THE UNIVERSITY OF CHICAGO

NEW APPROACHES TO THE DESIGN OF
ALLOSTERIC PROTEINS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

BY
LLOYD DEVIN STRICKLAND

CHICAGO, ILLINOIS
MARCH 2009
In memory of
the grandmother,
Doris Heinz
# Table of Contents

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of figures ................................................................. vii</td>
</tr>
<tr>
<td>List of tables ................................................................. ix</td>
</tr>
<tr>
<td>List of abbreviations ......................................................... x</td>
</tr>
<tr>
<td>Acknowledgments ................................................................. xi</td>
</tr>
</tbody>
</table>

## Chapter

1. **Approaches to the synthetic control of proteins** .................................................. 1
   1.1 Introduction .................................................................................. 1
   1.2 Novel globular proteins ................................................................. 2
   1.3 Conformational switches ................................................................. 4
   1.4 Functional adaptation .................................................................... 6
   1.5 Photosensors .............................................................................. 7

2. **Design of a light-controlled DNA binding protein** ............................................. 13
   2.1 Introduction .................................................................................. 13
   2.2 The allosteric lever-arm ................................................................. 15
   2.3 Design of LovTAP ......................................................................... 17
   2.4 Isolation of a photoswitching protein ............................................. 20
     2.4.1 The LOV domain photocycle and gene repression in *E. coli* ........ 20
     2.4.2 Isolation of LovTAP using an *in vitro* DNA binding assay ....... 20
2.5 Experimental validation of the design .......................................................... 24
  2.5.1 Local unfolding of the shared helix ......................................................... 24
  2.5.2 Structure and oligomeric state ............................................................... 28
  2.5.3 Conformational exchange and weak signal transfer ................................ 41
2.6 Materials and methods .............................................................................. 44

3 Improving signal transfer by mutating the \( \alpha \) helix .................................. 51
  3.1 Energetic model of signal transfer ............................................................ 51
  3.2 Improving photoswitching by rational mutation ........................................ 54
    3.2.1 Analysis of helix-stabilizing mutations ............................................. 54
    3.2.2 Mutations decrease dark-state DNA binding .................................... 58
    3.2.3 Mutations increase lit-state LOV-\( \alpha \) docking ................................. 60
  3.3 Discussion .................................................................................................. 64
  3.4 Materials and methods .............................................................................. 66

4 Conclusions and future directions ................................................................. 67
  4.1 Interdomain signal transduction through linkers ....................................... 67
    4.1.1 Signal transduction through helical linkers ....................................... 67
    4.1.2 Leashes and levers ............................................................................. 68
  4.2 Future directions in optical control of proteins .......................................... 69

Appendix: Design of a ligand-controlled DNA-binding protein ....................... 76
  A.1 Introduction ................................................................................................ 76
  A.2 Design of Zen ............................................................................................ 77
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.3 Experimental validation</td>
<td>82</td>
</tr>
<tr>
<td>A.3.1 Co\textsuperscript{2+} and Zn\textsuperscript{2+} binding</td>
<td>82</td>
</tr>
<tr>
<td>A.3.2 Conformational switching</td>
<td>82</td>
</tr>
<tr>
<td>A.3.3 DNA binding</td>
<td>90</td>
</tr>
<tr>
<td>A.4 Conclusions</td>
<td>94</td>
</tr>
<tr>
<td>A.5 Materials and methods</td>
<td>94</td>
</tr>
</tbody>
</table>

*References cited*: 98
List of figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5.1</td>
<td>The LOV2 domain of A. sativa phototropin 1</td>
<td>11</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Conceptual model of an allosteric lever arm</td>
<td>16</td>
</tr>
<tr>
<td>2.3.1</td>
<td>The E. coli TrpR dimer bound to operator DNA</td>
<td>18</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Sequence details of LOV2-TrpR fusion constructs</td>
<td>19</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Dark-state recovery in LOV2-TrpR fusion constructs</td>
<td>21</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Gene repression in vivo by LOV2-TrpR fusion constructs</td>
<td>22</td>
</tr>
<tr>
<td>2.4.3</td>
<td>DNA binding in vitro by LOV2-TrpR fusion constructs</td>
<td>23</td>
</tr>
<tr>
<td>2.4.4</td>
<td>Light-induced binding to operator DNA</td>
<td>25</td>
</tr>
<tr>
<td>2.4.5</td>
<td>Quantitative analysis of DNA binding</td>
<td>26</td>
</tr>
<tr>
<td>2.5.1</td>
<td>CD studies of the shared helix</td>
<td>29</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Structural detail of Trp 19 of TrpR</td>
<td>30</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Analysis of oligomeric state by SEC</td>
<td>32</td>
</tr>
<tr>
<td>2.5.4</td>
<td>AUC measurement of monomer-dimer equilibrium</td>
<td>33</td>
</tr>
<tr>
<td>2.5.5</td>
<td>Structural model showing steric clashes</td>
<td>34</td>
</tr>
<tr>
<td>2.5.6</td>
<td>SAXS analysis of tertiary and quaternary structure</td>
<td>36</td>
</tr>
<tr>
<td>2.5.7</td>
<td>Computational simulations of structural ensembles</td>
<td>39</td>
</tr>
<tr>
<td>2.5.8</td>
<td>Ab initio reconstruction from SAXS data</td>
<td>40</td>
</tr>
<tr>
<td>2.5.9</td>
<td>NMR studies of dark-state conformation</td>
<td>42</td>
</tr>
<tr>
<td>2.5.10</td>
<td>EPR studies of light-induced conformational changes</td>
<td>43</td>
</tr>
<tr>
<td>3.1.1</td>
<td>Model of activation and DNA binding</td>
<td>53</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Analytical model of photoswitching</td>
<td>55</td>
</tr>
</tbody>
</table>
3.2.1 AGADIR analysis and modeling of point mutants ........................................56
3.2.2 DNA binding by Jα mutants .....................................................................59
3.2.3 CD of Jα mutants ..................................................................................61
4.2.1 Experimental applications of a photoswitchable adaptor ......................71
4.2.2 Design scheme ......................................................................................72
4.2.3 Proposed experiments ...........................................................................74
A.2.1 Structural modeling ...............................................................................78
A.2.2 Sequence alignment ...............................................................................79
A.2.3 AGADIR analysis of helical propensity ................................................81
A.3.1 Spectroscopic analysis of Zn\(^{2+}\)-binding and conformational changes ..........83
A.3.2 Chemical denaturation ..........................................................................85
A.3.3 NMR analysis of Zn\(^{2+}\)-induced conformational changes .....................88
A.3.4 Chemical shift analysis of secondary structure ......................................89
A.3.5 Comparison of spectral quality in protein G and Zen .........................91
A.3.6 DNA binding studies ............................................................................92
List of tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dark- and lit-state DNA binding</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>Small-angle scattering results</td>
<td>37</td>
</tr>
<tr>
<td>3</td>
<td>Jα helix stabilization results</td>
<td>62</td>
</tr>
<tr>
<td>4</td>
<td>Chemical denaturation results</td>
<td>86</td>
</tr>
<tr>
<td>5</td>
<td>DNA binding</td>
<td>93</td>
</tr>
</tbody>
</table>
List of abbreviations

LOV  light, oxygen, voltage
AsLOV2  the LOV2 domain of *Avena sativa* phototropin 1
TrpR  the *trp* repressor of *Escherichia coli*
SAXS  small angle x-ray scattering
CD  far-UV circular dichroism spectroscopy
AUC  analytical ultracentrifugation
NMR  nuclear magnetic resonance
indel  insertion or deletion (of genomic DNA)
DNA  deoxyribonucleic acid
RNA  ribonucleic acid
Acknowledgments

First and foremost, I thank Tobin Sosnick for his insight, enthusiasm and trust. I also thank the members of the Sosnick Lab, and of the entire Department of Biochemistry and Molecular Biology for providing an atmosphere in which one can do the best science possible. In particular, Sean Crosson, Phoebe Rice and Shohei Koide kept me on track to graduate, Keith Moffat, Tao Pan and Ron Rock provided key advice at the right time, and Tony Kossiakoff and the entire BMB staff made things run smoothly. Many others helped and advised me with specific ideas or experiments—their contributions are acknowledged in the chapter headings. For getting me to the doors of the University of Chicago, I am indebted to great teachers and mentors, Townley Chisholm, Maury Hancock, William Tate, Kimo Griggs, Jim Davis and Knut Langsetmo among them. I also thank many colleagues, friends and climbing partners, including: Adarsh Pandit, Ali Shandiz, Kim Dittmar, Katrine and Daniel Whiteson, Laura Satkamp, Rob White, Aaron Smith and Kelli Harsch, and the members of the Veggie Dinner Club and the Chicago Mountaineering Club. Finally, I am thankful for the love and support of my family, and of my wife, Jennifer Strickland.
Chapter 1

Approaches to the synthetic control of proteins

1.1 Introduction

Proteins perform or participate in every vital cellular process, including the extraction and storage of energy, the chemical synthesis of new building blocks, the sensing of and response to the changing environment, and the remodeling of cellular structures for development, locomotion and proliferation. Cellular behaviors are both highly complex and tightly regulated. Proteins control proteins that, in turn, control other proteins in multilayered cascades and in networks that are almost incomprehensibly intertwined. The evolutionary appearance of new functions, particularly in higher organisms, proceeds not so much by the creation of new functional domains as by the reorganization and reordering of these networks and cascades [1].

Our desire to purposefully manipulate the function of proteins is two-pronged. We would like to better understand the roles of proteins in the life of the cell, and we would like to engineer proteins that can perform in roles unanticipated by nature. Approaches to protein design may be divided into three categories: human design, selection and computation. The human designer attempts to build a novel structure or modify an existing one by using known facts about the physical properties of amino acids, polypeptide fragments or complete proteins. Although the designer may have a specific idea of how particular structural features will lead to the desired function, this approach is necessarily limited to problems of low complexity, such as the folding of small peptides or minor modifications to existing structures. Selection builds upon human design. The
designer makes informed assumptions as to what structural features will give rise to the desired function. Based on these assumptions, the designer constructs a large library (containing up to $10^{10} - 10^{15}$ unique sequences) and isolates from it the proteins displaying the desired function [2]. Finally, using computation, the designer can explore complex models of structure and function, and screen a large number of designs in silico. Each of these approaches has strengths and weaknesses, and there is a considerable amount of complementarity and overlap among them. Computation and selection do not replace human design—rather they add leverage to it by vastly increasing the number of prototypes the designer can reasonably test.

1.2 Novel globular proteins

Among the first successful de novo designs of small globular proteins, and examples of the effectiveness of human design, were minimalist bundles of amphipathic α helices [3, 4]. These proteins were designed with hydrophobic cores composed primarily of leucine. As such, they tended to undergo hydrophobic collapse to a compact fold with one (or perhaps two) topologies, but side-chain packing was degenerate. Handel and DeGrado subsequently introduced histidine Zn$^{2+}$ ion binding residues at partially buried sites to create highly specific tertiary interactions among the helices. Adding Zn$^{2+}$ locked the protein into a single topology and globally increased the specificity of side-chain interactions [5].

Considerable progress has been made towards the computational design of well-folded proteins [6, 7]. Methods have been developed to a level of sophistication that new proteins may be designed around existing backbone topologies or new topologies within known
architectures. Dahiyat and Mayo redesigned the sequence of a small ββα (zinc-finger-like) protein [6]. The authors used a physics-based forcefield and the dead-end elimination algorithm to search a large space ($10^{62}$ unique rotamer-sequences) for the most stable sequence. They then verified the structure of the designed protein. Kuhlman, Dantas, et al. set out to design a protein with a novel topology [7]. In accordance with this goal, they first generated an ensemble of backbone-only structures based on fragments sampled from the PDB. They then used a physics-based potential to rank $>10^{186}$ rotamer-sequence combinations for each backbone and finally cycled through iterations of backbone and sequence optimization to arrive at the final design. As with the Dahiyat and Mayo study, the structure of the resulting protein was verified to match the design model.

From assembling defined units of secondary structure into globular domains, the next logical step is building larger multidomain proteins. Zhou and coworkers addressed this challenge by fusing together two well-studied helical bundles: the fast-folding B-domain of protein A (bdpA) and slower-folding Rd-apocytochrome $b_{562}$ (Rd-apocy $b_{562}$) [8]. They fused the carboxy-terminal helix of bdpA to the amino-terminal helix of Rd-apocy $b_{562}$ such that the helices partially overlapped, but hydrophobic residues from each parent protein were retained. Overall, the fusion forms a stable structure, as expected from modeling the individual domains. Although the amino-terminal helix of Rd-apocy $b_{562}$ is unstructured in the folding transition state of this protein, the fusion protein folds with two-state kinetics with a faster folding rate and a slower unfolding rate than that of the parent. This result suggests that stability can be conferred across a helical domain linker, although the mechanism by which this happens is unclear.

