LOV2-F$\alpha$). In case of the Ser611-OG - Asp659-OD1-OD2 distance we observe that both amino-acids form a stable H-bond after KMC-MD step 50 the light-state, while the dark-state remains cleaved over the whole simulation time (see Fig. 13.17A). By contrast, the dark-state forms a H-bond between Ser611 and Gln265 after KMC-MD step 10, while the light state remains at a distance of 0.8 nm, as depicted in Fig. 13.17B. We conclude from this behavior, that Ser611 is subjected to an internal re-organization in the light-state and binds onto the C-terminal lobe of the kinase in the light-state, while this amino-acid binds actively onto the LOV2-F$\alpha$-helix in the dark-state. In Fig. 13.18, we display the APBS analysis on the isolated A-loop as well as the F$\alpha$- and J$\alpha$-helices of the LOV2-system. In the figure we deduce that the F$\alpha$-helix inhibits the kinase domain at

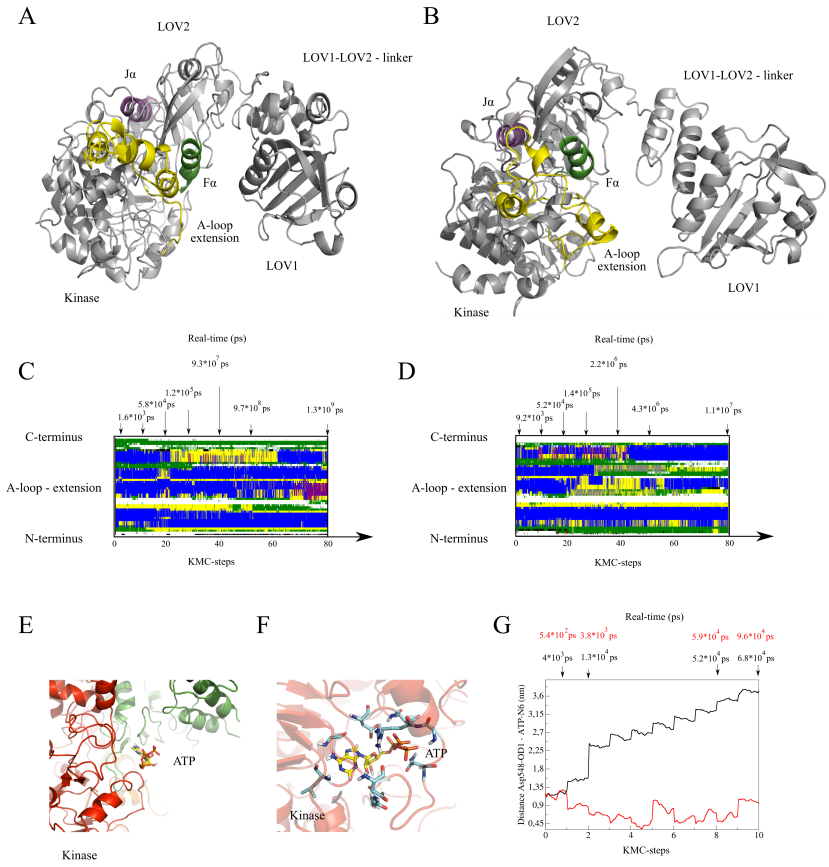


Figure 13.16.: Final structures of different parts of phot1 in the dark- and light-states after 80 KMC-MD steps. A: dark-state. B: light-state. Yellow: A-loop extension, Green: F$\alpha$-helix of LOV2-domain. C: Secondary structure analysis of A-loop extension in dark-state. D: Secondary structure analysis of A-loop extension in light-state. E: Starting structure of ATP-binding simulation with ATP located near Kinase-domain. F: ATP in binding pocket after 10 KMC-MD steps in light-state. G: Inter-atomic distance between Asp548-OD1 (ATP-binding pocket) and ATP-N6 in dark-state (black) and light-state (red).

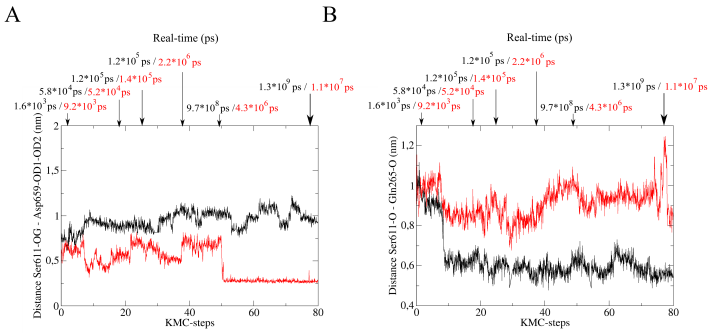


Figure 13.17.: A: Inter-atomic distance between Ser611-OG - Asp659-OD1-OD2, B: Inter-atomic distance Ser611-O - Gln265-O. dark-state (black). light-state (red).

the phosphorylation site Ser611 in the dark-state, while this helix is shifted towards the A-loop extension in case of the light-state. In Fig. 13.18E we show the inter-atomic distance between Ala344-O (LOV2-J$\alpha$) and Ser590-O (A-loop extension). For this quantity we observe in the light-state curve a down shift from a value of 2 nm down to 1 nm after the 23rd KMC-MD-step, while the dark-state curve remains at a constant distance of 2 nm. We conclude from these observations that the F$\alpha$ helix binds tightly onto the A-loop of the kinase domain in the dark-state, while this F$\alpha$-helix is cleaved from the A-loop in the light-state. Moreover, from the previous analysis we infer that Ser611 re-orients within the kinase domain and the J$\alpha$-helix binds onto the A-loop in the light-state. Next, we investigate the catalytic cleft, which is located within Asp547 and Ser611. From Fig. 13.18F we deduce that this cleft opens in the light-state, in contrast to the dark-state. In case of the interatomic distance between Asp547-O and Ser611-O we see a drift from 1.7 to 2.1 nm after the KMC-step 25, while the catalytic cleft remains closed in the dark-state reflected by a distance of 1.8 nm. These latter events lead to an unfolding process of the A-loop region of the kinase domain in case of the light-state, which we show in Fig. 13.16D. By contrast, the A-loop is stable throughout 80 KMC-MD steps in the dark-state, as confirmed by Fig. 13.16C. Moreover, we further conclude from both figures that the $\alpha$-helical content of the A-loop extension is in the dark-state twice as large as the light-state. We note in this context that Pfeifer *et al.* performed FTIR-experiments on full-length Crphot1 and mutants of Crphot1, in which they identified an unfolding process upon light irradiation in this A-loop region [165]. Moreover, we emphasize that we observe the same behavior in the secondary structure analysis of the simulations of both states, which is shown in Figs. 13.16C and 13.16D. In this analysis we observe that the largest changes in the light-state secondary structure occur in the sequential range between 587 and 602 within the A-loop region. We note that this behavior has been also observed in combined mutational and FTIR-experiments of Pfeifer *et al.*, where they identified an unfolding in the A-loop region between the amino-acids 551-605 [165]. We stress, that we observed this light-state unfolding of the A-loop extension in all of our 4 independent KMC-MD simulations, while the dark-state remained stable in this region.

In the following we continue with the analysis of the catalytic cleft. From Fig. 13.19 we deduce that the catalytic cleft opens and becomes solvent accessible in the light-state, while it remains almost not accessible for the solvent in the dark-state. In Fig. 13.19A we show the minimum distance between bulk-water and Asp547, which is located inside the catalytic cleft of the kinase domain, while in Fig. 13.19B we plot the histogram of this minimum distance. We conclude

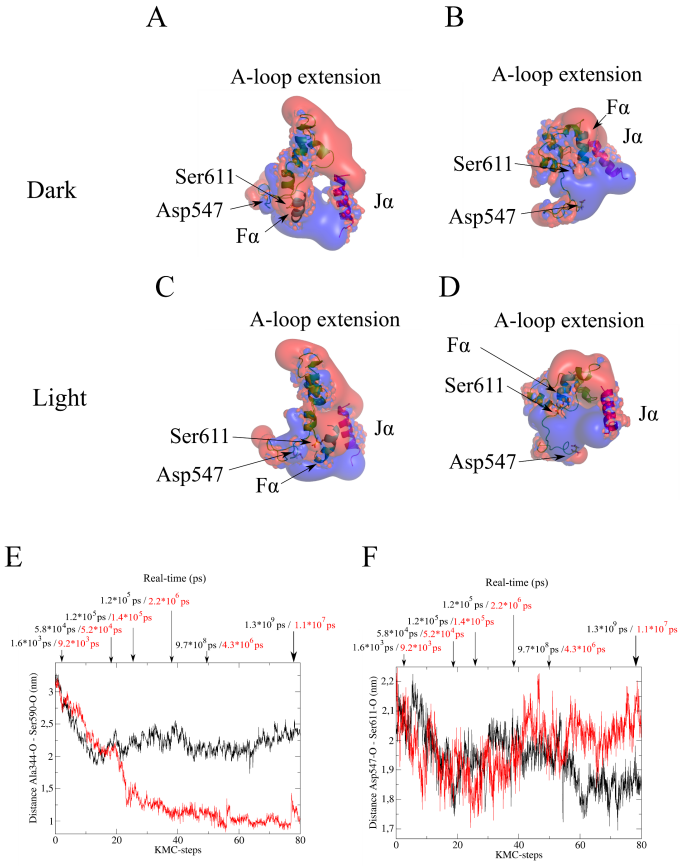


Figure 13.18.: APBS-analysis and inter-atomic distances at the CrLOV2-J$\alpha$ - kinase interface. A: APBS-analysis of A-loop extension, F$\alpha$-helix and J$\alpha$-helix in dark-state. B: APBS-analysis of A-loop extension, F$\alpha$-helix and J$\alpha$-helix in light-state. C: APBS-analysis of dark-state. D: APBS-analysis of light-state. E: Inter-atomic distance Ala344-O - Ser590-O over 80 KMC-MD steps. F: Inter-atomic distance Asp547-O - Ser611-O over 80 KMC-MD steps.

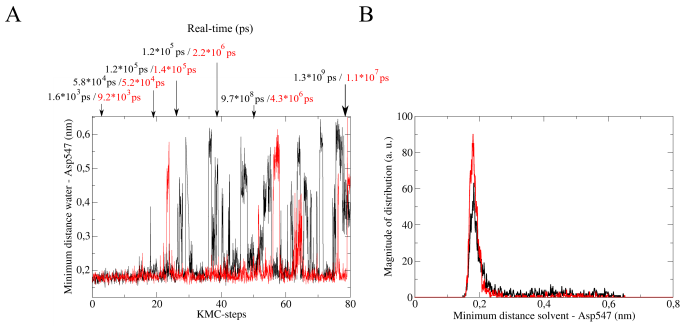


Figure 13.19.: Distribution of water and Asp547 (catalytic cleft). A: Minimum distance between water and Asp547 over 80 KMC-steps - dark-state (black), light-state (red). B: Histogram of minimum distance, shown in (A).

from these curves that the catalytic cleft is active in the light-state, due to its larger solvent accessibility throughout the simulation, compared to the dark-state. We further investigated the activation of the catalytic cleft through performing a biased-cofactor binding KMC simulation on both final structures of the dark and light-states of phot1, which we display in Fig. 13.16 E-G. We deduce from the ATP - Asp547-distance shown in Fig. 13.16G that with equal starting conditions of ATP outside of the kinase in both states, we observed that ATP intrudes into the catalytic cleft in the light-state, while the dark-state structure impedes intrusion.

Finally, we conclude from our simulations that the LOV1-domain controls the sterical interactions in the LOV1-LOV2-linker region, while the LOV2-domain controls the kinase activity. We find that the L1-L2 region acts as a binding partner onto the membrane and mRNA, which is state-dependent. Moreover, the LOV2-domain inhibits the kinase at Ser611, which is located at the A-loop. LOV2-F$\alpha$ as well as LOV2-J$\alpha$ play a deceisive role in the folding behavior of the A-loop of the kinase and alter their location in the light-state, which leads to kinase activation. This latter folding behavior finally leads to an opening of the kinase domain and activation of the catalytic cleft in the light state.

### 13.4.4. Scheme of in vivo function of phototropin1

To summarize, we show in Fig. 13.20 a schematic picture on the *in vivo* function of full-length phot1, inferred from our KMC-MD simulations. We propose that the L1-L2 region is responsible for migration and localization of dark-state phot1 at the negatively charged cell-membrane. Moreover, we demonstrate that the HTH-motif of L1-L2 might act as a motif for nucleotide-binding and nucleotide-control in the light-state. By contrast, in the dark-state the kinase domain is inhibited by the LOV2-domain, while the LOV1-domain dimerizes with the LOV2-domain. Upon illumination, we find that the LOV1-LOV2 distance is increased, while L1-L2 forms a HTH-motif. We point out that activation occurs at the kinase domain through disruption of the J$\alpha$-helix from the LOV2-domain. Subsequently, unfolding of the A-loop extension is initiated through cleavage between the LOV2-F$\alpha$ helix and the kinase. Initial phosphorylation occurs at Ser611, which is identical in sequence with the amino-acid Ser851 of the kinase domain from *Arabidopsis thaliana* [40], which induces large sterical changes. Subsequent auto-phosphorylation in the light-state occurs primarily in the C- and N-terminal Ser residues at the LOV2-domain

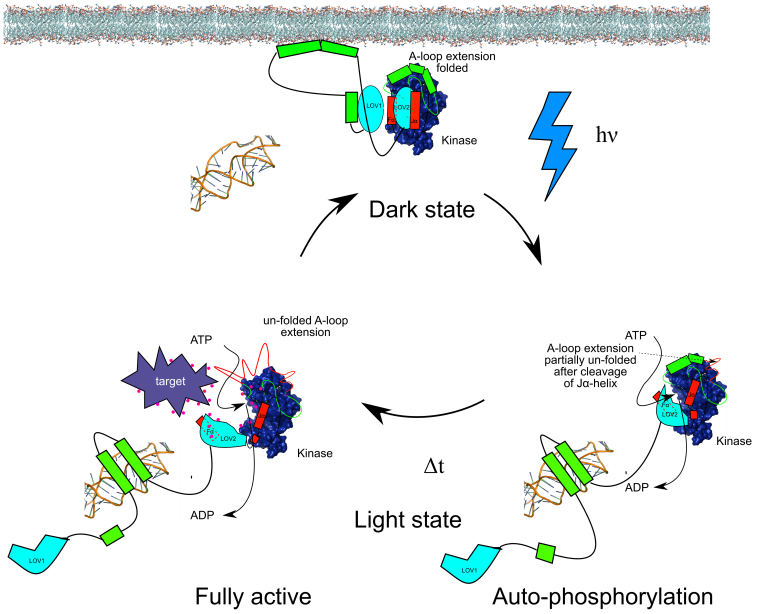


Figure 13.20.: Schematical description of light-induced cycle of phototropin and activation.

in *Arabidopsis thaliana*, as we know from experiments of Sullivan *et al.* [39]. After the initial autophosphorylation, repulsive electrostatic forces lead to a deceisive cleavage of the LOV2-domain from the kinase and target proteins as well as other phototropins within the cell cytoplasm become phosphorylated. Finally we note that the phosphorylation targets, which lead to final response of the organism are unknown. We stress that during phosphorylation of these targets, Crphot1 might bind onto the mRNA-nucleotide, which is regulated by LOV1.

# 14. Molecular dynamics on the Vivid-LOV-domain of Neurospora crassa

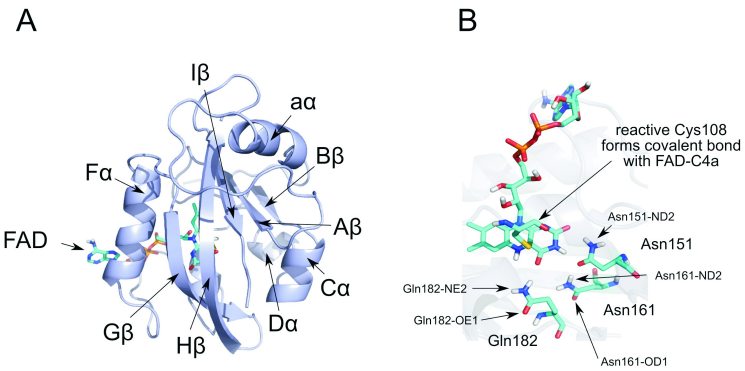


Figure 14.1.: A: Crystal structure of the VVD-LOV-domain from *Neurospora Crassa* with secondary structural elements and FAD-chromophore (MD-starting structure). B: FAD-chromophore in its binding pocket surrounded by relevant amino acids Gln182 (I$\beta$-strand), Asn161 (H$\beta$-strand) and Asn151 (G$\beta$-strand).

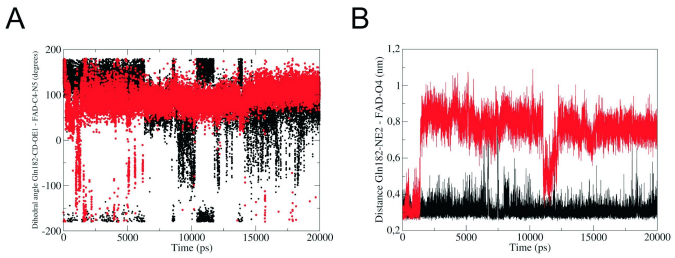


Figure 14.2.: Dihedral angles and inter-atomic distances of amino acids surrounding FAD versus simulation time of dark-state (black) and the light-state (red) [A: dihedral angle Gln182-(CD-OE1) - FAD-(C4-N5). B: inter-atomic distance Gln182-NE2 - FAD-C4=O.].

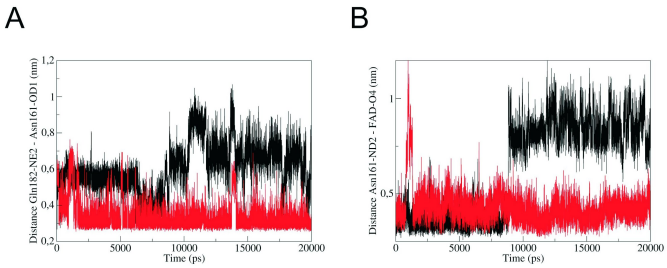


Figure 14.3.: Inter-atomic distances of amino acids surrounding FAD versus simulation time of dark-state (black) and the light-state (red) [A: inter-atomic distance Gln182-NE2 - Asn162-OD1. B: inter-atomic distance Asn161-ND2 - FAD-C4=O].

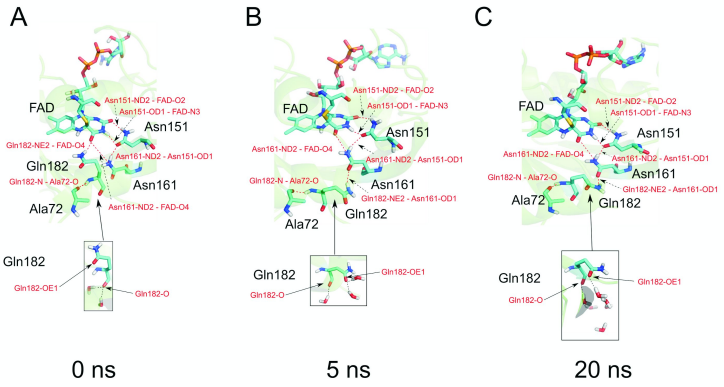


Figure 14.4.: Representative configurations of the amino-acid environment around FAD at different simulation times [A: 0 ns of production phase. B: 5 ns. C: final configuration at 20 ns].

