# Dissertation

submitted to the
Combined Faculties for the Natural Sciences and Mathematics
of the Ruperto-Carola University
Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

Robert Lindner, M.Sc.
from Opole, Poland

date of examination: _____

# Blue-Light-Regulated Adenylyl Cyclases

**optogenetic tools and model systems for inter-domain communication**

Referees: Prof. Dr. Ilme Schlichting
PD Dr. Jochen Reinstein


*Randall Munroe, http://xkcd.com/1430*

*"It is somewhat ironic that we approach a solution to the problem of de novo protein structure prediction just when structural biology promises to provide most of the answers."*

— Aloy *et al.* Proteins 53 Suppl 6 (2003) 436–456

# Contents

# Acknowledgments

Many people liken their doctoral studies to a journey, and indeed, this was a long one. It was not only the destination that kept me going, but it was the support from numerous people along the way that made this work possible. I would like to express my deep gratitude to...

...Ilme Schlichting for being my supervisor, department director, referee, co-author, teacher and mentor, in one person. Thank you for the tremendous support, invaluable lessons in critical scientific reasoning, clear communication, and for encouraging me to swim against the current when I need to.

...Jochen Reinstein for being my second referee and committee member, and more importantly, for being the *advisor next door* for any question.

...Barbara di Ventura, Frauke Gräter and Oliver Gruß for being very interested committee members, for scientific discussions and personal advice.

...Miroslaw Tarnawski for sharing data, great discussions in the lab, and lots of advice. Thanks for handling even the tiniest crystals and for the good times at the synchrotron.

...Elisabeth Hartmann for her tireless efforts in the lab, for teaching me how to handle proteins and for keeping the lab running.

...Udo Heintz and Diana Born, for being great colleagues, lab mates, co-authors, reviewers, *chefs de cuisine* and most importantly, friends.

...Andreas Winkler for mentoring in the beginning of the project, and for becoming a friend and a role model.

...Anton Meinhart, Thomas Barends and Andreas Menzel, for teaching me practical aspects of synchrotron data collection and analysis.

...the members of the Kuhlman Lab at UNC Chapel Hill, especially to Brian Kuhlman, Andrew Leaver-Fay and Hayretin Yumerefendi who taught me about Rosetta modeling.

...Chris Roome for much more than excellent IT support. Thanks for keeping lechuck running so smoothly.

...Tatiana Domratcheva for discussions about the parallels of photochemistry and psychology. Thanks for teaching me that some things have to be complicated.

...Robert Shoeman, for keeping things spinning in our department and for lots of advice how to drive all the mass-spectrometric racing cars.

...Melanie Müller and Marion Gradl for excellent mass-spectrometric sample handling and analysis.

...Kerstin-Anikó Seifert for preparing TEV protease and, more importantly, for bringing good vibes and music to sing along to in the lab.

...Martina Blochmann, for mastering all administrative aspects.

...the media kitchen and the glass washing facility at the MPI for Medical Research, for making my life easier.

...my office mates Cathleen Zeymer, Martine Hähsler and Sebastian Fischer for the good atmosphere, and for putting all the candy into my drawer.

...Lukas Reuter, Michael Heimes and Felix Schübel, for ensuring my supply of caffeine, for introducing me to the right people, and for making sweatpants look classy.

...Armin Kubis, Florian Iser and Kai Horny who joined me for their lab rotations and thesis research. Thanks for your trust and your excellent work. I can only hope that you learned half as much from me as I learned from you.

...the Hartmut Hoffmann-Berling International Graduate School (HBIGS) for a fellowship and a great curriculum.

...Kristina, Thomas, Kazu, Gabriela, Clara, Florence, Matthias, Egle, Max, Andrea, Jule, Julian, Iris, Sabine, Sophie, Tiia, Veronika, Madeleine, Alexa, Stephan, Akram, Andreas, Bruce, Marco, Marie, Miriam, Stephanie, Leo, Alex and Maren who contributed to the productive atmosphere at the institute.

...the volleyball family at HTV, for great volleyball matches and for trusting me as a coach.

...the *Neckarkrokodile* canoe polo team at WHW, for welcoming me and introducing me to the underwater world of the Neckar river.

...Severin Schneider and Alexandra Becker for making home a wonderful place.

...my sister Katharina for being able to tell entire stories without speaking a single word. You will understand best when I say that a PhD is very much like mountaineering.

...my parents for their sustained support and patience, for their love, their unique sense of humor, and for showing me that growing up is not about taking things too seriously.

# Abstract

Nature has evolved photosensory proteins that enable organisms to adapt their behavior to the prevailing light conditions. BLUF (sensor of blue light using FAD) and LOV (light, oxygen, voltage) photoreceptor domains are small globular domains that sense blue light using a flavin chromophore. These domains regulate diverse biological functions in a modular fashion, either as part of multidomain proteins or through protein–protein interactions. Natural and synthetic BLUF- and LOV-regulated proteins are popular tools for the targeted and non-invasive manipulation of biological systems by light and there is high demand for synthetic photoreceptor proteins to subject novel systems to optical control. The challenges encountered with the design of synthetic BLUF and LOV-regulated proteins illustrate the shortcomings of our current understanding of BLUF and LOV signaling mechanisms and allosteric regulation of multidomain proteins in general.

This work uses a wide array of biophysical methods, including X-ray crystallography, hydrogen-deuterium exchange mass spectrometry, small angle X-ray scattering and enzymology to characterize the structure and the photoactivation mechanism of two blue-light activated adenylyl cyclases (ACs): the BLUF-domain regulated bPAC from *Beggiatoa sp.* and the LOV-domain regulated mPAC from *Microcoleus chtonoplastes*. These two proteins contain a photoreceptor (BLUF or LOV) domain followed by an AC regulator domain and serve as model systems for allosteric regulation in multidomain proteins. As experimental determination of multidomain protein structure is a challenging task, this work also explores computational routes to predict multidomain protein structure from known domain structures and unknown domain linkers. A Rosetta modeling protocol was developed and thoroughly benchmarked to investigate if multidomain protein structures can be predicted or even rationally designed.

Crystal structures and high-resolution models of bPAC and mPAC revealed that the cores of the photoreceptor domains do not share an interface with the AC effector domains. Instead, this work demonstrates for the first time that the domains are linked by a coiled coil and a *handle*-like helical structure that constitute a conserved regulatory extension of the AC domain. Signals from the photoreceptors propagate through the coiled coil and the handles that interact with a prominent extrusion of the AC domain (the *tongue*). The tongue is established as a master regulator of the AC that controls both the opening angle of the AC active site cleft and the position of catalytic residues.

A model is derived, explaining how the fundamentally different signal generation mechanisms of BLUF- and LOV domains are transmitted to the coiled coil where they eventually lead to similar small-scale transient motions. Analysis of the interactions between the coiled coil, the handles and the tongues demonstrates that small and transient motions are sufficient to increase AC activity several hundred-fold through slight shifts of the AC conformational equilibrium. Together, this results in a new model of allosteric AC activation that unifies many previous, seemingly contrasting, observations. It is further shown that the Rosetta modeling protocol performs very well on static assemblies with

large interdomain interfaces, and can be used to predict and design stable multidomain structures. However, in the light of the subtleties that nature came up with for allosteric multidomain switches such as bPAC or mPAC, structural modeling is still far away from rational design of functional linkers.

# Zusammenfassung

Lichtempfindliche Proteine, sogenannte Photorezeptoren, ermöglichen Organismen, sich den gegebenen Lichtbedingen anzupassen. BLUF (*sensor of blue light using FAD*) und LOV (*light, oxygen, voltage*) Photorezeptoren sind kleine globuläre Domänen, die mithilfe eines Flavin-Cofaktors empfindlich für den blauen Bereich des sichtbaren Lichtspektrums sind. Diese Domänen sind modularer Bestandteil zahlreicher Sensor–Effektor-Systeme, in denen sie eine Reihe unterschiedlicher biologischer Funktionen regulieren, entweder durch Protein–Protein Wechselwirkungen oder als regulatorische Domäne in Multidomänenproteinen. Natürliche und synthetische BLUF und LOV Proteine erfreuen sich großer Beliebtheit für die gezielte und nichtinvasive Manipulation biologischer Systeme durch Licht. Entsprechend groß ist das Interesse, mithilfe synthetischer BLUF und LOV Photorezeptorproteine neue Anwendungen zu erschließen. Die Entwicklung neuer BLUF- und LOV-regulierter Photorezeptoren ist jedoch äußerst schwierig und zeugt von den Lücken im Verständnis dieser Domänen im Speziellen und von allosterischer Regulation in Multidomänenproteinen im Allgemeinen.
In dieser Arbeit findet eine Reihe biophysikalischer Techniken Anwendung, darunter Röntgenkristallographie, H/D-Austausch-Massenspektrometrie, Röntgen-Kleinwinkelstreuung und Enzymologie, um zwei blaulichtregulierte Adenylatzyklasen (AC) strukturell und mechanistisch zu charakterisieren: Die BLUF-regulierte AC bPAC aus *Beggiatoa sp.* und die LOV-regulierte AC mPAC aus *Microcoleus chtonoplastes*. Diese Proteine enthalten eine Photorezeptordomäne (BLUF oder LOV), gefolgt von einer AC Domäne und dienen als Modellsysteme für die allosterische Regulation in Multidomänenproteinen. Kristallstrukturen und hochaufgelöste Modelle der untersuchten Proteine und insbesondere der *Linker* zwischen den Domänen sind für die Untersuchung der Mechanismen von entscheidender Bedeutung, jedoch ist die Strukturaufklärung von Multidomänenproteinen ein schwieriges Unterfangen. Aus diesem Grund werden zusätzlich zum experimentellen Ansatz auch theoretische Methoden zur Proteinstrukturvorhersage ausgehend von einzelnen Domänenstrukturen und *Linker*-Sequenzen erforscht. Ein *Rosetta* Protokoll wurde für diesen Zweck entwickelt und an einem Benchmark-Datensatz auf seine Eignung zur Strukturvorhersage oder sogar zum rationalen Design von Multidomänenproteinen getestet.
Kristallstrukturen und Homologiemodelle von bPAC und mPAC zeigen, dass die konservierten Teile der Photorezeptordomänen nicht direkt mit den AC-Domänen interagieren. Stattdessen konnte gezeigt werden, dass ein *Coiled Coil* und henkelförmige helikale Strukturen (*Handles*) wie ein Adapter zwischen Photorezeptor und AC liegen. Diese Strukturen werden erstmals als konservierte Erweiterungen von AC-Domänen identifiziert, wo sie maßgeblich an der Weiterleitung der Signale verschiedener Regulatoren an AC-Domänen beteiligt sind. Lichtinduzierte Strukturänderungen in den Photorezeptordomänen werden auf die *Coiled Coils* und die *Handles* übertragen und erreichen die AC Domäne durch Interaktion der *Handles* mit einem auffälligen Struk-

turelement der AC-Domäne direkt unterhalb des aktiven Zentrums, der Zunge (*Tongue*). Durch die Interaktion zwischen den *Handles* und der *Tongue* werden gleichzeitig der Öffnungswinkel des aktiven Zentrums sowie die Orientierung katalytischer Reste beeinflusst. Dieses Regulationsprinzip ist von den löslichen bakteriellen ACs bPAC und mPAC bis hin zu menschlichen Transmembran-ACs erhalten.
Hieraus ergibt sich ein Modell, welches zeigt, wie die grundlegend unterschiedlichen Signale aus den BLUF und LOV-Domänen die gleichen transienten Fluktuationen in der Konformation des *Coiled Coil* auslösen. Dies führt zu einem Modell der AC-Aktivierung in dem kleine Strukturänderungen ausreichen, um durch kleinste Verschiebungen des konformationellen Gleichgewichts die Aktivität der AC-Domäne um zwei Größenordnungen zu erhöhen. Weiterhin wird gezeigt, dass das Rosetta-Protokoll zu sehr guten Vorhersagen für statische Strukturen mit großflächigen Domäneninteraktionen kommt. In Anbetracht der Feinheiten in den Strukturen allosterischer Multidomänen-Schalter wie bPAC und mPAC muss jedoch festgestellt werden, dass Strukturmodellierung zu diesem Zeitpunkt noch weit davon entfernt ist, zum rationalen Design funktioneller *Linker* beizutragen.

# List of Abbreviations

| | |
|---|---|
| aa | amino acids |
| AC | adenylyl cyclase |
| ApCpp | adenosine-5′-[(α-β)-methyleno]triphosphate |
| ATP | adenosine-5′-triphosphate |
| BLUF | (sensor of) blue light using FAD |
| BSA | bovine serum albumin |
| bp | base pairs |
| cAMP | cyclic adenosine-3′,5′-monophosphate |
| c-di-GMP | 3′,5′-cyclic diguanylic acid |
| cGMP | cyclic guanosine-3′,5′-monophosphate |
| DTE | dithioerythritol |
| DTT | dithiothreitol |
| EDTA | ethylene diamine tetraacetic acid |
| ESI | electrospray ionization |
| FAD | flavin adenine dinucleotide |
| FMN | flavin mononucleotide |
| GFP | green fluorescent protein |
| HDX | hydrogen-deuterium exchange |
| HEPES | 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid |
| HPLC | high performance liquid chromatography |
| HTH | helix turn helix |
| IPTG | isopropyl-β-D-1-thiogalactopyranoside |
| LC | liquid chromatography |
| LOV | light oxygen voltage (photoreceptor domain) |
| MALS | multi-angle light scattering |
| MOPS | 3-(N-morpholino)propanesulfonic acid |
| MS | mass spectrometry |
| NMR | nuclear magnetic resonance |
| nt | nucleotides |
| NTA | 2,2′,2″-nitrolotriacetic acid |
| $OD_{600}$ | optical density measured at 600 nm |
| PAS | Per-ARNT-Sim (domain) |
| PCR | polymerase chain reaction |
| PNK | polynucleotide kinase |
| ppm | parts per million |
| RMSD | root-mean-square deviation |

*continued on next page*

| | |
|---|---|
| SAD | single-wavelength anomalous dispersion |
| SAXS | small angle X-ray scattering |
| SDS | sodium dodecyl sulfate |
| SEC | size exclusion chromatography |
| TCEP | tris(2-carboxyethyl)phosphine |
| TEV | tobacco etch virus |
| Tris | tris(hydroxymethyl)-aminomethan |
| TOF | time of flight |

# Introduction

Sunlight is the primary source of energy for life on earth. Nature has evolved mechanisms that not only harness solar energy but also carry out specific behavioral responses to all relevant parts of the solar spectrum, giving rise to phenomena as simple as bacterial light avoidance or as complex as vision and circadian rhythm. The quality and intensity of light are sensed by specialized *photoreceptor* proteins that translate photon absorption of a *chromophore* into changes of tertiary or quaternary structure of an *effector*.
Interest in photoreceptor proteins has soared after the use of light-gated ion channel proteins to depolarize neurons in living tissues with unprecedented spatiotemporal precision [13, 112, 140]. Numerous other applications using light-activated ion channels, enzymes or transcription factors followed. This new field of research, named *optogenetics*, not only uses naturally occurring photoreceptors but significant efforts have been made to exploit the modular nature of some photoreceptor families to subject a wide range of biological processes to optogenetic control. At this point, the design of optogenetic tools is still far from rational and most synthetic optogenetic proteins are products of heuristic engineering and extensive screening. To take full advantage of nature's rich toolbox one must obtain a better understanding of allostery in multidomain proteins in general, and how regulatory signals are generated and transmitted to an effector in the specific case of photoreceptor proteins.

## 1.1. Photoreceptor Proteins

All known biological responses to light ranging from UV to infrared are carried out by only seven families of photoreceptor proteins. Similar to the widely known fluorescent proteins (e.g., green fluorescent protein, GFP [16, 151]), photoreceptors contain a light-absorbing chromophore (mostly a bound cofactor). However, in addition to decay of the light-excited state through fluorescence, photoreceptors are characterized by the presence of non-fluorescent decay channels that result in conformational changes of the

chromophore and the protein, producing the *lit state*. Photoreceptor proteins are classified according to their chromophore, the protein structure, and the reactions leading to the lit state. Rhodopsins are the only class of transmembrane photoreceptors and constitute a diverse family of proteins. They bind a retinal chromophore which can absorb nearly the entire range of visible light and isomerizes upon illumination. The UV-B sensor UVR8 is the only photoreceptor that exclusively uses amino acids residues as chromophore. Here, a cluster of three conserved tryptophan residues senses UV-B light. Illumination triggers dissociation of the dimer, followed by nuclear import and UV damage gene expression response [27, 165, 24]. Four families of photoreceptors have specialized in sensing the blue spectral region of light (around 450 nm): photoactive yellow protein is a small globular protein with an 120 aa PAS (Per-ARNT-Sim [136]) fold that senses blue light by isomerization of a covalently linked *p*-coumaric acid chromophore. LOV (light, oxygen, voltage) domains employ a similar PAS fold, however their blue light reception is based completely different photochemistry with a noncovalently bound flavin cofactor. Upon illumination with blue light LOV domains form a characteristic intermediate that contains a thioether bond between the flavin chromophore and a nearby cysteine residue [173, 174]. Similar to LOV domains, BLUF domains are small α-β domains with a central five-stranded β-sheet but a different topology [6, 84]. The photoactivated state has a red-shifted absorption spectrum indicative of changed hydrogen bonding around the flavin chromophore [126, 127, 51]. Cryptochromes are flavin-binding all-alpha proteins that signal through changes in redox state of the chromophore. CarH type photoreceptors are cobalamin-binding transcription factors that are sensitive to UV, blue and green light [152, 83]. Orange carotenoid protein (OCP) [92] is a highly specialized photoreceptor in cyanobacteria where it senses the presence of green- to blue light using a carotenoid cofactor [93, 110]. In the light-activated state OCP binds to the light-harvesting antennae [62] and contributes to nonphotochemical quenching of excess energy. Phytochromes were initially discovered in plants where they bind a linear tetrapyrrole chromophore that switches between a red light and a far-red light absorbing state to discriminate between direct sunlight and shade. Homologous proteins with similar spectral properties have been found in bacteria and named bacteriophytochromes [32]. Cyanobacteriochromes are distantly related homologs that cover the entire visible spectrum [79].

The overarching theme in all these proteins is that light-induced structural changes of the

chromophore, e.g., double bond isomerization or changes of its hydrogen bonding pattern, result in conformational change of the surrounding protein, the lit state. The lit state intermediate reverts to the ground state either through thermal decay or through enzymatic or photochemical conversion of the chromophore. Light-induced conformational changes are diverse and include opening of an ion channel or pump [146, 141]), change of oligomeric status [29, 98, 134, 152, 207, 230], DNA binding [71, 126, 215, 143] or modulation of enzymatic activity [7, 80, 171, 159, 164].
The flavin-binding BLUF and LOV blue light photoreceptor domains studied in this work occur as so-called short BLUF/LOV proteins that act as light-dependent interaction modules *in trans* [29, 84, 98, 207] or as allosteric regulators within multidomain proteins [80, 171, 159, 143, 152, 215, 71, 164]. The functions regulated by these small globular photoreceptors are remarkably diverse, hence BLUF and LOV proteins have received much attention as modules for optogenetic protein design.

### 1.1.1. Structure and Function of BLUF Photoreceptors

BLUF domains are almost exclusively found in bacteria and a few unicellular eukaryotes. The isoalloxazine ring of the flavin chromophore is sandwiched between two α-helices atop a five-stranded β-sheet. Three conserved residues are essential for BLUF photochemistry and signaling (in numbering of the archetypical BLUF protein AppA [54, 55, 126, 6]): Tyr21 (β1), Gln63 (β2) and Met106 (β5) (Figure 1.1a). The core of every BLUF domain is extended by C-terminal capping helices, however, the orientation, length and number of capping helices are extremely diverse [125]. These helices are thought to mediate interactions with effectors and it has been shown that photoregulation remains intact in chimeric proteins containing a heterologous BLUF core domain as long as the appropriate capping helix is still present [66, 163]. The light-regulated phosphodiesterase BlrP1 from *Klebsiella pneumoniae* is the only multidomain BLUF protein that has been structurally and mechanistically analyzed in the context of its effector [7, 216]. There, the capping helices interact with the active site through the phosphodiesterase dimerization interface. Further crystal structures of complete BLUF sensor–effector systems will be required to deduce general principles of BLUF signaling through capping helices.
Unlike the conformational changes required for downstream signaling, the photochem-

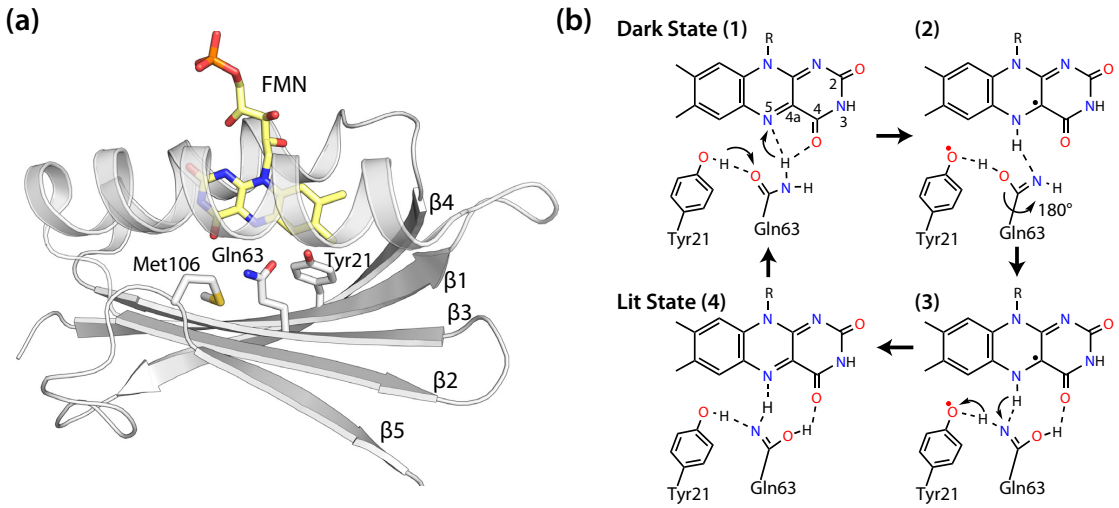



**Figure 1.1.:** Structure and photocycle of the BLUF domain. **(a)** Crystal Structure of AppA BLUF [215]. The BLUF core domain consists of a five-stranded β-sheet flanked by two helices between which a flavin cofactor is sandwiched. The conserved Tyr21 and Gln63 are essential for photochemistry and Met106 is essential for signal transduction. **(b)** Photocycle of the BLUF domain. Illumination of the dark state BLUF domain results in proton coupled electron transfer from Tyr21 to the flavin, concomitant with tautomerization of Gln63, forming a biradical pair (2). The tautomeric Gln63 side chain rotates and is stabilized by hydrogen bonds to flavin N5 and O4 (3). The radical pair recombines and the Tyr21 hydroxyl radical is protonated to form the lit state intermediate (4).

ical events within the BLUF domain have been studied in great detail. The consensus mechanism of the BLUF photoreaction involves a proton-coupled electron transfer from a conserved tyrosine to the excited isoalloxazine ring of the flavin chromophore, followed by tauromerization and likely rotation of a conserved glutamine [190, 37, 38]. The flavin semiquinone radical recombines with the tyrosyl radical to form the long-lived signaling state (Figure 1.1b). Hydrogen bonding of the Z-Z imidic acid form of the glutamine to the flavin results in the observed red shift of the flavin absorption maximum and the downshift of the flavin C4=O stretching vibration [127, 190, 95] observed in the infrared spectrum. Light-induced changes of hydrogen bonding are propagated through the BLUF β-sheet to the capping helices on the domain surface where they ultimately elicit downstream effects [215, 216, 57, 221, 220, 231]. The lit state reverts back to the dark state with lifetimes ranging between seconds [149] and minutes [126].

The initial use of truncated BLUF proteins [6, 85, 57] for the study of structure, photochemistry and signaling has led to an accumulation of conflicting and misleading data in the field [206]. AppA turned out to be a poor model system as it serves as regulator of

gene expression in a ternary AppA-PpsR-DNA complex, hence its functional output can only be observed indirectly and requires extensive controls [215]. Consequently, BLUF function has long been discussed in terms of visible and infrared spectroscopy, leading to further misinterpretation. The shortcomings of this one-dimensional approach are illustrated by the mutations Y21F and M106A. Both variants are spectroscopically silent in all investigated BLUF proteins (Y21F is dark-locked and M106A undergoes a normal photocycle) and only functional assays showed that they lock PixD and bPAC in pseudolit conformations [128, 231, 192]. Although the conservation of these residues suggests some mechanistic importance, there is no explanation why these amino acid substitutions produce a functionally lit structure. For this reason it is vital that studies on the BLUF signaling mechanism not only consider the spectroscopically observable photocycle but they should also take into account the conformational changes leading to activation of the effector domain.

### 1.1.2. Structure and Function of LOV Photoreceptors

LOV domains are a subclass of the ubiquitous PAS family and occur in bacteria, archaea, fungi and plants. Similar to BLUF domains, the isoalloxazine ring of the flavin cofactor is sandwiched between two α-helices and lies atop a five-stranded β-sheet (Figure 1.2a). The excited $S_1$ state of the flavin chromophore undergoes rapid intersystem crossing to an excited triplet intermediate. With time constants ranging from 100 ns to several ms, the flavin N5 atom is protonated and a nearby cysteine residue forms a covalent thioether bond to the flavin C4(a) atom. Reduction of the flavin by the cysteine, to form a neutral semiquinone radical and a cysteinyl radical are likely intermediate steps of the reaction [8], however they have not been consistently observed [105]. The thioether adduct to the protonated flavin constitutes the signaling state that is characterized by the loss of absorbance in the blue spectral range (450 nm). Protonation of the N5 atom leads to rotation of a conserved glutamine residue on the Iβ-strand, ultimately causing downstream signaling (Figure 1.2b). Accordingly, it has been shown that naturally occurring or engineered LOV domains lacking the adduct-forming cysteine form a long-lived neutral semiquinone radical intermediate that is protonated at N5 and therefore still capable of signaling [229].

Deprotonation of the chromophore is likely the rate-limiting step of lit state decay [2,

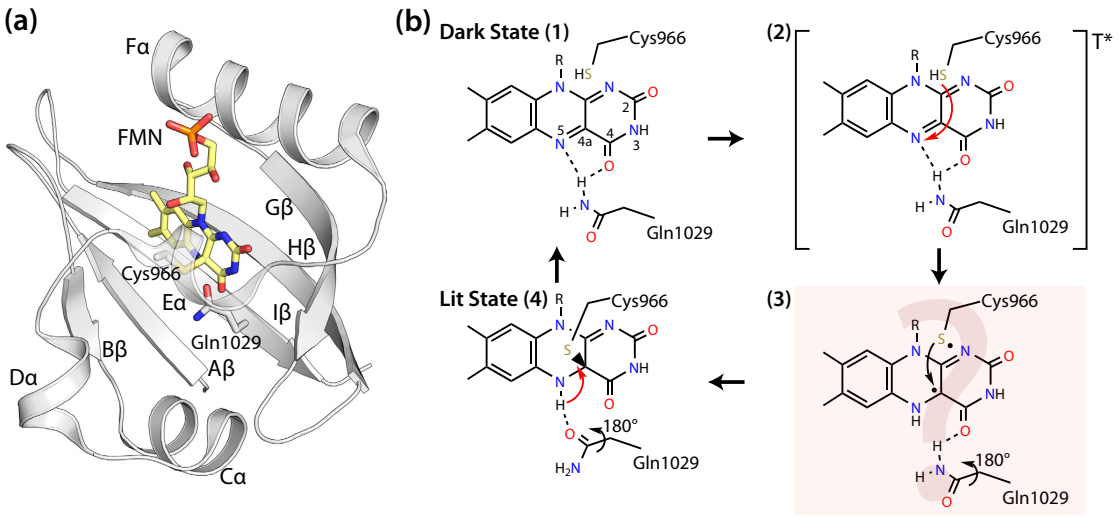



**Figure 1.2.:** Structure and photocycle of the LOV domain. **(a)** Crystal Structure of the *Adiantum capillus-veneris* PHY3 LOV2 domain in the lit state [31]. The LOV core domain consists of a five-stranded β-sheet flanked by two helices between which a flavin cofactor is sandwiched. PAS and LOV nomenclature uses consecutive uppercase letters for secondary structure elements. The conserved Cys966 and Gln1029 are essential for photochemistry and signaling. **(b)** Photocycle of the LOV domain. Reaction steps that are likely but not consistently confirmed are indicated in red. The excited $S_1$ state undergoes intersystem crossing to yield an excited triplet state (2). Intermediates are subject of debate but the excited triplet intermediate likely decays through proton coupled electron transfer and a neutral semiquinone radical [8] species (3) that recombines to form the lit state intermediate (4). The lit state intermediate is characterized by a thioether bond between the conserved Cys966 and flavin C4a. Deprotonation of flavin N5 is likely the rate-limiting step for lit state decay [237, 119]. It is not known if deprotonation of flavin N5 and reprotonation of Cys966 are mediated by an external base.

237]. LOV lit state lifetimes range from seconds [25] to days [82] and although mutations leading to acceleration of retardation of lit state decay have been rationalized, e.g., by facilitating solvent access to the chromophore [119], consistent structural determinants of lit state lifetime are by no means established.

The diversity of LOV protein structures and signaling mechanisms has been widely studied in the field of LOV photoreceptors [52]. A significant amount of structural information has been determined in the context of the effector domains [222, 143, 164, 36, 46] and substantial efforts have been made to derive the mechanism of effector modulation from studies of short or truncated LOV proteins [167, 214, 71, 207, 29]. Inspection of the architectural diversity in LOV-regulated proteins shows interaction of one or more α-helices with the LOV core β-sheet as recurring theme (Figure 1.3). The helices precede (e.g. YtvA/YF1 [36] or PpsB1 [26]), follow (e.g. LOV2 [68]) the LOV

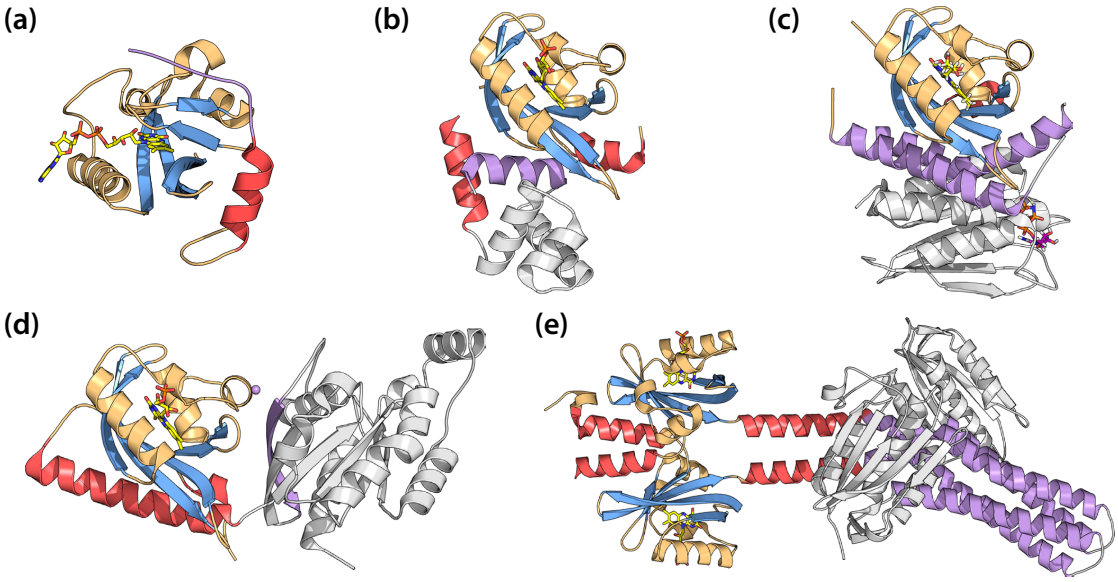



**Figure 1.3.:** Diversity of LOV domain proteins. LOV core domains are colored according to their secondary structure elements (α-helices yellow, β-sheets blue). N- and C-terminal helical extensions (denoted A'α and Jα, respectively) are colored in red, and elements involved in signal transduction are colored in purple. **(a)** VIVID from *Neurospora crassa*. Illumination releases the N-terminal A'α helix from the LOV β-sheet. The latch (purple) mediates dimerization in the lit state [236]. **(b)** Light-regulated transcription factor EL222 from *Erythrobacter litoralis*. The LOV β-sheet does not pack against an N- or C- terminal extension but instead directly interacts with a helix of the DNA-binding helix-turn-helix (HTH) motif. Illumination releases the HTH motif from the β-sheet and thereby enables DNA binding [143]. **(c)** Light-regulated histidine kinase EL346 from *Erythrobacter litoralis*. The catalytic/ATP-binding (CA) domain is coupled to the LOV domain via a dimerization/histidine phosphotransfer-like (DHpL) domain of which two α-helices pack against the β-sheet of the LOV core domain. Illumination loosens the LOV–DHpL packing and thereby promotes CA domain activation [164]. **(d)** Synthetic fusion protein of *Avena sativa* phototropin 1 LOV2 to human Rac1 GTPase: photoactivatable Rac (*PA-Rac*) [222]. The C-terminal LOV Jα helix packs against the LOV β-sheet. Illumination increases the dynamics of the Jαhelix and loosens the LOV-Rac1 interface [214]. **(e)** Synthetic light-regulated histidine kinase YF1 [135, 36]. The *Bacillus subtilis* YtvA LOV domain replaces the oxygen-sensing PAS domain of the *Bradyrhizobium japonicum* FixL histidine kinase. The N-terminal LOV A'α helices are sandwiched between the LOV core domains of the dimeric protein. The C-terminal Jα helices stem from FixL and extend away from the LOV domain to form a coiled coil that is continued in the dimerization/histidine phosphotransfer (DHp) domain on which the catalytic CA domains reside. Illumination changes the packing of the A'α helices against the LOV cores, leading to splaying of the Jα-helices [46] which is assumed to induce helix rotation in the DHp domain [10].

domain in primary sequence, or they may be entirely unrelated to the LOV domain and part of the effector (e.g. EL222 [143] and EL346 [164]).

A common theme observed in LOV signaling is an altered sheet–helix interaction in the light-activated state. The structural consequences of light-activation are manifold

and depend on the specific protein architecture: The helix may dissociate from the β-sheet to release the effector from a caged conformation [143, 71, 222] or to expose or destroy an oligomerization interface [71, 207, 29]. Other oligomeric LOV proteins undergo more subtle changes in quaternary structure that specifically change the effector conformation [164, 36, 167, 46].

### 1.1.3. LOV and BLUF based Optogenetic Tools

The description of LOV and BLUF photoreceptor domains as modular building bricks has fueled visions of mix-and-match like assembly of novel multidomain photoreceptors to subject any cellular function to optogenetic control. Almost twenty years down the road, vital questions concerning rational design, such as the appropriate choice of photoreceptor domain, the design of the interdomain linker and the domain boundaries, still lack general answers. It is unclear how to modulate the lit state lifetime of BLUF domains and reliable determinants of LOV lit state life time are not known either. The *dynamic range* of an optogenetic tool, i.e., the maximum stimulation achievable by light, remains particularly challenging to modify. Thermodynamic considerations predict that suppression of dark state activity is the most promising route to achieve a high dynamic range [227, 194], however this remains a highly theoretical concept as most effector activation mechanisms are not sufficiently explored for activity tuning.

As a consequence, the reality of synthetic multidomain photoreceptor proteins lags behind the high expectations. To date, the BLUF domain has not been used to confer de-novo photosensitivity to any system. Merely naturally occurring BLUF-regulated adenylyl cyclases [80, 171] and GTP-accepting modifications thereof [96] have made their way into the optogenetic toolbox.

The finding that the signaling state of the model system LOV2 from *Avena sativa* phototropin involves dissociation and unfolding of the C-terminal capping helix from the domain core [68, 67] has provided a paradigm for rational design of photocaged peptides that has been exploited with tremendous success [156]. This approach has also been applied to photocage an entire domain [222], however the proposed principle only reliably works for peptides and is unlikely to serve as general strategy for domain caging [214]. Other intrinsic properties such as dimerization of the Aureochrome LOV domain [134, 59] and Vivid [207, 91] have been used to generate light-induced dimer-

izers. There exist only few examples in which the LOV domain has been employed to substitute the regulatory domain of an otherwise regulated protein: the most extensively studied one is the synthetic LOV-regulated histidine kinase YF1 [135]. For generation of this protein, the similarity between the oxygen-sensing PAS domain of the *Bradyrhizobium japonicum* histidine kinase FixL and the LOV domain of the *Bacillus subtilis* transcriptional regulator YtvA has been exploited to specifically mimic the activated PAS domain by light-activated LOV domain. Despite the availability of crystal structures of the YtvA LOV domain [137], the FixL PAS domain [44] and the fusion protein YF1 [36], the structural mechanism of light-induced histidine kinase activation remains unclear.

Characterization of light activation mechanisms is a challenging endeavor for several reasons: Multidomain proteins containing photoreceptor and effector domains may be difficult to handle *in vitro*. High resolution crystal structures are readily available for each domain but often the critical linker regions are either disordered or do not allow crystallization. This is even aggravated for the characterization of the light-activated state because it is not only of transient nature but it also often shows increased secondary structure dynamics [116], adding further structural heterogeneity. In addition to the photoreceptor and the linker, the effector itself serves as a tool to study its activation mechanism by the photoreceptor. Consequently, it should provide a directly measurable output signal that can be correlated with structural and spectroscopic data. There are several LOV and BLUF domain-regulated adenylyl cyclases that meet these criteria and are therefore suitable for mechanistic investigation of light-induced allosteric regulation in addition to being popular optogenetic tools.

## 1.2. Structure and Regulation of Adenylyl Cyclase

Synthesis and breakdown of the universal second messenger cyclic adenosine-3',5'-monophosphate (cAMP) by adenylyl cyclases (AC) and phosphodiesterases are tightly regulated in order to maintain homeostasis. Mutation or pharmacological intervention leading to dysregulation of AC has been linked to various pathologies, including learning defects [223], neuropathic pain [210], dyskinesia [20, 131] or behavioral disorders [223, 19]. Although mostly studied in the neurosciences, tight regulation of AC is also required for non-neuronal processes such as cardiac function and stress response [208] or sperm

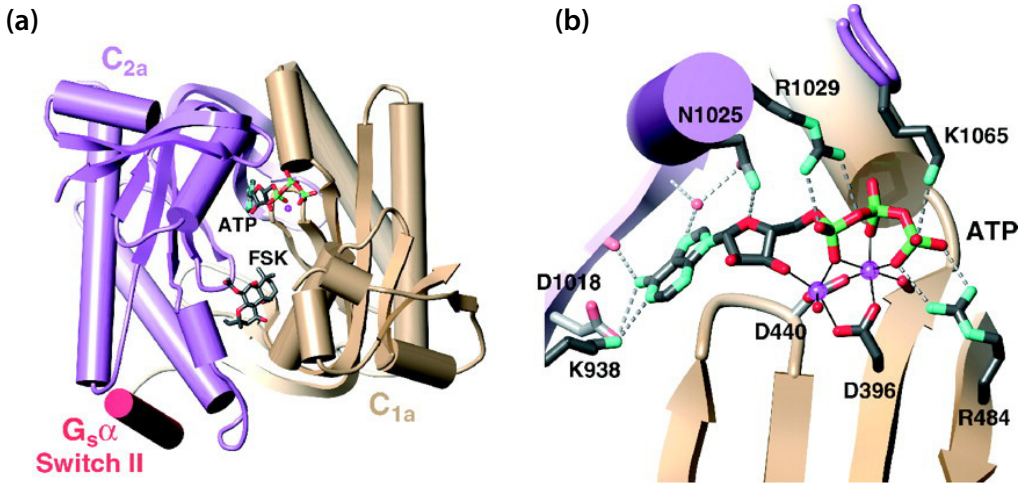



**Figure 1.4.:** Structure of Adenylyl Cyclase. **(a)** Soluble construct of chimeric mammalian transmembrane adenylyl cyclase (tmAC). The two homologous catalytic subunits (C1, ochre and C2, purple) form an active site in the interface. Asymmetric ACs have one catalytic active site that can bind ATP and one degenerate active site that can bind allosteric activators such as forskolin (FSK). Activation of tmACs occurs through interaction with the α subunit of heterotrimeric $G_s$ proteins (pink). Most bacterial ACs are symmetric homodimers and have two active sites. **(b)** Close-up of the AC active site in the C1/C2 subunit interface with modeled ATP nucleotide. C1 contributes the metal-binding Asp396 and Asp440 that bind the phosphate moiety of ATP through two bound $Mg^{2+}$ or $Mn^{2+}$ ions, as well as transition state stabilizing residues such as Arg484. Residues from C2 mediate base-specificity (Lys938, Asp1018), contact the ribose (Asn1029) and balance the negative charge of the phosphates (Arg1029, Lys1065). cAMP is formed through intramolecular nucleophilic attack of the ribose 3'-OH at the α-phosphate. From Tesmer *et al.* [204]. Reprinted with permission from AAAS.

maturation and motility [47, 75].

Type III AC is the most ubiquitous class, employed for signaling in bacteria, lower eukaryotes and animals. Humans possess ten type III AC isoforms, nine of which are transmembrane proteins and one that is soluble. Catalysis occurs at the interface of two homologous subunits in eukaryotes that are thought to originate from gene duplication of an ancestral bacterial homodimeric AC [186]. Two magnesium ions attach the triphosphate moiety of ATP to two conserved aspartate residues of one subunit and assist the nucleophilic attack of the ribose 3'-hydroxyl at the α-phosphate (Figure 1.4). Base-specificity is mainly conferred by residues from the other subunit. A conserved lysine interacts with the adenine N1 nitrogen and an aspartate (subclass IIIa, refer to [114] for a comprehensive classification) or serine/threonine (subclass IIIb) residue interacts with the N6 exocyclic amine. ATP binding and catalysis have been studied extensively by crystallographic [203, 204, 189] kinetic [33, 197] and theoretical analysis [64],

however only little is known about the mechanisms of allosteric regulation, i.e., how inhibited or pre-catalytic states commonly observed in experiments [203, 189, 99, 149], are transformed into catalytically competent ones.
Regulation and activation of adenylyl cyclase has been notoriously difficult to study in vitro because many accessible constructs lacked important regulatory elements and all nucleotide-bound crystal structures showed some features of an inactive conformation. Consequently, no common molecular model of AC activation exists and regulation has been described to occur at various length scales. Mammalian transmembrane ACs are stimulated by interaction with the α-subunit of a heterotrimeric G-protein ($G_s\alpha$) that was proposed to act through adjustment of the angle between the AC subunits [203]. Soluble human AC has been crystallized in the absence and presence of its activator bicarbonate, showing that bicarbonate enhances domain dynamics and serves as scavenger for an arginine residue that otherwise occludes one of the catalytic aspartates [99]. Much larger effects have been described for the mycobacterial pH-regulated AC Rv1264 that was crystallized under low and high pH [205]. A coil to helix transition in the linker between the N-terminal pH sensor and the AC domain disrupts the active site, placing catalytic residues up to 40 Å apart.

### 1.2.1. Optogenetic Control of AC Activity

A number of blue-light regulated type III adenylyl cyclases have been described, utilizing either BLUF [80, 171, 193, 149] or LOV [159] domains as regulators. The first described photoactivatable AC (PAC) originates from the unicellular flagellate *Euglena gracilis* [80] and was named EuPAC. EuPAC is a 400 kDa heterotetrameric complex consisting of two α and two β subunits that each contain two AC and two BLUF domains. Photoregulated AC activity has been shown for each of the subunits [179, 120], and the α subunit is more commonly used as optogenetic tool [15, 9, 212]. EuPAC is difficult to characterize in vitro, hence only little is known about the mechanism of this protein. It appears that at least one BLUF domain is inactive and either not involved in AC regulation [120] or unable to bind flavin [81].
Smaller bacterial PACs have been discovered, including bPAC [171, 193] from *Beggiatoa sp* that combines several favorable properties and is the most widely used PAC to date. bPAC comprises only 350 residues encoding one BLUF and one AC domain that assem-

ble in a homodimer, and it has a 300-fold dynamic range of light-activation [171, 193]. The rather long lit state half-life of 20 s has been reported to flood some systems with cAMP [191], hence shorter half-lives are desirable for some applications. OaPAC [149] is another BLUF-regulated AC homologous to bPAC, and shows one the shortest lit state Although these proteins are promising candidates for optogenetic applications, little is known about the mechanisms of light-induced AC activation. Consequently, there is no starting point for rational modifications of the BLUF domain photocycle kinetics, the sensitivity and dynamic range of activation, or the kinetics and base specificity of the AC domain. In particular, attempts to change the nucleotide specificity of bPAC [171] or to accelerate the dark state recovery of bPAC [192] had only mixed success.

The discovery of a LOV domain regulated AC in *Microcoleus chtonoplastes* (mPAC) held great promise for photocycle tuning. mPAC is a three-domain protein comprising an N-terminal PAS domain of unknown function, a central LOV domain and a C-terminal adenylyl cyclase domain. The suitability of mPAC for in vivo optogenetics has been demonstrated [21], however mPAC can be photoactivated merely 15–20-fold and suffers from very high dark state activity that precludes most applications [159].

Attempts to engineer bacteriophytochrome-regulated adenylyl cyclase have produced an enzyme that can be stimulated sixfold by near infrared light [170]. Although the low dynamic range limits the utility of this protein, phytochrome-based PACs allow noninvasive deep tissue stimulation and their lit state can be specifically switched off by light of a different wavelength, eliminating the need for fast dark state recovery.

In summary, an ideal optogenetic adenylyl cyclase should have a high range of light activation, fast or switchable dark state recovery and it should be manageable *in vivo* as well as *in vitro* for engineering. An enzyme that combines these properties has not yet been found or engineered. In this respect, the small BLUF and LOV regulated ACs bPAC and mPAC are not only useful optogenetic tools but they also offer the advantage that spectroscopic data can be accompanied directly by measurement of enzyme kinetics. This provides a self-consistent functional readout for mechanistic studies that will ultimately serve as basis for rational photoreceptor engineering.

## 1.3. Prediction of Multidomain Protein Structure

Approximately 50 % of the *protein universe* [111], i.e., the entirety of all protein sequences, is found to constitute *domains*: units of conserved sequence and structure that can fold, function and evolve independently of their protein context [213]. Remarkably, while the number of protein sequences is growing exponentially with massive DNA sequencing efforts, and the protein data bank (PDB) containing high-resolution experimental structures is growing at increasing pace, the number of newly identified domains is saturating [111, 175]. In fact, out of 17444 structures (containing 73141 domains) added to the PDB between March 2013 and January 2015 (CATH releases 4.0 to 4.1 [184]), not one single unknown fold was found. This reflects that protein function predominantly evolves through rearrangement and repurposing of present folds, and the novelty brought about by these structures does not lie within the domains. Instead, the arrangement and the interaction of domains in multidomain proteins or complexes as well as the structure of the interdomain linkers, have moved into the focus of attention. While the experimental determination of globular domain structures is now routinely done by X-ray crystallography or nuclear magnetic resonance spectroscopy, the size and inherent flexibility of multidomain proteins with potentially disordered linkers confront structural biologists with a formidable challenge. The computational prediction of protein structure is by no means less challenging than its experimental counterpart. Irrespective of the chosen method for structure prediction, the major bottleneck is to search the vast conformational space that a protein is free to occupy. There is, however, hope for the prediction of multidomain protein assemblies: With the large body of domain structures available from experiments, the search space shrinks to a tractable size if experimental data is appropriately integrated into the modeling process.

### 1.3.1. Domain Assembly Modeling

The task of modeling the tertiary structure of a multidomain protein from the structures of the individual domains is referred to as *domain assembly*. The simplest approach to this problem is to dock the domains by rigid body perturbations, ignoring the potential constraints imposed by the interdomain linker. Several studies demonstrated that integration of distance constraints as a function of linker sequence length significantly

improves prediction performance [22, 224]. Such methods break down for longer linkers with many conformational degrees of freedom as the distance constraint becomes a very weak one when no linker structure is assumed. The opposite approach to constrained docking is explicit folding of the linker. Wollacott *et al.* followed this strategy and used a full Rosetta folding protocol (cf. section 1.3.2) to put the linker into the center of their considerations [218]. Performance of this protocol was no better than that of constrained docking and attributed by the authors to insufficient sampling of linker conformations. The corresponding Rosetta protocol is poorly documented and cannot be investigated in more detail.

As full treatment of the linker conformation is the only approach that allows modeling of structures or interactions that mediate allosteric communication, a new domain assembly protocol is developed in this work. The strategy remains centered on folding of the interdomain linker. A flexible framework is implemented to allow testing of different perturbation methods, including one that attempts to reproduce the results from Wollacott *et al.*. Thorough analysis of successful and failed predictions is carried out to learn which aspects of multidomain structure prediction and design can be reliably handled by the protocol.

### 1.3.2. Rosetta Protein Structure Prediction

Rosetta, like any other computational protein structure prediction (CPSP) algorithm, builds upon two fundamental functions: scoring and sampling. The score function assigns an (actual or arbitrary) energy value to a particular conformation and the sampling function determines the next perturbation to the structure in search of a global energy minimum. Score functions can be physics-based or knowledge-based, or a combination of both. Physics-based score functions in CPSP typically do not consider quantum mechanics for the sheer size of the problem, but instead force fields are used to assign potentials to interatomic interactions and bond geometry that are derived from a combination of quantum mechanical calculations and observation of experimental data [63, 211]. Knowledge-based score functions utilize statistics from known protein structures to determine whether a conformation or an interaction is favorable. Score terms can include secondary structure propensities of residues or peptides, implicit side chain solvation terms, scores from Ramachandran analysis or tabulated pairwise residue

interaction terms [232, 166].
The Rosetta modeling library uses an adaptive score function that combines physics-based and knowledge-based terms in a manner that adapts to the given modeling task. Sampling in Rosetta is mostly based on *Monte Carlo* (MC) algorithms. MC sampling performs random perturbations to the structure and uses a decision criterion to accept or reject the resulting conformation. As the perturbation is completely arbitrary, Monte Carlo algorithms have the potential to sample the conformational space with large steps, avoiding sampling of local minima. This requires a meaningful perturbation method to avoid futile sampling of unphysical conformations, while still covering a large portion of the energy landscape. Several of the most successful perturbation methods are based on insertion of peptide coordinates (*fragments*) derived from structure databases [166, 169, 225]. Fragment insertion bears two advantages: on the one hand, impossible geometries are avoided and on the other hand, appropriate homology-based choice of fragments confines the search space to conformations that the modeled peptide is likely to assume [219, 65]. Over the years, Rosetta has evolved from a single fragment-based structure prediction algorithm to a modular library containing hundreds of *movers* to perturb a structure, a customizable score function, and dozens of *protocols* that apply a sequence of movers and score functions for any given modeling purpose.
The effective incorporation of complementary data to obtain more reliable models of multidomain assemblies is likely to become a major task in computational protein structure prediction. The modular Rosetta framework is ideally suited to integrate various sources of prior knowledge into the modeling process, ranging from experimentally determined domain structures to distance constraints between residues known to interact, or secondary structure predictions.

## 1.4. Aim of the Project

The mechanisms of LOV and BLUF photoreceptors have been extensively studied, yet little is known about their mechanical output and its coupling to the regulated effector domains. This manifests itself in the challenges encountered when designing or tuning proteins for optogenetic applications. Nonetheless, the fact that one type of effector domain can be regulated by different sensors and the fact that one type of photoreceptor can regulate various outputs suggest the existence of a common mechanical *language* at

the interfaces of multidomain proteins that is conserved within and beyond photoreceptors. To decipher this language, in the specific case of two blue light regulated adenylyl cyclases as model systems, as well as on a more general scale, is the aim of this work.

Specifically, one major goal is to determine the structures of bPAC and mPAC to

*i* understand the mechanisms of BLUF and LOV signal generation. The light-induced conformational changes that link conserved residues in the BLUF active site with the domain outside are unknown, and so is the role of aliphatic side chains in signaling and lit state lifetime of LOV domains. Structural details revealing the inner workings of the photoreceptor domains are therefore of particular interest.

*ii* investigate the linker structure that allows propagation of conformational changes from the photoreceptor to the AC domain. Comparison of bPAC and mPAC linker structures and their functional motions shall be used to determine if there exist common – perhaps universal – mechanisms of allosteric AC regulation.

*iii* explore the light-induced conformational changes in the AC domain. This may provide answers to the open question what distinguishes an activated AC from an inactive one.

High-resolution structures are correlated with in-solution measurements of shape and structural dynamics, as well as spectroscopy and enzyme kinetics of selected variants, to capture events in the photoreceptor and in the enzyme domain. In combination, this provides a comprehensive picture of cause and consequence, i.e. the *mechanism* [124].

As outlined above, multidomain proteins, in particular those containing dynamic interdomain linkers, are not always amenable to structural analysis by X-ray crystallography. This work takes advantage of the large number of available domain structures and aims to develop a Rosetta structure prediction protocol that can generate reliable models of multidomain assemblies from domain structures and linker sequences. This approach is not limited to photoreceptor proteins, hence the protocol is benchmarked on a set of 216 two-domain proteins with known crystal structures. Several modeling strategies are explored with a special emphasis on the prediction of interdomain linkers.

Knowing which aspects of multidomain protein structure can be predicted has implications for the extent to which structural modeling can help understand and potentially design the function of multidomain signaling proteins.

# Materials and Methods

## 2.1. Materials

### 2.1.1. Chemicals and Consumables

Chemicals and consumables used in this work are listed in Table 2.1. Aqueous solutions were prepared using water from an ELGA PURELAB Ultra system (ELGA LabWater, Veolia Water Technologies, Celle).

**Table 2.1.:** Laboratory materials used in this work. Aqueous solutions were prepared using water from an ELGA PURELAB Ultra system (ELGA LabWater, Veolia Water Technologies, Celle).

| Name | Manufacturer/Supplier |
|---|---|
| **Chemicals** | |
| acetonitrile (LC-MS grade), proteinogenic L-amino acids, ammonium formate, ampicillin sodium salt, ATP disodium salt, Bis-Tris, $CaCl_2$, FMN sodium salt, formic acid, D-glucose, glacial acetic acid, glycer(ol-$d_3$), 85 % $H_3PO_{4aq}$, HEPES, imidazole, methanol (LC-MS grade), $MnCl_2$, MOPS Na/K-tartrate, $NH_4I$, PEG 3350, $RbCl_2$, sucrose, TBAP, TCEP, TEMED, thiamine, thrombin (bovine plasma), thymine, Tris | Sigma-Aldrich (Steinheim, Germany) |
| APS, $CaCl_2$, tri-sodium citrate · 2 $H_2O$, $HCl_{aq}$, $KOH_{aq}$, $K_2HPO_4$, $KH_2PO_4$, $(NH_4)_2SO_4$, LiCl, β-mercaptoethanol, methanol, $MgCl_2$, NaCl, $NaOH_{aq}$, ethylene glycol, isopropanol, potassium acetate | Merck (Darmstadt, Germany) |
| Bacto agar, Bacto tryptone, Bacto yeast extract, agar | Becton Dickinson Biosciences (Heidelberg, Germany) |
| DTE, DTT, EDTA, glycerol, IPTG, urea | GERBU (Heidelberg, Germany) |
| agarose, bromophenol blue, BSA (Fraction V, pH 7.0), SDS, kanamycin sulfate | Serva (Heidelberg, Germany) |
| ApCpp, cAMP | Jena Bioscience (Jena, Germany) |

*(continued on next page)*

**Table 2.1** *(continued)*

| Name | Manufacturer/Supplier |
|---|---|
| ATP-αS (Rp) | Biolog (Bremen, Germany) |
| GelRed nucleic acid stain | Biotium (Hayward, CA, USA) |
| InstantBlue (coomassie stain) | Expedeon (Over, UK) |
| GeneRuler 1kb DNA ladder | Thermo Fisher Scientific (Waltham, MA, USA) |
| L-selenomethionine | Acros Organics (Geel, Belgium) |
| cOmplete EDTA-free protease inhibitor | Roche Diagnostics (Mannheim, Germany) |
| ProtoGel (acrylamide : bis-acrylamide 37.5:1, 30 % in $H_2O$) | National Diagnostics (Atlanta, GA, USA) |
| Standard-I medium | Carl Roth (Karlsruhe, Germany) |
| **Enzymes and Related Reagents** | |
| *Dpn*I, *Nco*I, *Nde*I, *Not*I, Q5 DNA polymerase, QuickLigase, T4 PNK, CutSmart, T4 PNK buffer, Q5 reaction buffer, Quick Ligation Buffer, deoxynucleotide (dNTP) solution mix (10 mM each), 6 ×purple loading dye | New England Biolabs (Ipswich, MA, USA) |
| TEV protease | provided by Kerstin Seifert |
| **Laboratory Consumables** | |
| Amicon centrifugal filter units | Merck (Darmstadt, Germany) |
| CrystalQuick LP 96-well crystallization trays | Greiner Bio-One (Kremsmünster, Austria) |
| SuperClear 24-well pre-greased crystallization trays | CrystalGen (Commack, NY, USA) |
| MembraCell dialysis tube (0.02 mm regenerated cellulose) | Carl Roth (Karlsruhe, Germany) |
| PuraDisc 1.45 µm syringe filters | GE Healthcare (München, Germany) |
| HisPur Ni-NTA agarose resin | Thermo Fisher Scientific (Waltham, MA, USA) |
| QIAquick PCR Purification Kit, QIAprep Spin Miniprep Kit, PEG Suite, Classic II crystallization screens | Qiagen (Hilden, Germany) |
| Protein LoBild Tubes (1.5 mL) | Eppendorf (Hamburg, Germany) |
| Zymoclean Gel DNA Recovery Kit | Zymo Research (Irvine, CA, USA) |

Buffers and microbial growth media are shown in Table 2.2. Liquid growth media and agar plates were produced and autoclaved by the institute's media facility. Antibiotics were added to the agar-containing media after autoclavation when the temperature had reached 60 °C. Liquid media were supplemented with antibiotics immediately before use.

**Table 2.2.:** Buffers and microbial growth media used in this work. Growth media were prepared by the institute's media facility. Aquaeous solutions were prepared using water from an ELGA PURELAB Ultra system (ELGA LabWater, Veolia Water Technologies, Celle).

| Name | Composition |
|---|---|
| **Growth Media** | |
| LB | 10 g/L tryptone<br>5 g/L yeast extract<br>10 g/L NaCl<br>pH 7.0 |
| LB-Agar | 25 g/L Standard-I<br>15 g/L agar |
| SOC | 20 g/L tryptone<br>5 g/L yeast extract<br>10 mM NaCl<br>10 mM $MgCl_2$<br>10 mM $MgSO_4$<br>2.5 mM KCl 20 mM D-glucose<br>pH 7.0 |
| SeMet medium | 5.6 g/L $K_2HPO_4$<br>1.6 g/L $KH_2PO_4$<br>0.8 g/L $(NH_4)_2SO_4$<br>0.2 g/L tri-sodium citrate · 2 $H_2O$<br>0.2 g/L $MgSO_4$ · 7 $H_2O$<br>5.8 g/L D-glucose<br>40 mg/L each of amino acids ACDEGHNPQRSWY<br>100 mg/L each of amino acids FIKLTV<br>40 mg/L thiamine<br>40 mg/L thymine<br>50 mg/L selenomethionine (immediately before inoculation) |
| **Buffers** | |

*(continued on next page)*

**Table 2.2** *(continued)*

| Name | Composition |
|---|---|
| Tfb-I | 30 mM potassium acetate pH 5.8<br>100 mM RbCl<br>50 mM $MnCl_2$<br>10 mM $CaCl_2$<br>15 % (w/v) glycerol |
| Tfb-II | 10 mM MOPS pH 6.5<br>75 mM $CaCl_2$<br>10 mM RbCl<br>15 % (w/v) glycerol |
| TAE | 40 mM Tris<br>20 mM acetic acid<br>1 mM EDTA |
| Laemmli buffer | 25 mM Tris/HCl pH 8.0<br>200 mM glycine<br>0.1 % (w/v) SDS |
| 4 × SDS loading dye | 25 mM Tris/HCl pH 6.8<br>400 mM DTE<br>8.2 % (w/v) SDS<br>40 % (w/v) glycerol<br>0.1 % (w/v) bromophenol blue |
| IMAC A | 20 mM HEPES<br>300 mM NaCl<br>25 mM imidazole<br>5 % (w/v) glycerol<br>pH 7.5 |
| IMAC B | 20 mM HEPES<br>300 mM NaCl<br>250 mM imidazole<br>5 % (w/v) glycerol<br>pH 7.5 |
| mPAC dialysis | 20 mM HEPES<br>300 mM NaCl<br>5 mM $CaCl_2$<br>2 mM DTE newline 5 % (w/v) glycerol<br>pH 7.5 |
| IEC A | 20 mM HEPES<br>100 mM NaCl<br>5 % (w/v) glycerol pH 7.5 |

*(continued on next page)*

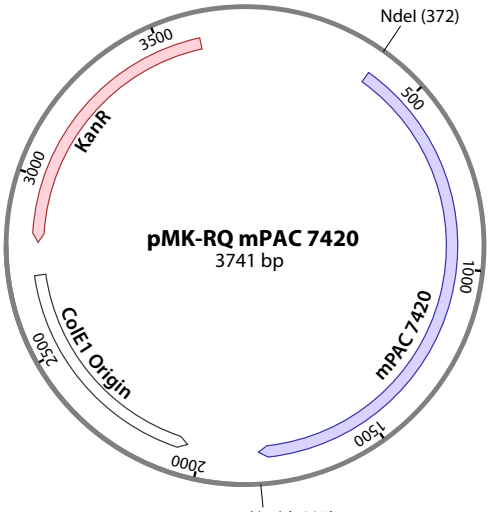
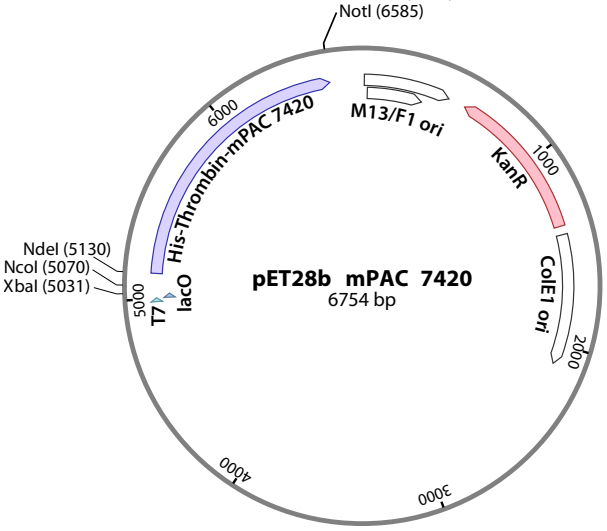
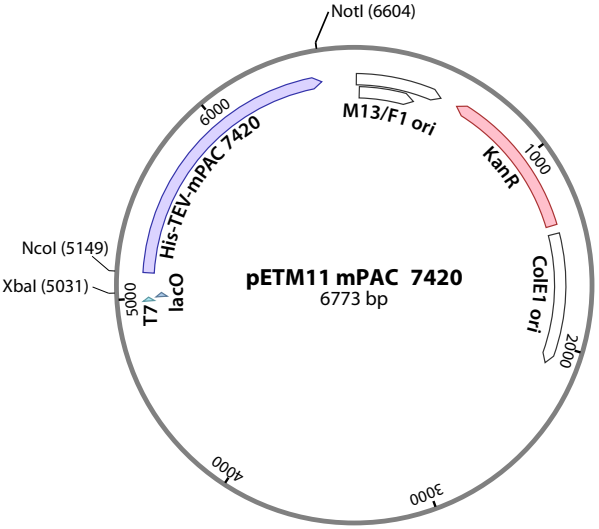
**Table 2.2** *(continued)*

| Name | Composition |
|---|---|
| IEC B | 20 mM HEPES<br>1 M NaCl<br>5 % (w/v) glycerol pH 7.5 |
| SEC | 20 mM HEPES<br>150 mM NaCl<br>5 mM $MgCl_2$<br>2 mM DTE<br>5 % (w/v) glycerol<br>pH 7.5 |
| mPAC HDX Labeling | 25 mM HEPES<br>150 mM NaCl<br>5 mM $MgCl_2$<br>5 % (w/v) glycer(ol-$d_3$)<br>in $D_2O$, pH 7.6 |
| bPAC HDX Labeling | 25 mM HEPES<br>100 mM NaCl<br>10 mM $MgCl_2$<br>2 mM DTE<br>1 mM EDTA<br>5 % (w/v) glycer(ol-$d_3$)<br>in $D_2O$, pD 7.5 |
| mPAC HDX Q | 200 mM ammonium formate<br>2 M Urea<br>pH 2.6 |
| bPAC HDX Q | 200 mM ammonium formate<br>2.8 M Urea<br>pH 2.6 |
| mPAC HPLC | 60 mM $KH_2PO_4$<br>0.5 mM TBAP<br>pH 3.2 |

## 2.1.2. DNA Constructs

The gene encoding mPAC 7420 was synthesized with *E. coli* optimized codons (GeneArt, Regensburg, Germany), *Nde*I and *Not*I restriction sites at the 5' and 3' ends, respectively, and obtained in a pMK-RQ cloning vector. The protein sequence of the UniProt entry

**Table 2.3.:** DNA Constructs Used in this Work.

| Name | Plasmid Map |
|---|---|
| pMK-RQ mPAC 7420 |  |
| pET28b mPAC 7420 |  |
| pETM11 mPAC 7420 |  |

B4VKN6 was used. pET28b and pETM11 vectors were used for protein expression. Plasmid maps of the vectors used in this work are shown in Table 2.3. Site-directed mutagenesis (cf. section 2.2.2) was used to generate truncation variants and point mutants. Variants used in this work are listed in Table 2.4.