Undoubtedly these examples are only a glimpse at things to come as we approach the goal of being able to design particular protein structures at will. However, structure, at the
level of architecture, topology, and side-chain packing, is only one aspect of the protein design problem. The purpose of a protein, whether created by nature or by design, is to carry out a given function. Towards the end of creating proteins that are useful for science and industry, approaches aimed at modifying existing proteins have met with greater success.

### 1.3 Conformational switches

One way of controlling protein function is to create a conformational switch that responds to a change in conditions. Often, such switches are created by combining two domains in such a way that their folding is mutually exclusive—in a sense, the opposite strategy of that employed by Zhou and coworkers. In one example of this approach, Radley and coworkers created a temperature controllable barnase variant by inserting the ubiquitin sequence into a surface loop [9]. In another variation, Ha and coworkers inserted a GCN4 sequence into barnase, effectively making DNA an allosteric inhibitor of enzyme activity [10]. Hori and Sugiura created a homeodomain variant in which metal destabilizes the homeodomain structure, driving the adoption of a zinc finger fold and destroying DNA binding [11, 12]. These projects all follow the human design approach, and rely on global unfolding of the protein domains. Thus it is necessary that the protein domains employed be able to unfold reversibly, and the energy input is expected to be quite large on the scale of typical protein conformational changes.

Guntas, Ostermeier and colleagues systematically examined a series domain insertion libraries of increasing complexity [13-16]. They began by inserting a DNA fragment encoding the TEM-1 β-lactamase gene (BLA, which confers ampicillin resistance) at
random locations in the *E. coli* maltose-binding protein (MBP) gene. This approach is feasible because the termini of BLA are quite close in the 3D structure, thus the insertion should not disrupt the structure of MBP very much in most cases. They isolated from this library fusions with switching ratios of BLA activity (the ratio of $k_{cat}/K_M$ values, measured with and without maltose) of up to 1.8. Interestingly, many of the functional fusions had the BLA insert near or internal to an $\alpha$ helix in the MBP structure. Because the termini of BLA are also $\alpha$ helices, the authors speculated that the helices were long and continuous in some of the constructs [14].

In a second generation library of randomly circularly-permuted BLA inserted randomly into MBP, Guntas and coworkers recovered a fusion with up to 16-fold switching in $k_{cat}$. They also demonstrated that this fusion was able to confer greater ampicillin resistance on *E. coli* in the presence of maltose [13]. Finally, upon further refinement of library construction, they were able to increase switching up to 170-fold [15]. This work shows that it is possible to create allosteric switches by simple genetic recombination of proteins that are not normally allosterically regulated, without refinement of the primary sequence of either of the parent proteins.

Modular recombination is one of the most important means of diversification in the evolution of signaling systems in higher organisms [17]. Within this paradigm, the dominant signaling events are inter- and intramolecular binding interactions between a wide variety of conserved domains [18]. These binding interactions sometimes are, in turn modulated by posttranslational modifications, especially phosphorylation. Because the binding domains are typically small and their structures are self-contained, they constitute a kit of parts that evolution has shuffled into myriad signaling networks. The Lim lab has
applied this principle to the challenge of designing proteins with novel input-output relationships [19, 20].

1.4 Functional adaptation

One of the simplest types of functional adaptation has been the engineering of scaffolds to bind novel substrates. A common protein binding motif consists of a β sandwich with flexible loops at the ends of the β strands. This motif is found in proteins such as antibodies, fibronectin type III (FN3) domains, and extracellular hormone binding domains such as growth hormone receptors [21]. The flexible loops form a binding surface that can change conformation to be complementary to a variety of substrates. This conformational diversity, coupled with sequence diversity introduced by random mutagenesis of the loops, allows one to create tight, specific binders to potentially any substrate. Huang and coworkers adapted this idea to create a multidomain affinity clamp in which a FN3 domain is appended to a naturally-peptide-binding PDZ domain. These authors randomized the binding surface of the FN3 domain to bind to the PDZ-peptide complex, creating a clamshell structure that binds the peptide substrate more tightly than the PDZ domain alone [22].

The clamshell motif has also been exploited by nature to create tight, specific binding proteins. Bacterial periplasmic binding proteins, such as maltose binding protein (MBP) consist of two domains connected by a flexible hinge. The substrate binds in a groove between the domains, inducing a conformational change in the protein [23]. This conformational change was exploited by Guntas and colleagues to modulate BLA activity, as described above. Other studies have used FRET sensors as a way of reporting on the
binding-induced conformational change, effectively creating a fluorescent biosensor [24]. Furthermore, as the MBP structure is amenable to redesign of the binding pocket to create non-natural substrate specificity, there appears to be a high degree of flexibility in the MBP protein in particular, and in clamshell-type architectures in general.

Recently, computational methods have been used to introduce novel enzymatic activity into existing scaffolds. In a pair of articles, Baker and colleagues reported the introduction of retro-aldolase and Kemp elimination activity on nonnatural substrates into multiple, previously catalytically inert scaffolds of two different folds [25, 26]. Significantly, the majority of successful designs were based on scaffolds with a TIM-barrel fold. This fold, which features a deep active site surrounded by flexible loops, proliferated during the natural evolution of enzymes and is thought to be particularly well-suited for enzymatic activity. Nevertheless, these studies emphasize the promise of computational methods because the authors created completely novel enzymes with activities not yet found in nature. This feat is far beyond the capacity of human intuition, unaided by computation. It also exceeds the current state-of-the-art in \textit{in vivo} selection methods, because these enzymatic activities cannot be easily linked to survival of a host organism. Indeed, the major bottleneck preventing the further improvement of novel enzymes may turn out to be the lack of effective selection strategies to complement computational design.

\subsection*{1.5 Photosensors}

A major goal of protein design efforts is the creation of novel molecules that can be used for \textit{in vivo} biological experiments. Light is an especially attractive means of
perturbing biological systems because commercially-available light microscopes allow for convenient illumination of cells, tissues and even intact organisms with high spatial and temporal precision.

In vitro chemical modification of proteins has proven to be a successful strategy for inducing optical control of biological systems. In particular, the small molecule azobenzene (AZB) and its derivatives have been widely used to create photoswitchable variants of peptides and proteins [27]. Typically, a bi-functional cysteine-reactive AZB moiety is covalently attached to two cysteines on a peptide or protein. Photoisomerization of the AZB changes the geometry of the peptide, reversibly disrupting its structure and function. The geometry of this type of AZB modification is especially compatible with helical structures such a GCN4 [28]. A limitation of AZB modification for biological experiments is that the protein must be engineered, purified, chemically modified and reintroduced into the cell. However, this limitation is neatly avoided in the case of AZB-modified membrane proteins—the AZB can be reacted with extracellular domains on living cells. Several variations of this approach have been used to create synthetic, light-gated channels for use in electrophysiology experiments [29].

Several classes of proteins mediate biological responses to light: rhodopsins, Per-Arnt-Sim (PAS) domains, including photoactive yellow protein (PYP) and Light, Oxygen or Voltage (LOV) domains, BLUF domains, cryptochromes and phytochromes [30]. Levskaya and coworkers created one of the first genetically-encoded, synthetic, optically controlled systems by engineering the two-component EnvZ-OmpR system to respond to light instead of the natural osmotic response. They replaced the periplasmic sensory domain of the EnvZ histidine kinase with a bacterial phytochrome (phy) Cph1 [31]. Illumination of the E. coli carrying this engineered sensor causes EnvZ
autophosphorylation, activation of the OmpR response regulator, and transcription of a lacZ reporter gene. Although this project was remarkable as a feat of synthetic biology, it highlights an engineering limitation of phytochrome domains. These photosensors require specific linear tetapyrrole cofactors (phycocyanobilin, PCB, in this case) to function. Many organisms do not have phytochromes, and thus are not likely to produce the necessary cofactor for transgenic phytochrome expression. Levskaya et al. readily overcame this problem by introducing additional genes for PCB biosynthesis—nevertheless, this additional step could represent a significant hurdle for routine experimental expression of phy-based photoswitches.

Several groups have taken advantage of the natural binding interaction between plant phy domains and the PIF3 domain [32]. Shimizu-Sato and coworkers used this binding interaction in a standard yeast two-hybrid system to control transcription of a reporter gene [33]. Leung and coworkers used the interaction to enforce the proximity of a Rac1 GTPase and a WASP effector domain [34]. They showed that the activated Rac1-WASP complex was able to stimulate actin assembly in a light-dependent manner. As with the Levskaya et al. study, these designs could not be functional in vivo experimental systems without simultaneous expression of cofactor biosynthetic genes.

Similar to the phy domain, the LOV (light-oxygen-voltage) domain is a naturally occurring photosensory receptor. LOV domains were first identified as the molecular sensor of light in plant phototropism and are widely distributed in signaling pathways [35-37]. The α/β core of the LOV domain binds a cofactor, flavin, that is found in most cell types. The cofactor has a broad absorption spectrum, with a maximum at 450 nm. LOV domains respond to blue light, in contrast to the red-light sensitivity of phy domains [38]. Photon absorption leads to a photoexcited state that promotes the formation of a covalent
adduct between the 4a carbon of the flavin and an invariant cysteine residue in the LOV domain, which results in conformational rearrangements in the protein [39].

The core of the LOV domain is often flanked by amino- or carboxy-terminal helices, termed A’α and Ja, respectively (Fig. 1.5.1) [40-42]. These helices pack against the central sheet, directly opposite the flavin-binding pocket. In at least two divergent LOV domains, photoexcitation is accompanied by displacement and unfolding of an α helix that lies either amino- or carboxy-terminal to the core domain. Displacement of this helix mediates signal transduction to effector domains in situ [40, 41]. In the case of AsLOV2, both the folded and unfolded conformations of the amino-terminal helix are populated in the dark and photoexcited (lit) states. However the balance of these two populations shifts dramatically between the dark- and lit-states, changing from ~98% folded in the dark to ~10% folded when fully lit [43]. Thus, the LOV domain can potentially regulate the activity of an effector by 600-fold.

LOV domains bind a flavin-mononucleotide (FMN) or flavin-adenine-dinucleotide (FAD) cofactor. These small molecules are used in a wide variety of metabolic pathways as cofactors in redox reactions and, conveniently, are available in most organisms for incorporation into heterologously-expressed LOV domains.

Lee, Natarajan and coworkers in the Ranganathan and Benkovic labs have demonstrated a potential means of conveniently engineering LOV-based photoswitches. Using a conceptually similar approach to that of Gunatas and Ostermeier, they inserted the A. sativa LOV2 (AsLOV2) domain into two sites in E. coli dihydrofolate reductase (DHFR) [44]. One of these sites was identified using statistical coupling analysis (SCA) as an allosterically active region of the protein. SCA and similar methods of covariation
Figure 1.5.1 The LOV2 domain of A. sativa phototropin 1. A: An X-ray crystal structure of the LOV domain (PDB ID 2VOU, light blue ribbon) shows the amino-terminal A´α and the carboxy-terminal Ja helix (dark blue ribbon) [62]. B: Photoexcitation of LOV2 is accompanied by displacement and unfolding of the Ja helix. Although both the Ja-docked (inactive) and Ja-displaced (active) conformations are detectably populated in the dark and lit states, photoexcitation increases the equilibrium constant of helix displacement ($K_{helix}$) by 600-fold.
analysis have been used to aid in the design of small proteins, and to identify residues which are important in protein-protein interactions [45, 46]. The other site was a control site not identified by SCA. The authors found that at least one insertion at the predicted allosteric site showed a two-fold increase in enzymatic activity upon photoactivation, but no change for insertions at the control site. This result highlights the fact that the LOV domain is a good candidate for domain insertion, and could potentially be used to great effect in conjunction with the method of Guntas and Ostermeier [15].