## 14.1. Introduction

In this chapter, we elucidate the pathway of signal-transduction of the Vivid (VVD) LOV-domain. As we described previously, this photosensory protein is essential for adjusting the circadian clock of *Neurospora crassa*. Malzahn *et al.* and Chen *et al.* recently elucidated the *in vivo* functionality of VVD-LOV [63, 62]. Although these works successfully demonstrated the *in vivo* role of VVD-LOV, a molecular based description of the dynamics of signal transduction of VVD-LOV on a molecular level is still lacking. In the follwing sections, we present the results of our simulation on the dark and the light-state of VVD-LOV. In agreement with the experimental results, we will elucidate the pathway of signal transduction and will further discuss the molecular basis of signal-transduction from the vicinity of the FAD-chromophore to the peripheral a$\alpha$-helix.

### 14.1.1. Generation of starting structures

As starting structure for our dark-state simulations for the VVD-LOV-monomer as well as dimer, we used the dark-state crystal structure of VVD-LOV-domain from *Neurospora crassa* (PDB-

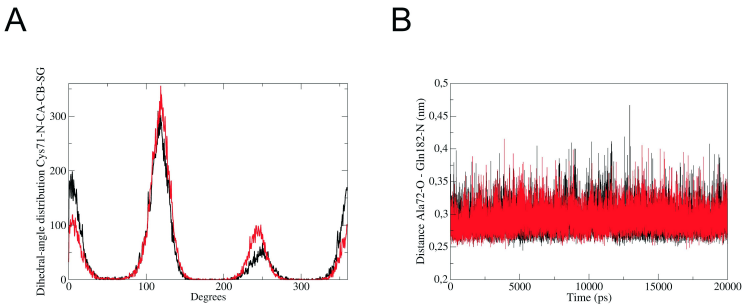


Figure 14.5.: Histogram of characteristic dihedral angle of Cys71 and inter-atomic distances of characteristic amino acids on I$\beta$-strand and a$\alpha$-A$\beta$-linker versus simulation time for the dark-state (black) and light-state (red) [A: histogram of dihedral angle Cys71-(N-CA-CB-SG). B: inter-atomic distance Gln182-N - Ala72-O.].

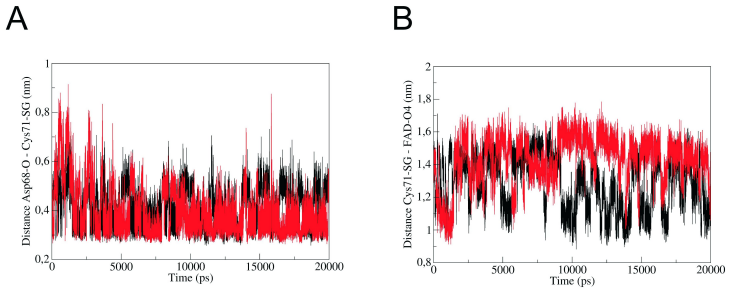


Figure 14.6.: Inter-atomic distances of characteristic amino acids on I$\beta$-strand and a$\alpha$-A$\beta$-linker versus simulation time for the dark-state (black) and light-state (red) [ A: inter-atomic distance Asp68-O - Cys71-SG. B: inter-atomic distance Cys71-SG - FAD-O4].

code: 2PD7), which was determined through x-ray diffraction measurements by Zoltowski *et al.* [65]. Its tertiary structure is visualized in Fig. 14.1A and the relevant amino acids in the vicinity of the chromophore in Fig. 14.1B. To create the initial structure for our simulation of the cysteinyl-FAD (CFD) adduct of the wildtype VVD LOV-domain, we employed the dark-state crystal structure previously mentioned and generated a CFD-adduct by forming a covalent bond between the Cys108-S and FAD-C4a. We point out in this context that the covalent linkage between the reactive cysteine and FAD is required to transmit the stress from the reaction center to the protein and trigger the protein signal, as we have demonstrated through the comparison of the methyl-mercaptan adduct and the cysteinyl-FMN adduct state of the LOV1-domain of *Chlamydomonas reinhardtii* [249] as well as the AsLOV2-J$\alpha$-Rac1 fusion protein [247]. Simulation parameters were chosen, as previously described in section 3.6.

## 14.2. Results

We start the analysis of our simulation results by considering the spatio-dynamical behavior of relevant amino acids adjacent to the FAD-chromophore. In Fig. 14.2A we show the dihedral

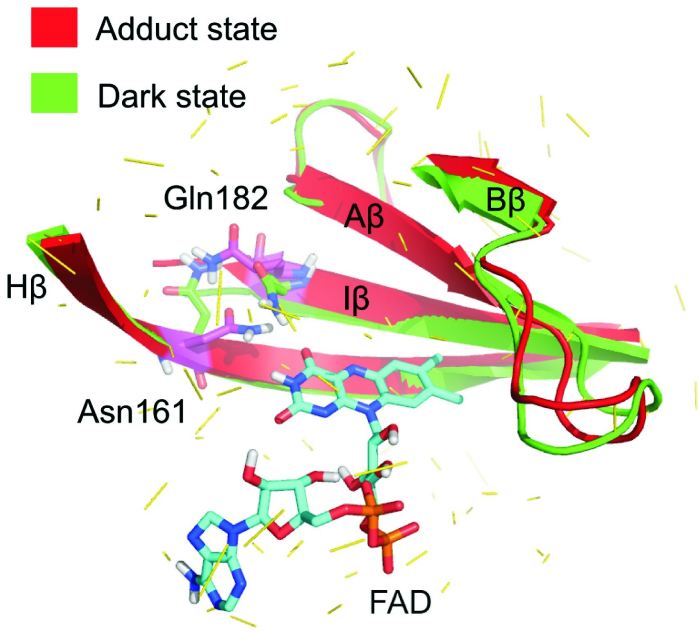


Figure 14.7.: Light-induced conformational changes taking place at H$\beta$-, I$\beta$- and A$\beta$-strands after 20 ns of simulation [green: dark-state. red: light-state].

angle, encompassing the atoms Gln182-(CD-OE1) of the Gln182-sidechain and FAD-(C4-N5) of the FAD-isoalloxazine ring, as a function of simulation time. We observe that in the time-range from 0 up to 2.5 ns the light-state (red) curve is subjected to a large perturbation, which reaches a maximum amplitude of about 240 degrees at around 2 ns and then relaxes back to an average value at around 100 degrees. This switch of the dihedral angle correlates with the jump in the distance between the amino-group Gln182-NE2 and FAD-C4=O from an average value of 2.5 Å up to 8 Å at the same simulation time, as we can easily deduce from Fig. 14.2B. By contrast, we note that the dark-state curve for the distance between FAD-C4=O and Gln182-NE2 fluctuates around a stable H-bonding distance over the whole simulation time. Next, we investigate the effect of the switching of Gln182 on the adjacent asparagine residue Asn161, located in vicinity to the FAD-chromophore. To this end, we consider the distance between the atoms Asn161-OD1 and Gln182-NE2 of the light-state in Fig. 14.3A, which drops to H-bond distance after 2 ns. From this observation, we conclude that, subsequent to the light-state dihedral switch of Gln182 and reorientation of Gln182 with regard to FAD-C4=O shown in the Figs. 14.2A and B, a coupling between Gln182 and its neighboring amino-acid residue Asn161 takes place at around 2 ns through H-bond formation, which persists until the end of simulation. Both amino-acids are respectively located on the I$\beta$- and H$\beta$-strands and, thus, the process results in $\beta$-sheet tightening and weak interlinking of the $\beta$-strands. This contrasts to the dark-state, where we observe no H-bond formation between the sidechains of these two amino-acid residues. To further illustrate the time-dependent changes taking place in the amino-acid environment of FAD after CFD-adduct formation, we visualize in Fig. 14.4 representative snapshots of the light-state configurations at characteristic simulation times. In Fig. 14.4A we first display the configuration of the amino-acid environment of FAD at 0 ns of the production phase. We notice that at this stage of the simulation the carbonyl-oxygen FAD-C4=O is in average twice H-bonded by the amino-acids Gln182

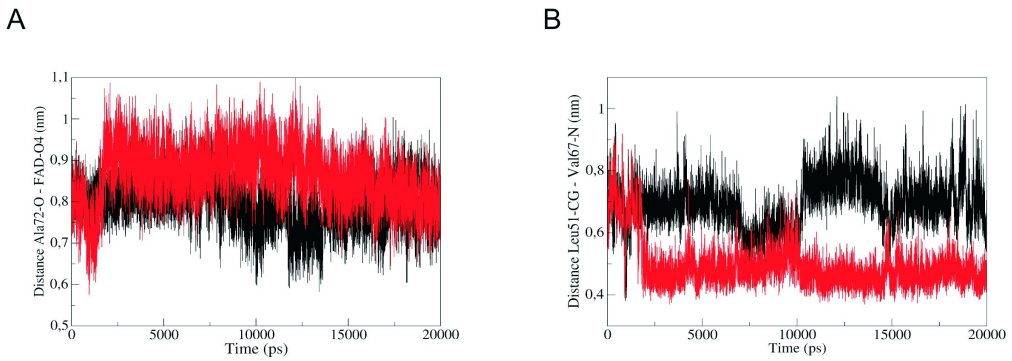


Figure 14.8.: Inter-atomic distances of amino-acid residues located at the N-cap with regard to characteristic residues or FAD at the protein core versus simulation time [A: Ala72-O (A$\beta$-strand)- FAD-O4. B: Leu51-CG (a$\alpha$-helix) - Val67-N (a$\alpha$-A$\beta$-linker).].

and Asn161 and no H-bonding exists between these two amino acids, whereas FAD-C2=O is singly H-bonded by Asn151. Moreover, we deduce from the additional snapshot of the Gln182-environment in Fig. 14.4A that the sidechain-carbonyl group with Gln182-OE1 is not H-bonded at 0 ns, whereas the backbone-carbonyl group with Gln182-O is stabilized through the presence of two H-bonds formed with two water molecules from the solvent. Next, let us examine the amino-acid configuration at 5 ns shown in Fig. 14.4B, which is a representative configuration after which the previously discussed Gln182-switch at 2 ns has taken place. At this stage of the simulation, the H-bond between FAD-C4=O and Gln182 is disrupted and the sidechain of Gln182 has rotated, to form a new H-bond between the atoms Gln182-NE2 and Asn161-OD1 as well as supplementary H-bonds between the sidechain-carbonyl group of Gln182-OE1 with solvent water molecules (see snapshots in Figs. 14.4B and C). This results in a tightening of the $\beta$-sheet through coupling of the amino acids Gln182 and Asn161, respectively located on the I$\beta$- and H$\beta$-strands. Finally, we show in Fig. 14.4C the final configuration at a simulation time of 20 ns. At this stage the FAD-environment is very similar to the one at 5 ns. This observation confirms our analysis of the inter-atomic distances shown in Fig. 14.2 and 14.3, where we found that after the switching process at 2 ns the amino-acid environment remains stable and, therefore, the $\beta$-strands stay coupled up to the end of the simulation. In this context it is worth mentioning that the flipping of the glutamine residue was already suspected in several earlier experimental as well as theoretical studies with other LOV-domains to be a triggering step of their signaling pathway [35, 246, 65, 48, 244]. For example, x-ray diffraction experiments on AsLOV2 by Crosson *et al.* [10] and Halavaty *et al.* [55] showed the importance of the Gln513 near to the FMN-chromophore in the signaling pathway. They demonstrated that illumination of the crystal sample of AsLOV2 causes a rotation of its dihedral angle Gln513-(CA-CB-CD-OE1) of about 180 degrees, which may be followed by a subsequent H-bond formation of the residue with FMN-N5. Nash *et al.* [56] showed the importance of Gln513 in the signaling pathway of AsLOV2 through generating several point mutations and using UV-, CD- and N15-HSQC-NMR-spectroscopy as well as limited proteolysis. In particular, they demonstrated that the singly H-bonded Gln513-AsLOV2-Asn mutant resembles the light-state without illumination of AsLOV2. Their work revealed large differences in the spatio-temporal relaxation of the system between investigations performed in solution at room temperature and the ones in the crystal at cryogenic temperatures. They found

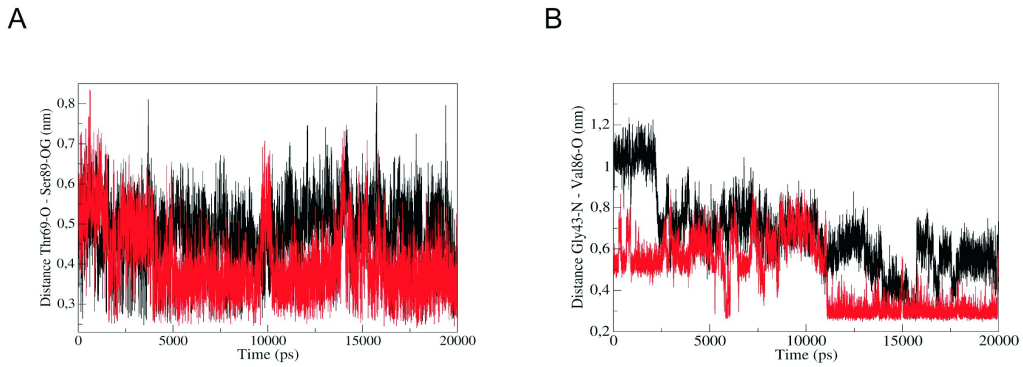


Figure 14.9.: Inter-atomic distances of amino-acid residues located at the N-cap with regard to characteristic residues or FAD at the protein core versus simulation time [A: Thr69-O (a$\alpha$-A$\beta$-linker)- Ser89-OG (B$\beta$-strand). B: Gly43-N (N-terminal coil)- Val86-O (B$\beta$-strand)].

in the latter case that the close packing of the proteins on the crystal lattice as well as the low temperatures allow only small displacements of amino acids after CFD-adduct formation, which favors the formation of a H-bond between Gln513-OE1 and FMN-N5-H as proposed in the work of Crosson *et al.* [10] and Halavaty *et al.* [55]. However, these conclusions contrast to the results obtained in solution by Alexandre *et al.* [46, 197] from FTIR-measurements on AsLOV2. Their study revealed that in solution an equilibrium between different dark-state conformers does exist, which can inter-convert by undergoing significant structural rearrangement. More specifically, they found that the FAD-C4=O-carbonyl-oxygen is twice H-bonded in the so-called crystal-like conformation, whereas in the second conformation the FAD-C4=O is singly H-bonded. Alexandre *et al.* interpreted this latter state as a pseudo-lit state, which agrees well with the singly H-bonded light-state configurations found by us in the time-range after 2 ns as well as our dark-state conformer after 8 ns, in which FAD-C4=O remains only singly H-bonded with Asn161 or Gln182 respectively (see Figs. 14.2B and 14.3B). Moreover, Alexandre *et al.* further concluded from their work [46] that a loop-tightening process according to an amide-I-vibrational signal occurs after adduct formation, which concords with the $\beta$-sheet-tightening process through H-bond formation between Gln182-NE2 and Asn161-OD1 on the I$\beta$- and H$\beta$-strands at 2 ns (see Fig. 14.3A). Finally, the important role of the glutamine residue Gln182 was also confirmed by the x-ray diffraction measurements of Zoltowski *et al.* [65]. They demonstrated through size-exclusion chromatography and x-ray diffraction that the initial Gln182-switch induces changes in the H-bonding pattern around Cys71 at the a$\alpha$-A$\beta$-loop, resulting in an increase of the size of the domain and enhancement of the dimerization tendency. However, since these experiments have been performed with proteins subjected to crystal packing at cryogenic temperatures, it is likely that changes may also implicate other residues of VVD-LOV in solution at room temperature, as discussed in detail in the following.

To investigate this possibility, we study next how the changes in the H-bonding pattern around FAD affects the peripheral $\alpha$-helices and $\beta$-strands, such as the Ncap-region that includes the a$\alpha$-helix and a$\alpha$-A$\beta$-linker domain (residues 37 to 70). To this end, we display in Fig. 14.5A the distribution of the sidechain dihedral angle Cys71-(N-CA-CB-SG), determined over the whole

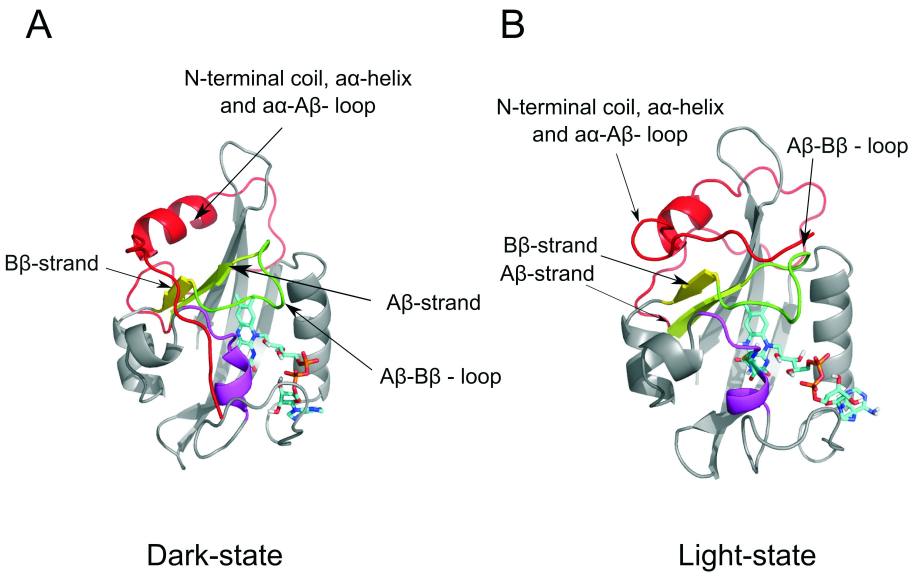


Figure 14.10.: Different perspectives of the final configurations, obtained after 20 ns of simulation [A: dark-state. B: light-state].

simulation trajectory, from the Cys71-residue, located on the a$\alpha$-A$\beta$-linker. We observe an increase in magnitude of the dark-state curve compared to the light-state curve around an angle of 0 degrees, whereas at an angle of 240 degrees this tendency is reversed. From this observation, we conclude that upon adduct formation Cys71 rotates about an angle of 120 (or 240) degrees, which confirms the implication of this residue in the signaling pathway of VVD-LOV as shown by Zoltowski *et al.* [65]. More specifically, these authors demonstrated through substituting Cys71 with Ser and performing x-ray-diffraction measurements with the mutant that Ser71 forms a tighter H-bond with the Asp68-carbonyl than Cys71 in the wildtype VVD-LOV. Moreover, they found that this point mutation completely prevents the N-terminal conformational change and, thus, concluded that Cys71 plays an important role in controlling the binding affinity between the Ncap and VVD-LOV-core. Based on these investigations, they proposed the following two-step signaling mechanism: (1) Gln182 on the I$\beta$-strand rotates to improve interactions between the Gln182-amide and the Ala72-carbonyl group on the a$\alpha$-A$\beta$-linker; (2) the Cys71-thiol group breaks a buried H-bond to the Asp68-carbonyl and swivels to form a new H-bond with the Asp68-amide nitrogen. To study this possible mechanism, we first visualize in Fig. 14.5B the distances between Gln182-N and Ala72-O in the dark and light-states. We see that both curves do not differ and that they remain stable over the whole simulation time. This clearly demonstrates that the disruption of the H-bond between FAD-C4=O and Gln182 as well as subsequent rotation of Gln182 at 2 ns, observed in the Figs. 14.2A and 14.2B, does not affect the interaction between Gln182 and Ala72 and, therefore, the signal propagation does not take place through the pathway of Ala72 and Asp68 in solution. Moreover, we infer from Fig. 14.6A that in the light-state the H-bond between Asp68-O and Cys71-SG is strengthened compared to the dark-state. This observation also contrasts to the second step in the mechanism suggested by Zoltowski *et al.*, in which the Cys71-thiol group changes its H-bond connectivity from Asp68-carbonyl to Asp68-amide nitrogen. However, we deduce from the same figure as well as the Fig. 14.6B, in which we plot the inter-atomic distance between Cys71-SG and FAD-C4=O, that Cys71 is