**Table 2.4.:** Variants of mPAC 7420 used in this work.

| Plasmid Construct | Short Name |
|---|---|
| **Basic Constructs** | |
| pET28b $H_6$-Thrombin-mPAC 7420 | mPAC p28 |
| pETM11 $H_6$-TEV-mPAC 7420 N2D | mPAC |
| pETM11 $H_6$-TEV-mPAC 7420 M-$^{133}$G–$^{483}$E | $mPAC_{\Delta 132}$ |
| pETM11 $H_6$-TEV-mPAC 7420 M-$^{157}$V–$^{483}$E | $mPAC_{\Delta 156}$ |
| pETM11 $H_6$-TEV-mPAC 7420 M-$^{133}$G–$^{283}$N | $mPAC_{LOV}$ |
| pET28b $H_6$-Thrombin-mPAC 7420 M-$^{301}$L–$^{483}$E | $mPAC_{AC}$ |
| pETM11 $H_6$-TEV-mPAC 7420 V-$^{289}$I–$^{483}$E | $mPAC_{AC+Handle}$ |
| **Variants of mPAC** | |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468L | $mPAC_{K468L}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468Q | $mPAC_{K468Q}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468R | $mPAC_{K468R}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S | $mPAC_{KS}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S G158A | $mPAC_{KS,G158A}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S V160T | $mPAC_{KS,V160T}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S L171A | $mPAC_{KS,L171A}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S L171G | $mPAC_{KS,L171G}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S L171V | $mPAC_{KSLV}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S N226A | $mPAC_{KS,N226A}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S N226S | $mPAC_{KS,N226S}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S Q257D | $mPAC_{KS,Q257D}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S Q257H | $mPAC_{KS,Q257H}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S Q257N | $mPAC_{KS,Q257N}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S L171V V160G | $mPAC_{KSLV,V160G}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S L171V I214L | $mPAC_{KSLV,I214L}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S L171V I214V | $mPAC_{KSLV,I214V}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S L171V L238S | $mPAC_{KSLV,L238S}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S L171V L238T | $mPAC_{KSLV,L238T}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S L171V I240L | $mPAC_{KSLV,I240L}$ |
| pETM11 $H_6$-TEV-mPAC 7420 N2D K468S L171V I240V | $mPAC_{KSLV,I240V}$ |
| **Variants of $mPAC_{LOV}$** | |
| pETM11 $H_6$-TEV-mPAC 7420 M-$^{133}$G–$^{283}$N V160I | $mPAC_{LOV,V160I}$ |
| pETM11 $H_6$-TEV-mPAC 7420 M-$^{133}$G–$^{283}$N L171I | $mPAC_{LOV,L171I}$ |

### 2.1.3. Oligonucleotides

DNA Oligonucleotides were obtained from MWG Biotech (Ebersberg, Germany), with HPSF purification when applicable and HPLC purification otherwise. Oligonucleotides used in this work are listed in Table 2.5.

**Table 2.5.:** DNA Oligonucleotides used in this work.

| Name | Sequence |
|---|---|
| **Subcloning Primers** | |
| p28_NcoI_F | GTT TAA CTT TAA GAA GGA GAT ATA CCA TGG |
| KpnI_TEV_F | TTA GGT ACC GAG AAT CTT TAT TTT CAG GGC G |
| 7420_NcoI-mPAC_F | AAT CCA TGG ATC CGA GCT GCG AAG AAA ATG |
| p28_NotI_R | GCG GCC GCA CTC GAG CAC |
| 7420_dN132_F | TTA CCA TGG GTA GCC AGG CAT ATC G |
| 7420_dN156_F | TAA TCC ATG GTT GGT ATT GTT ATT GCA GAT GC |
| 7420_dN288_F | AAT CCA TGG TGA TTG TGG ATC AGC TGA AAC |
| 7420_dN299_NdeI_F | TTA CAT ATG GCC CAG CAG TTT ACC GAA G |
| 7420_dC283R | TTA AGC GGC CGC AAG CTT AAT TCA GCA GCA GAC GTT CG |
| **Mutagenic Primers (5'-phosphorylated)** | |
| V467_RTH_R | TCA CCA CAA TCG CAC CAC |
| K468S_RTH_F | GCG GCA AAG GTG AAA TGA TTA AC |
| G469_RTH_F | GGC AAA GGT GAA ATG ATT AAC |
| K468R_RTH_R | ACG CAC CAC AAT CGC ACC ACG |
| K468L_RTH_R | CAA CAC CAC AAT CGC ACC ACG |
| K468Q_RTH_R | CTG CAC CAC AAT CGC ACC ACG |
| I256_RTH_R | AAT ACC CAC AAA ATG GGT CAG |
| Q257H_RTH_F | CAT AGC GAT ATT AGC GAT CGT ATT AAA GC |
| Q257D_RTH_F | GAT AGC GAT ATT AGC GAT CGT ATT AAA GC |
| Q257N_RTH_F | AAC AGC GAT ATT AGC GAT CGT ATT AAA GC |
| G158A_RTH_R | AAC AAT CGC AAC GCT GCT TGC TGC AAT TGC |
| I161_RTH_F | ATT GCA GAT GCA CGT CTG C |
| V160T_RTH_R | GGT AAT ACC AAC GCT GCT TGC TGC |
| V160I_RTH_R | GAT AAT ACC AAC GCT GCT TGC TGC |
| V160G_RTH_R | GCC AAT ACC AAC GCT GCT TGC TGC |
| I172_RTH_F | ATT TAT GTT AAT CCG GCA TTT GAA G |
| L171V_RTH_R | CAC CGG CAT ATC CGG CAG ACG TG |
| L171A_RTH_R | CGC CGG CAT ATC CGG CAG ACG TG |

*(continued on next page)*

**Table 2.5** *(continued)*

| Name | Sequence |
|---|---|
| L171G_RTH_R | GCC CGG CAT ATC CGG CAG ACG TG |
| L225_RTH_R | CAG CAG GGT AAC GGT ACA ATT TTC |
| N226S_RTH_F | GCG TAT CGC AAA GAT GGC ACC |
| N226A_RTH_F | AGC TAT CGC AAA GAT GGC ACC |
| K215_RTH_F | AAA GCC GGT GAA AAT TGT ACC G |
| I214L_RTH_R | CAG TGC TGC ACG CAG CTG ATC AAC |
| I214V_RTH_R | CAC TGC TGC ACG CAG CTG ATC AAC |
| S241_RTH_F | AGC CCG ATT TAT GAT GAT CAC |
| I240V_RTH_R | CAC GGT CAG TTC ATT CCA AAA CGG |
| I240L_RTH_R | CAG GGT CAG TTC ATT CCA AAA CGG |
| L238S_RTH_R | AAT GGT GCT TTC ATT CCA AAA CGG GGT GCC |
| L238T_RTH_R | AAT GGT GGT TTC ATT CCA AAA CGG GGT GCC |

# 2.2. Experimental Methods

## 2.2.1. Microbiological Procedures

### Preparation of Chemically Competent Cells

200 mL LB-medium were inoculated with 2.5 mL of a dense overnight culture and grown at 37 °C, shaking at 120 rpm. At an $OD_{600}$ of 0.5, the culture was cooled on ice for 15 min. All subsequent handling and centrifugation was performed at 4 °C. Cells were kept on ice for handling and transfer. Cells were pelleted by 10 min of centrifugation at 4500 × *g* and resuspended in a total volume of 20 mL ice-cold buffer TfbI. After 15 min of incubation on ice, cells were pelleted for 5 min at 3000 rpm and resuspended in a total volume of 3 mL ice-cold TfbII buffer. Aliquots of of 60 µL were distributed into pre-cooled tubes and flash-frozen in liquid nitrogen. Cells were stored at -80 °C.

### Chemical Transformation

60 µL of competent cells were thawed on ice and mixed with 10-50 ng of plasmid DNA. After 10 min of incubation on ice cells were heat-shocked at 42 °C for 40 s and placed

back on ice for 2 min. 150 µL of pre-warmed (37 °C) SOC-medium were added and cells were shaken at 800 rpm and 37 °C for 1 h. Cells were plated on LB-agar plates supplemented with the respective antibiotics and incubated overnight at 37 °C.

### Plasmid Propagation and Preparation

For plasmid propagation, 10 mL of pre-warmed LB medium supplemented with respective antibiotics was inoculated with a single colony of *E. coli* carrying the plasmid. Cells were grown at 37 °C and 120 rpm, harvested after 8 h or overnight growth. Plasmid DNA was isolated using the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

### Cell Culture for Native Protein Expression

*E. coli* BL21 (DE3) cells were transformed with the pET28b or pETM11 plasmid containing the insert of interest (Table 2.4) and a single colony was used to inoculate 100mL of LB medium supplemented with 40 µg/mL kanamycin. The starter culture was grown overnight at 37 °C, shaking at 120 rpm. 5 × 2 L of LB medium containing 40 µg/mL kanamycin were inoculated with each 20 mL the overnight starter culture. Cultures were grown at 37 °C in baffled flasks, shaking at 90 rpm and growth was monitored hourly. Cultures with retarded growth due to cAMP production were grown to an $OD_{600}$ of 0.6 and cultures with doubling times under 1 h were grown to an OD of 0.4 before temperature was reduced to 20 °C. At an $OD_{600}$ of 0.8, expression of the construct was induced with 0.25 mM IPTG. Expression was done overnight and cells were harvested by centrifugation at 4,500 × *g* for 15 min. Pellets were flash-frozen in liquid nitrogen and stored at −80 °C until purification.

### Cell Culture for Expression of Selenomethionine Derivatives

Cultures of *E. coli* BL21 (DE3) carrying pET28b-mPAC$_{AC}$ or pETM11-mPAC$_{AC+Handle}$ were grown overnight in LB-medium containing 40 µg/mL kanamycin. Flasks containing SeMet minimal medium were inoculated 1:100 with dense overnight culture and incubated at 37 °C, shaking at 90 rpm. Temperature was reduced to 20 °C at an $OD_{600}$ of 0.4 and expression was induced with 0.25 mM IPTG at an $OD_{600}$ of 0.8. After overnight

expression, cells were harvested by centrifugation at 4,500 × *g* for 15 min. Pellets were flash-frozen in liquid nitrogen and stored at −80 °C until purification.

## 2.2.2. Cloning of DNA Constructs

### Agarose Gel Electrophoresis

1 % (w/v) agarose was dissolved in TAE buffer by microwave heating. The dissolved molten agarose was stored at 60 °C until used. For one gel, 40 mL of agarose were supplemented with 1 × GelRed Nucleic Acid Gel Stain (Biotium, Hayward, CA, USA) and cast into a PerfectBlue Mini S (7 × 8 cm) gel system (PeqLab, Erlangen, Germany) containing a comb. When the gel had solidified, the comb was removed and the gel system was filled with 1 × TAE buffer. DNA samples were mixed with 1 × purple loading dye (New England Biolabs, Ipswich, MA, USA) and loaded into the wells alongside a GeneRuler 1kb DNA ladder (Thermo Fisher Scientific, Waltham, MA, USA). Electrophoresis was performed at 100 V for 30 min or until the required separation was achieved. Gel imaging was done using a ChemiDoc MP system (Biorad, Hercules, CA, USA) and if necessary, bands were excised on a UV tray (Bioblock Scientific, Illkirch, France).DNA was extracted from the gel using the Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, CA, USA) following the manufacturer's instructions.

### Restriction Digest and Ligation

Each 2 µg of pMK-RQ mPAC 7420 and the pET28b vector were digested using *Nde*I and *Not*I (New England Biolabs) for 15 min in CutSmart buffer (New England Biolabs) at 37 °C. The insert was separated from the vector using agarose gel electrophoresis. The bands corresponding to the mPAC 7420 insert and the empty pET28b vector, respectively, were excised from the gel and DNA was isolated. To obtain the construct encoding mPAC p28 (Table 2.4), the mPAC 7420 insert was ligated into the pET28b vector. Ligation was done using Quick Ligase (New England Biolabs) with 25 ng of vector DNA and a five-fold molar excess of insert DNA in 1 × Quick Ligase buffer. After 15 min of incubation at room temperature, 3 µL of the ligation mixture were transformed into chemically competent *E. coli* TOP10 or BL21 (DE3) Gold cells.

### Polymerase Chain Reaction (PCR)

Starting from the construct encoding mPAC p28, all other basic constructs were generated by PCR. Each reaction was set up on ice in a total volume of 30 µL containing 1 × Q5 polymerase buffer (New England Biolabs), 0.5 µM of the forward and reverse primers, respectively (Table 2.6), 250 µM dNTPs, 2-5 ng of template DNA and 1 U of Q5 DNA polymerase (New England Biolabs). Reactions were carried out in a PTC-200 cycler (MJ Research, St Bruno, QC, Canada) using the following program:

| Temperature | Time | |
|---|---|---|
| 98 °C | 30 s | |
| 98 °C | 10 s | 30 × |
| 56 °C | 10 s | |
| 72 °C | 20 s/kb | |
| 72 °C | 2 min | |
| 4 °C | until processing | |

Reaction products were purified using the QIAquick PCR Purification Kit (Qiagen) following the manufacturer's instructions and digested with the appropriate restriction enzymes. Primers and restriction enzymes used for cloning the respective constructs are shown in Table 2.6. The digested PCR product was isolated by agarose gel electrophoresis. After extraction from the gel, the digested insert was ligated into the appropriate vector.

**Table 2.6.:** Primers and restriction enzymes used for cloning of basic constructs. mPAC p28 was used as template for all reactions.

| Forward Primer | Reverse Primer | Restriction Enzymes | Target Vector | Yields Construct |
|---|---|---|---|---|
| 7420_NcoI-mPAC_F | p28_NotI_R | *Nco*I, *Not*I | pETM11 | mPAC |
| 7420_dN132_F | p28_NotI_R | *Nco*I, *Not*I | pETM11 | $mPAC_{\Delta 132}$ |
| 7420_dN156_F | p28_NotI_R | *Nco*I, *Not*I | pETM11 | $mPAC_{\Delta 156}$ |
| 7420_dN288_F | p28_NotI_R | *Nco*I, *Not*I | pETM11 | $mPAC_{AC+Handle}$ |
| 7420_dN299_NdeI_F | p28_NotI_R | *Nde*I, *Not*I | pET28b | $mPAC_{AC}$ |
| 7420_dN132_F | 7420_dC283_R | *Nco*I, *Not*I | pETM11 | $mPAC_{LOV}$ |

### Site-Directed Mutagenesis

Site-directed mutagenesis was done using a protocol similar to the method described by Hemsley *et al.* [72]. The protocol requires two primers that are phosphorylated at their 5' ends (Table 2.5). Primers were phosphorylated in a total volume of 50 µL containing 1 × T4 Polynucleotide Kinase buffer (New England Biolabs), 10 µM of the respective primer, 3.5 mM ATP and 5 U of T4 Polynucleotide Kinase (New England Biolabs). Phosphorylation was carried out for 1 h at 37 °C followed by inactivation of the kinase during 25 min at 60 °C. Mutagenic PCR was run as described above, using nonoverlapping primers that amplify the entire plasmid. Template DNA was degraded using *Dpn*I and the reaction product was isolated by agarose gel electrophoresis. 50 ng of the isolated DNA were ligated and transformed into chemically competent *E. coli* TOP10. Table 2.7 lists the combinations of primers used to generate the respective mPAC variants.
Coding sequences of all plasmids were verified by sequencing (GATC Biotech, Konstanz, Germany) before protein expression.

**Table 2.7.:** Primer pairs used for site-directed mutagenesis.

| Template | Forward Primer | Reverse Primer | Yields Construct |
|---|---|---|---|
| mPAC | G469_RTH_F | K468L_RTH_R | $mPAC_{K468L}$ |
| mPAC | G469_RTH_F | K468Q_RTH_R | $mPAC_{K468Q}$ |
| mPAC | G469_RTH_F | K468R_RTH_R | $mPAC_{K468R}$ |
| mPAC | K468S_RTH_F | V468_RTH_R | $mPAC_{KS}$ |
| $mPAC_{KS}$ | I161_RTH_F | G158A_RTH_R | $mPAC_{KS,G158A}$ |
| $mPAC_{KS}$ | I161_RTH_F | V160T_RTH_R | $mPAC_{KS,V160T}$ |
| $mPAC_{KS}$ | I172_RTH_F | L171A_RTH_R | $mPAC_{KS,L171A}$ |
| $mPAC_{KS}$ | I172_RTH_F | L171G_RTH_R | $mPAC_{KS,L171G}$ |
| $mPAC_{KS}$ | I172_RTH_F | L171V_RTH_R | $mPAC_{KSLV}$ |
| $mPAC_{KS}$ | N226A_RTH_F | L225_RTH_R | $mPAC_{KS,N226A}$ |
| $mPAC_{KS}$ | N226S_RTH_F | L225_RTH_R | $mPAC_{KS,N226S}$ |
| $mPAC_{KS}$ | Q257D_RTH_F | I256_RTH_R | $mPAC_{KS,Q257D}$ |
| $mPAC_{KS}$ | Q257H_RTH_F | I256_RTH_R | $mPAC_{KS,Q257H}$ |
| $mPAC_{KS}$ | Q257N_RTH_F | I256_RTH_R | $mPAC_{KS,Q257N}$ |
| $mPAC_{KSLV}$ | I161_RTH_F | V160G_RTH_R | $mPAC_{KSLV,V160G}$ |
| $mPAC_{KSLV}$ | K215_RTH_F | I214L_RTH_R | $mPAC_{KSLV,I214L}$ |

*(continued on next page)*

**Table 2.7** *(continued)*

| Template | Forward Primer | Reverse Primer | Yields Construct |
|---|---|---|---|
| $mPAC_{KSLV}$ | K215_RTH_F | I214V_RTH_R | $mPAC_{KSLV,I214V}$ |
| $mPAC_{KSLV}$ | S241_RTH_F | L238S_RTH_R | $mPAC_{KSLV,L238S}$ |
| $mPAC_{KSLV}$ | S241_RTH_F | L238T_RTH_R | $mPAC_{KSLV,L238T}$ |
| $mPAC_{KSLV}$ | S241_RTH_F | I240L_RTH_R | $mPAC_{KSLV,I240L}$ |
| $mPAC_{KSLV}$ | S241_RTH_F | I240V_RTH_R | $mPAC_{KSLV,I240V}$ |
| $mPAC_{LOV}$ | I161_RTH_F | V160I_RTH_R | $mPAC_{LOV,V160I}$ |
| $mPAC_{LOV}$ | I172_RTH_F | L171I_RTH_R | $mPAC_{LOV,L17I}$ |

## 2.2.3. Protein Purification

All purification steps were carried out at ambient temperatures between 4 and 12 °C. Samples were kept on ice and variants containing a photoreceptor domain were handled under safe-light conditions.

### Preparation of Full Length mPAC

Cell pellets were thawed and resuspended in 100 mL buffer IMAC A containing two Tablets of cOmplete EDTA-free protease inhibitor (Roche Diagnostics, Mannheim, Germany). Cells were lysed by four passages through a Microfluidizer (Microfluidics, Westwood, MA, USA) and the lysate was cleared by 1 h of ultracentrifugation at 185,000 × *g*. The cleared lysate was filtered through a 1.45 µm cellulose acetate syringe filter (Puradisc FP30/1.2, GE Healthcare, München, Germany) and loaded on a 5 mL $Ni^{2+}$-NTA Agarose (HisPur, Thermo Fisher Scientific) column equilibrated with IMAC A. The column was washed with 10 volumes of IMAC A and each 10 volumes of 20 % and 50 % of IMAC B in IMAC A. The protein was eluted from the column with 5 column volumes of 75 % IMAC B in IMAC A. mPAC concentration was estimated from absorbance at 280 nm and 447 nm. For constructs expressed from the pETM11 vector, TEV protease was added in a 1:20 molar ratio for proteolytic removal of the hexahistidine tag. Constructs expressed from pET28b were cleaved with 800 U thrombin under otherwise identical conditions.

The mixture was dialyzed overnight in a MembraCell tube (0.02 mm regenerated cellulose, Carl Roth, Karlsruhe, Germany) against 1 L of mPAC dialysis buffer. Subsequently, the solution was applied to a $Ni^{2+}$-NTA agarose column equilibrated with IMAC A, and the flowthrough was retained.
The solution was concentrated to a total volume of 2 mL or a concentration of 300 µM in a centrifugal filter concentrator (Amicon, 100 kDa cutoff, Merck, Darmstadt, Germany), whichever was reached first and incubated with excess FMN for at least 10 min. Fractions of 2 mL were applied to a Superdex 200 16/60 (GE Healthcare) gel filtration column equilibrated with mPAC SEC buffer. Material eluting from the column was fractionated and analyzed by SDS-PAGE (section 2.2.4). Fractions containing the protein of interest were pooled and concentrated to approximately 300 µM. Aliquots of 50 µL were flash-frozen in liquid nitrogen and stored at −80 °C.

### Preparation of $mPAC_{\Delta 132}$

Lysis, $Ni^{2+}$-NTA chromatography, dialysis and re-chromatography were carried out as described for full length mPAC. Following elution from the second $Ni^{2+}$-NTA column the sample was diluted by two volumes of 20 mM HEPES pH 7.5 and applied to a MonoQ 10/100 GL (GE Healthcare) column equilibrated with IEC A buffer. The column was washed with 10 volumes of 15 % IEC B before a gradient to 40 % IEC B was run over 10 column volumes. $mPAC_{\Delta 132}$ elutes at a conductivity of approximately 20 mS $cm^{-1}$. The protein was concentrated in a centrifugal filter concentrator (Amicon, 50 kDa cutoff, Merck) to a total volume of 2 mL or a concentration of 350 µM, whichever was reached first and incubated with excess FMN for at least 10 min. Fractions of 2 mL were applied to a Superdex 200 16/60 (GE Healthcare) gel filtration column equilibrated with mPAC SEC buffer. Fractions eluting from the column were analyzed by SDS-PAGE (section 2.2.4) and those containing $mPAC_{\Delta 132}$ were pooled and concentrated to approximately 350 µM. Aliquots of 50 µL were flash-frozen in liquid nitrogen and stored at −80 °C.

### Preparation of $mPAC_{AC}$, $mPAC_{AC+Handle}$ and Selenomethionine Derivatives

Lysis, $Ni^{2+}$-NTA chromatography, dialysis and re-chromatography were carried out for $mPAC_{AC+Handle}$ and its selenomethionine derivative as described for full length mPAC. $mPAC_{AC}$ and its selenomethionine derivative were not thrombin-treated, hence the dial-

ysis and re-chromatography steps were omitted. Protein concentration was estimated based on absorption at 280 nm wavelength only and the protein was concentrated in a centrifugal filter concentrator (Amicon, 30 kDa cutoff, Merck) to a total volume of 2 mL or a concentration of 500 µM, whichever was reached first. Fractions of 2 mL were applied to a Superdex 75 16/60 (GE Healthcare) gel filtration column equilibrated with mPAC SEC buffer. Fractions eluting from the column were analyzed by SDS-PAGE (section 2.2.4) and those containing $mPAC_{AC}$ variants were pooled and concentrated to approximately 500 µM.

#### Preparation of bPAC Variants

All bPAC Variants were expressed and purified by Elisabeth Hartmann. The procedure consisted of $Ni^{2+}$-NTA chromatography, removal of the hexahistidine tag by thrombin during dialysis, a second round of $Ni^{2+}$-NTA chromatography in which the flowthrough was collected, followed by size exclusion chromatography using a Superdex 75 16/60 (GE Healthcare) column. Detailed procedures are described in the corresponding publication [115].

### 2.2.4. Basic Biochemical Procedures

#### Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Denaturing and reducing electrophoretic analysis of proteins was done by discontinuous SDS-PAGE as described by Laemmli [106]. The separating gel contained 375 mM Tris/Cl pH 8.8, 0.1 % SDS, 10–15 % acrylamide:bis-acrylamide (37.5:1) mixture, 1.5 mM APS and 0.05 % TEMED. The stacking gel contained 125 mM Tris/Cl pH 8.8, 0.1 % SDS, 4.5 % acrylamide:bis-acrylamide (37.5:1) mixture, 2.2 mM APS and 0.1 % TEMED.
Prior to loading, samples were denatured and reduced in 1 × SDS-PAGE loading dye at 95 °C for 3 min. Electrophoresis was performed in Laemmli buffer at a constant current of 35 mA per gel for 40 min. Gels were stained with InstantBlue (Expedeon, Over, UK).

**Table 2.8.:** Molar extinction coefficients of mPAC and bPAC variants at 280 nm wavelength. Coefficients were estimated using ProtParam [50], and 19,000 (21,000) $M^{-1}$ $cm^{-1}$ was added for variants containing a LOV (BLUF) domain.

| Construct | ε [$M^{-1}$ $cm^{-1}$] |
|---|---|
| **mPAC (LOV)** | |
| mPAC (all full-length) | 87,870 |
| $mPAC_{\Delta 132}$ | 50,400 |
| $mPAC_{AC}$ | 16,960 |
| $mPAC_{AC+Handle}$ | 16,960 |
| $mPAC_{\Delta 156}$ | 48,910 |
| $mPAC_{LOV}$ | 33,440 |
| **bPAC (BLUF)** | |
| bPAC (all full-length) | 52,860 |
| $bPAC_{1\text{-}103}$ | 32,460 |
| $bPAC_{1\text{-}127}$ | 33,950 |
| $bPAC_{AC}$ | 17,420 |

### Determination of Protein Concentration by UV/Vis Absorbance

mPAC concentrations were estimated by the absorbance at 280 and 447 nm wavelength using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Thermo Fisher Scientific) according to Lambert-Beer's Law. bPAC concentrations for spectroscopic and HDX experiments were determined by absorbance at 440 nm. Concentrations used in kinetic experiments were provided by Elisabeth Hartmann, and were based on absorbance at 280 nm. To account for the large fraction of other proteins present in the bPAC preparations (cf. section 2.2.7), concentrations determined by absorbance at 280 nm were retrospectively adjusted by the mass fraction of bPAC in the preparation, as judged by SDS-PAGE.

Molar extinction coefficients at 280 nm wavelength were estimated using ProtParam [50]. Full flavin occupancy (FMN for mPAC, FAD for bPAC) in 1:1 stoichiometry was assumed for variants containing photoreceptor domains, hence a value of 19,000 $M^{-1}$ $cm^{-1}$ was added to the calculated molar extinction coefficient at 280 nm for mPAC. The extinction coefficient of free FAD at 280 nm is 21,000 $M^{-1}$ $cm^{-1}$ and was added to the calculated extinction coefficients at 280 nm for bPAC. Resulting molar extinction coefficients are

shown in Table 2.8. Absorption at 447 (or 440 nm, in the case of bPAC) is attributable to protein-bound flavin only and was therefore assumed to be identical for all constructs bearing a LOV or BLUF photoreceptor domain ($\varepsilon$ = 12,500 $M^{-1}$ $cm^{-1}$).

### 2.2.5. Multi Angle Light Scattering

Analytical size exclusion chromatography coupled to multi-angle light scattering was performed to determine the molecular weight and thereby the oligomeric state of mPAC variants in solution. The setup consisted of a HPLC pump (model 626, Waters, Milford, MA, USA), a Superdex 200 10/300 GL size exclusion column (GE Healthcare), an absorbance detector (model 2996, Waters), a refractive index detector (model 2414, Waters), and a Dawn Heleos 18-angle light scattering detector (Wyatt, Santa Barbara, CA, USA). The system was equilibrated and run with mPAC SEC buffer ± 0.5 mM ATP at a flowrate of 0.5 mL $min^{-1}$.
mPAC and $mPAC_{\Delta 132}$ were diluted to 60 µM in mPAC SEC buffer ± 0.5 mM ATP, and 40 µL of each sample were injected. The experiment was carried out at room temperature under safe-light conditions. The molecular weight of the eluting species was calculated using the ASTRA software (Wyatt).

### 2.2.6. Dark State Recovery Kinetics

Dark state recovery kinetics were measured in a custom-built setup that minimizes the influence of probe light on the photoreceptor (Figure 2.1). Probe light was provided by a balanced deuterium-halogen lamp (DH-2000-BAL, Ocean Optics, Ostfildern, Germany) with an OD1 neutral density filter. Photoreceptor excitation was done with collimated light from a royal blue LED ($\lambda_{max}$ = 455 nm, Doric Lenses, Québec, QC, Canada) installed perpendicular to the absorption light path, providing 6 mW $cm^{-2}$. The sample was placed in a Peltier-thermostated 4-way QPOD (Ocean Optics, Ostfildern, Germany) cuvette holder and kept at 20 °C. Spectra were generated using Shamrock SR303i spectrograph (Andor Technology, Belfast, UK) connected to a Newton 970 camera (Andor Technology). Spectrograph and camera settings are shown in Listing 2.1.
Protein was diluted to reach an absorbance of approximately 0.1 at the maximum near 440 nm (bPAC: 440 nm, mPAC: 447 nm), and a path length of 3 mm. Blue light was

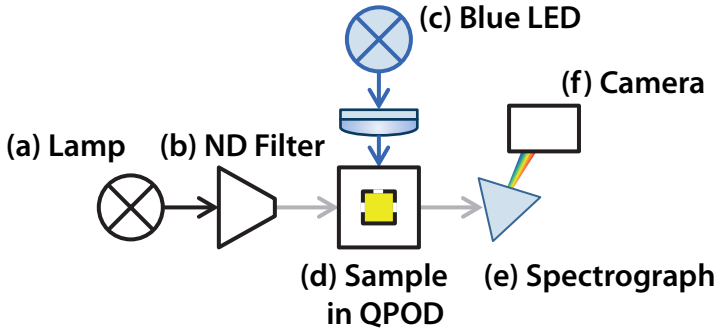



**Figure 2.1.:** Setup for measuring photoreceptor dark state recovery kinetics. **(a)** Probe light is provided by a DH-2000-BAL deuterium-halogen lamp (Ocean Optics) with an OD1 neutral density filter **(b)**.**(c)** Photoreceptor excitation is done with collimated light from a blue LED (Doric Lenses) installed perpendicular to a thermostated QPOD (Ocean Optics) 4-way sample mount **(d)**. Spectra are generated by a Shamrock SR303i spectrograph (**e**, Andor Technology) connected to a Newton 970 camera (**f**, Andor Technology).

switched on for 30 s, and absorption spectra were monitored after the illumination was switched off with a frequency appropriate for the expected dark state recovery time. At least three consecutive exposure–recovery cycles were recorded per sample.

**Listing 2.1:** Spectrograph and camera settings for photoreceptor dark state recovery measurement. Exposure time was held constant at 0.1 s, however acquisition frequency was varied according to the expected recovery time.

```
Temperature (C):              -60
Data Type:                    Optical Density
Acquisition Mode:             Kinetics
Trigger Mode:                 Internal
Exposure Time (secs):         0.10001
Accumulate Cycle Time (secs): 0.10001
Number of Accumulations:      variable
Kinetic Cycle Time (secs):    variable
Readout Mode:                 Full Vertical Binning
Vertical Shift Speed (usecs): 19.35 # 38.55 for bPAC
Pixel Readout Rate (MHz):     2.5
Baseline Clamp:               OFF
Clock Amplitude:              Normal
Output Amplifier:             Conventional
Pre-Amplifier Gain:           1
Wavelength (nm):              425.935
```

```
Grating Groove Density (l/mm):300
Grating Blaze:                500
Filter:                       no filter
Input Side Slit Width (um):   2500
```

Spectra were smoothed using a moving average filter (bandwidth of 6 nm). The absorption maximum was determined, and recovery was fitted to the temporal evolution of a 2 nm wide absorption slice around the maximum wavelength as monoexponential function. The recovery of some bPAC variants was better described as bi-exponential process.

## 2.2.7. Steady State cAMP Production Kinetics

### mPAC Steady State Kinetics

mPAC AC reactions were carried out in a total volume of 250 µL in 1.5 mL Protein LoBind tubes (Eppendorf, Hamburg, Germany) that had been rinsed with 1 mg ml$^{-1}$ BSA before use. mPAC stocks were diluted to 20–40 µL and protein concentration was determined spectroscopically at this stage. An appropriate volume was added to the reaction buffer to reach final mPAC concentrations of 200–500 nM in the reaction. Reactions were initiated by addition of an appropriate volume of an approximately 10 × stock solution of $Mg_2^{2+}$·ATP (concentration determined spectroscopically, $\varepsilon_{259}$ = 15,400 M$^{-1}$ cm$^{-1}$). Unless otherwise indicated, reaction velocities were measured in mPAC SEC buffer at 20 °C and ATP concentrations of 100 µM, either under safe-light conditions or constant illumination from the top of the open reaction vessel with 7 mW cm$^{-2}$ at the sample position. Aliquots of 35 µL were taken, placed in thin-walled vessels and quenched by boiling in a water bath for 3–10 min, followed by flash-freezing in liquid nitrogen. Samples were kept in liquid nitrogen or at −80 °C until processing. Samples were thawed and precipitated protein was removed by centrifugation at 16000 × *g* for 20 min before analysis by HPLC. The cAMP/ATP ratio of the mPAC AC reaction was quantified using absorbance measurement at 254 nm during reversed-phase HPLC, similar to the protocol described by Tekkanat and Fox [201]. Separation of ATP and cAMP was achieved on a Prontosil Hypersorb ODS column (250 × 4.6 mm, 5.0 µm particle size, Bischoff, Leonberg, Germany) with isocratic elution of mPAC HPLC buffer and 5 % acetonitrile at a total flow rate of 1.5 mL min$^{-1}$ at an oven temperature of 20 °C.

For determination of initial velocities, the cAMP/ATP ratio was evaluated and a line was fitted through the maximum number of time points (at least three) showing linear increase of cAMP concentration. The slope [% cAMP $s^{-1}$] was then adjusted to the starting ATP concentration and divided by the enzyme concentration to yield apparent catalytic rate constants.

#### bPAC Steady State Kinetics

bPAC AC reactions and HPLC quantification were carried out by Elisabeth Hartmann [115]. Pronounced differences in the purity of bPAC preparations and day-to-day variation of determined reaction velocities required elaborate controls and post-normalization. Catalytic rates for each bPAC variant were measured multiple times and normalized by the rates measured for a wild type control sample measured under identical conditions on the same day. Furthermore, as different bPAC variants could be enriched to different degrees ranging from 30 to 60 %, catalytic rate constants were additionally corrected by the fractional contribution of bPAC as judged by SDS-PAGE.

### 2.2.8. Hydrogen/Deuterium Exchange Mass Spectrometry

#### HDX-MS of mPAC

Aliquots of mPAC labeling buffer (38 µL ± 2.5 mM ApCpp) were prepared and equilibrated at 20 °C. Labeling of the protein was initiated by addition of 2 µL of mPAC (220 µM ± 2.5 mM ApCpp in mPAC SEC buffer). For light-state labeling, the protein stock was pre-equilibrated under constant illumination at 6 mW $cm^{-2}$ from a royal blue collimated LED (Thorlabs, Newton, NJ, USA) before dilution into labeling buffer, and illumination was sustained throughout the labeling reaction. Aliquots of 3 µL were removed after 15, 30, 300, 1200 and 3600 s, and quenched by mixing with 12 µL ice-cold mPAC Q buffer, immediately followed by flash-freezing in liquid nitrogen. The sample was kept in liquid nitrogen until analysis by HPLC-MS. At least three independent labeling reactions were carried out for each data point.
For HPLC-MS analysis, samples were thawed within 1 min by mixing with 80 µL of ice-cold mPAC Q buffer and completely injected into a cooled HPLC-MS setup (0.5 ± 0.1 °C)

using an ice-cold glass syringe. Samples were digested on an in-line pepsin column (Applied Biosystems, Darmstadt, Germany) in 5 % methanol$_{aq}$ and 0.6 % formic acid at 10 °C during a contact time of 4 min at 0.05 mL min$^{-1}$. Resulting peptides were desalted on a 2 cm C18 guard column (Discovery Bio C18, Supelco, Bellefonte, PA, USA) during 1 min at 0.5 mL min$^{-1}$, followed by a binary gradient of 0 to 50 % acetonitrile$_{aq}$ in the presence of 0.6 % formic acid at 0.1 mL min$^{-1}$ over 6 min. Peptides were separated on a reversed-phase column (XR ODS 75 mm × 3 mm, 2.2 µm particle size; Shimadzu, Duisburg, Germany) prior to infusion into a maXis ESI-UHR-TOF mass spectrometer (Bruker, Bremen, Germany).

### HDX-MS of bPAC

HDX-MS for bPAC wild type in the presence and absence of ATP-αS was carried out by Andreas Winkler. Detailed procedures are described in the corresponding publication [115]. The variants bPAC$_{R121S}$ and bPAC$_{Y7F}$ were analyzed in a slightly modified fashion: The stock concentration was 200 µM and the labeling reaction was initiated by addition of 2 µL bPAC stock into 38 µL bPAC HDX labeling buffer. Aliquots of 7 µL were taken after 10, 45, 180, 900 and 3600 seconds, quenched with 7 µL ice-cold bPAC Q buffer and frozen in liquid nitrogen where they were kept until analysis. At least three independent labeling reactions were carried out for each data point. Frozen aliquots of the labeling reaction were thawed within 1 minute by mixing with 90 µL ice-cold bPAC Q buffer and 100 µL were injected into the HPLC. LC-MS conditions were identical to those used for mPAC.

### Data Analysis

All data analysis tasks were performed using Hexicon 2 [117]. The Hexicon 2 result browser was used for visualization and manual curation of results based on inspection of raw mass spectra. Ambiguous peptide sequence assignments were validated using targeted peptide sequencing by collision-induced dissociation tandem mass spectrometry of protonated peptides using the respective digestion protocols described above and a chromatographic separation time of 45 min.

## 2.2.9. Crystallization and Structure Determination

### Crystallization of $mPAC_{AC}$ and $mPAC_{AC+Handle}$

Sparse matrix screens were set up at 20 °C in 96-well vapor diffusion geometry (CrystalQuick LP, Greiner Bio-One, Kremsmünster, Austria), using 70 µL reservoir solution and 200 nL drops containing 100 nL reservoir solution and 100 nL protein at a stock concentration of 450 µM.

Initial crystals were obtained for $mPAC_{AC}$ and $mPAC_{AC+Handle}$ in several conditions of the PEG Suite screen (Qiagen) containing 20 % (w/v) PEG 3350 and 200 mM of various salts. Drops containing crystals were harvested and diluted in total volume of 30 µL 20 % (w/v) PEG 3350, 200 mM $MgCl_2$. Crystals were sheared by vigorous pipetting and vortex mixing, and serial dilutions of this seed stock were prepared. Optimized crystals of $mPAC_{AC}$ were obtained in 20 % (w/v) PEG 3350 and 200 mM LiCl, growing in sitting drops of 330 nL volume, containing 150 nL $mPAC_{AC}$ stock (450 µM), 150 nL reservoir solution and 30 nL of a 1:100 dilution of the seed stock. Crystals appeared overnight and continued to grow for approximately four days. They were harvested 6–10 days after setting up the drops. Small cuboid crystals measuring up to 50 µm in the largest dimension were mounted by Miroslaw Tarnawski and frozen in liquid nitrogen without further cryoprotection. Crystals of $mPAC_{AC+Handle}$ were obtained under identical conditions, however, they were cryoprotected in a mixture of the reservoir solution with 25 % (v/v) ethylene glycol prior to freezing in liquid nitrogen.

The selenomethionine derivative of $mPAC_{AC}$ crystallized in 20 % (w/v) PEG 3350, 200 mM $NH_4I$, using the same strategy as for the native protein and seeding with native $mPAC_{AC}$ as described above. For cryoprotection, 100 mg sucrose/glucose 1:1 mixture were dissolved in 1 mL of water, mixed thoroughly and centrifuged briefly. Crystals were incubated in a 1:1 mixture of the sugar suspension's supernatant with reservoir solution prior to flash-freezing in liquid nitrogen.

A different crystal form of selenomethionine-substituted $mPAC_{AC}$ was obtained by Kai Horny in the Classic II 96-well sparse matrix screen (Qiagen), in a condition containing 20 % (w/v) PEG 3350 and 200 mM Na/K-tartrate. Octahedral crystals of approximately 100 µm in the largest dimension were obtained in hanging drop vapor diffusion geometry (SuperClear 24-well plates, CrystalGen, Commack, NY) using 750 µL reservoir solution and 1 + 1 µL (reservoir + 480 µM protein stock) crystallization drops. The reservoir

contained 20 % (w/v) PEG 3350, 200 mM Na/K-tartrate and 100 mM Bis-Tris pH 5.5. A 1:100 dilution of the drop containing the initial hit was used for streak seeding. Crystals appeared overnight and were harvested after 7 days. Crystals were mounted and flash-frozen in liquid nitrogen by Ilme Schlichting without further cryoprotection.

### Crystallization of bPAC

Crystallization and mounting of the various bPAC variants was done by Elisabeth Hartmann, Ilme Schlichting, Miroslaw Tarnawski and Andreas Winkler. Crystallization conditions and specifics of handling are listed in Table 2.9.

### Data Collection

X-ray data collection was done at the beamline X10SA of the Swiss Light Source (Paul Scherrer Institut, Villigen, Switzerland) at 100 K in a laminar stream of liquid nitrogen. bPAC data were collected by different data collection teams from the Max Planck Institutes in Heidelberg and Dortmund. mPAC data collection was done together with Ilme Schlichting, Thomas Barends and Anton Meinhart.
Native mPAC crystals were measured at 1.0 Å wavelength, using fine φ slicing ($\Delta\varphi = 0.1°$) and a rotation range of at least 180°. Selenomethionine derivatives were used for experimental phasing by single wavelength anomalous dispersion (SAD): X-ray fluorescence was measured at excitation energies of 12.6–12.7 keV and datasets were collected at the fluorescence peak energy using fine φ slicing and a rotation range of 360 °. If crystals were sufficiently large, multiple datasets were recorded at different positions of the crystal.

### Data Processing, Model Building and Refinement

All datasets were indexed, integrated and scaled using XDS [86]. The set of test reflections was inherited whenever appropriate, otherwise 5 % of the reflections were assigned as test reflections.
An initial model of bPAC was obtained by Daniel Frey and Anton Meinhart using molecular replacement with the coordinates of the CyaC AC domain [189] and the BlrB BLUF domain [84]. Further crystal structures (e.g., illuminated crystals or Y7F mutant)

**Table 2.9.:** bPAC crystallization conditions. Unless otherwise indicated, crystals were cryoprotected by brief incubation in 50 mM Tris-Cl pH 7.6, 50 mM $MgCl_2$, 25 % (w/v) PEG 3350, 25 % (w/v) glycerol. Crystallization and crystal handling was done by Elisabeth Hartmann, Ilme Schlichting, Miroslaw Tarnawski and Andreas Winkler [115].

| Variant | Geometry | Protein | Reservoir |
|---|---|---|---|
| bPAC ($P1$), $Ca^{2+}$ form | 2 µL hanging drop | 212 µM | 0.1 M Na-MES pH 6.5, 0.2 M $CaCl_2$, 0.6 M NaCl, 20 % (w/v) PEG 3350, 3 % (w/v) 1,6-diaminohexane |
| bPAC ($P1$) | 2 µL sitting drop | 250 µM + 1 mM ApCpp | 0.1 M Na-MES pH 6.5, 0.6 M NaCl, 20 % (w/v) PEG 3350 |
| bPAC ($P1$), illuminated | as bPAC ($P1$) | as bPAC ($P1$) | as bPAC ($P1$), rotated in front of a blue LED for 10–30 sec before freezing |
| bPAC ($P2_12_12_1$) | 2 µL sitting drop | 212 µM | 0.1 M Na-MES pH 6.5, 0.2 M Mg-acetate, 20 % (w/v) PEG 3000, 3 % (w/v) 1,6-diaminohexane |
| $bPAC_{AC}$ | 0.2 µL sitting drop | 340 µM | 0.1 M Na-Hepes pH 7.0, 0.2 M NaI, 20 % (w/v) PEG 3350 |
| $bPAC_{Y7F}$ | 2 µL sitting drop | 200 µM | 0.1 M Na-Hepes pH 7.0, 20 % (w/v) PEG 3350 |
| $bPAC_{Y7F}$, ATP soak | as $bPAC_{Y7F}$ | as $bPAC_{Y7F}$ | as $bPAC_{Y7F}$, brief soak in 50 mM Tris-Cl pH 7.6, 115 mM $MgCl_2$, 48 mM ATP, 35 % (w/v) PEG 3350 |
| $bPAC_{I8V,I116V}$ | 0.2 µL sitting drop | 200 µM | 0.1 M Tris-Cl pH 8.0, 0.3 M $Mg(NO_3)_2$, 23 % (w/v) PEG 2000. Cryoprotection in 0.1 M Tris-Cl pH 8.0, 50 mM $MgCl_2$, 35 % (w/v) PEG 3350 |
| $bPAC_{Y126F,N257A}$ | 0.2 µL sitting drop | 200 µM | 0.1 M Tris-Cl pH 8.5, 0.2 M $MgCl_2$, 30 % (w/v) PEG 4000 |

were phased by molecular replacement with Phaser [129], searching for individual bPAC chains, followed by simulated annealing in Phenix [1]. Models were refined by iterative manual model building in COOT [45] and refinement in Phenix. Translation, libration, screw (TLS) parameters were refined for three groups per chain, including the BLUF domain (1–126), the handles and the amino-terminus of the AC domain (127–260), and

the carboxy-terminal part of the AC domain (261–350).
Native $mPAC_{AC}$ and $mPAC_{AC+Handle}$ crystallized in space group $C222_1$ and were phased using Phaser for molecular replacement with the coordinates of the isomorphous selenomethionine derivative (see below). In addition, the selenomethionine-derivative of $mPAC_{AC}$ crystallized in space group $P2_12_12_1$ which was not observed in any crystals of the native protein. Se-SAD phasing of $mPAC_{AC}$ was done in a hybrid approach: the selenium substructures were determined using SHELXD [183] and fed to Phenix AutoSol [202, 1] together with the protein sequence. Initial models were improved by iterative cycles of manual model building in COOT and refinement using Phenix. TLS parameters were refined for one TLS group per chain.

## 2.2.10. Small Angle X-Ray Scattering (SAXS)

### Sample Preparation and Data Collection

SAXS samples and corresponding buffers were prepared by dilution and concentration in mPAC SEC buffer (±1 mM ApCpp) using centrifugal filter tubes (Amicon, Merck). The flowthrough was retained as solvent background sample. Protein concentrations were determined spectroscopically by their absorbance at 447 nm. Sample descriptions are listed in Table 2.10. SAXS data collection was done together with Ilme Schlichting, Miroslaw Tarnawski and Thomas Barends at the beamline X12SA of the Swiss Light Source (Paul Scherrer Institut, Villigen, Switzerland). The beamline was set up by Andreas Menzel who provided initial instructions and assisted with data collection and quality control.
Quartz capillaries (inner diameter 1 mm) were mounted in a sample holder and filled with buffer. The sample holder was cooled to 10 °C for all measurements. Scattering of the buffer was determined at 20 positions along the capillary (spacing 0.5 mm), exposing each position with 11.2 keV photons for 0.5 s. Each position was measured 10 times and scattering intensities were averaged. Scattered X-rays were detected by a Pilatus 2M detector (Dectris, Baden-Dättwil, Switzerland) at the end of an evacuated flight tube, 2.13 m from the sample position. Buffer was replaced by the protein sample without moving the sample holder or the capillaries. For measurement of the dark samples, lights in the experimental hutch were switched off and after 5 minutes, scattering of the protein-containing solutions was measured at the same positions using the procedure

**Table 2.10.:** mPAC SAXS Samples. Samples were prepared by dilution and concentration cycles in SAXS buffer ± 1 mM ApCpp. If indicated, samples were pre-illuminated with a royal-blue LED at approximately 5 mW $cm^{-2}$ for 2 min and measured immediately thereafter.

| Sample | Construct | Concentration [mg $mL^{-1}$] | ApCpp | Light |
|---|---|---|---|---|
| Δ– 1 | $mPAC_{\Delta 132}$ | 7.5 | – | – |
| Δ– 2 | $mPAC_{\Delta 132}$ | 6.5 | – | – |
| Δ– 3 | $mPAC_{\Delta 132}$ | 4.7 | – | – |
| Δ+ 1 | $mPAC_{\Delta 132}$ | 9.1 | + | – |
| Δ+ 2 | $mPAC_{\Delta 132}$ | 5.2 | + | – |
| Δ+ L1 | $mPAC_{\Delta 132}$ V160I | 9.1 | + | + |
| Δ+ L2 | $mPAC_{\Delta 132}$ V160I | 6.2 | + | + |
| Δ+ L3 | $mPAC_{\Delta 132}$ V160I | 4.1 | + | + |
| F– 1 | mPAC | 9.3 | – | – |
| F– 2 | mPAC | 3.8 | – | – |
| F+ 1 | mPAC | 11.0 | + | – |
| F+ 2 | mPAC | 8.4 | + | – |

described above. Samples containing the variant $mPAC_{\Delta 132}$ V160I (dark state recovery half-life 2900 s) were pre-illuminated in the capillaries with a royal-blue LED (Thorlabs) at 5 mW $cm^{-2}$ for 2 min and measured immediately thereafter.

### Data Analysis

Scattering intensities were radially integrated and the distance from the beam center was converted to momentum transfer $q = 4\pi \sin\theta\lambda^{-1}$ [$Å^{-1}$] with $\theta$ being the scattering angle. Scattering intensities at the 20 measured positions were visually inspected and positions containing obvious outliers in the buffer or protein sample were omitted from analysis. Remaining positions were averaged into one scattering curve and the buffer scattering intensity was subtracted from the protein solution's scattering intensity.
Data processing and analysis were done using tools from the ATSAS [49] package and the SCATTER program developed by Robert Rambo (www.bioisis.net). Data were normalized to their concentration using `datop`, truncated in the very low and very high angular range using `datcrop`, followed by scaling and merging, as well as extrapolation of scattering intensity to zero concentration by `almerge` using at least 50 overlapping data

points from each matching dataset. Distance distribution functions P(r) were calculated with `gnom` and with the regularized indirect Fourier transform of SCATTER using 3500 refinement rounds and a regularization parameter $\lambda$ of 0.1. The spherical radius of gyration $R_g$ was determined from the Guinier plot and from the P(r) distribution [53] Porod volumes and the resulting density assuming a protein dimer were calculated using `datporod`. Ab-initio shape reconstruction was done for $mPAC_{\Delta 132}$ + ApCpp using `gasbor` and the P(r) distribution returned from `gnom` processing.

## 2.3. Rosetta Modeling

### 2.3.1. Preparation of Benchmarking Dataset

A set of high-quality experimentally determined structures were obtained from the PDB to evaluate the performance of the method. The PDB was filtered by the query shown in Listing 2.2 on 25 July 2014 and the resulting 11,957 structures were further filtered. PFAM to PDB mappings containing 229,425 annotated domains in 93,271 unique structures were obtained from PFAM and used to search for PDB files that have exactly two annotated domains that do not overlap by more than 20 % of their total sequence. This resulted in 1,752 PDB files. To avoid bias from multiple copies of identical or highly similar proteins, sequences were extracted from the remaining PDB files and pairwise sequence alignments were generated with ClustalW2 [107]. Sequences with an (arbitrary) distance below 0.1 were clustered together using a greedy agglomerative strategy, resulting in 785 clusters. Domain boundaries for these clusters were assigned based on the PFAM annotation and curated manually by visual inspection in PyMol [180]. Clusters with high structure similarity to other clusters, no obvious interdomain linker or cross-domain disulfides were discarded, resulting in a final set of 158 PDB files (Table A.7).

In addition to the structures selected here, a set of structures from a previous domain assembly benchmarking study [218] was included. As the publication contained only linker lengths but not the actual linker positions, linkers were manually re-assigned based on visual inspection, attempting to preserve the published linker length whenever possible. Duplicate (homologous) structures were removed, resulting in a final set of 58 PDB files (Table A.7). All PDB files were split into N and C-terminal domains, entirely

discarding linker coordinates.

**Listing 2.2:** PDB advanced search query to generate the initial set of benchmarking structures. The query was run on 25 July 2014 and returned 11,957 structures.

```
Chain Type:
  there is a protein chain but not and DNA or RND or Hybrid
Number of Chians Search :
  Min Number of Chains=1
  Max Number of Chains=4
Oligomeric state Search :
  Min number of oligomeric state=1
  Max number of oligomeric state=1
CloseContactsQuery:
  derivedStructConn.count.comparator=
    between derivedStructConn.count.min=0
    derivedStructConn.count.max=10 # limit number of disulfides
Secondary structure has:
  between 15 - 75 percent of elements are Alpha Helical
  between 15 - 75 percent of elements are Beta Sheet
Molecular Weight Search:
  Min Molecular Weight=15000
  Max Molecular Weight=80000
Experimental Method is X-RAY
Resolution is between 0.0 - 2.6
```

## 2.3.2. Domain Assembly

Dedicated `DomainAssemblyMover` and `DomainAssemblyJobInputter` classes were added to the C++ codebase of Rosetta for the task of domain assembly modeling. The `DomainAssemblyJobInputter` constructs a starting pose from the coordinates of the N- and C- terminal domains and the linker sequence: the sequence of the linker is read from a configuration file and individual amino acids are appended to the carboxy-terminus of the N-terminal domain with extended chain parameters ($\varphi$ = -150° and -60° for proline, $\psi$ = 150°, $\omega$ = 180°) and ideal bond lengths and angles. The C-terminal domain is appended to the newly inserted domain linker using ideal extended chain

geometry at the fusion site. This starting pose is fed to the `DomainAssemblyMover` that uses a two-step protocol consisting of a coarse-grained centroid stage and a fullatom refinement. Three strategies of different complexity are available for the centroid stage (cf. below), two of which require 9-mer or 3-mer fragments. All strategies were run to produce 5000 decoys unless otherwise indicated.

### Fragment Generation

Five-hundred 3-mer and 200 9-mer fragments were generated for each linker residue using Rosetta's `fragment_picker` application [58] and the precompiled `vall.jul19-.2011` fragment database. Fragment picking was based on secondary structure and solvent accessibility predictions from `psipred` [130] and `sparks-x` [226]. Homologous sequences were found using `psiblast` [4]. Close homologs were excluded at an E-value threshold of 0.05 against all sequences in the PDB. Score weights for fragment selection are shown in Listing 2.3.

**Listing 2.3:** Score weights for the Rosetta `fragment_picker` application.

```
# 3-mers
# score name         priority  wght  min_allowed  extras
ProfileScoreL1            700   1.0     -
ProfileScoreStructL1      100   1.4     -
SolventAccessibility      500   0.5     -
Phi                       300   3.9     -
Psi                       200   0.9     -
SecondarySimilarity       600   1.0     -        psipred
RamaScore                 400   6.0     -        psipred
FragmentCrmsd               0   6.0     -

# 9-mers
# score name         priority  wght  min_allowed  extras
ProfileScoreL1            700   1.0     -
ProfileScoreStructL1      100   4.0     -
SolventAccessibility      500   1.5     -
Phi                       300   1.0     -
Psi                       200   0.6     -
SecondarySimilarity       600   1.0     -        psipred
```

| | | | | |
|---|---|---|---|---|
| RamaScore | 400 | 0.8 | - | psipred |
| FragmentCrmsd | 0 | 0 | - | |

### Centroid Stage

Three different strategies are available for the centroid stage: *(i)* fragment-free backbone moves, *(ii)* 3-mer insertion followed by small backbone moves and *(iii)* a full ab initio protocol using 9-mer and 3-mer insertion. All strategies use a centroid representation of the pose in which the backbone remains in a fullatom representation while each side chain is replaced by a pseudo Cβ-atom that has residue-specific properties (position, size, polarity). The fragment-free protocol uses a smooth centroid score function (`score3_smooth`) and repeatedly perturbs the linker torsion angles by up to 180° using Rosetta's `SmallMover`. The number of repetitions was set to 50 (outer) × 500 (inner). A Metropolis Monte Carlo algorithm (`TrialMover` with $kT$ = 0.8) is used to generate a canonical ensemble. After the said number of repetitions, the best-scoring pose is recovered. This coarse sampling is followed by another 30 × 200 torsion angle perturbations of up to 2° in helical segments and 4° otherwise, before the best-scoring pose is recovered and subjected to fullatom refinement. The 3-mer insertion protocol inserts 3-mer fragments from the PDB into the linker (cf. section 2.3.2) using 50 × 500 repetitions of the `ClassicFragmentMover` and scores them using the `score3_smooth` scorefunction and a Metropolis Monte Carlo algorithm with $kT = 0.8$. Fragment insertion is followed by 30 × 200 rounds of torsion angle perturbations (`SmallMover`) up to 2° in helical segments and 4° otherwise, before the best-scoring pose is recovered and subjected to fullatom refinement. The full ab initio protocol relies on the `ClassicAbinitio` protocol reviewed in [166]. The protocol is run with default parameters and score functions, except for a five-fold increased number of cycles (`-abinitio:increase_cycles 5`).

### Fullatom Refinement

The pose is switched to fullatom representation, restoring sidechains and applying the `ref2015` score function. Severe clashes are resolved by global coarse repacking of all sidechains using a simulated-annealing strategy and a rotamer library [43]. Following global repacking, only the linker and the domain interface sidechains are relaxed in four repetitions of the following simulated annealing procedure: Drop the repulsive score

term to 10 % and ramp it back up during seven steps while repeatedly optimizing all linker torsion and side chain angles as well as interface side chain angles by gradient-based minimization (`MinMover` using the `lbfgs_armijo_nonmonotone` algorithm and a convergence threshold of 0.001), followed by rotamer library [43] based side chain repacking.

### 2.3.3. Evaluation of Results

Decoys were evaluated by their RMSD from the experimentally determined (*native*) structure. To evaluate the performance of the protocol in blind prediction cases, the Rosetta score was taken into account to generate the following score metrics: best RMSD of the five best scoring decoys (b5), best RMSD of the best one percent of decoys (b1p).

RMSD-based scores were compared to a total of 38 structural features (Table 2.11) to learn which properties of a structure influence the performance of the domain assembly protocol. To test for association between the features and prediction accuracy, assembly problems were grouped according to their b5 score: those having a b5 score of less than 3 Å were marked as successfully solved whereas problems with a b5 score of more than 5 Å were marked as clearly failed. Borderline cases were not considered. The nonparametric Kolmogorov-Smirnov test was applied to determine whether the empirical cumulative distribution functions of a feature value shows significant differences between solved and failed assembly problems. Resulting p-values were not adjusted for multiple testing due to the high mutual dependence of the test variables (e.g., the number of interface residues, hydrogen bonds and the $\Delta\Delta G$ value of an interface). Instead, redundant test variables were used to validate each other.