In the field of protein design number of significant practical challenges remain. The ability to design a sequence that can efficiently perform an arbitrarily chosen function—building an ideal protein photosensor for biological experiments from scratch, for example—likely will not be attained for years, or even decades. Nevertheless, smaller, practical goals can and are being met. For example, Huang and coworkers used FN3 as a small antibody mimic, or “minibody” to selectively purify active estrogen receptor ligand binding domain from a mixture of active and inactive proteins [47]. We are rapidly approaching a point at which we can design, at will, genetically-encoded, optically-controlled proteins for routine experimental purposes.
Chapter 2

Design of a light-controlled DNA binding protein

Additional, unpublished data have been added, and the text has been revised to include discussion of these results. I thank Charles Yanofsky and Jannette Carey for advice on TrpR experiments and interpretations, Calin Guet for advice on E. coli reporter experiments, Liang Guo, Nathan Baird, Venky Pingali and Pappannan Thiyagarajan for advice and assistance with SAXS experiments, Walter Stafford for advice on AUC experiments, Josh Kurutz for advice on NMR experiments, and Kevin Gardner and Xiaolan Yao for helpful discussions. H. Raghuraman performed the EPR experiments. Alexander Augustyn performed the monte-carlo simulations.

2.1 Introduction

A substantial test of our understanding of protein structure is the design of new molecules with controllable functions, including the rewiring of protein-mediated signaling networks. The modular nature of proteins makes such engineering straightforward in principle. However, the mechanisms by which the components can be allosterically connected are not fully within our control. In one prevalent model, allostery results from intramolecular binding between discrete domains that are linked by regions of undefined structure [18]. Although this model has considerable appeal and explanatory power, the discovery of additional mechanisms that are compatible with modular
architectures would increase our understanding of signal transduction and facilitate protein design efforts [48].

New functional architectures evolve through the shuffling of modular domains by genomic insertion and deletion (indels) [17, 49]. In one type of indel, domain insertion, two covalent connections between the domains constrain their relative orientation. Allostery in these chimeras is probably related to structural changes at the junctions [16, 50]. Protein designers have used domain insertion to create allosteric switches with mechanisms ranging from subtle conformational changes to mutually exclusive folding [9, 15, 51]. A conceptually similar approach has been used successfully to engineer highly modular RNA switches [52]. The other type of indel, end-to-end fusion, imposes fewer constraints on the relative orientation of the domains, especially if the linker region is unstructured. On the one hand, few structural constraints may favor easy evolvability; on the other, possible mechanisms of allosteric coupling may be excluded.

Allostery, once regarded as an evolutionarily-refined property occurring only in oligomeric proteins, is now understood to be common to proteins and readily exploited by evolution or by design [53–56]. Reinforcing this idea, especially with respect to modular signaling proteins, is the observation that the deactivation of a protein can be accomplished by any perturbation that disrupts the active site [57]. Allosteric effectors may be small molecules, peptides, or other proteins that bind to a site distant from the active site.

Alternatively, allostery may be associated with inter- and intra-molecular binding interactions that cause either steric occlusion of the active site or domain rearrangements acting on short linkers such that strain is introduced into the structure [18, 19, 58]. Nevertheless, designers have created systems which do not fit easily with this framework.
Ostermeier, Guntas and colleagues recovered an allosteric end-to-end fusion of \( \beta \)-lactamase and maltose binding protein, domains which do not naturally associate, from a library [14]. It remains unclear whether allostery in end-to-end fusions can be associated with structural changes in the linker regions, in the absence of conserved interdomain interactions.

### 2.2 The allosteric lever-arm

We reasoned that an allosteric switch could be created by joining two domains such that they share a continuous helix (Fig. 2.2.1). By itself, this type of fusion can result in a single well-folded protein, as shown by Bai, Zhou and coworkers [8]. In order to create an allosteric switch with a bistable energy surface having two alternative wells, we built the fusions to have a steric domain-domain overlap if the shared helix assumes its normal position in both domains. Because residues in regular helices are confined to a narrow region of the Ramachandran map of backbone dihedral angles, bending of an \( \alpha \) helix is energetically unfavorable [59]. As a consequence, the shared helix acts as a rigid lever arm, and the overlap is most readily relieved by the disruption of contacts between the shared helix and one domain or the other.

Because these helical contacts are integral to the structure of the domains, their disruption will cause a global shift in the conformational ensemble [53, 60, 61]. Conversely, a perturbation such as ligand binding or photoexcitation, which changes the conformational ensemble of the protein, will also change the stability of the helix-domain contacts. This change shifts the relative affinity of the shared helix for each of the two domains, thereby allowing a signal sensed by one domain to be allosterically propagated
Figure 2.2.1 Conceptual model of an allosteric lever arm. Joining two domains across terminal $\alpha$ helices creates a bistable system in which steric overlap (red star) is relieved by the disruption of contacts between the shared helix and one or the other of the domains. A perturbation ($\Delta$) such as ligand binding or photoexcitation alters the energy surface of the system (black line) to favor a new conformational ensemble (dashed line) with different functional properties.
to the other domain. With this concept in mind, we designed a light-controlled DNA binding protein.

2.3 Design of LovTAP

As a light-sensitive input module, we chose the photoactive LOV2 domain of *Avena sativa* phototropin 1 (Fig. 1.5.1). LOV domains, widely distributed in modular, multidomain signaling proteins, contain a cofactor-binding PAS domain flanked by amino- or carboxy-terminal helices [35, 37, 41, 42, 62]. In AsLOV2, absorption of a photon leads to the formation of a covalent adduct between the FMN cofactor and a conserved cysteine residue [63]. This adduct decays spontaneously to the ground state in tens of seconds. Adduct formation is accompanied by the displacement and unfolding of the 20-residue, carboxy-terminal Jα helix from the LOV domain, an event which likely mediates signal propagation in its biological context [40, 64].

As an output module, we chose the bacterial transcription factor *trp* repressor (TrpR, Fig. 2.3.1). TrpR, with its L-tryptophan cofactor, binds its operator DNA as a homodimer [65]. Mutations throughout the protein, including those in the 21-residue, amino-terminal helix, affect cofactor and operator binding, suggesting the presence of many allosterically sensitive sites [66-69]. Isolated TrpR domains occur widely in bacteria, but are not known to participate in modular architectures.

We ligated AsLOV2 (residues 404–543) via its carboxy-terminal Jα helix to a succession of 13 amino-terminal truncations of TrpR (residues 11–108, Fig. 2.3.2). We chose residue 11 as the amino-terminal boundary of TrpR because this amino acid is the first helical residue in the crystal structure [70].
Figure 2.3.1 The E. coli TrpR dimer bound to operator DNA. TrpR (PDB ID 1TRR) is shown as an orange ribbon [70]. Operator DNA is shown as a gray surface. The amino-terminus of the protein is an α helix (red).
Figure 2.3.2  Sequence details of LOV2-TrpR fusion constructs. The amino acid sequence of the Jα helix of LOV2, through Ala 543, is shown in blue. The sequence of the amino-terminal helix of TrpR, beginning with Ala 2, is shown in red. Trp 19 of TrpR is indicated with an arrow. For this study, we created a series of constructs in which the LOV2 domain, intact through Ala 543, is fused to successive truncations of the amino-terminal helix of TrpR, beginning with Met 11. Constructs are identified by the first included TrpR residue at the point of fusion.

\[
\begin{array}{ll}
A. \textit{sativa LOV2} & ---\text{AAERGVLikkTAENIDEAA }543 \downarrow \\
E. \textit{coli trpR} & 2 \text{AQSPYSAAEMABQRHQEWLRFVDLKNAYQN}--- \\
\text{Met 11} & ---\text{AAERGVLikkTAENIDEAA}AMABQRHQEWLRFVDLKNAYQN--- \\
\text{Ala 12} & ---\text{AAERGVLikkTAENIDEAA}AMBQRHQEWLRFVDLKNAYQN--- \\
\text{Glu 13} & ---\text{AAERGVLikkTAENIDEAA}AEQRHQEWLRFVDLKNAYQN--- \\
\text{Phe 22 (LovTAP)} & ---\text{AAERGVLikkTAENIDEAA}AFVDLKNAYQN--- \\
\text{Leu 25} & ---\text{AAERGVLikkTAENIDEAA}ALLKNAYQN--- \\
\end{array}
\]
2.4 Isolation of a photoswitching protein

2.4.1 The LOV domain photocycle and gene repression in E. coli

We considered the possibility that some fusions of LOV2 and TrpR might abolish the activity of one or both domains. Upon purification, we found that FMN binding and reversible photobleaching are preserved in all constructs, suggesting that the activity of the LOV domain is preserved (Fig. 2.4.1) [63]. We also tested whether the LOV2-TrpR constructs are able to repress a β-galactosidase (LacZ) reporter gene in E. coli, and whether this repression is light dependent [71]. Bacteria expressing the constructs from a plasmid produced low levels of LacZ, comparable to bacteria expressing wild-type TrpR from the same plasmid (Fig. 2.4.2). This suggests that TrpR tolerates fusion with the LOV domain well. None of the constructs showed any light sensitivity in this assay—expression levels in the dark and under broad-spectrum fluorescent lights were the same. This is a false negative result in the case of one of the constructs. Most likely this is due to a combination of the low intensity of the fluorescent lamps, and high expression levels of the construct that resulted in a high background repression.

2.4.2 Isolation of LovTAP using an in vitro DNA binding assay

We used a more sensitive in vitro protection assay to test for light sensitivity in DNA binding [72]. Similar to the in vivo assay, we found that all constructs specifically protect cognate DNA against nuclease digestion under ambient light, although they do so to differing degrees depending on the truncation of the TrpR domain (Fig. 2.4.3). One construct, which we refer to as the LOV and Tryptophan Activated Protein, or LovTAP, preferentially protects cognate DNA when illuminated (Fig. 2.4.4). This construct joins
Figure 2.4.1 Dark-state recovery in LOV2-TrpR fusion constructs. Recovery after photoexcitation was monitored by absorbance at 447 nm. Constructs are identified by the first included TrpR residue at the point of fusion. Kinetic traces are normalized to an amplitude of 1.
**Figure 2.4.2** Gene repression in vivo by LOV2-TrpR fusion constructs. Transcription of *lacZ* from a *trp* promoter was monitored by β-galactosidase activity. Cultures were grown either in the dark or under a broad-spectrum fluorescent light, as indicated. Constructs are identified by the first included TrpR residue at the point of fusion.
Figure 2.4.3  DNA binding in vitro by LOV2-TrpR fusion constructs. A: Map of plasmid used in the RsaI protection assay. B: DNA binding under ambient light. Concentrations, indicated by black wedges, were 320, 100 and 30 nM monomer (100 and 30 nM for constructs 21 and 25). Specific DNA binding to the trp operator protects the large (1890 bp) fragment from digestion to the small (1410 bp) fragment by RsaI. Constructs are identified by the first included TrpR residue at the point of fusion.
the carboxy-terminus of the Jα helix of the LOV domain to the middle of the amino-terminal helix of TrpR at Phe 22 (Fig. 2.3.2). At saturating photoexcitation (20 mW·cm⁻² irradiance at 470 nm) and micromolar LovTAP concentration, the apparent rate of RsaI digestion of cognate DNA is decreased by about 2-fold compared to the rate in the dark. DNA binding is specific for the trp operator and requires free L-tryptophan (data not shown), suggesting that LovTAP binds DNA in a manner that is characteristic of the TrpR domain. Mutation of the photoactive cysteine of the LOV domain to serine prevents the normal photocycle and abolishes the light-sensitivity of DNA protection. Therefore, the activation of LovTAP originates with photochemical events involving the FMN cofactor [63].

To better quantify the effectiveness of the LovTAP design, we determined its binding affinity for operator DNA in the dark and lit states. In a modified RsaI protection assay, with a DNA substrate having only one cleavage site, the observed cleavage rate was the intrinsic rate multiplied by the fractional occupancy of LovTAP at the cleavage site (Fig. 2.4.5 and Table 1). For nanomolar LovTAP concentrations, the observed cleavages rates indicated average DNA binding affinities of 130 ± 10 nM and 780 ± 300 nM in the lit and dark states, respectively. Therefore, photoexcitation of LovTAP increases DNA binding by 6-fold.