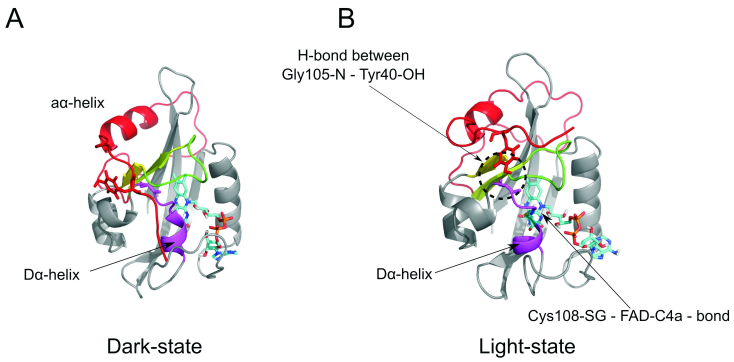


Figure 14.11.: Different perspectives of the final configurations, obtained after 20 ns of simulation [A: dark-state. B: light-state].

perturbed at around 2 ns by the rotation of Gln182 and subsequent coupling between the I$\beta$- and H$\beta$-strands through H-bond formation between Gln182-NE2 and Asn161-OD1. We conclude from this analysis that the signaling pathway of VVD-LOV, triggered by adduct formation and subsequent rotation of Gln182, causes a distortion of the I$\beta$-strand and subsequent coupling between the I$\beta$- and H$\beta$-strands, as depicted in Fig. 14.7. These latter steps then causes the breakage and weakening of H-bonds between the I$\beta$- and A$\beta$-strands, leading to a reorientation of the A$\beta$-B$\beta$-loop with respect to the I$\beta$- and H$\beta$-strands as well as the FAD-chromophore. This is confirmed by considering the increase in the distance after the Gln182-switch at 2 ns between Ala72-O, located on the A$\beta$-strand, and FAD-O4 in Fig. 14.8A. To further track the evolution of the signal to the peripheral secondary structure elements, we next analyze the changes taking place in the Ncap-region between the A$\beta$-strand and the a$\alpha$-helix. This region has been suspected by Zoltowski *et al.* to decisively affect the dimerization tendency of VVD-LOV upon light activation [65, 66]. To this end, we show in Fig. 14.8B the distances between the backbone atom Val67-N, located on the a$\alpha$-A$\beta$-linker, and the sidechain atom Leu51-CG, located on the a$\alpha$-helix. We deduce from the figure that after the Gln182-switch at 2 ns a sharp drop of about 2 Å in the inter-atomic distance between Val67-N and Leu51-CG takes place, which remains stable up to the end of the simulation time. This event describes a contraction process of the a$\alpha$-helix onto the a$\alpha$-A$\beta$-linker, which goes along with the reorientation of the A$\beta$- and B$\beta$-strands. The light state conformation of the VVD-LOV-surface is then stabilized by the formation of new H-bonds between the B$\beta$-strand and a$\alpha$-A$\beta$-linker, as can be concluded from the drop in the distance between Ser89-OG on the B$\beta$-strand and Thr69-O on the a$\alpha$-A$\beta$-linker at around 4 ns in Fig. 14.9A. To further elucidate how the changes at the A$\beta$- and B$\beta$-strands are affecting the N-terminal region as well as the a$\alpha$-helix at the Ncap, we show in Fig. 14.9B the inter-atomic distance between the backbone atoms Gly43-N and Val86-O, located respectively at the coil on the N-terminal end and the A$\beta$-B$\beta$-loop. We observe that at around 11.5 ns a sharp drop of the light-state curve of about 2.5 Å down to H-bond distance takes place, which remain stable up to the end of the simulation time. This indicates that the new position of the a$\alpha$-helix in the light-state is further stabilized through the formation of new H-bonds between the N-terminal end and the A$\beta$-B$\beta$-loop, as demonstrated on the example of the H-bond between Gly43 and Val86. To further illustrate the light-induced conformational changes occurring at the Ncap and the A$\beta$-B$\beta$-loop, we show in the Figs. 14.10A and 14.10B the

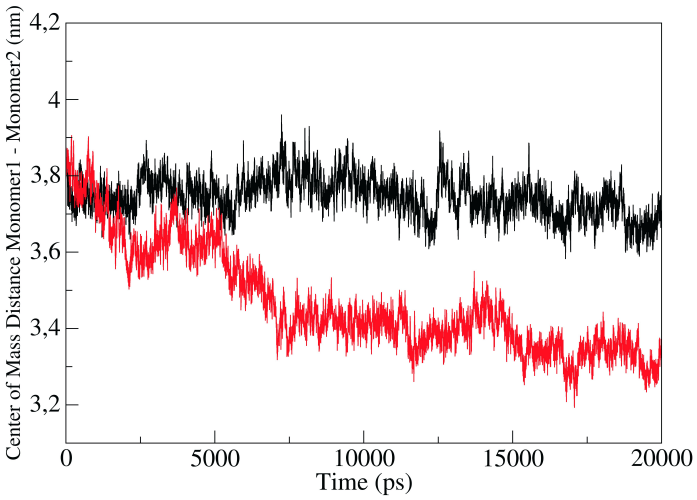


Figure 14.12.: Center of mass distance between monomer A and monomer B of dark- (black) and light-state (red) dimer as a function of simulation time.

final configurations of VVD-LOV on the hydrophilic side in case of the dark and light-state. By comparing both structures, we conclude that in the light-state the Gln182-switch and subsequent coupling between the I$\beta$- and H$\beta$-strands induces a reorientation of the A$\beta$-B$\beta$-loop with respect to the VVD-LOV-core, which goes along with a contraction in the distance between the a$\alpha$-A$\beta$-linker and a$\alpha$-helix as well as an unfolding of the a$\alpha$-helix. Both conformational changes at the VVD-LOV-surface are ultimately stabilized by the formation of new H-bonds between the B$\beta$-strand and the a$\alpha$-A$\beta$-linker as well as between the N-terminal coil and the A$\beta$-B$\beta$-loop. Moreover, we can deduce from the Figs. 14.11A and D that an additional stabilization of the light-state conformation of VVD-LOV takes place through H-bond formation between the residues Gly105 and Tyr40, respectively located on the D$\alpha$-E$\alpha$-loop and the a$\alpha$-helix. It is worth emphasizing in this regard that our conclusions concord well with the x-ray measurements of Zoltowski *et al.* [65], which revealed that the Gln182-switch leads to conformational changes around the Ncap with a shift of about 2 Å of the a$\alpha$-A$\beta$-linker domain toward the LOV-core and a reorientation of the N-terminal end. Moreover, we point out that a tightening of the $\beta$-sheet and an increase in the mobility of the A$\beta$-B$\beta$-loop have also been suspected to be implicated in the signal transduction pathway of other LOV-domains. For example, Iwata *et al.* [45] demonstrated on the LOV2-domain of *Adiantum* Phytochrome3 using FTIR- and UV-visible-spectroscopy that immediately upon adduct formation the system looses the local structure of amino acids around the FMN-chromophore, which causes in subsequent stages a $\beta$-sheet tightening, including the A$\beta$- and B$\beta$-strands, as well as a rearrangement of the $\alpha$-helices on the hydrophilic side of the protein. In a later work Halavaty *et al.* [55] showed through x-ray crystallography on the AsLOV2-J$\alpha$-system that, upon irradiating the dark-state crystal, distance shifts occur in the region around the A$\beta$ and B$\beta$-strands, which strongly influence the J$\alpha$-helix. In a recent theoretical investigation on the signal transduction pathway of the LOV1-domain from *Chlamydomonas reinhardtii* we demonstrated that the adduct formation between Cys57

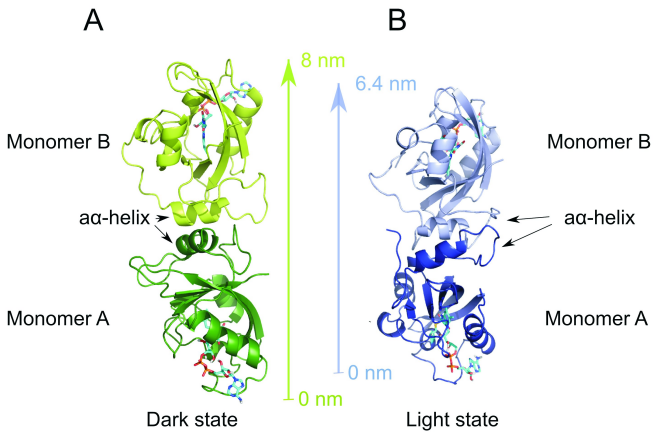


Figure 14.13.: Final dimer structures of VVD-LOV after 20 ns of simulation [dark-state (monomer A: dark green, monomer B: bright green). adduct state (monomer A: dark blue, monomer B: bright blue)].

and FMN-C4a induces a $\beta$-sheet tightening process through coupling of the A$\beta$- and B$\beta$-strands [249]. This process goes along with an increase of the van-der-Waals contacts, which are formed within a dimer on the $\beta$-sheet surfaces of the LOV1-domains allowing a stronger hydrophobic interaction and increase of the dimerization tendency. Similarly as in case of the LOV1-domain, we suspect the conformational changes taking place at the $\beta$-scaffold and surface of VVD-LOV to play a decisive role in controlling its dimerization tendency with partner domains. In Fig. 14.12 we show the center-of-mass distances between both monomers in case of dark and light-state. We observe in case of the light-state a constant decrease in the monomer-monomer center-of-mass distance starting from a value of 3.8 nm at the beginning of the production phase down to a value of 3.3 nm after 20 ns, whereas the dark-state curve remains nearly stable over the whole simulation time. At the end of the simulation the difference between the light-state and dark-state curve reaches a maximum of 4.5 Å . Next, in Fig. 14.13 we show the final configurations, obtained after 20 ns of MD-simulation for the dark and light-states. In case of the light-state dimer we observe that the a$\alpha$-A$\beta$-linker and a$\alpha$-helix of both monomers are more closely aggregated to each other than in case of the dark-state dimer. Moreover, we observe a longitudinal contraction of both monomers of about 1.6 nm in the former case, which indicates that the light-state possesses a higher dimerization tendency than the dark-state. We point out that our conclusions agree well with the x-ray measurements of Zoltowski *et al.* [65, 66], which revealed that in the crystal the VVD-LOV-monomers are dimerizing through the Ncap- as well as N-terminal region and that these regions undergo important conformational changes upon light excitation. Moreover, they showed through size-exclusion chromatography that the dimerization tendency of VVD-LOV strengthens significantly in the light-state. Finally, our findings also confirm several recent experimental studies [61, 62, 63], in which the increased affinity of VVD-LOV to other PAS-domains, such as e.g. WCC, in response to light was evidenced. Therefore, we believe that our findings will also contribute to the elucidation of the interaction and functionality of VVD-LOV with its partner domains *in vivo*.

## 14.3. Conclusions

In summary, we showed through molecular dynamics simulations of the fungal photoreceptor Vivid from *Neurospora crassa* that the primary steps after cysteinyl-adduct formation involve a switch of Gln182 in vicinity of the FAD-chromophore, which results in the coupling between the I$\beta$- and H$\beta$-strands through H-bond formation and subsequent breakage as well as weakening of H-bonds between the I$\beta$- and A$\beta$-strands. This latter process then causes a reorientation of the A$\beta$-B$\beta$-loop with respect to the VVD-LOV-core and a simultaneous contraction of the partially unfolded a$\alpha$-helix onto the a$\alpha$-A$\beta$-linker. Both conformational changes at the VVD-LOV-surface are ultimately stabilized by the formation of new H-bonds between the B$\beta$-strand and the a$\alpha$-A$\beta$-linker as well as between the N-terminal coil and the A$\beta$-B$\beta$-loop. The signaling pathway proposed by us, which relies on simulations performed in solution, contrasts to the mechanism inferred from crystallographic experiments, in which the Gln182-switch triggers the signal from the I$\beta$-strand to the a$\alpha$-A$\beta$-linker via a direct route involving Gln182, Ala72, Cys71 and Asp68. Moreover, we demonstrate through additional dimer simulations that the light-induced conformational changes, observed in the monomeric case, play a decisive role in controlling the dimerization tendency of Vivid with its partner domains. In case of the light-state dimer we observe that both the a$\alpha$-A$\beta$-linker as well as the a$\alpha$-helix of both monomers are aggregating onto the LOV-core, which results in a longitudinal contraction of the dimer structure and indicates that the light-state possesses a higher dimerization tendency than the dark-state. Finally, our findings also provide additional insight about the functioning of VVD-LOV and its modulatory effect on partner domains *in vivo*. Further investigations will be required to study different dimerization modes of VVD-LOV in solution and to enlighten the crosstalk between VVD-LOV and WCC at the molecular level.

# Part IV.

# Application of artificial LOV-based photosensors

# 15. Photoactivable Rac1-GTPase

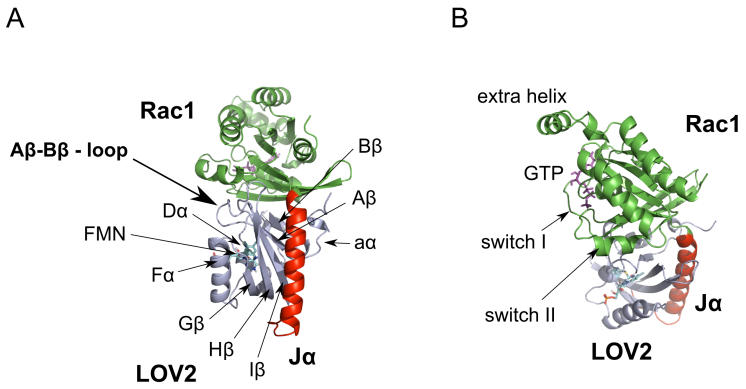


Figure 15.1.: Crystal structure of the AsLOV2-J$\alpha$-Rac1-photoenzyme, obtained from the small-angle-x-ray scattering measurements of Wu *et al.* [72]. Both figures show from different perspectives the AsLOV2-protein core region (grey) with adjacent J$\alpha$-helix (red), which are both aggregated to the Rac1-GTPase (green). Moreover, they include the FMN-chromophore (yellow) and the substrate GTP (magenta).

## 15.1. Introduction

In the following sections we elucidate the early stages of the signal transduction pathway of the AsLOV2-J$\alpha$-Rac1-photoenzyme on the molecular level, using the MD-simulation method. Our study shows that the initial process after formation of the Cys450-FMN-adduct involve the breakage of a H-bond between the carbonyl-oxygen FMN-C4=O and the amino group of Gln513, followed by a rotational re-orientation of its sidechain. This triggering event near the FMN-chromophore initiates the creation of a new H-bond between Gln513 and Asn492, coupling the I$\beta$- and H$\beta$-strands together, and causes a progressive extension of the I$\beta$-strand, transmitting the stress towards the LOV2-J$\alpha$-interface. Afterwards, these movements leads to a partial unwinding of the J$\alpha$-helix at its N-terminal end and subsequent disruption of the J$\alpha$-helix from the LOV2-domain, reflected by the breakage of the two H-bonds Asp522-Glu525 as well as Asp522-Arg526. These latter processes induce the cleavage of two supplementary H-bonds at the LOV2-Rac1-interface, namely Gln507-Asp619 as well as Glu409-Asn595, ultimately causing the release of the Rac1-GTPase from the LOV2-domain. Moreover, we demonstrate that our simulation technique can be used to predict the signaling behavior of novel photoenzymes altered by mutations and in this way provide guidance for experimentalists. The details of the simulation as well as the generation of the starting structures of our dark and adduct state simulations (see Figs. 15.1A and 15.1B) are provided in the method section given below.

### 15.1.1. Generation of starting structures

As a starting structure for our dark-state simulation, we used the dark-state crystal structure of the AsLOV2-Jα-Rac1-photoenzyme from Wu *et al.* (PDB-code: 2wkp.pdb), which was determined by x-ray diffraction measurements [72] and whose secondary structure is visualized in Figs. 15.1A and 15.1B. To create the initial structures of our adduct state simulation, we employed the dark-state crystal structure previously mentioned and generated a cysteinyl-FMN (CFN) adduct by forming a covalent bond between the Cys450-S and FMN-C4a. Similarly, we created a methyl-mercaptan (MM) adduct of its Cys57Gly-mutant by using the same dark-state crystal structure and forming a covalent bond between the MM and FMN-C4a. This latter state solely differs from the CFN-adduct state through the lack of the covalent linkage between the reactive cysteine residue Cys450 and the FMN-chromophore. We point out that covalent linkage between the reactive cysteine and FMN is required to transmit the stress from the reaction center to the protein and trigger the protein signal, as we will demonstrate through the comparison of our simulations with the MM-adduct and CFN-adduct state.

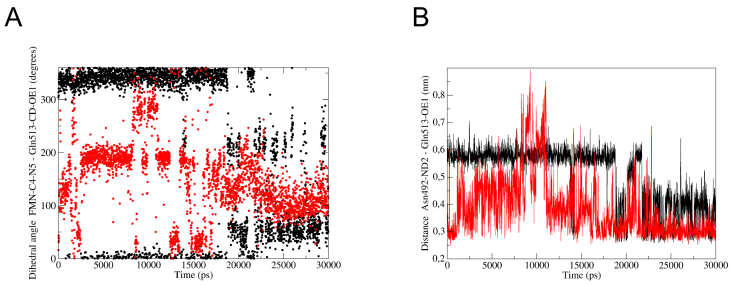


Figure 15.2.: Dihedral angle or distance of characteristic atoms on the FMN-chromophore or on surrounding amino acids to atoms of neighboring amino acid residues in the dark (black) and adduct (red) states [(A) dihedral angle Gln513-(CD-OE1) - FMN-(C4-N5). (B) distance Gln513-OE1 - Asn492-ND2. (C) distance Gln513-NE2 - FMN-O4. (D) distance Asn482-ND2 - FMN-O4].

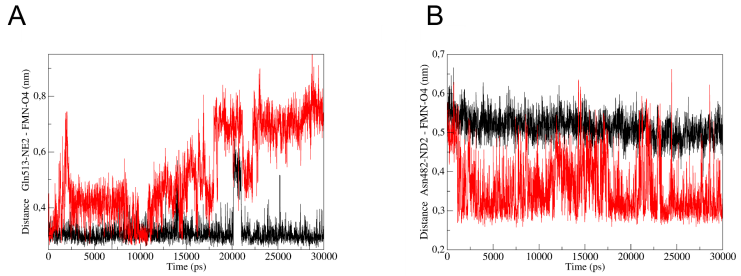


Figure 15.3.: Dihedral angle or distance of characteristic atoms on the FMN-chromophore or on surrounding amino acids to atoms of neighboring amino acid residues in the dark (black) and adduct (red) states [(A) dihedral angle Gln513-(CD-OE1) - FMN-(C4-N5). (B) distance Gln513-OE1 - Asn492-ND2. (C) distance Gln513-NE2 - FMN-O4. (D) distance Asn482-ND2 - FMN-O4].

## 15.2. Results and discussion

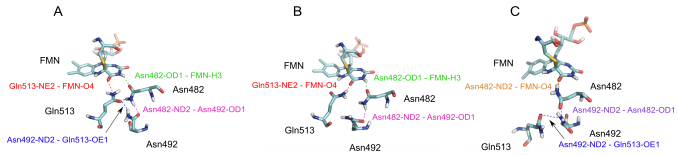


Figure 15.4.: Representative local configurations of amino acids Gln513, Asn492 and Asn482 surrounding the FMN-chromophore at (A) 0 ps, (B) 10000 ps and (C) 30000 ps [cyan: C-atom; red: O-atom; blue: N-atom; white: H-atom; orange: P-atom; yellow: S-atom].