The fraction of decoys with an RMSD below 3 Å from native was defined as *empirical hit rate* and used to estimate from a negative binomial distribution how many decoys have to be sampled in order to have a 95 % chance of obtaining at least one decoy with less than 3 Å RMSD from the native structure. This number was then multiplied with a Monte Carlo based estimate of the average runtime for each strategy and each assembly problem to obtain an *average waiting time*.

**Table 2.11.:** Structural properties determined for analysis of domain assembly performance. Properties were compared to the quality of domain assembly predictions to learn if they are predictors of modeling performance. Secondary structure was determined by DSSP [87]. Surface areas, ΔΔG values and the number of interdomain interactions were determined by PISA [103]. Interface clashes and rotamer outliers were determined by the the MolProbity [18] tools `clashscore` and `rotalyze`, respectively, as implemented in the Phenix environment [1].

| Name | Description |
|---|---|
| **Linker Sequence Properties** | |
| LENGTH | Linker length |
| PDLI | Profile domain linker propensity index [39] |
| ARMADILLO | Armadillo linker propensity score [42] |
| **Compactness Measures** | |
| CA_DIST | Average Cα distance in the linker |
| RG_FULL | Radius of gyration (entire chain) |
| RG_FULL_NORM | Radius of gyration normalized by (chain length)$^{0.4}$ |
| RG_LINKER | Linker radius of gyration |
| RG_LINKER_NORM | Linker radius of gyration normalized by (linker length)$^{0.4}$ |
| CHAIN_VOLUME | Total volume of protein chain |
| LINKER_VOLUME | Total volume of the linker |
| **Secondary Structure** | |
| D_HELIX | Helix fraction of both domains |
| D_STRAND | Strand fraction of both domains |
| L_HELIX | Helix fraction of the linker |
| L_STRAND | Strand fraction of the linker |
| **Domain Interface** | |
| BSA_AB | Buried surface area between domains |
| BSA_L | Buried surface area between both chains and linker |
| DDG_AB | ΔΔG of the domain interface |
| DDG_L | ΔΔG of the linker-domain interfaces |
| NHB_AB | Number of hydrogen bonds between the domains |
| NHB_L | Number of hydrogen bonds between the linker and the domains |
| NSB_AB | Number of salt bridges between the domains |
| NSB_L | Number of salt bridges between the linker and the domains |
| LARGEST_X_INTERFACE | Largest crystallographic interface |
| BEST_X_DDG | Lowest crystallographic interface ΔΔG |
| SASA_AB | Total solvent accessible surface area of the domains |
| SASA_L | Solvent accessible surface area of the linker |
| EXP_RES_AB | Total number of solvent-exposed residues in the domains |
| EXP_RES_L | Total number of solvent-exposed residues in the linker |
| INTERFACE_RES_AB | Number of interface residues between the domains |

*(continued on next page)*

**Table 2.11** *(continued)*

| Name | Description |
|---|---|
| INTERFACE_RES_L | Number of interface residues between the linker and the domains |
| INT_CLASH | Number of clashes (overlap $> 0.4$ Å) in the domain interface |
| INT_ROTAMER_OUTLIER | Number of rotamer outliers in the domain interface |
| **Fragment Quality** | |
| FRMS_9 | 2.5 % quantile of 9-mer RMSD to the respective experimental peptide |
| DD_CA_9 | 2.5 % quantile of 9-mer end-to-end distance difference to the respective experimental peptide |
| DD_ALPHA_9 | 2.5 % quantile of 9-mer end-to-end vector angle difference to the respective experimental peptide |
| FRMS_3 | 2.5 % quantile of 3-mer RMSD to the respective experimental peptide |
| DD_CA_3 | 2.5 % quantile of 3-mer end-to-end distance difference to the respective experimental peptide |
| DD_ALPHA_3 | 2.5 % quantile of 3-mer end-to-end vector angle difference to the respective experimental peptide |

# Results

## 3.1. Structural Mechanism of bPAC Photoactivation

### 3.1.1. Background

When I joined this project, the crystal structure of bPAC had been determined by Ilme Schlichting, Elisabeth Hartmann, Daniel Frey and Anton Meinhart in our department. A large number of further crystal structures of mutants, nucleotide-treated or illuminated crystals followed; however structural differences were minor between the various treatments and no systematic analysis taking advantage of the large number of crystallographic datasets had been carried out. Mutants had been devised by Ilme Schlichting and Andreas Winkler to probe structural hypotheses by steady-state kinetic measurements. Finally, hydrogen–deuterium exchange mass spectrometry (HDX-MS) experiments had been carried out by Andreas Winkler, to investigate secondary structure dynamics in the dark and under illumination, in the presence and absence of nucleotide.

During the preparation of a manuscript on the structure and function of bPAC, an article was published, describing the crystal structure of a homologous BLUF-AC from *Oscillatoria acuminata* (OaPAC) [149]. This structure is highly similar to bPAC and shows the same arrangement of the BLUF and AC domains. The authors proposed that the central coiled coil transmits structural changes from the activated BLUF domain to the AC domain, however, they investigated only fragments of the signaling pathway and provided no specifics on the structural changes emerging from the BLUF domain and how these lead to an active AC structure. The central questions thus remained: How do light-induced changes in hydrogen bonding near the flavin chromophore propagate to the BLUF domain surface and how do these changes contribute to the formation of an active AC conformation?

To address these questions, I re-analyzed present and newly collected crystallographic data to create consistent statistics and re-refined structural models with the aim to

minimize phase bias. I used superposition-independent methods to analyze structural differences between the crystal structures and to identify mechanical hinges and levers along the signaling pathway. I re-analyzed steady-state AC kinetics and performed time-resolved spectroscopy to characterize structure-based point mutations with twofold functional data: AC activity and BLUF light state lifetime. Hypotheses from kinetic experiments were corroborated by additional HDX-MS measurements that I carried out on selected variants.

The combination of structural and functional analysis allowed me to propose a more complete picture of BLUF signaling and helped identify previously overlooked elements of the AC domain that unify all hitherto proposed models of AC activation. Large parts of this chapter have been published in the following peer-reviewed article:

**R. Lindner**, E. Hartmann, M. Tarnawski, A. Winkler, D. Frey, J. Reinstein, A. Meinhart, I. Schlichting, *Photoactivation Mechanism of a Bacterial Light-Regulated Adenylyl Cyclase*, J. Mol. Biol. 429 (9) (2017) 1336–1351

Experimental data has been jointly collected and analyzed by the authors of the article and is marked accordingly. All text and figures used in this thesis were prepared by myself whereas the final article, as published in the *Journal of Molecular Biology*, has undergone major editing by Ilme Schlichting.

### 3.1.2. Enzymatic and Spectroscopic Properties of bPAC

bPAC was cloned, expressed and purified by Elisabeth Hartmann who also carried out enzymatic measurements and HPLC analysis. bPAC was obtained as soluble homodimeric protein bearing three additional amino acids after proteolytic removal of the amino-terminal hexahistidine tag. The BLUF domain binds a flavin cofactor and exhibits the characteristic UV/Vis absorption spectrum with a maximum at 440 nm [54]. Illumination with blue light induces a 15-nm red shift that decays with a half-life of 19.5 s in the dark (Figure 3.1).
Similar to previous reports [171, 193, 192], bPAC could only be enriched to approximately 60 % purity. Pronounced variation of determined catalytic rate constants did not allow absolute quantification of cAMP catalytic rates and hampered comparison of different bPAC variants. Replicates measured on the same day showed low variation (Table A.1), hence differences in protein purity and experimental inter-day variation were identified as the major sources of error that needed to be adjusted for. Determined rate constants were divided by the fraction of bPAC in the preparation (as determined by densitometric analysis of SDS-PAGE gels, Table A.2) and expressed on a qualitative scale, relative to the corresponding wild type's rate constant measured under identical conditions on the same day (Table 3.1). Details of the correction procedure and raw data are shown in Appendix A.1. The average corrected apparent dark state AC rate constant at 100 µM ATP is 0.008 ± 0.001 $s^{-1}$ (standard error of the mean, SEM, n=10) and increases 300-fold to 2.6 ± 0.3 $s^{-1}$ (SEM, n=10) under constant illumination at 7 mW $cm^{-2}$.

Two principal mechanisms can be devised for light-stimulation of AC activity: the BLUF domain can inhibit an intrinsically high AC activity in the dark or it can promote the formation of a catalytically active conformation in the light, that an intrinsically inactive AC cannot assume on its own. To address these two possibilities, the $bPAC_{AC}$ construct containing the isolated AC domain was expressed and characterized. Further constructs containing parts of the interdomain linker did not express ($bPAC_{\Delta 100}$ and $bPAC_{\Delta 105}$) or were poorly soluble ($bPAC_{\Delta 113}$, $bPAC_{\Delta 128}$). No cAMP production was detected by any of these constructs from which can be concluded that the isolated AC domain of bPAC is inactive. The BLUF domain is therefore a true allosteric activator of the AC domain in bPAC.

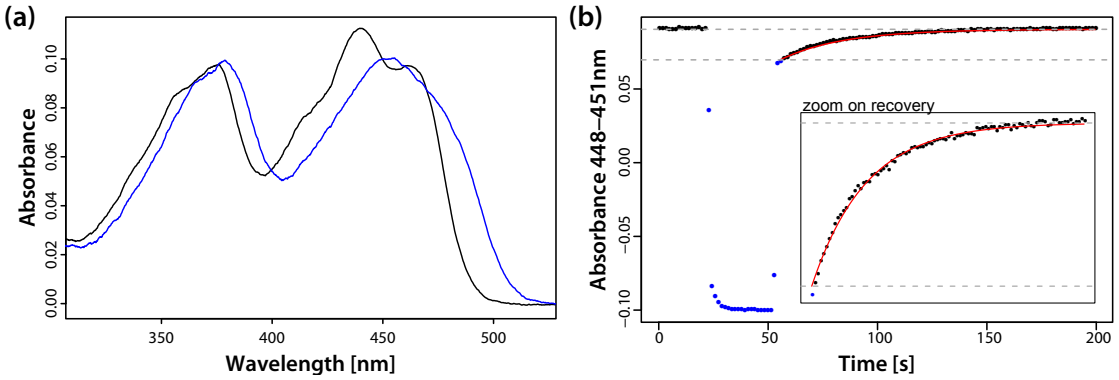



**Figure 3.1.:** Dark state recovery of bPAC. **(a)** Absorption spectrum of bPAC in the dark (black) and red-shifted spectrum immediately after illumination (blue). **(b)** Time trace of absorbance band between 448 and 451 nm. First segment (black) indicates the steady state absorbance before illumination, the segment colored in blue indicates absorbance during illumination with blue light and the following segment (black) shows recovery of dark state absorbance after the light has been shut off. The red line shows a monoexponential fit to the recovery trace (see inset for zoom). The recovery half-life of bPAC is 19.5 s (Table 3.1).

### 3.1.3. Structure of bPAC in the Dark

Wild type bPAC crystallized in space groups $P2_12_12_1$ and $P1$ with a dimer in the asymmetric unit. The initial structure determined by Ilme Schlichting, Daniel Frey and Anton Meinhart was used as template for all subsequent data processing and refinement. The two crystal forms differ mainly in the relative arrangement of the AC domains (described below). Data collection and refinement statistics are shown in Table A.4.

The description of structural features common to both crystal forms is based on the triclinic crystal form as it diffracted to significantly higher resolution (1.8 Å) than the orthorhombic one (3.3 Å). bPAC forms a parallel dumbbell-shaped dimer with an antiparallel arrangement of N-terminal BLUF and C-terminal AC domains, respectively (Figure 3.2). The core BLUF domains are followed by C-terminal capping helices ($\alpha3_{\mathrm{BLUF}}$) that form a coiled coil between both BLUF β-sheets (Figure 3.3). The C-terminal part of the coiled coil and a handle-like platform (made up by $\alpha4_{\mathrm{BLUF}}$ and the loop $\alpha4_{\mathrm{BLUF}}$–$\beta1_{\mathrm{AC}}$) form a helical bundle on which the AC domains reside. There is no direct interface between the core BLUF and AC domains.

BLUF domains show a highly conserved fold in the core and vary mostly in the arrangement of their C-terminal extensions [125]. bPAC and OaPAC [149] are unique in using $\alpha3_{\mathrm{BLUF}}$ for dimerization through formation of an intermolecular coiled coil. As

**Table 3.1.:** Kinetic and spectroscopic properties of bPAC wild type and variants. $\lambda_{max}$ denotes the BLUF absorption maximum in the blue range (near 440 nm). AC activity is given qualitatively with respect to the wild type in the light: "+++", "++", and "+" indicate activities of $\geq$ 50 %, $\geq$ 10 %, and $\geq$ 5 % of the corresponding wild type sample in the light, respectively, and "–" indicates less than 5 % of wild-type light state activity. Underlying kinetic data is shown in Table A.3.

| Construct | AC Activity | | $\lambda_{max}$ [nm] | | Recovery | Description |
|---|---|---|---|---|---|---|
| | Dark | Light | Dark | Light | $t_{1/2}$ [s] | |
| **wild type** | – | +++ | 440 | 455 | $19.5 \pm 0.4$ | Wild type |
| **Y7F** | ++ | ++ | 436 | 438[a] | n/a | Pseudolit, decouples $\beta1_{BLUF}$ from flavin |
| **I8V** | – | +++ | 440 | 446 | $11.8 \pm 0.9$ | Packing between $\beta1_{BLUF}$ and $\alpha3_{BLUF}$ |
| **I116V** | – | +++ | 440 | 450 | $16.7 \pm 0.9$ | Packing between $\beta1_{BLUF}$ and $\alpha3_{BLUF}$ |
| **I8V/I116V** | – | +++ | 440 | 453 | $6.6 \pm 0.3$, $37 \pm 2$ [b] | Packing between $\beta1_{BLUF}$ and $\alpha3_{BLUF}$ |
| **R121S** | ++ | +++ | 448 | 449 | $13.0 \pm 0.5$ | $\alpha3_{BLUF}$ sidechain hydrogen bonds, interaction with $\beta4_{BLUF}$ |
| **K78/T115C** | ++ | ++ | 442 | 453 | $61 \pm 4$ | $\beta4_{BLUF}$ crosslink with capping helix |
| **K78/T115C +DTE** | + | +++ | 442 | 453 | $61 \pm 4$ | $\beta4_{BLUF}$ crosslink with $\alpha3_{BLUF}$ |
| **L123C (±DTE)** | – | +++ | 443 | 452 | $15.0 \pm 0.5$ | Crosslink within $\alpha3_{BLUF}$ |
| **bPAC 1-103** | n/a | n/a | 443 | 447 | $7.7 \pm 0.2$ | BLUF domain without $\alpha3_{BLUF}$ helix |
| **bPAC 1-127** | n/a | n/a | 440 | 448 | $7.0 \pm 0.2$ | BLUF domain with $\alpha3$ capping helix |
| **bPAC 145-350** | –[c] | n/a | n/a | n/a | n/a | Isolated AC domain |
| **Y126F** | – | – | 440 | 448 | $8.7 \pm 0.1$ | Coupling of $\alpha3_{BLUF}$ to handle/tongue |
| **P141G** | –[c] | – | 441 | 446 | $14.1 \pm 0.8$ | Coupling of $\alpha3_{BLUF}$ to handle/tongue |
| **N257A** | – | ++ | n.d. | n.d. | n.d. | Coupling of $\alpha3_{BLUF}$ to handle/tongue |
| **N257K** | – | ++ | 440 | 453 | $26 \pm 3$ | Coupling of $\alpha3_{BLUF}$ to handle/tongue |
| **Y126F/N257A** | – | +++ | 440 | 452 | $22.4 \pm 0.4$ | Coupling of $\alpha3_{BLUF}$ to handle/tongue |

[a] $\lambda_{max}$ of $bPAC_{Y7F}$ shifts slightly, likely due to the emergence of a long-lived radical species that abosrbs above 500 nm

[b] $bPAC_{I8V,I116V}$ recovers its dark state spectrum in a process that is best described by two exponential with 35 % (fast) and 65 % (slow) amplitude

[c] No activity could be detected for $bPAC_{145\text{-}350}$ and $bPAC_{P141G}$ under various conditions

the central coiled coil constitutes the only link between the BLUF and AC core domains, its interactions with both domains were investigated in greater detail. The coiled coil is stabilized by hydrophobic interactions between heptad repeats in the helices (Ile106, Ile109, Leu113, Ile116, Leu123) and by insertion into a trough between two rows of hydrophobic side chains that line up on the β-sheet of the opposing BLUF domain (Figure 3.3b). This suggests a tight packing between the $\alpha3_{BLUF}$ helices and the BLUF β-sheet, however alternative conformations of Val39, Leu41 (both $\beta2_{BLUF}$) and Ile116

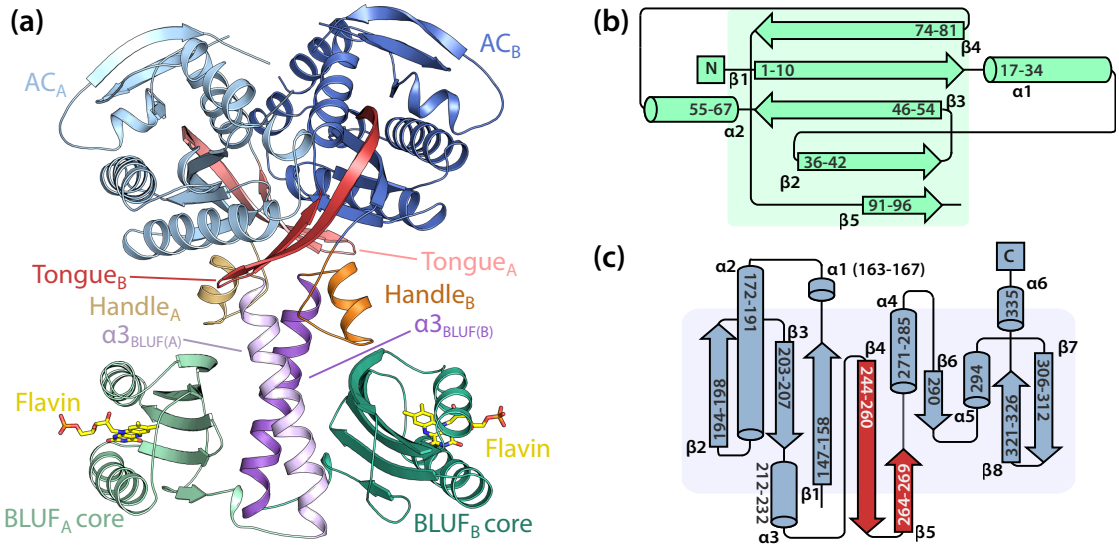



**Figure 3.2.:** Structure and domain architecture of bPAC. **(a)** bPAC is a parallel homodimer (chains A and B colored in light and dark shades, respectively) in which the individual domains face each other in an antiparallel fashion. The N-terminal flavin-binding BLUF core domains are colored in green. The BLUF domains dimerize through their $\alpha3_{BLUF}$ capping helices (purple) that form an intermolecular coiled coil. $\alpha4_{BLUF}$ and the loop $\alpha4_{BLUF}$–$\beta1_{AC}$ make up a handle-like structure (orange) and form a four-helix bundle with the C-termini of the capping helices. The AC domains (blue) reside on the helix bundle and interact with it mainly through a tongue-like extrusion (red) consisting of the $\beta4_{AC}$–$\beta5_{AC}$ hairpin that reaches to the other AC subunit and is sandwiched between the helix bundle and the AC core domain of the other chain. **(b)** Domain topology of the BLUF domain. **(c)** Domain topology of the AC domain.

($\alpha3_{BLUF}$) indicate the presence of packing defects and conformational freedom in the central part of the coiled coil.

One rotamer of Ile116* (asterisks indicate residues in the dimer-related chain) is in van der Waals contact to His120 that is part of an extensive hydrogen bond network around Glu124 (Figure 3.3c). Its carboxylate group interacts with His120* on one side and forms a two-pronged salt bridge with the guanidinium group of Arg121 on the other side. The latter is anchored at the backbone carbonyl oxygen of Asn73 ($\beta4_{BLUF}$), and His120 stabilizes the coiled coil through a hydrogen bond to Ser119*. Leu75 and Leu77, located in a conserved kink in $\beta4_{BLUF}$, insert between the helices of the $\alpha3_{BLUF}$ coiled coil and are in van der Waals distance of Glu124 and His120 (Figure 3.3c). Hydrophobic packing of the C-terminal part of $\alpha3_{BLUF}$ and the handles results in formation of a helical bundle that interacts with the $\beta4_{AC}$–$\beta5_{AC}$ (*tongue*) region lining the AC active site.

The AC domains exhibit the canonical type III fold and face each other in an antiparallel

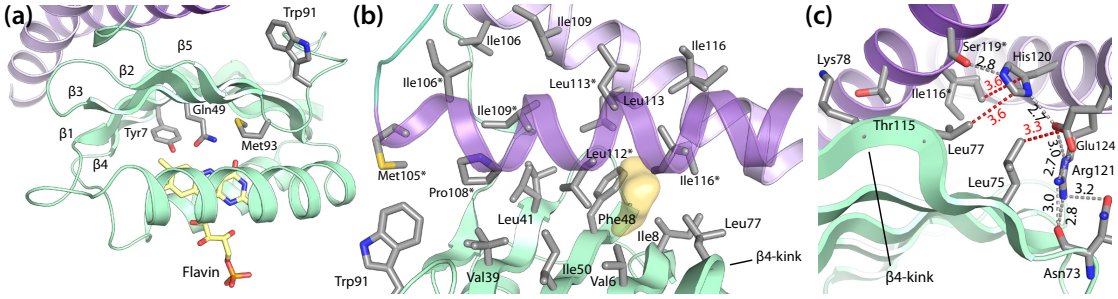



**Figure 3.3.:** Close-up view of the BLUF domain. **(a)** BLUF core (green) with the conserved Tyr7, Gln49, Trp91 and Met93. Trp91 points out of the flavin-binding pocket, while Met93 points inwards and interacts with Gln49 ($Trp_{out}$ conformation). **(b)** The $\alpha3_{BLUF}$ capping helix (purple) inserts between two rows of hydrophobic residues on the surface of the opposing BLUF domain's β-sheet. Val39, Leu41 and Ile116 have alternative side chain conformations. One rotamer of Ile116* (asterisks indicate amino acids in the dimer-related chain) leaves a void of 6.7 Å$^3$ (outlined in yellow) lined by Val6, Ile8 ($\beta1_{BLUF}$, flanking the photoactive Tyr7), Phe48 ($\beta3_{BLUF}$, opposite of the essential Gln49), and Leu112* in the coiled coil (7.6 Å$^3$ for the corresponding rotamer of Ile116). The other rotamer of Ile116* packs against Leu113 in the coiled coil and Ile8 ($\beta1_{BLUF}$), Phe48 ($\beta3_{BLUF}$), and Leu77 ($\beta4_{BLUF}$) on the face of the BLUF β-sheet. **(c)** Side chain hydrogen bond network around Glu124 at the C-terminus of $\alpha3_{BLUF}$. $\beta4_{BLUF}$ interacts extensively with this hydrogen bond network through the Asn73 carbonyl and van der Waals interactions of Leu75 and Leu77. One of the Asn73 side chain conformations further interacts with Arg121. Distances of polar and charged interactions are given in black; van der Waals contacts are indicated in red.

arrangement to form two active sites in the dimer interface. Each active site is made up by the metal-binding residues Asp157 and Asp201, the ribose-contacting Asn274*, and the major determinants of base specificity Lys197* and Thr267* [118, 204]. Several features of the AC active site indicate that the observed conformation represents an inactive state: the nucleotide-binding elements $\alpha4_{AC}$, $\alpha4^*_{AC}$ and $\beta5_{AC}$ are very close to each other and to the $\beta2_{AC}$–$\beta3_{AC}$ hairpin (containing Asp201), obstructing binding of the adenine base and ribose. Thr267 is turned away from the nucleotide binding site and the side chain of Lys197 interacts with the carbonyl oxygens of Phe198* and His266, locking it in a conformation unable to interact with nucleotide.

A tongue-like extrusion ($\beta4_{AC}$–$\beta5_{AC}$) extends away from the AC domain into the other monomer where it is located between the C-terminus of $\alpha3_{BLUF}$, $\alpha2_{AC}$ and $\beta2_{AC}$, i.e., between the handle and the AC core (Figure 3.4). The homodimeric arrangement of bPAC in the triclinic crystals displays conformational asymmetry (1.4 Å RMSD between chains A and B) that is most pronounced in the interaction of the tongues with the handles. The tongues are either rotated "down" toward the BLUF domain, interacting with the handles ($tongue_{DOWN}$ conformation), or detached from the handles ($tongue_{UP}$,

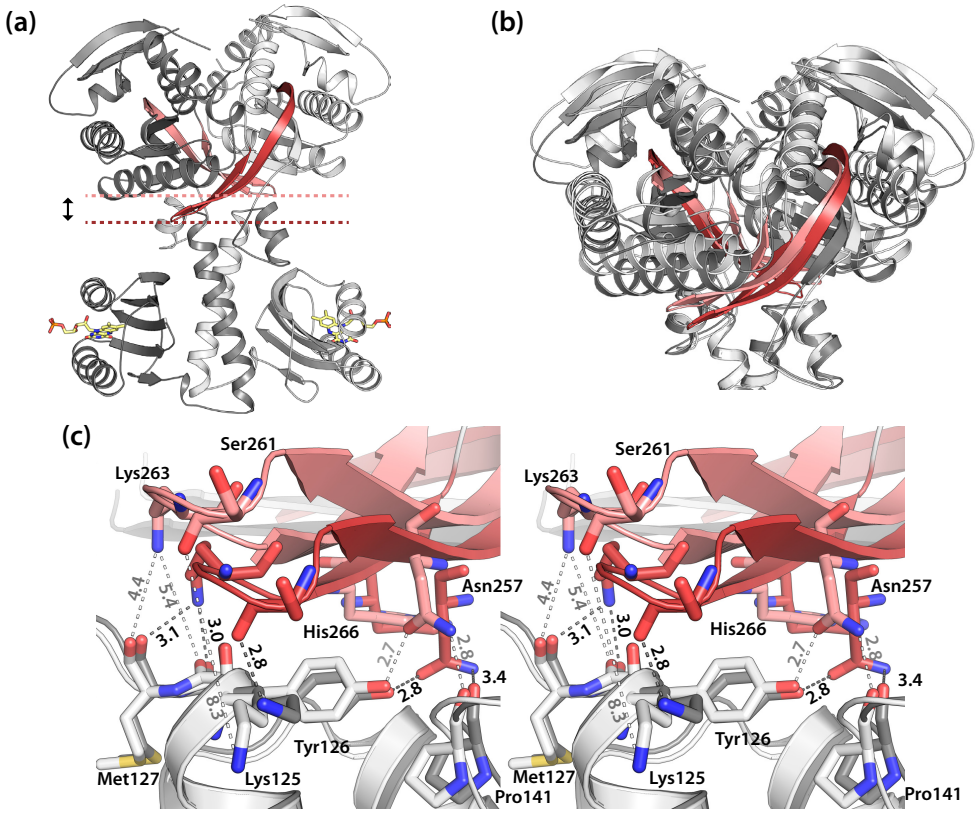



**Figure 3.4.:** Asymmetry of $\beta4_{AC}$–$\beta5_{AC}$ tongues in the *P*1 crystal structure. **(a)** Crystal structure color-coded by tongue conformer: the $tongue_{UP}$ conformer (light gray and red) comprises the handle and tongue region of chain A together with the remaining BLUF and AC domains of chain B. The tongue and handle of chain B together with the remaining parts of chain A make up the $tongue_{DOWN}$ conformer (dark gray and red). Dashed lines indicate the location of the tongue in the $tongue_{UP}$ (light red) and $tongue_{DOWN}$ (dark red) conformation, respectively. **(b)** Dimers of the $tongue_{UP}$ (light gray and red) and $tongue_{DOWN}$ (dark gray and red) conformer were modeled, and their $\alpha3_{BLUF}$ helices were superimposed. The AC domains are closer to the BLUF domain in the $tongue_{DOWN}$ conformation than in the $tongue_{UP}$ conformation and open in a wedge-like manner. **(c)** Stereoscopic close-up of the interactions between the tongue and the helix bundle in the $tongue_{UP}$ (light colors) and $tongue_{DOWN}$ (dark colors) conformations. Distances indicated in gray correspond to the $tongue_{UP}$ conformation, and distances indicated in black correspond to the $tongue_{DOWN}$ conformation. The distance between the AC and the BLUF domains is shorter in the $tongue_{DOWN}$ conformation than in the $tongue_{UP}$ conformation (e.g., Cα distance of Leu262–Asn73* is 11.4 Å versus 9.7 Å for Leu262*–Asn73).

Figure 3.4a and b). Both chains exhibit the $tongue_{DOWN}$ conformation in the orthorhombic crystals.

This structural flexibility is likely of functional relevance as the $\beta4_{AC}$–$\beta5_{AC}$ constitute the only direct contact between the $\alpha3_{BLUF}$ capping helices, the handles and the AC core. As the tongue is firmly attached to the opposing core's AC core domain, the angle at which

it extends from the AC domain determines the relative position of the other chain. The $\text{tongue}_{\text{DOWN}}$ is therefore linked to clamshell-like opening of the AC domain interface. Opening of the AC domain can be quantified in terms of the domain opening angle that is defined as the angle between the Cα atoms of Ala277 ($\alpha4_{\text{AC}}$), Pro146 (denoted the *hinge*) and Ala277*. The angle varies slightly between the two possible choices of Pro146 (chain A or B, denoted $\text{hinge}_{\text{A}}$ and $\text{hinge}_{\text{B}}$, respectively), and differences increase with the asymmetry of the dimer. The opening angles of the described $P1$ and $P2_12_12_1$ structures measure (26.9° / 28.4°) and (32.9° / 33.3°) at $\text{hinge}_{\text{A}}$ / $\text{hinge}_{\text{B}}$, respectively.

## 3.1.4. Structure of Illuminated bPAC Crystals

bPAC did not crystallize under constant illumination, hence dark-grown bPAC crystals were illuminated for different times prior to cryo-cooling to learn about the structural changes resulting in AC activation. In general, illuminated crystals diffracted to lower resolution than crystals from the same batch kept in the dark (Table A.4).

While no gross overall structural changes occur upon illumination (Cα RMSD 0.88 Å), there are significant differences in quaternary structure consistent with a transition of the UP position of the $\beta4_{\text{AC}}$–$\beta5_{\text{AC}}$ tongue of chain A toward the DOWN position, while the tongue of chain B remains in the DOWN position (Figure 3.5). This asymmetric $\beta4_{\text{AC}}$–$\beta5_{\text{AC}}$-mediated motion of the AC domain toward the BLUF domains increases the AC domain opening angle from 26.9° / 28.4° to 28.7° / 28.8° for $\text{hinge}_{\text{A}}$ / $\text{hinge}_{\text{B}}$, respectively, moving the AC domain centers of mass of 0.7 Å apart. Although only the tongue of chain A shows a significant conformational change of the backbone, the side chain of His266 ($\beta5_{\text{AC}}$) rotates toward Asn257 ($\beta4_{\text{AC}}$) in both chains. Taken together, illumination results in a clamshell-like opening of the ATP-binding site mediated by new interactions in the $\beta4_{\text{AC}}$–$\beta5_{\text{AC}}$ tongue that move the AC core of the opposing molecule toward the BLUF domain (Figure 3.6).

The BLUF core domains and the flavin chromophore exhibit substantial disorder in illuminated crystals, hence molecular details of illumination-induced changes are difficult to assess. Nonetheless, one can observe a striking increase of the backbone interaction distance of Tyr7 and Gln49 from 3.0 to 3.4 Å in illuminated bPAC. A similar, yet less pronounced light-induced increase of the $\beta1_{\text{BLUF}}$–$\beta3_{\text{BLUF}}$ backbone distance was observed in AppA [85].

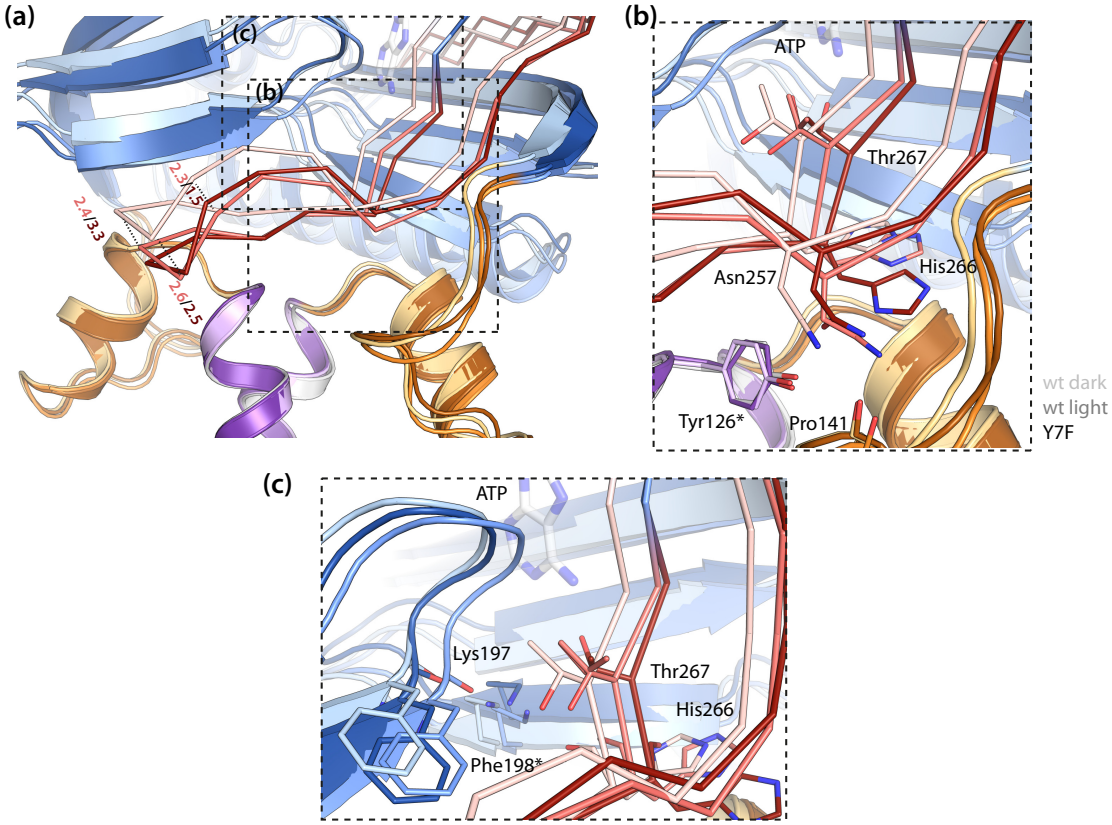



**Figure 3.5.:** Structure of the $\beta4_{AC}$–$\beta5_{AC}$ tongue and AC active site in dark state and illuminated bPAC crystals as well as the pseudo-lit $bPAC_{Y7F}$ variant (light, medium and dark colors, respectively). $\alpha3_{BLUF}$ of all structures was superimposed. **(a)** The $\beta4_{AC}$–$\beta5_{AC}$ tongue of chain A ($tongue_{UP}$ conformer in the dark, shown as light-red Cα trace) moves into the DOWN position in the illuminated crystal. Both tongues are in the down position in the symmetric $bPAC_{Y7F}$ structure. **(b)** Close-up of residues with changing interaction distances in illuminated crystals: Tyr126(OH)*–Asn257(OD1): 2.7 to 3.0 Å; Asn257(ND2)–Pro141(O): 2.8 to 2.3 Å. The movement of His266 toward Asn257 (6.5 to 3.9 Å) moves the Thr267 side chain away from the other subunit and toward the ATP base in illuminated wild type bPAC. His266 rotates further towards Asn257 in $bPAC_{Y7F}$ (2.5 Å) which positions the substrate-specifying Thr267 closer to the ATP binding site. **(c)** The pronounced movement of the AC subunits away from each other moves Lys197 away from the carbonyl of Phe198* in $bPAC_{Y7F}$ (2.4 Å to 3.0 Å) where it is held in its inactive conformation in the dark. The ATP base was modeled modeled into the active site based on Tesmer *et al.* [204].

### 3.1.5. Structure of a Pseudo-Lit State Variant

The constitutively active variant $bPAC_{Y7F}$ (Table 3.1 and [192]) was crystallized to investigate how it compares to illuminated crystals. $bPAC_{Y7F}$ crystallized in space group $C2$ with one chain per asymmetric unit and a two-fold crystallographic symmetry axis producing the dimer-related chain.

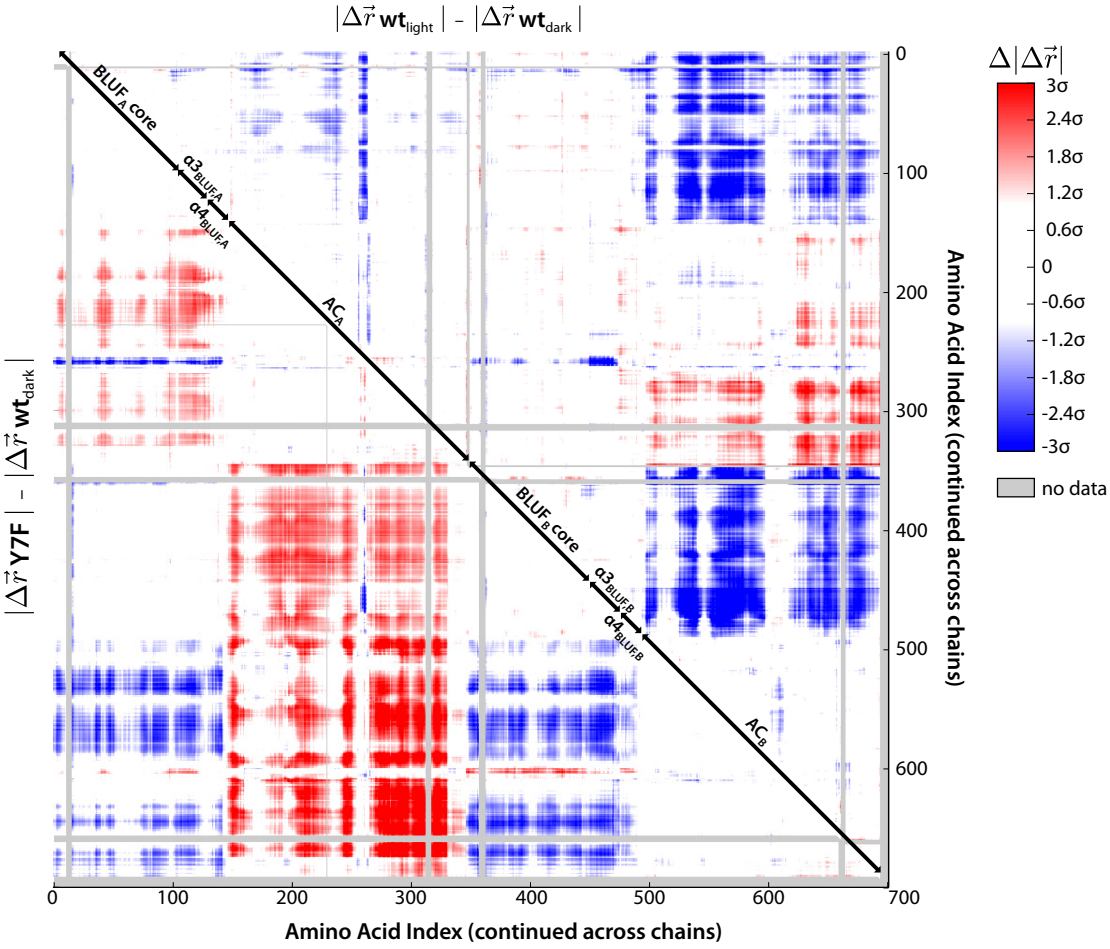



**Figure 3.6.:** Distance difference map. Differences between structures are visualized as a matrix of differences between pairwise Cα-distances expressed as multiples of positional uncertainty (σ) according to [176]. The lower left half shows differences between bPAC in the dark and the pseudolit bPAC$_{Y7F}$. The upper right half shows differences between bPAC in the dark and illuminated bPAC. Blue colors indicate lower distances between Cα atoms in the lit or pseudo-lit structures, respectively. Red colors indicate larger distances. Distance differences of less than 1σ are regarded as insignificant and shown in white. Parts of the protein lacking atomic coordinates in the respective crystal structures are colored gray.

Both β4$_{AC}$–β5$_{AC}$ tongues are in the DOWN conformation, producing a larger AC opening angle than in illuminated wild type bPAC (Table A.6 and Figure 3.6). The likely presence of alternate conformations in the tongue and in the handle complicates interpretation of the electron density. In either case, Asn257 is hydrogen-bonded to Tyr126* and Pro141 (2.7 and 2.4 Å, respectively), while His266 is oriented toward Asn257 and interacts with its side chain carbonyl at a distance of 2.5 Å. The conformational change

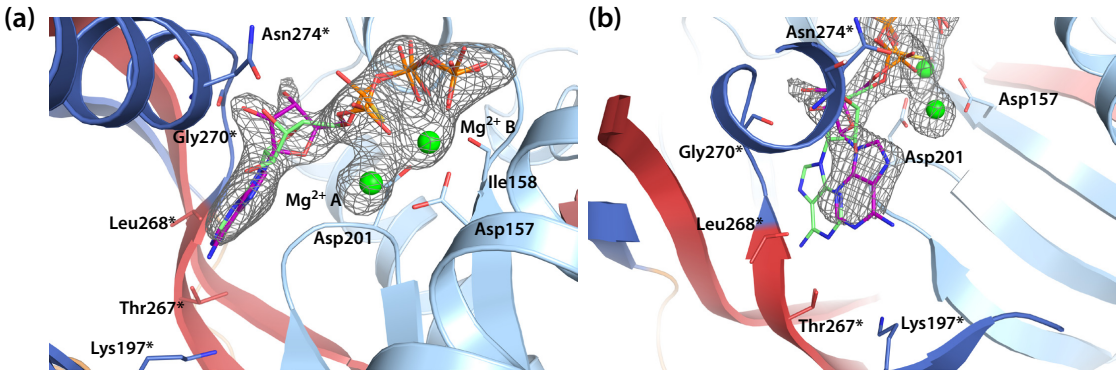



**Figure 3.7.:** Close-up of $bPAC_{Y7F}$ active site with bound (purple) and modeled (green) ATP. The experimental ATP conformation (purple) can be clearly assigned without model bias, as shown by the Fo-Fc omit map contoured at 3σ. **(a)** Overview of ATP coordination in the active site. The phosphate moiety is tightly coordinated by two magnesium ions bound to the catalytic aspartates and to the backbone carbonyl of Ile158. A model of a productively bound ATP molecule is shown in green. **(b)** Detailed view of base-specifying residues Thr267, Leu268 and Lys197. The base and the ribose are flipped by 180° around the adenine N1–N9 axis with respect to the proposed productive orientation [204]. Only one active site is shown as the other active site can be generated by two-fold crystallographic symmetry.

of His266 compared to the wild type twists the $\beta5_{AC}$ backbone and thereby positions the substrate-specifying Thr267 further away from the opposing subunit (Figure 3.5c). Through motion of the subunits away from each other, Lys197 is freed from the interaction with the backbone carbonyl of Phe198* that positions the Lys197 side chain away from the active site in the dark state. Taken together, the $bPAC_{Y7F}$ variant extends the conformational changes observed in illuminated wild type crystals. Not only is the opening angle larger but also residues involved in catalysis are affected to a larger extent than in illuminated wild type crystals.

The open lit-state structure appears more suitable for nucleotide binding, but dark-grown crystals of bPAC dissolved rapidly upon illumination in the presence of ATP. Soaking of $bPAC_{Y7F}$ crystals in the dark with high concentrations of $Mg^{2+}_2$·ATP produced interpretable electron density for the entire nucleotide bound in an unproductive conformation (Figure 3.7). The triphosphate moiety is bound to Asp157 and Asp201 via two $Mg^{2+}$ ions. Ion A is tetrahedrally coordinated by the two aspartate carboxylates, the Rp oxygen of the α-phosphate, and one water molecule. Ion B is octahedrally coordinated by both aspartates, oxygens from all three phosphates and the backbone carbonyl oxygen of Ile158. The ribose and the base are rotated by 180° with respect to the suggested produc-

tive hydrogen bonding pattern [203, 204]. The ribose is in a C2'-endo conformation with the 2'-hydroxyl group in hydrogen bonding distance to the backbone amide of Asn274 and the carbonyl oxygen of Gly270 while the 3'-hydroxyl points away from the phosphates, similar to the structure of mammalian AC in complex with ATP-αS (Rp) [204]. The adenine base is oriented such that the hydrogen bond-donating C6 amino group is directed at Lys197 whereas the hydrogen bond-accepting N3 points in the direction of Thr267 and the carbonyl oxygen of Leu268. As a result of these potential interactions, the base and the ribose would need to flip around the N1–N9 axis in order to enter the hydrophobic pocket more deeply and to allow closure of the active site. A nearly identical nucleotide conformation has been described for 3'-O-(N-methylanthraniloyl)-GTP (MANT-GTP) which is a potent inhibitor of mammalian AC [138]. A model of a productively bound ATP molecule in the $bPAC_{Y7F}$ crystal structure (Figure 3.7) shows that the $\beta5_{AC}$ strand is too close to the metal-binding aspartates to allow binding of the ribose in this conformation.

### 3.1.6. Structural Dynamics of bPAC

Crystal structures of dark-adapted and illuminated protein provide insight into the molecular interactions that couple the BLUF domain to the AC domain; however lattice constraints prevented observation of a catalytically competent state. To better understand the molecular snapshots obtained from crystallography, structural dynamics during illumination and catalysis in solution were investigated using hydrogen–deuterium exchange coupled to mass spectrometry (HDX-MS). Measurements on the wild type protein and initial data analysis were carried out by Andreas Winkler.

Relative deuteration rates of bPAC in the dark indicate the presence of two domains with protected secondary structures and less protected regions in the C-terminal parts of the BLUF domain (residues 80–105) and the AC domain (residues 300–350), respectively (Figure 3.8a). The dynamic region of the BLUF domain corresponds to the $\beta4_{BLUF}$–$\beta5_{BLUF}$ loop, the $\beta5_{BLUF}$ strand and the linker to the capping helix. The dynamic C-terminus of the AC domain contains the mobile $\beta7_{AC}$–$\beta8_{AC}$ hairpin [189]. The $\alpha3_{BLUF}$ capping helix, the $\alpha4_{BLUF}$ helix of the handle, and most significantly the peptide Gln114–Glu124 at the C-terminus $\alpha3_{BLUF}$ incorporate deuterium at remarkably low rates.

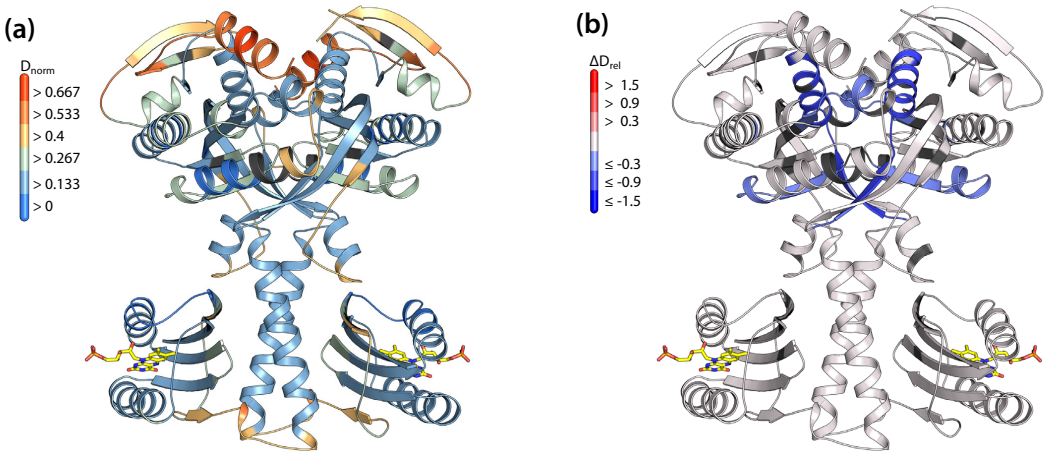



**Figure 3.8.:** Hydrogen–deuterium exchange of bPAC in the dark after 900 s. Regions colored in black indicate lack of coverage. **(a)** Relative deuteration normalized with respect to maximum deuteration. Colors range from blue (low relative deuteration) to red (high relative deuteration). The BLUF and AC domain cores and $\alpha3_{BLUF}$ and $\alpha4_{BLUF}$ are protected from exchange, whereas the domain C termini ($\beta5_{BLUF}$–$\alpha3_{BLUF}$ loop and the AC domain after $\alpha4_{AC}$) show high levels of deuteration. A map showing underlying peptides is given in Figure A.1. **(b)** Deuteration differences between bPAC in the presence and absence of the competitive inhibitor ATP-αS (Rp). Decreased deuterium uptake in the presence of ATP-αS (Rp) is marked in blue, whereas increased deuterium uptake is marked in red. Nucleotide-binding elements $\beta1_{AC}$–$\alpha1_{AC}$ (Asp157), $\alpha1_{AC}$–$\beta2_{AC}$ (Asp201), $\beta5_{AC}$ (Thr267), and $\alpha4_{AC}$ (Asn274) are protected from exchange in the presence of the inhibitor. A map showing underlying peptides is given in Figure A.2. Figure A.5 shows relative deuteration time series of all analyzed peptides.

To faithfully mimic an ATP-bound state, the protein was incubated with ATP-αS (Rp) that had been identified as competitive inhibitor. In the dark state of the protein, ATP-αS (Rp) mainly stabilizes nucleotide-binding elements in several parts of the AC active site: the metal-binding $\beta1_{AC}$–$\alpha1_{AC}$ loop, $\alpha2_{AC}$ close to the metal-binding Asp201, and the $\beta5_{AC}$–$\alpha4_{AC}$ region containing Thr267, Asn274 and Arg278, all of which have been implicated in catalysis [118, 203]. Fast-exchanging amides of the $\beta7_{AC}$–$\beta8_{AC}$ hairpin experience some protection (Figure A.5). Elements outside the AC domain are hardly affected by nucleotide in the dark. The $\alpha4_{BLUF}$ helix in the handle region (e.g. peptide Lys125–Tyr133) constitutes one exception and displays slight deprotection at long incubation times.

Illumination of the protein (Figure 3.9) increases the exchange rate of the BLUF $\alpha1_{BLUF}$–$\beta2_{BLUF}$ loop and the C-terminal $\beta5_{BLUF}$ strand, both of which are commonly observed signatures of BLUF domain illumination [221, 220, 57, 231, 215, 216]. Analysis of overlapping peptides highlights residues Ala92–Leu98 following the conserved Trp91

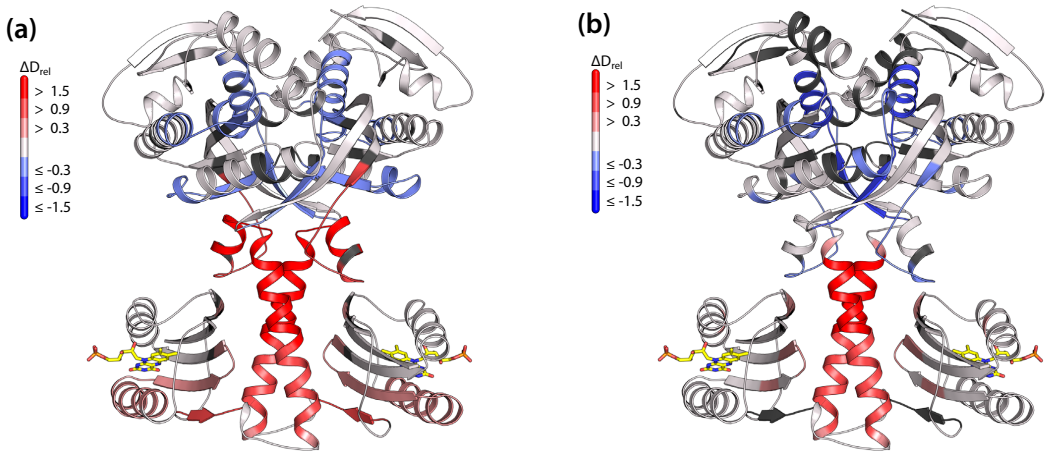



**Figure 3.9.:** Hydrogen–deuterium exchange of bPAC in the light after 900 s. Decreased deuterium uptake in the light is marked in blue, whereas increased deuterium uptake is marked in red. Regions colored in black indicate lack of coverage. **(a)** Deuteration differences between bPAC in the dark and in the light. Illumination increases the exchange rate of the $\alpha1_{BLUF}$ helix, the $\alpha1_{BLUF}$–$\beta2_{BLUF}$, and $\beta2_{BLUF}$–$\beta3_{BLUF}$ loops, parts of the $\beta4_{BLUF}$, and the C-terminal $\beta5_{BLUF}$ strand. The entire $\alpha3_{BLUF}$ coiled coil, particularly at the C-terminus, and the handle region show increased deuterium incorporation in the light. Nucleotide-binding elements in the AC domain are protected from exchange in the light. A map showing underlying peptides is given in Figure A.3. **(b)** Deuteration differences between bPAC in the dark and in the light, in the presence of the competitive inhibitor ATP-αS (Rp). Effects on deuteration uptake are similar to the effects in the absence of nucleotide for most parts of the protein. The protective effect on the AC domain is more pronounced in the presence of nucleotide. The handle shows a different response to illumination when nucleotide is present and experiences slight protection in the $\alpha4_{BLUF}$–$\beta1_{AC}$ region upon illumination. A map showing underlying peptides is given in Figure A.4. Figure A.5 shows relative deuteration time series of all analyzed peptides.

as site of increased deuteration. In addition to these elements, the $\beta2_{BLUF}$–$\beta3_{BLUF}$ loop shows increased deuterium incorporation under illumination. The entire $\alpha3_{BLUF}$ capping helix and the handle show deprotection in the lit state and slow-exchanging amides of the highly protected Gln114–Glu124 peptide are among the most strongly deprotected. The effects on the AC domain are similar to what was observed upon addition of nucleotide. Illumination of bPAC in the presence of ATP-αS (Rp) produced larger changes in deuterium incorporation than in the absence of nucleotide (Figure 3.9b). In addition to the light- and nucleotide-induced changes described above, the helical bundle exhibits distinct exchange dynamics that occur exclusively when bPAC is illuminated in the presence of ATP-αS (Rp): Only a short part of the capping helix (e.g., peptide Gln114–Glu124) shows strong destabilization while peptides leading into the handle show no difference (e.g., His120–Tyr133) or moderate stabilization (Leu134–Asn140) upon illumination

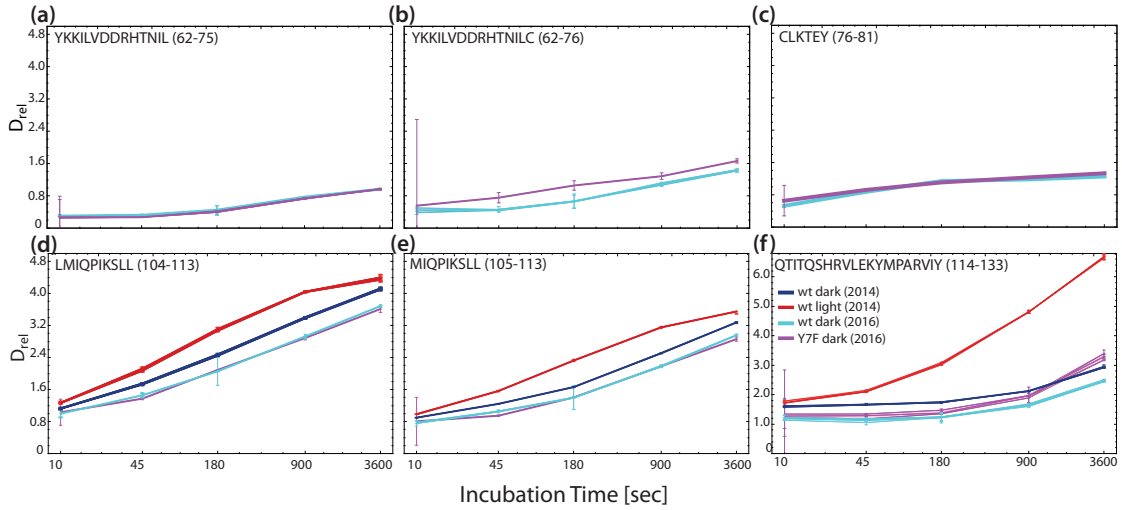



**Figure 3.10.:** Deuterium incorporation in wild type bPAC and the Y7F variant. **(a–c)** Deuteration of overlapping peptides in the $\beta4_{BLUF}$ kink. The peptides Tyr62–Leu75 **(a)** and Cys76–Tyr81 **(c)** show no deuteration difference whereas Tyr62-Cys76 **(b)** does. Taking into account rapid back-exchange of the first two amides during work-up, deuteration differences between the wild type and $bPAC_{Y7F}$ in the $\beta4_{BLUF}$ strand can be localized to the amides of Cys76 and Leu77. Panels **(d)** and **(e)** show peptides from the middle of the $\alpha3_{BLUF}$ helix that show no difference between wild type and Y7F variant. For comparison, the corresponding peptides of dark-adapted and illuminated wild type are shown, indicating clear deuteration differences. **(f)** Peptide Q114–Y133 (different deuteration scale) covering the C-terminal part of $\alpha3_{BLUF}$ and the handle shows increased deuteration in illuminated wild type and the Y7F variant. Note that data from 2014 and 2016 measurement series have different back-exchange offsets due to different work-up, and can therefore not be compared to each other quantitatively.

in the presence of nucleotide. Analysis of overlapping peptides (Figures A.4 and A.5) allows assignment of the destabilization to the amides of the sequence $^{116}$ITQSH. Notably, this part of the coiled coil contains the unusual packing defects described above. Inspection of deuterium incorporation distributions in this part of the coiled coil shows strict EX2 deuteration kinetics [102] that diverge strongly after long incubation times. Such exchange behavior is indicative of a short-lived unprotected state that occurs with a higher frequency in the light than in the dark.

The variant $bPAC_{Y7F}$ cannot engage in BLUF photochemistry, yet it exhibits constitutively high AC activity (Table 3.1 and [192]). To investigate the mechanism of photochemistry-independent AC activation in this variant, a separate HDX-MS experiment was carried out comparing wild type bPAC to $bPAC_{Y7F}$. While secondary structure dynamics of "downstream" elements, e.g., the AC domain, are similar to illuminated

wild type protein, the BLUF domain showed a distinct signature: Only the $\beta4_{BLUF}$ strand around residues 76-77 (Figure 3.10a-c) shows increased deuteration when compared to wild type in the dark. This region forms a prominent kink and has extensive backbone interactions with $\beta1_{BLUF}$ near Phe7 (Tyr7). Release of Phe7 from the flavin hydrogen bond network in $bPAC_{Y7F}$ may result in a similar effect on $\beta4_{BLUF}$ as illumination does. Leu75 and Leu77 in $\beta4_{BLUF}$ are in direct vicinity of His120 and Glu124. Perturbation of the hydrogen bonding network in this region is a key feature of illumination (Figure 3.9). Unlike the illuminated wild type protein, $bPAC_{Y7F}$ does not display increased secondary structure dynamics in other parts of the BLUF domain, e.g., $\beta5_{BLUF}$ near the conserved methionine or the adjacent the N-terminal part of $\alpha3_{BLUF}$ (Figure 3.10d, e) whereas dynamics around Glu124 increase (Figure 3.10f). This suggests that AC activation in $bPAC_{Y7F}$ occurs predominantly through interaction of $\beta4_{BLUF}$ with adjacent parts of $\alpha3_{BLUF}$, rather than affecting the BLUF methionine.

## 3.1.7. Probing BLUF-AC Signaling with Mutagenesis

### BLUF Domain Constructs

$bPAC_{\Delta103}$ and $bPAC_{\Delta127}$ are truncation constructs that contain only the BLUF core up to $\beta5_{BLUF}$ and the BLUF core with the $\alpha3_{BLUF}$ capping helix, respectively. Both constructs yield soluble protein. Spectroscopic properties of the truncation constructs were investigated to assess the influence of the allosterically coupled AC domain or of the capping helix alone on the BLUF photocycle. The absorption spectrum of $bPAC_{\Delta127}$ in the dark state is nearly identical to the full-length protein whereas the $bPAC_{\Delta103}$ absorption is slightly red-shifted to 443 nm in the dark (wild type 440 nm). The signaling state decays at a markedly increased rate in both truncated constructs (7.7 ± 0.2 s in $bPAC_{\Delta103}$ and 7.0 ± 0.2 s in $bPAC_{\Delta127}$).

### Packing Defects between the BLUF β-sheet and $\alpha3_{BLUF}$: $bPAC_{I8V}$, $bPAC_{I116V}$, $bPAC_{I8V,I116V}$

This set of amino acid substitutions was introduced to investigate the role of the packing defect between $\beta1_{BLUF}$ and the $\alpha3_{BLUF}$ capping helix. Ile8 is adjacent to the photoactive Tyr7 in $\beta1_{BLUF}$ and faces Ile116* in the capping helix of the opposing subunit. Both

isoleucine residues show conformational flexibility (as evidenced by the presence of alternate conformations) in the dark state of bPAC and line a packing defect between the BLUF core and the coiled coil, underneath the sidechain hydrogen bond network around Glu124. Similar packing defects between the BLUF core β-sheet and $\alpha3_{BLUF}$ are present in BlrP1 [7] and BlrB [84].