2.5 Experimental validation of the design

2.5.1 Local unfolding of the shared helix

We tested whether the shared helix is implicated in the allosteric behavior of LovTAP by measuring the helical content in the dark and lit states using CD spectroscopy. The
Figure 2.4.4 Light-induced binding to operator DNA. A: DNA protection in the light (L) and dark (D) at 50 nM protein monomer. Only the 1890 bp (upper) and the 1410 bp (lower) bands are shown. The examples are representative of all constructs except LovTAP. B: DNA protection in the light and dark at 130 nM LovTAP monomer. C: Light (dashed lines) and dark (solid lines) activity of LovTAP. Digestion is the intensity ratio of the sum of the product bands to the sum of the product bands plus the reactant band. The concentrations in the legend indicate LovTAP monomer concentration in reaction mixture.
Figure 2.4.5  Quantitative analysis of DNA binding. The concentrations in the legend indicate LovTAP dimer concentration in reaction mixture. Digestion is the intensity ratio of the sum of the product bands to the sum of the product bands plus the reactant band. Red lines are single exponential fits to the data.
<table>
<thead>
<tr>
<th>[LovTAP] (nM)</th>
<th>( k_0 ) Dark (min(^{-1}))</th>
<th>( k_0 ) Lit (min(^{-1}))</th>
<th>( k_{\text{obs}} ) Dark (min(^{-1}))</th>
<th>( k_{\text{obs}} ) Lit (min(^{-1}))</th>
<th>( K_d ) Dark (nM)</th>
<th>( K_d ) Lit (nM)</th>
<th>Fold-change</th>
</tr>
</thead>
<tbody>
<tr>
<td>285</td>
<td>0.127 ± 0.004</td>
<td>0.139 ± 0.007</td>
<td>0.093 ± 0.004</td>
<td>0.044 ± 0.002</td>
<td>788 ± 61</td>
<td>132 ± 29</td>
<td>5.9 ± 1.4</td>
</tr>
<tr>
<td>440</td>
<td>0.083 ± 0.005</td>
<td>0.091 ± 0.004</td>
<td>0.059 ± 0.005</td>
<td>0.018 ± 0.002</td>
<td>1074 ± 148</td>
<td>107 ± 79</td>
<td>10 ± 7.6</td>
</tr>
<tr>
<td>690</td>
<td>0.13 ± 0.02</td>
<td>0.12 ± 0.02</td>
<td>0.052 ± 0.005</td>
<td>0.025 ± 0.0001</td>
<td>502 ± 232</td>
<td>186 ± 163</td>
<td>2.7 ± 2.7</td>
</tr>
<tr>
<td>Average</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>788 ± 94</td>
<td>142 ± 61</td>
<td>5.6 ± 2.5</td>
</tr>
</tbody>
</table>

* \( k_0 \) is the intrinsic rate of digestion of DNA by RsaI.

\(^{\dagger}\) \( k_{\text{obs}} \) is the rate of digestion of DNA by RsaI in the presence of LovTAP and L-tryptophan.

\(^{\ddagger}\) \( K_d \) is the dissociation constant of the binding reaction of LovTAP to cognate DNA.
dark-state CD spectrum of LovTAP is close to the residue-weighted average of the dark-state AsLOV2 spectrum and the TrpR spectrum (Fig. 2.5.1) [68, 73]. The difference of these two spectra indicates that some of the α helix present in AsLOV2 or TrpR has been lost in the fusion and replaced with random coil (Fig. 2.5.1). This result is expected because 11 helical TrpR residues were truncated in making the construct, and at least some of the remaining helical residues are predicted to be unfolded in the dark state.

Mutating residues Ile 532, Ala 536, or Ile 539 of AsLOV2, which lie on the Jα helix and contact the β sheet, to a charged glutamic acid leads to a pseudo-lit-state structure [64]. In the dark state, these mutants have a less negative CD signal at 222 and 207 nm than LovTAP, consistent with the mutations partially unfolding the shared helix (Fig. 2.5.1, I532E and A536E not shown). Under saturating photoexcitation, all three variants have similar CD signals at 222 nm (−8,000 deg·cm²·dmol⁻¹) and at 207 nm (−9,500 deg·cm²·dmol⁻¹), suggesting similar structures (Fig. 2.5.1). In full-length phototropin 1, the I532E, A536E and I539E mutations in the LOV2 domain decouple kinase activity from photoexcitation [64]. Similarly in LovTAP, these mutations decouple DNA protection from photoexcitation, implicating the Jα helix in the mechanism of allosteric signal propagation (Fig. 2.4.4, I532E and A536E not shown).

### 2.5.2 Structure and oligomeric state

Wild-type TrpR is a tight, intertwined dimer. LovTAP contains a destabilizing substitution of Trp 19 in the hydrophobic core for glutamic acid (Fig. 2.3.2, 2.5.2). Trp 19, the first conserved residue in bacterial TrpR domains and the first residue in the primary sequence of *E. coli* TrpR to participate in the hydrophobic core, pins the amino-terminal helical arm to the body of the protein. TrpR, alone and in LOV2-TrpR constructs,
Figure 2.5.1  CD studies of the shared helix. A: Dark-state CD spectra. LovTAP and the I539E mutant are shown, as are AsLOV2 and TrpR. The green line (Average) is the residue-weighted average of LOV2 and TrpR. The dotted line (Difference x3) is the difference of the LovTAP spectrum and the residue-weighted average spectrum multiplied by three. B: Dark-state recovery, monitored by CD, for LovTAP and the I539E mutant. Black lines are exponential fits to the data.
Figure 2.5.2 *Structural detail of Trp 19 of TrpR.* One monomer of TrpR is shown as an orange ribbon, and the other is shown as a brown ribbon. Trp 19 is shown as spheres, the L-tryptophan cofactor is shown as sticks, and operator DNA is shown as a gray surface.
tolerates changes at this highly conserved position, but always with a diminution of DNA-binding activity (Fig. 2.4.3) [67]. Shao and coworkers showed that the L39E mutation causes TrpR to be monomeric in solution ($K_d = 111 \, \mu\text{M}$), indicating that a single mutation can dramatically disrupt its structure [74]. Size exclusion chromatography (SEC) of different LOV2-TrpR constructs suggests that the mutation of W19 causes a similar shift to monomer (Fig. 2.5.3). Also by SEC, LovTAP, with its W19E mutation, is partially monomeric at concentrations below 20 $\mu\text{M}$ (Fig. 2.5.3). Using sedimentation equilibrium analytical ultracentrifugation (AUC) we measured the $K_d$ in LovTAP to be $\sim 2 \, \mu\text{M}$ (Fig. 2.5.4) [75]. It is unclear whether light-induced modulation of the monomer-dimer equilibrium plays a role in the mechanism. It is reasonable to think that photoactivation causes a change in the equilibrium, but this idea remains untested.

We built a working model of LovTAP in the dark-state by assuming that the carboxy-terminal $\alpha$ helix of LOV2 and the amino-terminal helix of TrpR form a structurally continuous, shared helix across the point of ligation (Fig. 2.5.5). Steric overlaps occur between the LOV domains and the TrpR domain, suggesting that contacts between the shared helix and one or both of the parent domains must be disrupted to relieve the strain. Given that the $\alpha$ helix readily dissociates from the core of AsLOV2, while wild-type TrpR is a tight, intertwined dimer, the shared helix should strongly associate with the TrpR domain in the context of LovTAP [66]. Upon the destabilizing W19E substitution in LovTAP, a few residues of the shared helix presumably dissociate from the TrpR domain and dock against the LOV domain. In this configuration, the steric overlap is relieved, but the TrpR domain has decreased DNA binding affinity, a weaker monomer-dimer equilibrium, or both (Fig. 2.5.3) [74].
Figure 2.5.3 Analysis of oligomeric state by SEC. The L-tryptophan cofactor is not present. 

A: Comparison of LOV2-TrpR constructs. LovTAP is in black, and non-allosteric constructs are in gray. The numbers above the traces indicate the approximate elution volume, in mL. Of the non-allosteric constructs, all with Trp19 intact are in the group eluting at 49 mL, while all with Trp19 substituted are in the group eluting at 55 mL. B: Concentration dependence of LovTAP elution. The numbers above the traces indicate the peak monomer concentration at elution, in µM. The 46 mL (20 µM) and 55 mL (0.1 µM) peaks are ~60 and ~30 kDa proteins, respectively, corresponding to dimeric and monomeric species. The 50 mL peak (1 µM) is a mixture in equilibrium on the time scale of the experiment. The 63 mL peak (*) is a stable proteolytic fragment of the LOV domain.
Figure 2.5.4  AUC measurement of monomer-dimer equilibrium. A global fit by SEDANAL is shown as green lines. Monomer concentrations (0.9, 1.8, 3.5 and 7 µM) and rotor speeds (22K, 32K and 39K rpm) are indicated. The L-tryptophan cofactor is not present. The fitted monomer-dimer association constant is 2 µM.
Figure 2.5.5  Structural model showing steric clashes. LOV2 is shown in blue, TrpR is shown in orange, and DNA is shown as a gray surface. LOV2 or TrpR residues having steric clashes (atoms occupying the same space as atoms of other residues) are shown as red spheres.
We investigated the overall size and shape of LovTAP using synchrotron-based small angle x-ray scattering (SAXS). We recorded scattering profiles at 4, 8, and 16 µM LovTAP, both in the dark and under saturating photoexcitation. In all cases, the radius of gyration (R_g) is invariant at 29–30 Å (Fig. 2.5.6, Table 2). In decreasing concentration from 16 µM to 8 µM, the forward scattering intensity (I_0) decreases by a factor of two, as expected. However, further decreasing the concentration to 4 µM decreases I_0 by slightly more than a factor of two, perhaps reflecting an increased monomer concentration. Although photoexcitation does not cause a change in R_g, I_0 decreases by ~10%, indicating that some structural change does occur. The difference in I_0 may reflect a change in solvation, with the TrpR domain having more accessible surface area in the dark state than in the photoexcited state.

The experimental P(r) pair-distribution functions observed in the lit- and dark-states compare well with a distribution calculated from our working model (Fig. 2.5.6) [76, 77]. This agreement indicates that LovTAP is structurally compact in both states. However, the working model is not physically plausible due to the large regions of steric overlap between the LOV2 and TrpR domains. Furthermore, it fails to account for the partial unfolding of the shared helix as predicted by the helical lever-arm hypothesis. We asked whether lit-state unfolding of the Jα region of the shared helix is compatible with the observation of a compact structure in both the dark- and lit-states. Assuming the shared helix is the only linker between the LOV2 and TrpR domains, one might reason that a change from helix-coil transition of ~20 residues of this linker should be accompanied by a detectable change in R_g.

To address this discrepancy, Alexander Augustyn and I performed computational simulations of LovTAP. We used the initial model in a monte-carlo simulation to explore
**Figure 2.5.6** SAXS analysis of tertiary and quaternary structure. *A:* Guinier plots of dark and lit states. The black lines are fits for $R_g$ and $I_o$. *B:* $P(r)$ pair-distribution functions. The distributions calculated from the experimental data (red lines) are compared with those calculated from the ensemble models shown in Fig. 2.5.7 (yellow and blue lines) and with one calculated from the initial model shown in Fig. 2.5.5 (black line).
**Table 2  Small-angle scattering results.**

<table>
<thead>
<tr>
<th>[LovTAP]</th>
<th>$R_g$ (Guinier)*</th>
<th>$R_g$ (P(r))</th>
<th>$I_0$/conc.†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D</td>
<td>L</td>
<td>D</td>
</tr>
<tr>
<td>4 µM</td>
<td>29.9 ± 0.6 Å</td>
<td>28.7 ± 0.7 Å</td>
<td>29.2 ± 0.5 Å</td>
</tr>
<tr>
<td>8 µM</td>
<td>29.5 ± 0.2 Å</td>
<td>29.0 ± 0.2 Å</td>
<td>29.8 ± 0.2 Å</td>
</tr>
<tr>
<td>16 µM</td>
<td>29.2 ± 0.1 Å</td>
<td>29.0 ± 0.1 Å</td>
<td>29.8 ± 0.1 Å</td>
</tr>
</tbody>
</table>

* $R_g$ is calculated from both the Guinier analysis of the scattering intensity and the second moment of the P(r) pair-distribution function.

† The zero angle scattering, $I_0$, divided by the protein concentration is proportional to the molecular weight of the molecule, and is normalized to the dark-state value at [LovTAP] = 16 µM.
conformations of shared helix [78]. During the simulations, we held rigid the TrpR domain and the LOV domains, with the exception of the Jα region of the shared helix. We allowed the ($\phi$, $\psi$) angles of the Jα region to move in order to eliminate steric overlap. Conformations were accepted or rejected based on a statistical potential. We divided the resulting trajectory into ensembles of compact and extended structures, and calculated their P(r) functions (Fig. 2.5.7). The P(r) distribution of the compact ensemble agrees well with the experimental distributions, whereas that of the extended ensemble does not agree (Fig. 2.5.6). Interestingly, in the compact ensemble, the LOV domain appears to form an intramolecular binding interaction with the TrpR domain via the hydrophobic β sheet surface normally occupied by the Jα. Such a conformation is reminiscent of one proposed for LOV2 signaling in the full-length phototropin [64]. It is unlikely that two proteins that have not coevolved would interact with high enough affinity to be detectable in a dimeric system. However, such an interaction could occur between two tethered domains in a single polypeptide because the entropic cost of association is greatly reduced [79]. The interaction observed in the computational model occurs on a region of the surface of the TrpR domain distinct from the area occluded in the initial model. In reality, LovTAP may sample several such conformations—the modeling demonstrates, however, that the protein may remain compact while the Jα region of the shared helix unfolds as a loop.