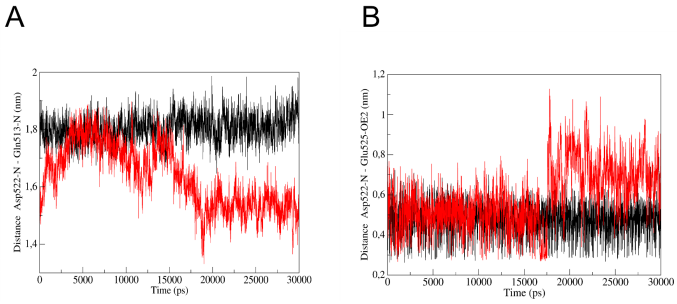


Figure 15.5.: Inter-atomic distance from characteristic amino acids located at the LOV2-J$\alpha$-interface, i.e. at the linker region between the I$\beta$-strand and the J$\alpha$-helix, in the dark (black) and adduct (red) states [(A) Asp522-N - Gln513-N. (B) Asp522-N - Glu525-OE2.].

We start the analysis of our simulation results by considering the spatio-dynamical behavior of relevant amino acids adjacent to the FMN-chromophore. In Fig. 15.2A we show the dihedral angle, enclosing the atoms Gln513-(CD-OE1)-FMN-(C4-N5), as a function of simulation time. We observe that in the CFN-adduct state, starting from an initial angle of 120 degrees up to 2500 ps, the quantity grows stepwise and reaches a maximum angle of 300 degrees at around 10000 ps, until it gradually falls back to an average value of 100 degrees. The rotational re-orientation of the glutamine residue is visualized in Fig. 15.4, in which we compare the representative configurations of the amino acids Gln513, Asn492 and Asn482 surrounding the FMN-chromophore at characteristic simulation times. In case of the configuration at 0 ps (Fig. 15.4A) the FMN-C4=O-carbonyl oxygen is singly H-bonded by the residue Gln513, whereas the Gln513-OE1-carbonyl oxygen on the Gln513-sidechain is twice H-bonded by the amino-groups of the sidechains of Asn492 and Asn482. Comparing this configuration with the configuration at 10000 ps (Fig. 15.4B), we see that the switch of the dihedral angle of about 180 degrees in this time range corresponds to a rotation of the carbonyl-oxygen of Gln513-OE1 about the axis formed by the sidechain atoms Gln513-NE2 and Gln513-CD of the glutamine residue. This process initiates the temporary release of the amino acid Asn492 through breakage of the H-bond between Gln513 and Asn492, as can be deduced by the increase in the distance between

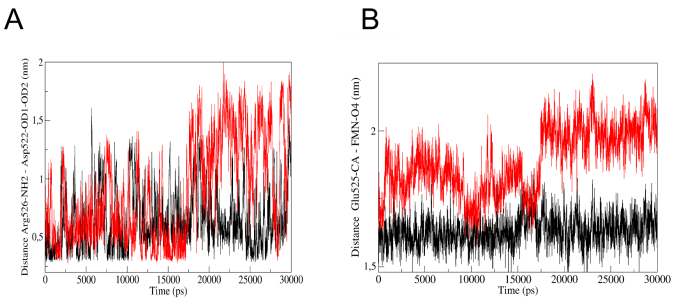


Figure 15.6.: Inter-atomic distance from characteristic amino acids located at the LOV2-J$\alpha$-interface, i.e. at the linker region between the I$\beta$-strand and the J$\alpha$-helix, in the dark (black) and adduct (red) states [(A) Arg526-NH2 - Asp522-OD1(OD2). (B) Glu525-CA - FMN-O4].

Asn492-ND2 and Gln513-OE1 in Fig. 15.2B in the same time range. Moreover, we infer from the distance between Gln513-NE2 and FMN-O4 in Fig. 15.2A and from the configuration at 10000 ps in Fig. 15.4B that during this phase the H-bond between Gln513 and FMN-C4=O-carbonyl oxygen remains stable. From the former graph, we further conclude that the H-bond between Gln513 and FMN-C4=O-carbonyl oxygen is successively released starting from 10000 ps, while a new H-bond between Gln513-OE1 and the amino group of Asn492-ND2 is formed leading to the coupling of the H$\beta$- and I$\beta$-strands. The latter process can be deduced from the drop in the corresponding distance to a value of about 3Å in Fig. 15.2B. We note that the formation of this new H-bond causes a stable coupling between both $\beta$-strands, persisting over the remaining simulation time. This coupling process is consistent with the early stage of a $\beta$-sheet tightening process, inferred from FTIR-experiments on LOV2- [228, 202, 46] and LOV2-J$\alpha$-systems [56]. From the distance between Asn482-ND2 and FMN-O4 in Fig. 15.3B, we deduce that the local configuration of the FMN-chromophore in the binding pocket is stabilized through the creation of a new H-bond between Asn482-ND2 and FMN-O4, leaving the FMN-C4=O singly H-bonded with Asn482. This observation is confirmed by comparing the local configuration of amino acids near to the FMN-chromophore at 30000 ps in Fig. 15.4C. In this context it is worth pointing out that the glutamine residue adjacent to the FMN-chromophore is highly conserved and has previously been suggested to be crucial for the signaling process of the LOV-domains. Several early experimental studies on the isolated LOV2-domain suggested that the primary process after CFN-adduct formation involves the breakage of a H-bond between the glutamine residue and the FMN-O4 [52], possibly followed through rotation of its sidechain by the formation of a new H-bond with the FMN-N5 [10, 11]. In a later crystallographic study on the AsLOV2-J$\alpha$-system Halavaty and Moffat [55] proposed a local re-orientation of the conserved glutamine, associated with a disruption of a H-bond between an asparagine and aspartic acid on the surface of the protein. In a series of spectroscopical investigations with point mutations of the LOV2-J$\alpha$-system [56, 227] as well as the full-length phototropin [35] it was demonstrated that this glutamine residue plays a central role in both spectral tuning and propagating the signal via the I$\beta$-strand from the LOV-core to the J$\alpha$-helix, while its lack attenuates light-induced autophosphorylation of the phototropin. In a recent theoretical work on the isolated AsLOV2-domain Freddolino *et al.* [49] deduced from MD simulations that CFN-adduct formation induces the breakage of a H-bond between the conserved glutamine and FMN-O4, followed by the creation of a new H-bond with FMN-N5. In addition, their simulation results suggested a second conformation, in which the

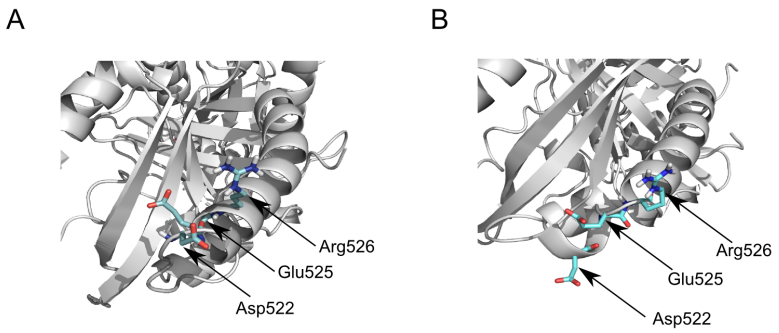


Figure 15.7.: Representative local configurations of amino acids Arg526, Glu525 and Asp522, located at the N-terminal end of the LOV2-J$\alpha$-interface in the dark (A) and (B) adduct states at the final simulation time of 30 ns. In addition, we depict the structural arrangement of the H$\beta$-, I$\beta$-strands and the J$\alpha$-helix with respect to the FMN-chromophore and the surrounding amino acids Gln513 (I$\beta$) and Asn492 (H$\beta$) in the dark (C) and adduct (D) states at 30 ns.

glutamine interacted with neighboring residues on the I$\beta$-strand leading to an increased mobility in the LOV2-J$\alpha$-interface region. However, a clear molecular picture of the signaling process could not be established in this work, due to the neglect of the J$\alpha$-helix. From our simulation results, we conclude that the primary steps after CFN-adduct formation involve the breakage of a H-bond between Gln513-NE2 and FMN-O4, followed by a rotational re-orientation of the glutamine residue Gln513 and subsequent creation of a new H-bond between the amino group of Gln513-OE1 and Asn492-ND2. This latter process causes the coupling and tightening of the I$\beta$ and H$\beta$-strands. Moreover, it permits to propagate the stress to the LOV2-J$\alpha$-interface, which ultimately leads to the detachment of the J$\alpha$-helix from the LOV2-domain, as we will demonstrate in the following. We emphasize that the early steps of the signaling pathway of the AsLOV2-J$\alpha$-Rac1-system show very close analogies to the pathway of the isolated AsLOV2-J$\alpha$-photosensor [246]. The main events, however, are delayed in the former case due to the covalent linkage of the Rac1-GTPase at the C-terminal end of the J$\alpha$-helix.
To analyze the detachment process of the J$\alpha$-helix from the LOV2-domain, we present in Fig. 15.5 and 15.6 data from our dark and CFN-adduct state simulations of characteristic inter-atomic distances from residues, located in the LOV2-J$\alpha$-interface region, as a function of simulation time. The respective residues are depicted in Fig. 15.7A. In Fig. 15.5A we first display the minimum distance between the backbone nitrogen atoms of Asp522 and Gln513, located respectively on the J$\alpha$-helix and the I$\beta$-strand, as a function of simulation time. In case of the CFN-adduct state curve we observe a steady increase of the distance between both residues from the start of the production phase up to a maximum of 1.8 nm. This relates to the propagation of stress along the I$\beta$-strand as a consequence of the rotation of the carbonyl-oxygen of Gln513-OE1 about the axis formed by the sidechain atoms Gln513-NE2 and Gln513-CD of the glutamine residue, depicted in Figs. 15.4A and 15.4B. The distance between Asp522-N and Gln513-N then relaxes back to the original distance of 1.5 nm at around 17500 ps, which coincides with the breakage of the H-bond between Gln513-NE2 and FMN-O4, shown in Fig. 15.3A, as well as the subsequent rotational re-orientation of the glutamine residue Gln513 in the same time range, visualized in Figs. 15.4B and 15.4C. This move leads to an unwinding of the J$\alpha$-helix at its N-terminal end,

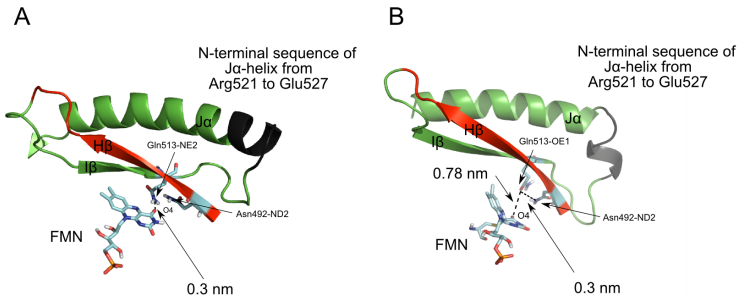


Figure 15.8.: Structural arrangement of the H$\beta$-, I$\beta$-strands and the J$\alpha$-helix with respect to the FMN-chromophore and the surrounding amino acids Gln513 (I$\beta$) and Asn492 (H$\beta$) in the dark (A) and adduct (B) states at 30 ns.

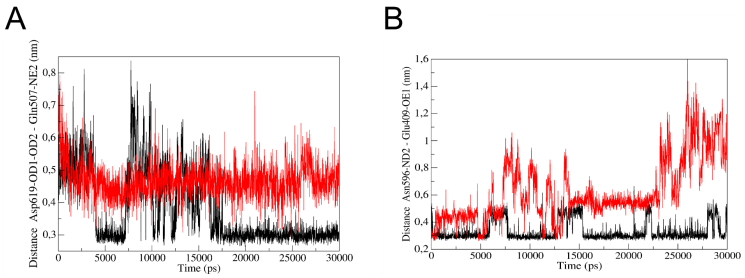


Figure 15.9.: Inter-atomic distances from characteristic amino acids at the interface between LOV2-core and the Rac1-GTPase in the dark (black) and adduct (red) states. [(A) Asp619-OD1(OD2) - Gln507-NE2. (B) Asn595-ND2 - Glu409-OE1.].

leading to the breakage of a H-bond between Asp522-N and Glu525-OE2 as well as another between Asp522-OD1(OD2) and Arg526-NH2, as a result of the drag force. This can be concluded from the abrupt jumps in the corresponding distances at 17500 ps, plotted in Figs. 15.5B and 15.6A. The effect of the breakage of the two H-bonds on the helical conformation of the J$\alpha$-helix is visualized in Fig. 15.7B, where we show the N-terminal end of the J$\alpha$-helix of the CFN-adduct state at the final simulation time of 30000 ps. We see that the helix is unfolded in this residue range, whereas in the dark-state it remains folded, as can be deduced from the final dark-state configuration in Fig. 15.7A. In Figs. 15.8A and 15.8B we show next the corresponding orientations of the amino acids Gln513 on the I$\beta$-strand and Asn492 on the H$\beta$-strand with regard to the J$\alpha$-helix as well as the FMN-chromophore in the dark and CFN-adduct states, respectively. By comparing both figures, we observe a significant distortion of the H$\beta$- and I$\beta$-strands in the N-terminal region of the J$\alpha$-helix in the CFN-adduct state, caused by the rotational reorientation of Gln513 as well as the coupling of the H$\beta$- and I$\beta$-strands through the creation of a new H-bond between Gln513-OE1 and the amino group of Asn492-ND2 (see Fig. 15.2B). In this situation a significant portion of J$\alpha$-helix unwinds at its N-terminal end and starts successively to disrupt from the LOV2-domain, as can be deduced from the sequence of the J$\alpha$-helix (Arg521 up to Glu527) highlighted in black color in Fig. 15.8B. The possibility that the disruption of the J$\alpha$-helix might be the critical event in the signaling pathway has first been proposed by Harper *et al.* in their seminal paper on phototropin switches [53]. To show this, they compared the dark

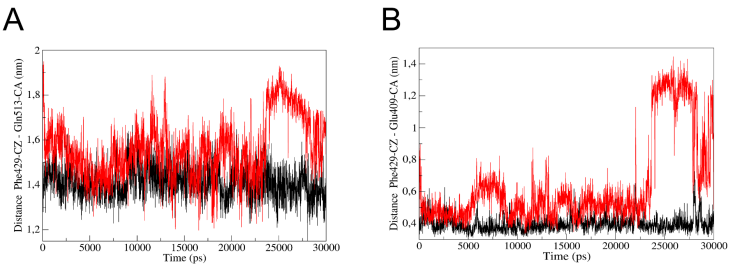


Figure 15.10.: Inter-atomic distances from characteristic amino acids at the interface between LOV2-core and the Rac1-GTPase in the dark (black) and adduct (red) states. [(A) Phe429-CZ - Gln513-CA. (B) Phe429-CZ - Gln409-CA.].

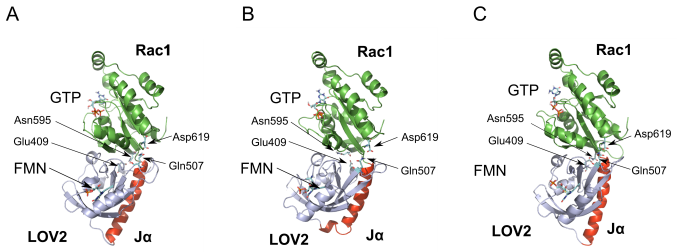


Figure 15.11.: Configurations of the AsLOV2-J$\alpha$-Rac1-system in the dark-state (A), CFN-adduct state (B) as well as MM-adduct state (C), obtained at the final simulation time at 30 ns [AsLOV2-protein core region (grey). J$\alpha$-helix (red). Rac1-GTPase (green)].

and CFN-adduct state 3D-HNCO-spectra of the isolated AsLOV2-J$\alpha$-system, using the minimum chemical-shift-difference method to quantify the spectral changes. They observed within the LOV2-core large differences for the residues close to the FMN-chromophore as well as for the ones on the G$\beta$-, H$\beta$- and I$\beta$-strands. Significant perturbations were also observed at sites on the E$\alpha$-helix and the J$\alpha$-helix over 15 Å away from the FMN-chromophore. They concluded from their work that the structural re-arrangements, induced by the formation of the CFN-adduct, propagate through the LOV-domain, ultimately causing the disruption of the J$\alpha$-helix. To determine which of the protein segments at the LOV2-J$\alpha$-interface could be critical in the detachment process, they further created a truncated protein lacking the last 24 residues of the AsLOV2-J$\alpha$-system (AsLOV2$\Delta$J$\alpha$), starting from residue Glu537 up to the end of the J$\alpha$-helix. This deletion removes a key portion of the C-terminal J$\alpha$-helix, including the H-bond forming residue Asp540. The comparison of the dark-state $^{15}$N/$^{1}$H-HSQC-spectra of the AsLOV2-J$\alpha$- and AsLOV2$\Delta$J$\alpha$-system revealed significant chemical-shift differences for residues located inside the core of the protein domains. Further removal of C-terminal residues, starting from Asp522 and leaving only the LOV2-core (AsLOV2$\Delta$), produced no additional spectral changes in the dark-state spectra. These results indicate that the J$\alpha$-helix possesses a critical interaction with the LOV2-core in the protein part containing Asp540 that is removed by truncation. From a later crystallographic study on the AsLOV2-J$\alpha$-system, Halavaty and Moffat [55] inferred that the formation of the CFN-adduct causes the local re-arrangement of the H-bonding network in the FMN-binding pocket. These re-arrangements are associated with a disruption of the H-bond