Amino acid substitutions increasing the size of the void from either side did not influence the kinetics of the adenylyl cyclase in the dark or in the light. The $bPAC_{I8V,I116V}$ double mutant crystallized in space group *P*1 (Table A.4) with both chains exhibiting the $tongue_{DOWN}$ conformation. His266 is rotated away from Asn257, leaving Thr267 in the locked dark state position. Several residues lining the site of mutation (Phe48, Leu112, Phe41) move slightly towards the cavity that is larger in $bPAC_{I8V,I116V}$ than in the wild type. The cavity between the BLUF β-sheet of chain A and the $\alpha3_{BLUF}$ coiled coil has a total volume of 9.9 $Å^3$ whereas two smaller cavities with volumes of 7.0 $Å^3$ and 5.6 $Å^3$, respectively, are located between the BLUF domain of chain B and the coiled coil. Both single substitution variants $bPAC_{I8V}$ and $bPAC_{I116V}$ are less soluble than the wild type while the double substitution variant $bPAC_{I116V}$, $bPAC_{I8V,I116V}$ restores solubility.

The red-shifted BLUF signaling states of $bPAC_{I8V}$ and $bPAC_{I116V}$ decay with a markedly reduced half-life (11.8 ± 0.9 s and 16.7 ± 0.4 s, respectively) while substitution of both isoleucine residues by valine produces a double exponential dark state recovery with half-lives of 6.9 ± 0.3 s and 33.3 ± 0.7 s and amplitudes of 35 % and 65 %, respectively. The fast recovery rate is reminiscent of the isolated BLUF constructs $bPAC_{\Delta103}$ and $bPAC_{\Delta127}$, possibly reflecting a state in which the BLUF core is partially detached from the $\alpha3_{BLUF}$ helix.

### Crosslink between $\beta4_{BLUF}$ and $\alpha3_{BLUF}$: $bPAC_{K78C,T115C}$

The K78C/T115C double substitution ($bPAC_{K78C,T115C}$) was designed for disulfide crosslinking to prevent motion of the BLUF core domain relative to the $\alpha3_{BLUF}$ coiled coil. Lys78 in the $\beta4_{BLUF}$ strand is in close proximity to Thr115* in the $\alpha3_{BLUF}$ helix. The variant has a slightly red-shifted absorption spectrum with its maximum at 442 nm (wild type: 440 nm). After illumination with blue light, the spectrum is red-shifted by 11 nm and decays slowly with a half-life of ~1 min. When reducing agent is present in the buffer, $bPAC_{K78C,T115C}$ exhibits a slightly increased but light-responsive adenylyl cyclase

activity. In the absence of reducing agent, crosslinking was observed only in the light. The size of the crosslinked species is consistent with a bPAC dimer on non-reducing SDS-PAGE and the band was confirmed to contain bPAC by mass spectrometry. Adenylyl cyclase activity is constitutively high and barely light-responsive in the absence of reducing agent. This indicates that illumination either brings $\beta4_{BLUF}$ closer to $\alpha3_{BLUF}$ or it enhances conformational flexibility of these regions, increasing the likelihood of crosslinking events.

### Side Chain Interaction Network in $\alpha3_{BLUF}$: $bPAC_{R121S}$

Arg121 connects the peripheral $\beta4_{BLUF}$ strand to the $\alpha3_{BLUF}$ coiled coil by hydrogen bonds to the Asn73 carbonyl oxygen and the carboxylate group of Glu124. Substitution of Arg121 with serine severely perturbs secondary structure stability in the C-terminal part of $\alpha3_{BLUF}$ as shown by marked increase of deuteration in hydrogen–deuterium exchange experiments. $bPAC_{R121S}$ exhibits high dark state AC activity that further increases to near wild type levels in the light. The absorbance spectrum of $bPAC_{R121S}$ exhibits a red-shifted maximum at 448 nm, lacks the dark state fine structure and is reminiscent of the illuminated wild type (Figure 3.11). Illumination slightly red-shifts the absorption maximum to 449 nm which decays with a half-life of 13.0 ± 0.5 s.

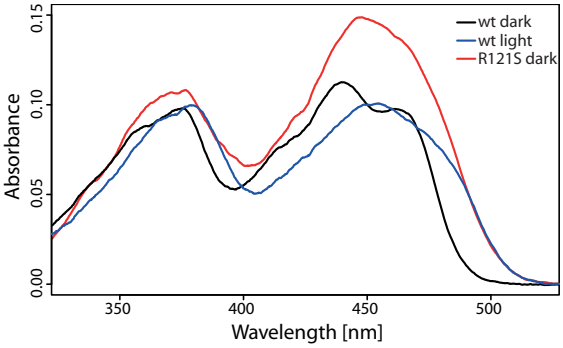



**Figure 3.11.:** Absorption spectra of bPAC wild type and $bPAC_{R121S}$. The wild type spectrum exhibits two peaks around the maximum at 440 nm in the dark (black line) whereas the illuminated protein has a red-shifted absorption spectrum lacking this fine structure (blue line). The spectrum of $bPAC_{R121S}$ is red-shifted with respect to the wild type and does not display any fine structure in the dark state (red line).

The red shifted spectrum of $bPAC_{R121S}$ in the dark cannot be attributed to the loss of interactions between the BLUF core and the $\alpha3_{BLUF}$ capping helix as the isolated BLUF domain construct $bPAC_{\Delta103}$ displays a typical dark state absorption spectrum with a slight red shift to a maximum of 442 nm. Therefore, it appears that the R121S substitution not only activates the AC domain but also specifically affects hydrogen bonds around the flavin chromophore of the BLUF domain.

### Interaction of the $\alpha3_{BLUF}$ Coiled Coil, the Handles and the $\beta4_{AC}$–$\beta5_{AC}$ Tongue: Substitution of Y126, P141 and N257

Tyr126 is located at the C-terminal end of the $\alpha3_{BLUF}$ helix and interacts with His266 in the $\beta4_{AC}$–$\beta5_{AC}$ tongue as well as Pro141 in the handle through Asn257, also located in the $\beta4_{AC}$–$\beta5_{AC}$ tongue. This region integrates structural changes in the side chain hydrogen bond network around Glu124 and the handle. Structural dynamics are translated to the $\beta4_{AC}$–$\beta5_{AC}$ tongues that control the AC domain angle as well as the orientation of catalytic residues such as Thr267 and Lys197. A series of variants were created to perturb the interaction between $\alpha3_{BLUF}$, the handles and the $\beta4_{AC}$–$\beta5_{AC}$ tongue in different ways:

$bPAC_{Y126F}$ lacks an important link between the $\alpha3_{BLUF}$ helix and the $\beta4_{AC}$–$\beta5_{AC}$ tongue. This variant shows severely reduced solubility and has no detectable AC activity in the dark or in the light. Spectroscopic properties of $bPAC_{Y126F}$ differ from the wild type protein in several aspects: the signaling state absorbance is only red-shifted by ~8 nm (wild type: 15 nm) and decays with a half-life of 8.7 ± 0.1 s. Increased recovery rates have also been observed for truncated constructs lacking the AC domain, hence the accelerated dark state recovery of $bPAC_{Y126F}$ may indicate loss of coupling between the AC domain and the BLUF domain.

Pro141 is part of a $^{140}$NPLT motif that forms a rigid type I β-turn. Interaction of the Pro141 carbonyl oxygen with Asn257 or Tyr126 attaches the handle firmly to $\alpha3_{BLUF}$ and the AC domain. The coupling between BLUF and AC in this region is demonstrated by the complex hydrogen-deuterium exchange kinetics in response to nucleotide and light: In the absence of nucleotide, illumination destabilizes amide hydrogen bonds of the peptide 138–145. Although hydrogen-deuterium exchange in the dark is not affected, illumination has a protective effect in the presence of ATP-αS (Rp), hence this region responds differently to illumination depending on the nucleotide state of the AC active site. Substitution of Pro141 by glycine changes the β-turn architecture and likely perturbs the coupling to Tyr126 and Asn257. Spectroscopic and kinetic properties of $bPAC_{P141G}$ are reminiscent of $bPAC_{Y126F}$. The protein shows no detectable AC activity in the light or in the dark. The dark state recovery half-life is shorter than in the wild type: 14.1 ± 0.8 s.

The variants $bPAC_{N257A}$ and $bPAC_{N257K}$ exhibit substantial residual AC activity upon

blue light illumination. In contrast to $bPAC_{Y126F}$ that appears decoupled, $bPAC_{N257K}$ is spectroscopically very similar to the wild type with a 13 nm red-shift of the light-adapted absorption spectrum and a dark state recovery half-life of 27 ± 3 s.
The double substitution variant $bPAC_{Y126F,N257A}$ crystallizes as symmetric dimer with the $tongue_{DOWN}$ conformation. In contrast to the singly-substituted variants that have buried polar groups, residues 126 and 257 are coupled through hydrophobic interactions (Phe126CZ–Ala257CB distance 3.9 Å) in $bPAC_{Y126F,N257A}$. Solubility of the protein is increased compared to the single mutants. It is not clear how Pro141 and His266 are involved in the interactions; however introduction of the second substitution restores the light-state adenylyl cyclase activity to near wild type levels.

## 3.2. Investigations and Engineering on mPAC

### 3.2.1. Background

Before the discovery of mPAC through genome screening in the laboratory of Wolfgang Gärtner, only BLUF domain-regulated photoactivatable adenylyl cyclases (PAC) had been known. Attempts to tune the photocycle (e.g., to accelerate dark state recovery) or enzyme kinetics of bPAC, the most widely used PAC in optogenetic applications,met only limited success [192]. In contrast to BLUF domains, LOV domains had been successfully used for photoreceptor engineering and a large body of knowledge exists on photocycle tuning [156]. For this reason, the LOV-domain regulated *Microcoleus* photoactivatable adenylyl cyclase mPAC held great promise as a new, potentially tunable optogenetic tool for the manipulation of cAMP levels.
The sequence of mPAC was shared among collaborators in the DFG Forschergruppe 1279 and I began working on in vitro characterization of mPAC before any information had been published about it. Soon thereafter, the sequence and basic enzymological parameters of mPAC were published by the laboratory of Wolfgang Gärtner [159]. Initial excitement about mPAC ceased following first reports about prohibitively high dark state AC activity in vitro and in vivo [159, 21].
My role in this project was to characterize the structural mechanism of photoactivation in mPAC, and to deduce structure-function relationships from comparison with bPAC. I transferred insights obtained from bPAC to suppress the high dark state activity of mPAC and screened for variants of the LOV domain with faster dark state recovery. In combination, this produced an mPAC-based optogenetic tool that allows precise regulation and meets many requirements of in vivo studies.

All experiments presented in this thesis were planned and conducted by myself. I received assistance in crystal and SAXS sample handling from Ilme Schlichting and Miroslaw Tarnawski, and advice for SAXS data collection and processing by Andreas Menzel and Thomas Barends. The K468S variant of mPAC was devised in consultation with Ilme Schlichting.

### 3.2.2. Enzymatic and Spectroscopic Properties of mPAC

mPAC was obtained as soluble homodimeric protein bearing two additional amino acids after proteolytic removal of the hexahistidine tag. Preparations were largely pure except for minor contaminations by the host chaperone DnaK and a degradation product lacking the AC domain. The LOV domain is fully occupied with FMN and exhibits an absorption maximum at 447 nm. Illumination with blue light bleaches the absorption band at 447 nm which recovers in the dark with a half-life of $11.2 \pm 0.2$ s (Figure 3.12). The apparent AC rate constant at 100 µM ATP is $0.049 \pm 0.005\ s^{-1}$ (SEM) in the dark and increases 16-fold to $0.83 \pm 0.04\ s^{-1}$ under illumination with blue light at 7 mW $cm^{-2}$. Michaelis Menten parameters were difficult to determine in the dark and yielded a $K_M$ of $20 \pm 8$ µM, and $v_{max}$ of $0.040 \pm 0.002\ s^{-1}$. Under constant illumination $K_M$ is $160 \pm 20$ µM and $v_{max}$ equals $2.9 \pm 0.1\ s^{-1}$. The AC rate constant increases with pH between 6.8 and 8.5 whereas the dark state recovery rate of the LOV domain is nearly constant in this pH range (Table 3.2).

To investigate the role of each domain, the truncation constructs $mPAC_{\Delta 132}$ and $PAC_{\Delta 156}$, $mPAC_{AC}$ and $mPAC_{LOV}$ were expressed and characterized. SEC-MALS experiments showed that the PAS domain mainly promotes dimerization (Figure 3.13) and it appears dispensable for photoregulation of mPAC: deletion ($mPAC_{\Delta 132}$) had no effect on lit state

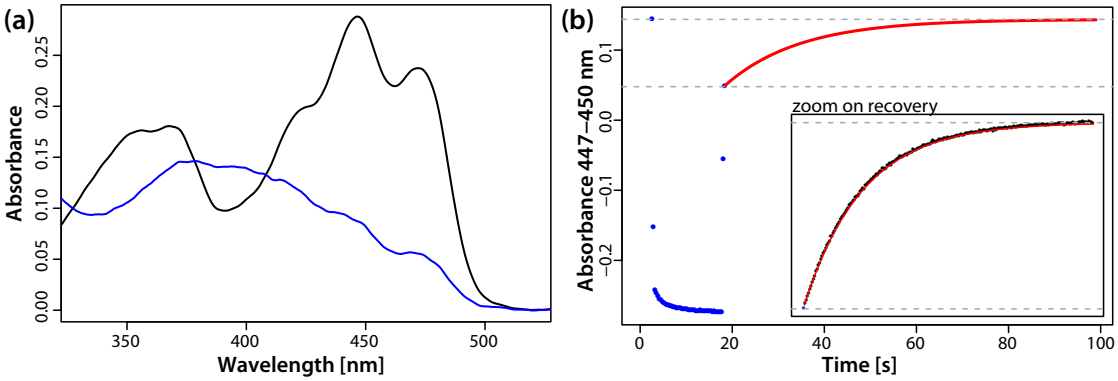



**Figure 3.12.:** Dark state recovery of mPAC. **(a)** Absorption spectrum of mPAC in the dark (black) and immediately after illumination (blue). **(b)** Time trace of absorption band between 447 and 450 nm. First segment (black) indicates the steady state absorption before illumination, the segment colored in blue indicates absorbance during illumination with blue light and the following segment (black) shows recovery of dark state absorption after the light has been shut off. The red line shows a monoexponential fit to the recovery trace (see inset for zoom). The recovery half-life of mPAC is 11.2 s (Table 3.2).

**Table 3.2.:** Kinetic and spectroscopic properties of mPAC wild type and variants. AC activity is reported as the apparent catalytic rate constant at 100 µM ATP [$s^{-1}$]. The dark state absorption maximum of all variants was at 447 nm.

| Construct | pH | AC Activity [$s^{-1}$] Dark | Light | Recovery t½ [s] | Description |
|---|---|---|---|---|---|
| **wild type** | 6.8 | 0.016 ± 0.005 | 0.74 ± 0.07 | 11.09 ± 0.05 | Wild type |
| **wild type** | 7.5 | 0.049 ± 0.005 | 0.83 ± 0.04 | 11.2 ± 0.2 | Wild type |
| **wild type** | 8.5 | 0.060 ± 0.008 | 0.99 ± 0.07 | 11.1 ± 0.2 | Wild type |
| **KS** | 6.8 | 0.0003 ± 0.0002 | 0.05 ± 0.07 | – | K468S variant |
| **KS** | 7.5 | 0.0006 ± 0.0002 | 0.15 ± 0.03 | 11.2 ± 0.3 | K468S variant |
| **KS** | 8.5 | 0.0017 ± 0.0002 | 0.56 ± 0.01 | 11.4 ± 0.1 | K468S variant |
| **K468Q** | 7.5 | < 0.001[a] | 0.13 ± 0.01 | – | K468Q variant |
| **K468L** | 7.5 | < 0.001[a] | 0.097 ± 0.005 | – | K468L variant |
| **Δ132** | 7.5 | 0.09 ± 0.01 | 0.8 ± 0.3 | 11.0 ± 0.2 | LOV-AC |
| **Δ132** | 8.5 | – | – | 11.1 ± 0.3 | LOV-AC |
| **Δ156** | 7.5 | – | – | 8.1 ± 0.1 | LOV-AC lacking $\alpha 1_{LOV}$ |
| **LOV** | 6.8 | n/a | n/a | 10.0 ± 0.3 | Isolated LOV |
| **LOV** | 7.5 | n/a | n/a | 10.3 ± 0.2 | Isolated LOV |

[a] Upper bound estimate. There were not enough time points at which cAMP was detected for a reliable fit.

decay and the AC rate constant was barely affected. The activity of $mPAC_{AC}$ was not determined in vitro, however growth of the $mPAC_{AC}$ expression culture was similarly retarded by cAMP-induced starvation stress as that of cultures expressing the wild type protein, suggesting some activity of the truncated AC construct.

The photocycle kinetics of the LOV domain are largely independent of the PAS and AC domains. The isolated LOV domain recovers from illumination with a half-life of 10.3 ± 0.2 s which is only slightly shorter than in the wild type. Conversely, the integrity of the LOV domain and its helical extensions appears important: A construct lacking the N-terminal $\alpha 1_{LOV}$ helix (also referred to as A'α or N-cap [136, 70]) of the LOV domain ($mPAC_{\Delta 156}$) has a markedly reduced lit state half-life of 8.1 ± 0.1 s.

## 3.2.3. Structures of the mPAC AC Domain

### Structures of Native $mPAC_{AC+Handle}$ and $mPAC_{AC}$

Both $mPAC_{AC+Handle}$ and $mPAC_{AC}$ crystallized in space group $C222_1$ with nearly identical unit cell parameters and four molecules in the asymmetric unit arranged similar

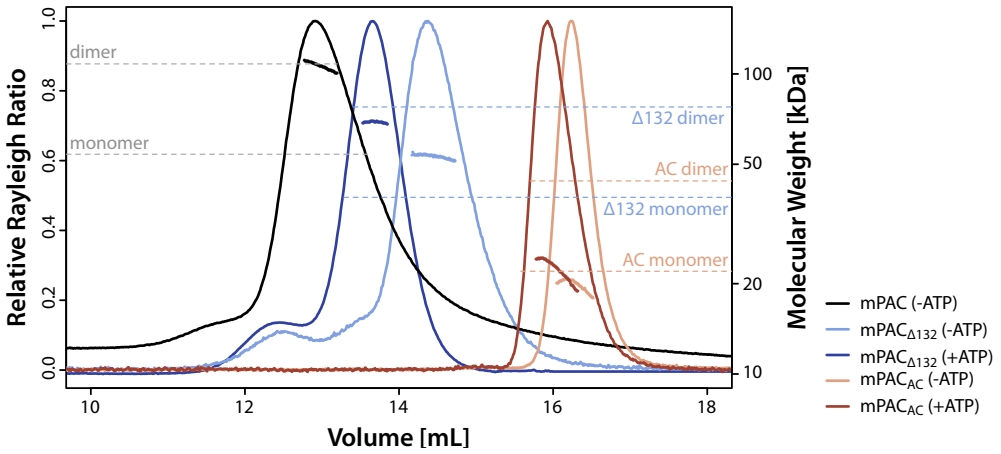



**Figure 3.13.:** Oligomeric state of mPAC in the presence and absence of 0.5 mM ATP determined by SEC-MALS. Elution was monitored by scattering (solid lines, left axis) and the determined molecular weight is shown as dots under the respective peaks (right axis). Dashed lines indicate the expected molecular weight for the respective monomer and dimer species. mPAC (black) elutes as dimer in the absence of ATP. Pronounced tailing suggests a dimer-monomer equilibrium. mPAC$_{Δ132}$ (shades of blue) elutes as a mixture of monomers and dimers in the absence of ATP, and is shifted towards dimers in the presence of ATP. The isolated AC domain (shades of brown) elutes as monomer and is shifted towards a small fraction of dimers in the presence of ATP. All proteins had a concentration of approximately 1 µM at the light scattering detector.

enough to be phased by the same model of SeMet-substituted mPAC$_{AC}$ obtained from Se-SAD phasing. As the structures of AC$_{AC+Handle}$ and mPAC$_{AC}$ are highly similar and electron density was only interpretable starting at residue 300, description of the structures is based on crystals of mPAC$_{AC}$ that diffracted to better resolution. Data collection and refinement statistics are shown in Table A.5.

The asymmetric unit contains two dimers related to each other via a two-fold noncrystallographic symmetry axis. The monomers of each dimer are related by a two-fold noncrystallographic axis that runs perpendicular to the axis relating the dimers and is located between the β4$_{AC}$–β5$_{AC}$ tongues of both molecules (Figure 3.14a). The dimers bury surface areas of 1355 and 1304 Å$^3$ in the interface between chains A/B and C/D, respectively, and are judged significantly noncrystallographic by PISA analysis [103]. mPAC$_{AC}$ contains all elements of a typical type IIIa AC fold [203, 114], however the β4$_{AC}$–β5$_{AC}$ tongues extend away from the AC cores in an angle that is vastly different from other structures of catalytically competent AC dimers (e.g., [203, 189, 205, 99, 115]). Crystal contacts hold the tongues in a conformation that is bent back towards their own AC core instead of reaching into the other monomer to contact α2$^*_{AC}$ and the β2$^*_{AC}$–β3$^*_{AC}$

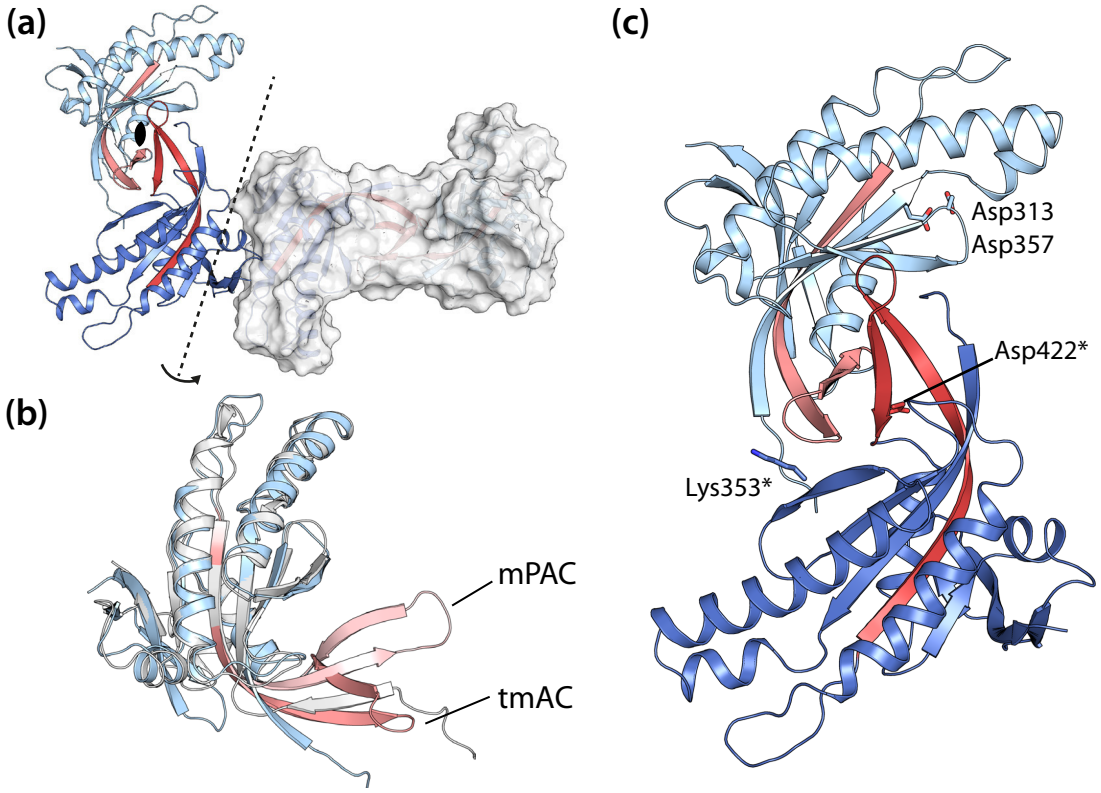



**Figure 3.14.:** Dimeric crystal structure of $mPAC_{AC}$. **(a)** $mPAC_{AC}$ crystallized as dimer of dimers in the asymmetric unit. Monomers are related through a two-fold noncrystallographic symmetry axis (perpendicular to the paper plane, approximate location indicated by ⧫) and dimers are related through a two-fold noncrystallographic axis indicated by a dashed line. One dimer is shown in cartoon representation and the second dimer is shown as surface. AC domain cores of each chain are colored in shades of blue and the $\beta4_{AC}$–$\beta5_{AC}$ tongues are indicated in red. **(b)** Superposition of a $mPAC_{AC}$ monomer (blue) with the structure of mammalian transmembrane AC (tmAC, gray, PDB 1CJK [204]). While most of the core domain fold is nearly identical, the angle at which the $\beta4_{AC}$–$\beta5_{AC}$ tongues (mPAC light red; tmAC dark red) extend away from the domain to interact with the dimer-related chain is vastly different. **(c)** The AC active site is composed by Asp313 ($\beta1_{AC}$) and Asp357 ($\beta2_{AC}$–$\beta3_{AC}$) of one chain (shown on chain A, light blue/red), and Lys353* ($\beta2^*_{AC}$) and Asp422* ($\beta5^*_{AC}$, shown on chain B, darker shades) of the other chain. The arrangement of $mPAC_{AC}$ observed in this crystal structure does not allow formation of a composite active site at the dimer interface.

hairpin (Figure 3.14b, asterisks indicate elements in the dimer-related chain). As a consequence, the $\beta4_{AC}$–$\beta5_{AC}$ tongues of the observed $mPAC_{AC}$ dimer interact mainly with each other and with the $\beta3^*_{AC}$ strand of the dimer-related chain. This assembly is not capable of forming an active site at the dimer interface because the metal-binding Asp313 ($\beta1_{AC}$) and Asp357 ($\beta2_{AC}$–$\beta3_{AC}$ loop) are too far away from the base-specifying residues Lys353* ($\beta2^*_{AC}$) and Asp422* ($\beta5^*_{AC}$ provided by the other chain (Figure 3.14c).

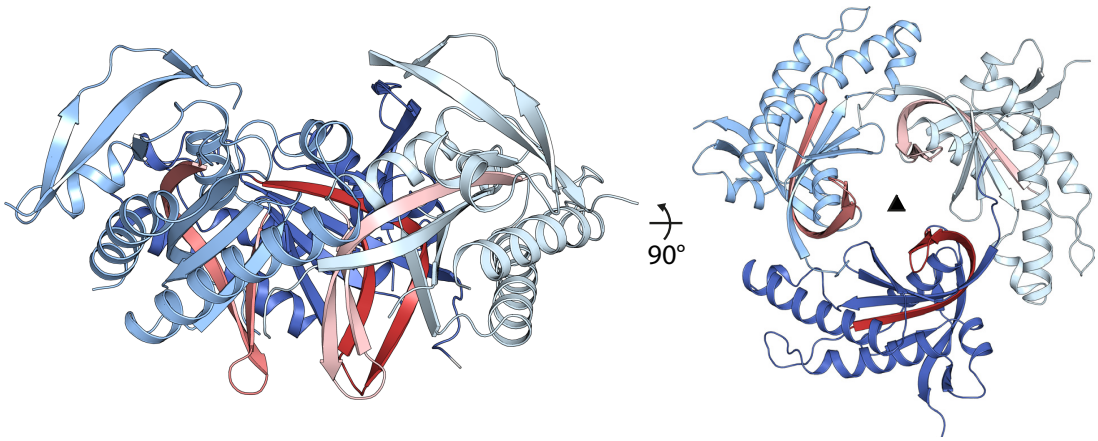



**Figure 3.15.:** Trimeric crystal structure of $mPAC_{AC}$. The SeMet derivative of $mPAC_{AC}$ crystallized as trimer in the asymmetric unit. AC domain cores of each chain are colored in shades of blue (light to dark shades for chain A through C) and the $\beta4_{AC}$–$\beta5_{AC}$ tongues are indicated in corresponding shades of red. The trimers are related by a three-fold noncrystallographic symmetry axis perpendicular to the central pore (right, location indicated by ▲). The assembly of $mPAC_{AC}$ observed in this crystal structure does not allow formation of a composite active site at any binary dimer interface.

### Trimeric Arrangement in Crystals of the $mPAC_{AC}$ SeMet Derivative

A second crystal form was found exclusively in the SeMet derivative of $mPAC_{AC}$. Octahedral bipyramidal crystals of space group $P2_12_12_1$ diffracted to a resolution of 2.0 Å. Phase estimates were obtained experimentally using Se-SAD. Data collection and refinement statistics are shown in Table A.5.

The asymmetric unit contains three molecules related to each other through three-fold noncrystallographic symmetry (Figure 3.15). Comparison with the structure of mammalian transmembrane AC [204] shows that the type IIIa AC fold is largely conserved except for the $\beta4_{AC}$–$\beta5_{AC}$ tongue that folds back towards the AC core domain instead of extending away from the domain, similar to the dimeric structure of $mPAC_{AC}$. The $\beta4_{AC}$–$\beta5_{AC}$ tongue of each molecule is packed against a crystallographically related molecule where it inserts between the $\alpha1_{AC}$–$\alpha2_{AC}$ loop and a $^{392}$TPQG motif that forms a rigid type I ST-turn [41] in the $\alpha3_{AC}$–$\beta4_{AC}$ loop. The trimeric structure shows no binary interface at which the metal-binding residues Asp313 and Asp357 are in proximity to the base-specifying Lys353 and Asp422, to form an active site.

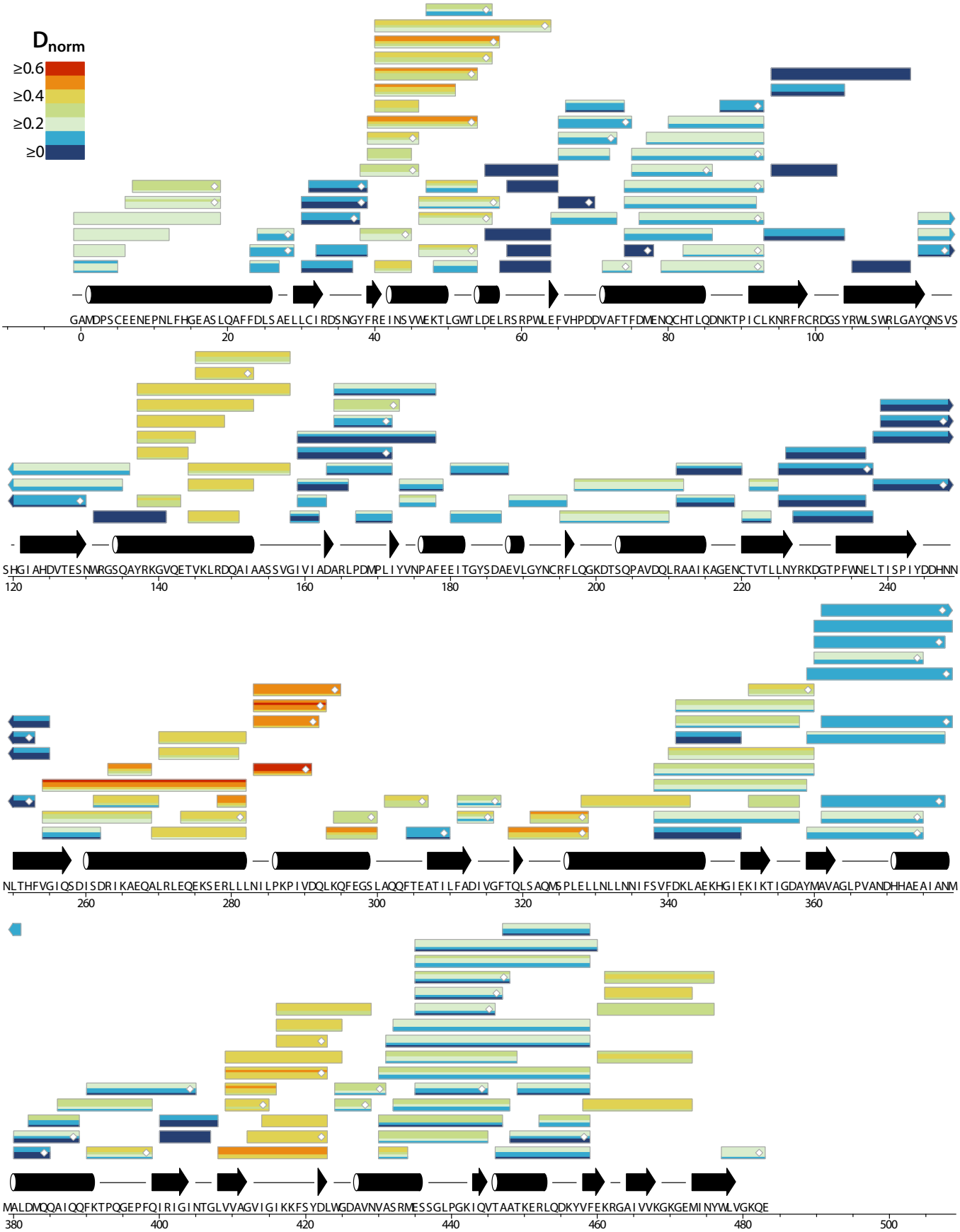



**Figure 3.16.:** Relative deuteration of dark-adapted mPAC. Deuterium incorporation is normalized to the maximum number of exchangeable protons in a peptide and color-coded accordingly from blue (low) to red (high).Each box indicates one peptide and contains up to five different colors for incubation times of 15,60,300,1200 and 3600 s, respectively, bottom-up. MS/MS confirmed peptides are marked with diamonds. Secondary structure is shown as predicted by Jpred [28]. Numbering follows the wild type sequence (Uniprot B4VKN6).

## 3.2.4. Structural Dynamics of mPAC

Hydrogen-deuterium exchange mass spectrometry (HDX-MS) was carried out to investigate how nucleotide binding and illumination affect the secondary structure dynamics of mPAC and to trace the path of structural changes that connect the flavin chromophore to the AC active site. For this purpose, four series of steady-state HDX-MS measurements were performed in the dark and in the light, with and without 2.5 mM ApCpp.

### Structural Dynamics of mPAC in the Dark

Relative deuteration rates of dark-adapted mPAC show three domains with highly dynamic and likely unstructured linkers. The boundaries of the linker regions can be assigned to Lys139–Gly158 (PAS–LOV) and Arg263–Leu300 (LOV–AC). In addition to the interdomain linkers, two sites of markedly increased deuteration were identified within the domains. Amino acids Phe39–Val45 corresponding to the $\alpha2_{PAS}$ (or $C\alpha_{PAS}$) helix and Val410–Leu423 in the $\beta4_{AC}$–$\beta5_{AC}$ tongue incorporate deuterium at higher rates than the surrounding peptides (Figure 3.16). The latter peptide corresponds to the part of the $\beta4_{AC}$–$\beta5_{AC}$ tongue that protrudes from the AC core domain and shows the largest degree of variation in the crystal structures.
Addition of ApCpp leads to pronounced stabilization of nucleotide binding elements in the AC domain, including $\beta1_{AC}$ around Asp313, the $\beta2_{AC}$–$\beta3_{AC}$ loop harboring Asp357, $\alpha4_{AC}$ and most significantly, the dynamic part of the $\beta4_{AC}$–$\beta5_{AC}$ tongue is protected by nucleotide binding (Figure 3.17). Exchange kinetics of the AC domain largely fall into the EX2 regime [102] except for the peptide Leu351–Met360 that exists in an equilibrium between a slowly exchanging conformation and one that exchanges faster. Conversion of the slow-exchanging conformation into the faster one is severely slowed by ApCpp binding via Asp357 in this peptide, giving rise to bimodal EX1 exchange kinetics (Figure 3.17b).
The PAS–LOV linker exhibits moderate protection at early exchange time points, indicating a more structured conformation that is in rapid equilibrium with a less protected one (mostly localized to the amides of the peptide 145–158). The effect of nucleotide on the LOV–AC linker region is much more pronounced. Analysis of overlapping peptides assigns the most significant deuteration differences to the amides 260–293 with the peptide Leu270–Leu282 showing a difference of 4.3 deuterons after 300 seconds of

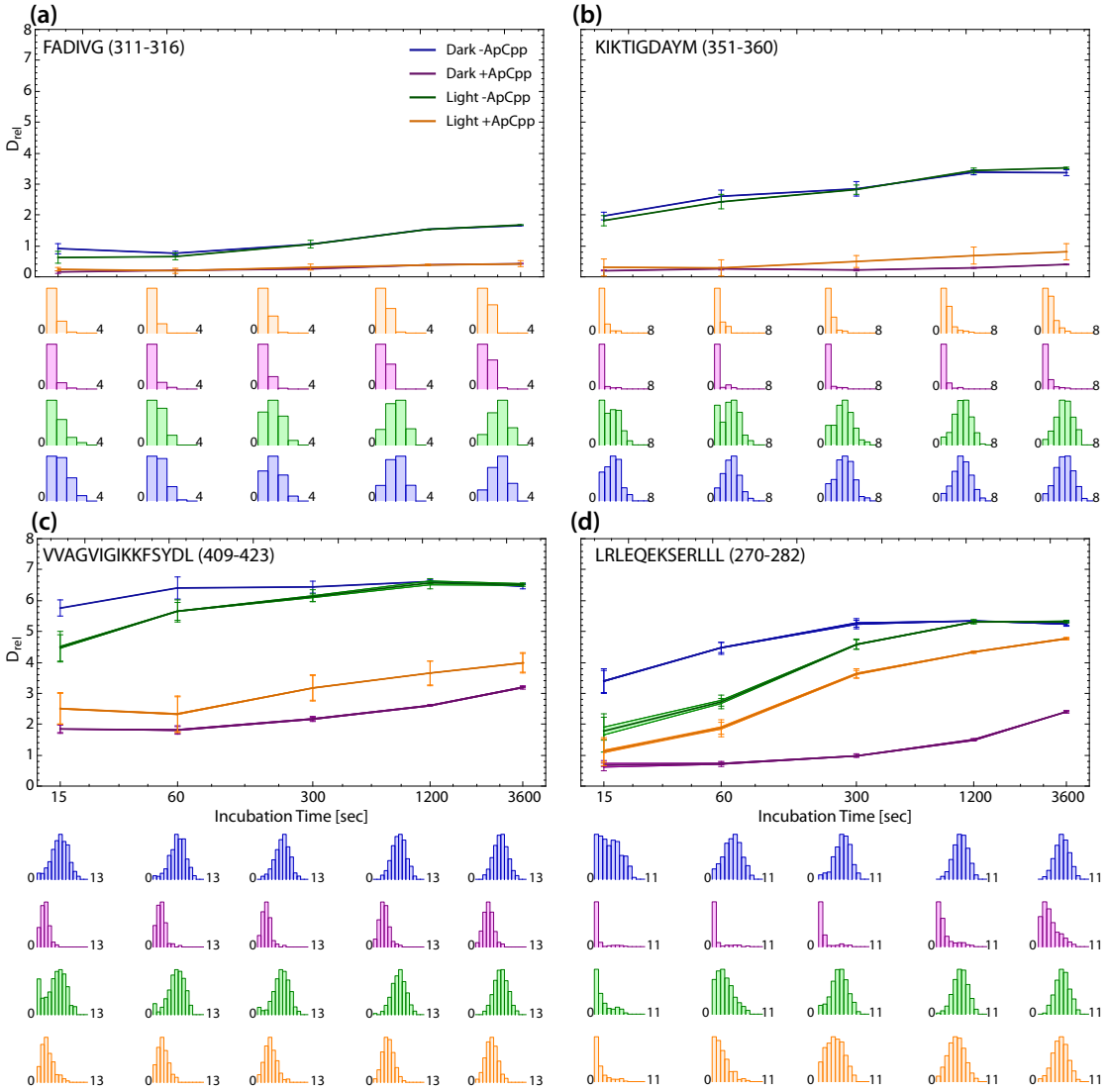



**Figure 3.17.:** Deuteration of selected mPAC peptides. Four labeling experiments were carried out: mPAC was kept in the dark in the absence (blue) and presence (purple) of ApCpp, or under constant blue light in the absence (green) or presence (orange) of ApCpp. Relative deuteration is plotted against the incubation time in $D_2O$ containing buffer. Estimated deuterium incorporation distributions for each timepoint are shown below the respective graphs. **(a)** The peptide Phe311–Gly316 ($\beta1_{AC}$) contains the metal-binding Asp313. **(b)** The peptide Lys351–Met360 ($\beta2_{AC}$–$\beta3_{AC}$) harbors the metal-binding Asp357 through which ApCpp is bound. **(c)** Peptide Val409–Leu423 corresponds to the part of the $\beta4_{AC}$–$\beta5_{AC}$ tongue that protrudes from the AC domain. **(d)** The peptide Leu270–Leu282 is predicted to constitute the C-terminus of $\alpha6_{LOV}$ ($J\alpha_{LOV}$) helix linking the LOV-domain to the AC domain. This region is strongly protected by binding of ApCpp to the AC domain and shows bimodal deuteration distributions. Peptide maps showing deuteration differences upon nucleotide binding and illumination in the presence or absence of ApCpp are shown in Figures A.7–A.9. All time series of deuterium incorporation are shown in Figure A.10.

exchange (Figure 3.17d). These peptides are predicted to belong to the C-terminus of the $\alpha6_{LOV}$ ($J\alpha_{LOV}$) helix, a short loop and a short helix. This region corresponds to the $^{280}$LLLNILPXXI[AV] motif that is highly conserved in type IIIa AC. This motif, termed cyclase transducer element (CTE) [234] has been proposed to form a coiled coil (broken by Pro286) and a handle-like structure, similar to bPAC and other type IIIb ACs in which the sequence is not as strictly conserved [100, 115].
Peptides belonging to the C-terminus of the predicted coiled coil clearly show bimodal deuteration distributions. This indicates an EX1 deuteration regime with two slowly interconverting conformations, one of which rapidly incorporates deuterium whereas the other is largely protected from exchange. Assuming strict EX1 kinetics [102], one can draw conclusions about the lifetime of the protected conformation: In the absence of ApCpp the protected conformation is only observable in the first exchange timepoint (15 s), indicating that a large fraction of the mPAC population has sampled the unprotected confirmation after this time. Addition of ApCpp strongly stabilizes the protected conformation, as the majority of mPAC molecules has not sampled the unprotected conformation after 3600 s.

### Structural Dynamics During Illumination and Catalysis

Illumination of mPAC in the absence of ApCpp causes a moderate change in deuterium incorporation that is mostly limited to the LOV domain. The flanking $\alpha1_{LOV}$ and $\alpha6_{LOV}$ ($A'\alpha_{LOV}$ and $J\alpha_{LOV}$, respectively) helices experience protection in the light. This contradicts the common conception of $J\alpha_{LOV}$ unfolding upon illumination that is mainly based on NMR studies on oat phototropin LOV2 [68, 67] and has become a paradigm in LOV signaling.
Illumination protects the highly dynamic $\alpha2_{PAS}$ ($C\alpha_{PAS}$) helix from exchange at early timepoints. Inspection of deuteration distributions shows that the protective effect is caused by stabilization of a conformation that is protected from exchange. This conformation must be relatively short-lived as it is no longer detectable after 60 s of exchange. A similar effect is observed in the peptide Val409–Leu423 that corresponds to the part of the $\beta4_{AC}$–$\beta5_{AC}$ tongue that protrudes from the AC domain: Illumination stabilizes a protected conformation that remains detectable for at least 300 s (Figure 3.17c). The exchange kinetics of the protected conformation are reminiscent of nucleotide-bound

mPAC, indicating that illumination may slightly stabilize a conformation of the $\beta4_{AC}$–$\beta5_{AC}$ tongue that resembles a nucleotide-bound conformation. The $\beta1_{AC}$ strand with the metal binding Asp313 is not affected by illumination in the absence of ApCpp and the $\beta2_{AC}$–$\beta3_{AC}$ loop harboring Asp357 shows only a slight shift of the deuteration distribution, indicating a protected population at early time points.

In order to obtain the most representative view of AC photoactivation by the LOV domain, nucleotide-bound mPAC was illuminated with blue light during $D_2O$ labeling. Comparison of deuterium incorporation into nucleotide-bound mPAC in the dark and in the light highlights the C-terminus of the predicted coiled coil as major site of light-induced unprotection (Figure 3.17d). Deuteration kinetics of the coiled coil in the nucleotide-bound protein are significantly different from nucleotide-free protein that also incorporates deuterium at high rates. Illumination of nucleotide-bound mPAC does not decrease the lifetime of the protected conformation but instead it increases the frequency at which hydrogen bonds in the protected conformation are transiently breaking.

The $\beta4_{AC}$–$\beta5_{AC}$ tongue of the AC domain is the only part of the AC domain that shows significant illumination-induced changes in deuteration (Figure A.9). Interestingly, the effect of illumination on deuteration kinetics of the $\beta4_{AC}$–$\beta5_{AC}$ tongue is almost identical to the coiled coil and the handles as the kinetics correspond to more frequent breathing of the nucleotide-stabilized protected conformation (Figure 3.17c and d). Nucleotide-binding secondary structure elements near Asp313 and Asp357 show no distinct response to illumination in the presence of nucleotide, indicating that light is not required for nucleotide binding to the protein.

### 3.2.5. Low-Resolution Solution Structure of mPAC

The overall domain arrangement of mPAC and its response to nucleotide binding and illumination were investigated using small angle X-ray scattering (SAXS). Data was collected for full length mPAC and $mPAC_{\Delta132}$ in the presence and absence of ApCpp. The $mPAC_{\Delta132}$ V160I variant behaves as the $mPAC_{\Delta132}$ construct except for its markedly increased lit state half-life of 2900 s, and was used to measure scattering of the lit state. Scattering curves and Kratky plots are shown in Figure 3.18a–d.

Samples containing ApCpp show more pronounced features in the scattering curves and Kratky plots than those without nucleotide. Illumination of the nucleotide-bound pro-

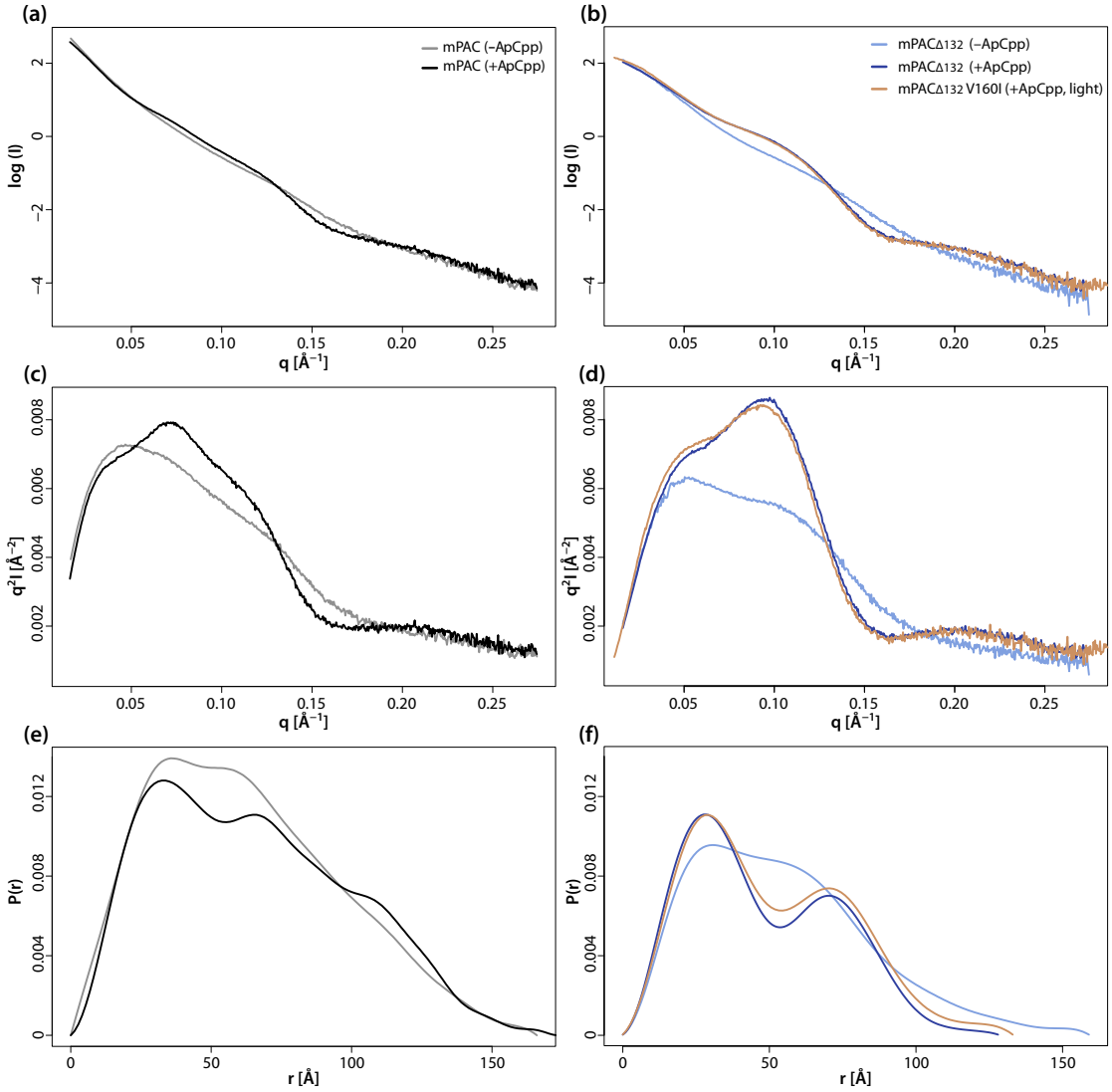



**Figure 3.18.:** Small angle X-ray scattering of mPAC (panels **a**,**c**,**e**) and mPAC$_{\Delta 132}$ (panels **b**,**d**,**e**). **(a)** Logarithmic scattering curve of mPAC in the absence (gray) and presence (black) of ApCpp. **(b)** Logarithmic scattering curve of mPAC$_{\Delta 132}$ in the absence (light blue) and presence (dark blue) of ApCpp, and of pre-illuminated mPAC$_{\Delta 132}$ V160I in the presence of ApCpp (orange). **(c)** Kratky plot of mPAC in the absence (gray) and presence (black) of ApCpp. **(d)** Kratky plot of mPAC$_{\Delta 132}$ in the absence (light blue) and presence (dark blue) of ApCpp, and of pre-illuminated mPAC$_{\Delta 132}$ V160I in the presence of ApCpp (orange). **(e)** Estimated pairwise distance distribution for mPAC in the absence (gray) and presence (black) of ApCpp. **(f)** Estimated pairwise distance distribution for mPAC$_{\Delta 132}$ in the absence (light blue) and presence (dark blue) of ApCpp, and of pre-illuminated mPAC$_{\Delta 132}$ V160I in the presence of ApCpp (orange).

**Table 3.3.:** Structural parameters of mPAC derived from small angle X-ray scattering. $R_g$ corresponds to the real space radius of gyration determined by GNOM [198] analysis. ApCpp concentration was 1 mM and the $mPAC_{\Delta 132}$ V160I variant was pre-illuminated for 5 min in the reaction vessels and for another 1 min in the capillary with 6 mW cm$^{-2}$ at 450 nm, and measured immediately.

| Construct | ApCpp | Light | $R_g$ [Å] | Porod Volume [Å$^3$] |
|---|---|---|---|---|
| **mPAC** | – | – | $50.2 \pm 0.6$ | 237,000 |
| | + | – | $52 \pm 1$ | 179,000 |
| **$mPAC_{\Delta 132}$** | – | – | $43.7 \pm 0.2$ | 141,000 |
| | + | – | $38.1 \pm 0.7$ | 127,000 |
| **$mPAC_{\Delta 132}$ V160I** | + | + | $39.3 \pm 0.2$ | 125,000 |

tein had only little effect. The Kratky plot shows one peak and two shoulders for mPAC whereas the Δ132 variant gives rise to one peak and one shoulder (Figure 3.18c,d). The occurrence of multiple maxima in the Kratky plot is often observed in proteins with multiple globular domains. A slow decline of $q^2I$ at higher $q$ values in the Kratky plot, as observed in the samples without ApCpp, is often indicative of flexible or unfolded parts [161].

In order to obtain a structural interpretation of the scattering data, an indirect Fourier transform was computed using GNOM [198]. The resulting pair distance distribution functions are shown in Figure 3.18e and f. Resulting real space radii of gyration and associated Porod volumes are shown in Table 3.3. Radii of gyration of globular proteins scale with approximately $N^{0.6}$ (with N being the number of amino acids) and varying offsets [76, 187]. Assuming dimers of mPAC and $mPAC_{\Delta 132}$, the determined radii of gyration suggest a non-globular, elongated shape as globular equivalents of mPAC and $mPAC_{\Delta 132}$ are expected to show $R_g$ values of less than 35 and 30 Å, respectively.

The pairwise distance distribution for mPAC shows two maxima, followed by a flat decline to a maximum particle diameter of 170 Å. The sample containing ApCpp shows three more defined distance maxima at 30–35 Å, 60–65 Å and 100–105 Å. Distance distributions of $mPAC_{\Delta 132}$ show a similar response to ApCpp, however the distance distribution of nucleotide-bound $mPAC_{\Delta 132}$ contains only two very clearly defined maxima at 30 Å and 70 Å. Distance distribution maxima are slightly less clearly separated in the illuminated sample which may indicate more conformational flexibility or an overall

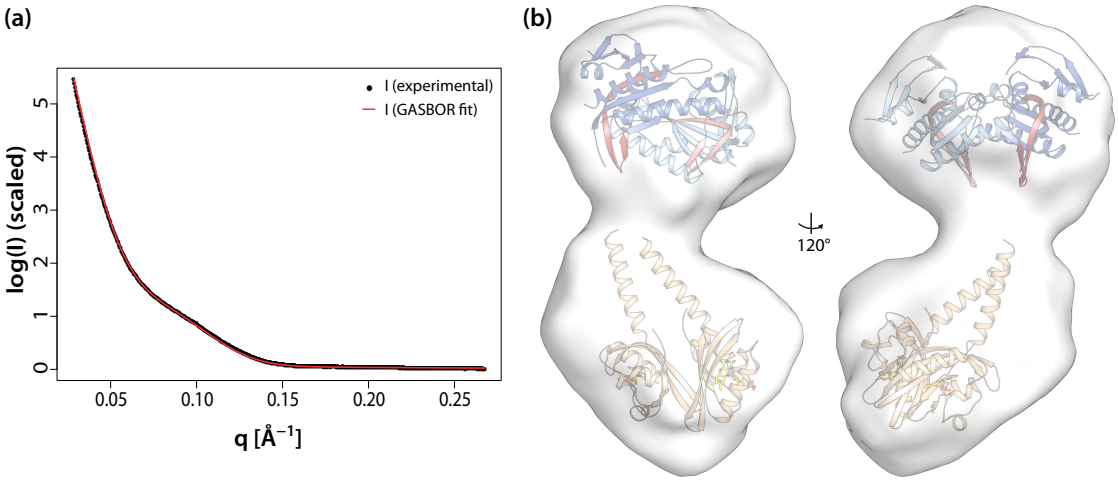



**Figure 3.19.:** SAXS-based 3D reconstruction of mPAC$_{\Delta 132}$. **(a)** Scaled experimental data (black dots) and fit (red line) from GASBOR [199] reconstruction. **(b)** The molecular envelope is based on an ensemble of ten GASBOR runs that were averaged. Structures of an AC domain dimer (blue, with the β4$_{AC}$–β5$_{AC}$ tongues colored in red) and a LOV domain dimer (yellow) were manually placed in the density for visualization of domain size and scale. The AC structure is based on the trimeric structure of the mPAC$_{AC}$ SeMet derivative and was modeled as a dimer based on the structure of mammalian transmembrane AC [203]. The LOV domain dimer shows the structure of YtvA that has high sequence similarity to the mPAC LOV domain [137].

poorer sample quality. These frequently occurring distances likely represent intra- and inter-domain distances. The shortest distance corresponds to the distances within the LOV, PAS and AC domains; the intermediate distance indicates the average distance between neighboring domains and the largest distance, only present in full length mPAC, describes the average distance of the PAS and AC domains.

It can be assumed that the AC domain is a globular dimer because the protein conformation clearly responds to ApCpp. Given an AC dimer, the occurrence of three most frequent distances in mPAC and two preferred distances in mPAC$_{\Delta 132}$ is only compatible with a parallel N-to-C terminal domain arrangement in nucleotide-bound mPAC. Pronounced minima in the distance distributions further indicate that mPAC domains likely do not have extensive interfaces but are instead connected by rigid linkers that keep domains at a well-defined distance. Broad distance distributions with less pronounced maxima in the absence of ApCpp indicate that the rigid linkers likely only form upon nucleotide binding.

The pairwise distance distribution of mPAC$_{\Delta 132}$ in the dark showed very sharp features

and provided sufficient quality for a bead model shape reconstruction. GASBOR [199] was used to generate an ensemble of ten reconstructions that were averaged. The resulting molecular envelope is shown in Figure 3.19 together with manually fitted models of LOV and AC domain dimers. The molecular envelope shows two globular domains linked by an elongated region that is compatible with a coiled coil as predicted by Jpred [28] and COILS [121].

### 3.2.6. High-Resolution Homology Models of mPAC$_{\Delta 132}$

#### Homology-Based Model of LOV Photoactivation in mPAC

The extended stalk in the low resolution SAXS reconstruction of mPAC$_{\Delta 132}$ and homology considerations strongly suggest that the Jα helix extends away from the LOV domain dimer and forms a coiled coil, resulting in an architecture that is similar to YtvA [137], PpsB1 [26, 167] or YF1 [36]. These proteins dimerize via the N-terminal A′α helix that is wedged between the β-sheets of both LOV domains. mPAC likely also dimerizes through A′α as the predicted A′α helix contains hydrophobic residues in positions that can interact both with each other and the LOV β-sheets.
Combination of published data on homologous LOV domains with results from this work allows proposing a signaling pathway from the flavin chromophore to the coiled coil in mPAC. Homology models of mPAC were generated based on the crystal structures of YF1 and PpsB1 (Figure 3.20) [36, 26, 167]. Protonation of the flavin N5 in the lit state causes the sidechain of the conserved Gln257 to rotate by 180° and to move due to the lost interaction with the flavin C4 carbonyl oxygen [167, 31, 145]. The shifted Gln257 likely engages in repulsive interactions with Phe178 (Cα) that is within van der Waals distance of Gln257 (3.9 Å) in the dark. The corresponding residue of YtvA (Phe46) flips in illuminated crystals [137] and experiences large chemical shift changes in illuminated PpsB1 [167]. Rotation of Gln257 may further affect Val160 (Aβ) in a similar fashion. The Aβ-strand is located between Iβ and Cα and repulsive interactions of the sidechains likely push the backbones apart as seen in the loss of hydrogen bonding between the backbone of Gln257 (Iβ) and Gly158 (Aβ) [167]. Although suggestive from the experimental structures of PpsB1, YtvA and the mPAC model, it is not clear if distortion of the β-sheet geometry is exclusively caused by steric repulsion because (*i*) mutation of Gln257 to asparagine, reducing the volume of the sidechain, locks EL222 (likely also

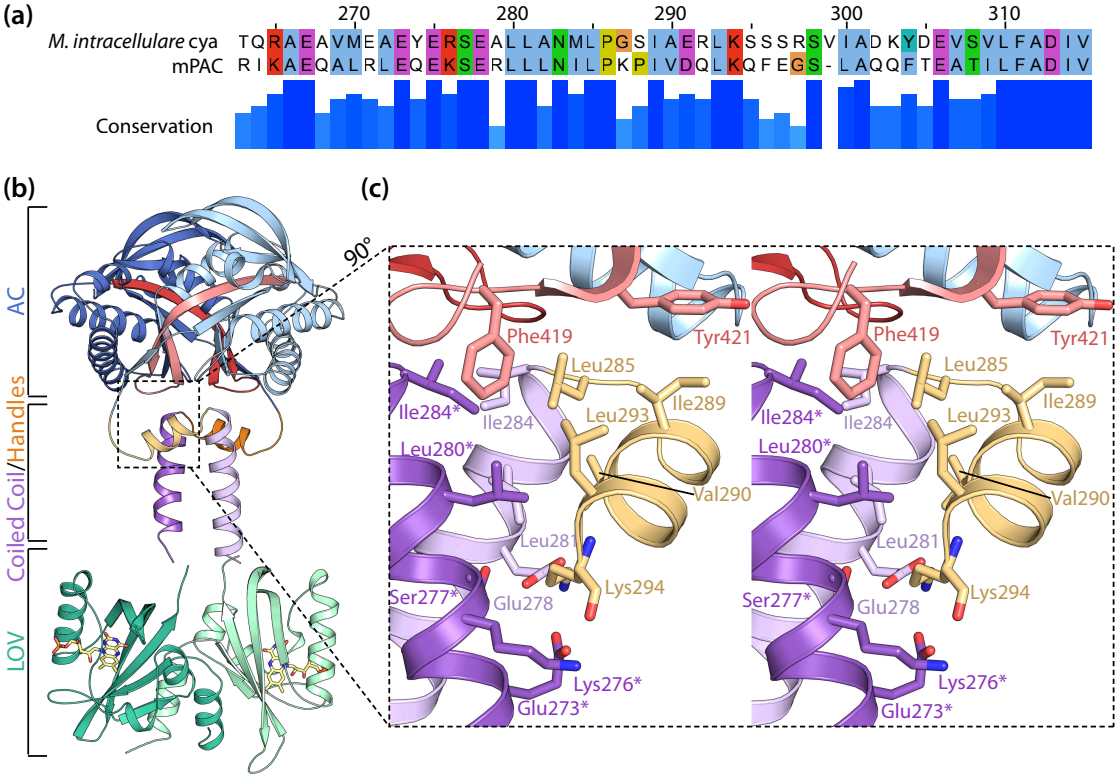



**Figure 3.20.:** High-resolution homology model of $mPAC_{\Delta 132}$. **(a)** Sequence alignment of the *Mycobacterium intracellulare* cya coiled coil and handle region with mPAC. Residue numbering follows mPAC. **(b)** Overall domain architecture of $mPAC_{\Delta 132}$ based on the LOV domain dimer of YF1 [36] (green, PDB 4GCZ) and the coiled coil (purple), handles (orange) and AC domain (blue) of *M. intracellulare* cya [209] (PDB 5O5L). Chains A and B of the homodimer are indicated in light and dark shades of the respective colors. **(c)** Stereoscopic detail view of the coiled coil and handle with mPAC sidechains. Glu273*, Lys276*, Ser277* in the central part of the coiled coil form a network of polar interactions within the coiled coil (Glu278) and the handle (Lys294). The C-terminal part of the coiled coil has a hydrophobic core formed by residues from both chains: Leu280, Leu281, Ile284, Leu285, Ile289, Val290 and Leu293 from both chains constitute an extensive network of hydrophobic interactions. Phe419 ($\beta 5_{AC}$) inserts into this cluster and functions as a lid to protect the hydrophobic core from solvent. Tyr421 opposite the base-specifying Asp422 likely also shields the coiled coil from solvent.

mPAC as $mPAC_{Q257N}$ was toxic to the cells) in a pseudolit conformation [142], and (*ii*) mPAC variants with increased sidechain volumes of Gly158 and Val160 show increased lit-state stability as evidenced by slow dark state recovery (cf. section 3.2.7).