The molecular envelopes reconstructed from the dark- and lit-state data also agree well with the compact ensemble model (Fig. 2.5.8) [80]. The LOV domains do not directly occlude the DNA binding surface of the TrpR domain in either state, reinforcing our proposed mechanism in which inactivation is accomplished by deformation of the TrpR domain.
Figure 2.5.7  Computational simulations of structural ensembles. A: Extended ensemble. B: Compact ensemble. LOV2 is shown in blue and TrpR is shown in orange. Bound DNA (gray surface) is shown for orientation, but was not included in the simulations or in subsequent calculations.
Figure 2.5.8 Ab initio reconstruction from SAXS data. A: Reconstruction from dark-state data. B: Reconstruction from lit-state data. C: Dark- and lit-state reconstructions docked onto the compact ensemble. Reconstructions are shown as green meshes. Otherwise, the colors are the same as in Fig. 2.5.7.
2.5.3 Conformational exchange and weak signal transfer

In a TROSY-HSQC NMR experiment, relatively few dispersed resonances are visible. More resonances are detectable at 37°C than at lower temperatures, perhaps reflecting faster tumbling. Overall, the LovTAP spectrum resembles that of the I539E variant of LOV2 [64]. This suggests that most of the dispersed resonances derive from the LOV domain, and that at 37°C, the shared helix is largely dissociated from the LOV domain in LovTAP (Fig. 2.5.9). At lower temperatures, the LOV domain may be exchanging between helix-bound and helix-dissociated conformations on an intermediate time scale. Since few, if any, dispersed peaks arise from the TrpR domain, it is likely that it undergoes intermediate conformational exchange at all temperatures.

We attempted to use electron paramagnetic resonance (EPR) spectroscopy as a way to probe light-induced conformational changes in the shared helix [81]. We introduced cysteine mutations at several sites on the shared helix and conjugated (1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate (MTSSL) spin labels at these sites. Overall, the EPR spectra acquired from these samples indicated that the sites were surface-exposed, as expected, and that photoexcitation caused only minimal changes in lineshape (Fig. 2.5.10). Spin labels placed at i, i+7 positions on the shared helix should encounter detectable changes in the interspin distance and geometry upon a helix-coil transition. The helix-coil transition may result in a smaller or larger interspin distance, depending on the conformational preference of the unfolded peptide. However, photoexcitation resulted in little spectral change for such probes. Biochemical RsaI protection experiments and CD spectroscopy data showed that incorporation of spin labels on the shared helix unfolds it in a manner similar to the helix-disrupting I532E, A536E and I539E mutations (data not shown) [64]. Therefore, in wild-type LovTAP, the
**Figure 2.5.9** NMR studies of dark-state conformation. A: HSQC spectra taken at three temperatures (11, 30, and 37 °C). B: The 37 °C spectrum (green) has numerous resonances in common with the dark-state spectrum of I539E AsLOV2 (right panel, black contours), but not with the lit-state I539E spectrum, or with either wild-type spectrum. (AsLOV2 spectra reproduced from [64]. The 37 °C spectrum of LovTAP was scaled linearly in the proton and nitrogen dimensions to best overlay.)
Figure 2.5.10  EPR studies of light-induced conformational changes. A: Dark-state (red) and lit-state (pink) spectra for spin labels placed a position 523 and 530. B: Dark- and lit-state spectra for spin labels placed a position 530 and 538. Colors same as in A. Both sets of spectra are compared with a dark-state spectrum for a spin label placed at position 530 (black).
shared helix is only weakly associated with the LOV domain in the dark, and this interaction is easily disrupted by mutations. This weak LOV-helix association suggests that the poor photoswitching in LovTAP is predominantly due to poor dark-state suppression.

2.6 Materials and methods

Cloning, expression and purification

The DNA fragment encoding *E. coli* TrpR was obtained by colony PCR. A clone of *A. sativa* phot1 LOV2 was generously provided by Dr. Kevin Gardner. The fusion proteins were created from these templates using overlap extension PCR and subcloned into the expression vector pCal-n (Stratagene) so as to be in frame with the amino terminal calmodulin binding peptide. The results were confirmed by DNA sequencing.

Rosetta 2(DE3)pLysS *E. coli* (Novagen) carrying these vectors were grown in LB media at 37° C to an OD$_{600}$ ~ 0.5 and induced at 20° C by the addition of 0.3 mM IPTG. After ~ 16 hours, the cells were harvested by centrifugation and resuspended in B-buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 1 mM imidazole, 1 mM MgCl$_2$ and 2 mM CaCl$_2$, 10 mM β-mercaptoethanol and EDTA-free protease inhibitor cocktail (Roche)). The cells were lysed by the addition of lysozyme and deoxyribonuclease-I. The lysate was cleared by centrifugation and applied to equilibrated calmodulin affinity resin (Stratagene). The resin was washed with W-buffer (50 mM Tris, pH 8.0, 500 mM NaCl, 1 mM imidazole, 1 mM MgCl$_2$ and 2 mM CaCl$_2$) and the protein eluted in 3-5 mL with E-buffer (50 mM Tris, pH 8.0, 500 mM NaCl, 2 mM EGTA). The eluate was applied to a HiPrep 16/60 Sephacryl S-100 HR column (GE Healthcare) wrapped with aluminium foil and equilibrated with G-buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 1 mM EDTA) running at 1 mL/min. Fractions
corresponding to peaks in $A_{447}$ were collected and analyzed by SDS-PAGE. Concentration was determined by $A_{447}$ using $\varepsilon = 13,800 \text{ M}^{-1}\text{cm}^{-1}$ for the monomer [82].

For the NMR experiment, the DNA fragment encoding LovTAP was subcloned into a modified version of pET-28a. The vector was modified by mutating the Gly-Ser codons in the thrombin recognition sequence to a BamHI restriction site, and the LovTAP DNA fragment was cloned between the new BamHI site and HindIII, resulting in a LovTAP construct with an amino-terminal polyhistidine tag in place of the calmodulin binding peptide. BL-21(DE3) E. coli carrying this plasmid was grown in M9 media, with $^{15}$NH$_4$Cl as the sole nitrogen source, at 37° C to an OD$_{600}$ ~ 0.5 and induced at 20° C by the addition of 0.3 mM IPTG. After ~ 16 hours, the cells were harvested by centrifugation. Cell pellets from 3 L of culture were resuspended in 25 mL buffer (50 mM Tris, pH 8.0, 500 mM NaCl, 10 mM imidazole, 10 mM $\beta$-mercaptoethanol and EDTA-free protease inhibitor cocktail (Roche) and 2 mM PMSF). The cells were lysed by sonication and the extract was cleared by centrifugation. The cleared extract was applied to a 5 mL HisTrap HP column (GE Healthcare), washed with lysis buffer and eluted over a 12.5–250 mM imidazole gradient. The FMN-containing fractions were pooled, concentrated, and further purified by gel filtration (HiPrep 16/60 Sephacryl S-100 HR column (GE Healthcare) wrapped with aluminum foil and equilibrated with 50 mM KPO$_4$, pH 6.4, 100 mM KCl, 1 mM EDTA). The FMN-containing fractions were pooled and concentrated to 150 µM.

**Illumination**

Blue AlGaInP LEDs (theledlight.com, 20° viewing angle, 8000 mcd, 468 nm $\lambda_{\text{max}}$ at 3.4 V) were powered by a 3.3 V, 4 A AC adaptor (Mouser). The radiant power delivered to the
samples was estimated using a hand-held power meter (New Focus). Irradiance was calculated using 0.12 cm$^2$ as the cross sectional area of the reaction tube.

**Dark-state recovery**

Samples in buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 1 mM EDTA) were bleached using light from a 75 W xenon arc source (Oriel). A liquid light guide (Oriel) was used to direct the illumination onto the sample. Recovery was monitored by $A_{447}$ using a Hewlett Packard 8452A diode array spectrometer.

**β-galactosidase reporter assay**

DNA fragments encoding LOV2-TrpR constructs were subcloned into the vector pZA32 [83] and the resulting plasmids were transformed into the *E. coli* strain CY15058 [71] obtained from the *E. coli* Genetic Resource center (http://cgsc.biology.yale.edu). Overnight cultures were diluted into minimal media [69] and grown either in the dark or under broad spectrum fluorescent lights. β-galactosidase activity was measured as described [84].

**RsaI Protection Assay** [72].

The KpnI/RsaI site of plasmid pUC19 was deleted and the duplex oligos 5’–AATCGAACTCGCTAGCGAGTACG–3’ and 5’–ATTGTACTCGCTAGCGAGTTCG–3’ (IDT) were ligated into the resulting plasmid between the TfiI sites. The deletion and insert were confirmed by restriction analysis and sequencing. The resulting plasmid, pUC19-KpnIΔ-trpR$_s$(3A), contains two natural RsaI sites and one site buried in a trp operator that binds only one TrpR dimer [85]. TrpR activity protects the 1890 bp fragment
from digestion by RsaI to the 1410 bp and 480 bp products. A 676 bp internal control results from digestion at unprotected sites. For simplicity, only the 1890 bp and 1410 bp fragments are shown in the figures. Protection assays were performed with 19 nM pUC19-KpnIΔ-trpR₃(3A), 1X NEB Buffer 2 (10 mM Tris·HCl, pH 7.9, 10 mM MgCl₂, 50 mM NaCl, 1 mM DTT), 0.1 mg/mL BSA (Promega), 0.1 mM L-tryptophan (Sigma). Protein in G-buffer, G-buffer supplemented with 350 mM NaCl, or E-buffer, was diluted 10-fold into this mixture and incubated for 1 minute under ambient conditions. RsaI was diluted to a concentration of 0.5-2 U/μL in dilution buffer (50% v/v glycerol, 1X NEB Buffer 2, 0.2 mg/mL BSA) and then diluted 20-fold into the reaction mixture. Reactions were performed in 0.5 mL thin walled tubes placed in the block of a thermocycler set at 25° C. The reaction was quenched at the desired time points by adding loading buffer (Promega) to 1X and EDTA to 15mM. The samples were run on a 1.5% agarose gel in 1X TAE, stained with ethidium bromide and imaged. ‘Digestion’ was determined as the ratio of the intensity of the product bands (1410 bp and 480 bp) to that of the product bands plus the reactant band (1890 bp). The bands were quantified using ImageJ (http://rsb.info.nih.gov/ij/), and the results plotted using Kaleidagraph (Synergy Software).

Modified RsaI Protection Assay

The plasmid pUC19-KpnIΔ-trpR₃(3A) was digested with PvuI and the resulting linear fragment containing only a single RsaI restriction site buried in the trp operator was gel-purified. This 1680 bp reactant is digested to a 375 bp and a 1305 bp product by RsaI. Other than the substitution of DNA substrate, the assay was performed essentially as described above. The amount of digestion was determined as the ratio of the intensity of the product bands to that of the product bands plus the reactant band. The bands were
quantified using ImageJ (http://rsb.info.nih.gov/ij/), and the data was analyzed using IGOR Pro (WaveMetrics, Lake Oswego, OR). For all fits, the amplitude was constrained to unity and only the rate was allowed to vary. The dissociation constant ($K_d$) in the lit and dark states was calculated using the equation:

$$K_d = \frac{L}{(R - 1)}$$

where $L$ is the concentration of LovTAP and $R$ is the ratio of the intrinsic rate of RsaI digestion ($k_0$) and the rate in the presence of LovTAP ($k_{obs}$).

**Circular Dichroism (CD) Spectroscopy**

CD measurements were performed using a Jasco J-715 spectropolarimeter equipped with a temperature controlled cuvet holder. The conditions used were 1–7 μM protein in 150 mM KPO$_4$, pH 7.0, 1 mM EDTA, 25° C, 0.1 cm path length cuvet. Wavelength scans were the average of 10 measurements taken at 1 nm increments with a 2 nm bandwidth, 4 s averaging time and a speed of 20 nm per minute. Data were converted to mean residue ellipticity and plotted using Kaleidagraph. For kinetic experiments, illumination was from a single blue LED connected to an external switch. The sample was illuminated for 30 s to saturated photoexcitation, then the LED was switched off and data recorded for 250 s. CD at 222 nm and at 207 nm were recorded in triplicate, averaged, and fit to a single exponential function using IGOR Pro.