Asn414-Asp515 on the surface of the protein and displacement of the N- and C-terminal flanking regions of the LOV2-domain, which constitutes the structural signal. Although all these studies demonstrate that the disruption of the J$\alpha$-helix is critically involved in the signal transduction pathway of the isolated AsLOV2-J$\alpha$-photosensor, our simulation results show that all key molecular processes are significantly delayed or even suppressed when the photosensor is linked to an enzyme. By next comparing the pathway of the AsLOV2-J$\alpha$-Rac1-photoenzyme to the isolated AsLOV2-J$\alpha$-system, we note that the disruption and unwinding of the J$\alpha$-helix in the former case starts with the breakage of two H-bonds at the N-terminal end at a simulation time of 17500 ps. By contrast, the disruption of the J$\alpha$-helix in the second case takes place through an abrupt cleavage of a H-bond at the C-terminal end at around 3000 ps, which goes along with a slow successive cleavage of another H-bond at its N-terminal end in the time range from 2500 ps up to 20000 ps [246].
Next, we show in Figs. 15.9 and 15.10 inter-atomic distances from characteristic residues, located in the interfacial region between the LOV2-core and Rac1, as a function of simulation time. From the dark-state curves in Figs. 15.9A and 15.9B, we deduce that the pair of residues Asp619-OD1(OD2) - Gln507-NE2 as well as Asn595-ND2 - Glu409-OE1 form on average stable H-bonds in the dark, which are subjected to equilibrium fluctuations originating from different local configurations of the FMN-chromophore in the binding pocket [246]. By contrast, from the CFN-adduct state curve in Fig. 15.9A we conclude that the distance between Asp619-OD1(OD2) and Gln507-NE2 is shifted to a larger average value at about 4.7 Å compared to the dark-state curve, which shows that this H-bond is broken in the CFN-adduct state. Moreover, in Fig. 15.9B we observe that in the CFN-adduct state the distance between the amino group of Asn595-ND2 and Glu409-OE1 shows a steady rise up to 23000 ps with a significant perturbation in the range between 7500 ps and 14000 ps. In the latter time span the H-bond between both residues successively breaks due to a slow displacement of the $\beta$-sheet with regard to the J$\alpha$-helix away from Rac1, until it becomes fully disrupted after 14000 ps. This period is followed by an accelerated increase of the quantity after 23000 ps, where the Rac1 completely detaches from the LOV2-core. The latter conclusion can be drawn from Figs. 15.11A and 15.11B, in which we compare the secondary structure of the final configuration at 30 ns of the dark and CFN-adduct states, respectively. We deduce from the figures that the cleavage of the H-bonds Gln507-Asp619 and Glu409-Asn595 at the LOV2-Rac1-interface goes along with the detachment and unwinding of the J$\alpha$-helix at its N-terminal end. To elucidate the pathway of stress propagation from the center of the LOV2-domain to the LOV2-Rac1-interface, we consider next in Fig. 15.10A the distance between Phe429-CZ and Gln513-CA, located respectively on the A$\beta$- and I$\beta$-strands, as well as in Fig. 15.10B the distance between Phe429-CZ and Glu409-CA, located respectively on the A$\beta$-strand and the protein segment at the LOV2-Rac1-interface containing the H-bond forming residue Glu409. We observe in both figures an abrupt jump in the distances at around 23000 ps, which correlates with the accelerated increase of the distance between Asn595-ND2 and Glu409-OE1 in Fig. 15.9B. In this time range the $\beta$-sheet of the LOV2-domain starts to glide away at increased distance from the J$\alpha$-helix, which ultimately leads to the full release of the LOV2-domain from Rac1. We can draw the latter conclusion from the final CFN-adduct state configuration of the LOV2-J$\alpha$-Rac1-system in Fig. 15.11B by considering the position of the $\beta$-sheet of the LOV2-domain with regard to the J$\alpha$-helix and the Rac1. By contrast from the final dark-state configuration in Fig. 15.11A, we deduce that the Rac1 stays attached to the LOV2-core over the whole simulation time, maintained by stable H-bonds at the LOV2-Rac1-interface. This shows that the LOV2-J$\alpha$-Rac1-system in its dark-state form adopts a stable closed conformation, which blocks the binding of the effector protein PAK1. This binding site is released upon CFN-adduct formation. In this context it is worth mentioning that the regulation of enzymatic activity of Rac1 by light-induced control of effector protein binding through the

LOV2-Jα-photosensor has been confirmed in the experimental study of Wu *et al.* [72]. They showed through pull-down assay experiments that the AsLOV2-Jα-Rac1-photoenzyme possesses a greatly reduced affinity for its effector protein PAK1 in the dark, similarly as in case of a AsLOV2-Jα-Rac1-protein fusion containing the light-insensitive LOV2-Cys450Ala-mutant. By contrast, effector protein binding was restored in the AsLOV2-Jα-Rac1-construct containing the LOV2-Ile539Glu-mutant, which mimics the unfolded lit state. They deduced from these experiments that the LOV2-domain in its closed conformation blocks the binding of the effector protein to Rac1, and that light-induced detachment of the Jα-helix from the LOV2-core releases steric inhibition, leading to Rac1-activation. In order to show that the binding affinity of the PAK1-effector domain to the Rac1-enzyme can decisively be controlled through mutation of specific amino acids and that our simulation technique can be used to predict the signaling behavior of such photoenzymes, we present in Fig. 15.11C the final structure after 30 ns of simulation of the MM-adduct state of the Cys450Gly-mutant in comparison to the dark-state as well as CFN-adduct state in Figs. 15.11A and 15.11B respectively. The MM-adduct state solely differs from the CFN-adduct state through the lack of the covalent linkage between the reactive cysteine residue Cys450 and the FMN-chromophore. It can be created experimentally through mutation by replacing Cys450 by glycine, and generating an adduct through addition of methyl-mercaptan and irradiation with blue light [2]. From the figure, we infer that the Jα-helix of the MM-adduct remains folded and the entire system in its closed conformation similar as in case of the dark-state. This demonstrates that through suppression of the covalent linkage between Cys450 and the FMN the signal of the Rac1-photoenzyme can be inhibited, which concords well with the observations made by Wu *et al.* for the photoinactive LOV2-Cys450Ala-mutant. In this context, it is also worth mentioning that similar structural changes upon light-activation have been observed in case of phototropin kinases. Through *in vitro* phosphorylation experiments with phototropin 2 from *Arabidopsis*, Matsuoka *et al.* [218] proposed first that the LOV2-domain acts as an inhibitor to the kinase activity. From subsequent docking studies on the same system, Tokutomi *et al.* [212] concluded that the LOV2-domain inhibits the kinase by binding to the catalytic cleft between the C- and N-terminal lobes of the kinase. In a very recent work Pfeifer *et al.* [165] deduced from their FTIR-measurements on the full-length phototropin from *Chlamydomonas reinhardtii* that the changes in the turn structure reflect the light-induced opening of the kinase by release of the inhibitory LOV-domain. Our study complements these experimental works by resolving the early stages of the activation mechanism of the AsLOV2-Jα-based photoenzyme AsLOV2-Jα-Rac1 on a molecular level and provides guidance for the creation of novel protein fusions based on this photosensor. Besides the light-induced regulation of enzymatic reactivity, these systems show great potential for the spatio-temporal control of gene function and gene protection [169, 170], as has recently been impressively demonstrated in the pioneering work of Strickland *et al.* [81, 225]. These authors connected the AsLOV2-Jα-photosensor with the *Escheria coli* trp-repressor (TrpR) protein, which enabled light-induced coupling of operator DNA and prevented it from nuclease digestion.

## 15.3. Conclusions

In the previous sections we have demonstrated that the signal transduction pathway after adduct formation of the photoenzyme AsLOV2-Jα-Rac1 begins with the alteration of the H-bond network through a rotational re-orientation of the glutamine Gln513 adjacent to the FMN-chromophore, inducing the coupling of the Iβ- and Hβ-strands as well as the propagation of the stress via the Iβ-strand to the LOV2-Jα-interface. This then leads to a partial unwinding of the Jα-helix at its N-terminal end and to the subsequent disruption of the Jα-helix from the LOV2-

domain, reflected by the breakage of the two H-bonds Asp522-Glu525 as well as Asp522-Arg526. These latter processes induce the cleavage of two supplementary H-bonds at the LOV2-Rac1-interface, namely Gln507-Asp619 as well as Glu409-Asn595, ultimately causing the release of the Rac1-GTPase from the LOV2-domain. In addition, our work shows that in the dark-state the AsLOV2-J$\alpha$-Rac1-system adopts a closed conformation, in which the LOV2-domain blocks the binding site of the effector protein PAK1 on Rac1. By contrast in the adduct state, the unwinding of the J$\alpha$-helix and its detachment from the LOV2-domain releases steric inhibition, enabling Rac1-activation through effector protein binding. Furthermore, the signaling of the Rac1-photoenzyme can be inhibited through suppression of the covalent linkage between Cys450 and the FMN, which demonstrates that our simulation technique possesses significant predictive power to assess the influence of specific point mutations on the signaling pathway of this photoenzyme and has the potential to provide guidance for experimentalists for engineering novel artificial protein constructs. This will open new routes for investigating light-induced regulation of enzymatic reactivity, allowing the modulation of cell behavior and cell metabolism [169], as well as gene function, enabling the control of transcription processes [170].

# 16. MD-simulation on light-activable Trp-Repressor LOV-TAP

## 16.1. Introduction

In the following sections, we will present the results of our study on the light-activable DNA-switch LOV-TAP. Sosnick *et al.* generated this DNA-photoswitch through fusion of the AsLOV2-J$\alpha$ photosensor with TrpR of *E. coli* [81]. They found through SAXS-, biochemical transformations and structural modelling, that LOV-TAP is in a light-controlled equilibrium with DNA. This equilibrium is shifted to a DNA-bound state in the light-state, while LOV-TAP dissociates from the DNA in the dark. We modelled the DNA-photoswitch LOV-TAP and found that the signal of AsLOV2-J$\alpha$ is triggered through coupling between Gln513 and Asn492, as we have also described in the previous examples of AsLOV2-J$\alpha$ regulated systems in the chapters 4 and 12. Subsequently, N-terminal disruption occurs at the J$\alpha$-helix of this system, which is caused by a coupling between the H$\beta$- and I$\beta$-strands. This latter process releases the tension along the so-called hairpin-region coupling the C-terminal J$\alpha$-helix and the Trp-repressor protein, where both proteins are fused together. Finally, this leads to an aggregation of the whole LOV-TAP system onto the DNA in the light-state. By contrast, this is not the case for the dark-state, where the LOV-TAP remains folded at the hairpin-region and dissociates from the DNA. Finally, we conclude that the tension at the previously mentioned hairpin-region leads to a dissociative movement of LOV-TAP from the DNA in the dark form, while in the light-state this hairpin-region unfolds and LOV-TAP aggregates onto the DNA, which is in agreement with the experiments of Sosnick *et al.* [81].

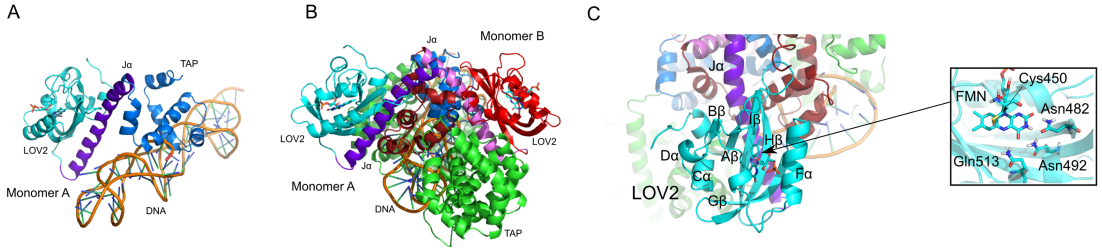


Figure 16.1.: A: Starting structure of LOV-TAP monomer, attached onto DNA. B: Overall structure of LOV-TAP dimer attached onto DNA and wild-type TrpR (green). C: AsLOV2-J$\alpha$-system with secondary structure elements and amino-acid environment of FMN-chromophore.

### 16.1.1. Model

We prepared the model of LOV-TAP for our simulations as follows. As starting structures for the LOV-TAP, we took the structures of the *E.coli* Tryptophan-repressor (TrpR) (pdb-code: 1TRR)

[78] and AsLOV2-Jα (pdb-code: 2V0U) [55] determined by X-ray diffraction measurements, as shown in Fig. 16.1. Subsequently, we generated the model of the LOV-TAP dimer through connection of the AsLOV2-Jα-system at Ala543 with the TrpR-structure at Phe22 at the chains D and E of TrpR [81] (see Fig. 16.1). Hereupon, the system was minimized using the l-BFGS minimizer with the GROMOS96-43a1 forcefield to describe the interactions. Then, we centered the minimized structure in a cubic box with box-length 14.9 nm and solvated the protein-DNA-complex with 111324 SPC-water molecules. Additionally, we neutralized the system by adding 9 sodium ions. Finally, we relaxed the modelled system in a short MD-run of 0.5 ps with a timestep of 0.5 fs. During this relaxation-phase, a pressure of 1 atm and a temperature of 100 K were maintained using the berendsen-thermostat and -barostat. We validated our model through measurement of the radial-distribution-function (RDF) and the radius of gyration of the dark-state form. Then, we compared the resulting data with the experiments of Strickland *et al.* [81], as shown in Fig. 16.2. We observe that our LOV-TAP model has a maximum at 3.1 nm in the DNA-protein RDF, while the experimental RDF has a peak at 3.0 nm. Moreover, the radius of gyration of our model is at an average value of 2.75 nm and the experimentally determined radius of gyration is at 2.9 nm (see Fig. 16.2). From these values we conclude, that our model fits well to the experimental values of Strickland *et al.*. In the following, we present the results of our 20 ns MD-simulation of the LOV-TAP DNA-model in the dark- and light-state. For our simulations we employed the parameters given in section 3.6.

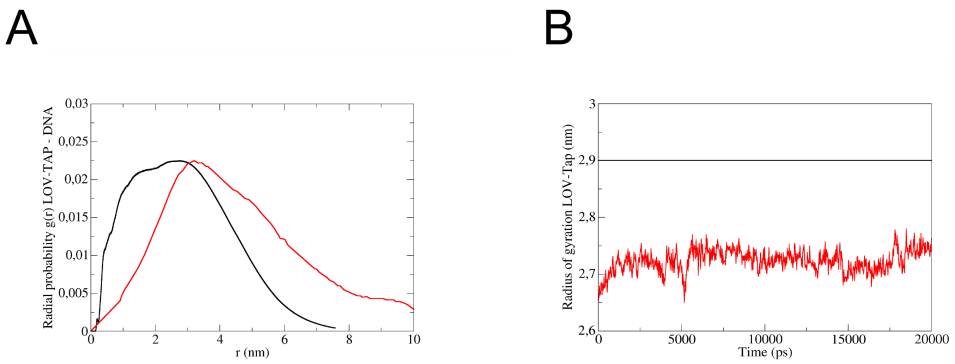


Figure 16.2.: (A) Radial distribution between DNA and LOV-TAP, as well as (B) radius of gyration of LOV-TAP. [ Experiment (black). LOV-TAP dark-state (red) ]

## 16.2. Results and discussion

We start with the analysis of the amino-acid environment in vicinity of the FMN-chromophore, which consists of three amino-acids Gln513, Asn492 and Asn482. In Fig. 16.3 we display the time-dependent inter-atomic distances and dihedral angles as well as the final configurations of the amino-acids nearby the FMN-chromophore. In Fig. 16.3A, we show the dihedral angle between atoms Gln513-CA-CB - CD-OE1 of the Gln513-sidechain in the light-state as a function of simulation time. In case of this angle a jump occurs from -100 degrees up to 100 degrees in

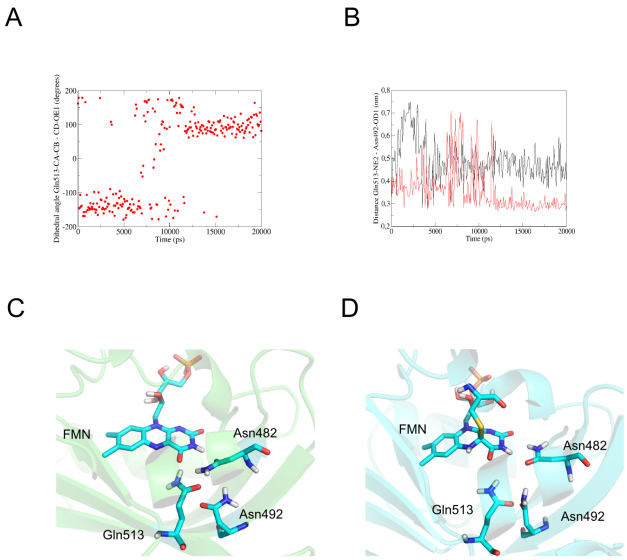


Figure 16.3.: Dihedral angles and inter-atomic distances in vicinity of FMN-chromophore as a function of simulation time. A: Gln513-CA-CB - CD-OE1 dihedral angle (light-state). B: Gln513-NE2 - Asn492-OD1 inter-atomic distance (red: light-state, black: dark-state). C: Final configuration in vicinity of FMN-chromophore in dark-state. D: Final configuration in vicinity of FMN-chromophore in light-state.

a time-range of the simulation from 7.5 ns to 12 ns. We point out that we observe an H-bond formation process between the atoms Gln513-OE1 and Asn492-ND2 at the same simulation time of 12 ns, as shown in Fig. 16.3B. In the same figure we see that in contrast to the light-state no H-bond formation occurs between these atoms in the dark-state. Several experimental studies have highlighted the essential role of Gln513 in the signaling of phototropin1-systems. Christie *et al.* showed through site-directed mutagenesis, fluorescence spectroscopy and western blot analysis, that Gln513 is essential for signaling of the LOV2-domain in full-length phototropin1 from *Arabidopsis* [35]. Halavaty *et al.* and Crosson *et al.* found through X-ray diffraction experiments on the isolated AsLOV2-J$\alpha$-domain, that Gln513-switching plays an important role in the transduction of the signal onto the peripheral J$\alpha$-helix in the light state [10, 55]. Alexandre *et al.* performed FTIR-spectroscopy on the dark- and light-state of the AsLOV2-J$\alpha$ system and found that a doubly as well as a singly bonded state of the FMN-C4=O carbonyl oxygen exists [46]. They attributed the singly bonded state to a cleavage-process between the atoms Gln513-NE2 and FMN-C4=O, which can be interpreted as a dihedral-switch of Gln513. Moreover, we demonstrated through our previous simulation study on AsLOV2-J$\alpha$, that coupling between Gln513 and Asn492 is essential for the disruption of the J$\alpha$-helix in the light-state [246], as we also have described in chapter 12. In a further simulation study on the AsLOV2-J$\alpha$ fusion-protein PA-Rac1, we observed the high affinity for coupling between Gln513 and Asn492 in the light-state, which triggers disruption of the J$\alpha$-helix [245] (see chapter 15). In Figs. 16.3C and 16.3D we illustrate the configurations in vicinity of the FMN-chromophore in the dark- (green) and light-state (blue). We observe, that in the dark-state no coupling occurs between Gln513 and Asn492, whereas in the light-state an H-bond is formed between Gln513 and Asn492 at 12.5

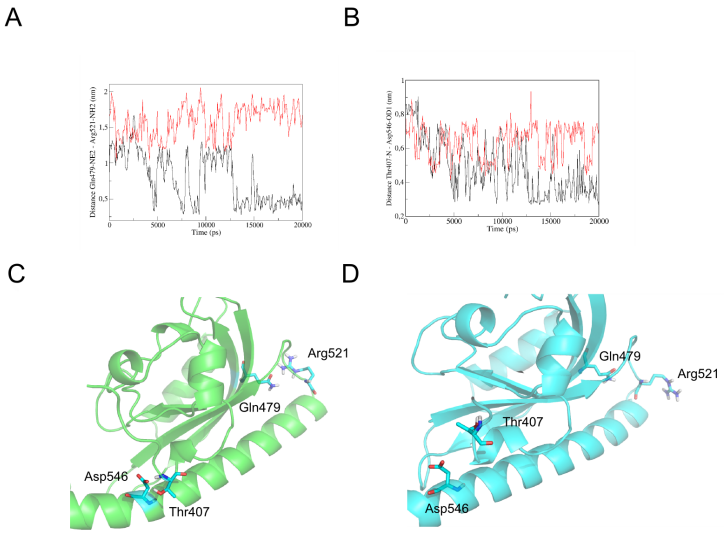


Figure 16.4.: Inter-atomic distances at AsLOV2 - J$\alpha$-interface as a function of simulation time. A: Gln479-NE2 - Arg521-NH2 inter-atomic distance vs. simulation time (red: light-state, black: dark-state). B: Thr407-N - Asp546-OD1 inter-atomic distance vs. simulation time (red: light-state, black: dark-state). C: AsLOV2-J$\alpha$ configuration of dark-state. D: AsLOV2-J$\alpha$ configuration of light-state.