Regardless of the origin of structural changes in Aβ, Ile159 is of particular interest as it is located in center of the conformational changes created by rotation of Gln257. Its

outward-facing sidechain inserts into the rather loosely packed hydrophobic core formed between the A′α helices, Iβ and Aβ at distances between 4.3 and 4.7 Å. Even small conformational changes of Ile159 may lead to the collapse of the packing as observed in the crystal structure and NMR experiments of light-adapted PpsB1 [167]. The A′α helices shift by one-third of a turn with respect to each other and the opposing molecule's β-sheet, resulting in a newly formed hydrophobic interface. The lit-state interface involves a higher A′α crossing angle and interaction of the β-sheets with each other. Crystal structures of PpsB1, SAXS-guided molecular dynamics simulations and EPR spectroscopy show that a change in the A′α angle acts as a lever that moves the Iβ–Jα junctions apart by 2–3 Å [167, 46, 11]. While these structural changes are largely conserved and backed up by solid experimental data, compelling evidence for a molecular model of downstream events is lacking. Simulations attempting to recapitulate time-resolved SAXS measurements on YF1 suggest that separation of the Jα N-termini induces torque on the coiled coil and leads to supercoiling of its C-terminal segment [10, 46] accompanied by a register shift. While the hydrophobic face of the Jα coiled coil in mPAC is in principle large enough to accommodate such a register shift, it is not clear if this is compatible with the dimeric assembly of the AC domain that is tightly linked to the coiled coil. SAXS experiments showed no major structural differences between $mPAC_{\Delta 132}$ in the dark and in the light. Similarly, no large conformational changes were observed in SAXS measurements of the homologous YtvA protein [157].

### Structure and Function of the LOV–AC Linker

The sequence of the LOV–AC linker region contains a highly conserved motif recently described as cyclase transducer element (CTE) [234]. Based on crystal structures of full-length type IIIb ACs in which the sequence motif is less strictly conserved, Ziegler *et al.* proposed that all type IIIa and IIIb cyclases possess a handle-like structure preceding the AC domain as observed in bPAC and other type IIIb ACs [234, 115, 149, 100]. Although the low-resolution solution structure of mPAC and the secondary structure prediction suggest a coiled coil followed by handles in mPAC, a molecular model could not be constructed because no crystal structure of a full-length type IIIa AC was available and the proposed coiled coil sequence shows no preferred heptad pattern. The available type IIIb handle structures are not suitable as templates for a homology model with molecular

resolution because they constitute a diverse group with little sequence similarity to the highly conserved type IIIa CTE sequences [234].
At the time of writing, a crystal structure of a soluble construct of the type IIIa AC cya from *Mycobacterium intracellulare* was published [209]. This construct contained the handles and parts of the coiled coil, thereby providing a template for the regulatory element linking the LOV and the AC domain of mPAC. Using the coordinates of *Mycobacterium intracellulare* cya and of the YF1 [36] protein, a high-resolution model of $mPAC_{\Delta 132}$, the minimal photoregulated mPAC construct, was generated. The overall shape and size of the molecule are in good agreement with the low-resolution data from SAXS experiments on $mPAC_{\Delta 132}$ in the presence of ApCpp (Figure A.6).
Figure 3.20 shows an overview and details of the regulatory interface in mPAC. The coiled coil is almost entirely conserved between mPAC and the template. In particular, the $^{273}$EXX[R/K]SE$^{278}$ motif connecting the helices to each other and to the side chain and the backbone of Lys294 in the handle, and the extensive cluster of hydrophobic residues in the helical bundle (Figure 3.20c) are nearly identical. The handles are coupled to the $\beta4_{AC}$–$\beta5_{AC}$ tongue through Phe419 ($\beta5_{AC}$) that inserts into the hydrophobic core of the helical bundle from the distal side and thereby serves as a lid that shields the hydrophobic core from water. Tyr421 next to the base-specifying Asp422 likely serves the same purpose.

## 3.2.7. Mechanism-Based Tuning of Enzyme Activity and Photocycle

Despite the favorable properties of the LOV photoreceptor domain with its short and tunable lit state half-life, mPAC is of limited use as optogenetic tool because of its high dark state activity and the moderate range of light activation. Reduction of mPAC dark state activity is therefore an essential prerequisite for *in vivo* application. Only when this is achieved, it may be useful to tune the LOV domain photocycle to match the kinetic requirements of specific experiments.

### Tuning the AC Dark State Reaction

The high dark state activity of mPAC is rooted either in an intrinsically high AC activity or in the inability of the LOV domain to effectively control the AC domain. Both factors appear to play a role as the isolated mPAC AC domain shows some activity in

the absence of the LOV domain and the interface between the LOV and the AC domain (specifically the C-terminal $J\alpha_{LOV}$ helix and the $\beta4_{AC}$–$\beta5_{AC}$ tongue) is highly dynamic in the dark, resembling lit-like conformational dynamics. There is no evidence that stabilization of secondary structures at the domain interface will increase the range of light activation. In fact, investigations of the bPAC photoactivation mechanism have shown that illumination does not affect nucleotide binding or creates an entirely novel conformation but instead, light shifts the conformational equilibrium of the nucleotide bound AC domain toward a near-attack conformation that occurs at a lesser frequency in the dark [14, 115]. Consequently, the issue of high dark state AC activity was addressed directly in the AC domain at the level of the dark state activation mechanism without interfering with nucleotide binding or light-activation.

The conserved Lys468 in the mobile $\beta7_{AC}$–$\beta8_{AC}$ hairpin was shown to move toward the active site upon nucleotide binding and is thought to assist in proton shuttling after deprotonation of the ribose 3'-OH, thereby effectively increasing its acidity [189, 64]. Interestingly, hydrogen–deuterium exchange of the $\beta7_{AC}$–$\beta8_{AC}$ hairpin responds weakly to nucleotide binding but not to illumination, hence it appeared to be a suitable target for addressing the dark reaction rate.

Indeed, substitution of Lys468 by serine (producing $mPAC_{KS}$) resulted in a 100-fold decrease of the dark state AC activity compared to wild type mPAC (Table 3.2). At the same time, the apparent light state rate constant at pH 7.5, 100 µM ATP and 6 mW $cm^{-2}$ blue light was only reduced approximately five-fold, resulting in a 20-fold improvement of the range of light activation. Under these conditions, the activity of $mPAC_{KS}$ can be increased 250-fold by illumination. At saturating ATP concentrations ($K_M$ = 108 ± 9 µM in the light and 20 ± 10 µM in the dark, Figure 3.21b), $mPAC_{KS}$ activity increases more than 700-fold in the light, reaching a light state maximum catalytic rate constant of 0.36 ± 0.01 $s^{-1}$ (12 % of maximum wild type velocity at ATP saturation). The $K_M$ value is similar to that of wild type mPAC, hence the mutation does not affect nucleotide binding properties of mPAC.

Consistent with a role of Lys468 in ribose deprotonation, the catalytic rate constant of $mPAC_{KS}$ was much more sensitive to changes in pH than the wild type. Activity was especially decreased under slightly acidic conditions (0.05 $s^{-1}$ in the light at pH 6.8) and less affected under more basic conditions (0.56 $s^{-1}$ in the light at pH 8.5, Figure 3.21a). Ser468 does not appear to exert a specific role in catalysis as the kinetic properties of the

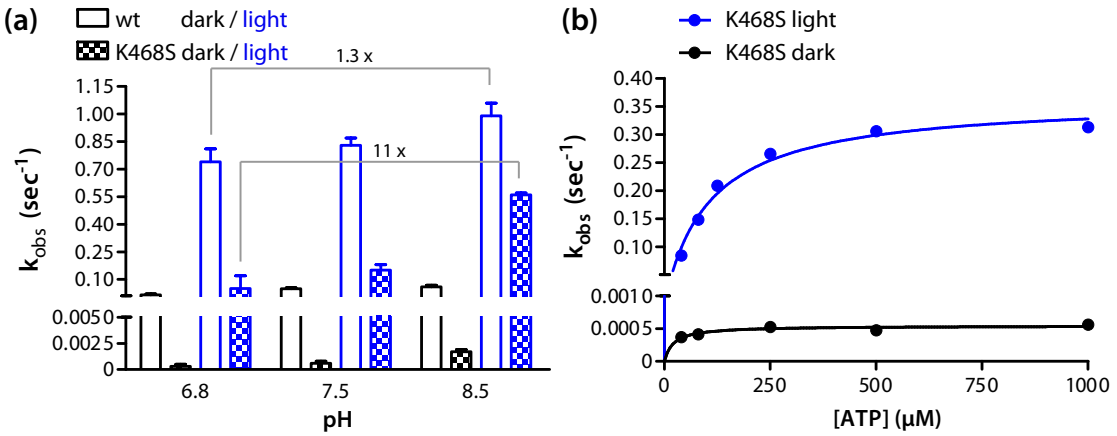



**Figure 3.21.:** Activity of mPAC K468S ($mPAC_{KS}$). **(a)** Apparent catalytic rate constants of mPAC and $mPAC_{KS}$ in the dark and in the light at 100 µM ATP and pH values of 6.8, 7.5 and 8.5. The activity of $mPAC_{KS}$ is affected much more strongly by pH than the wild type's activity. **(b)** ATP-dependent activity of $mPAC_{KS}$. $K_M$ was determined to be $20 \pm 10$ µM in the dark and $108 \pm 9$ µM in the light, respectively. $v_{max}$ is $0.0005 \pm 0.0003\,s^{-1}$ in the dark and $0.36 \pm 0.01\,s^{-1}$ in the light, respectively.

variants K468Q and K468L are similar to those of $mPAC_{KS}$ (Table 3.2). Nonetheless, substitution of Lys468 by a polar residue resulted in slightly higher residual AC activity than in the K468L variant, possibly due to weak stabilization of a water molecule.

### Tuning of LOV Lit State Decay

The lifetime of the LOV lit state spans five orders of magnitude in different proteins [156, 25, 26] and there has been considerable interest in rational tuning of signaling state lifetime for optogenetic applications. As the LOV domain cannot be switched off with visible light, there is particular interest in accelerating dark state recovery for improved temporal control.

Lit state decay shows a pronounced solvent isotope effect [200, 30] and some LOV domains show pH-dependent decay kinetics [101], hence, deprotonation of the flavin N5 is commonly discussed as rate-limiting step. Accordingly, access of solvent molecules and small planar bases to the flavin binding pocket has been shown to affect the lit state lifetime, however results are mixed: lit state decay of LOV domains from Aureochrome or phototropin can be accelerated by imidazole several thousand-fold [2, 73] whereas other LOV domains such as mPAC are only moderately affected [94]. In the light of

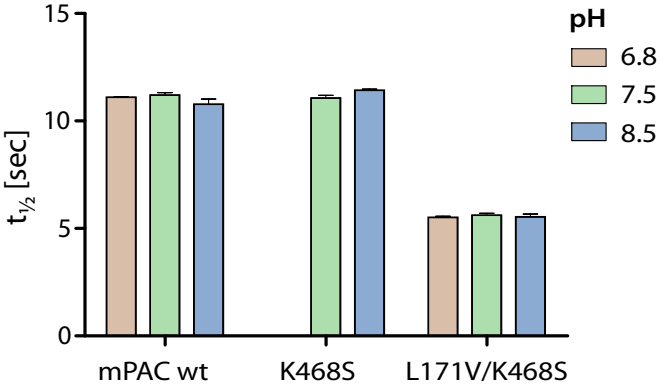



**Figure 3.22.:** Lit state half-lives of mPAC variants and pH dependence. mPAC wild type and mPAC K468S have identical lit state half lives of 11.2 s. The lit state of mPAC K468S L171V decays approximately twice as fast with a half-life of 5.6 s. None of the variants show pH-dependent dark state recovery kinetics in the range of pH 6.8–8.5.

this heterogeneity, it is not clear which structural factors govern lit state decay kinetics and how to modify them by rational engineering. For this reason, variants of $mPAC_{KS}$ corresponding to those reported to affect the lit state lifetime were screened in cleared lysate. Table 3.4 lists the properties of the produced variants and the lit state half lives determined in lysate.

The probably best-described decay-accelerating variant L171V originates from large-scale screening on LOV2 (I427V) [25] and consistently accelerates dark state recovery in all tested LOV domains [156], including mPAC ($t_{½}$ = 5 s). The corresponding mutation (I85V) renders VVD lit state decay kinetics sensitive to pH from which the authors concluded that it acts by facilitating access of water to the active site [237, 17], however no such effect could be detected in $mPAC_{L171V}$ (Figure 3.22). Instead, it appears that the flexibility of the flavin binding pocket and the rigidity of flavin binding to play an important role for the stability of the lit state. Using a very invasive strategy, Raffelberg *et al.* could accelerate the lit state decay of YtvA up to 45-fold by removing the double hydrogen bond between the chromophore and the conserved residue corresponding to Asn226 [158]. Neither of the corresponding mPAC variants N226A and N226S bind flavin, hence it might be more appropriate for mPAC to target residues that are less conserved or that do not directly interact with the photoactive isoalloxazine moiety of the flavin chromophore. The terminal phosphate of the FMN chromophore is typically

**Table 3.4.:** Photocycle screening in mPAC. Two rounds of screening were conducted. The first round explored literature-based variants and the second structure-based round aimed at further facilitating solvent access to the flavin N5 atom in mPAC K468S L171V. Culture growth, soluble mPAC expression and yellow color of the soluble protein fraction were monitored qualitatively on a scale from – (negative, absent) to +++ (very strong). Dark state recovery times were measured in cleared lysate.

| Variant | Growth | Soluble Expression | Yellow Color | $t_{½}$ [s] |
|---|---|---|---|---|
| K468S | +++ | ++ | +++ | 14 |
| **First Round: Literature-Based** | | | | |
| K468S G158A | + | +/++ | ++ | 17 |
| K468S V160T | + | + | +++ | 37 |
| K468S V160I | + | ++ | ++ | 2900 |
| K468S L171V | +++ | +++ | +++ | 5 |
| K468S L171A | +++ | ++ | – | n/a |
| K468S L171G | +++ | ++ | – | n/a |
| K468S N226S | + | + | – | n/a |
| K468S N226A | + | +/- | – | n/a |
| K468S Q257D | +/- | +/- | +/- | 14 |
| K468S Q257H | – | n/a | n/a | n/a |
| K468S Q257N | – | n/a | n/a | n/a |
| **Second Round: Structure-Based** | | | | |
| K468S L171V V160G | ++ | ++ | – | n/a |
| K468S L171V I214L | ++ | ++ | + | 20 |
| K468S L171V I214V | ++ | ++ | + | 5 |
| K468S L171V I238S | ++ | ++ | ++/+++ | 160 |
| K468S L171V I238T | ++ | ++ | +++ | 160 |
| K468S L171V I240L | ++ | ++ | ++/+++ | 14 |
| K468S L171V I240V | ++ | ++ | + | 4 |

coordinated by two arginine residues (Arg195 and Arg211 in mPAC). Two additional arginines were found in PpsB1 that recovers extremely slowly ($t_{½}$ = 100,000 s) [82]. Substitution of these residues resulted in up to 280-fold acceleration of lit state decay and conversely, introduction of an arginine into the corresponding positions of the homologous PpsB2 protein resulted in up to 10-fold prolonged lit state lifetimes [82, 26]. mPAC contains only the two canonical arginine residues and modification of these universally conserved residues was not attempted.

The strategy chosen for photocycle tuning of mPAC focused on nonconserved aliphatic

sidechains in the flavin binding pocket. The variants L171I, V160T and G158A increase lit state half-life by a factor of 1.5–2. Dramatic effects were for V160I that has a 260-fold increased lit state half-life. Val160 bridges the interaction network between the rotated conformation of Gln257, Leu171 and Phe178. Introduction of an isoleucine side chain likely fills the space and allows only little conformational flexibility between Gln257 and the protonated flavin chromophore. It would be interesting to investigate if the V160I substitution affects the stability of the Gln257–flavin N5 hydrogen bond and thereby delays deprotonation of N5. Conversely, shortening of nonconserved aliphatic sidechains to remove steric constraints in the flavin binding pocket has produced the double mutant $mPAC_{L171V,I214L}$ that resulted in 2–3-fold acceleration of lit state decay ($t_{½}$ = 4 s). Several variants such as L171A, L171G and L171V/V160G do not bind flavin, suggesting a fine balance between the plasticity of the flavin binding cavity and its ability to bind flavin. Given that $mPAC_{L171V,I214L}$ has one of the shortest lit state half-lives ever reported for a LOV domain and already shows slightly impaired flavin binding, it is unclear whether this balance can be pushed any further without losing chromophore binding.

# 3.3. Domain Assembly Modeling

## 3.3.1. Background

Experimental determination of multidomain protein structure faces many challenges and its computational prediction is likely to become a major task in theoretical structural biology. With the growing number of experimentally determined domain structures, the focus of the problem shifts from in-silico protein folding to the assembly of domains into tertiary and quaternary structures [3].

The domain assembly project presented in this work is based on a Rosetta modeling protocol developed by Oana Lungu in the laboratory of Brian Kuhlman at the University of North Carolina in Chapel Hill. It originally served to demonstrate that the crystal structure of a synthetic fusion protein, photoactivatable Rac (PA-Rac) [222], can be predicted, raising hopes for the rational design of further synthetic photoactivatable proteins. This prospect captured my attention which is why I spent two months in the group of Brian Kuhlman to learn about the basics of Rosetta modeling and to reactivate the domain assembly protocol for a broader range of applications.

I re-wrote the protocol to perform well not only at the task it had originally been designed for but on a more general set of problems. The protocol predicts the structure of a multidomain protein from structures of the individual domains and a linker sequence by explicit folding of the linker and evaluation of the entire structure's Rosetta score. This strategy is promising for the design of functional multidomain proteins for two reasons: Proper choice of the linker is considered the bottleneck of synthetic protein design [235], hence explicit treatment of interdomain linkers appears essential for modeling functional assemblies. Furthermore, as this protocol is implemented within the large Rosetta framework, it allows seamless integration of structural information from complementary techniques such as HDX-MS [34], SAXS [168, 188] or NMR spectroscopy [160, 178].

The performance of the protocol was assessed using a benchmarking dataset containing 216 experimentally determined structures. Thorough analysis of factors that determine prediction success or failure allows investigations about the type of assemblies that can be predicted, which principles of multidomain protein folding are captured by the Rosetta score function and whether such a protocol can help design synthetic multidomain proteins with desired structure and function.

### 3.3.2. Overall Prediction Performance

The first question relevant for the evaluation of a modeling protocol is whether the protocol is capable of retrieving the correct domain orientation. For this purpose, domain assembly predictions were run on a test set of 216 experimentally determined structures of two-domain proteins (*assembly problems*). 158 structures were selected in this work and 58 structures were taken from a previous publication on domain assembly prediction [218]. 5000 assemblies were predicted for each assembly problem and each method, for cases that were predicted using different sampling strategies. The quality of each predicted assembly (*decoy*) was quantified by its root mean square deviation (RMSD) from the experimental (*native*) structure. To judge whether an assembly problem was successfully solved by the set of 5000 decoys, the Rosetta score was used to generate two metrics: The *b5* score is the smallest RMSD from native out of the best five scoring decoys. The *b1p* score indicates the smallest RMSD among the best-scoring one percent of decoys. Using these measures, it can be evaluated whether low RMSD decoys are modeled and reliably identified as such by the Rosetta score function.

#### Accuracy of Predicted Domain Arrangement

The ability of the modeling protocol to correctly retrieve the relative domain orientation was quantified by the RMSD over all atoms. Table 3.5 contains a breakdown of b1p and b5 scores for the various datasets and modeling strategies. Overall, 136 out of 216 domain assembly problems (63 %) were predicted to a b5 value of less than 3 Å (Figure 3.23a). 48 assemblies were successfully modeled using a simple strategy employing small random moves, that was later discontinued. As the fragment insertion and ab initio protocols also contain small random perturbations, it can be assumed that all assemblies correctly predicted by small moves will also be correctly predicted by the other protocols.
Taken together, the accuracy the domain assembly protocol presented in this work compares favorably with previous studies of domain assembly: The ab initio based Rosetta protocol published by Wollacott *et al.* [218] yielded a b5 value of less than 3 Å in 55 % of the modeled assemblies; the restrained docking method AIDA [224] yields 68 % on the same test set. It has to be noted that the value of 74 % observed on the Wollacott/Zanghellini set in this study was obtained from a subset containing only 58 of the 76 original assembly problems that were modeled by the other studies. The

**Table 3.5.:** Accuracy of domain assembly predictions. Domain assembly was predicted for a total of 216 experimental structures (158 selected in this work and 58 from a previous study on domain assembly [218]). Assemblies that did not reach a b5 value below 3 Å RMSD using a simple small move protocol were modeled using either a 3mer insertion protocol or a full ab initio protocol. b1p indicates the lowest RMSD from native among the best scoring percentile of assemblies. b5 indicates the lowest RMSD from native among the five best scoring assemblies.

| Dataset | Assemblies | Assemblies with | | | |
|---|---|---|---|---|---|
| | | b1p < 3 Å | b1p < 5 Å | b5 < 3 Å | b5 < 5 Å |
| **Lindner Set** | | | | | |
| any method | 158 | 102 (65 %) | 121 (77 %) | 93 (59 %) | 106 (67 %) |
| 3mer insertion | 132 [a] | 70 (53 %) | 87 (66 %) | 62 (47 %) | 73 (55 %) |
| ab initio | 119 [b] | 70 (59 %) | 78 (66 %) | 64 (54 %) | 70 (59 %) |
| **Wollacott/Zanghellini Set** | | | | | |
| any method | 58 | 46 (79 %) | 48 (83 %) | 43 (74 %) | 46 (79 %) |
| 3mer insertion | 49 [c] | 34 (69 %) | 37 (76 %) | 29 (59 %) | 34 (69 %) |
| ab initio | 26 [d] | 14 (54 %) | 16 (62 %) | 13 (50 %) | 15 (58 %) |
| **Total** | 216 | 148 (69 %) | 169 (78 %) | 136 (63 %) | 152 (70 %) |

[a] not including 15 assemblies that were successfully modeled with a simpler, discontinued protocol
[b] not including 11 assemblies that were successfully modeled with a simpler, discontinued protocol
[c] not including 4 assemblies that were successfully modeled with a simpler, discontinued protocol
[d] not including 11 assemblies that were successfully modeled with a simpler, discontinued protocol

missing problems were removed due to quality issues with the experimental structures, including poor modeling, interdomain disulfides (e.g., 1C2A, 1LBU), posttranslational phosphorylation in the domain interface (e.g., 1EUD), structural homology (1GV1 and 1HYE) or unclear linker boundaries.

To compare the performance of 3mer insertion and ab initio sampling, the b5 values of 123 assembly problems modeled with both methods were evaluated (Figure 3.23b). Out of these 123 problems, 53 could be modeled to a b5 value of less than 5 Å by both methods, 5 were only solved by 3-mer insertion, 10 were only solved by ab initio sampling, and 55 could not be solved to a b5 value of less than 5 Å by either method. b5 values (the same applies for b1p values) were linearly correlated between the two methods (Pearson $r$=0.84). Although the ab initio method appears to retrieve the native assembly slightly more frequently (51 %) than 3mer insertion (47 %) in this subset, statistically significant differences between the methods' accuracies cannot be established.

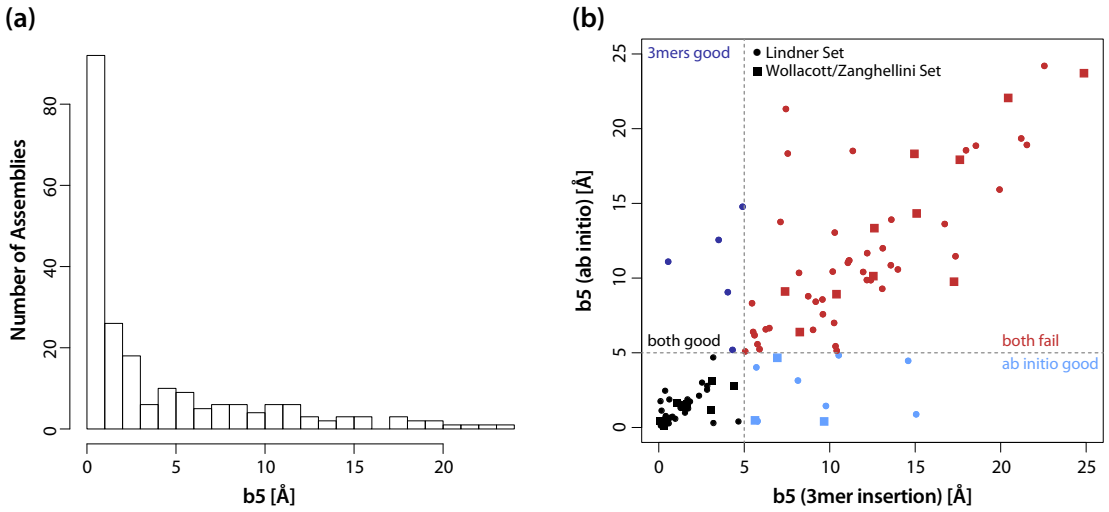



**Figure 3.23.:** Performance of different domain assembly strategies. **(a)** Distribution of the best b5 values reached for each of the 216 modeled domain assembly problems. 136 assemblies were modeled to a b5 value of less than 3 Å, 152 assemblies reached a b5 value of less than 5 Å. **(b)** Comparison of b5 values from the 3mer insertion strategy and the more expensive ab initio strategy. Circles (●) represent assembly problems from this work, squares (■) represent assembly problems from Wollacott *et al.* [218]. Each data point represents one assembly problem that has been run with 3mer insertion and ab initio sampling. Problems that were not solved to a b5 value of 5 Å by any of the two methods are indicated in red. Problems that were solved by ab initio sampling but not by 3mer insertion are indicated in light blue. Problems solved by 3mer insertion but not by ab initio sampling are indicated in dark blue. Problems solved by both sampling methods are indicated in black. b5 values obtained by both methods are highly correlated (Pearson $r$=0.84). The data displayed in panel **(b)** are a non-representative subset of all tested domain assembly problems.

### Accuracy of Predicted Linker Structure

The domain assembly protocol developed in this work models the domain orientation by explicit treatment of the linker that is evaluated by the same score function as the rest of the assembly. This is done in the attempt to obtain an accurate model of the linker structure that might be relevant for domain interaction. Accuracy of linker modeling was quantified by the linker b5 score that indicates the lowest linker Cα RMSD of the five best overall scoring decoys.

Overall b5 scores hold no predictive value for linker b5 scores (Figure 3.24a). In particular, good predictions (indicated by low overall b5) can be achieved irrespective of the linker's RMSD form native. This is in line with the observation that in some cases, prediction of the overall assembly could be improved by using a linker sequence that did

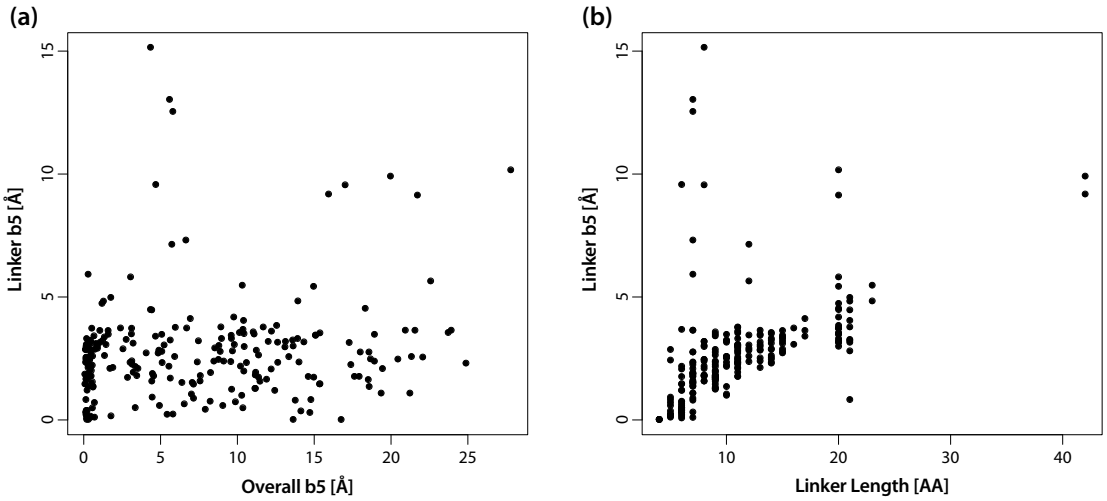



**Figure 3.24.:** Accuracy of linker predictions. **(a)** Relationship of overall b5 scores and linker b5 scores. There is no correlation between the accuracy of the predicted overall assembly and that of the linker. **(b)** The linker length, i.e., the number of degrees of freedom is the most significant predictor of the linker b5 value.

not correspond to the native one [218]. The linker b5 score correlated significantly with the linker length (Figure 3.24b). Analysis of variance indicates that linker length, and not linker volume, compactness, buried surface area or secondary structure content, is the most significant predictor of the linker b5 score, suggesting that the protocol exploits the linker's conformational degrees of freedom to the full extent, with little consideration of its contribution to the overall Rosetta score.

## 3.3.3. Protocol Benchmarking

In the following section, the domain assembly protocol is dissected to address three questions: (*i*) what makes a problem easy or hard, (*ii*) which stage of the assembly protocol sets the course for success or failure and (*iii*) how much computing power is required to reliably solve a problem?

### Features Affecting Predictability of an Assembly

It has been observed in this work and by others [224] that the accuracy of predictions depends strongly on the chosen test set. In particular, the Wollacott/Zanghellini [218]

test set appears to contain a disproportionately large number of "easy" problems as both the AIDA method [224] and the `DomainAssemblyMover` developed in this work perform significantly better on this set than on the sets of assembly problems chosen in the respective studies. It is therefore sensible to investigate which features of an assembly problem affect the performance of the modeling protocol.

Intuition suggests that a larger search space, in this case represented by linker length, should adversely affect the accuracy of the prediction. While this was observed in the restrained docking method AIDA [224], there is no indication that the degrees of freedom in the linker have an effect on the accuracy of the `DomainAssemblyMover`. Merely the accuracy of the predicted linker structure decreases with linker length (cf. section 3.3.2 and Figure 3.24b), hence this simple assumption must be dismissed.

The assembly problems were therefore characterized by a set of 37 additional structural measures (Table 2.11) to obtain quantitative insight into structural determinants of prediction accuracy. Statistical testing was performed to detect features that differ between clearly solved (b5 < 3 Å) and clearly failed (b5 > 5 Å) prediction problems. Figure 3.25 shows the distribution of all features that were found to be associated with prediction success. The most significant predictors of modeling success are related to interactions at the interdomain interface: buried interdomain surface area (BSA_AB, $p = 4 \times 10^{-6}$), interface $\Delta\Delta G$ (DDG_AB, $p = 3 \times 10^{-5}$), the total number of interdomain interface residues (INTERFACE_RES_AB, $p = 2 \times 10^{-3}$) and hydrogen bonds (NHB_AB p = 0.02, but not salt bridges, NSB_AB, p = 1.0). Interestingly, the surface area buried in the linker–domain interface (BSA_L) is not associated with prediction accuracy although it is in the same size range (interquartile range 580–900 $Å^2$) as the buried surface area between the domains (interquartile range 450–950 $Å^2$). The same applies to interface binding energies (interquartile ranges of -11.6 to -3.0 kcal $mol^{-1}$ between the domains and -11.2 to -5.0 kcal $mol^{-1}$ between the linker and the domains). Strong interactions at the interdomain interface clearly facilitate modeling already at the low-resolution centroid stage of the protocol, by forming a deep well in the energy landscape that guides the sampling.

Two features associated with fragment quality highlight the importance of proper sampling at the centroid stage: both 9-mer and 3-mer fragment RMSD from the corresponding native linker peptides were significantly associated with modeling success (FRMS_9, p = 0.02, and FRMS_3, p = 0.03, respectively). In particular, the 9-mer RMSD shows a

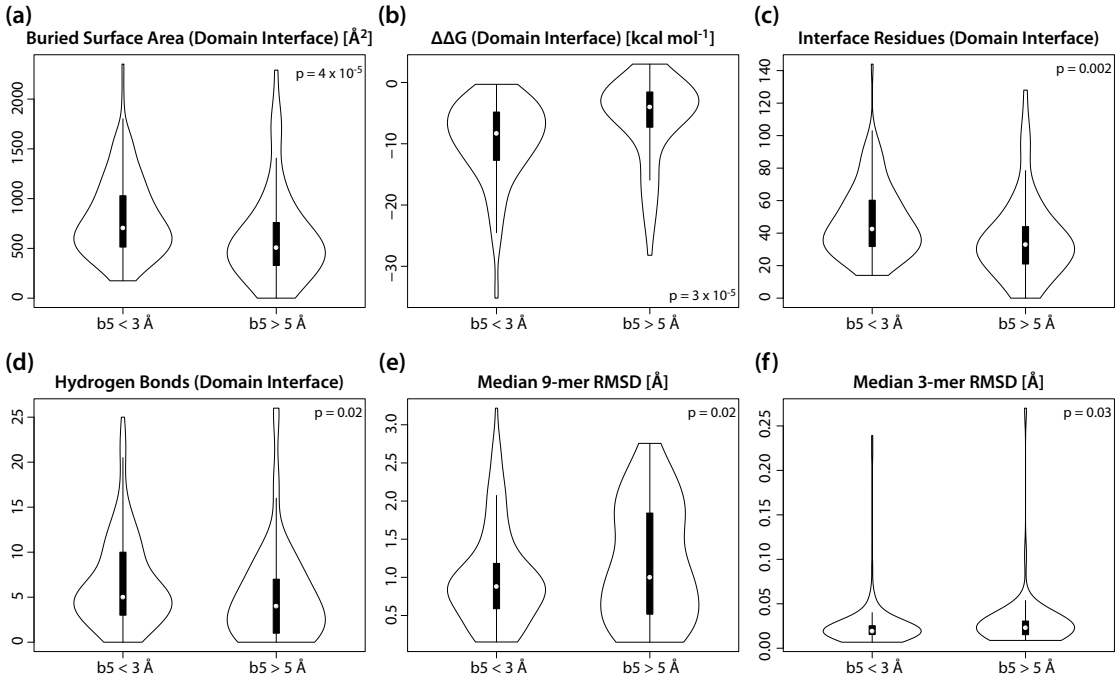



**Figure 3.25.:** Structural features significantly associated with domain assembly prediction success. Success was defined as a b5 score under 3 Å whereas predictions with b5 $>$ 5 Å were counted as failures. Borderline cases were not considered. The outline of each violin plot shows a kernel density estimate of the feature distribution across all assembly problems. The box inside each violin extends from the 25 % to the 75 % quantile of each distribution and the median is marked by a white dot. **(a)** Successfully predicted assembly problems have a significantly higher buried surface area in the domain interface ($p = 4 \times 10^{-5}$), concurrent with **(b)** a higher negative $\Delta\Delta G$ value,($p = 3 \times 10^{-5}$) **(c)** a higher number of residues involved in the domain interface ($p = 0.002$) and more specifically **(d)** a higher number of hydrogen bonds in the domain interface ($p = 0.02$), than assembly problems with b5 scores above 5 Å. **(e)** Assembly problems that reach b5 $<$ 3 Å have a lower median 9-mer ($p = 0.02$) and **(f)** 3-mer ($p = 0.03$) fragment RMSD from the experimental peptides at the corresponding linker positions. In particular, a subset of assembly problems with b5 $>$ 5 Å had a median 9-mer RMSD around 2 Å that was rarely found in successfully modeled assembly problems.

bimodal distribution (Figure 3.25e) with a subset of failed assembly problems that had a median fragment RMSD of over 2 Å, likely thwarting any attempt of successful ab initio sampling. It is customary in the evaluation of fragment-based Rosetta protocols to remove homologous proteins from the fragment picking process, to avoid bias from problems that have many homologs in the PDB. This exclusion, however, uses an arbitrary sequence similarity threshold and it is therefore difficult to determine whether poor fragment RMSD in this case stems from improper homolog exclusion (that would be an artifact of the benchmarking method) or from poor fragment sampling, indicating need

for improvement on this rather ancient part of Rosetta.
Additional clues about weaker associations were obtained from inspection of feature distributions: the group of failed problems contained an accumulation of structures with five or more clashes at the interdomain and domain-linker interfaces. Furthermore, the sequence composition of the linker appears to be related to assembly success, as indicated by the PDLI [39] and the Armadillo [42] scores that predict the "linker-likeness" of a sequence. Although it appears intuitive that linkers with a more canonical sequence composition lead to better overall assemblies, the underlying mechanisms are unclear in the light of the finding that assembly success is not related to the linker structure.

### Critical Stages of Structure Prediction

Structural predictors of assembly success have been identified, yet in order to improve prediction accuracy from this knowledge, it is important to learn at what stage of the protocol the course towards success or failure is set. The influence of fragment quality suggests that sampling is limiting at least in some cases. On the other hand, the strong dependence on a large and energetically favorable interdomain interface points at shortcomings of the scoring function to identify more subtle interactions.
The quality of the score function was assessed by running the `DomainAssemblyMover` with simple small move sampling, starting on 127 native structures. If the score function correctly identifies the native structure as the global minimum, the need for sampling should be eliminated. Native structures were recovered to a b5 score below 3 Å in 102 cases (80 %), compared to 71 cases (56 %) reached by the protocol when starting from an extended linker. 35 of the 56 failures (63 %) were successfully solved when starting from the native structure, hence in these cases, failure is likely attributable to insufficient sampling.
This leaves 21 cases in which the protocol yielded b5 scores above 3 Å irrespective of the starting structure. These assembly problems share the common property of an extremely small interdomain surface area (median 210 $Å^2$), a very low number of interface residues (median 16) and consequently only low free energy of domain association (median $\Delta\Delta G$ -2.5 kcal $mol^{-1}$). For comparison, successfully modeled prediction problems had a median surface area of 705 $Å^2$, a median of 43 interacting residues and a median $\Delta\Delta G$ of -8.3 kcal $mol^{-1}$. At the same time, fragment RMSD values were nearly twice as high

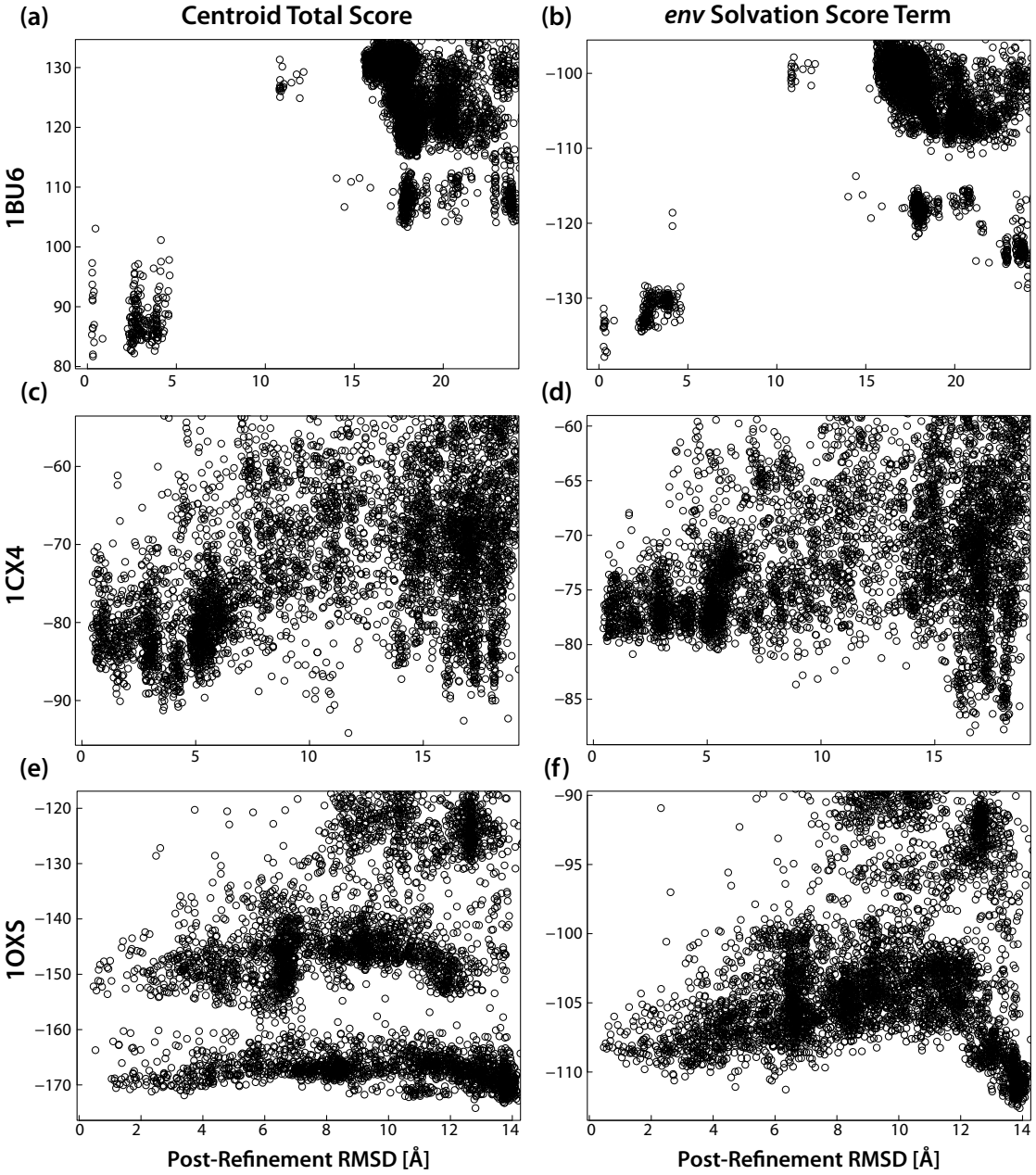



**Figure 3.26.:** Centroid scores during domain assembly. Centroid total scores (panels **(a)**, **(c)** and **(e)**) and the *env* one-body solvation score term (panels **(b)**, **(d)** and **(f)**) are shown to exemplify three different centroid scoring scenarios that all lead to successful refinement (b5 $< 3$ Å). For the assembly problem 1BU6 (b5 $= 0.273$ Å), both the total centroid score **(a)** and the *env* term **(b)** clearly identify decoys that refine to low RMSD. For 1CX4 (b5 $= 0.711$ Å), neither the total centroid score **(c)** nor the *env* score term **(d)** show any clustering with respect to RMSD. In fact, low *env* scores occur only in decoys that refine to more than 15 Å RMSD. For 1OXS (b5 $= 0.558$ Å), both the total centroid score **(e)** and the *env* term **(f)** show clustering and a shallow funnel that converges to a score minimum at 14 Å RMSD.

as for successfully solved problems (median 9-mer RMSD of 1.67 *versus* 0.88 Å and median 3-mer RMSD of 0.027 *versus* 0.019 Å). Taken together, approximately two-thirds of the failed assemblies are attributable to insufficient sampling whereas the remaining one-third fails due to a combination of poor scoring and insufficient sampling. The latter group of failures, however, can be foreseen if it is known that the domains interact only weakly.

Large-scale sampling is done at the centroid stage of the protocol, hence structures and scores at this stage of the protocol were investigated in more detail. Overall, there is no consistent correlation of total score at the centroid stage with the final score. Correlations may be as low as zero and as high as 0.81 (1TUA) but they have no predictive value for the overall performance of the protocol. Neither the total centroid score nor any of the individual score terms (reviewed in detail elsewhere [166]) consistently predict low RMSD such that re-weighting of centroid score terms would not lead to a generally improved solution (Figure 3.26). The presence of clusters and funnels in the centroid score landscape (Figure 3.26a, b, e and f) suggests that score barriers exist between different conformations and that near-native conformations are sampled during the centroid stage.

Given this picture, the centroid score function can in some cases identify what is later in the protocol going to be a low RMSD decoy, however this is not required for good performance of the protocol. Hence the major task of the centroid stage is to sample low RMSD conformations and to pass these to the refinement stage. Therefore, robust improvement of the protocol can only be achieved by modification of the sampling strategy (i.e., the smoothness of the score function and the scale of the perturbations). An ideal centroid stage protocol finds the optimal balance between a large diversity of conformations and robust convergence towards near-native conformations. The next section will introduce objective measures to assess different centroid stage sampling strategies in terms of their efficiency.

### Protocol Efficiency

Efficient sampling is a major problem of computational structure prediction. It has been shown above that two-thirds of the domain assembly prediction failures are due to poor sampling during the centroid stage. The `DomainAssemblyMover` offers three centroid

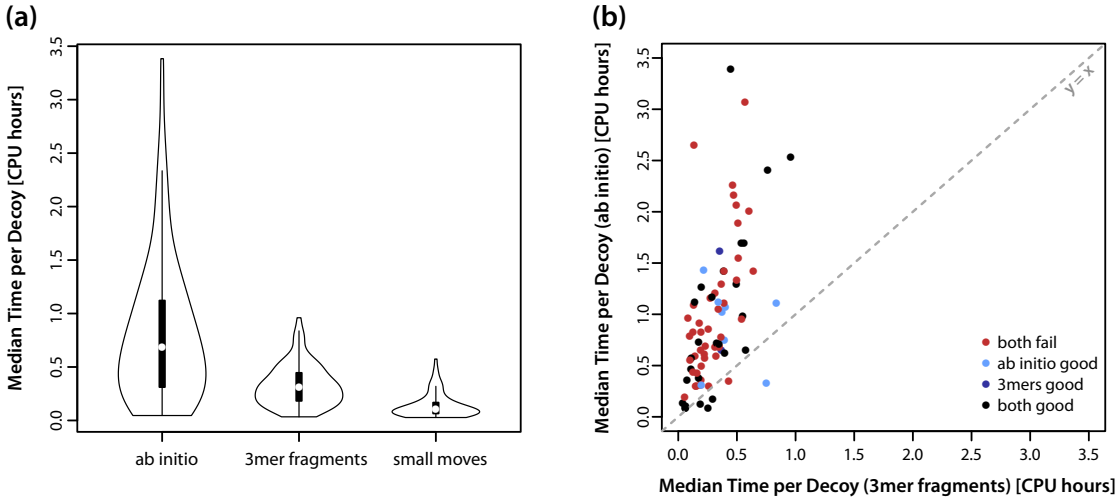



**Figure 3.27.:** Runtime of different domain assembly sampling strategies. **(a)** Distribution of median CPU time to generate one decoy using ab initio sampling, 3mer insertion and small moves in the centroid stage of the domain assembly protocol. The outline of each violin plot shows a kernel density estimate of the runtime distribution. The box inside the each plot extends from the 25 % to the 75 % quantile of each distribution and the median is marked by a white dot. **(b)** Comparison of 3mer insertion runtime with ab initio runtime, color-coded by prediction performance. Each data point represents one domain assembly prediction problem that has been run with both methods. Problems that were not solved (b5 $\geq$ 3 Å) by either method are indicated in red. Problems that were solved by ab initio sampling but not by 3mer insertion are indicated in light blue. Problems solved by 3mer insertion but not by ab initio sampling are indicated in dark blue. Problems solved by both sampling methods are indicated in black. There is no correlation between runtime of either protocol and prediction success. The median runtime of ab initio sampling is approximately three times the median runtime of 3mer insertion. The displayed data in panel **(b)** are a non-representative subset of all tested domain assembly problems.

mode sampling strategies that come with different computational cost: small random perturbations (discontinued at a later stage), insertion of 3-mer fragments sampled from the PDB and the classic ab initio strategy that uses 9-mers and 3-mers in an extensive simulated annealing protocol [166]. As shown above, the ab initio protocol slightly outperforms the 3-mer insertion protocol in terms of accuracy when the Monte Carlo algorithm is stopped after 5000 decoys (Figure 3.23b). To judge whether application of the more expensive protocol justifies this minor gain in accuracy, the runtime performance of the respective protocols was investigated and balanced against their accuracy. Median runtimes per decoy of all domain assembly protocols range between two minutes (1W9Q) and more than three hours (1JPM) with a long-tailed distribution (Figure 3.27a). The median runtime for one decoy using ab initio sampling was 40 min, 3mer insertion

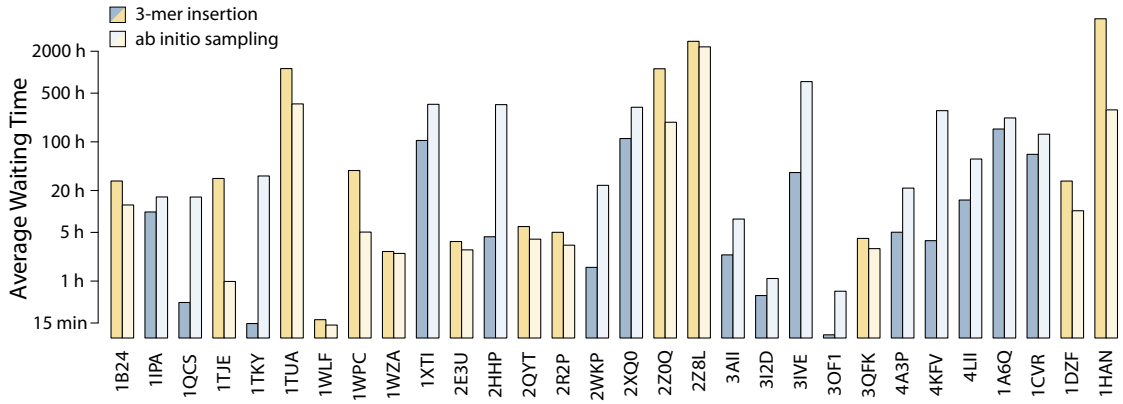



**Figure 3.28.:** Estimated waiting times to produce at least one decoy with RMSD $< 3$ Å with 95 % probability, using 3-mer insertion (darker shades of blue and yellow) and ab initio sampling (lighter shades). The color indicates the strategy with the shorter average waiting time: blue colors indicate shorter waiting times using 3-mer insertion and yellow colors indicate shorter waiting times using ab initio sampling. The median ratio of waiting times (ab initio over 3-mers) is 1.5, hence ab initio sampling takes on average 1.5-fold more time to produce a decoy with RMSD $< 3$ Å.

had a median runtime of 18 min and small move sampling had a median runtime of 6 min for one decoy. When applied to the same problem, ab initio sampling took approximately three times longer than 3-mer insertion (Figure 3.27b).

A statistical measure was devised to determine whether the more thorough, yet more time-consuming sampling of the ab initio strategy outperforms the less accurate 3-mer insertion that can produce roughly three times as many decoys in the same time: the *average waiting time* is an estimate of how much CPU time is needed until a decoy under 3 Å RMSD is produced with at least 95 % probability, by one strategy on one assembly problem. Figure 3.28 shows the average waiting times for 30 assembly problems that were modeled with both 3-mer insertion and ab initio sampling using comparable parameters. Waiting times range from 10 minutes (3OF1, 3-mer insertion) to 5855 hours (1HAN, 3-mer insertion). This large difference does not arise from different runtimes of the protocol (3OF1 has a median runtime of 10 minutes, 1HAN of 24 minutes) but rather from a different frequency of decoys with RMSD < 3 Å: over 98 % of the decoys for 3OF1 have an RMSD under 3 Å, hence one decoy suffices to reach 3 Å RMSD with a probability of 95 %. On the contrary, only one out of 5000 decoys for 1HAN reaches RMSD < 3 Å, hence nearly 15,000 decoys need to be produced to reach 3 Å RMSD with

a probability of 95 %.

In addition to different waiting times for different assembly problems, waiting times also differ between the 3-mer insertion and ab initio strategies when applied to the same problem (e.g., 1TKY where ab initio sampling takes 130-fold more time than 3-mer insertion to reliably produce low RMSD decoys). Nevertheless, there is no generally superior strategy as the median ratio of average waiting times for both methods is near unity (ab initio sampling takes approximately 1.5-fold more time to produce one decoy with RMSD < 3 Å). The number of assembly problems that reach a solution faster with 3-mer insertion (n = 16) is not significantly different from the number of problems that are faster solved with ab initio sampling (n = 14). The small sample size of 30 assembly problems makes statistical associations difficult to identify, however it appears that smaller and more compact proteins are better modeled by ab initio sampling (assembly problems that were faster solved by ab initio sampling were found to have smaller radii of gyration, p = 0.02).

Taken together, 3-mer insertion is approximately three times faster than ab initio sampling, whereas ab initio sampling produces a larger fraction of low-RMSD decoys, resulting in almost equal overall efficiency. Efficiency calculations did not take into account that the ab initio strategy reached b5 scores under 3 Å for a larger number of assembly problems than 3-mer insertion,however statistically significant differences in accuracy could not be established and therefore cannot serve as basis for a general recommendation.

# Discussion

The modular nature of protein domains suggests that interdomain communication uses a universal mechanical language that constitutes the basis for evolution of novel biological function through domain rearrangement. To investigate this language, two blue light regulated adenylyl cyclases were used as model systems with a regulatory input that can be precisely administered, and a functional output that can be readily quantified. As knowledge of the structure is essential for understanding multidomain protein signaling, but at the same time very challenging to obtain experimentally, the potential of a computational approach for the prediction of multidomain protein structure and function was evaluated.

The experimentally characterized model systems bPAC and mPAC appear very similar at the first glance: a blue photon comes in and cyclic AMP comes out. Only on a closer look one appreciates the different photochemistries, the different domain linker architectures and subtle differences in the adenylyl cyclase domains. This detail view in turn is required to discern truly common principles from those specific to the respective photoreceptor. Always keeping the entire sensor–effector system in mind, the structural mechanism of bPAC and mPAC photoactivation can be broken down into three central questions:

*i* how do changes around the flavin chromophore propagate to the photoreceptor domain surface,

*ii* how is this structural signal received and propagated to the AC domain, and

*iii* how do light-induced conformational changes in the AC domain affect its catalytic activity?

Parallels between mPAC and bPAC, implications for other regulator–effector pairs, as well as the prospects for computational design of allosteric multidomain proteins, are discussed.

## 4.1. Mechanisms of Photoreceptor Activation and Signaling

The largest difference in the photoactivation mechanisms of bPAC and mPAC lies at the origin of the signal. In addition to the different photochemistries and lit state intermediates, the vastly different tertiary structures of the bPAC and the mPAC coiled coils mandate different mechanisms of signal transduction from the photoreceptor domain to the coiled coil: bPAC and other BLUF-regulated ACs are unique in using the $\alpha3_{BLUF}$ capping helix to form an intermolecular coiled coil that is sandwiched between the BLUF core β-sheets. BLUF signaling must therefore affect the interaction *between* the BLUF β-sheet and the coiled coil. The capping helices of the mPAC LOV domain also form a coiled coil, however they extend away from the LOV domain core and share no interface with the photoreceptor domain. Light-induced conformational changes of the LOV domain must therefore propagate *along* the coiled coil rather than via an interface as in bPAC. For this reason, the discussion of the initial signal is treated separately for bPAC and mPAC, with its respective implications for signal generation in BLUF and LOV (PAS) domains.

### 4.1.1. A New Twist in BLUF Photoactivation

bPAC as model system with structural, spectroscopic and kinetic readout provided unprecedented insights that allow proposing a comprehensive and self-consistent model of BLUF signaling. By combining new information from bPAC and previous results from other BLUF proteins, two conserved pathways can be described that couple residues in the chromophore binding pocket to the coiled coil: one pathway acts on the N-terminal part of the $\alpha3_{BLUF}$ coiled coil through the conserved methionine (Met93 in bPAC) whereas the second, less explored pathway affects the C-terminal part of $\alpha3_{BLUF}$ through a conserved kink in $\beta4_{BLUF}$.

Met93 resides in a conserved, strained position that is determined by interaction with Gln93 [12, 139] and tight packing against aliphatic residues inside the BLUF core. Upon formation of the lit state, the tautomeric and rotated Gln49 displaces the side chain of Met93 by up to 1 Å as observed in structures of illuminated AppA crystals [85]. In the manner of a light-triggered spring, Met93 relaxes and triggers a chain of conforma-

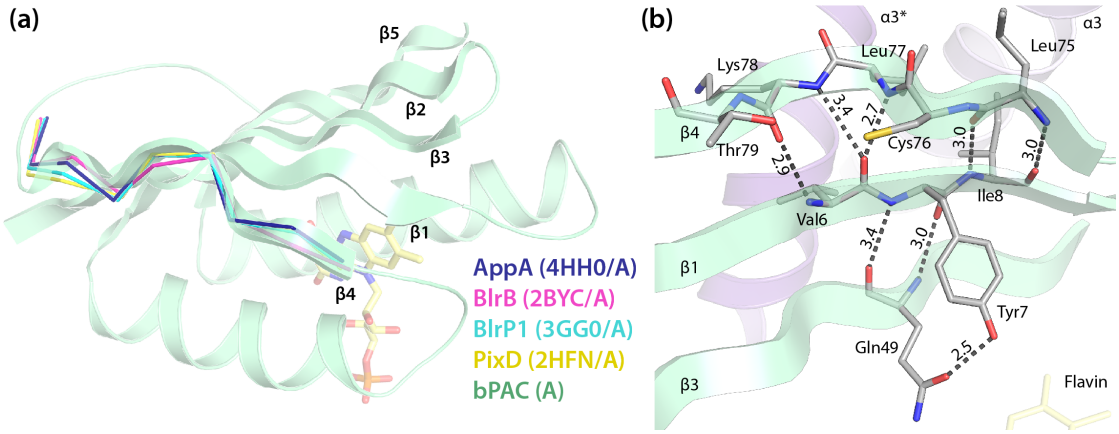



**Figure 4.1.:** Kink in $\beta4_{BLUF}$. **(a)** The kink in $\beta4_{BLUF}$ is conserved throughout BLUF domains. The structure of the bPAC BLUF domain is shown as cartoon and a Cα-trace is shown for AppA (dark blue [85]), BlrB (pink [84]), BlrP1 (cyan [7]) and PixD (yellow [230]). **(b)** Signaling pathway linking photoactive BLUF residues to the $\alpha3_{BLUF}$ helices. Ile8 opposite Tyr7 interacts directly with $\alpha3_{BLUF}$. Changes in the backbone interactions between $\beta1_{BLUF}$ and $\beta4_{BLUF}$ propagate to $\alpha3_{BLUF}$ through the conserved $\beta4_{BLUF}$ kink.

tional changes of conserved residues in $\beta2_{BLUF}$ and $\beta3_{BLUF}$ that relay its impact to the N-terminal part of the $\alpha3_{BLUF}$ helix. The complex chain of interactions from Met93 to the domain surface is conserved in BLUF signaling and can be traced by combining structural and functional data from several BLUF proteins (Appendix A.2).

A second, parallel signaling pathway is related to the conformation of the photoactive tyrosine (Tyr7 in bPAC). Substitution of the tyrosine by phenylalanine has been associated with a functionally pseudolit conformation in several BLUF proteins including bPAC [231, 133, 192], however the origin of the pseudolit conformation has remained unclear. Interestingly, HDX-MS measurements on $bPAC_{Y7F}$ revealed that this mutation does not act through the canonical Met93 pathway and the N-terminus of the coiled coil, but instead directly affects the C-terminal part of $\alpha3_{BLUF}$ through interaction with residues that form a conserved kink in $\beta4_{BLUF}$ (Figure 4.1a). This kink in turn directly interacts with the backbone of the residues neighboring Tyr7/Phe7 and thereby relays conformational changes of this residue (Figure 4.1b). Several lines of evidence suggest that this pathway is not only relevant in the Y7F variant but it is also employed in wild type bPAC and other BLUF proteins: Illumination of the BLUF domain distorts the β-sheet geometry in crystals of bPAC and AppA [85], and NMR experiments showed

light-induced chemical shift changes in $\beta4_{BLUF}$ residues of BlrB [221] and BlrP1 [220] that confirm its independent involvement in BLUF signaling.
The interface between the BLUF β-sheet and the central part of the coiled coil shows packing defects whereas the N- and C-terminal parts are tightly attached, hence the BLUF core domains clamp the coiled coil in a fashion that resembles a pair of brackets. The illustrated signaling pathways affect precisely this clamp in a parallel and independent fashion. Conformational changes or amino acid substitutions loosening one (Y7F, M93A [192]) or both (illumination) arms of the bracket can propagate throughout the coiled coil via transient conformational fluctuations near the sites of imperfect packing [77]. Such fluctuations are in line with the strict EX2 kinetics observed in HDX-MS measurements of light-activated bPAC. Imperfect packing between the BLUF core and parts the capping helix also occurs in BlrP1 [7] and BlrB [84], and studies on PixD have shown that high pressure decreases the quantum yield of light-activated photodissociation through suppression of conformational fluctuations [104].