**Structural Modeling**

Atomic coordinates of *E. coli* TrpR (PDB ID: 1TRR) and *A. sativa* phot1 LOV2 (PDB ID: 2VOU) were analyzed in DeepView. Structural models were built by aligning the $C_\alpha$ atoms of several residues at the end of the LOV2 Jα helix with the $C_\alpha$ atoms of same
number of residues on the amino-terminal helix of TrpR, immediately preceding the intended junction. To illustrate the degree of steric overlap accompanying a continuous shared helix, no further adjustments to the structure were made.

*Size Exclusion Chromatography*

For analysis of the concentration dependence of elution, 0.5 mL of sample in G-buffer was applied to a HiPrep 16/60 Sephacryl S-100 HR column, wrapped with aluminium foil to exclude light, equilibrated with G-buffer and running at 1 mL/min. The elution was monitored by $A_{447}$.

*Analytical ultracentrifugation*

Experiments were performed in a Beckman Optima XL-A using a 6-channel centerpiece. The protein concentration, $C(r)$, was measured using $A_{447}$. Data were analyzed using SEDANAL [75].

*Small Angle X-ray Scattering*

All SAXS data were collected at the BioCAT beamline at the Advanced Photon Source, Argonne National Laboratory. LovTAP samples were prepared to 0.125, 0.25, and 0.5 mg/mL in 50 mM Tris, pH 8.0, 500 mM NaCl, 1 mM EDTA, and the sample and buffer blanks were filtered to 0.22 µm. Buffer blanks and samples were flowed through a 1.5 mm capillary at a rate of 2 µL/s and 15 exposures of ~ 1 s duration were collected. The exposures were averaged, buffer blanks subtracted, and Guinier analysis performed using IGOR Pro. $I(Q)$ data were converted to $P(r)$ data using GNOM [76]. Experimental $P(r)$ data were compared to the distribution calculated from our dark state model plus amino-
terminal calmodulin binding peptide tags modeled as a random walk [77]. Ten *ab initio* reconstructions were generated from the [LovTAP] = 16 μM data using DAMMIN and averaged using DAMAVER [80, 86]. No assumptions about the final shape were enforced in the reconstruction, save for a 100 Å D_max determined in the GNOM P(r) calculation. DAMMIN reconstructions were performed using data collected at [LovTAP] = 4, 8, and 16 μM, and the shape was found to be independent of concentration.

*NMR spectroscopy*

TROSY-HSQC spectra were acquired on a Varian Unity Inova 600 MHz spectrometer. nmrPipe and nmrViewJ were used for data processing and analysis [87, 88]. Three spectra were acquired, over 12 hours each, at 11, 30 and 37 °C.

*EPR spectroscopy*

Protein was purified by calmodulin affinity chromatography. Before elution, the sample was washed with binding buffer containing 5 mM TCEP and washed again to remove the TCEP. Immediately after elution, a 10-fold excess of (1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate (MTSSL) was added and the sample was incubated on ice for 1 hour. Excess MTSSL was removed by gel filtration.

Continuous-wave spectra were acquired on a Bruker EMX spectrometer. Illumination was from a single LED and directed onto the sample using a 0.4 mM, high NA optical fiber.
Chapter 3

Improving signal transfer by mutating the Jα helix

The data presented in this chapter are unpublished. I thank Grzegorz Gawlak for assistance purifying proteins and performing experiments, and Xiaolan Yao, Michael Rosen and Kevin Gardner for stimulating discussions.

3.1 Energetic model of signal transfer

In isolated LOV2 under both dark and lit conditions, the Jα helix detectably populates both the folded, LOV-docked conformation and the locally unfolded, undocked conformation. The population fraction of the undocked conformation, however, shifts dramatically from 1.6% in the dark state, to 91% in the lit state (Fig. 1.5.1) [43]. In Arabidopsis phototropin 1, the unfolding of the Jα helix from the LOV2 domain has been shown to activate the serine/threonine kinase domain. Although the mechanism by which this activation occurs remains unresolved, it has been proposed that helix unfolding exposes a hydrophobic surface on the Jα or on its binding surface on LOV2, that participates in inter- or intra-molecular binding interactions with the kinase domain [64].

These efforts towards understanding signal transduction in AsLOV2 have paved the way for the facile use of the domain as a design tools [44, 89]. Nevertheless, these initial attempts at designed photoswitches suffer from a low dynamic range. A more efficient coupling between the conformational changes in LOV2 and the output domains would greatly enhance the usefulness of these proteins.
Incorporation of the LOV domain into a functional signaling protein, natural or designed, introduces competition for the Jα helix or for its binding surface on LOV2 (Fig. 3.1.1). This competition reduces the population of molecules in which the helix is docked onto LOV2. Compensating mutations to the LOV domain or the helix that increase the LOV-Jα affinity should repopulate the LOV-docked conformation. We use the dimensionless parameter $x_{\text{eff}}$ to represent the stabilization of LOV-Jα docking imparted by mutation or fusion. $x_{\text{eff}} = x_{\text{mut}} x_{\text{fus}}$, where

$$x_{\text{mut}} = \frac{K_{\text{mut}}^{\text{wt}}}{K_{\text{mut}}^{\text{mut}}}$$

and

$$x_{\text{fus}} = \frac{K_{\text{fus}}^{\text{wt, iso}}}{K_{\text{fus}}^{\text{mut}}}$$

reflecting stabilization or destabilization due to mutations or fusion (e.g., with TrpR). The superscripts wt, mut, iso, and fus refer to wild-type, mutated, isolated and fused LOV domains, respectively. A number of mutations that destabilize the LOV-Jα interaction have been described—for these mutations, $x_{\text{eff}} = x_{\text{mut}} < 1$ [64]. The equilibrium constant of helix undocking, $K_{\text{helix}}$, is divided by $x_{\text{eff}}$ to account for these modifications. We assume for simplicity that $x_{\text{eff}}$ is the same in the lit and dark states.

The observed affinity constants for DNA binding in the dark- and lit-states, respectively, are the intrinsic DNA binding affinity, $K_{\text{DNA}}$, in the active conformation multiplied down by the fraction of the population which is binding competent:

$$K_{\text{obs}}^{\text{lit}} = \frac{K_{\text{DNA}}}{1 + \frac{K_{\text{obs}}^{\text{lit}}}{K_{\text{helix}}} x_{\text{eff}}}$$

and

$$K_{\text{obs}}^{\text{dark}} = \frac{K_{\text{DNA}}}{1 + \frac{K_{\text{obs}}^{\text{dark}}}{K_{\text{helix}}} x_{\text{eff}}}$$

(1, 2)

The ratio of the observed DNA binding activity of the effector in the lit- and dark-states is:
Figure 3.1.1  Model of activation and DNA binding. LovTAP is in equilibrium between an inactive conformation (left), in which the shared helix is docked on the LOV domain (blue), and an active conformation (center), in which it is docked on the TrpR domain (orange). Due to competition by TrpR for docking of the helix, the microscopic equilibrium constant of this reaction, $K_{\text{helix}}$, is increased by a factor $1/x_{\text{eff}}$. As in the isolated LOV domain, photoexcitation shifts this equilibrium to the right by 600-fold. In the active conformation, LovTAP binds DNA (right) with an intrinsic association constant $K_{\text{DNA}}$. 
Thus, mutations that increase $x_{\text{eff}}$ increase $\beta$ up to the limiting value of 600 [43].

In LovTAP, the ratio of lit- to dark-state DNA binding activity is ~6. This value equates to $x_{\text{eff}} = 0.075$ (Eq. 1). For the dark-state, then, $K_{\text{helix}} / x_{\text{eff}} = 0.017 / 0.075 = 0.23$, meaning that the inactive conformation is only ~20% of the total population in the dark. We reasoned that dark-state suppression could be improved by promoting docking of the helix with the LOV domain. Ideally, one would like to increase only the dark-state docking of the helix, so that the lit state activity remains high. However, our model predicts that the system is in a regime where $\beta$ increases rapidly with $x_{\text{eff}}$ (Fig. 3.1.2). Thus, even if our assumption that $x_{\text{eff}}$ is the same in the lit and dark states holds true, mutations which increase $x_{\text{eff}}$ should improve dark-state suppression.

### 3.2 Improving photoswitching by rational mutation

#### 3.2.1 Analysis of helix-stabilizing mutations

Our initial efforts to increase the J$\alpha$ helix affinity to the body of the LOV2 domain focused on increasing the intrinsic helix propensity. Because the helix is unfolded when it is undocked, helix-stabilizing mutations should increase the docking infinity. We used the program AGADIR [90] to identify point mutations in the J$\alpha$ region of the shared helix that would be likely to increase its helical propensity. We calculated the helical propensity of the J$\alpha$ helix for single point mutations (Fig. 3.2.1). With the exception of Gly528, we did not include positions identified by Harper et al. as part of the conserved hydrophobic interface between the J$\alpha$ helix and the $\beta$ sheet [40]. We also did not calculate scores for...
Figure 3.1.2  Analytical model of photoswitching. Equations 1, 2 & 3 are plotted as functions of $x_{\text{eff}}$. In the upper panel, $K_{\text{obs}}$ is plotted as a function of $x_{\text{eff}}$. $K_{\text{DNA}}$ is set equal to the lit-state $K_{\text{obs}}$ for wild-type LovTAP. In the lower panel, $\beta$ is plotted as a function of $x_{\text{eff}}$. 

\[ K_{\text{obs}} = \frac{K_{\text{DNA}}}{1 + \frac{x_{\text{eff}}}{K_{\text{helix}}}} \]

\[ x_{\text{eff}} = x_{\text{fus}} x_{\text{mut}} \]

\[ \beta = \frac{1 + \frac{x_{\text{eff}}}{K_{\text{helix}}}}{1 + \frac{x_{\text{eff}}}{K_{\text{helix}}}} \]
Figure 3.2.1 AGADIR analysis and modeling of point mutants. A: Histogram of predicted helicity (% helix) for Jα point mutations. The wild-type helicity is 1.53% (dashed red line). The positions of several mutations in the histogram are indicated. The mutations R526D, G528A and N538E (black letters) were tested in this study. A, inset: Helicity predicted at the residue level for the R526D, G528A and N538E mutations. B: The R526D, G528A and N538E mutations (CPK spheres) shown on a model of the LOV domain (gray surface; Jα, blue ribbon).
mutations to proline or cysteine. AGADIR returned a value of 1.53% for the average helicity for the 21 residues of the Jα helix (residues 523–543). 158 out of 187 single mutations were within 0.7% of this value. Of these mutations, 78 were stabilizing and 79 were destabilizing.

Of the remaining 29 mutations, all increased helicity by more than 0.7%. 17 replaced the intrinsically helix-destabilizing glycine at position 528. Gly528 is on the hydrophobic interface between the Jα, and the position is one of the most conserved in LOV2 Jα helices. Both points suggest that the position may not tolerate mutations. However, Harper et al. observed a helix-docked structure in a G528E variant of the LOV2 domain, and inspection of the high-resolution crystal structure suggests that the position could accommodate some rotamers of non β-branched amino acids without excessive steric clashes (Figure 3.2.1) [62]. AGADIR scores the G528A mutation at 6.45% helicity. Although this value is considerably less than for a G528L substitution (15.9%), we reasoned that the mutation to alanine is least likely to introduce steric clashes.

Aλer Gly528, mutations at Arg526 showed the next largest increases in helicity according to AGADIR (Fig. 3.2.1). 14 out of 17 mutations at this position showed some improvement, with the largest value of 3.6% for R526D. This position is surface-exposed in the crystal structure of LOV2, and is not conserved in LOV2 domains, suggesting that mutation would be unlikely to introduce unfavorable interactions (Fig. 3.2.1).

The R526D and G528A mutations are both near the amino-terminal end of the Jα helix. We wanted to test mutations near the carboxy-terminus. Mutations of Asp540 and Glu541 to arginine show the greatest improvement in AGADIR scores, at 2.49% helix and 2.95% helix, respectively. However, these residues contact the TrpR domain in LovTAP, and the mutations might have unpredictable effects on the domain's structure. N138,
further away from the TrpR domain, is less likely to introduce complicating interactions. N538E, at 1.99% helix, is the best-scoring mutation at this position (Fig. 3.2.1).