ns, which is stable until the final stage of simulation. As we described previously, this coupling starts with a side-chain rotation of Gln513, which has also been observed in the experiments of Crosson *et al.* and Halavaty *et al.*. We note, that Gln513 is located at the I$\beta$-strand, while Asn492 is located at the H$\beta$-strand. We infer from the Gln513-Asn492 coupling process as well as our previous simulation results and experiments of Iwata *et al.*, that coupling between Gln513 and Asn492 leads to tightening between the H$\beta$- and I$\beta$-strands, as well as changes in the N-terminal A$\beta$-B$\beta$-strand region [45, 246, 245, 247]. We further illustrate the effect of this tightening process on the AsLOV2-J$\alpha$ interface in Figs. 16.4A and 16.4B, where we show the inter-atomic distances and configurations of the amino-acids located at the N- and C-terminal parts of the J$\alpha$-helix. In Fig. 16.4A, we visualize the inter-atomic distance between the atoms Gln479-NE2 and Arg521-NH2, which are located at the G$\beta$-strand of the AsLOV2-domain and the N-terminal part of the J$\alpha$-helix. We observe, that both amino-acids form an H-bond in the dark-state at the same time when Gln513-NE2 - Asn492-OD1 reach a stable fluctuation-range at 4900 ps, while in the light-state both amino-acids do not undergo H-bond formation. We deduce from this fact that $\beta$-sheet tightening between the H$\beta$- and I$\beta$-strands is crucial for disruption of H-bonds in the N-terminal part of the J$\alpha$-helix in the light-state. In the dark state, where no coupling occurs between the H$\beta$- and I$\beta$-strands, the N-terminal region of the J$\alpha$-helix is stabilized. Additionally, we measured the inter-atomic distances between the atoms Asp546-OD1 and Thr407-N, which are located at the N-terminal a$\alpha$-helix of the AsLOV2-domain and the C-terminal region of the J$\alpha$-helix, as shown in Fig. 16.4B. In case of this distance we observe, that these amino-acids are cleaved in the light state, whereas in the dark-state H-bond formation occurs at a simulation time of 4.9 ns (see Fig. 16.4B). We note that this this H-bond

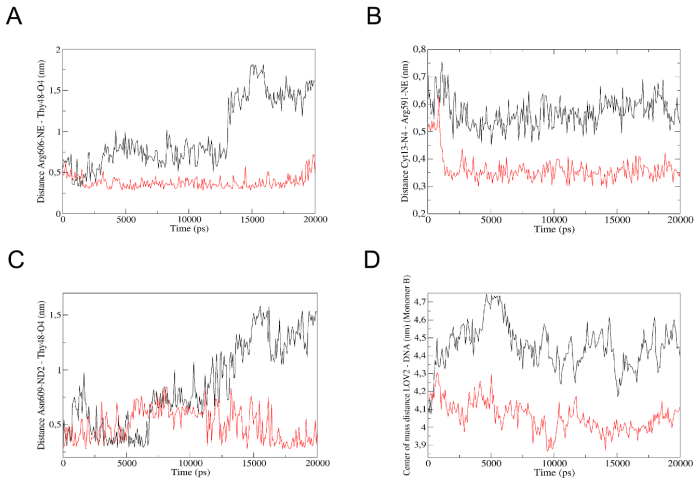


Figure 16.5.: Inter-atomic as well as center-of-mass distances at LOV-TAP - DNA-interface as a function of simulation time. A: Arg606-NE - Thy48-O4 inter-atomic distance (nm) (red: light-state, black: dark-state). B: Cyt13-N4 - Arg591-NE inter-atomic distance (nm) (red: light-state, black: dark-state). C: Asn609-ND2 - Thy48-O4 inter-atomic distance (nm) (red: light-state, black: dark-state). D: LOV2-DNA center-of-mass distance vs. simulation time (red: light-state, black: dark-state).

between Asp546-OD1 and Thr407-N becomes stablized at the same time, when the amino-acids Gln479 and Arg521 in the N-terminal region of the J$\alpha$-helix form an H-bond. We point out that we have observed this behavior in previous studies on PA-Rac1 as well as on the isolated AsLOV2-J$\alpha$-system [247, 246]. Harper *et al.* demonstrated through 2D-solution HSQC-NMR, limited proteolysis and CD-spectroscopy, that the signal of AsLOV2-J$\alpha$ is triggered through disruption of the J$\alpha$-helix in the light-state [53]. We stress that we observe the same behavior in our simulation of LOV-TAP in the light-state. However, we note that the signaling pathway in the LOV-TAP system might differ in the late stages of the disruption from the signaling pathway of the isolated AsLOV2-J$\alpha$ domain, due to presence of TrpR and the DNA in LOV-TAP. Next, we show representative configurations of the AsLOV2-J$\alpha$ interface in the final stage of simulation for both states in Figs. 16.4C and 16.4D, confirming the cleavage behavior in the light-state and a stabilized J$\alpha$-helix in the dark-state. In these figures we observe that Asp546 and Thr407 (C-terminal J$\alpha$-helix) as well as Gln479 and Arg521 (N-terminal J$\alpha$-helix) are H-bonded in the dark-state, while these amino-acids are cleaved in the light-state. We point out that a similar sequence of events occurs at the FMN-chromophore and at the LOV2-J$\alpha$ interface in the second monomer, as we show in Fig. 16.11. We continue with the analysis of the effects of the triggering events in the light-state onto the TrpR-DNA interface, which we illustrate in Figs. 16.5 and 16.6. We observe that the distance between atoms Arg606-NE (TrpR) and the DNA-base Thy48-O4 is stable in case of the light-state, whereas the same amino-acid cleaves from the DNA in the dark-state, as shown in Fig. 16.5A. In case of this distance we see that the dark-state curve (black) performs two up-shifts at two different simulation times at 2500 ps and 12500 ps, where the TrpR-DNA-distance is up-shifted by 0.3 nm and 0.7 nm. We note that the N-terminal part of the J$\alpha$-helix is stabilized at the same time in the dark-state, as the TrpR-DNA interface becomes disrupted at 12500 ps. Moreover, we observe an increase in the distance between the atoms Asn609-ND2 - Thy48-O4 from 0.7 nm up to 1.1 nm at a simulation

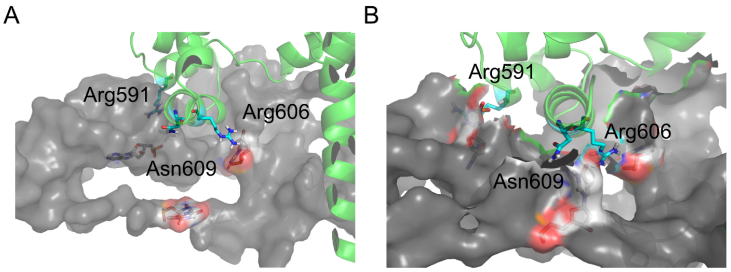


Figure 16.6.: Final configurations at LOV-TAP DNA-interface. A: dark-state. B: light-state.

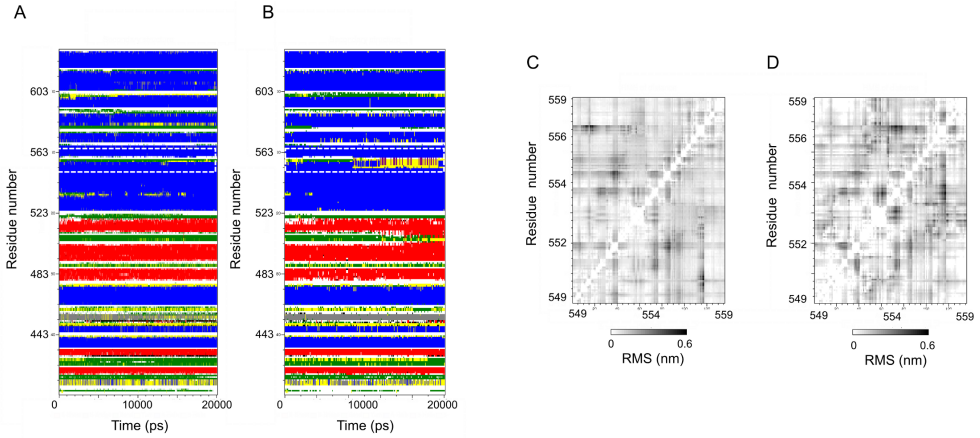


Figure 16.7.: A: Secondary structure analysis vs. simulation time of dark-state. B: Secondary structure analysis vs. simulation time of light-state. C: RMS-distance matrix of dark-state. D: RMS-distance matrix of light-state.

time of 12500 ps in the dark state, as shown in Fig. 16.5C. From the distance between Cyt13-N4 and Arg591-NE, we infer that both residues form an H-bond in the light state at a simulation time of 3 ns, while the dark state remains at a distance of 0.55 nm (see Fig. 16.5B). In Fig. 16.5D, we display the center-of-mass (COM) distance between the AsLOV2-domain and DNA for dark- and light-state. This distance shows an initial increase from 4.1 nm up to 4.5 nm at a simulation time of 2 ns in the dark state. Then, it undergoes a further shift from 4.5 nm up to 4.7 nm at a simulation time of 4.9 ns in the dark-state, when the C- and the N-terminal end of the J$\alpha$-helix becomes stabilized. By contrast, the light-state remains at an average LOV2-DNA COM-distance of 3.9 nm over the whole simulation time of 20 ns. We conclude from the behavior of both states at the DNA-TrpR interface, that the triggering events of the light state lead to a tight binding between TrpR and DNA. By contrast, the stabilization of the J$\alpha$-helix in the dark-state leads to the disruption at the TrpR-DNA interface. We continue with the illustration of the changes at the DNA-TrpR interface by analyzing the final amino-acid configurations in Fig. 16.6A (dark-state) and Fig. 16.6 B (light-state). In these figures we observe that the DNA-protein interface in the dark-state is disrupted, while the same interface is conserved in the light-state. We note that the second monomer has the same behavior at the TrpR-DNA interface, confirmed by the inter-atomic distances between DNA and TrpR in case of the second monomer in Figure 16.11. Next, we show the secondary structure analysis of LOV-TAP in Figs.

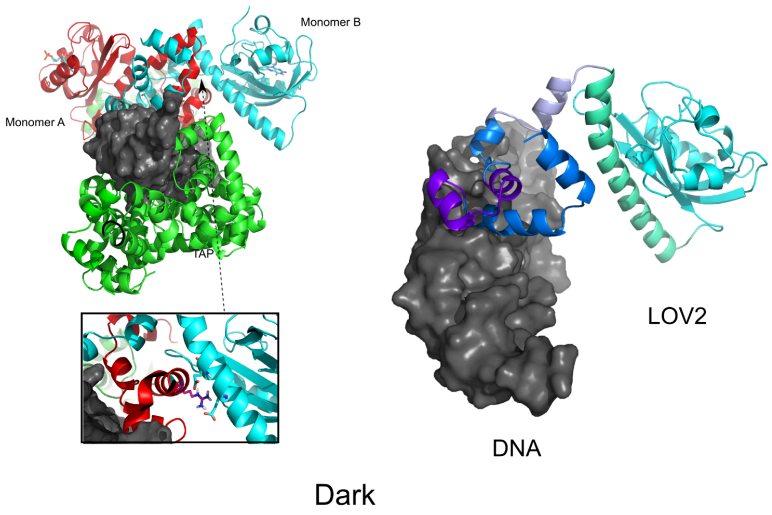


Figure 16.8.: Final structure of LOV-TAP - DNA complex of dark-state.

16.7A (dark-state) and 16.7B (light-state). In these secondary-structure analyses we focussed on the residue-regions between the C-terminal end of the J$\alpha$-helix and the N-terminal TrpR-protein in the residue-range from 549 to 559, i.e. the so-called hairpin-region. We observe that in this region an unfolding-process occurs in the light-state at a simulation time of 12 ns. We note that at the same simulation time coupling occurs between residues Asn492 and Gln513 in the light state, as shown previously in Fig. 16.3B. We conclude that this initial mechanism, which leads to a sheet tightening between the H$\beta$- and I$\beta$-strands, leads to unfolding in the hairpin region at 12 ns. By contrast the hairpin-region remains folded in the dark-state, as shown in Fig. 16.7A. From this light-dependent folding behavior we infer, that the folding-state of the hairpin region has an imminent influence on the amino-acids at the TrpR-DNA interface. We further analyzed the hairpin-region through measurement of the root-mean-square deviation (RMS) as a function of the residue-numbers (see Figs. 16.7C and 16.7D). We observe that the RMS-matrix of the light-state has huge RMS-values of 0.6 nm in the residue-ranges 552-554 and 555-556 (see Fig. 16.7D). By contrast, we see that the dark-state has overall lower RMS-values in these regions, as we see in Fig. 16.7C. We conclude from this RMS-analysis, that the hairpin-region has an overall larger mobility in the light-state, due to the unfolding-process in the residue-range from 549 to 559. Moreover, we find that the increased mobility in the light-state enables TrpR to bind onto the DNA. The mechanism of TrpR-binding in *E. coli* has been experimentally investigated by several authors. Lawson *et al.* and Otwinowski *et al.* determined through X-ray diffraction measurements the structural affinity of TrpR for binding to DNA and clarified the role of the so-called co-repressor tryptophan [78, 242]. This co-repressor binds tightly onto TrpR, which tilts as a consequence and binds to DNA [79]. We note that our LOV-TAP dark-state system resembles the state in which the co-repressor tryptophan is missing, which leads to a cleavage of wild-type TrpR from DNA. This is caused by the induced stress along the hairpin region, which de-regulates the initial orientation of TrpR in the dark state of LOV-TAP. Finally, this leads to a disruptive process at the TrpR-DNA interface. By contrast, the light-state is unfolded in the hairpin-region and TrpR remains bonded at the DNA-surface. This state resembles a state in wild-type TrpR where the co-repressor tryptophan is present and TrpR binds to DNA. Finally, we show in Figs. 16.8 and 16.9 the overall structures in the final stage of simulation for dark- and

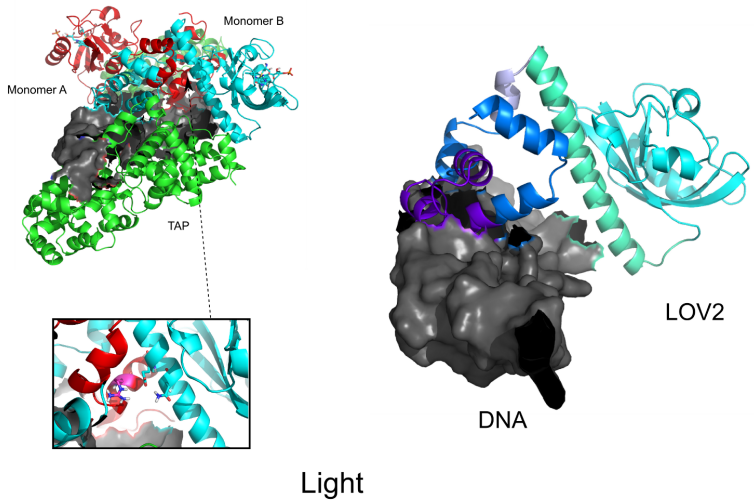


Figure 16.9.: Final structure of LOV-TAP - DNA complex of light-state.

light-state respectively. We oberve in case of the dark-state final structure that both LOV-TAP Monomers are cleaved from the DNA. Moreover, we see that the AsLOV2-J$\alpha$ system is stabilized in the N-terminal region of the J$\alpha$-helix and the hairpin-region remains folded in the dark state. We note that in the final dark-state structure the monomer is at a COM-distance of 4.4 nm to the DNA-surface, which is similar for the other monomer. By contrast, both light-state monomers of LOV-TAP bind onto the DNA-surface, as visualized in Fig. 16.9. We finally show a schematic representation of the different states adopted by LOV-TAP during its photocycle. First of all in Fig. 16.10A we visualize the starting structure of LOV-TAP. In this state LOV-TAP is attached to the DNA, while conformational stress acts in the hairpin-region between TrpR and the J$\alpha$-helix. Next, we display the final stage of the dark-state simulation in Fig. 16.10B, where this stress is released through cleavage of LOV2 from the DNA-surface. In this state the hairpin remains folded. By contrast, the hairpin-region unfolds in the light-state and LOV-TAP remains stable at the TrpR-DNA interface, as we show in Fig. 16.10C. In Fig. 16.10D we display the state of LOV-TAP, where the cleaved dark-state undergoes adduct-formation after irradiation. This state is in equilibrium with the light-state structure of Fig. 16.10C, where LOV-TAP is bound onto the DNA-surface.

## 16.3. Conclusion

In conclusion the light-regulated behavior of LOV-TAP can be summarized as follows. The first steps of signaling in the light-state take place in the vicinity of the FMN-chromophore with a dihedral-switch of the Gln513-sidechain and subsequent coupling between Gln513 and Asn492. This latter coupling leads to a tightening process between the I$\beta$- and H$\beta$-strands and subsequent cleavage of the J$\alpha$-helix, which remains stable in the dark-state. Mediated through the J$\alpha$-helix, the major signal in LOV-TAP is transduced from LOV2 onto the TrpR-DNA interface along the hairpin-region. In the dark-state, this region remains folded in the C-terminal part of the J$\alpha$-helix and induces a stress along this hairpin-region, leading to cleavage of TrpR from the DNA-surface. By contrast in the light-state, unfolding of the J$\alpha$-helix occurs C-terminally and

the hairpin-region becomes flexible, resulting in the tight binding between DNA and the TrpR-protein. Finally, we prove through our simulations the dynamic light-controlled cycle of binding and cleavage between LOV-TAP and DNA, which shows that computer simulation techniques can be useful complementary tools to elucidate and optimize functional behavior of complex proteins in opto-genetical applications.

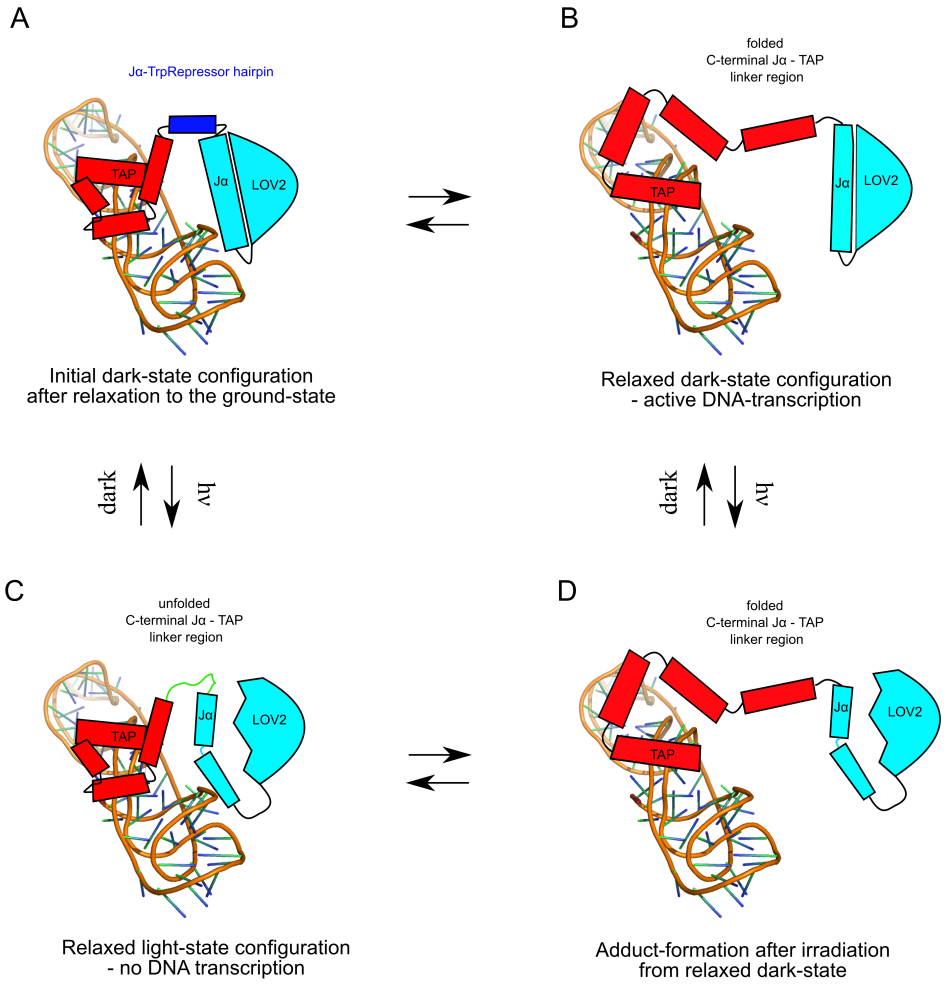


Figure 16.10.: Model of light controlled thermodynamic cycle of LOV-TAP A: Starting structure. B: dark-state final structure (relaxed). C: light-state final structure (bound onto DNA-surface). D: light-state interconverted from relaxed dark-state.