### 4.1.2. LOV Photoactivation in mPAC

Studies on diverse LOV-regulated proteins suggest that there exist distinct architectural families, each of which has a conserved mechanism of propagating light-induced conformational changes to the effector [74, 52]. This work presents compelling experimental and theoretical evidence that the structure of the mPAC LOV domain and its mechanism of photoactivation resemble those described for YF1 or PpsB1 [36, 26, 167]. The origin of the structural signal in mPAC is the protonation of the flavin N5 atom and the formation of a flavin C4a–cysteinyl adduct. Ensuing structural changes of the conserved glutamine (Gln257 in mPAC) are relayed via Ile159 to the A′α helices that can pack between the LOV β-sheets in at least two stable conformations. The conformation preferred in the lit state involves a register shift of the A′α helices with respect to each other and the β-sheet of the opposing monomer. This model based on PpsB1 [26, 167] and YF1 [36, 11] is consistent with HDX-MS data of mPAC that show slightly changed secondary structure dynamics of A′α upon illumination (Figure A.9). In addition to illustrating the pathway of signal transmission from the chromophore to the A′α helix, the high-resolution homology model rationalizes the experimentally determined LOV lit state lifetimes that largely depend on the rigidity of the flavin binding pocket and the

stability of the flavin N5–Gln257 hydrogen bond.
Resulting from the A'α shift, the lit-state structure has a larger distance between the N-terminal ends of the Jα-helices that extend away from the LOV domain core and form a coiled coil. The part of the coiled coil that is most strongly destabilized by illumination is at the distal end of Jα, hence for the structural signal to propagate along the coiled coil, some rigidity is required for mechanical transmission. Indeed, no light-induced signal reaches the AC domain in the absence of nucleotide because it likely dissipates in in the unstructured linker (Figure A.8).

### 4.1.3. Role and Origin of BLUF and LOV Capping Helices

Helical elements play an essential role in BLUF and LOV signaling. Even short BLUF or LOV proteins that regulate their effector in *trans* [215, 84] or act through oligomerization [230, 236, 134, 71] are only functional in the context of their capping helices. In fact, none of the BLUF or LOV proteins characterized so far interact with their effectors through the core photoreceptor domain. Despite their universal importance, sequences and structures of capping helices are extremely diverse, with each family of BLUF or LOV proteins having a different size, number and orientation of capping helices with respect to the β-sheet.
bPAC, its homolog OaPAC [149] and likely all other BLUF-regulated adenylyl cyclases use their $\alpha 3_{\mathrm{BLUF}}$ capping helix to form an intermolecular coiled coil that ends in a handle-like helical bundle. Interestingly, nearly identical handles have also been observed in crystal structures of more distantly related adenylyl cyclases, including a bicarbonate-regulated human soluble AC [99, 172], soluble guanylyl cyclase [123] and the soluble fragment of a mycobacterial transmembrane AC with high similarity to mPAC [209]. Sequence analysis predicts conservation of this handle structure in all type IIIa and IIIb cyclases [234], including such that are regulated by PAS, HAMP and phytochrome PHY domains [90, 88, 89, 217, 170]. It is therefore suggestive to think of capping helices in general as conserved parts of the respective effectors that evolved to interface with the photoreceptor domain, rather than variable parts of the photoreceptor domain.
While this notion is certainly true for LOV domains (e.g., EL346 in which the LOV domain packs against a helix that database search identifies as conserved histidine kinase sequence), BLUF capping helices are more specialized for interaction with the BLUF

core domain and similar sequences are only found in homologous BLUF proteins. Regardless of the degree of adaptation to LOV or BLUF signaling, this implies that a discussion of the structure and signaling mechanism through helical elements is only meaningful in the context of the specific effector.

## 4.2. Mechanism of Adenylyl Cyclase Regulation

The combination of crystal structures, measurements of secondary structure dynamics and site-directed mutagenesis allows to trace and validate a pathway of molecular interactions from the photoreceptor domain to the AC domain. In bPAC, photoactivation culminates in small scale conformational fluctuations at the C-terminus of the $\alpha 3_{BLUF}$ coiled coil. An accurate high-resolution homology model of mPAC shows that despite the different signaling pathway, illumination eventually perturbs a hydrogen bond network between C-terminus of the Jα coiled coil and the handles, resulting in increased secondary structure dynamics in this region, similar to bPAC. Crystal structures that accompanied measurements of secondary structure dynamics showed large conformational changes in the photoreceptor and AC domains of illuminated or pseudolit bPAC. Strikingly, no conformational change could be made out in the coiled coil although it showed the largest light-induced perturbation of secondary structure dynamics in bPAC and mPAC.

To resolve this dichotomy, the structure and possible functional motions of the coiled coil, as well as their effect on AC domain structure and activity are discussed. On a more general scale, the conserved regulatory interactions in photoactivatable ACs allow proposing an AC regulation mechanism that unifies all hitherto observed length scales of AC activation and explains why a fully activated nucleotide-bound crystal structure will be difficult — if not impossible — to obtain.

### 4.2.1. Architecture and Mechanism of Coiled Coil Signaling

This work shows for the first time that a coiled coil followed by two handles is a conserved structural motif that integrates signals from different input domains (LOV and BLUF in this case) and translates them into AC activation. This finding is in line with sequence analysis showing that type IIIa and IIIb ACs are preceded by a conserved sequence motif

termed cyclase transducer element (CTE) [234], underlining the general importance of the handles for AC regulation. The architecture and the light-induced molecular events within the coiled coil and the handles therefore constitute the basis of modularity in BLUF, LOV and AC domains.

The coiled coils of bPAC and mPAC differ in sequence but both deviate from the canonical heptad repeat pattern to form intricate side chain hydrogen bond and salt bridge networks. While the function of the polar network in the center of the coiled coil is not clear, it is strictly conserved in type IIIa AC [234] and in BLUF-regulated type IIIb ACs [149]. It is tempting to speculate that light-induced breakage of polar interactions provides a more binary ON/OFF structural response than reorganization of a hydrophobic interface as the free energy of association is localized to a small number of residues [60, 147]. This hypothesis is in line with the observation that substitution of one single amino acid in bPAC (R121S) destabilizes the hydrogen bond network to an extent that is sufficient for constitutive AC activation.

The coiled coil of bPAC is sandwiched between the BLUF β-sheets and thereby additionally stabilized. In mPAC, it projects away from the LOV domain and only the very C-terminal part is buried in a hydrophobic interface between the handles that are arranged perpendicular to the coiled coil (forming a helical bundle). While this architectural difference determines the pathway of signal propagation, it is unlikely to cause the observed differences in secondary structure stability of the coiled coil. HDX-MS measurements of mPAC show highly correlated structural dynamics of the coiled coil, the handles and the $\beta4_{AC}$–$\beta5_{AC}$ tongues: In the absence of nucleotide Phe419 (the lid) is only transiently attached to the coiled coil and the handles. Whenever the lid is detached from the core, water can penetrate into the helical bundle, leading to massive destabilization of the hydrophobic interface. Nucleotide binding stabilizes the AC dimer interface and $\beta4_{AC}$–$\beta5_{AC}$ in particular. With the active site occupied by nucleotide, the lid has less conformational freedom and is more tightly attached to the helical bundle, resulting in less frequent sampling of the unprotected conformation. This effect is not observed in bPAC because its helical bundle does not have a dynamic hydrophobic lid and the $\beta4_{AC}$–$\beta5_{AC}$ tongue shows low secondary structure dynamics regardless of nucleotide binding.

It was shown that an unprotected conformation with high secondary structure dynamics of the helical bundle underneath the $\beta4_{AC}$–$\beta5_{AC}$ tongue correlates with activation of the

AC domain in both investigated systems. Discussion of coiled coil motions allows to propose a model how conformational fluctuations that are potentially invisible to X-ray crystallography can lead to AC activation by the photoreceptor.

### Structural Models of Coiled Coil Motions

Coiled coils are versatile signal transducers that allow a wide range of functional motions in response to small structural perturbations at remote sites [122]. A very simple and appealing model for cyclase activation has been proposed for the pH-regulated mycobacterial AC Rv1264. Crystals grown at different pH values showed either a rigid α-helix keeping the AC domains apart in the inactive state or a flexible loop that allows self-organized formation of an AC active site dimer interface [205]. A similar transition has been observed in the HAMP domain regulated phosphodiesterase 2a [154] and generalized into a model of *regulated unfolding* [181] (Figure 4.2a). The model of regulated unfolding, especially in the context of LOV domain regulated proteins, is reminiscent of light-induced domain uncaging [222] where the photoreceptor sterically occupies a functional interface. As none of these proteins contain coiled coils or handle-like structures, these models are likely not valid for activation of type IIIa and IIIb cyclases: the structures of bPAC and mPAC rule out a caged interface between the photoreceptor capping helix and the AC, as well as a large-scale disruption of the AC domain interface by the coiled coil. Moreover, this work has ruled out an inhibitory role of the photoreceptor domains both for mPAC and bPAC, which is in contrast to the type IIIc AC Rv1264 that is fully active in the absence of its regulatory domain [113].
The opposite transition, namely the reorganization of the helical bundle into a continuous coiled coil has been proposed by Zielger *et al.* based on sequence analysis, arguing that the handle sequence constitutes a continuation of the coiled coil heptad repeats [234] (Figure 4.2b). From a structural perspective, it is highly questionable if a continuous helix could be formed over proline residues, including the universally conserved and essential helix breaker $Pro286_{mPAC}$. Furthermore, an extension of the helix by at least 10 residues (13 in mPAC, producing 2.8–3.6 turns) would extend the rod-shaped molecules by at least 15 Å in their longest axis. Such a dramatic structural change is not compatible with the SAXS measurements of illuminated $mPAC_{\Delta 132}$.

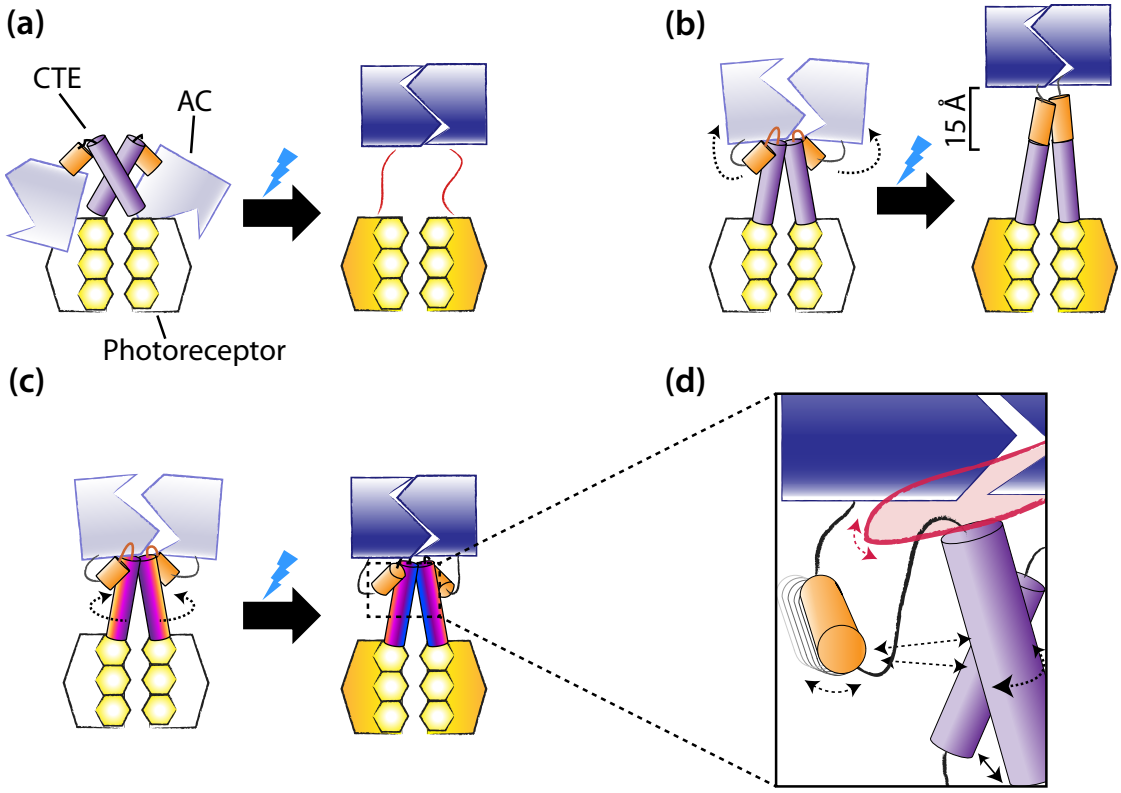



**Figure 4.2.:** Mechanisms of signal transduction through coiled coils. **(a)** The mechanisms of *regulated unfolding* [181] or *domain uncaging* [222] assume that the coiled coil (purple) and the handles (orange) form a rigid structure that prevents formation of a catalytically active AC dimer in the dark. Upon activation, the linker unfolds and allows the intrinsically active AC domain to assemble in an active conformation. **(b)** The mechanism of *helix straightening* was proposed based on the observation that the coiled coil heptad pattern continues into the handle helix. Upon activation, the entire handle assembles into one continuous α-helix [234]. This conformational change increases the length of the coiled coil by at least 15 Å. **(c)** The mechanism of *helix rotation* is based on the observation of register switching in HAMP domains [78] and on the prediction that N-terminal splaying of the coiled coil leads to downstream supercoiling of the helices [46, 10]. The coiled coils of bPAC and mPAC can likely assume two registers. Helix rotation can lead to alternate packing of the coiled coil, resulting in a different relative orientation of the handles and the AC domains. **(d)** Rather than switching to a defined active conformation of the coiled coil, destabilization of the interactions between the polar residues in the coiled coil and the handles (or the BLUF domain, in the case of bPAC) can also lead to increased dynamics, e.g., through rotational sampling of the helices in the coiled coil. The tight coupling between the coiled coil, the handles and the $\beta4_{AC}$–$\beta5_{AC}$ tongue (red) predicts that helix dynamics directly translate into the AC active site.

HDX-MS indicates locally decreased stability of hydrogen bonds around the network of polar residues in the middle of the coiled coil of mPAC and bPAC, however in both cases deuteration differences between the dark and the lit state only arise after long

incubation times, indicating that the lit state structure is only slightly different from the dark state; it merely spends more time in a transient state of unprotected backbone hydrogen bonds. Such small-scale or transient structural changes in combination with increased secondary structure dynamics have been described in conjunction with helix rotation and register shifts in HAMP domains [233, 195, 108]. The simplest of such models has been described as the *gearbox model* [78]. This model predicts that the four-helix coiled coil can alternate between two stable forms of packing by rotation of the helices in the coiled coil (Figure 4.2c). There is evidence that the helical register also plays a role in signaling through two-helical coiled coils such as the S-helix [5], the DHp domain of histidine kinases [48, 150] or the linker of diguanylate cyclases [56]. In the latter case it could be shown that the coiled coil has two non-ideal registers and that mutations stabilizing either register could lock the enzyme in the ON or in the OFF state. While the coiled coil of bPAC and mPAC could accommodate a rotated conformation, it is not clear whether splaying of the helices at the N-terminus actually leads to directed rotation as predicted for YF1 [11, 10], nor is it clear how a rotated conformation shifts the AC active site residues toward an active conformation. Moreover, crosslinking of residues within the coiled coil of $bPAC_{L123C}$ to fix the helical register had no effect on photoactivation, hence register shifts within the coiled coil are likely not relevant for AC activation.

Rather than promoting a shift from one register of the coiled coil to the other, activation may increase the frequency of sampling between the different registers or other transient conformations. Release of the coiled coil from the handles (in mPAC) or the BLUF domain (in bPAC) removes brackets that constrain conformational sampling of the helices (Figure 4.2d). Increased sampling of the coiled coil and the handles (both are tightly coupled to the $\beta4_{AC}$–$\beta5_{AC}$ tongue) may promote more frequent sampling of the active conformation, rather than stabilizing it. This is in line with the observation that allosteric signaling need not proceed through directed conformational change but may well occur through changed dynamics of conformational sampling [61]. In fact, modulation of coiled coil dynamics through a sensor domain has been shown for HAMP domains [233] and for the PAS-regulated histidine kinase BvgS from *Bordatella pertussis* [109].

### Coiled Coil Motions are Relayed through the $\beta4_{AC}$–$\beta5_{AC}$ Tongue

Although it is not entirely clear what the coiled coil motions look like and whether they involve a change of the equilibrium conformation, structures obtained from illuminated bPAC crystals illustrate the ensuing conformational changes in the AC domain. Irrespective of illumination, crystal structures of bPAC, OaPAC [149, 148] and the various nucleotide- and activation states of human soluble AC [99, 100] show a clamshell-like movement of the AC active site cleft that correlates with the position of the $\beta4_{AC}$–$\beta5_{AC}$ tongues (UP or DOWN, Figure 4.3a). Despite this variation that illustrates thermally accessible states of the AC domain, illumination of bPAC crystals caused a distinct and significant change of the tongue position toward the DOWN conformation when compared to structures determined from dark state crystals grown in the same drop. Therefore, increased dynamics or conformational changes of the coiled coil lower the barrier for conformational transition or shift the equilibrium between closed and open AC conformations.

In the dark-state crystal structures of bPAC, including those showing an open $\text{tongue}_{DOWN}$ conformation, the substrate-determining residues Lys197 ($\beta2_{AC}$) and Thr267 ($\beta5_{AC}$) are in an inactive conformation, with Thr267 turned away from the nucleotide-binding pocket and Lys197 arrested through interactions with the carbonyl oxygen atoms of Phe198* and His266. Comparison of structures from dark-state and illuminated wild type crystals, and $bPAC_{Y7F}$ shows that the interaction between the coiled coil and the tongue changes such that Thr267 moves further away from the opposing subunit which is required for interaction with the adenine base. At the same time, Lys197 is freed from its resting position near Phe198* in the dark state, allowing the phosphate-binding $\beta2_{AC}$–$\beta3_{AC}$ hairpin to move away from the neighboring subunit and providing more space for the adenine base to enter the active site more deeply (Figure 4.3b).

In conclusion, activation through the coiled coil, the handles and the attached $\beta4_{AC}$–$\beta5_{AC}$ tongue promotes sampling of conformations that are present but rarely sampled in the ground state. This manifests itself in opening of the AC domains and importantly, owing to the tongue's central location underneath the ATP binding pocket, it additionally removes locks on active site residues that are arrested in an unproductive conformation even in the open AC form. Both forms of structural change, adjustment of the AC opening angle ([203, 189, 148]) and release of catalytic residues from locked positions

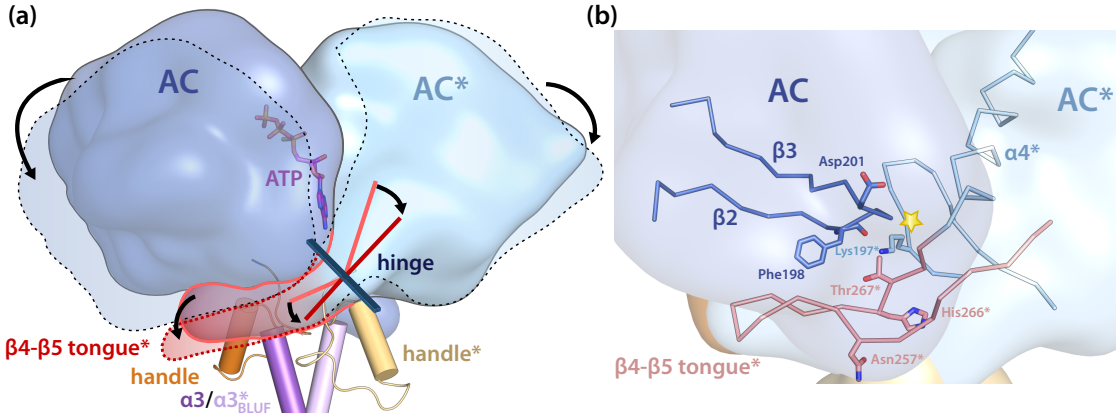



**Figure 4.3.:** Regulation of adenylyl cyclase by the $\beta4_{AC}$–$\beta5_{AC}$ tongue. **(a)** The clamshell-like movement of the AC domain can be described by rotation around a hinge across the plane of $\beta1_{AC}$, $\beta4_{AC}$ and $\beta5_{AC}$. The $\text{tongue}_{UP}$ conformation is indicated in light red and corresponds to a closed AC angle while the $\text{tongue}_{DOWN}$ conformation (dark red, dashed lines) moves the AC domains apart and opens the active site cleft. The tongue conformation is adjusted by interaction with $\alpha3_{BLUF}$ (purple) and the handles (orange) in bPAC. The location of the substrate ATP has been modeled according to Tesmer *et al.* [203]. **(b)** Interactions in the closed AC active site. The phosphate-binding $\beta2_{AC}$–$\beta3_{AC}$ loop and the base-specifying residues Lys197* ($\beta2^*_{AC}$) and Thr267* ($\beta5^*_{AC}$) are too close to each other to accommodate ATP in the active site (marked by a star) in the locked conformation. Interaction of the $\alpha4_{AC}$–$\beta5_{AC}$ tongue (e.g. through Asn257* and His266*) with the coiled coil and the handles moves the $\beta4^*_{AC}$–$\beta5^*_{AC}$ tongue towards the DOWN position, moving both the $\beta4^*_{AC}$–$\beta5^*_{AC}$ tongue and the AC* domain core away from the other chain's $\beta2_{AC}$–$\beta3_{AC}$ loop in a clamshell-like motion (cf. panel **(a)**). This opens the active site cleft and removes locks from active site residues, allowing them to assume a catalytically competent conformation.

[189, 99, 100] have been discussed as mechanisms of AC regulation. In contrast to the previously discussed AC activation mechanisms that assume a conformational switch between distinct conformations, this work suggests that activation only changes the frequency at which the catalytically active conformation is sampled.

### 4.2.2. Active Conformation of Adenylyl Cyclase

A catalytically active conformation is defined by two features: (*i*) an AC opening angle that allows not only binding of the phosphate moiety but also of the nucleoside, and (*ii*) presence of the expected interactions between conserved active site residues and the nucleoside [203]. According to these criteria, a fully activated conformation of bPAC was not observed. The same applies to other AC structures such as the short isoform of human soluble AC (solAC) [99, 172]. Independent of the AC opening angle and even in

the presence of the allosteric activator bicarbonate, the base-specifying lysine (Lys334) points away from the ATP analog that is bound with low occupancy. As in bPAC, the lysine is locked in an inactive conformation by interaction with the backbone carbonyl of a tongue residue (Tyr404). Similarly, the base-specifying Thr405 is too far for a base-discriminating interaction despite pointing at the adenine base [99, 172]. Further structural changes of bPAC and solAC are required to assume a catalytically active conformation. However, in both cases, these structural changes are not compatible with the crystal lattice as bPAC crystals dissolved in the presence of ATP and light. Similarly, a significant drop in resolution was observed when soaking nucleotide-bound crystals of solAC in bicarbonate [99]. Thus, the extent of structural differences between inactive and AC structures remains unknown.

bPAC and mPAC have similar $K_{\mathrm{M}}$ values for ATP in the dark and in the light. This surprising finding contradicts the initial expectation that light-induced structural changes prepare the active site for nucleotide binding and thereby increase substrate affinity. Instead, it has been shown that bPAC and other ACs sample a variety of conformations in the dark (ground) state and binding of the nucleotide is dominated by the phosphate–metal interaction. Depending on the opening angle, the entire nucleotide can enter the binding pocket but in thermal equilibrium, the nucleoside-specifying residues remain locked or too far for interaction. Occasionally, the enzyme-substrate complex will sample a near attack conformation [14] of similar or higher energy, that is catalytically competent. This represents the dark state reaction observed as "leakiness" of the photoactivatable ACs. Stimulation by light (or any other activator, indicated by ↯ in scheme 4.1) lowers the barrier for the conformational transition to this near attack conformation or shifts the equilibrium between the inactive and active conformations (scheme 4.1):

$$\mathrm{E} + \mathrm{S} \overset{K_{\mathrm{D}}}{\rightleftharpoons} \mathrm{ES} \overset{\lightning}{\rightleftharpoons} \mathrm{ES}^{*} \xrightarrow{k_{cat}} \mathrm{EP} \longrightarrow \mathrm{E} + \mathrm{P} \tag{4.1}$$

This model is in line with structural and kinetic data presented in this work and elsewhere. It extends the previous explanation for the observation of an inactive substrate complex, stating that base-discriminating interactions form only late during catalysis and are relevant for turnover rather than substrate binding [162, 100]. It is further supported by the similar affinities of solAC for ATP and GTP [100] and the observation that mutants of

the base-specifying threonine retain only residual activity by mainly affecting turnover velocity but not $K_M$ [88, 171].

Importantly, this implies that there exists no distinct activated conformation of AC, ready to catalyze cyclization of an incoming nucleotide, that is not present in the ground state. The mPAC variant K468S was devised as a consequence of this model. The mutation targets the dark state reaction of mPAC, rather than attempting to manipulate nucleotide binding or a supposed conformation along the photoactivation pathway. At saturating ATP concentrations, the turnover number of $mPAC_{KS}$ is increased more than 700-fold by light, a factor that is unprecedented in the field of photoactivatable ACs [80, 171, 193, 149, 170]. Together with fast-recovering variants of the LOV domain, this will empower novel applications for the optogenetic manipulation of cAMP levels.

### 4.2.3. AC Activation through Handles is Highly Conserved

The proposed mechanism of AC activation was corroborated by comparative analysis between bPAC and mPAC, as well as site-directed mutagenesis coupled to enzymological analysis. Sequence analysis has shown a regulatory motif (cyclase transducer element, CTE) that corresponds to the handles introduced in this work [234]. The sequence of the handles is highly conserved in all type IIIa ACs, allowing to extrapolate the structural mechanism of the bPAC and mPAC model systems to medically relevant and difficult to study systems such as mammalian transmembrane ACs (tmAC) of which the handles could not be crystallized.

Not all of the conserved regulatory interactions are present in both domains of asymmetric pseudoheterodimeric tmAC, indicating that activation occurs in an asymmetric fashion to target exclusively the functional active site [234]. For instance, the strictly conserved $Glu278_{mPAC}$–$Lys294_{mPAC}$ interaction is only present in the first cytoplasmic domain of tmAC where it attaches one helix to the handle of the same chain. $Glu273^{*}_{mPAC}$ that is predicted from mPAC to join the interaction network from the other helix is not conserved in the second cytoplasmic domain of mammalian tmAC.

A number of pathologically relevant mutations can be rationalized by inspection of the handles and their interactions. The variants M279I ($Leu281_{mPAC}$) in mouse AC3 [155] and R418W (maps to residue 267 in mPAC that is part of the LOV domain) in human

AC5 [20] likely disrupt coiled coil geometry or packing against the handles, both resulting in hyperactivation of the AC. The disruptive variant P858S ($Pro286_{mPAC}$) was found to eliminate activity of retinal-specific guanylyl cyclase 2D in cases of Leber's congenital amaurosis [35]. Interestingly, concomitant with the asymmetry of conservation patterns, activating variants in the $\beta4_{AC}$–$\beta5_{AC}$ tongue of mammalian AC are enriched in the second cytoplasmic domain that contributes the substrate-determining aspartate and lysine residues and pairs with $\beta1^*_{AC}$ and $\beta2^*_{AC}$ of the first cytoplasmic domain that contribute the metal-binding aspartates. The variants I1010M ($Ile414_{mPAC}$), K1014N ($Lys418_{mPAC}$) and P1015Q ($Phe419_{mPAC}$) have been found to constitutively activate human AC2 in a genetic screen [69]. While I1010M disturbs the packing of $\beta4_{AC}$ to the other subunit and thereby pushes the $\beta4_{AC}$–$\beta5_{AC}$ tongue toward the DOWN conformation, the variants K1014N and P1015Q ($\beta4_{AC}$–$\beta5_{AC}$ loop and $\beta5_{AC}$, respectively) likely change the relative positioning of the $\beta4_{AC}$–$\beta5_{AC}$ tongue to the handles.
With the knowledge of bPAC and mPAC photoactivation, such homology considerations can provide insight into cyclase-activating mutations that map to regions outside the crystallized constructs [203]. Moreover, the asymmetric distribution of activating mutations in asymmetric cyclases corroborates that activation through direct interaction between the handles and the $\beta4_{AC}$–$\beta5_{AC}$ tongue lining the active site (but not the degenerate site) is a universally conserved strategy.

## 4.3. Prediction and Design of Multidomain Protein Assemblies

Description of the photoactivation mechanisms of bPAC and mPAC relied heavily on the availability of high-resolution structural models of the proteins. Crystal structures of LOV, BLUF and AC domains had long been available but only the knowledge of their relative arrangement and of the intricate linker structures connecting the domains provided insight into the allosteric activation mechanism. Domain assembly modeling addresses precisely this problem: the prediction of multidomain protein structure from the combination of homologous or known domain structures. This is tightly linked to the unresolved problem of rational design of synthetic multidomain proteins as a protocol that can reliably model naturally occurring assemblies will also be useful for the design

of synthetic ones. With the increasing number of domain structures becoming available, domain assembly is likely to become an important task for computational protein structure prediction [3, 175].

This work characterized a prototype Rosetta protocol that can assemble two-domain proteins with a linker sequence. The overall prediction accuracy of the `DomainAssemblyMover` compares favorably with other approaches to domain assembly modeling [218, 22, 224], and explicit modeling of the linker lifted the requirement on linker length that had limited the accuracy of constrained docking methods [22, 224]. Nevertheless, the notion that explicit treatment of linkers would lead to accurately modeled linkers could not be confirmed. Assembly was mainly driven by the domain interface, hence modular proteins with little interdomain interface and highly structured and functional linkers such as bPAC or mPAC are not likely to be modeled reliably. One should bear in mind that fragments from homologous proteins were excluded from modeling to reduce knowledge bias. As seen in bPAC or mPAC, functional linkers are highly specialized with conserved sequence and structure. Moreover, as fragments are clearly the premier source for the fold and secondary structure of small peptides [219, 65], the protocol's failure to accurately model structured linkers must be taken with caution when homologous fragments are artificially excluded.

These findings predict that the protocol presented in this work is not suitable for the prediction and design of functional domain linkers because the linker structure was largely irrelevant for the overall assembly. The complex interaction networks in the linkers of bPAC and mPAC and their subtle structural rearrangements that affect the dynamics of the domain coupling are invisible to the Rosetta score function. This eliminates applications for the prediction and design of switchable sensor-effector systems. At this point, fragment-based computational protein structure prediction remains no more than a sensitive tool to predict and design interdomain interfaces of rigid multidomain protein assemblies as it is required for capsid-like proteins or synthetic protein nanomaterials [153, 228, 97].

# References


[1] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta Crystallogr. D Biol. Crystallogr. 66 (Pt 2) (2010) 213–221.

[2] M. T. A. Alexandre, J. C. Arents, R. van Grondelle, K. J. Hellingwerf, J. T. M. Kennis, A base-catalyzed mechanism for dark state recovery in the Avena sativa phototropin-1 LOV2 domain, Biochemistry 46 (11) (2007) 3129–3137.

[3] P. Aloy, A. Stark, C. Hadley, R. B. Russell, Predictions without templates: new folds, secondary structure, and contacts in CASP5, Proteins 53 Suppl 6 (2003) 436–456.

[4] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (17) (1997) 3389–3402.

[5] V. Anantharaman, S. Balaji, L. Aravind, The signaling helix: a common functional theme in diverse signaling proteins, Biology Direct 1 (1) (2006) 25.
URL `http://www.biology-direct.com/content/1/1/25/abstract`

[6] S. Anderson, V. Dragnea, S. Masuda, J. Ybe, K. Moffat, C. Bauer, Structure of a novel photoreceptor, the BLUF domain of AppA from Rhodobacter sphaeroides, Biochemistry 44 (22) (2005) 7998–8005.

[7] T. R. M. Barends, E. Hartmann, J. J. Griese, T. Beitlich, N. V. Kirienko, D. A. Ryjenkov, J. Reinstein, R. L. Shoeman, M. Gomelsky, I. Schlichting, Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase, Nature 459 (7249) (2009) 1015–1018.
URL `http://www.ncbi.nlm.nih.gov/pubmed/19536266`

[8] C. Bauer, C.-R. Rabl, J. Heberle, T. Kottke, Indication for a radical intermediate preceding the signaling state in the LOV domain photocycle, Photochem. Photobiol. 87 (3) (2011) 548–553.

[9] D. Bellmann, A. Richardt, R. Freyberger, N. Nuwal, M. Schwärzel, A. Fiala, K. F. Störtkuhl, Optogenetically Induced Olfactory Stimulation in Drosophila Larvae Reveals the Neuronal Basis of Odor-Aversion behavior, Front Behav Neurosci 4 (2010) 27.

[10] O. Berntsson, R. P. Diensthuber, M. R. Panman, A. Björling, E. Gustavsson, M. Hoernke, A. J. Hughes, L. Henry, S. Niebling, H. Takala, J. A. Ihalainen, G. Newby, S. Kerruth, J. Heberle, M. Liebi, A. Menzel, R. Henning, I. Kosheleva, A. Möglich, S. Westenhoff, Sequential conformational transitions and $\alpha$-helical supercoiling regulate a sensor histidine kinase, Nat Commun 8 (1) (2017) 284.

[11] O. Berntsson, R. P. Diensthuber, M. R. Panman, A. Björling, A. J. Hughes, L. Henry, S. Niebling, G. Newby, M. Liebi, A. Menzel, R. Henning, I. Kosheleva, A. Möglich, S. Westenhoff, Time-Resolved X-Ray Solution Scattering Reveals the Structural Photoactivation of a Light-Oxygen-Voltage Photoreceptor, Structure 25 (6) (2017) 933–938.e3.

[12] H. S. Biswal, E. Gloaguen, Y. Loquais, B. Tardivel, M. Mons, Strength of NH···S Hydrogen Bonds in Methionine Residues Revealed by Gas-Phase IR/UV Spectroscopy, J. Phys. Chem. Lett. 3 (6) (2012) 755–759.
URL http://dx.doi.org/10.1021/jz300207k

[13] E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel, K. Deisseroth, Millisecond-timescale, genetically targeted optical control of neural activity, Nature Neuroscience 8 (9) (2005) 1263–1268.
URL http://www.nature.com/neuro/journal/v8/n9/full/nn1525.html

[14] T. C. Bruice, S. J. Benkovic, Chemical basis for enzyme catalysis, Biochemistry 39 (21) (2000) 6267–6274.

[15] D. Bucher, E. Buchner, Stimulating PACalpha increases miniature excitatory junction potential frequency at the Drosophila neuromuscular junction, J. Neurogenet. 23 (1-2) (2009) 220–224.

[16] M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward, D. C. Prasher, Green fluorescent protein as a marker for gene expression, Science 263 (5148) (1994) 802–805.

[17] R. H. Chan, R. A. Bogomolni, Structural water cluster as a possible proton acceptor in the adduct decay reaction of oat phototropin 1 LOV2 domain, J Phys Chem B 116 (35) (2012) 10609–10616.

[18] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr 66 (1) (2010) 12–21.
URL http://scripts.iucr.org/cgi-bin/paper?dz5180

[19] X. Chen, H. Cao, A. Saraf, L. S. Zweifel, D. R. Storm, Overexpression of the Type 1 Adenylyl Cyclase in the Forebrain Leads to Deficits of Behavioral Inhibition, J. Neurosci. 35 (1) (2015) 339–351.
URL http://www.jneurosci.org/content/35/1/339

[20] Y.-Z. Chen, J. R. Friedman, D.-H. Chen, G. C.-K. Chan, C. S. Bloss, F. M. Hisama, S. E. Topol, A. R. Carson, P. H. Pham, E. S. Bonkowski, E. R. Scott, J. K. Lee, G. Zhang, G. Oliveira, J. Xu, A. A. Scott-Van Zeeland, Q. Chen, S. Levy, E. J. Topol, D. Storm, P. D. Swanson, T. D. Bird, N. J. Schork, W. H. Raskind, A. Torkamani, Gain-of-function ADCY5 mutations in familial dyskinesia with facial myokymia, Ann. Neurol. 75 (4) (2014) 542–549.

[21] Z.-H. Chen, S. Raffelberg, A. Losi, P. Schaap, W. Gärtner, A cyanobacterial light activated adenylyl cyclase partially restores development of a Dictyostelium discoideum, adenylyl cyclase a null mutant, J. Biotechnol. 191 (2014) 246–249.

[22] T. M. K. Cheng, T. L. Blundell, J. Fernandez-Recio, Structural assembly of two-domain proteins by rigid-body docking, BMC Bioinformatics 9 (2008) 441.

[23] S. Chowdhury, M. C. Lee, G. Xiong, Y. Duan, Ab initio folding simulation of the Trp-cage mini-protein approaches NMR resolution, J. Mol. Biol. 327 (3) (2003) 711–717.

[24] J. M. Christie, A. S. Arvai, K. J. Baxter, M. Heilmann, A. J. Pratt, A. O'Hara, S. M. Kelly, M. Hothorn, B. O. Smith, K. Hitomi, G. I. Jenkins, E. D. Getzoff, Plant UVR8 Photoreceptor Senses UV-B by Tryptophan-Mediated Disruption of Cross-Dimer Salt Bridges, Science 335 (6075) (2012)

1492–1496.
URL http://science.sciencemag.org/content/335/6075/1492

[25] J. M. Christie, S. B. Corchnoy, T. E. Swartz, M. Hokenson, I.-S. Han, W. R. Briggs, R. A. Bogomolni, Steric interactions stabilize the signaling state of the LOV2 domain of phototropin 1, Biochemistry 46 (32) (2007) 9310–9319.

[26] F. Circolone, J. Granzin, K. Jentzsch, T. Drepper, K.-E. Jaeger, D. Willbold, U. Krauss, R. Batra-Safferling, Structural Basis for the Slow Dark Recovery of a Full-Length LOV Protein from Pseudomonas putida, Journal of Molecular Biology 417 (4) (2012) 362–374.
URL http://www.sciencedirect.com/science/article/pii/S0022283612001258

[27] C. Cloix, G. I. Jenkins, Interaction of the Arabidopsis UV-B-specific signaling component UVR8 with chromatin, Mol Plant 1 (1) (2008) 118–128.

[28] C. Cole, J. D. Barber, G. J. Barton, The Jpred 3 secondary structure prediction server, Nucleic Acids Research 36 (Web Server) (2008) W197–W201.
URL http://nar.oxfordjournals.org/content/36/suppl_2/W197.abstract

[29] K. S. Conrad, A. M. Bilwes, B. R. Crane, Light-induced subunit dissociation by a light-oxygen-voltage domain photoreceptor from Rhodobacter sphaeroides, Biochemistry 52 (2) (2013) 378–391.

[30] S. B. Corchnoy, T. E. Swartz, J. W. Lewis, I. Szundi, W. R. Briggs, R. A. Bogomolni, Intramolecular proton transfers and structural changes during the photocycle of the LOV2 domain of phototropin 1, J. Biol. Chem. 278 (2) (2003) 724–731.

[31] S. Crosson, K. Moffat, Photoexcited structure of a plant photoreceptor domain reveals a light-driven molecular switch, Plant Cell 14 (5) (2002) 1067–1075.

[32] S. J. Davis, A. V. Vener, R. D. Vierstra, Bacteriophytochromes: Phytochrome-Like Photoreceptors from Nonphotosynthetic Eubacteria, Science 286 (5449) (1999) 2517–2520.
URL http://science.sciencemag.org/content/286/5449/2517

[33] C. W. Dessauer, A. G. Gilman, Purification and Characterization of a Soluble Form of Mammalian Adenylyl Cyclase, J. Biol. Chem. 271 (28) (1996) 16967–16974.
URL http://www.jbc.org/content/271/28/16967

[34] D. Devaurs, D. A. Antunes, M. Papanastasiou, M. Moll, D. Ricklin, J. D. Lambris, L. E. Kavraki, Coarse-Grained Conformational Sampling of Protein Structure Improves the Fit to Experimental Hydrogen-Exchange Data, Front Mol Biosci 4.
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5344923/

[35] S. R. Dharmaraj, E. R. Silva, A. L. Pina, Y. Y. Li, J. M. Yang, C. R. Carter, M. K. Loyer, H. K. El-Hilali, E. K. Traboulsi, O. K. Sundin, D. K. Zhu, R. K. Koenekoop, I. H. Maumenee, Mutational analysis and clinical correlation in Leber congenital amaurosis, Ophthalmic Genet. 21 (3) (2000) 135–150.

[36] R. P. Diensthuber, M. Bommer, T. Gleichmann, A. Möglich, Full-length structure of a sensor histidine kinase pinpoints coaxial coiled coils as signal transducers and modulators, Structure 21 (7) (2013) 1127–1136.

[37] T. Domratcheva, B. L. Grigorenko, I. Schlichting, A. V. Nemukhin, Molecular models predict light-induced glutamine tautomerization in BLUF photoreceptors, Biophysical Journal 94 (10) (2008) 3872–3879.

[38] T. Domratcheva, E. Hartmann, I. Schlichting, T. Kottke, Evidence for Tautomerisation of Glutamine in BLUF Blue Light Receptors by Vibrational Spectroscopy and Computational Chemistry, Sci Rep 6 (2016) 22669.

[39] Q. Dong, X. Wang, L. Lin, Z. Xu, Domain boundary prediction based on profile domain linker propensity index, Comput Biol Chem 30 (2) (2006) 127–133.

[40] Y. Duan, P. A. Kollman, Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution, Science 282 (5389) (1998) 740–744.
URL http://science.sciencemag.org/content/282/5389/740

[41] W. J. Duddy, J. W. M. Nissink, F. H. Allen, E. J. Milner-White, Mimicry by asx- and ST-turns of the four main types of $\beta$-turn in proteins, Protein Sci 13 (11) (2004) 3051–3055.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2286581/

[42] M. Dumontier, R. Yao, H. J. Feldman, C. W. V. Hogue, Armadillo: domain boundary prediction by amino acid composition, J. Mol. Biol. 350 (5) (2005) 1061–1073.

[43] R. L. Dunbrack Jr., M. Karplus, Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains, Nature Structural and Molecular Biology 1 (5) (1994) 334.
URL https://www.nature.com/articles/nsb0594-334

[44] C. M. Dunham, E. M. Dioum, J. R. Tuckerman, G. Gonzalez, W. G. Scott, M. A. Gilles-Gonzalez, A distal arginine in oxygen-sensing heme-PAS domains is essential to ligand binding, signal transduction, and structure., Biochemistry 42 (25) (2003) 7701–7708.
URL http://europepmc.org/abstract/MED/12820879

[45] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr. D Biol. Crystallogr. 66 (Pt 4) (2010) 486–501.

[46] C. Engelhard, R. P. Diensthuber, A. Möglich, R. Bittl, Blue-light reception through quaternary transitions, Scientific Reports 7 (1) (2017) 1385.
URL https://www.nature.com/articles/s41598-017-01497-7

[47] G. Esposito, B. S. Jaiswal, F. Xie, M. A. M. Krajnc-Franken, T. J. A. A. Robben, A. M. Strik, C. Kuil, R. L. A. Philipsen, M. v. Duin, M. Conti, J. A. Gossen, Mice deficient for soluble adenylyl cyclase are infertile because of a severe sperm-motility defect, PNAS 101 (9) (2004) 2993–2998.
URL http://www.pnas.org/content/101/9/2993

[48] H. U. Ferris, S. Dunin-Horkawicz, N. Hornig, M. Hulko, J. Martin, J. E. Schultz, K. Zeth, A. N. Lupas, M. Coles, Mechanism of Regulation of Receptor Histidine Kinases, Structure 20 (1) (2012) 56–66.
URL http://www.sciencedirect.com/science/article/pii/S0969212611004540

[49] D. Franke, M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H. D. T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries, D. I. Svergun, ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions, J Appl Cryst, J Appl Crystallogr 50 (4).
URL http://scripts.iucr.org/cgi-bin/paper?ge5042

[50] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. Wilkins, R. Appel, A. Bairoch, Protein Identification and Analysis Tools on the ExPASy Server, in: J. Walker (ed.), The Proteomics Protocols Handbook, Humana Press, 2005, pp. 571–607, dOI: 10.1385/1-59259-890-0:571.
URL http://dx.doi.org/10.1385/1-59259-890-0%3A571

[51] M. Gauden, S. Yeremenko, W. Laan, I. H. M. van Stokkum, J. A. Ihalainen, R. van Grondelle, K. J. Hellingwerf, J. T. M. Kennis, Photocycle of the flavin-binding photoreceptor AppA, a bacterial transcriptional antirepressor of photosynthesis genes, Biochemistry 44 (10) (2005) 3653–3662.

[52] S. T. Glantz, E. J. Carpenter, M. Melkonian, K. H. Gardner, E. S. Boyden, G. K.-S. Wong, B. Y. Chow, Functional and topological diversity of LOV domain photoreceptors, PNAS 113 (11) (2016) E1442–E1451.
URL http://www.pnas.org/content/113/11/E1442

[53] O. Glatter, Data Treatment, in: O. Glatter, O. Kratky (eds.), Small Angle X-ray Scattering, chap. 4, Academic Press, 1982, p. 155.

[54] M. Gomelsky, S. Kaplan, AppA, a Redox Regulator of Photosystem Formation in Rhodobacter sphaeroides 2.4.1, Is a Flavoprotein, J. Biol. Chem. 273 (52) (1998) 35319–35325.
URL http://www.jbc.org/content/273/52/35319

[55] M. Gomelsky, G. Klug, BLUF: a novel FAD-binding domain involved in sensory transduction in microorganisms, Trends in Biochemical Sciences 27 (10) (2002) 497–500.
URL http://www.cell.com/article/S0968000402021813/abstract

[56] G. Gourinchas, S. Etzl, C. Göbl, U. Vide, T. Madl, A. Winkler, Long-range allosteric signaling in red light-regulated diguanylyl cyclases, Sci Adv 3 (3) (2017) e1602498.

[57] J. S. Grinstead, S.-T. D. Hsu, W. Laan, A. M. J. J. Bonvin, K. J. Hellingwerf, R. Boelens, R. Kaptein, The Solution Structure of the AppA BLUF Domain: Insight into the Mechanism of Light-Induced Signaling, ChemBioChem 7 (1) (2006) 187–193.
URL http://onlinelibrary.wiley.com/doi/10.1002/cbic.200500270/abstract

[58] D. Gront, D. W. Kulp, R. M. Vernon, C. E. M. Strauss, D. Baker, Generalized Fragment Picking in Rosetta: Design, Protocols and Applications, PLoS One 6 (8).
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3160850/

[59] M. Grusch, K. Schelch, R. Riedler, E. Reichhart, C. Differ, W. Berger, Á. Inglés-Prieto, H. Janovjak, Spatio-temporally precise activation of engineered receptor tyrosine kinases by light, EMBO J. 33 (15) (2014) 1713–1726.

[60] M. Guharoy, P. Chakrabarti, Conservation and relative importance of residues across protein-protein interfaces, PNAS 102 (43) (2005) 15447–15452.
URL http://www.pnas.org/content/102/43/15447

[61] J. Guo, H.-X. Zhou, Protein Allostery and Conformational Dynamics, Chem. Rev. 116 (11) (2016) 6503–6515.

[62] M. Gwizdala, A. Wilson, D. Kirilovsky, In vitro reconstitution of the cyanobacterial photoprotective mechanism mediated by the Orange Carotenoid Protein in Synechocystis PCC 6803, Plant Cell 23 (7) (2011) 2631–2643.

[63] A. T. Hagler, E. Huler, S. Lifson, Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals, J. Am. Chem. Soc. 96 (17) (1974) 5319–5327.

[64] D. K. Hahn, J. R. Tusell, S. R. Sprang, X. Chu, Catalytic Mechanism of Mammalian Adenylyl Cyclase: A Computational Investigation, Biochemistry 54 (40) (2015) 6252–6262.

[65] K. F. Han, D. Baker, Global properties of the mapping between local amino acid sequence and local structure in proteins, PNAS 93 (12) (1996) 5814–5818.
URL http://www.pnas.org/content/93/12/5814

[66] Y. Han, S. Braatsch, L. Osterloh, G. Klug, A eukaryotic BLUF domain mediates light-dependent gene expression in the purple bacterium Rhodobacter sphaeroides 2.4.1, Proc. Natl. Acad. Sci. U.S.A. 101 (33) (2004) 12306–12311.
URL http://www.ncbi.nlm.nih.gov/pubmed/15292515

[67] S. M. Harper, L. C. Neil, I. J. Day, P. J. Hore, K. H. Gardner, Conformational changes in a photosensory LOV domain monitored by time-resolved NMR spectroscopy, J. Am. Chem. Soc. 126 (11) (2004) 3390–3391.

[68] S. M. Harper, L. C. Neil, K. H. Gardner, Structural basis of a phototropin light switch, Science 301 (5639) (2003) 1541–1544.

[69] M. E. Hatley, B. K. Benton, J. Xu, J. P. Manfredi, A. G. Gilman, R. K. Sunahara, Isolation and characterization of constitutively active mutants of mammalian adenylyl cyclase, J. Biol. Chem. 275 (49) (2000) 38626–38632.

[70] U. Heintz, A. Meinhart, A. Winkler, Multi-PAS domain-mediated protein oligomerization of PpsR from Rhodobacter sphaeroides, Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr 70 (3) (2014) 863–876.
URL http://scripts.iucr.org/cgi-bin/paper?cb5045

[71] U. Heintz, I. Schlichting, Blue light-induced LOV domain dimerization enhances the affinity of Aureochrome 1a for its target DNA sequence, Elife 5 (2016) e11860.

[72] A. Hemsley, N. Arnheim, M. D. Toney, G. Cortopassi, D. J. Galas, A simple method for site-directed mutagenesis using the polymerase chain reaction., Nucleic Acids Res 17 (16) (1989) 6545–6551.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC318348/

[73] E. Herman, M. Sachse, P. G. Kroth, T. Kottke, Blue-Light-Induced Unfolding of the J$\alpha$ Helix Allows for the Dimerization of Aureochrome-LOV from the Diatom Phaeodactylum tricornutum, Biochemistry 52 (18) (2013) 3094–3101.
URL http://dx.doi.org/10.1021/bi400197u

[74] J. Herrou, S. Crosson, Function, structure and mechanism of bacterial photosensory LOV proteins, Nat. Rev. Microbiol. 9 (10) (2011) 713–723.

[75] K. C. Hess, B. H. Jones, B. Marquez, Y. Chen, T. S. Ord, M. Kamenetsky, C. Miyamoto, J. H. Zippin, G. S. Kopf, S. S. Suarez, L. R. Levin, C. J. Williams, J. Buck, S. B. Moss, The "soluble" adenylyl cyclase in sperm mediates multiple signaling events required for fertilization, Dev. Cell 9 (2) (2005) 249–259.

[76] L. Hong, J. Lei, Scaling law for the radius of gyration of proteins and its dependence on hydrophobicity, J. Polym. Sci. B Polym. Phys. 47 (2) (2009) 207–214.
URL http://onlinelibrary.wiley.com/doi/10.1002/polb.21634/abstract

[77] S. J. Hubbard, P. Argos, Cavities and packing at protein interfaces, Protein Sci. 3 (12) (1994) 2194–2206.

[78] M. Hulko, F. Berndt, M. Gruber, J. U. Linder, V. Truffault, A. Schultz, J. Martin, J. E. Schultz, A. N. Lupas, M. Coles, The HAMP domain structure implies helix rotation in transmembrane signaling, Cell 126 (5) (2006) 929–940.

[79] M. Ikeuchi, T. Ishizuka, Cyanobacteriochromes: a new superfamily of tetrapyrrole-binding photoreceptors in cyanobacteria, Photochem. Photobiol. Sci. 7 (10) (2008) 1159–1167.
URL http://pubs.rsc.org/en/content/articlelanding/2008/pp/b802660m

[80] M. Iseki, S. Matsunaga, A. Murakami, K. Ohno, K. Shiga, K. Yoshida, M. Sugai, T. Takahashi, T. Hori, M. Watanabe, A blue-light-activated adenylyl cyclase mediates photoavoidance in Euglena gracilis, Nature 415 (6875) (2002) 1047–1051.
URL http://www.ncbi.nlm.nih.gov/pubmed/11875575

[81] S. Ito, A. Murakami, K. Sato, Y. Nishina, K. Shiga, T. Takahashi, S. Higashi, M. Iseki, M. Watanabe, Photocycle features of heterologously expressed and assembled eukaryotic flavin-binding BLUF domains of photoactivated adenylyl cyclase (PAC), a blue-light receptor in Euglena gracilis, Photochem. Photobiol. Sci. 4 (9) (2005) 762–769.

[82] K. Jentzsch, A. Wirtz, F. Circolone, T. Drepper, A. Losi, W. Gärtner, K.-E. Jaeger, U. Krauss, Mutual Exchange of Kinetic Properties by Extended Mutagenesis in Two Short LOV Domain Proteins from Pseudomonas putida, Biochemistry 48 (43) (2009) 10321–10333.
URL http://dx.doi.org/10.1021/bi901115z

[83] M. Jost, J. Fernández-Zapata, M. C. Polanco, J. M. Ortiz-Guerrero, P. Y.-T. Chen, G. Kang, S. Padmanabhan, M. Elías-Arnanz, C. L. Drennan, Structural basis for gene regulation by a $B_{12}$-dependent photoreceptor, Nature 526 (7574) (2015) 536.
URL https://www.nature.com/articles/nature14950

[84] A. Jung, T. Domratcheva, M. Tarutina, Q. Wu, W.-H. Ko, R. L. Shoeman, M. Gomelsky, K. H. Gardner, I. Schlichting, Structure of a bacterial BLUF photoreceptor: insights into blue light-mediated signal transduction, Proc. Natl. Acad. Sci. U.S.A. 102 (35) (2005) 12350–12355.
URL http://www.ncbi.nlm.nih.gov/pubmed/16107542

[85] A. Jung, J. Reinstein, T. Domratcheva, R. L. Shoeman, I. Schlichting, Crystal structures of the AppA BLUF domain photoreceptor provide insights into blue light-mediated signal transduction, J. Mol. Biol. 362 (4) (2006) 717–732.
URL http://www.ncbi.nlm.nih.gov/pubmed/16949615

[86] W. Kabsch, XDS, Acta Crystallogr D Biol Crystallogr 66 (Pt 2) (2010) 125–132.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2815665/

[87] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (12) (1983) 2577–2637.

[88] T. Kanacher, A. Schultz, J. U. Linder, J. E. Schultz, A GAF-domain-regulated adenylyl cyclase from Anabaena is a self-activating cAMP switch, EMBO J 21 (14) (2002) 3672–3680.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC126120/

[89] K. Kanchan, J. Linder, K. Winkler, K. Hantke, A. Schultz, J. E. Schultz, Transmembrane signaling in chimeras of the Escherichia coli aspartate and serine chemotaxis receptors and bacterial class III adenylyl cyclases, J. Biol. Chem. 285 (3) (2010) 2090–2099.

[90] M. Katayama, M. Ohmori, Isolation and characterization of multiple adenylate cyclase genes from the cyanobacterium Anabaena sp. strain PCC 7120., J Bacteriol 179 (11) (1997) 3588–3593.
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC179152/

[91] F. Kawano, H. Suzuki, A. Furuya, M. Sato, Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins, Nature Communications 6 (2015) 6256.
URL https://www.nature.com/articles/ncomms7256

[92] T. Kay Holt, D. W. Krogmann, A carotenoid-protein from cyanobacteria, Biochimica Et Biophysica Acta-bioenergetics - BBA-BIOENERGETICS 637 (1981) 408–414.

[93] C. A. Kerfeld, M. R. Sawaya, V. Brahmandam, D. Cascio, K. K. Ho, C. C. Trevithick-Sutton, D. W. Krogmann, T. O. Yeates, The Crystal Structure of a Cyanobacterial Water-Soluble Carotenoid Binding Protein, Structure 11 (1) (2003) 55–65.
URL http://www.sciencedirect.com/science/article/pii/S096921260200936X

[94] S. Kerruth, P. Langner, S. Raffelberg, W. Gärtner, J. Heberle, Characterization of the Blue-Light-Activated Adenylyl Cyclase mPAC by Flash Photolysis and FTIR Spectroscopy, Photochem. Photobiol. 93 (3) (2017) 857–864.

[95] M. G. Khrenova, A. V. Nemukhin, T. Domratcheva, Photoinduced electron transfer facilitates tautomerization of the conserved signaling glutamine side chain in BLUF protein light sensors, J Phys Chem B 117 (8) (2013) 2369–2377.

[96] T. Kim, M. Folcher, M. D.-E. Baba, M. Fussenegger, A Synthetic Erectile Optogenetic Stimulator Enabling Blue-Light-Inducible Penile Erection, Angew. Chem. Int. Ed. 54 (20) (2015) 5933–5938.
URL http://onlinelibrary.wiley.com/doi/10.1002/anie.201412204/abstract

[97] N. P. King, W. Sheffler, M. R. Sawaya, B. S. Vollmar, J. P. Sumida, I. André, T. Gonen, T. O. Yeates, D. Baker, Computational design of self-assembling protein nanomaterials with atomic level accuracy, Science 336 (6085) (2012) 1171–1174.

[98] A. Kita, K. Okajima, Y. Morimoto, M. Ikeuchi, K. Miki, Structure of a cyanobacterial BLUF protein, Tll0078, containing a novel FAD-binding blue light sensor domain, J. Mol. Biol. 349 (1) (2005) 1–9.

[99] S. Kleinboelting, A. Diaz, S. Moniot, J. van den Heuvel, M. Weyand, L. R. Levin, J. Buck, C. Steegborn, Crystal structures of human soluble adenylyl cyclase reveal mechanisms of catalysis and of its activation through bicarbonate, Proc. Natl. Acad. Sci. U.S.A. 111 (10) (2014) 3727–3732.

[100] S. Kleinboelting, J. van den Heuvel, C. Steegborn, Structural analysis of human soluble adenylyl cyclase and crystal structures of its nucleotide complexes-implications for cyclase catalysis and evolution, FEBS J. 281 (18) (2014) 4151–4164.

[101] T. Kottke, J. Heberle, D. Hehn, B. Dick, P. Hegemann, Phot-LOV1: photocycle of a blue-light receptor domain from the green alga Chlamydomonas reinhardtii, Biophys. J. 84 (2 Pt 1) (2003) 1192–1201.

[102] M. M. Krishna, L. Hoang, Y. Lin, S. Englander, Hydrogen exchange methods to study protein folding, Methods 34 (1) (2004) 51–64.
URL http://www.sciencedirect.com/science/article/pii/S1046202304000519

[103] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, J. Mol. Biol. 372 (3) (2007) 774–797.

[104] K. Kuroi, K. Okajima, M. Ikeuchi, S. Tokutomi, T. Kamiyama, M. Terazima, Pressure-Sensitive Reaction Yield of the TePixD Blue-Light Sensor Protein, J. Phys. Chem. B 119 (7) (2015) 2897–2907.
URL http://dx.doi.org/10.1021/jp511946u

[105] R. J. Kutta, K. Magerl, U. Kensy, B. Dick, A search for radical intermediates in the photocycle of LOV domains, Photochem. Photobiol. Sci. 14 (2) (2015) 288–299.

[106] U. K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 (5259) (1970) 680–685.

[107] M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins, Clustal W and Clustal X version 2.0, Bioinformatics 23 (21) (2007) 2947–2948.

[108] E. Lesne, E. Dupré, C. Locht, R. Antoine, F. Jacob-Dubuisson, Conformational Changes of an Interdomain Linker Mediate Mechanical Signal Transmission in Sensor Kinase BvgS, J. Bacteriol. 199 (18).

[109] E. Lesne, E.-M. Krammer, E. Dupre, C. Locht, M. F. Lensink, R. Antoine, F. Jacob-Dubuisson, Balance between Coiled-Coil Stability and Dynamics Regulates Activity of BvgS Sensor Kinase in Bordetella, MBio 7 (2) (2016) e02089.

[110] R. L. Leverenz, M. Sutter, A. Wilson, S. Gupta, A. Thurotte, C. B. d. Carbon, C. J. Petzold, C. Ralston, F. Perreau, D. Kirilovsky, C. A. Kerfeld, A 12 Å carotenoid translocation in a photoswitch associated with cyanobacterial photoprotection, Science 348 (6242) (2015) 1463–1466.
URL http://science.sciencemag.org/content/348/6242/1463

[111] M. Levitt, Nature of the protein universe, Proc. Natl. Acad. Sci. U.S.A. 106 (27) (2009) 11079–11084.

[112] X. Li, D. V. Gutierrez, M. G. Hanson, J. Han, M. D. Mark, H. Chiel, P. Hegemann, L. T. Landmesser, S. Herlitze, Fast noninvasive activation and inhibition of neural and network activity by vertebrate rhodopsin and green algae channelrhodopsin, PNAS 102 (49) (2005) 17816–17821.
URL http://www.pnas.org/content/102/49/17816

[113] J. U. Linder, A. Schultz, J. E. Schultz, Adenylyl cyclase Rv1264 from Mycobacterium tuberculosis has an autoinhibitory N-terminal domain, J. Biol. Chem. 277 (18) (2002) 15271–15276.

[114] J. U. Linder, J. E. Schultz, The class III adenylyl cyclases: multi-purpose signalling modules, Cellular Signalling 15 (12) (2003) 1081–1089.
URL http://www.sciencedirect.com/science/article/pii/S089865680300130X

[115] R. Lindner, E. Hartmann, M. Tarnawski, A. Winkler, D. Frey, J. Reinstein, A. Meinhart, I. Schlichting, Photoactivation Mechanism of a Bacterial Light-Regulated Adenylyl Cyclase, J. Mol. Biol. 429 (9) (2017) 1336–1351.

[116] R. Lindner, U. Heintz, A. Winkler, Applications of hydrogen deuterium exchange (HDX) for the characterization of conformational dynamics in light-activated photoreceptors, Front. Mol. Biosci 2 (2015) 33.
URL http://journal.frontiersin.org/article/10.3389/fmolb.2015.00033/full

[117] R. Lindner, X. Lou, J. Reinstein, R. L. Shoeman, F. A. Hamprecht, A. Winkler, Hexicon 2: Automated Processing of Hydrogen-Deuterium Exchange Mass Spectrometry Data with Improved Deuteration Distribution Estimation, J. Am. Soc. Mass Spectrom. 25 (6) (2014) 1018–1028.
URL http://link.springer.com/article/10.1007/s13361-014-0850-y

[118] Y. Liu, A. E. Ruoho, V. D. Rao, J. H. Hurley, Catalytic mechanism of the adenylyl and guanylyl cyclases: modeling and mutational analysis, Proc. Natl. Acad. Sci. U.S.A. 94 (25) (1997) 13414–13419.