3.2.2 Mutations decrease dark-state DNA binding

In LovTAP we found that the G528A mutation produced a 2-fold decrease in dark-state DNA binding but no change in the lit-state binding (Fig. 3.2.2). This is in agreement with our model prediction that there should be no change in DNA binding in the lit-state, regardless of whether the mutation increases lit-state helix docking. The N538E mutation causes a 3.8-fold decrease in the dark-state DNA affinity, and, like the G528A mutation, does not change the lit-state affinity (Fig. 3.2.2). This larger-than-expected dark-state change may be explained by the tertiary interactions of the mutated residue. The glutamic acid can form a salt bridge with Lys413 of LOV2, which may promote helix docking to the LOV domain (Fig. 3.2.1). Together, the G528A and N538E mutations decrease dark-state DNA binding affinity by 9-fold, without changing the lit-state binding (Fig. 3.2.2).

In contrast to G528A and N538E, the R526D mutation does not change the dark-state DNA affinity and slightly lowers the lit-state DNA affinity (not shown). The double mutant R526D/M530R, with a score of 8.47% helix, would be expected to introduce a favorable salt-bridge into the exposed surface of the helix—yet it has no effect on DNA binding (not shown). Based on these mutations, it appears that the AGADIR-predicted helical propensity is not a good predictor of dark-state suppression. This may be because any increase in helical propensity stabilizes the inactive (LOV-bound) and active (TrpR-bound) conformations equally. Inspection of the LOV2 structure suggests that stabilizing LOV-Jα tertiary interactions may be more important than helical propensity. Further improvement in the dynamic range of photoswitching through improving LOV-Jα
**Figure 3.2.2** DNA binding by Ja mutants. A: Effects of the G528A and N538E mutations. In the upper panel, observed affinity constants ($K_{obs}$) are plotted against $x_{eff}$. The data are compared to the model presented in Fig. 3.1.2 (black lines) and to a similar model in which both LOV domains must be in the active (undocked) conformations (gray lines). For wild-type LovTAP, $x_{eff} = x_{fus}$ is calculated using equation 3. For mutants, $x_{mut}$ is determined by CD in the lit state, and $x_{eff} = x_{mut} x_{fus}$. In the lower panel, $\beta$ is plotted as a function of $x_{eff}$. B: The experimentally measured dark-state $K_{obs}$, plotted against the dark-state $K_{obs}$ predicted from CD measurements. The predicted values are calculated assuming that $x_{mut}$ is the same in the lit and dark states. The identity line is plotted for reference. C: Comparison of predicted and actual dark-state $K_{obs}$ values, where the dark-state $x_{mut}$ values differ from the lit-state values by a factor of two. D: Comparison of predicted and actual dark-state $K_{obs}$ values, where $K_{obs}$ is predicted using the single- and double-undocking models.
docking will likely come from more sophisticated computational methods that can account for these tertiary interactions, or from library selection experiments aimed at maximizing dark-state suppression.

3.2.3 Mutations increase lit-state LOV-Jα docking

We asked whether the observed decreases in dark-state DNA binding derive from stabilization of the LOV-Jα docking interaction. CD spectroscopy can accurately measure $K_{\text{helix}}$ values in the range of 0.1–10. Within this range, the fractional populations of helix-docked molecules is between ~90% and ~10%, and CD is sharply dependent on $K_{\text{helix}}$. Outside this range, $K_{\text{helix}}$ can vary considerably, say from 100–1000, with no observable change in CD. It would therefore be difficult to measure dark-state stabilization directly ($K_{\text{helix}} < 0.017$). However, if our assumption that mutations stabilize the lit and dark states equally, then we can reliably measure the dark-state stabilization in the lit state (0.1 < $K_{\text{helix}}$ < 10). Comparing dark-state DNA affinity values predicted from these measurements with observed values would also serve as a stringent test of the assumptions underlying our model.

In the dark, all variants of LOV2 have a CD signal at 222 nm ($[\Theta]_{222}$) of −5900 deg·cm²·dmol⁻¹, corresponding to 98.5% of the molecules having the Jα helix docked ($K_{\text{helix}} = 0.017$, Fig. 3.2.3). Under illumination, wild-type LOV2 has a $[\Theta]_{222}$ of −3253 deg·cm²·dmol⁻¹, corresponding to 9% of the molecules having the Jα helix docked ($K_{\text{helix}} = 10$). For illuminated LOV2 G528A, $[\Theta]_{222} = −3782$ deg·cm²·dmol⁻¹. By linear interpolation from the wild-type data, 27% of the molecules have the Jα helix docked ($K_{\text{helix}} / x_{\text{mut}} = 2.7$, $x_{\text{mut}} = 3.7$, $x_{\text{eff}} = 0.28$). By similar calculations, for the N538E mutant $x_{\text{eff}} = 0.44$ and for the G528A/N538E double mutant $x_{\text{eff}} = 0.95$ (Table 3).
**Figure 3.2.3** CD of Jα mutants. *A:* Representative traces of kinetic recovery from photoexcitation, as monitored by CD at 222 nm. Wild-type LOV2 and the N538E mutant are shown. Fits are shown as solid lines. *B:* Fitted parameters from kinetic recovery, plotted against the population percentage of molecules with the Jα helix docked. The solid line is determined by the wild-type parameters, with the fully-photoexcited CD set to 9% docked, and the fully-recovered CD set to 98.5% docked. The percent populations for the mutants are calibrated using this line, and used to calculate $x_{mut}$. 

\[
\begin{align*}
N538E & \quad x_{mut} = 5.85 \\
G538A / N538E & \quad x_{mut} = 12.6 \\
WT \text{ lit} & \quad x_{mut} = 3.68 \\
WT \text{ dark} & \quad x_{mut} = 12.4 
\end{align*}
\]
Table 3  \( \alpha \) helix stabilization results.

<table>
<thead>
<tr>
<th>mutant*</th>
<th>([\Theta]_{222})</th>
<th>% ( \alpha ) docked†</th>
<th>( K_{\text{helix}} / x_{\text{mut}} )</th>
<th>( x_{\text{mut}} )</th>
<th>( x_{\text{eff}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt (dark)</td>
<td>-5903</td>
<td>98.5</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>wt (lit)</td>
<td>-3253</td>
<td>9</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G528A</td>
<td>-3782</td>
<td>26.9</td>
<td>2.72</td>
<td>3.68</td>
<td>0.28</td>
</tr>
<tr>
<td>N538E</td>
<td>-4077</td>
<td>36.9</td>
<td>1.71</td>
<td>5.85</td>
<td>0.44</td>
</tr>
<tr>
<td>G528A / N538E</td>
<td>-4642</td>
<td>55.9</td>
<td>0.79</td>
<td>12.6</td>
<td>0.95</td>
</tr>
</tbody>
</table>

* Data represents lit-state values, unless otherwise specified.
† The mean residue ellipticity at 222 nm, in deg·cm\(^2\)·dmol\(^{-1}\).
‡ Equal to \((|\Theta|_{222} + 2986.2) / -29.6\).
Using our model, we calculated dark-state DNA binding affinities from the CD-derived values of $x_{\text{eff}}$. The calculated values agree with the measured values within the experimental error of the latter (Fig. 3.2.2). Initially, we assumed that mutations stabilize helix docking in the lit and dark states equally. Our results appear to validate this assumption, as we are able to use stability parameters measured in the lit state to predict dark-state DNA affinities. To test whether differences in lit- and dark-state stabilization would be detectable in this experiment, we re-calculated the DNA binding affinities using alternative assumptions: that the helix-docking stability imparted by the mutations is two-fold greater in the dark than in the light, and that it is two-fold smaller. We found that these predicted affinities did not agree well with the measured values (Fig. 3.2.2).

We also asked whether an alternative model of photoswitching might agree with our data equally well. The model presented above does not account for the possibility that LovTAP might only be fully activated once both shared helices undock from their respective LOV domains. We created an alternative model in which the helices undock independently with equilibrium constants $K_{\text{helix}} / x_{\text{eff}}$. This model, where

$$K_{\text{obs}} = \frac{K_{\text{DNA}}}{\left(1 + \frac{x_{\text{eff}}}{K_{\text{helix}}} \right)^2}$$

is plotted in Figure 3.2.2. The lit-state DNA binding data agree with both models for the range of $x_{\text{eff}}$. However, the model does not agree with the dark-state data. We used the model to predict $K_{\text{obs}}$ values from the CD-derived values of $x_{\text{eff}}$. Again, these predicted values do not agree with the observed values (Fig. 3.2.2).
3.3 Discussion

We have identified two mutations to the LOV2 domain of *A. sativa* phototropin 1 that stabilize the docking of the Ja helix. We have shown that this stabilization improves photoswitching in a designed protein, and we have presented an analytic model that quantitatively accounts for this improvement. Our model is based on the idea that fusion of the LOV and TrpR domains in LovTAP creates a competition between the two domains for binding of this helix. The binding of the Ja is only marginally stable, and competition effectively depopulates it even in the dark state. Because dark-state suppression of DNA binding requires docking of this helix, the original design is nearly fully active in the dark.

The Ja helix has a low intrinsic helicity as predicted by AGADIR. In principle, the docking stability of the Ja could be improved by increasing the intrinsic helicity. This could be accomplished by reducing the entropy of the unfolded conformation, i.e., with a Gly-to-Ala mutation, or by introducing electrostatic interactions between adjacent turns of the helix, i.e., with an $i, i+4$ salt bridge. The docking stability could also be increased by introducing stabilizing tertiary contacts, such as salt bridges or by improving packing of hydrophobic residues between the Ja helix and the $\beta$-sheet.

Although both mutations substantially improve photoswitching, the G128A mutation is predicted by AGADIR to have a substantially increased intrinsic helicity, while the N138E mutation is not. Inspection of the mutations in the LOV2 structure suggests that both may stabilize tertiary interactions between the Ja helix and the $\beta$-sheet. A third mutation, R526D, is also predicted to increase helicity, but does not make any tertiary contacts in the LOV2 structure, and does not improve photoswitching. While we cannot rule out the possibility that increased helical propensity contributes to the increased
stability of the G528A mutation, it appears that stabilizing tertiary interactions are more effective for improving photoswitching. Increased helix propensity may stabilize both the LOV-docked and TrpR docked conformations, leading to a net-zero effect on photoswitching.

The LOV-Jα docking-undocking reaction likely contributes to the modulation of effector activity in the natural phototropin and in the synthetic LOV2-based photoswitch described by Lee, Natarajan, et al. [44, 64]. Although the latter authors did not examine the role of this conformational change in photoswitching, it is reasonable to suppose that it occurs, and Ostermeier has described how such a change could mediate allosteric control in domain-insertion architectures [16]. Because the mutations described here are in the LOV domain and exert their effects there, they are likely to improve most, if not all, designed LOV2-based photoswitches.

Natural cellular signaling systems frequently employ feedback, amplification, and multiple layers of regulation [91-93]. Synthetic systems can be similarly designed to integrate multiple signals and respond with switch-like behavior, even at the level of an individual protein [19, 94]. However, these criteria may not always be desirable for molecules that are to be used in in vivo experiments. In the lab, one would like to change a single, convenient variable and thereby generate a continuously graded perturbation with a large dynamic range. In principle, synthetic LOV2-based photoswitches can be created that, when illuminated, respond with at least a 600-fold increase in effector activity. In LovTAP, we have increased the dynamic range of photoswitching from 6-fold to 70-fold by making two mutations that together stabilize the LOV-Jα interaction by 17-fold. The introduction of one or more additional mutations that independently stabilize the LOV-Jα
interaction by the same amount would increase the dynamic range to nearly its limiting value of 600-fold.

3.4 Materials and Methods

Prediction and structural analysis of mutations.

Lists of mutated LOV2 Jα helix (residues 523-543) sequences were generated in Excel. All single mutations except cysteine and proline for residues 525–528, 530, 531, 534, 537, 538, 540, and 541 were included. The sequences were submitted to AGADIR (http://www.embl-heidelberg.de/Services/serrano/agadir/agadir-start.html) and scored with the following settings: pH = 7, temp. = 278, ionic strength = 0.1, Nterm = acetylated, Cterm = amidated. The scores were plotted and analyzed using Igor Pro.

Modified RsaI protection assay

RsaI protection assays were performed essentially as described in Chapter 2.

Preparation of LOV2 samples

LOV2 (404-560) samples were prepared essentially as described in [40], and exchanged into assay buffer (10 mM NaPO₄, pH 8.0, 65 mM NaCl, 0.1 mM EDTA) by gel filtration.