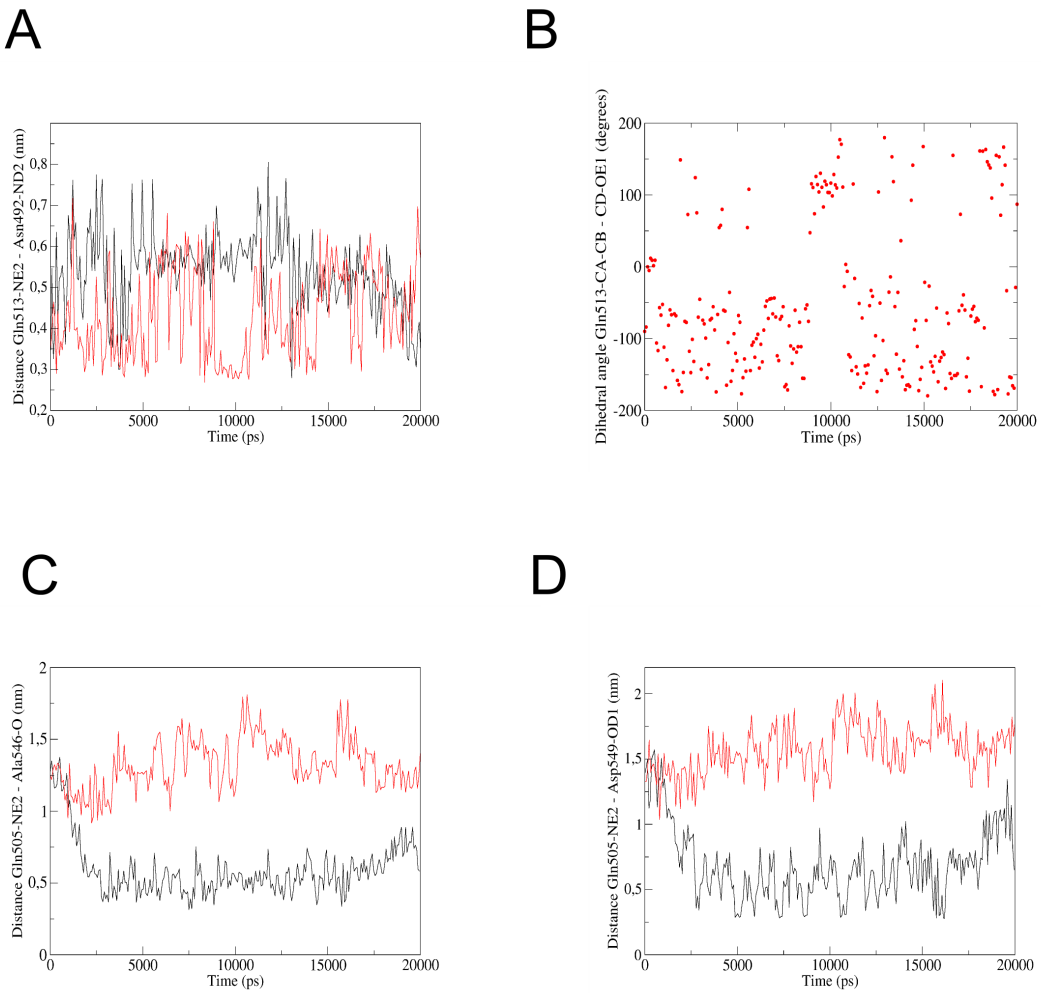


Figure 16.11.: Inter-atomic distances and dihedral-angles of second LOV-TAP monomer as a function of simulation time. A: Distance Gln513-NE2 - Asn492-OD1. B: Dihedral-angle Gln513-CD-OE1 - FMN-C4-N5. C: Distance Gln505-NE2 - Ala546-O at AsLOV2-J$\alpha$-interface. D: Distance Gln505-NE2 - Asp549-OD1 at AsLOV2-J$\alpha$-interface.

## References


1. T. Kottke, J. Heberle, D. Hehn, B. Dick, P. Hegemann, *Biophys. J.* **84**, 1192-1201 (2003).
2. K. Lanzl, M. von Sanden-Flohe, R-J. Kutta, B. Dick, *Phys Chem Chem Phys* **12**: 6594-6604 (2010).
3. R-J. Kutta, E. S. A. Hofinger, H. Preuss, G. Bernhardt, B. Dick, *ChemBioChem* **9** 1931-1938 (2008).
4. M. R. Evans, P. B. Card, K. H. Gardner, *Proc. Natl. Acad. Sci.* **106**, 2617-2622 (2009).
5. B. L. Taylor, I. B. Zhulin, *Microbiol. Mol. Biol. Rev.* **63**, 479-506 (1999).
6. J. R. Nambu, J. O. Lewis, K. A. Wharton, S. T. Crews, *Cell* **67**, 1157-1167 (1991).
7. L. J. Pellequer, K. A. Wagner-Smith, S. A. Kay, E. D. Getzoff, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5884-5890 (1998).
8. W. Gong, B. Hao, S. S. Mansy, G. Gonzalez, M. A. Gonzalez, M. K. Chan, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 15189-15193 (1998).
9. J. H. Morais Cabral, A. Lee, S. L. Cohen, B. T. Chait, M. Li, R. Mackinnon, *Cell* **95**, 649-655 (1998).
10. S. Crosson, K. Moffat, *Plant Cell* **14**, 1067-1075 (2002).
11. R. Fedorov, I. Schlichting, E. Hartmann, T. Domratcheva, M. Fuhrmann, P. Hegemann, *Biophys. J.* **84**, 2474-2482 (2003).
12. C. Bertolucci, L. J. Ming, G. Gonzalez, M. A. Gilles-Gonzalez, *Chem. Biol.* **3**, 561-566 (1996).
13. M. A. Gilles-Gonzalez, G. Gonzalez, M. F. Perutz, L. Kiger, M. C. Marden, C. Poyart, *Biochemistry* **33**, 8067-8073 (1994).
14. M. A. Gilles-Gonzalez, G. Gonzalez, M. F. Perutz, *Biochemistry* **34**, 232-236 (1995).
15. E. K. Monson, G. S. Ditta, D. R. Helinski, *J. Biol. Chem.* **270**, 5243-5250 (1995).
16. K. R. Rodgers, G. S. Ludkat-Rodgers, J. A. Baron, *Biochemistry* **35**, 9539-9548 (1996).
17. W. W. Sprenger, W. D. Hoff, J. P. Armitage, K. J. Hellingwerf, *J. Bacteriol.* **175**, 3096-3104 (1993).
18. P. H. Quail, M. T. Boylan, B. M. Parks, T. W. Short, Y. Xu, D. Wagner, *Science* **268**, 675-680 (1997).
19. K. C. Yeh, J. C. Lagarias, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13976-13981 (1998).
20. H. A. Schneider-Poetsch, M. Braun, S. Marx, A. Schaumburg, *FEBS Lett.* **281**, 245-249 (1991).
21. S. A. Kay, *Science* **276**, 753-754 (1997).
22. U. Schibler, *Nature* **393**, 620-621 (1998).

23. D. P. King, Y. Zhao, A. M. Sangoram, L. D. Wilsbacher, M. Tanaka, M. P. Antoch, T. D. Steeves, M. H. Vilaterna, J. M. Kornhauser, P. L. Lowrey, F. W. Turek, J. S. Takahashi, *Cell* **89**, 641-653 (1997).

24. G. L. Semenza, *Physiology (Bethesda)* **19**, 176 (2004).

25. V. A. Bespalov, I. B. Zhulin, B. L. Taylor, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10084-10089 (1996).

26. A. M. Stock, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10487-10489 (1997).

27. V. A. Baryshev, A. N. Glagolev, V. P. Skulachev, *Nature* **292**, 338-340 (1981).

28. A. N. Glagolev, *J. Theor. Biol.* **82**, 171-185 (1980).

29. K. Jason, T. H. Scheuermann, P. C. Anderson, V. Daggett, K. H. Gardner, *J. Am. Chem. Soc.* **131** 17647-17654 (2009).

30. R. K. Bruick, S. L. McKnight, *Science* **294**, 1337 (2001).

31. P. Hegemann, *Annu. Rev. Plant. Biol.* **59**, 167-189 (2008).

32. J. M. Christie, *Annu. Rev. Plant. Biol.* **58**, 21-45 (2007).

33. W. R. Briggs, T-S. Tseng, H-Y. Cho, T. E. Swartz, S. Sullivan, R. A. Bogomolni, E. Kaiserli, J. M. Christie, *J Integr Plant Biol* **49** 4-10 (2007).

34. T. Kottke, P. Hegemann, B. Dick, J. Heberle, *Biopolymers* **82**, 373378 (2006).

35. M. A. Jones, K. A. Feeney, S. M. Kelly, J. M. Christie, *J. Biol. Chem.* **282**, 6405-6414 (2007).

36. M. Salomon, U. Lempert, W. Rüdiger, *FEBS Letters* **572**, 8-10 (2004).

37. J. M. Christie, T. E. Swartz, R. A. Bogomolni, W. R. Briggs *Plant J.* **32** 205-219 (2002).

38. T. Kagawa, M. Kasahara, T. Abe, S. Yoshida, M. Wada, *Plant Cell Physiol* **45** 416-426 (2004).

39. S. Sullivan, C. E. Thomson, D. J. Lamont, M. A. Jones, J. M. Christie, *Mol. Plant* **1** 178-191 (2008).

40. E. Kaiserli, S. Sullivan, M. A. Jones, K. A. Feeney, J. M. Christie, *Plant Cell* **21**, 3226-3244 (2009).

41. T. Bednarz, A. Losi, W. Gärtner, P. Hegemann, J. Heberle *Photochem. Photobiol. Sci.* **3**, 575-579 (2004).

42. E. Schleicher, R. M. Kowalczyk, C. W. M. Kay, P. Hegemann, A. Bacher, M. Fischer, R. Bittl, G. Richter, S. Weber, *J. Am. Chem. Soc.* **126** 11067-11076 (2004).

43. Y. Sato, M. Nabeno, T. Iwata, S. Tokutomi, M. Sakurai, H. Kandori, *Biochemistry* **46** 10258-10265 (2007).

44. T. Iwata, D. Nozaki, S. Tokutomi, H. Kandori, *Biochemistry* **44** 7427-7434 (2005).

45. T. Iwata, D. Nozaki , S. Tokutomi, T. Kagawa, M. Wada, H. Kandori, Biochemistry **42**, 8183-8191 (2003).

46. M. T. Alexandre, R. v. Grondelle, K. J. Hellingwerf, J. T. M. Kennis, *Biophys J.* **97**, 238-247 (2009).

47. C. Neiss, P. Saalfrank, *Photochem. Photobiol.* **77**, 101-109 (2003).

48. M. Dittrich, P. L. Freddolino, K. Schulten, *J. Phys. Chem. B* **109** 13006-13013 (2005).

49. P. L. Freddolino, M. Dittrich, K. Schulten, *Biophys. J.* **91**, 3630-3639 (2006).

50. C. Neiss, P. Saalfrank, *J Chem Inf Comput Sci* **44** 1788-1793 (2004).

51. S. Arai, M. Togashi, M. Shiozawa, Y. Inoue, M. Sakurai, *Chem. Phys. Lett.* **414**, 230-233 (2005).

52. D. Nozaki, T. Iwata, T. Ishikawa, T. Todo, S. Tokutomi, H. Kandori, *Biochemistry* **43**, 8373-8379 (2004).

53. S. M. Harper, L. C. Neil, K. H. Gardner, *Science* **301**, 1541-1544 (2003).

54. S. M. Harper, J. M. Christie, K. H. Gardner, *Biochemistry* **43**, 16184-16192 (2004).

55. A. S. Halavaty, K. Moffat, *Biochemistry* **46**, 14001-14009 (2007).

56. A. I. Nash, W.-H. Ko, S. M. Harper, K. H. Gardner, *Biochemistry* **47**, 13842-13849 (2008).

57. Edmunds LNJ. *Cellular and molecular bases of biological clocks.* New York: Springer Verlag; 1988.

58. C. Heintzen, J. J. Loros, J. C. Dunlap *Cell* **104** 453-464 (2001).

59. K. Lee, J. C. Dunlap, J. J. Loros *Genet* **163** 103-114 (2003).

60. G. F. Gardner, J. F. Feldman, *Genet* **96** 877-886 (1980).

61. S. M. Hunt, S. Thompson, M. Elvin, C. Heintzen, *Proc Natl Acad Sci USA* **107**, 16709-16714 (2010).

62. C. H. Chen, B. S. DeMay, A. S. Gladfelter, J. C. Dunlap, J. J. Loros, *Proc Natl Acad Sci USA* **107**, 16715-16720 (2010).

63. E. Malzahn, S. Ciprianidis, K. Kaldi, T. Schafmeier, M. Brunner, *Cell* **142** 762-772 (2010).

64. S. M. Hunt, M. Elvin, S. K. Crosthwaite, C. Heintzen, *Gene Dev* **21** 1964-1974 (2007).

65. B. D. Zoltowski, C. Schwerdtfeger, J. Widom, J. J. Loros, A. M. Bilwes, J. C. Dunlap, B. R. Crane, *Science* **316** 1054-1057 (2007).

66. B. D. Zoltowski, B. R. Crane *Biochemistry* **47** 7012-7019 (2008).

67. J. S. Lamb, B. D. Zoltowski, S. A. Pabit, Li Li, B. R. Crane, L. Pollack, *J Mol Biol* **393** 909-919 (2009).

68. J. D. Scott, T. Pawson, *Science* **326**, 1220-1224 (2009).

69. J. Lee, Natarajan M, Nashine VC, Socolich M, Vo T, Russ WP, Benkovic SJ, Ranganathan R. *Science* **322**, 438-442 (2008).

70. V. J. Vinson, *Sci. Signal* **1**: ec365 (2008).

71. A. Möglich, R. A. Ayers, K. Moffat, *Structure* **17** 1282-1294 (2009).

72. Y. I. Wu, D. Frey, O. I. Lungu, A. Jaehrig, I. Schlichting, B. Kuhlman, K. M. Hahn, *Nature* **461**, 104-108 (2009).

73. P. Ballario, P. Vittorioso, A. Magrelli, C. Talora, A. Cabibbo, G. Macino, *EMBO J.* **15**, 1650-1657 (1996).

74. E. J. Rebar, C. O. Carlo, *Science* **263**, 671-673 (1994).

75. K. M. Murphy, V. Ranganathan, M. L. Farnsworth, M. Kavallaris, R. B. Lock, *Cell Death Differ.* **7**, 102-111 (2000).

76. T. J. Petty, S. Emamzadah, L. Constantino, I. Petkova, E. S. Stavridi, J. G. Saven, E. Vauthey, T. D. Halazonetis, *EMBO J.* **30**, 2167-2176 (2011).

77. J. Carey, D. E. A. Lewis, T. A. Lavoie, J. Yang, *J. Biol. Chem.* **266**, 24509-24513 (1991).

78. C. L. Lawson, J. Carey, *Nature* **366**, 178-182 (1993).

79. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular Biology of the Cell*, *Garland Science* (2008).

80. A. I. Nash, R. McNulty, M. E. Shillito, T. E. Swartz, R. A. Bogomolni, H. Luecke, K. H. Gardner, *Proc. Natl. Acad. Sci. U.S.A.* **DOI: 10.1073/pnas.1100262108**, 1-6 (2011).

81. D. Strickland, K. Moffat, T. R. Sosnick, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10709-10714 (2008).

82. M. P. Allen, D. J. Tildesley, *Computer Simulations of Liquids* (Oxford Science Publications, 1987).

83. R. Hentschke, *Statistische Mechanik* **Wiley-VCH Weinheim** (2004).

84. A. B. Bortz, M. H. Kalos, J. L. Lebowitz, *J. Comput. Phys* **17**, 10-18 (1975).

85. B. J. Alder, T. E. Wainwright, *J. Chem. Phys.* **31**, 459-466 (1959).

86. A. M. Mathiowetz, A. Jain, N. Karasawa, W. A. Goddard, *Proteins* **20**, 227-247 (1994).

87. S. Miyamoto, P. A. Kollman, *J. Comput. Phys.* **13**, 952-962 (1992).

88. A. K. Mazur, *J. Comput. Chem.* **18**, 1354-1364 (1997).

89. J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **23**, 327-341 (1977).

90. G. Ciccotti, J. P. Ryckaert, *Comp. Phys. Rep.* **4**, 345-392 (1986).

91. B. Hess, H. Bekker, H. J. C. Berendsen, *J. Comput. Chem.* **18**, 1463-1472 (1997).

92. E. Lindahl, B. Hess, D. v.d. Spoel, *J. Mol. Mod.* **7**, 306-317 (2001).

93. M. Bergdorf, C. Peter, P. H. Hünenberger, *J. Chem. Phys.* **119** 9129-9144 (2003), and references therein

94. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **4** 187-217 (1983).

95. C.L. Brooks III, B. M. Pettitt, M. Karplus, *J. Chem. Phys.* **83** 5897-5908 (1985).

96. K. S. Kim, *Chem. Phys. Lett.* **156** 261-268 (1989).

97. R. J. Loncharich, B. R. Brooks, *Proteins Struct. Funct. Genet.* **6** 32-45 (1989).

98. P. J. Steinbach, B. R. Brooks, *J. Comput. Chem.* **15** 667-683 (1994).

99. J. Norberg, L. Nilsson, *Biophys. J.* **79** 1537-1553 (2000).

100. J. A. Barker, R. O. Watts, *Mol. Phys.* **26** 789-792 (1973).

101. I. G. Tironi, R. Sperb, P. E. Smith, W. F. van Gunsteren, *J. Chem. Phys.* **102** 5451-5459 (1995).

102. P. H. Hünenberger, W. F. van Gunsteren, *J. Chem. Phys.* **108** 6117-6134 (1998).

103. P. P. Ewald, *Ann. Phys.* (Leipzig) **64** 253-287 (1921).

104. T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **98** 10089-10092 (1993).

105. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **103** 8577-8593 (1995).