[119] J. Lokhandwala, R. I. Silverman y de la Vega, H. C. Hopkins, C. W. Britton, A. Rodriguez-Iglesias, R. Bogomolni, M. Schmoll, B. D. Zoltowski, A Native Threonine Coordinates Ordered Water to Tune Light-Oxygen-Voltage (LOV) Domain Photocycle Kinetics and Osmotic Stress Signaling in Trichoderma reesei ENVOY, J Biol Chem 291 (28) (2016) 14839–14850.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4938200/

[120] J. Looser, S. Schröder-Lang, P. Hegemann, G. Nagel, Mechanistic insights in light-induced cAMP production by photoactivated adenylyl cyclase alpha (PACalpha), Biol. Chem. 390 (11) (2009) 1105–1111.

[121] A. Lupas, M. V. Dyke, J. Stock, Predicting coiled coils from protein sequences, Science 252 (5009) (1991) 1162–1164.
URL http://science.sciencemag.org/content/252/5009/1162

[122] A. N. Lupas, J. Bassler, S. Dunin-Horkawicz, The Structure and Topology of $\alpha$-Helical Coiled Coils, in: Fibrous Proteins: Structures and Mechanisms, Subcellular Biochemistry, Springer, Cham, 2017, pp. 95–129, dOI: 10.1007/978-3-319-49674-0_4.
URL https://link.springer.com/chapter/10.1007/978-3-319-49674-0_4

[123] X. Ma, A. Beuve, F. van den Akker, Crystal structure of the signaling helix coiled-coil domain of the $\beta$1 subunit of the soluble guanylyl cyclase, BMC Struct Biol 10 (2010) 2.
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2828450/

[124] P. Machamer, L. Darden, C. F. Craver, Thinking about Mechanisms, Philosophy of Science 67 (1) (2000) 1–25.
URL http://www.journals.uchicago.edu/doi/abs/10.1086/392759

[125] S. Masuda, Light detection and signal transduction in the BLUF photoreceptors, Plant Cell Physiol. 54 (2) (2013) 171–179.

[126] S. Masuda, C. E. Bauer, AppA is a blue light photoreceptor that antirepresses photosynthesis gene expression in Rhodobacter sphaeroides, Cell 110 (5) (2002) 613–623.
URL http://www.ncbi.nlm.nih.gov/pubmed/12230978

[127] S. Masuda, K. Hasegawa, A. Ishii, T.-a. Ono, Light-induced structural changes in a putative blue-light receptor with a novel FAD binding fold sensor of blue-light using FAD (BLUF); Slr1694 of synechocystis sp. PCC6803, Biochemistry 43 (18) (2004) 5304–5313.

[128] S. Masuda, K. Hasegawa, H. Ohta, T.-a. Ono, Crucial role in light signal transduction for the conserved Met93 of the BLUF protein PixD/Slr1694, Plant Cell Physiol. 49 (10) (2008) 1600–1606.

[129] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software, J Appl Crystallogr 40 (Pt 4) (2007) 658–674.

[130] L. J. McGuffin, K. Bryson, D. T. Jones, The PSIPRED protein structure prediction server, Bioinformatics 16 (4) (2000) 404–405.

[131] N. E. Mencacci, R. Erro, S. Wiethoff, J. Hersheson, M. Ryten, B. Balint, C. Ganos, M. Stamelou, N. Quinn, H. Houlden, N. W. Wood, K. P. Bhatia, ADCY5 mutations are another cause of benign hereditary chorea, Neurology 85 (1) (2015) 80–88.

[132] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, Equation of State Calculations by Fast Computing Machines, The Journal of Chemical Physics 21 (6) (1953) 1087–1092.
URL http://aip.scitation.org/doi/10.1063/1.1699114

[133] S. Metz, J. Hendriks, A. Jäger, K. Hellingwerf, G. Klug, In Vivo Effects on Photosynthesis Gene Expression of Base Pair Exchanges in the Gene Encoding the Light-responsive BLUF Domain of AppA in Rhodobacter Sphaeroides, Photochemistry and Photobiology 86 (4) (2010) 882–889.
URL http://onlinelibrary.wiley.com/doi/10.1111/j.1751-1097.2010.00749.x/abstract

[134] D. Mitra, X. Yang, K. Moffat, Crystal structures of Aureochrome1 LOV suggest new design strategies for optogenetics, Structure 20 (4) (2012) 698–706.

[135] A. Möglich, R. A. Ayers, K. Moffat, Design and signaling mechanism of light-regulated histidine kinases, J. Mol. Biol. 385 (5) (2009) 1433–1444.

[136] A. Möglich, R. A. Ayers, K. Moffat, Structure and signaling mechanism of Per-ARNT-Sim domains, Structure 17 (10) (2009) 1282–1294.

[137] A. Moglich, K. Moffat, Structural Basis for Light-dependent Signaling in the Dimeric LOV Domain of the Photosensor YtvA, J Mol Biol 373 (1) (2007) 112–126.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2175523/

[138] T.-C. Mou, A. Gille, D. A. Fancy, R. Seifert, S. R. Sprang, Structural basis for the inhibition of mammalian membrane adenylyl cyclase by 2 '(3')-O-(N-Methylanthraniloyl)-guanosine 5 '-triphosphate, J. Biol. Chem. 280 (8) (2005) 7253–7261.

[139] V. R. Mundlapati, S. Ghosh, A. Bhattacherjee, P. Tiwari, H. S. Biswal, Critical Assessment of the Strength of Hydrogen Bonds between the Sulfur Atom of Methionine/Cysteine and Backbone Amides in Proteins, J Phys Chem Lett 6 (8) (2015) 1385–1389.

[140] G. Nagel, M. Brauner, J. F. Liewald, N. Adeishvili, E. Bamberg, A. Gottschalk, Light Activation of Channelrhodopsin-2 in Excitable Cells of Caenorhabditis elegans Triggers Rapid Behavioral Responses, Current Biology 15 (24) (2005) 2279–2284.
URL http://www.sciencedirect.com/science/article/pii/S0960982205014077

[141] G. Nagel, D. Ollig, M. Fuhrmann, S. Kateriya, A. M. Musti, E. Bamberg, P. Hegemann, Channelrhodopsin-1: A Light-Gated Proton Channel in Green Algae, Science 296 (5577) (2002) 2395–2398.
URL http://science.sciencemag.org/content/296/5577/2395

[142] A. I. Nash, W.-H. Ko, S. M. Harper, K. H. Gardner, A conserved glutamine plays a central role in LOV domain signal transmission and duration, Biochemistry 47 (52) (2008) 13842–13849.
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2630409/

[143] A. I. Nash, R. McNulty, M. E. Shillito, T. E. Swartz, R. A. Bogomolni, H. Luecke, K. H. Gardner, Structural basis of photosensitivity in a bacterial light-oxygen-voltage/helix-turn-helix (LOV-HTH) DNA-binding protein, Proc. Natl. Acad. Sci. U.S.A. 108 (23) (2011) 9449–9454.

[144] H. Nguyen, J. Maier, H. Huang, V. Perrone, C. Simmerling, Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent, J. Am. Chem. Soc. 136 (40) (2014) 13959–13962.

[145] D. Nozaki, T. Iwata, T. Ishikawa, T. Todo, S. Tokutomi, H. Kandori, Role of Gln1029 in the photoactivation processes of the LOV2 domain in adiantum phytochrome3, Biochemistry 43 (26) (2004) 8373–8379.

[146] D. Oesterhelt, W. Stoeckenius, Functions of a new photoreceptor membrane, Proc. Natl. Acad. Sci. U.S.A. 70 (10) (1973) 2853–2857.

[147] Y. Ofran, B. Rost, Protein-protein interaction hotspots carved into sequences, PLoS Comput. Biol. 3 (7) (2007) e119.

[148] M. Ohki, A. Sato-Tomita, S. Matsunaga, M. Iseki, J. R. H. Tame, N. Shibayama, S.-Y. Park, Molecular mechanism of photoactivation of a light-regulated adenylate cyclase, PNAS 114 (32) (2017) 8562–8567.
URL http://www.pnas.org/content/114/32/8562

[149] M. Ohki, K. Sugiyama, F. Kawai, H. Tanaka, Y. Nihei, S. Unzai, M. Takebe, S. Matsunaga, S.-I. Adachi, N. Shibayama, Z. Zhou, R. Koyama, Y. Ikegaya, T. Takahashi, J. R. H. Tame, M. Iseki, S.-Y. Park, Structural insight into photoactivation of an adenylate cyclase from a photosynthetic cyanobacterium, Proc. Natl. Acad. Sci. U.S.A. 113 (24) (2016) 6659–6664.

[150] R. Ohlendorf, C. H. Schumacher, F. Richter, A. Möglich, Library-Aided Probing of Linker Determinants in Hybrid Photoreceptors, ACS Synth. Biol. 5 (10) (2016) 1117–1126.
URL http://dx.doi.org/10.1021/acssynbio.6b00028

[151] M. Ormö, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien, S. J. Remington, Crystal Structure of the Aequorea victoria Green Fluorescent Protein, Science 273 (5280) (1996) 1392–1395.
URL http://science.sciencemag.org/content/273/5280/1392

[152] J. M. Ortiz-Guerrero, M. C. Polanco, F. J. Murillo, S. Padmanabhan, M. Elías-Arnanz, Light-dependent gene regulation by a coenzyme B12-based photoreceptor, PNAS 108 (18) (2011) 7565–7570.
URL http://www.pnas.org/content/108/18/7565

[153] J. E. Padilla, C. Colovos, T. O. Yeates, Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments, Proc. Natl. Acad. Sci. U.S.A. 98 (5) (2001) 2217–2221.

[154] J. Pandit, M. D. Forman, K. F. Fennell, K. S. Dillman, F. S. Menniti, Mechanism for the allosteric regulation of phosphodiesterase 2a deduced from the X-ray structure of a near full-length construct, Proc. Natl. Acad. Sci. U.S.A. 106 (43) (2009) 18225–18230.

[155] J. L. Pitman, M. C. Wheeler, D. J. Lloyd, J. R. Walker, R. J. Glynne, N. Gekakis, A gain-of-function mutation in adenylate cyclase 3 protects mice from diet-induced obesity, PLoS ONE 9 (10) (2014) e110226.

[156] A. Pudasaini, K. K. El-Arab, B. D. Zoltowski, LOV-based optogenetic devices: light-driven modules to impart photoregulated control of cellular signaling, Front Mol Biosci 2 (2015) 18.

[157] E. B. Purcell, C. A. McDonald, B. A. Palfey, S. Crosson, An Analysis of the Solution Structure and Signaling Mechanism of LovK, a Sensor Histidine Kinase Integrating Light and Redox Signals, Biochemistry 49 (31) (2010) 6761–6770.
URL http://dx.doi.org/10.1021/bi1006404

[158] S. Raffelberg, M. Mansurova, W. Gärtner, A. Losi, Modulation of the photocycle of a LOV domain photoreceptor by the hydrogen-bonding network, J. Am. Chem. Soc. 133 (14) (2011) 5346–5356.

[159] S. Raffelberg, L. Wang, S. Gao, A. Losi, W. Gärtner, G. Nagel, A LOV-domain-mediated blue-light-activated adenylate (adenylyl) cyclase from the cyanobacterium Microcoleus chthonoplastes PCC 7420, Biochem. J. 455 (3) (2013) 359–365.

[160] S. Raman, O. F. Lange, P. Rossi, M. Tyka, X. Wang, J. Aramini, G. Liu, T. A. Ramelot, A. Eletsky, T. Szyperski, M. A. Kennedy, J. Prestegard, G. T. Montelione, D. Baker, NMR structure determination for larger proteins using backbone-only data, Science 327 (5968) (2010) 1014–1018.

[161] R. P. Rambo, J. A. Tainer, Characterizing Flexible and Instrinsically Unstructured Biological Macromolecules by SAS using the Porod-Debye Law, Biopolymers 95 (8) (2011) 559–571.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3103662/

[162] A. Rauch, M. Leipelt, M. Russwurm, C. Steegborn, Crystal structure of the guanylyl cyclase Cya2, Proc. Natl. Acad. Sci. U.S.A. 105 (41) (2008) 15720–15725.

[163] S. Ren, M. Sawada, K. Hasegawa, Y. Hayakawa, H. Ohta, S. Masuda, A PixD–PapB chimeric protein reveals the function of the BLUF domain C-terminal $\alpha$-helices for light signal transduction, Plant Cell Physiol. 53 (9) (2012) 1638–1647.

[164] G. Rivera-Cancel, W.-h. Ko, D. R. Tomchick, F. Correa, K. H. Gardner, Full-length structure of a monomeric histidine kinase reveals basis for sensory regulation, Proc. Natl. Acad. Sci. U.S.A. 111 (50) (2014) 17839–17844.

[165] L. Rizzini, J.-J. Favory, C. Cloix, D. Faggionato, A. O'Hara, E. Kaiserli, R. Baumeister, E. Schäfer, F. Nagy, G. I. Jenkins, R. Ulm, Perception of UV-B by the Arabidopsis UVR8 protein, Science 332 (6025) (2011) 103–106.

[166] C. A. Rohl, C. E. Strauss, K. M. Misura, D. Baker, Protein Structure Prediction Using Rosetta, in: Ludwig Brand and Michael L. Johnson (ed.), Methods in Enzymology, vol. Volume 383 of Numerical Computer Methods, Part D, Academic Press, 2004, pp. 66–93.
URL http://www.sciencedirect.com/science/article/pii/S0076687904830040

[167] K. Röllen, J. Granzin, V. Panwalkar, V. Arinkin, R. Rani, R. Hartmann, U. Krauss, K.-E. Jaeger, D. Willbold, R. Batra-Safferling, Signaling States of a Short Blue-Light Photoreceptor Protein PpSB1-LOV Revealed from Crystal Structures and Solution NMR Spectroscopy, J. Mol. Biol. 428 (19) (2016) 3721–3736.

[168] P. Rossi, L. Shi, G. Liu, C. M. Barbieri, H.-W. Lee, T. D. Grant, J. R. Luft, R. Xiao, T. B. Acton, E. H. Snell, G. T. Montelione, D. Baker, O. F. Lange, N. G. Sgourakis, A hybrid NMR/SAXS-based approach for discriminating oligomeric protein interfaces using Rosetta, Proteins 83 (2) (2015) 309–317.

[169] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction, Nat Protoc 5 (4) (2010) 725–738.
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2849174/

[170] M.-H. Ryu, I.-H. Kang, M. D. Nelson, T. M. Jensen, A. I. Lyuksyutova, J. Siltberg-Liberles, D. M. Raizen, M. Gomelsky, Engineering adenylate cyclases regulated by near-infrared window light, Proc Natl Acad Sci U S A 111 (28) (2014) 10167–10172.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4104894/

[171] M.-H. Ryu, O. V. Moskvin, J. Siltberg-Liberles, M. Gomelsky, Natural and engineered photoactivated nucleotidyl cyclases for optogenetic applications, The Journal of biological chemistry 285 (53) (2010) 41501–41508.

[172] S. M. Saalau-Bethell, V. Berdini, A. Cleasby, M. Congreve, J. E. Coyle, V. Lock, C. W. Murray, M. A. O'Brien, S. J. Rich, T. Sambrook, M. Vinkovic, J. R. Yon, H. Jhoti, Crystal structure of human soluble adenylate cyclase reveals a distinct, highly flexible allosteric bicarbonate binding pocket, ChemMedChem 9 (4) (2014) 823–832.

[173] M. Salomon, J. M. Christie, E. Knieb, U. Lempert, W. R. Briggs, Photochemical and Mutational Analysis of the FMN-Binding Domains of the Plant Blue Light Receptor, Phototropin,, Biochemistry 39 (31) (2000) 9401–9410.
URL http://dx.doi.org/10.1021/bi000585+

[174] M. Salomon, W. Eisenreich, H. Dürr, E. Schleicher, E. Knieb, V. Massey, W. Rüdiger, F. Müller, A. Bacher, G. Richter, An optomechanical transducer in the blue light receptor phototropin from Avena sativa, PNAS 98 (22) (2001) 12357–12361.
URL http://www.pnas.org/content/98/22/12357

[175] A. Scaiewicz, M. Levitt, The language of the protein universe, Curr. Opin. Genet. Dev. 35 (2015) 50–56.

[176] T. R. Schneider, Objective comparison of protein structures: error-scaled difference distance matrices, Acta Crystallogr. D Biol. Crystallogr. 56 (Pt 6) (2000) 714–721.

[177] D. Schneidman-Duhovny, M. Hammel, J. A. Tainer, A. Sali, FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles, Nucleic Acids Res 44 (W1) (2016) W424–W429.
URL https://academic.oup.com/nar/article/44/W1/W424/2499365

[178] G. v. d. Schot, Z. Zhang, R. Vernon, Y. Shen, W. F. Vranken, D. Baker, A. M. J. J. Bonvin, O. F. Lange, Improving 3d structure prediction from chemical shift data, J Biomol NMR 57 (1) (2013) 27–35.
URL https://link.springer.com/article/10.1007/s10858-013-9762-6

[179] S. Schröder-Lang, M. Schwärzel, R. Seifert, T. Strünker, S. Kateriya, J. Looser, M. Watanabe, U. B. Kaupp, P. Hegemann, G. Nagel, Fast manipulation of cellular cAMP level by light in vivo, Nat. Methods 4 (1) (2007) 39–42.

[180] L. Schrödinger, The PyMOL Molecular Graphics System, Version 1.8 (Nov. 2015).

[181] J. E. Schultz, J. Natarajan, Regulated unfolding: a basic principle of intraprotein signaling in modular proteins, Trends Biochem. Sci. 38 (11) (2013) 538–545.

[182] 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, Millisecond-scale molecular dynamics simulations on Anton, in: Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis, 2009, pp. 1–11.

[183] G. M. Sheldrick, A short history of SHELX, Acta Cryst A 64 (1) (2008) 112–122.
URL http://scripts.iucr.org/cgi-bin/paper?sc5010

[184] I. Sillitoe, T. E. Lewis, A. Cuff, S. Das, P. Ashford, N. L. Dawson, N. Furnham, R. A. Laskowski, D. Lee, J. G. Lees, S. Lehtinen, R. A. Studer, J. Thornton, C. A. Orengo, CATH: comprehensive structural and functional annotations for genome sequences, Nucleic Acids Res 43 (D1) (2015) D376–D381.
URL https://academic.oup.com/nar/article/43/D1/D376/2439515

[185] K. T. Simons, C. Kooperberg, E. Huang, D. Baker, Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions11edited by F. E. Cohen, Journal of Molecular Biology 268 (1) (1997) 209–225.
URL http://www.sciencedirect.com/science/article/pii/S0022283697909591

[186] S. C. Sinha, M. Wetterer, S. R. Sprang, J. E. Schultz, J. U. Linder, Origin of asymmetry in adenylyl cyclases: structures of Mycobacterium tuberculosis Rv1900c, EMBO J. 24 (4) (2005) 663–673.

[187] D.-M. Smilgies, E. Folta-Stogniew, Molecular weight–gyration radius relation of globular proteins: a comparison of light scattering, small-angle X-ray scattering and structure-based data, J Appl Crystallogr 48 (Pt 5) (2015) 1604–1606.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4603275/

[188] P. Sønderby, Å. Rinnan, J. J. Madsen, P. Harris, J. T. Bukrinski, G. H. J. Peters, Small-Angle X-ray Scattering Data in Combination with RosettaDock Improves the Docking Energy Landscape, J Chem Inf Model 57 (10) (2017) 2463–2475.

[189] C. Steegborn, T. N. Litvin, L. R. Levin, J. Buck, H. Wu, Bicarbonate activation of adenylyl cyclase via promotion of catalytic active site closure and metal recruitment, Nat Struct Mol Biol 12 (1) (2005) 32–37.
URL http://www.nature.com/nsmb/journal/v12/n1/full/nsmb880.html

[190] A. L. Stelling, K. L. Ronayne, J. Nappa, P. J. Tonge, S. R. Meech, Ultrafast structural dynamics in BLUF domains: transient infrared spectroscopy of AppA and its mutants, J. Am. Chem. Soc. 129 (50) (2007) 15556–15564.

[191] W. Steuer Costa, S.-C. Yu, J. F. Liewald, A. Gottschalk, Fast cAMP Modulation of Neurotransmission via Neuropeptide Signals and Vesicle Loading, Curr. Biol. 27 (4) (2017) 495–507.

[192] M. Stierl, A. Penzkofer, J. T. M. Kennis, P. Hegemann, T. Mathes, Key residues for the light regulation of the blue light-activated adenylyl cyclase from Beggiatoa sp, Biochemistry 53 (31) (2014) 5121–5130.

[193] M. Stierl, P. Stumpf, D. Udwari, R. Gueta, R. Hagedorn, A. Losi, W. Gärtner, L. Petereit, M. Efetova, M. Schwarzel, T. G. Oertner, G. Nagel, P. Hegemann, Light modulation of cellular cAMP by a small bacterial photoactivated adenylyl cyclase, bPAC, of the soil bacterium Beggiatoa, J. Biol. Chem. 286 (2) (2011) 1181–1188.

[194] D. Strickland, X. Yao, G. Gawlak, M. K. Rosen, K. H. Gardner, T. R. Sosnick, Rationally improving LOV domain-based photoswitches, Nat. Methods 7 (8) (2010) 623–626.

[195] N. Sukomon, J. Widom, P. P. Borbat, J. H. Freed, B. R. Crane, Stability and Conformation of a Chemoreceptor HAMP Domain Chimera Correlates with Signaling Properties, Biophys. J. 112 (7) (2017) 1383–1395.

[196] C. M. Summa, M. Levitt, Near-native structure refinement using in vacuo energy minimization, Proc Natl Acad Sci U S A 104 (9) (2007) 3177–3182.
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1802011/

[197] R. K. Sunahara, C. W. Dessauer, R. E. Whisnant, C. Kleuss, A. G. Gilman, Interaction of Gs$\alpha$ with the Cytosolic Domains of Mammalian Adenylyl Cyclase, J. Biol. Chem. 272 (35) (1997) 22265–22271.
URL http://www.jbc.org/content/272/35/22265

[198] D. I. Svergun, Determination of the regularization parameter in indirect-transform methods using perceptual criteria, J Appl Cryst, J Appl Crystallogr 25 (4) (1992) 495–503.
URL http://scripts.iucr.org/cgi-bin/paper?wi0087

[199] D. I. Svergun, M. V. Petoukhov, M. H. Koch, Determination of domain structure of proteins from X-ray solution scattering., Biophys J 80 (6) (2001) 2946–2953.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1301478/

[200] T. E. Swartz, S. B. Corchnoy, J. M. Christie, J. W. Lewis, I. Szundi, W. R. Briggs, R. A. Bogomolni, The Photocycle of a Flavin-binding Domain of the Blue Light Photoreceptor Phototropin, J. Biol. Chem. 276 (39) (2001) 36493–36500.
URL http://www.jbc.org/content/276/39/36493

[201] K. K. Tekkanat, I. H. Fox, Isocratic separation of ATP and its degradation products from biological fluids by automated liquid chromatography, Clin. Chem. 34 (5) (1988) 925–932.

[202] T. C. Terwilliger, P. D. Adams, R. J. Read, A. J. McCoy, N. W. Moriarty, R. W. Grosse-Kunstleve, P. V. Afonine, P. H. Zwart, L.-W. Hung, Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard, Acta Cryst D, Acta Cryst Sect D, Acta

Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr 65 (6) (2009) 582–601.
URL http://scripts.iucr.org/cgi-bin/paper?ea5095

[203] J. J. Tesmer, R. K. Sunahara, A. G. Gilman, S. R. Sprang, Crystal structure of the catalytic domains of adenylyl cyclase in a complex with Gsalpha.GTPgammaS, Science 278 (5345) (1997) 1907–1916.

[204] J. J. G. Tesmer, R. K. Sunahara, R. A. Johnson, G. Gosselin, A. G. Gilman, S. R. Sprang, Two-Metal-Ion Catalysis in Adenylyl Cyclase, Science 285 (5428) (1999) 756–760.
URL http://www.sciencemag.org/content/285/5428/756

[205] I. Tews, F. Findeisen, I. Sinning, A. Schultz, J. E. Schultz, J. U. Linder, The structure of a pH-sensing mycobacterial adenylyl cyclase holoenzyme, Science 308 (5724) (2005) 1020–1023.
URL http://www.ncbi.nlm.nih.gov/pubmed/15890882

[206] A. Udvarhelyi, T. Domratcheva, Glutamine rotamers in BLUF photoreceptors: a mechanistic reappraisal, J Phys Chem B 117 (10) (2013) 2888–2897.

[207] A. T. Vaidya, C.-H. Chen, J. C. Dunlap, J. J. Loros, B. R. Crane, Structure of a light-activated LOV protein dimer that regulates transcription, Sci Signal 4 (184) (2011) ra50.

[208] S. F. Vatner, M. Park, L. Yan, G. J. Lee, L. Lai, K. Iwatsubo, Y. Ishikawa, J. Pessin, D. E. Vatner, Adenylyl cyclase type 5 in cardiac disease, metabolism, and aging, Am J Physiol Heart Circ Physiol 305 (1) (2013) H1–H8.
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3727099/

[209] I. Vercellino, L. Rezabkova, V. Olieric, Y. Polyhach, T. Weinert, R. A. Kammerer, G. Jeschke, V. M. Korkhov, Role of the nucleotidyl cyclase helical domain in catalytically active dimer formation, Proc. Natl. Acad. Sci. U.S.A. 114 (46) (2017) E9821–E9828.

[210] F. Wei, C. S. Qiu, S. J. Kim, L. Muglia, J. W. Maas, V. V. Pineda, H. M. Xu, Z. F. Chen, D. R. Storm, L. J. Muglia, M. Zhuo, Genetic elimination of behavioral sensitization in mice lacking calmodulin-stimulated adenylyl cyclases, Neuron 36 (4) (2002) 713–726.

[211] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, P. Weiner, A new force field for molecular mechanical simulation of nucleic acids and proteins, J. Am. Chem. Soc. 106 (3) (1984) 765–784.
URL http://dx.doi.org/10.1021/ja00315a051

[212] S. Weissenberger, C. Schultheis, J. F. Liewald, K. Erbguth, G. Nagel, A. Gottschalk, PAC$\alpha$– an optogenetic tool for in vivo manipulation of cellular cAMP levels, neurotransmitter release, and behavior in Caenorhabditis elegans, Journal of Neurochemistry 116 (4) (2011) 616–625.
URL http://onlinelibrary.wiley.com/doi/10.1111/j.1471-4159.2010.07148.x/abstract

[213] D. B. Wetlaufer, Nucleation, Rapid Folding, and Globular Intrachain Regions in Proteins, Proc Natl Acad Sci U S A 70 (3) (1973) 697–701.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC433338/

[214] A. Winkler, T. R. M. Barends, A. Udvarhelyi, D. Lenherr-Frey, L. Lomb, A. Menzel, I. Schlichting, Structural Details of Light Activation of the LOV2-based Photoswitch PA-Rac1, ACS Chem. Biol.

[215] A. Winkler, U. Heintz, R. Lindner, J. Reinstein, R. L. Shoeman, I. Schlichting, A ternary AppA–PpsR–DNA complex mediates light regulation of photosynthesis-related gene expression, Nat Struct Mol Biol advance online publication.
URL http://www.nature.com/nsmb/journal/vaop/ncurrent/abs/nsmb.2597.html

[216] A. Winkler, A. Udvarhelyi, E. Hartmann, J. Reinstein, A. Menzel, R. L. Shoeman, I. Schlichting, Characterization of Elements Involved in Allosteric Light Regulation of Phosphodiesterase Activity by Comparison of Different Functional BlrP1 States, Journal of Molecular Biology 426 (4) (2014) 853–868.
URL http://www.sciencedirect.com/science/article/pii/S0022283613007298

[217] K. Winkler, A. Schultz, J. Schultz, The S-helix determines the signal in a Tsr receptor/adenylyl cyclase reporter, Journal of Biological Chemistry 287 (19) (2012) 15479–15488.

[218] A. M. Wollacott, A. Zanghellini, P. Murphy, D. Baker, Prediction of structures of multidomain proteins from structures of the individual domains, Protein Sci 16 (2) (2007) 165–175.
URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2203296/

[219] P. E. Wright, H. J. Dyson, R. A. Lerner, Conformation of peptide fragments of proteins in aqueous solution: implications for initiation of protein folding, Biochemistry 27 (19) (1988) 7167–7175.
URL http://dx.doi.org/10.1021/bi00419a001

[220] Q. Wu, K. H. Gardner, Structure and Insight into Blue Light-Induced Changes in the BlrP1 BLUF Domain, Biochemistry 48 (12) (2009) 2620–2629.
URL http://dx.doi.org/10.1021/bi802237r

[221] Q. Wu, W.-H. Ko, K. H. Gardner, Structural requirements for key residues and auxiliary portions of a BLUF domain, Biochemistry 47 (39) (2008) 10271–10280.

[222] Y. I. Wu, D. Frey, O. I. Lungu, A. Jaehrig, I. Schlichting, B. Kuhlman, K. M. Hahn, A genetically encoded photoactivatable Rac controls the motility of living cells, Nature 461 (7260) (2009) 104–108.

[223] Z. L. Wu, S. A. Thomas, E. C. Villacres, Z. Xia, M. L. Simmons, C. Chavkin, R. D. Palmiter, D. R. Storm, Altered behavior and long-term potentiation in type I adenylyl cyclase mutant mice, PNAS 92 (1) (1995) 220–224.
URL http://www.pnas.org/content/92/1/220

[224] D. Xu, L. Jaroszewski, Z. Li, A. Godzik, AIDA: ab initio domain assembly for automated multi-domain protein structure prediction and domain-domain interaction prediction, Bioinformatics 31 (13) (2015) 2098–2105.

[225] D. Xu, Y. Zhang, Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field, Proteins 80 (7) (2012) 1715–1735.

[226] Y. Yang, E. Faraggi, H. Zhao, Y. Zhou, Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates, Bioinformatics 27 (15) (2011) 2076–2082.

[227] X. Yao, M. K. Rosen, K. H. Gardner, Estimation of the available free energy in a LOV2-J alpha photoswitch, Nat. Chem. Biol. 4 (8) (2008) 491–497.

[228] T. O. Yeates, J. E. Padilla, Designing supramolecular protein assemblies, Curr. Opin. Struct. Biol. 12 (4) (2002) 464–470.

[229] E. F. Yee, R. P. Diensthuber, A. T. Vaidya, P. P. Borbat, C. Engelhard, J. H. Freed, R. Bittl, A. Möglich, B. R. Crane, Signal transduction in light–oxygen–voltage receptors lacking the adduct-forming cysteine residue, Nat Commun 6.
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4682037/

[230] H. Yuan, S. Anderson, S. Masuda, V. Dragnea, K. Moffat, C. Bauer, Crystal Structures of the Synechocystis Photoreceptor Slr1694 Reveal Distinct Structural States Related to Signaling,, Biochemistry 45 (42) (2006) 12687–12694.
URL http://dx.doi.org/10.1021/bi061435n

[231] H. Yuan, V. Dragnea, Q. Wu, K. H. Gardner, C. E. Bauer, Mutational And Structural Studies Of The PixD BLUF Output Signal That Affects Light-Regulated Interactions With PixE, Biochemistry 50 (29) (2011) 6365–6375.

[232] Y. Zhang, A. Kolinski, J. Skolnick, TOUCHSTONE II: A New Approach to Ab Initio Protein Structure Prediction, Biophysical Journal 85 (2) (2003) 1145–1164.
URL http://www.sciencedirect.com/science/article/pii/S0006349503745512

[233] Q. Zhou, P. Ames, J. S. Parkinson, Mutational analyses of HAMP helices suggest a dynamic bundle model of input-output signalling in chemoreceptors, Mol. Microbiol. 73 (5) (2009) 801–814.

[234] M. Ziegler, J. Bassler, S. Beltz, A. Schultz, A. N. Lupas, J. E. Schultz, Characterization of a novel signal transducer element intrinsic to class IIIa/b adenylate cyclases and guanylate cyclases, FEBS J 284 (8) (2017) 1204–1217.
URL http://onlinelibrary.wiley.com/doi/10.1111/febs.14047/abstract

[235] T. Ziegler, A. Möglich, Photoreceptor engineering, Front. Mol. Biosci. 2.
URL https://www.frontiersin.org/articles/10.3389/fmolb.2015.00030/full

[236] B. D. Zoltowski, C. Schwerdtfeger, J. Widom, J. J. Loros, A. M. Bilwes, J. C. Dunlap, B. R. Crane, Conformational switching in the fungal light sensor Vivid, Science 316 (5827) (2007) 1054–1057.

[237] B. D. Zoltowski, B. Vaccaro, B. R. Crane, Mechanism-based tuning of a LOV domain photoreceptor, Nat Chem Biol 5 (11) (2009) 827–834.
URL http://www.nature.com/nchembio/journal/v5/n11/full/nchembio.210.html

# Appendix

## A.1. Evaluation of bPAC Steady State Kinetics

Apparent catalytic rate constants of bPAC were subject to high variation. Comparison of rate constants determined from wild type bPAC (different preparations and different days, Table A.1) revealed minor systematic differences between preparations and very high inter-day variation as major sources of error.

**Table A.1.:** Apparent catalytic rate constants of bPAC wild type at 100 µM ATP. Measurements and fitting of initial velocities were carried out by Elisabeth Hartmann.

| Date of Experiment | Preparation | Rate Constant [$s^{-1}$] | |
|---|---|---|---|
| | | Dark | Light |
| 130912 | A | 0 | 1.71 |
| 130912 | A | 0 | 1.87 |
| 130920 | A | 0.006 | 1.88 |
| 130920 | A | 0.006 | 2.00 |
| 130920 | B | 0.006 | 2.94 |
| 130920 | B | 0.009 | 3.34 |
| 151209 | C | 0 | 0.61 |
| 151216 | C | 0 | 2.01 |
| 151216 | C | 0 | 1.95 |
| 160420 | A | 0 | 0.56 |
| 160420 | B | 0 | 0.48 |
| 160420 | C | 0 | 0.57 |

Given the large inter-day variation of catalytic rate constants together with very low intra-day variation, it appeared sensible to determine the rate constant of a standard sample (in this case, wild type bPAC) and to normalize all other rate constants determined on the same day with respect to the standard. To account for different fractions of bPAC

in preparations of variant protein, catalytic rate constants were adjusted by a coarse estimate of the bPAC fraction, determined by densitometric analysis of SDS-PAGE gels (Table A.2).

**Table A.2.:** Fractions of bPAC in the preparations used for measurement of steady state kinetics. bPAC was expressed and purified by Elisabeth Hartmann and could not be highly enriched. The fraction of bPAC in the preparation was coarsely determined by densitometric analysis of SDS-PAGE gels. The amount of bPAC loaded in each well of the gel was adjusted by the absorbance at 440 nm whenever available.

| Variant | Preparation | Fraction bPAC [%] |
|---|---|---|
| Y7F | A | 20 |
| I8V | A | 30 |
| I8V/I116V | A | 50 |
| I116V | A | 30 |
| K78/T115C | A | 30 |
| R121S | A | 60 |
| L123C | A | 50 |
| Y126F | A | 50 |
| P141G | A | 5 |
| N257A | A | 30 |
| N257K | A | 50 |
| 144–350 (AC) | A | 100 |
| wild type | A | 50 |
| wild type | B | n/a |
| wild type | C | 60 |

This correction cannot eliminate the large errors inherent to experimental determination of bPAC catalytic rate constants. Therefore, the activity of bPAC variants in the dark and in the light is not discussed in terms of catalytic rate constants but only in qualitative categories, each spanning a large range of relative rate constants: The categories "+++", "++", and "+" indicate activities of $\geq 50\,\%$, $\geq 10\,\%$, and $\geq 5\,\%$ of the corresponding wild-type sample in the light, respectively, and "–" indicates less than 5 % of wild-type light state activity (Table A.3).

**Table A.3.:** Data underlying qualitative bPAC activity assignments. As kinetic data were subject to batch and day-to-day variation, catalytic rate constants for each mutant were measured multiple times and expressed with respect to a wild type control sample measured under identical conditions on the same day. bPAC and some mutants in particular were difficult to enrich. Therefore, activities are corrected by an estimate of the fraction of bPAC. Given these sources of error, only qualitative statements about the specific activity of bPAC mutants can be made. Therefore, activity was grouped into the categories "+++","++", "+" that indicate activities of $\geq 50\,\%$, $\geq 10\,\%$ and $\geq 5\,\%$ of the corresponding wild type sample in the light,as well as "–", indicating less than 5 % of wild type light state activity.

| variant | wild type rate [$s^{-1}$] | | variant rate [$s^{-1}$] | | variant % | | category | |
|---|---|---|---|---|---|---|---|---|
| | dark | light | dark | light | dark | light | dark | light |
| **Y7F** | 0 | 0.61 | 0.06 | 0.132 | 9.8 | 21.6 | + | ++ |
| **Y7F** | 0.009 | 1.39 | 0.21 | 0.36 | 15.1 | 25.9 | ++ | ++ |
| **Y7F** | 0.012 | 1.088 | 0.176 | 0.169 | 16.2 | 15.5 | ++ | ++ |
| **Y7F** | 0.012 | 1.088 | 0.141 | 0.209 | 13 | 19.2 | ++ | ++ |
| **I8V** | 0.011 | 0.98 | 0.003 | 0.605 | 0.3 | 61.7 | – | +++ |
| **I8V** | 0.011 | 0.98 | 0.004 | 0.522 | 0.4 | 53.3 | – | +++ |
| **I116V** | 0.011 | 0.98 | 0.004 | 0.547 | 0.4 | 55.8 | – | +++ |
| **I116V** | 0.011 | 0.98 | 0.006 | 0.492 | 0.6 | 50.2 | – | +++ |
| **I8V/I116V** | 0.011 | 0.98 | 0.0066 | 0.625 | 0.7 | 63.8 | – | +++ |
| **I8V/I116V** | 0.011 | 0.98 | 0.01 | 0.685 | 1 | 69.9 | – | +++ |
| **R121S** | 0.011 | 0.98 | 0.18 | 0.64 | 18.8 | 65.8 | ++ | +++ |
| **R121S** | 0.012 | 1.088 | 0.23 | 0.52 | 21.4 | 47.9 | ++ | ++ |
| **R121S** | 0.009 | 1.39 | 0.2101 | 1.11 | 15.1 | 79.9 | ++ | +++ |
| **Y126F** | 0.006 | 1.939 | 0 | 0.024 | 0 | 1.2 | – | – |
| **P141G** | 0.004 | 1.46 | 0 | 0 | 0 | 0 | – | – |
| **N257A** | 0.004 | 1.792 | 0 | 0.249 | 0 | 13.9 | – | ++ |
| **N257A** | 0.006 | 1.939 | 0 | 0.21 | 0 | 10.8 | – | ++ |
| **N257K** | 0.006 | 1.939 | 0 | 0.774 | 0 | 39.9 | – | ++ |
| **Y126F/N257A** | 0.004 | 1.792 | 0 | 0.9005 | 0 | 50.3 | – | +++ |
| **Y126F/N257A** | 0.006 | 1.939 | 0 | 1.034 | 0 | 53.3 | – | +++ |
| **K78/T115C –DTE** | 0.0084 | 2.9559 | 0.3465 | 1.0362 | 11.7 | 35.1 | ++ | ++ |
| **K78/T115C –DTE** | 0.0084 | 2.9559 | 0.3465 | 1.2738 | 11.7 | 43.1 | ++ | ++ |
| **K78/T115C –DTE** | 0.014 | 1.882 | 0.4686 | 0.6666 | 24.9 | 35.4 | ++ | ++ |
| **K78/T115C –DTE** | 0.014 | 1.882 | 0.49 | 0.755 | 26 | 40.1 | ++ | ++ |
| **K78/T115C +DTE** | 0.014 | 1.882 | 0.1 | 0.95 | 5.3 | 50.5 | + | +++ |
| **K78/T115C –DTE** | 0.008 | 2.4 | 0.485 | 1.655 | 20.2 | 69 | ++ | +++ |
| **K78/T115C +DTE** | 0.08 | 2.36 | 0.25 | 2.1 | 10.6 | 89 | ++ | +++ |
| **L123C –DTE** | 0.014 | 1.882 | 0.075 | 2.995 | 4 | 159.1 | – | +++ |
| **L123C –DTE** | 0.008 | 2.36 | x | 1.4675 | x | 62.2 | x | +++ |
| **L123C +DTE** | 0.008 | 2.4 | x | 1.3525 | x | 56.4 | x | +++ |

## A.2. Mechanism of Signaling through the BLUF Methionine

The combination of functional and structural data obtained from bPAC and other BLUF domains allows tracing the chain of interactions from the flavin chromophore through Met93 to the BLUF domain surface This signaling pathway is structurally conserved in all investigated BLUF domains.

Met93 is likely positioned by interaction with Gln93 [12, 139] and resides in a conserved strained position, as evidenced by Ramachandran analysis. The sulfur atom of Met93 is in van der Waals distance of a backbone atom in $\beta2_{BLUF}$ (e.g., Val39 C$\alpha$ in bPAC) that in turn is in direct contact with a residue on the $\alpha3_{BLUF}$ capping helix (Pro109 in bPAC). The terminal methyl group of Met93 is in van der Waals distance of a conserved leucine (Leu51 in bPAC) on $\beta3_{BLUF}$. Upon formation of the lit state, the tautomeric and rotated Gln49 displaces the side chain f Met93 by up to 1 Å as observed in structures of illuminated AppA crystals [85]. This is supported by large chemical shift changes of Met93 in nuclear magnetic resonance (NMR) experiments on illuminated PixD and BlrP [231, 221, 220]. Displacement and relaxation of the conserved Methionine not only explains the chemical shift changes of residues corresponding to Leu51 and Val39 but also those of the interacting residues in $\alpha3_{BLUF}$. Conversely, mutations removing the methionine sidechain (e.g. M93A) do not affect the spectroscopically observed photocycle of bPAC and PixD but instead lock the proteins in functionally pseudolit conformations [128, 192, 231]. The same is true for removal of aliphatic sidechains in the vicinity of Met93: the variant L41V in $\beta2_{BLUF}$ of the light-regulated phosphodiesterase BlrP1 increases the dynamics of the $\alpha3_{BLUF}$ helix to an extent similar to illumination [220] and leads to constitutively elevated phosphodiesterase activity (Elisabeth Hartmann, personal communication). In conclusion, the strained and tightly packed conformation of the conserved methionine acts as light-triggered spring that pushes on residues in $\beta2_{BLUF}$ and $\beta3_{BLUF}$ that relay its impact to the N-terminal part of the $\alpha3_{BLUF}$ helix.

# A.3. Supplemental Tables

**Table A.4.:** Crystallographic data collection and refinement statistics for bPAC. Numbers in parentheses indicate the highest resolution shell.

| | *x1_dark* | *x2_dark* | *x3_dark* | *x4_dark* | *x1_light* | *x2_light* | *Y7F* | *ac* | *126F_257A* | *8V_116V* | *Y7F_atp* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Data collection** | | | | | | | 2 crystals | | | | |
| Space group | *P*1 | *P*1 | *P*1 | $P2_12_12_1$ | *P*1 | *P*1 | *C*2 | $C222_1$ | *C*2 | *P*1 | *C*2 |
| *a, b, c* (Å) | 48.3 56.6 78.1 | 47.7 56.2 78.0 | 48.0 56.7 78.1 | 57.4 96.7 163.6 | 47.7 56.7 77.7 | 47.5 56.3 77.5 | 68.4 143.3 57.3 | 50.7 101.8 80.3 | 65.8 142.6 56.9 | 47.7 57.1 71.3 | 69.6 143.4 57.5 |
| α, β, γ (°) | 104.3 104.2 102.2 | 104.1 103.9 102.2 | 104.2 104 102.5 | 90 90 90 | 104.7 103.9 102.7 | 105.8 103.7 102.7 | 90 135.5 90 | 90 90 90 | 90 134.6 90 | 82.4 75.8 76.9 | 90 135.9 90 |
| Resolution (Å) | 44.8 - 2.0 (2.071 - 2.0) | 44.35 - 1.8 (1.864 - 1.8) | 44.47 - 1.8 (1.864 - 1.8) | 49.37 - 3.1 (3.211 - 3.1) | 44.19 - 2.25 (2.33 - 2.25) | 43.96 - 2.4 (2.486 - 2.4) | 45.5 - 2.5 (2.589 - 2.5) | 45.4 - 2.3 (2.382 - 2.3) | 44.47 - 2.4 (2.486 - 2.4) | 44.98 - 2.3 (2.382 - 2.3) | 45.88 - 2.4 (2.486 - 2.4) |
| Unique Reflections | 48658 (4794) | 64706 (6097) | 65175 (6258) | 17193 (1714) | 33433 (3285) | 27155 (2710) | 13360 (1330) | 9537 (919) | 14534 (1421) | 30125 (3007) | 17359 (1529) |
| $R_{meas}$ (%) | 9.2 (59.2) | 9.6 (62.0) | 10.8 (61.6) | 25.1 (151.2) | 8.3 (64.6) | 8.9 (62.3) | 15.9 (107.1) | 5.5 (44.5) | 9.2 (113.5) | 8.3 (101.6) | 10.6 (77.7) |
| $CC_{1/2}$ (%) | 99.8 (87.8) | 99.5 (65.1) | 99.2 (65.0) | 99.5 (71.6) | 99.7 (76.5) | 99.7 (79.4) | 99.3 (82.2) | 99.9 (85.8) | 100 (87.7) | 99.7 (43.2) | 99.6 (79.2) |
| $I$ / σ$I$ | 18.3 (4.2) | 9.4 (1.6) | 10.0 (1.6) | 10.8 (2.1) | 11.4 (3.2) | 11.3 (3.1) | 13.7 (2.5) | 17.4 (3.4) | 12.8 (1.6) | 8.5 (1.1) | 10.6 (2.3) |
| Completeness (%) | 97 (96) | 96 (90) | 95 (91) | 100 (100) | 96 (95) | 97 (96) | 100 (100) | 100 (99) | 100 (99) | 96 (96) | 100 (100) |
| Redundancy | 3.5 (3.6) | 3.3 (1.7) | 2.7 (1.7) | 11.6 (9.5) | 2.9 (2.9) | 3.5 (3.6) | 10.1 (10.4) | 12.4 (12.7) | 4.7 (4.4) | 1.8 (1.8) | 5.2 (5.3) |
| **Refinement** | | | | | | | | | | | |
| Unique Reflections | 48654 (4790) | 64688 (6097) | 65171 (6258) | 17182 (1714) | 33432 (3285) | 27153 (2710) | 13356 (1329) | 9534 (918) | 14528 (1420) | 30124 (3007) | 15281 (1529) |
| $R_{work}$ / $R_{free}$ (%) | 21.1 / 24.3 | 16.9 / 21.5 | 19.0 / 22.5 | 19.8 / 23.1 | 19.0 / 21.7 | 19.6 / 23.1 | 19.1 / 25.7 | 24.0 / 29.3 | 20.7 / 25.1 | 20.3 / 25.4 | 20.9 / 26.6 |
| **No. atoms** | 5879 | 6257 | 6158 | 5549 | 5630 | 5606 | 2746 | 1536 | 2788 | 5590 | 2810 |
| Protein | 5445 | 5657 | 5578 | 5487 | 5440 | 5435 | 2698 | 1504 | 2723 | 5477 | 2710 |
| Ligands/Ions | 70 | 69 | 65 | 62 | 64 | 62 | 61 | 4 | 31 | 31 | 33 |
| Water | 364 | 531 | 515 | 0 | 126 | 109 | 17 | 28 | 38 | 82 | 66 |
| ***B* factors** | 33.7 | 36.83 | 36.43 | 79.5 | 51.74 | 68.95 | 74.7 | 63.5 | 69.0 | 68.36 | 63.56 |
| Protein | 33.32 | 35.9 | 35.61 | 79.4 | 51.56 | 68.96 | 74.6 | 63.5 | 68.8 | 68.44 | 63.96 |
| Ligands/Ions | 52.13 | 45.91 | 49.59 | 87.2 | 77.36 | 99.79 | 91.7 | 102.0 | 103.5 | 94.38 | 53.42 |
| Water | 35.86 | 45.57 | 43.58 | n/a | 46.56 | 50.96 | 60 | 57.8 | 51.9 | 53.25 | 52.06 |
| **r.m.s. deviations** | | | | | | | | | | | |
| Bond lengths (Å) | 0.008 | 0.006 | 0.002 | 0.003 | 0.004 | 0.002 | 0.008 | 0.001 | 0.01 | 0.02 | 0.01 |
| Bond angles (°) | 0.9 | 0.74 | 0.45 | 0.56 | 0.7 | 0.45 | 0.96 | 0.37 | 1.23 | 1.54 | 1.36 |
| **Ramachandran Plot** | | | | | | | | | | | |
| Favored (%) | 98.4 | 99.3 | 99 | 98.1 | 98.4 | 98.5 | 97.6 | 99.5 | 96.1 | 96.5 | 94.6 |
| Allowed (%) | 1.6 | 0.7 | 1 | 1.9 | 1.6 | 1.5 | 2.4 | 0.5 | 3.9 | 3.2 | 5.1 |
| Outliers (%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.3 | 0.3 |
| **PDB Code** | 5M27 | 5M2A | 5MBC | 5MBB | 5MBD | 5MBE | 5NBY | 5MBG | 5MBH | 5MBJ | 5MBK |

**Table A.5.:** Crystallographic data collection and refinement statistics for $mPAC_{AC}$ and $mPAC_{AC+Handle}$. Friedel pairs were not merged in the processing of SeMet derivatives, hence the larger number of unique reflections in these crystals. Numbers in parentheses indicate the highest resolution shell.

| | *AC dimer (SeMet)* | *AC dimer (native)* | *AC+Handle dimer (native)* | *AC trimer (SeMet)* |
|---|---|---|---|---|
| **Data collection** | | | | |
| Wavelength (Å) | 0.978 | 1.000 | 1.000 | 0.979 |
| Space group | $C222_1$ | $C222_1$ | $C222_1$ | $P2_12_12_1$ |
| *a, b, c* (Å) | 79.0 80.5 265.4 | 78.1 78.9.263.5 | 79.5 79.6 264.4 | 79.1 86.6 87.5 |
| α, β, γ (°) | 90.0 90.0 90.0 | 90.0 90.0 90.0 | 90.0 90.0 90.0 | 90 90 90 |
| Resolution (Å) | 47.6 - 2.6 (2.80 - 2.70) | 43.9 - 2.5 (2.59 - 2.50) | 47.4 - 2.7 (2.80 – 2.70) | 48.6 - 2.0 (2.08 - 2.00) |
| Unique Reflections | 44432 (4119) | 28496 (2790) | 23495 (2308) | 40952 (3947) |
| $R_{meas}$ (%) | 14.6 (105.4) | 21.5 (103.5) | 24.3 (132.7) | 15.2 (97.7) |
| $CC_{1/2}$ (%) | 99.5 (55.4) | 98.5 (50.5) | 99.1 (46.2) | 99.9 (47.6) |
| $I$ / σ$I$ | 9.2. (1.2) | 5.3 (1.4) | 7.7 (1.7) | 18.0 (1.5) |
| Completeness (%) | 99 (92) | 99 (99) | 100 (100) | 100 (98) |
| Redundancy | 6.1 (3.4) | 5.3 (5.8) | 7.0 (7.2) | 29.6 (6.9) |
| **Refinement** | | | | |
| Unique Reflections | 44421 (4118) | 28495 (2790) | 23495 (2308) | 40951 (3948) |
| $R_{work}$ / $R_{free}$ (%) | 23.2 / 25.4 | 23.7 / 26.8 | 26.9 / 28.5 | 19.6 / 22.7 |
| **No. atoms** | 5702 | 5690 | 5459 | 4462 |
| Protein | 5672 | 5676 | 5453 | 4086 |
| Ligands/Ions | 0 | 1 | 0 | 0 |
| Water | 30 | 13 | 6 | 376 |
| ***B* factors** | 65.6 | 63.3 | 59.66 | 38.37 |
| Protein | 65.7 | 63.3 | 59.69 | 37.78 |
| Ligands/Ions | n/a | 82.4 | n/a | n/a |
| Water | 49.9 | 56.7 | 39.1 | 44.68 |
| **r.m.s. deviations** | | | | |
| Bond lengths (Å) | 0.002 | 0.006 | 0.005 | 0.004 |
| Bond angles (°) | 0.49 | 1.02 | 0.79 | 0.54 |
| **Ramachandran Plot** | | | | |
| Favored (%) | 98.1 | 98.8 | 98.3 | 99.0 |
| Allowed (%) | 1.9 | 1.2 | 1.7 | 1.0 |
| Outliers (%) | 0 | 0 | 0 | 0 |

**Table A.6.:** AC domain opening angles. The angle α is defined as the angle between the Cα atoms of the spanning residues A, HingeA and B. β is defined as the angle between A and HingeB and B.

| Protein | | | Angle [°] | | Spanning Residues | | | | Ref |
|---|---|---|---|---|---|---|---|---|---|
| **Description** | **PDB** | **SG** | **α** | **β** | **A** | **HingeA** | **HingeB** | **B** | |
| **bPAC** | | | | | | | | | |
| x1_dark | 5M27 | $P1$ | 27.0 | 28.5 | 277 (A) | 146 (A) | 146 (B) | 277 (B) | this work |
| x2_dark | 5M2A | $P1$ | 26.9 | 28.4 | 277 (A) | 146 (A) | 146 (B) | 277 (B) | this work |
| x3_dark | 5MBC | $P1$ | 26.9 | 28.4 | 277 (A) | 146 (A) | 146 (B) | 277 (B) | this work |
| x4_dark | 5MBB | $P2_12_12_1$ | 27.6 | 28.6 | 277 (A) | 146 (A) | 146 (B) | 277 (B) | this work |
| x1_light | 5MBD | $P1$ | 28.1 | 29.0 | 277 (A) | 146 (A) | 146 (B) | 277 (B) | this work |
| x2_light | 5MBE | $P1$ | 28.7 | 28.8 | 277 (A) | 146 (A) | 146 (B) | 277 (B) | this work |
| Y7F | 5NBY | $C2$ | 32.6 | 32.6 | 277 (A) | 146 (A) | 146 (A) | 277 (A) | this work |
| AC | 5MBG | $C222_1$ | 28.4 | 28.4 | 277 (A) | 146 (A) | 146 (B) | 277 (B) | this work |
| Y126F/N257A | 5MBH | $C2$ | 29.5 | 29.5 | 277 (A) | 146 (A) | 146 (A) | 277 (A) | this work |
| S27A | 5MBI | $P1$ | 29.0 | 30.4 | 277 (A) | 146 (A) | 146 (B) | 277 (B) | this work |
| I8V/I116V | 5MBJ | $P1$ | 31.5 | 31.9 | 277 (A) | 146 (A) | 146 (B) | 277 (B) | this work |
| Y7F/ATP | 5MBK | $C2$ | 32.8 | 32.8 | 277 (A) | 146 (A) | 146 (A) | 277 (A) | this work |
| **OaPAC** | | | | | | | | | |
| hexagonal | 4YUS | $P6_122$ | 30.9 | 30.9 | 276 (A) | 145 (A) | 145 (A) | 276 (A) | [149] |
| orthorhombic | 4YUT | $P2_12_12_1$ | 27.1 | 27.5 | 276 (A) | 145 (A) | 145 (B) | 276 (B) | [149] |
| **mammalian tmAC (VC1/IIC2 chimera), Gsα complex** | | | | | | | | | |
| apo | 1AZS | $P2_12_12$ | 42.3 | 38.3 | 515 (A) | 384 (A) | 879 (B) | 1028 (B) | [203] |
| ATP-αS | 1CJK | $P2_12_12$ | 40.5 | 38.2 | 515 (A) | 384 (A) | 879 (B) | 1028 (B) | [204] |
| MANT-GTP | 1U0H | $P2_12_12$ | 42.5 | 37.8 | 515 (A) | 384 (A) | 879 (B) | 1028 (B) | [138] |
| **pH-regulated Rv1264** | | | | | | | | | |
| active state[a] | 1Y11 | $P6_222$ | 27.0 | 27.0 | 322 (A) | 212 (A) | 212 (B) | 322 (B) | [205] |
| **human soluble AC** | | | | | | | | | |
| apo | 4CLF | $P6_3$ | 38.4 | 40.1 | 145 | 36 | 287 | 176 | [99] |
| ApCpp | 4CLK | $P6_3$ | 37.3 | 39.1 | 145 | 36 | 287 | 176 | [99] |
| bicarbonate | 4CLL | $P6_3$ | 38.3 | 40.5 | 145 | 36 | 287 | 176 | [99] |
| apo (cAMP soak) | 4CLP | $P6_3$ | 36.7 | 39.4 | 145 | 36 | 287 | 176 | [99] |
| PPi soak | 4CLS | $P6_3$ | 37.3 | 39.6 | 145 | 36 | 287 | 176 | [99] |
| apo (cAMP/PPi) | 4CLT | $P6_3$ | 36.9 | 39.8 | 145 | 36 | 287 | 176 | [99] |
| $Mg^{2+}$/ATP | 4CLU | $P6_3$ | 36.6 | 39.8 | 145 | 36 | 287 | 176 | [99] |
| bisulfite/ApCpp | 4CLW | $P6_3$ | 36.6 | 38.7 | 145 | 36 | 287 | 176 | [99] |
| biselenite | 4CLY | $P6_3$ | 37.8 | 39.3 | 145 | 36 | 287 | 176 | [99] |
| stilbene inhibitor | 4CLZ | $P6_3$ | 37.5 | 39.4 | 145 | 36 | 287 | 176 | [99] |
| bicarbonate/ApCpp | 4CM0 | $P6_3$ | 36.5 | 38.3 | 145 | 36 | 287 | 176 | [99] |
| bisulfite | 4CM2 | $P6_3$ | 38.2 | 39.9 | 145 | 36 | 287 | 176 | [99] |

[a] The N-terminal part of the Rv1264 AC domain differs from the other investigated ACs; hence the spanning residues chosen for comparison to the other values may not be ideal to capture the AC domain angle.