106. R. W. Hockney, J. W. Eastwood, *Inst. of Physics Pub.*, Bristol (1998).

107. D. van der Spoel, P. J. van Maaren, *J. Chem. Theory Comput.* **2** 1-11 (2006).

108. Y. Yonetani, *Chem. Phys. Lett.* **406** 49-53 (2005).

109. W. F. van Gunsteren, H. J. C. Berendsen, *Angew. Chem. Int. Ed. Engl.* **29** 992-1023 (1990).

110. G. Hummer, L. R. Pratt, A. E. Garcia, *J. Phys. Chem.* **100** 1206-1215 (1996).

111. G. Hummer, L. R. Pratt, A. E. Garcia, *J. Phys. Chem. A* **102** 7885-7895 (1998).

112. P. H. Hünenberger, J. A. McCammon, *J. Chem. Phys.* **110** 1856-1872 (1999).

113. N. A. Baker, P. H. Hünenberger, J. A. McCammon, *J. Chem. Phys.* **110** 10679-10692 (1999).

114. N. A. Baker, P. H. Hünenberger, J. A. McCammon, *J. Chem. Phys.* **113** 2510-2511 (2000).

115. R. D. Lins, U. Röthlisberger, *J. Chem. Theory Comput.* **2** 246-250 (2006).

116. W. Weber, P. H. Hünenberger, J. A. McCammon, *J. Phys. Chem. B* **104** 3668-3675 (2000).

117. P. H. Hünenberger, J. A. McCammon, *Biophys. Chem.* **78** 69-88 (1999).

118. C. Sagui, T. A. Darden, *Annu. Rev. Biophys. Biomol. Struct.* **28** 155-179 (1999).

119. B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, *J Chem Theor Comput* **4** 435-447 (2008).

120. E. Lindahl, B. Hess, D. van der Spoel, *J. Mol. Model.* **7**, 306-317 (2001).

121. T. Soares, X. Daura, C. Oostenbrink, L. Smith, W. Gunsteren, *J. Biomol. NMR* **30**, 407-422 (2004).

122. N. Todorova, F. S. Legge, H. Treutlein, I. Yarovsky, *J. Phys. Chem.* **112**, 11137-11146 (2008).

123. D. Frenkel, B. Smit,*Understanding molecular simulation: from algorithms to applications* (Academic Press, 2003).

124. J. B. Pereira-Leal, E. D. Levy, S. A. Teichmann, *Phil. Trans. R. Soc. B* **361**, 507-517 (2006).

125. K. Wennerberg, C. J. Der, *J. Cell Sci.* **117**, 1301-1312 (2004).

126. A. Singh, A. E. Karnoub, T. R. Palmby, E. Lengyel, J. Sondek and C. J. Der, *Oncogene* **23**, 9369-9380 (2004).

127. M. Karplus and J. Kuriyan, *Proc. Natl. Acad. Sci.* **102**, 6679-6685 (2005), and references therein.

128. S. A. Baeurle, *J. Math. Chem.* **46**, 363-426 (2009).

129. Y. Nakasone, T. Eitoku, D. Matsuoka, S. Tokutomi, M. Terazima, *J. Mol. Biol.* **367**, 432-442.

130. W. R. P. Scott, P. H. Hunenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Kruger, W. F. van Gunsteren, *J. Phys. Chem. A* **103**, 3596-3607 (1999).

131. L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, S. J. Marrink *J. Chem. Theory Comp.* **4**, 819-834 (2008).

132. S. O. Yesylevskyy, L. V. Schäfer, D. Sengupta, S. J. Marrink, *PLoS Comput. Biol.* **6**, DOI: e1000810. doi:10.1371/journal.pcbi.1000810, (2010).

133. H. M. Senn and W. Thiel, *Top Curr. Chem.* **268**, 173-290 (2007).

134. J. Noolandi, T. S. Davison, A. R. Völkel, X.-F. Nie, C. Kay, C. H. Arrowsmith, *Proc. Natl. Acad. Sci.* **97**, 9955-9960 (2000).

135. A. R. Völkel and J. Noolandi, *Biophys. J.* **80**, 1524-1537 (2001).

136. T. Coussaert, A. R. Völkel, J. Noolandi and A. P. Gast, *Biophys. J.* **80**, 2004-2010 (2001).

137. G. S. Ayton and G. A. Voth, *Curr. Opin. Struct. Biol.* **19**, 138-144 (2009), and references therein.

138. M. Tuckerman, B. J. Berne, G. J. Martyna, *J. Chem. Phys.* **97**, 1990-2001, (1992).

139. C. R. Sweet, P. Petrone, V. S. Pande and J. A. Izaguirre *J. Chem. Phys.* **128**, 145101 (2008).

140. J. A. Izaguirre, C. R. Sweet and V. S. Pande, *Pac. Symp. Biocomput.* **15**, 240-251 (2010).

141. D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, W. Wriggers, *Science* **330**, 341-346 (2010).

142. V. A. Voelz, G. R. Bowman, K. Beauchamp, V. S. Pande, *J. Am. Chem. Soc.* **132**, 1526-1528 (2010).

143. M. Shirts, V. S. Pande, *Science* **290**, 1903-1904 (2000).

144. D. E. Shaw, R. O. Dror, J. K. Salmon, J. P. Grossman, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, B. Batson, K. J. Bowers, E. Chow, M. P. Eastwood, D. J. Ierardi, J. L. Klepeis, J. S. Kuskin, R. H. Larson, K. Lindorff-Larsen, P. Maragakis, M. A. Moraes, S. Piana, Y. Shan, B. Towles, *Proceedings of the Conference on High Performance Computing, Networking, Storage and Analysis (SC09)*, ACM New York, NY, pp. 1-11 (2009); doi: 10.1145/1654059.1654099

145. J. A. Izaguirre, S. S. Hampton, *J. Comput. Phys.* **200**, 581-604 (2004).

146. H. Grubmueller, *Phys. Rev. E* **52**, 2893-2906 (1995).

147. A. Barducci, R. Chelli, P. Procacci, V. Schettino, F. L. Gervasio, M. Parrinello, *J. Am. Chem. Soc.* **128**, 2705-2710 (2006).

148. D. E. Makarov, H. Metiu, *J. Chem. Phys* **116**, 5205-5215 (2002).

149. D. E. Makarov, C. A. Keller, K. W. Plaxco, H. Metiu, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3535-3539 (2002).

150. O. Collet, *Phys. Rev. E* **67**, 061912 (2003).

151. D. E. Makarov, P. K. Hansma, H. Metiu, *J. Chem. Phys* **114**, 9663-9673 (2001).

152. P.-C. Li, D. E. Makarov, *J. Chem. Phys* **119**, 9260-9268 (2003).

153. M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, H. E. Gaub, *Science* **276**, 1109 (1997).

154. K. E. Sickafus, E. A. Kotomin, B. P. Uberuaga, *Radiation Effects in Solids* (NATO Science Series II: Mathematics, Physics and Chemistry, Vol. **235**).

155. J. Howard, Mechanics of motor proteins, in *Physics of bio-molecules and cells*, 75$^{\text{th}}$ *Les Houches Summer School*, edited by H. Flyvbjerg, F. Jülicher, P. Ormos and F. David (Springer-Verlag, Berlin, 2002), Vol. 75, pp. 69-94.

156. A. A. Gurtovenko, Y. Y. Gotlib, *J. Chem. Phys.* **115**, 6785-6794 (2001).

157. V. B. P. Leite, J. N. Onuchic, G. Stell and J. Wang, Biophys. J. **87**, 3633 (2004).

158. M. A. Jones, K. A. Feeney, S. M. Kelly, J. M. Christie, *J. Biol. Chem.* **282**, 6405-6414 (2007).

159. K. J. Laidler and M. C. King, J. Phys. Chem. **87**, 2657 (1983).

160. T. Keii, Heterogeneous kinetics: theory of Ziegler-Natta-Kaminsky polymerization (Springer Series in Chemical Physics, Berlin, 2004), pp. 59-60.

161. D. K. Roylance, Characterization of polymer deformation and fracture, in *Applications of polymer spectroscopy*, edited by E. G. Brame (Academic, New York, 1978), pp. 207-219.

162. D. C. Liu, J. Nocedal, *Math Program* **45**, 503-528 (1989).

163. D. Fincham, N. Quirke, D. J. Tildesley, *J. Chem. Phys.* **84**, 4535-4546 (1986).

164. Y. Zhang, M. Lagi, D. Liu, F. Mallamace, E. Fratini, P. Baglioni, E. Mamontov, M. Hagen and S.-H. Chen, J. Chem. Phys. **130**, 135101 (2009).

165. A. Pfeifer, T. Majerus, K. Zikihara, D. Matsuoka, S. Tokutomi, J. Heberle, T. Kottke, *Biophys. J.* **96**, 1462-1470 (2009).

166. C. DerMadirossian, A. Schnelzer, G. M. Bokoch, *Mol. Cell* **15**, 117-127 (2004).

167. D. Owen, L. J. Campbell, K. Littlefield, K. A. Evetts, Z. Li, D. B. Sacks, P. N. Lowe, H. R. Mott, *J. Biol. Chem.* **283**, 1692-1704 (2008).

168. J. Yamauchi, Y. Miyamoto, A. Sanbe, A. Tanoue, *Exp. Cell Res.* **312**, 2954-2961 (2006).

169. A. Deiters, *ChemBioChem* **11**, 47-53 (2010).

170. A. Deiters, *Curr Opin Chem Biol* **13**, 678-686 (2009).

171. A. Hall, *GTPases* (Oxford University Press 2000).

172. C. Vonrhein, H. Bonisch, G. Schafer, G. E. Schulz, *J. Mol. Biol.* **282**, 167-179 (1998).

173. A. Strasser, H. J. Wittmann, *J. Mol. Mod.* **13**, 209-218 (2007).

174. J. Meiler, D. Baker, *Proteins* **65**, 538-548 (2006).

175. W. D. Hoff, P. Dux, K. Hard, B. Devreese, I. M. Nugteren-Roodzant, W. Crielaard, R. Boelens, J. Kaptein, J. van Beeumen, K. J. Hellingwerf, *Biochemistry* **33** 13959 (1994).

176. T. H. Scheuermann, D. R. Tomchick, M. Machius, Y. Guo, R. K. Bruick, K. H. Gardner , *Proc. Natl. Acad. Sci. U.S.A.* **106** 450 (2009).

177. R. Baron, D. Trzesniak, A. H. de Vries, A. Elsener, S. J. Marrink, W. F. van Gunsteren, *Chem Phys Chem* **8**, 452-461 (2007).

178. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. de Vries, *J. Phys. Chem. B* **111**, 7812-7824 (2007).

179. J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, M. L. Klein, *J. Phys. Chem. B* **105**, 4464-4470 (2001).

180. S. Izvekov, G. A. Voth, *J. Chem. Phys. B* **109**, 2469-2473 (2005).

181. S. Izvekov, G. A. Voth, *J. Chem. Theor. Comput.* **2**, 637-648 (2006).

182. M. Praprotnik, S. Matysiak, L. D. Site, K. Kremer, C. Clementi, *J. Phys. Condens. Matter* **19**, 292201 (2007).

183. Q. Shi, S. Izvekov, G. A. Voth, *J. Phys. Chem. B* **110**, 15045-15048 (2006).

184. A. Rubinstein, R. F. Sabirianov, W. N. Mei, F. Namavar, A. Khoynezhad, *Phys. Rev. Lett. E* **82**, 021915 (2010).

185. K. Wood, M. Plazanet, F. Gabel, B. Kessler, D. Oesterhelt, D. J. Tobias, G. Zaccai, M. Weik, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18049-18054 (2007).

186. H. Stillinger, T. A. Weber, *Phys. Rev. B* **31**, 5262 (1985).

187. J. Frydman, F. U. Hartl, *Science* **272**, 1497-1502 (1996).

188. C. B. Millard, V. L. Shnyrov, S. Newstead, I. Shin, E. Roth, I. Silman, L. Weiner, *Protein Sci.* **12**, 2337-2347 (2003).

189. T. Mathes, C. Vogl, J. Stolz, P. Hegemann, *J. Mol. Biol.* **385**, 1511-1518 (2009).

190. W. Holzer, A. Penzkofer, P. Hegemann, *J. Lumin.* **112**, 444-448 (2005).

191. M. Lingenheil, R. Denschlag, R. Reichold, P. Tavan, *J. Chem. Theory Comput.* **4** 1293-1306 (2008).

192. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* **81** 3684-3690 (1984).

193. A. Mor, G. Ziv, Y. Levy, *J. Comput. Chem.* **29** 1992-1998 (2008).

194. E. Rosta, N-V. Buchete, G. Hummer, *J. Chem. Theory Comput.* **5** 1393-1399 (2009).

195. S. Nosé S, *J. Chem. Phys.* **81** 511-519 (1984).

196. W. G. Hoover, *Phys. Rev. A* **31** 1695-1697 (1985).

197. M. T. A. Alexandre, R. van Grondelle, K. J. Hellingwerf, B. Robert, J. T. M. Kennis, *Phys. Chem. Chem. Phys.* **10** 6693-6702 (2008).

198. Y. Nishina, T. Kitagawa, K. Shiga, K. Horiike, Y. Matsumura, H. Watari, T. Yamano, *J. Biochem.* **84** 925-932 (1978).

199. Y. Nishina, K. Shiga, K. Horiike, H. Tojo, S. Kasai, K. Yanase, K. Matsui, H. Watari, T. Yamano, *J. Biochem.* **88** 403-409 (1980).

200. P. K. Dutta, R. Spencer, C. Walsh, T. G. Spiro, *Biochim. Biophys. Acta* **623** 77-83 (1980).

201. L. M. Schopfer, M. D. Morris, *Biochemistry* **19** 4932-4935 (1980).

202. J. T. M. Kennis, S. Crosson, M. Gauden, I. H. M. van Stokkum, K. Moffat, R. van Grondelle, *Biochemistry* **42**, 3385-3392 (2003).

203. W. Holzer, A. Penzkofer, M. Fuhrmann, P. Hegemann, *Photochem. Photobiol.* **75** 479-487 (2002).

204. T. A. Schüttrigkeit, C. K. Kompa, M. Salomon, W. Rudiger, M. E. Michel-Beyerle, *Chem. Phys.* **294** 501-508 (2003).

205. Crosson S, Moffat K, *Proc. Natl. Acad. Sci. U. S. A.* **98** 2995-3000 (2001).

206. T. E. Swartz, S. B. Corchnoy, J. M. Christie, J. W. Lewis, I. Szundi, W. R. Briggs, R. A. Bogomolni, *J. Biol. Chem.* **276** 36493-36500 (2001).

207. A. Losi, E. Polverini, B. Quest, W. Gärtner, *Biophys. J.* **82** 2627-2634 (2002).

208. A. Losi, T. Kottke, P. Hegemann, *Biophys. J.* **86** 1051-1060 (2004).

209. S. Salzmann, J. Tatchen, C. M. Marian,*J Photochem Photobiol A* **198** 221-231 (2008).

210. S. Salzmann, V. Martinez-Junza, B. Zorn, S. E. Braslavsky, M. Mansurova, C. M. Marian, W. Gärtner, *J. Phys. Chem. A* **113** 9365-9375 (2009).

211. S. Salzmann, M. R. Silva-Junior, W. Thiel, C. M. Marian, *J. Phys. Chem. B* **113** 15610-15618 (2009).

212. S. Tokutomi, D. Matsuoka, K. Zikihara, *Biochim Biophys Acta* **1784**, 133-142 (2008).

213. S. A. Baeurle, M. G. Kiselev, E. S. Makarova, E. A. Nogovitsin, *Polymer* **50** 1805-1813 (2009).

214. S. S. Tatke, C. K. Loong, N. D'Souza, R. T. Schoephoerster, M. Prabhakaran, *Biopolymers* **89** 582-594 (2008).

215. S. Crosson, S. Rajagopal, K. Moffat, *Biochemistry* **42** 2-10 (2003).

216. X. Yao, M. K. Rosen, K. H. Gardner, *Nat. Chem. Biol.* **4** 491-497 (2008).

217. M. Vendruscolo, *Nat. Chem. Biol.* **4** 449-450 (2008).

218. D. Matsuoka, S. Tokutomi, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13337-13342 (2005).

219. K. Huang, C. F. Beck, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6269-6274, (2003).

220. M. Khrenova, T. Domratcheva, B. Grigorenko, A. Nemukhin, *J. Mol. Model* , in press; doi: 10.1007/s00894-010-0842-1

221. M. Nakasako, K. Zikihara, D. Matsuoka, H. Katsura, S. Tokutomi, *J. Mol. Biol.* **381** 718-733 (2008).

222. H-Y. Cho, T-S. Tseng, E. Kaiserli, S. Sullivan, J. M. Christie, W. R. Briggs, *Plant Physiol* **143** 517-529 (2007).

223. Hahn, K. & Kuhlman, *Nat. Methods* **7**, 595-597 (2010).

224. K. Lanzl, G. Nöll, B. Dick, *ChemBioChem* **9** 861-864 (2008).

225. D. Strickland, X. Yao, G. Gawlak, M. K. Rosen, K. H. Gardner, T. R. Sosnick, *Nat. Methods* **DOI: 10.1038/nmeth.1473** (2010).

226. Strickland, D., Moffat, K. & Sosnick, T. R. *Proc. Natl. Acad. Sci. USA* **105**, 10709-10714 (2008).

227. A. Yamamoto, T. Iwata, Y. Sato, D. Matsuoka, S. Tokutomi, H. Kandori, *Biophys. J.* **96**, 2771-2778 (2009).

228. M. T. Alexandre, T. Domratcheva, C. Bonetti, L. J. G. W. van Wilderen, R. van Grondelle, M.-L. Groot, K. J. Hellingwerf, J. T. M. Kennis, *Biophys J.* **97**, 227-237 (2009).

229. Ko, W.-H., Nash, A. I. & Gardner, K. H., *Nat. Chem. Biol.* **3**, 372-374 (2007).

230. Y-L. Wan, W. Eisinger, D. Ehrhardt, U. Kubitscheck, F. Baluska, W. Briggs, *Mol. Plant*, **1** 103-117 (2008).

231. K. M. Folta, L. S. Kaufmann, *Plant Mol. Biol.* **4**, 609-618 (2003).

232. S. Inada, M. Ohgishi, T. Mayama, K. Okada, T. Sakai, *Plant Cell* **16**, 887-896 (2004).

233. A. Onodera, S.G. Kong, M. Doi, K.I. Shimazaki, J. Christie, N. Mochizuki, A. Nagatani, *Plant Cell Physiol.* **2**, 367-374 (2005).

234. J. M. Christie, P. Reymond, G. K. Powell, P. Bernasconi, A. A. Raibekas, E. Liscum, W. R. Briggs, *Science* **282**, 16981701 (1998).

235. D. Chivian, D. E. Kim, L. Malmstroem, P. Bradley, T. Robertson, P. Murphy, C. E. Strauss, R. Bonneau, C. A. Rohl, D. Baker, *Proteins* **53** 524-533 (2003).

236. A. Moeglich, K. Moffat, *J. Mol. Biol.* **373**, 112-126.

237. M. C. Peitsch, *Bio/Technology* **13**, 658-660 (1995).

238. K. Arnold, L. Bordoli, J. Kopp, T. Schwede, *Bioinformatics* **22**, 195-201 (2006).

239. F. Kiefer, K. Arnold, M. Künzli, L. Bordoli, T. Schwede, *Nucleic Acids Res.* **37** 387-392.

240. C. A. Amezcua, S. M. Harper, J. Rutter, K. H. Gardner, *Structure* **10**, 1349-1361 (2002).

241. T. Mathes, *PhD-thesis*, (2007).

242. Z. Otwinowski, R. W. Schevitz, R.-G. Zhang, C. L. Lawson, A. Joachimiak, R. Q. Marmorstein, B. F. Luisi, P. B. Sigler, *Nature* **335** (1988).

243. V. R. Doltchinkova, M. D. Lambreva, *Electrophoresis* **23**, 2138-2143 (2002).

244. B. D. Zoltowski, B. Vaccaro, B. R. Crane BR, *Nat Chem Biol* **5** 827-834 (2009).

245. E. Peter, B. Dick, S. A. Baeurle, *J. Chem. Phys.*, **136** (12), 124112 (2012).

246. E. Peter, B. Dick, S. A. Baeurle, *Nat. Comm.* **1**, (2010).

247. E. Peter, B. Dick, S. A. Baeurle, *J. Chem. Biol.* doi: 10.1007/s12154-011-0060-z (2011).

248. E. Peter, B. Dick, S. A. Baeurle, *Prot. Struct. Funct. Bioinf.* doi: 10.1002/prot.24031 (2012).

249. E. Peter, B. Dick, S. A. Baeurle, *J Mol Model* doi: 10.1007/s00894-011-1165-6 (2011).

250. E. Peter, B. Dick, S. A. Baeurle, *Prot. Struct. Funct. Bioinf.* doi: 10.1002/prot.23213 (2011).