**Table A.7.:** Domain assembly problems used in this work.

| PDB | Chain | Linker | Sequence |
|---|---|---|---|
| **Selected in this work** | | | |
| 1A2O | A | 140–158 | PMAAPTTLKAGPLLSSEKL |
| 1B24 | A | 90–99 | ERIRLFNMRE |
| 1B9K | A | 117–125 | LPITLNKFF |
| 1E42 | A | 115–125 | LIPLNVLFVED |
| 1GQE | A | 115–125 | MFSGEYDSADC |
| 1H19 | A | 450–462 | SPGLPPIKPNYDM |
| 1HP1 | A | 323–329 | KAQLEVK |
| 1HT6 | A | 346–354 | NGITATSAL |
| 1IPA | A | 100–112 | RTLEEYRPSPDAL |
| 1J18 | A | 410–419 | KFKDLLGVTP |
| 1J5Y | A | 61–69 | VLAGGKSGV |
| 1JAL | A | 260–264 | ALLNL |
| 1KGS | A | 119–127 | RKSESKSTK |
| 1KL9 | A | 71–78 | SKRRVSPE |
| 1KNR | A | 410–425 | HDISTLPPWDESRVEN |
| 1L2F | A | 197–202 | SRRVPE |
| 1MIX | A | 111–117 | LKTYGVS |
| 1NE4 | A | 154–162 | RLTVADALE |
| 1NIJ | A | 195–217 | FNTNGFMLEENVVSTKPRFHFIA |
| 1O70 | A | 132–142 | IDHVLGVPYTT |
| 1OXS | C | 240–245 | LIGEIN |
| 1P2F | A | 114–122 | FLEREKKGL |
| 1Q79 | A | 340–346 | LFEAPNF |
| 1QCS | A | 84–93 | YSFDKAKQCI |
| 1QPK | A | 356–362 | GVRADSA |
| 1S2M | A | 203–213 | NLMEELTLKGI |
| 1TJE | A | 60–68 | ITAKDEEGL |
| 1TKY | A | 60–65 | ITAKDE |
| 1TUA | A | 82–87 | LLEEDQ |
| 1UMK | A | 115–119 | PSGLL |
| 1VAH | A | 399–407 | NVVDGEPFA |
| 1W3S | A | 71–79 | GALPTLAEP |
| 1W7K | A | 274–281 | AILPGRFQ |
| 1W9Q | A | 81–87 | RDRPFER |
| 1WLF | A | 81–89 | PCSHVVSCQ |
| 1WMF | A | 313–318 | SLNVAY |
| 1WPC | A | 387–397 | EARQKYAYGKQ |
| 1WZA | A | 406–411 | VFYTGK |

*(continued on next page)*

**Table A.7** *(continued)*

| PDB | Chain | Linker | Sequence |
|---|---|---|---|
| 1XTI | A | 205–218 | FVDDETKLTLHGLQ |
| 1YJQ | A | 165–170 | WHNNIR |
| 1YW5 | A | 55–68 | NNGYKPLVNEDGQV |
| 1Z2M | A | 73–80 | VDKSDEPL |
| 1ZAR | A | 86–92 | RSGKVDA |
| 2CVE | A | 118–126 | APKVPLVER |
| 2CXC | A | 67–72 | EYSSDL |
| 2D0A | A | 78–91 | QPSALAAHQLGSLS |
| 2D0H | A | 120–128 | IIPNFKTPD |
| 2DYI | A | 77–90 | EVADLPPLEEGRYY |
| 2E3U | A | 80–86 | LLNEGEY |
| 2FH7 | A | 285–294 | CGNTEVPARS |
| 2GC6 | A | 290–298 | ASHWPARLE |
| 2GJK | A | 398–406 | GIRPIRLQV |
| 2GPJ | A | 99–109 | LKKLINFEADW |
| 2GU1 | A | 184–225 | DRAGNSLERAFNRYPVDKAYRQITSGFNPKRKHPVTGRVVPH |
| 2GWR | A | 119–129 | LRRNDDEPAEM |
| 2HHP | A | 346–351 | FEKNDF |
| 2I1L | A | 152–160 | PAYLGRYFE |
| 2I1Q | A | 55–65 | LCDLGFKSGID |
| 2NLK | A | 291–299 | GKETEVSSN |
| 2NSM | A | 308–314 | EQVHQGI |
| 2PJD | A | 144–158 | RLEKQPVFDAEKFWG |
| 2QIK | A | 114–120 | SKSKNPI |
| 2QYT | A | 178–187 | AGIRAYNPTD |
| 2R2P | A | 93–99 | EYMENGS |
| 2V2U | A | 81–88 | IPHAGSHR |
| 2V9V | A | 60–64 | STERY |
| 2VAM | A | 207–212 | MSNKGS |
| 2VAP | A | 209–228 | KDGLINVDFADVKAVMNNGG |
| 2WKP | A | 137–144 | DEAAKELI |
| 2WN8 | A | 187–198 | VSSLDFKDDVSK |
| 2WXU | A | 237–248 | GNDPSVGKNVKE |
| 2XQ0 | A | 447–462 | WLYKPGMPPRPHFITA |
| 2Y68 | A | 290–299 | TFTGLPHRFE |
| 2Z0M | A | 189–192 | EACI |
| 2Z0Q | A | 190–200 | RLLYLEEGQKD |
| 2Z1A | A | 297–306 | QPVMALMQQV |
| 2Z1E | A | 117–129 | EHPVSDAGAKVGD |

*(continued on next page)*

**Table A.7** *(continued)*

| PDB | Chain | Linker | Sequence |
|---|---|---|---|
| 2Z8L | A | 84–98 | KKNQGAYYDYLNAPK |
| 2ZXR | A | 486–498 | RKPEALEGGIAPF |
| 3A3J | A | 248–253 | WGFANF |
| 3A7R | A | 245–251 | FGQAPAF |
| 3AB7 | A | 121–130 | APGEPVELEG |
| 3AII | A | 299–304 | LEEEAR |
| 3AXC | A | 71–78 | LRLRGGMQ |
| 3B7U | X | 278–289 | VTPTLLAGDKSL |
| 3BIQ | A | 166–172 | VWEVKDV |
| 3C7A | A | 195–200 | LAKHFL |
| 3CB6 | A | 174–181 | CLAIKDEQ |
| 3CHS | A | 451–464 | PGLPPIKPNYDMTL |
| 3CSG | A | 365–375 | KGSSVPTNLEV |
| 3DAY | A | 429–434 | RADDII |
| 3DHV | A | 391–395 | RLDFQ |
| 3EF6 | A | 304–316 | ILGKNVSAPQLPV |
| 3ETL | A | 59–67 | CDLGFKSGV |
| 3FEH | A | 253–260 | LSRNYLKE |
| 3FG2 | P | 307–319 | LTGDAKPYDGYPW |
| 3GHY | A | 194–200 | CSEAIQR |
| 3GN7 | A | 119–127 | GGSDDSSKD |
| 3GSL | A | 92–97 | RKPPAE |
| 3HLZ | A | 130–141 | EARKEGEKYPAE |
| 3HP7 | A | 56–67 | LKGEKLRYVSRG |
| 3I2D | A | 155–158 | MITP |
| 3IDV | A | 116–127 | QPDWTPPPEVTL |
| 3IVE | A | 301–306 | EVVQQT |
| 3LK7 | A | 304–312 | FGGVKHRLQ |
| 3LNU | A | 180–185 | YFENGL |
| 3M2W | A | 167–173 | ETTGEKY |
| 3M70 | A | 94–108 | KKEDYFSKKYNTTAI |
| 3OF1 | A | 145–149 | ADALD |
| 3P02 | A | 145–151 | LFNEFSG |
| 3P7N | A | 117–118 | VDDDQPNMG |
| 3PXA | A | 107–116 | SQDRKIFRGL |
| 3PYD | A | 158–164 | VLSRNTF |
| 3PYG | A | 159–165 | VLSRNTF |
| 3Q6B | A | 77–83 | DAGNRFY |
| 3QDF | A | 56–64 | LAPILASKV |

*(continued on next page)*

**Table A.7** *(continued)*

| PDB | Chain | Linker | Sequence |
|---|---|---|---|
| 3QFK | A | 297–303 | EDWLDYE |
| 3R44 | A | 396–402 | DRLKDMI |
| 3R8X | A | 200–209 | AEKLSKEEAK |
| 3RBV | A | 121–128 | NFMFLHHP |
| 3RGA | A | 121–134 | MWGVTDSSWTARPA |
| 3RKX | A | 60–66 | QLPDIWY |
| 3SR9 | A | 284–290 | GCGNTEV |
| 3SUW | A | 157–163 | DKPEYEY |
| 3TFL | A | 103–116 | TKKVEAPKTSKNQS |
| 3TMA | A | 150–156 | QLTERPL |
| 3TWL | A | 128–137 | LGPDALLEPM |
| 3V0F | A | 184–196 | KNYGGQLPPMKKL |
| 3V0I | A | 146–158 | KKNYGGQLPPMKK |
| 3V0J | A | 173–181 | YGGQLPPMK |
| 3V8J | A | 75–83 | TKSSALFDF |
| 3W5H | A | 117–121 | NGLLV |
| 3WGJ | A | 192–211 | TPGLINLDFADVKTIMSNQG |
| 3WHI | A | 64–68 | EEDHI |
| 4A3P | A | 118–123 | VEVYLT |
| 4AIE | A | 458–466 | ENPIVVDGD |
| 4AVA | A | 140–146 | AFVSPIP |
| 4B3F | X | 420–427 | LTVQYRMH |
| 4BMC | A | 94–102 | HLLPTLFKC |
| 4BPC | A | 283–288 | AGNTEV |
| 4DE8 | A | 101–115 | FTKKVEAPKTSKNQS |
| 4EMD | A | 158–164 | TVLARNV |
| 4GER | A | 144–151 | YTSNLEYY |
| 4GYI | A | 91–96 | RKDVYS |
| 4IDZ | A | 256–264 | RVAECSTGT |
| 4JVA | A | 150–164 | KSLEVSERLKVVDVI |
| 4K91 | A | 248–252 | FRFFE |
| 4KFV | A | 91–101 | CSFRRASEHVW |
| 4KHO | A | 177–181 | TLEDV |
| 4L6S | A | 135–139 | EKLKT |
| 4LAV | A | 124–134 | GEDLTEEEDGG |
| 4LII | A | 345–357 | MTGAAKPYWHQSM |
| 4M8I | A | 194–212 | SGEVNLDFADVKTIMSNQG |
| 4MEM | A | 116–121 | VEVYPI |
| 4ORF | A | 137–142 | AFITPI |

*(continued on next page)*

**Table A.7** *(continued)*

| PDB | Chain | Linker | Sequence |
|---|---|---|---|
| 4PQ0 | A | 141–152 | PGEYKSKYDQCA |
| 4PZV | A | 163–170 | LKQTLANT |
| 8TLN | E | 150–160 | DYTAGLIYQNE |
| **Wollacott/Zanghellini Set [218]** | | | |
| 1A62 | A | 45–50 | SGEDIF |
| 1A6Q | A | 289–298 | CFPNAPKVSP |
| 1A79 | C | 73–77 | VEERL |
| 1A8D | A | 245–265 | YLSITFLRDFWGNPLRYDTEY |
| 1AMM | A | 80–89 | LIPQHTGTFR |
| 1AVA | B | 343–352 | RHGIHNESKL |
| 1B63 | A | 218–225 | CGTAFLEQ |
| 1BAG | A | 344–350 | VMAGQPE |
| 1BG6 | A | 175–195 | SVLPQYVAVENVLHTSLTNVN |
| 1BI5 | A | 223–235 | VGSDPVPEIEKPI |
| 1BKB | A | 69–75 | EVPIIEK |
| 1BU6 | O | 251–258 | QLCVKEGM |
| 1CA1 | A | 245–257 | EGNDPSVGKNVKE |
| 1CJX | B | 140–156 | YLEGVERNPVGAGLKVI |
| 1CLC | A | 94–106 | KSVNFKIAMNVYE |
| 1CLI | B | 148–158 | EIIDGSKVSDG |
| 1CRZ | A | 132–138 | IKGAFRT |
| 1CTU | A | 170–189 | LEIKTLLMDEQDHGYALTGD |
| 1CVR | A | 337–357 | TVFGDPSLLVRTLVPTEMQVT |
| 1CX4 | A | 138–151 | SFIESLPFLKSLEV |
| 1D09 | B | 95–105 | SRPSLPERIDN |
| 1D5R | A | 171–179 | NHLDYRPVA |
| 1DZF | A | 132–148 | NEAALVVNITHHELVPK |
| 1EGA | B | 168–187 | KHLPEATHHFPEDYITDRSQ |
| 1EOV | A | 128–137 | SETPEALPIL |
| 1EV7 | A | 148–168 | AQVVKLWPDHGKLQENLLLHI |
| 1F1Z | A | 148–167 | QKQIPWFIFTDKEINPVVKE |
| 1F3A | A | 77–85 | KYDLYGKDM |
| 1F5N | A | 263–277 | NSKTKTLSGGIQVNG |
| 1FMT | A | 195–214 | VTYAEKLSKEEARIDWSLSA |
| 1FTS | A | 81–94 | KVDEPLNVEGKAPF |
| 1GCY | A | 352–359 | AGVRADSA |
| 1GV1 | B | 123–135 | SGLPKERVIGMAG |
| 1HAN | A | 123–142 | FEKPFLPGAAVSGFLTGEQG |
| 1HCL | A | 77–83 | EFLHQDL |

*(continued on next page)*

**Table A.7** *(continued)*

| PDB | Chain | Linker | Sequence |
|---|---|---|---|
| 1HYE | A | 145–148 | GLGT |
| 1I39 | A | 151–170 | ERLKEKFGDFGSGYASDPRT |
| 1I8D | B | 85–104 | ERAAKFSDEIGGHLMSGHIM |
| 1J8M | F | 79–98 | SNLFGGDKEPKVIPDKIPYV |
| 1JGT | B | 203–216 | RTWTPGLSRRILPE |
| 1JPM | A | 114–134 | MCGLPLYQMLGGYRDTLETDY |
| 1JPN | A | 83–102 | KEALGGEARLPVLKDRNLWF |
| 1K0M | A | 83–103 | LCPPRYPKLAALNPESNTAGL |
| 1KBW | C | 143–153 | EPKEGLPKVDK |
| 1KNY | A | 124–128 | KSVEA |
| 1KS9 | A | 159–178 | VLPDVAWHNNIRAELWRKLA |
| 1LAM | A | 144–163 | KLHGSEDQEAWQRGVLFASG |
| 1PII | A | 247–258 | AAVRRVLLGENK |
| 1PSZ | A | 144–164 | EFYEKNLKEYTDKLDKLDKES |
| 1QAM | A | 169–179 | RKKSRISHKDK |
| 1QFJ | C | 84–97 | IPHGEAWLRDDEER |
| 1QH4 | B | 97–116 | GGYKPTDEHKTDLNADNLQG |
| 1QTO | A | 52–63 | RTEHQIVADNTS |
| 1REC | A | 78–86 | HMTSAGKTN |
| 1SEI | B | 61–80 | IFLKYGPNERVITGLKRISK |
| 1SVP | A | 62–72 | QLPVNMRSEAF |
| 2REB | A | 238–245 | LYGEGINF |
| 4PGA | A | 194–214 | RLPAKRHTVNSEFDIKQISSL |

# A.4. Supplemental Figures

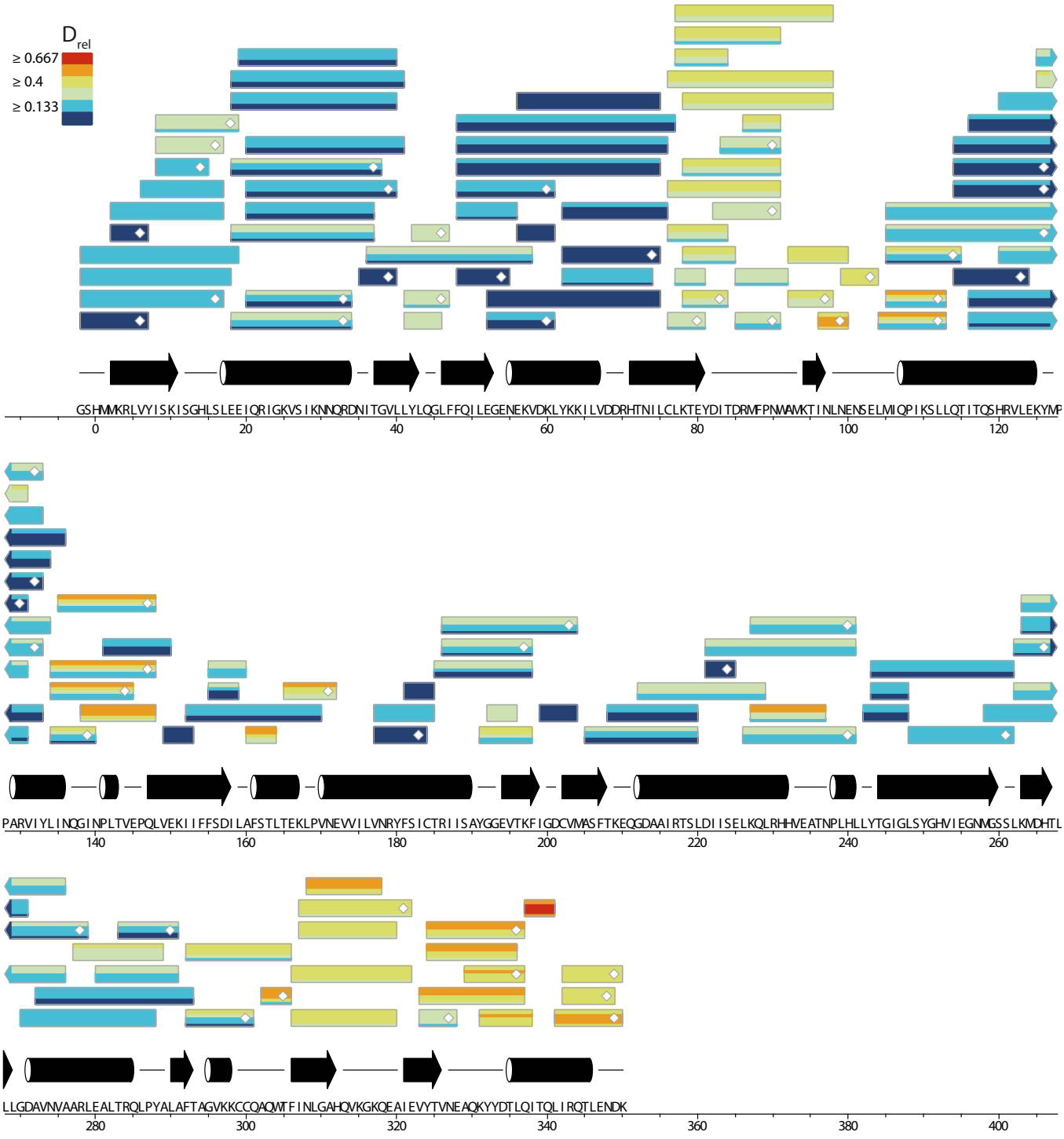



**Figure A.1.:** Relative deuteration of dark-adapted bPAC. Each box indicates one peptide and contains up to five different colors for incubation times of 10, 45, 180, 900 and 3600 s, respectively, bottom-up. MS/MS confirmed peptides are marked with diamonds. Secondary structure is shown as determined from the crystal structure using DSSP [87]. Numbering follows the wild type sequence (Uniprot A7BT71).

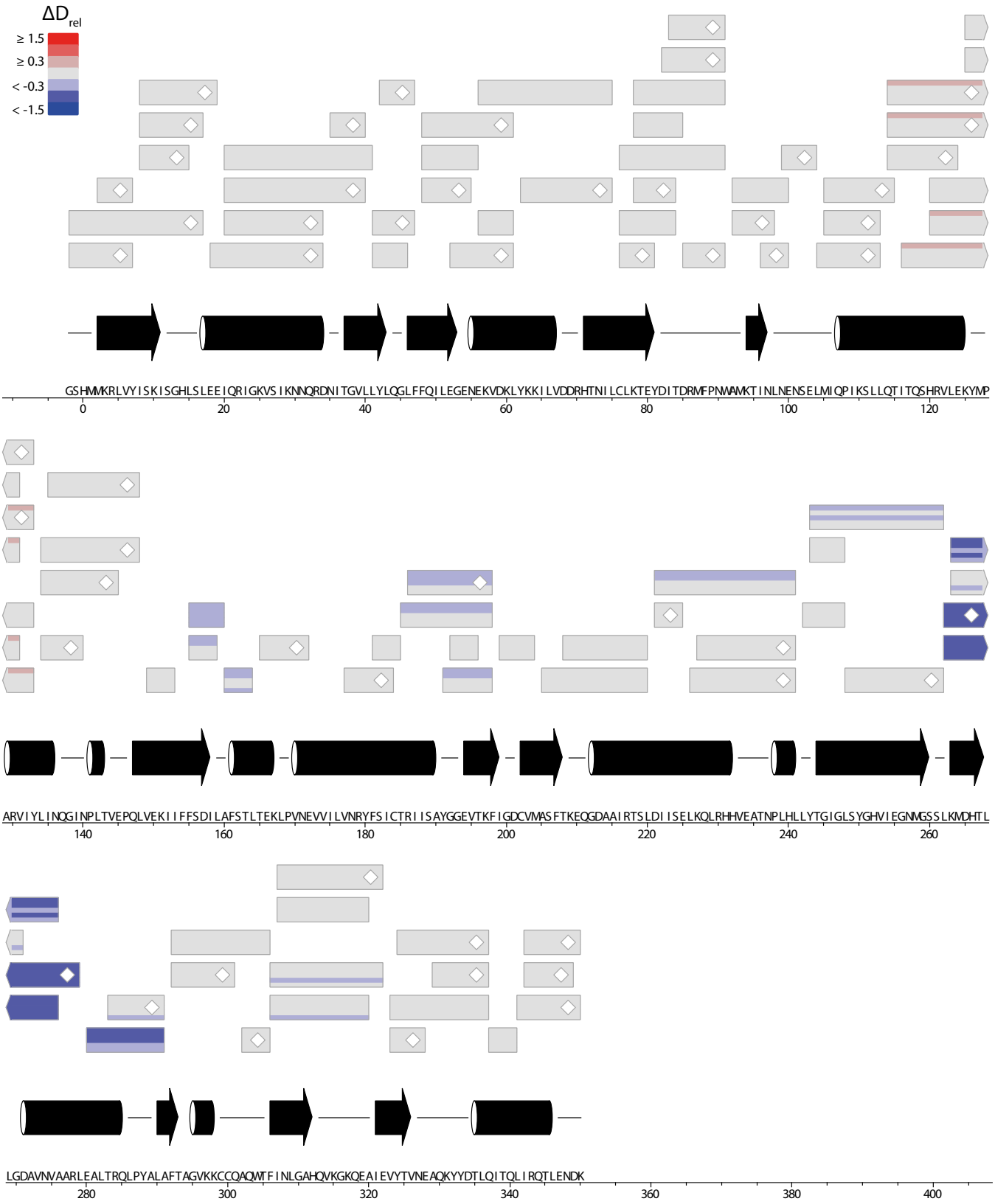



**Figure A.2.:** Differences in relative deuteration of dark-adapted bPAC in the presence and absence of ATP-αS (Rp). Each box indicates one peptide and contains up to five different colors for incubation times of 10, 45, 180, 900 and 3600 s, respectively, bottom-up. MS/MS confirmed peptides are marked with diamonds. Secondary structure is shown as determined from the crystal structure using DSSP [87]. Numbering follows the wild type sequence (Uniprot A7BT71).

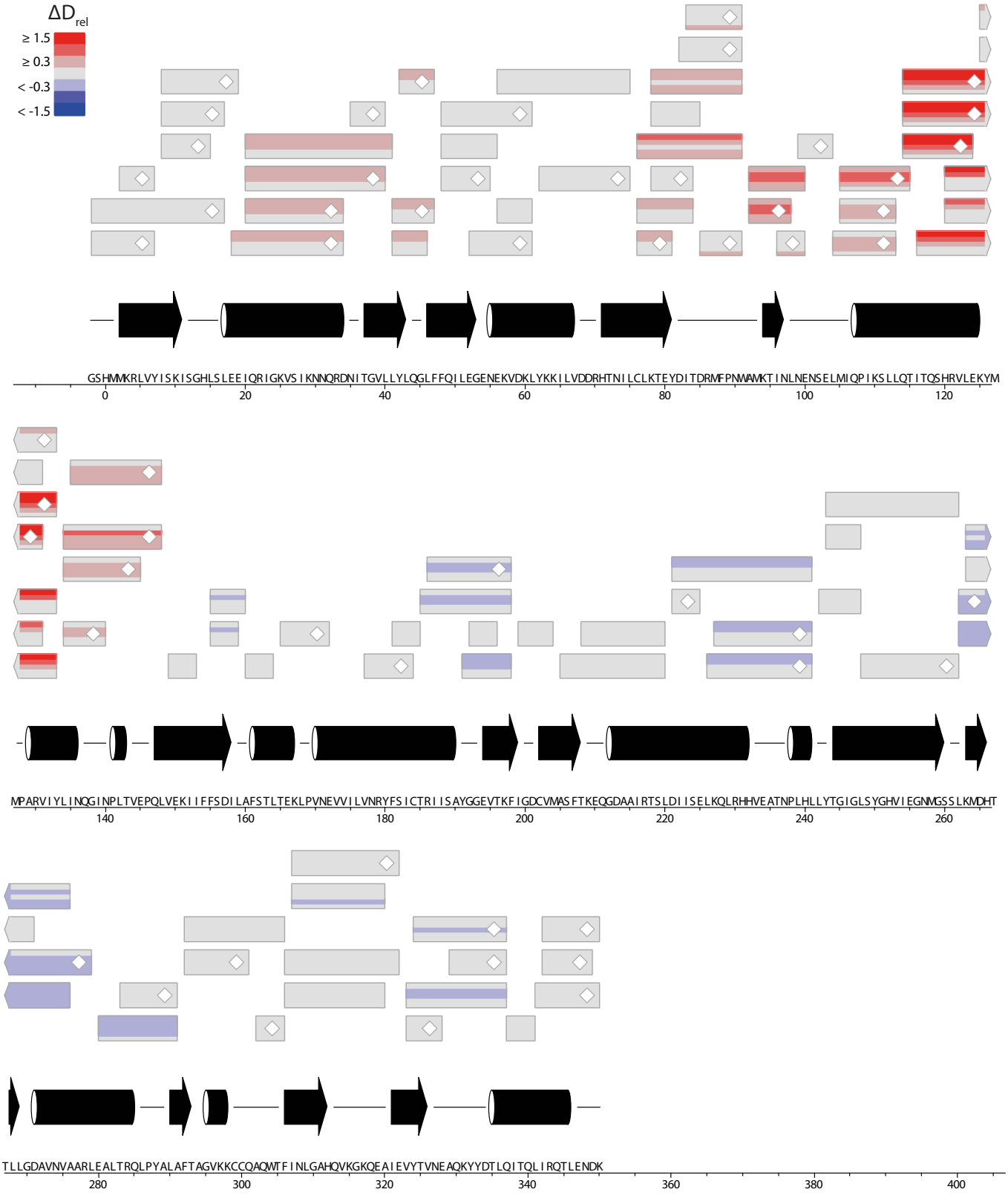



**Figure A.3.:** Differences in relative deuteration of dark- and light-adapted bPAC. Each box indicates one peptide and contains up to five different colors for incubation times of 10, 45, 180, 900 and 3600 s, respectively, bottom-up. MS/MS confirmed peptides are marked with diamonds. Secondary structure is shown as determined from the crystal structure using DSSP [87]. Numbering follows the wild type sequence (Uniprot A7BT71).

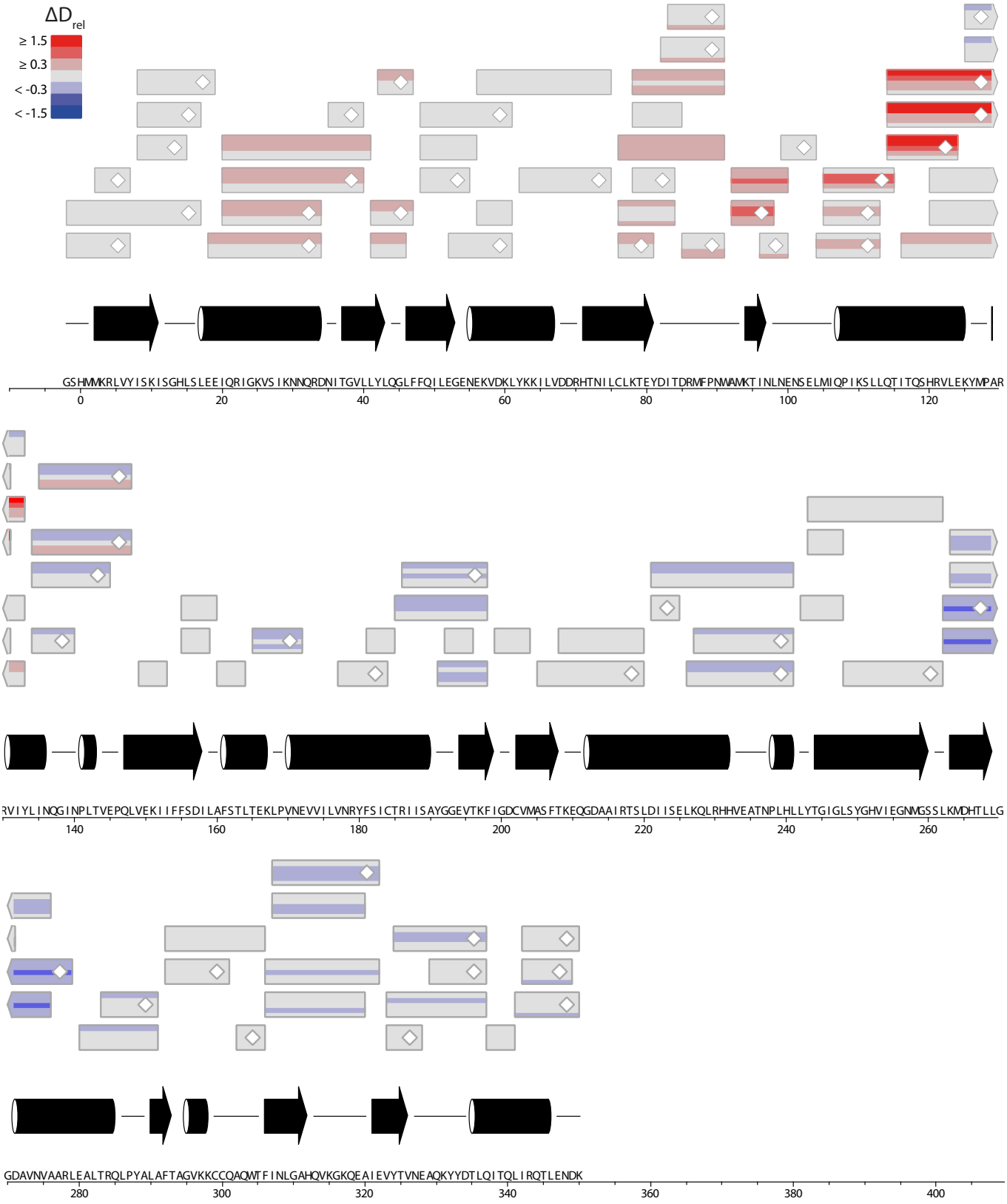



**Figure A.4.:** Differences in relative deuteration of dark- and light-adapted bPAC in the presence of ATP-αS (Rp). Each box indicates one peptide and contains up to five different colors for incubation times of 10, 45, 180, 900 and 3600 s, respectively, bottom-up. MS/MS confirmed peptides are marked with diamonds. Secondary structure is shown as determined from the crystal structure using DSSP [87]. Numbering follows the wild type sequence (Uniprot A7BT71).

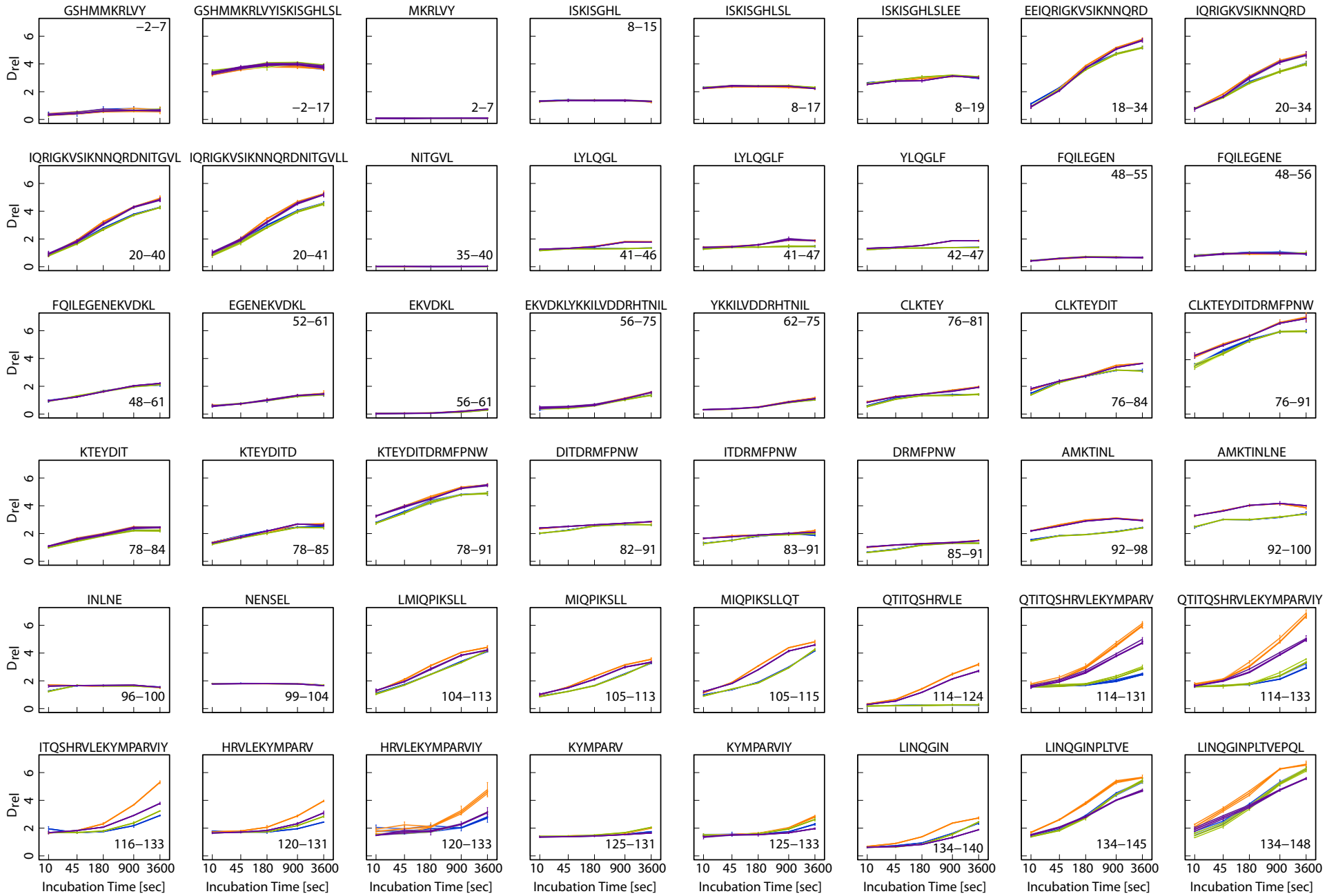



**Figure A.5.:**

*(continued on next page)*

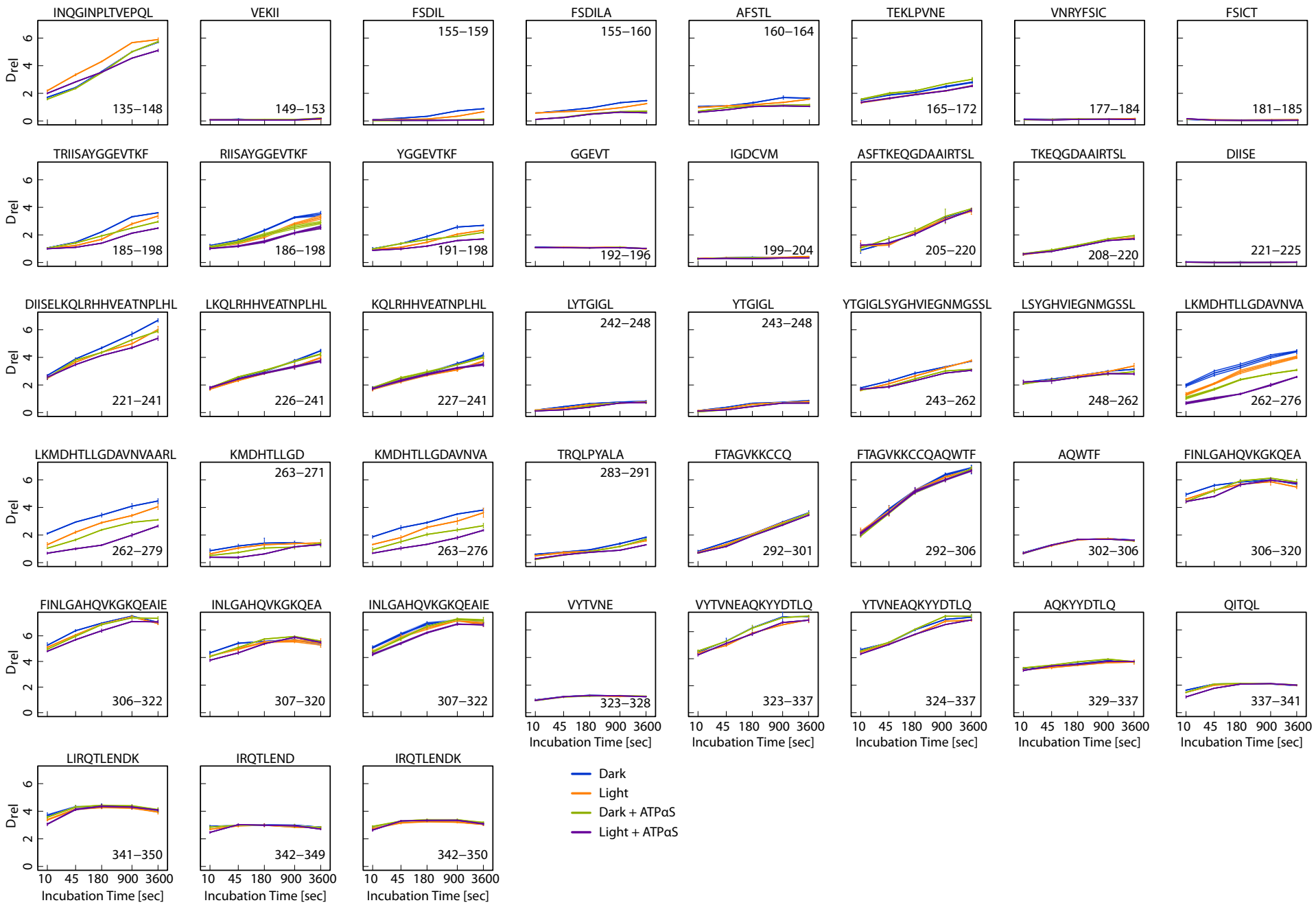



**Figure A.5.:** Deuterium incorporation time series of each peptide derived from bPAC. The temporal evolution of relative deuteration in four measured states is shown: blue – dark; green – dark in the presence of ATP-αS (Rp); orange – light; purple – light in the presence of ATP-αS (Rp).

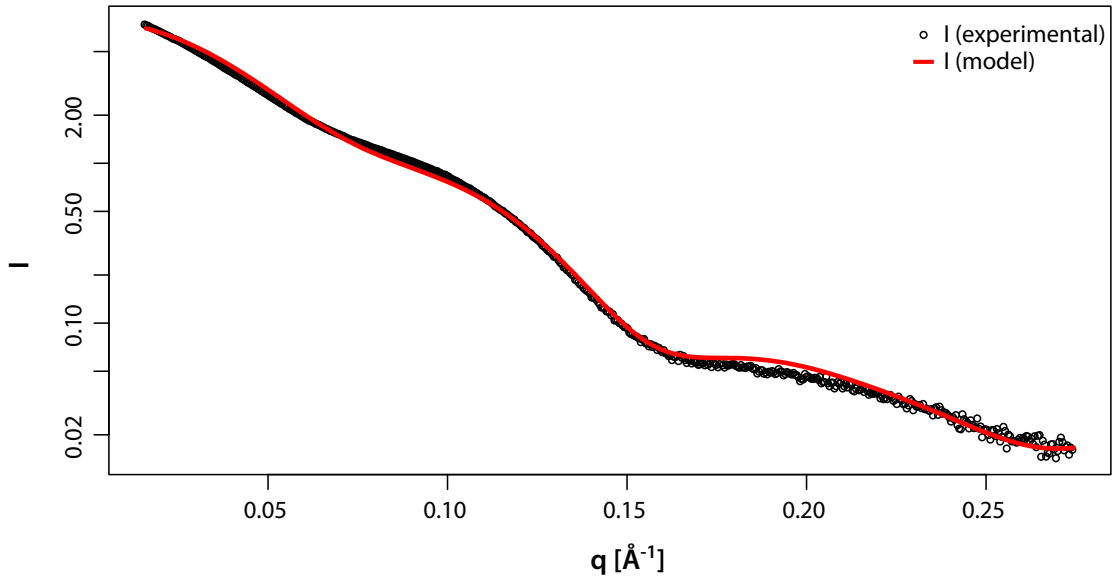



**Figure A.6.:** Agreement of $mPAC_{\Delta 132}$ high resolution model with SAXS data. The experimental scattering curve of $mPAC_{\Delta 132}$ in the presence of 1 mM ApCpp is shown in black. The theoretical scattering curve of the high-resolution model (red) was computed using FoXS [177].

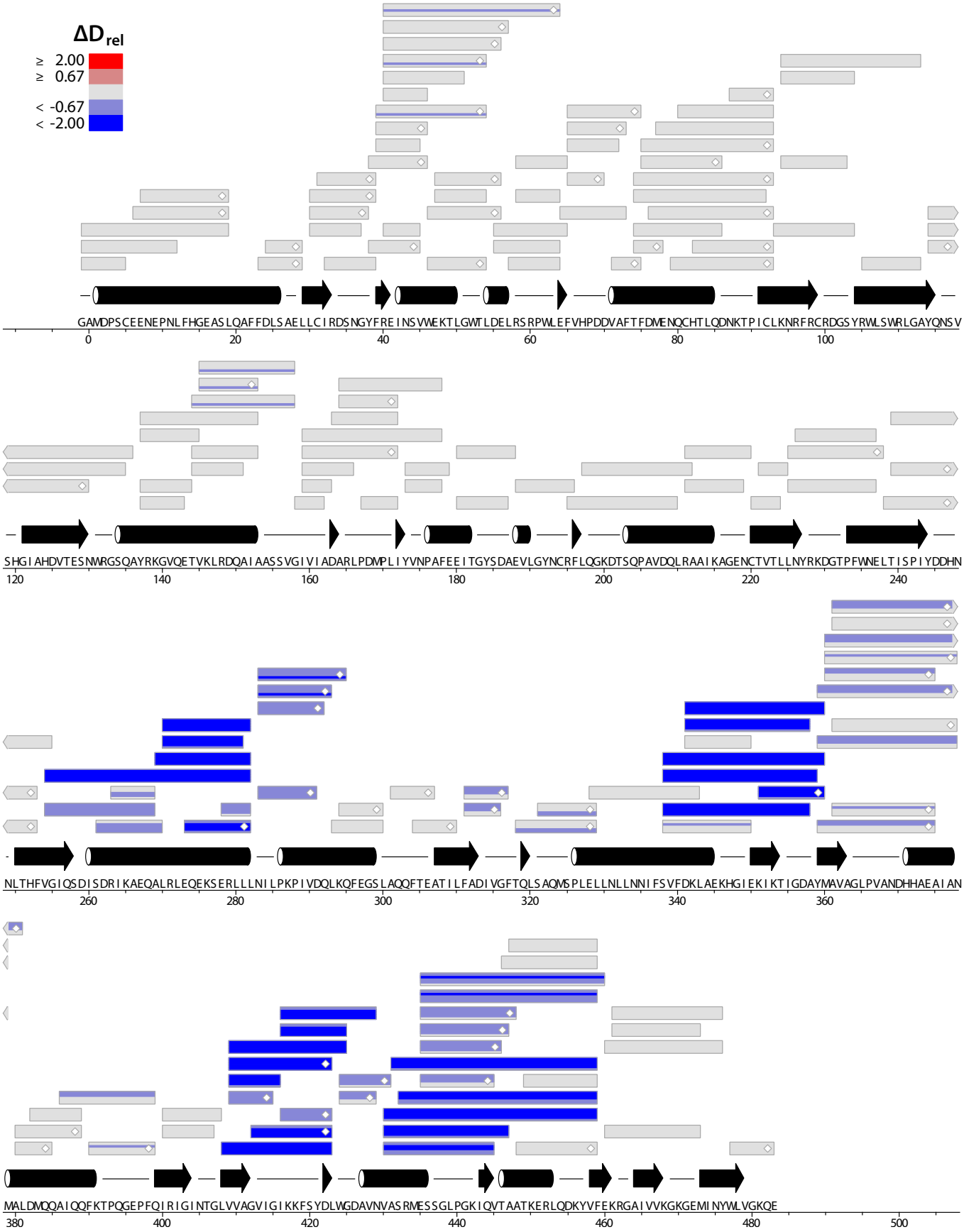



**Figure A.7.:** Differences in relative deuteration of dark-adapted mPAC in the presence and absence of ApCpp. Each box indicates one peptide and contains up to five different colors for incubation times of 15,60,300,1200 and 3600 s, respectively, bottom-up. MS/MS confirmed peptides are marked with diamonds. Secondary structure is shown as predicted by Jpred [28]. Numbering follows the wild type sequence (Uniprot B4VKN6).

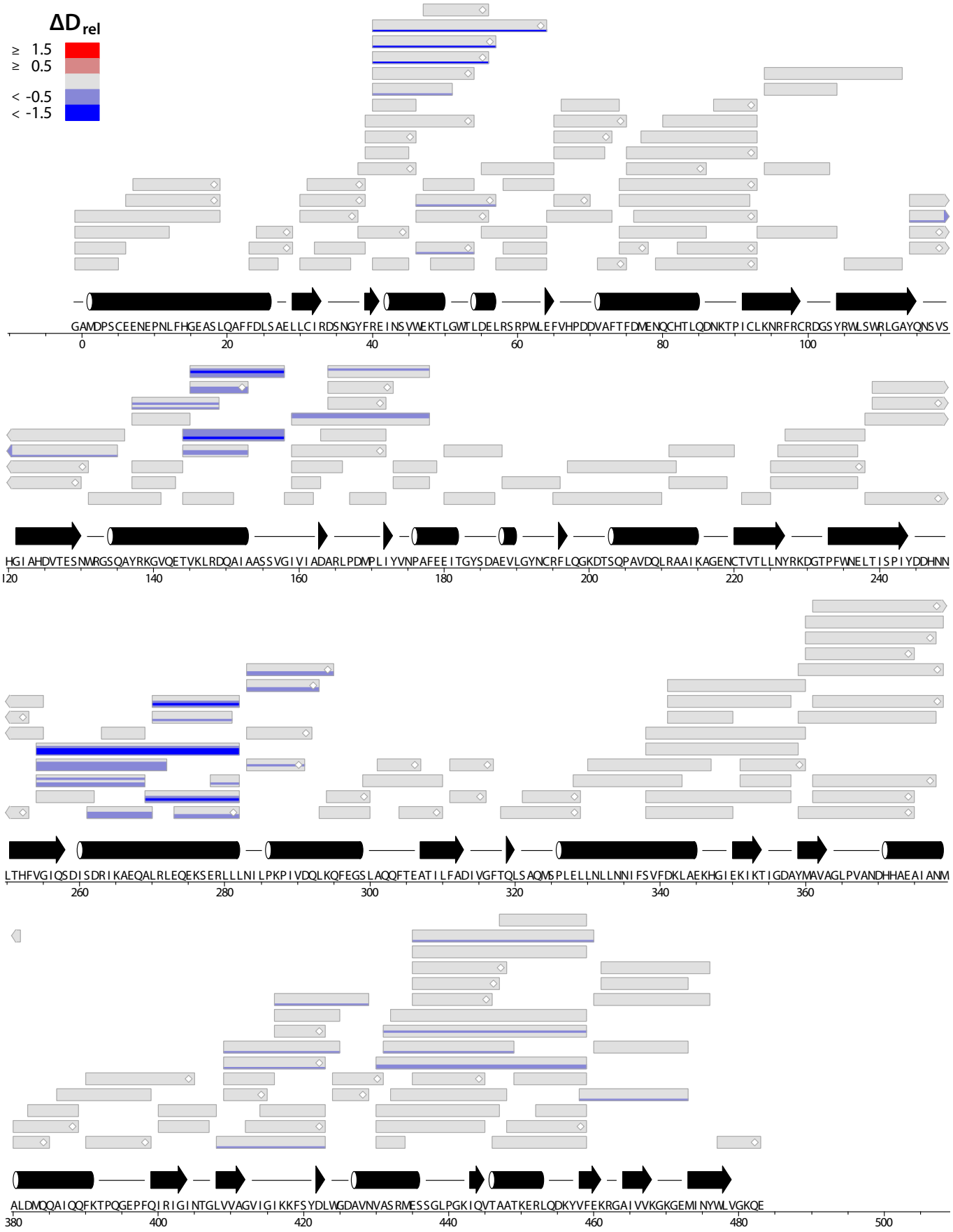



**Figure A.8.:** Differences in relative deuteration of apo mPAC upon illumination. Each box indicates one peptide and contains up to five different colors for incubation times of 15,60,300,1200 and 3600 s, respectively, bottom-up. MS/MS confirmed peptides are marked with diamonds. Secondary structure is shown as predicted by Jpred [28]. Numbering follows the wild type sequence (Uniprot B4VKN6).

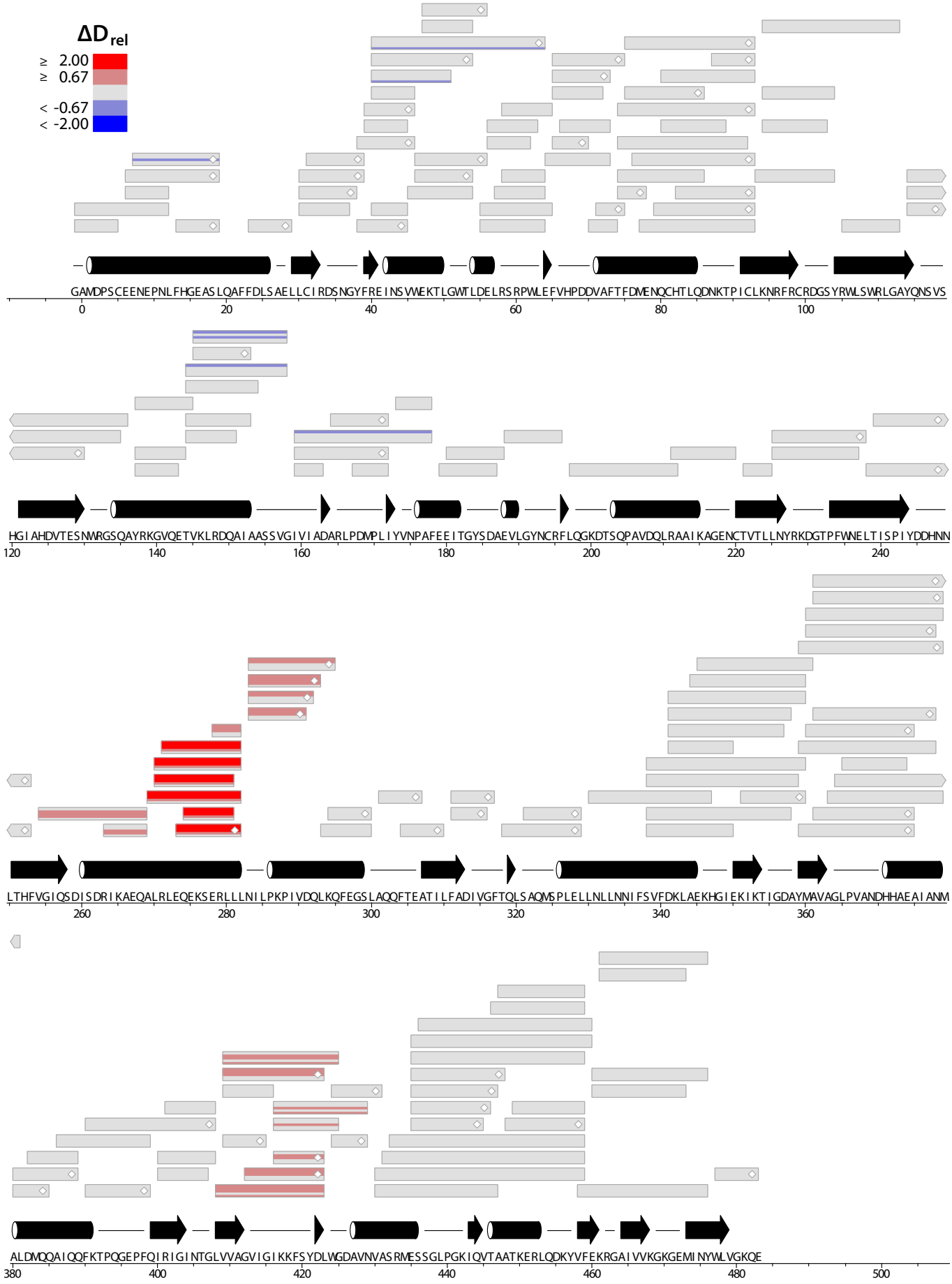



**Figure A.9.:** Differences in relative deuteration of ApCpp-bound mPAC upon illumination. Each box indicates one peptide and contains up to five different colors for incubation times of 15,60,300,1200 and 3600 s, respectively, bottom-up. MS/MS confirmed peptides are marked with diamonds. Secondary structure is shown as predicted by Jpred [28]. Numbering follows the wild type sequence (Uniprot B4VKN6).

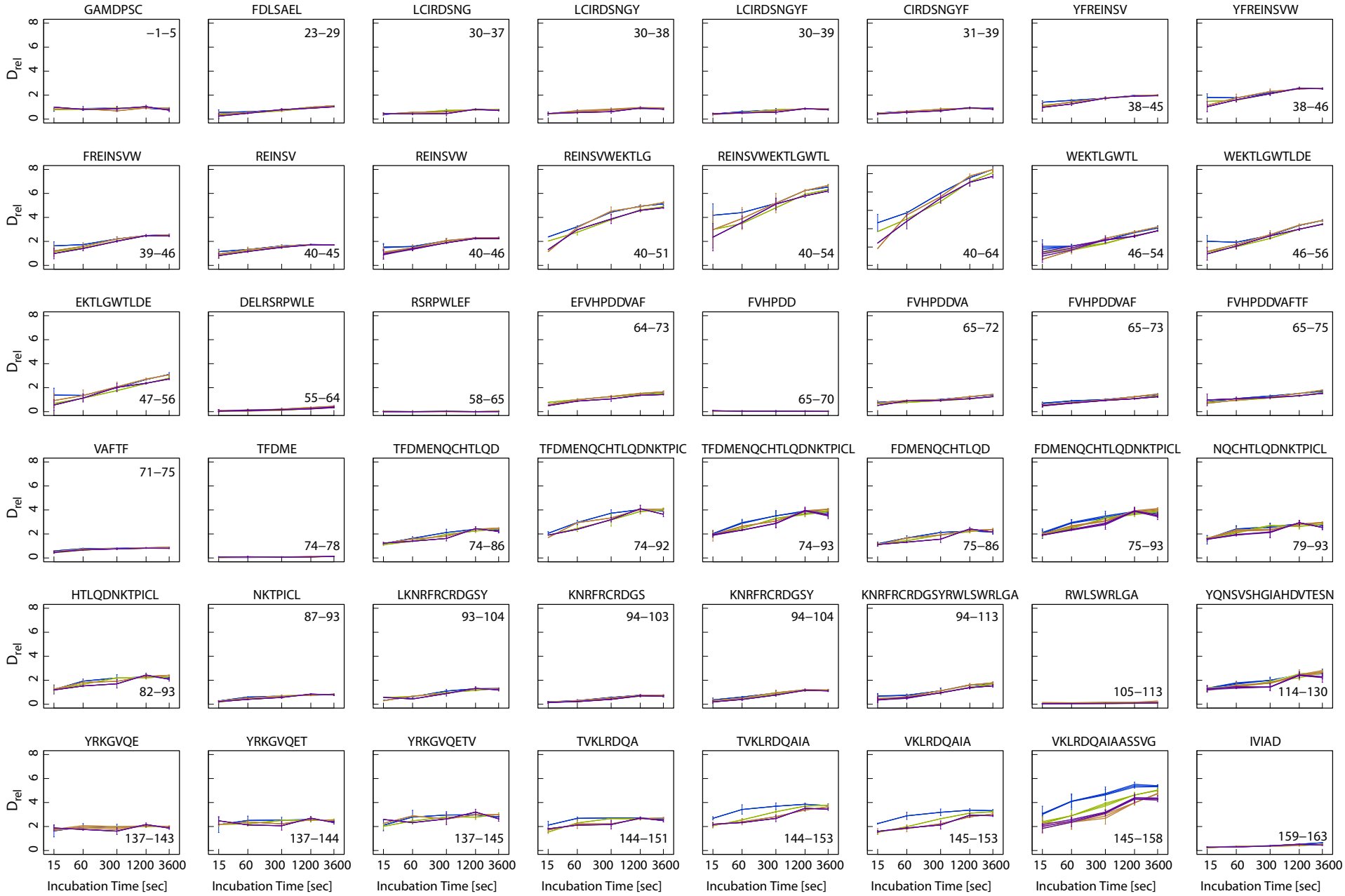



**Figure A.10.:**

*(continued on next page)*

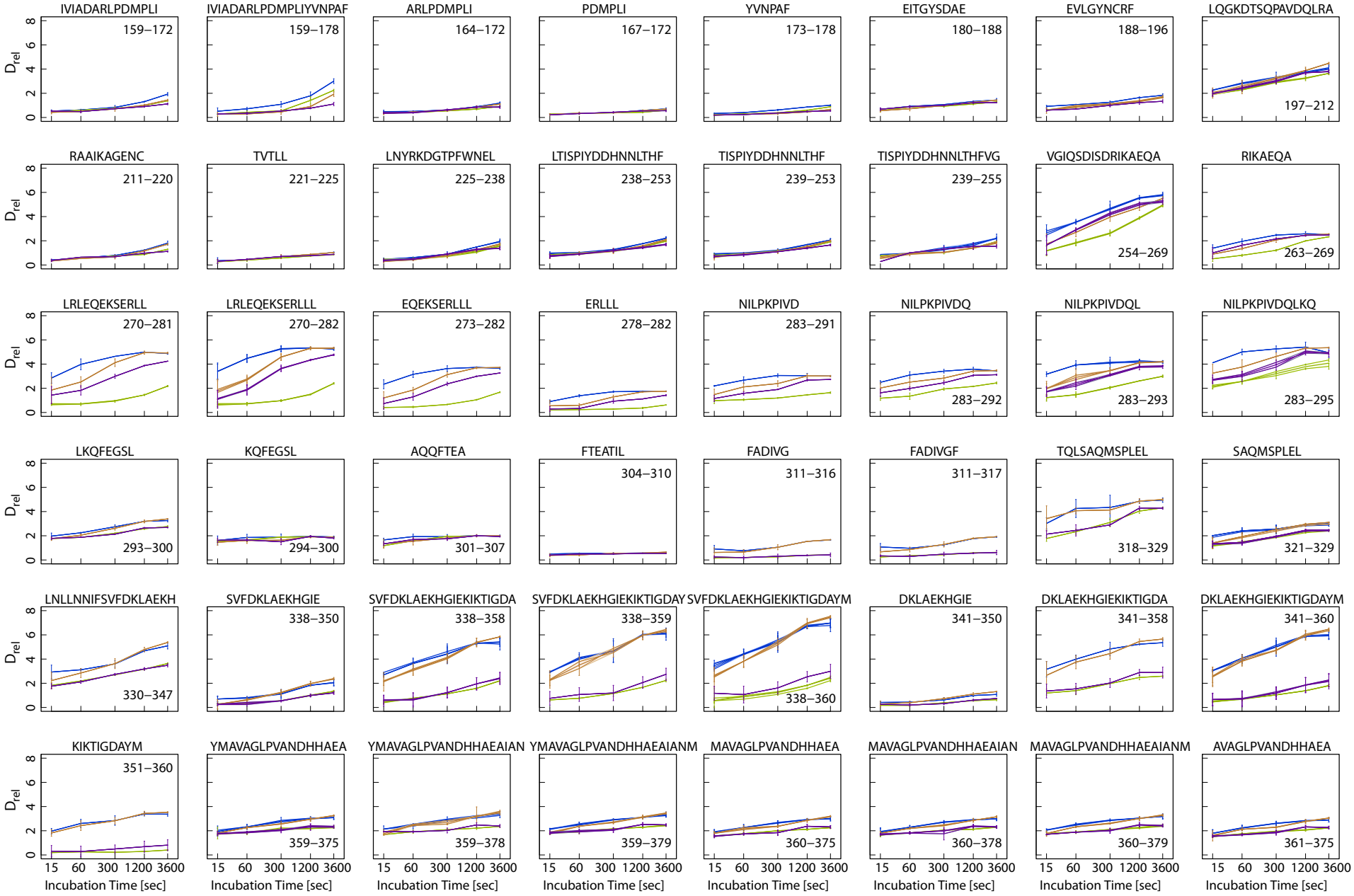



**Figure A.10.:** *(continued)*

*(continued on next page)*

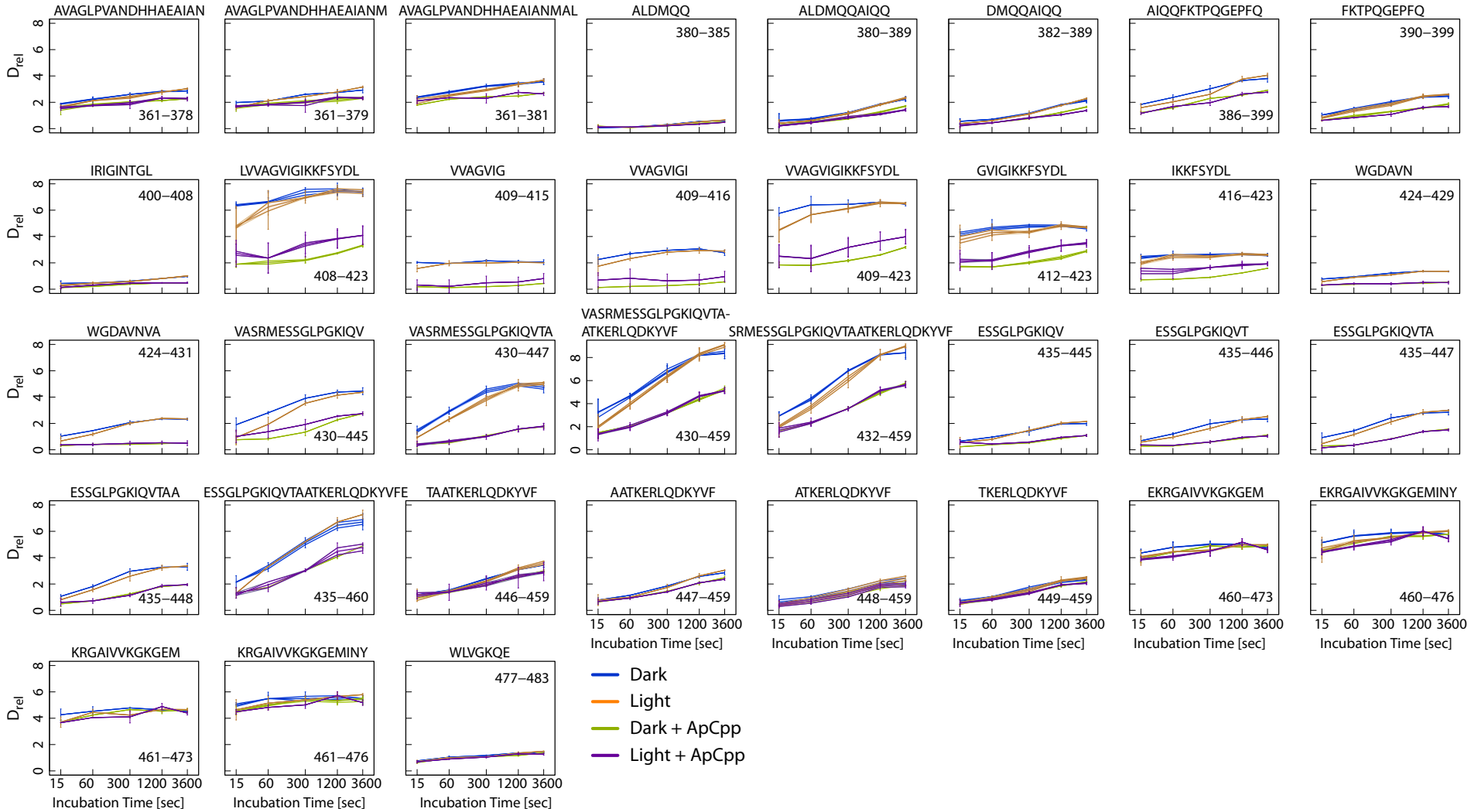



**Figure A.10.:** Deuterium incorporation time series of each peptide derived from mPAC. The temporal evolution of relative deuteration in four measured states is shown: blue – dark; green – dark in the presence of ApCpp; orange – light; purple – light in the presence of ApCpp.