CD spectroscopy

CD dark-state recovery experiments were performed as described in Chapter 2.
Chapter 4

Conclusions and future directions

I thank Bernard Liu for helpful discussions on Cbl, and Michael Glotzer for helpful discussions regarding the applications of designed photoswitches.

4.1 Interdomain signal transduction through linkers

4.1.1 Signal transduction through helical linkers

Our successful design of an allosteric lever arm and a bi-stable energy surface, along with the observation of natural analogues, suggest the existence of a general but largely unrecognized mode of connecting modular domains into a functionally integrated whole.

The amino-terminus of Cbl (Cbl-N), an archetypally modular protein, includes an SH2 domain, an EF-hand domain and a 4-helix bundle domain, in a compact, integrated structure [95]. Mutation of calcium ligands of the EF-hand reduces phosphopeptide binding, indicating allosteric coupling between the SH2 and EF-hand domains. A non-conserved, shared helix connects the SH2 domain and the EF hand domain, packing against both. Phosphopeptide binding rotates the SH2 domain towards the 4-helix bundle domain, twisting the shared helix and subtly distorting the EF hand domain.

More generally, coiled-coil linkers are recognized as an important motif in interdomain signal transduction [96]. Like AsLOV2, the LOV domain from the bacterial photosensor YtvA has a carboxy-terminal Jα helix. However, the LOV domain is dimeric, and the Jα helix, rather than docking against the β sheet, forms an arm like coiled-coil extension. Modeling and design studies, in which the LOV domain was fused to a
heterologous carboxy-terminal output domain, suggest that interdomain signaling occurs through this coiled-coil extension [97, 98]. The crystal structure of a the photosensory portion of *Pseudomonas aeruginosa* bacteriophytochrome reveals a coiled-coil backbone at the dimerization interface [99]. Modeling suggests that this coiled-coil is part of a more complex multi-helix structure linking the sensory and catalytic domains. Analysis of structural variation within the crystal suggests that conformational changes in the inter-domain helices may play a role in signal propagation.

### 4.1.2 Leashes and levers

We can separate domain linkers into two categories. In one category, allosteric effects result from intramolecular binding between domains connected by domain linkers, or “leashes,” of undefined structure [18]. A statistical mechanical treatment of this type of protein by Van Valen and coworkers employs a wormlike chain model for the linkers [100]. Within this framework, the sequence and even the precise conformation of the linker are largely neglected, and only the loop closure of the chain is considered as a condition for activation or deactivation. Naturally, loop closure must have some structural consequence, such as deformation of a binding surface, in order to lead to switching behavior [100].

Our model of LovTAP switching does not explicitly postulate loop closure in the same sense as the model of Van Valen *et al.* One might consider the intermolecular binding and loop unfolding observed in our computational simulation in the terms of a wormlike chain model. The synthetic photoswitch designed by Lee, Natarajan and coworkers might also have some features amenable to such treatment [44]. Nevertheless, it seems likely that the sequence and conformational details of the entire linker are important in both
systems, and there is no evidence that the primary effect of varying linker length is to modulate loop-closure entropy. Thus, we consider the linker as a helical lever-arm. Because a regular helix is resistant to bending and twisting, it can function as an rigid lever to transmit forces created by interdomain contacts, thereby generating a bistable system. In addition to the examples discussed here, we anticipate more examples of interdomain signal transduction through helices will be found in naturally-occurring proteins.

4.2 Future directions in optical control of proteins

Scientific discovery follows technological developments that make new experiments possible. One of the most striking examples has been the use of green fluorescent protein (GFP) and its variants. Biologists had been looking through microscopes to learn about the cellular world for hundreds of years before to the introduction of GFP—nevertheless, GFP fundamentally changed the way biologists use microscopes to study cells [101-103]. GFP has been used as a reporter of gene expression, to monitor the localization of single proteins in cells, to show their colocalization, and to quantify their dynamics. In contrast to prior fluorescence techniques, GFP can be used in live cells without the use of antibodies or microinjection.

Many processes within the cell are under tight spatial and temporal control. The cytoskeleton, for example, is continually remodeled in response to local fluctuations in signaling networks. These networks are typically studied by making genetic modifications to the cell, by microinjecting purified protein, or by using drugs that target specific proteins. Though useful, these approaches have long-lived effects on the entire cell, and
they are not spatially constrained to specific cellular compartments. Cellular signaling events, by contrast, occur in seconds to minutes within micron-sized regions of the cell.

A way of perturbing signaling networks locally and in real-time using genetically encoded, transiently photoactivatable variants of regulatory proteins could transform our understanding of cellular regulation. By activating a signaling protein at a particular time and location and then following its effects, one could understand its regulatory function in the context of specific cellular structures and compartments. Used in this way, photoactivatable variants of regulatory proteins could be used to probe many different cellular processes, including cytoskeletal remodeling, organization of membrane proteins into microdomains and release from cell cycle checkpoints (Fig. 4.2.1).

One can envision adapting the strategy outlined in this thesis to develop a light-regulated adaptor pair that can induce protein complex formation and localization (Fig. 4.2.2). Myriad signaling pathways are triggered by formation of multi-protein complexes or by changes in subcellular localization [79]. Ideally, one would like to create a dimerization adaptor with negligible affinity in the dark but with high (nanomolar) affinity when photoexcited. The well characterized MAP kinase adaptor proteins p14 and MP1 would be an ideal module for incorporation into a LOV-based photoswitch [104].

MP1 and p14 adopt the same fold and form a tight, pseudo-symmetric heterodimer (Fig. 4.2.2) [105]. Each protein contributes a central helix to the binding interface. Because the proteins are small (~125 aa), and their interaction is tight (K_d ~10 nM) and specific, they are ideal components for an engineered adaptor. In addition to the central interacting helix, MP1 and p14 have amino- and carboxy-terminal helices, rendering them well suited for the helix-fusion method we pioneered with the TrpR. The crystal structure of the p14-MP1 complex suggests that fusion of the LOV domain to the amino-
**Figure 4.2.1** Experimental applications of a photoswitchable adaptor. A photoswitchable adaptor would allow many questions to be directly addressed. A: Is local activation of a particular chemoattractant receptor sufficient to induce growth cone turning? B: What are the functional consequences of clustering of glycine receptors by the scaffolding molecule Gephyrin? C: Is recruitment of the Mad2 inhibitor p31comet to the kinetochore sufficient to cause mitotic exit?
**Figure 4.2.2** Design scheme. A: The α helix (red) of the LOV domain (gold) is predominantly folded in the dark and predominantly unfolded in the light. B: The general principle of using an adaptor protein pair, p14/MP1 to induce colocalization of proteins of interest X and Y. C: Fusion across a shared α helix can allosterically couple two domains through mutually exclusive capture of the helix. D: The p14/MP1 heterodimer. The N- and C-termini are in close proximity allowing for circular permutation, and exposing the N-terminus (N') of the central helix of p14 (red). E, F: A fusion of LOV and a circularly permuted p14. In the dark state, LOV captures the shared helix, blocking the p14/MP1 dimerization interface. In the lit state, LOV releases the helix, allowing dimerization.
terminus of MP1 or p14 should interfere with their dimerization in a light-regulated manner.

The N- and C-termini of p14 and MP1 are in close proximity indicating that they could be easily circularly permuted, placing the central helix at the N-terminus. Fusion of the carboxy-terminus of the LOV domain to the now terminal helix should interfere with the dimer binding interface, resulting in especially effective dark state suppression of binding. However, upon photoexcitation, this inhibition will be relieved, permitting complex formation (Fig. 4.2.2).

Upon isolation of an effective photoswitchable adaptor, one would test it in proof-of-concept experiments. An assay for light-dependent recruitment of RFP to the cell membrane would test whether the proteins can function well in vivo (Fig. 4.2.3). Photoexcitation of the membrane bound subunit, should stimulate membrane recruitment of its cytoplasmic binding partner. This assay would permit quantitative analysis of the in vivo binding of the two proteins in the lit and the dark states.

Once one achieved significant light-regulated recruitment of RFP to the membrane, one should test the ability of the light-regulated dimer to control a well-studied endogenous signaling pathway. The RFP tag could be readily extended with a GEF domain that can activate the GTPase Rac1 (Fig. 4.2.3). Local membrane recruitment of this domain is expected to locally activate Rac1, leading to easily assayed responses such as membrane ruffling, actin polymerization and lamellipodium extension.

The photoswitchable adaptor would be well suited to address an important question in cytokinesis. RhoA activation is required for assembly of the cleavage furrow. Active RhoA-GTP stimulates formin-mediated actin assembly and Rho-dependent kinase (ROCK) activation of myosin II, leading to an actomyosin-based contractile ring. ECT2, a
Figure 4.2.3 Proposed experiments. A: Light-dependent membrane targeting by a photoswitchable adaptor. B: Targeting of RFP to the plasma membrane. MP1-RFP is cytosolic in the dark. Illumination of a small spot with a 440 nm laser (dashed circle) induces binding of MP1-RFP to membrane-localized LOV-p14. C: Rac1 activation via membrane targeting of MP1-GEF. Photoexcitation induces proximity of the GEF and Rac1, activating Rac1 and leading to localized membrane ruffling and lamellipodium extension. D: Is localized activation of RhoA during anaphase sufficient to induce an ectopic cleavage furrow?
RhoGEF protein present at the central spindle stimulates the accumulation of a zone of activated, membrane-bound RhoA-GTP [106]. However, it is unknown whether a local zone of active RhoA is sufficient to initiate the formation of a productive cleavage furrow. One could address this question using the photoswitchable adaptor to target a RhoA GEF to the plasma membrane in early anaphase (Fig. 4.2.3). If local RhoA activation is sufficient to initiate actomyosin assembly, we should observe an ectopic cleavage furrow in the illuminated region, perpendicular to the normal furrow.

Because the photoswitchable adaptor leverages the generality of colocalization as a means of initiating cellular signaling events, it could be incorporated into a system of vectors that could be used as a general tool for cell biological applications. Vectors could be built with amino- and carboxy-terminal fusions, adaptors specifically designed for membrane targeting, and, potentially, adaptors of varying photoswitching dynamic range. Ideally, the adaptor would be easily deployable without consideration of structure, analogous to GFP tagging.
Appendix

Design of a ligand-controlled DNA-binding protein

The data presented in this appendix are unpublished. I thank Mohammed Yousef for assistance with the SPR experiments, and Josh Kurutz for assistance with the NMR experiments and for the protein G data. Benjamin Capraro performed the chemical denaturation experiment.

A.1 Introduction

We attempted to design a bifunctional protein with tight allosteric coupling between domains. We employed a novel strategy of fusing two independently functional parent domains at a homologous secondary structural element. We reasoned that this strategy would ensure that structural changes in one domain will be tightly coupled to changes in the other domain. This approach is grounded in ideas of evolution: just as modular multidomain proteins are comprised of previously existing single domains, single domains may be built from genetically encoded peptide fragments that are swapped in genetic shuffling events [107, 108]. Although we did not realized the goal of demonstrating allsoteric coupling between the two domains, we were successful in creating ligand-induced conformational changes and we applied lessons learned to subsequent projects.
**A.2 Design of Zen**

We coupled two small existing domains, engrailed homeodomain (Enh) and a consensus zinc finger (ZF), at areas of local secondary structure homology [109, 110]. Zinc fingers, a DNA binding motif having only about 30 residues, are unable to fold in the absence of a Zn$^{2+}$ ligand (Fig. A.2.1) [111]. The typical zinc finger fold consists of an amino-terminal $\beta$ hairpin and a carboxy-terminal $\alpha$ helix. The Zn$^{2+}$ ligand is coordinated by two cysteine residues in the hairpin, and two cysteine or histidine residues in the helix. Zinc fingers bind to DNA by the carboxy-terminal $\alpha$ helix, with sequence specific interactions occurring between only a few side chains and nucleotides. Multiple zinc finger domains can work in concert to increase specificity.

Homeodomains are the DNA recognition element of homeotic proteins, a large group of eukaryotic transcription factors that specify segmentation and anterior-posterior positioning during development [112]. The homeodomain's helix-turn-helix motif binds to 6 base pairs of cognate DNA with low sequence specificity (Fig. A.2.1). Like zinc fingers, homeodomains often work cooperatively in tandem to bind longer promoter sequences [113]. Of the ~60 residues of the homeodomain, seven are highly conserved, including four core residues and three residues involved in DNA recognition [112]. Homeodomains typically bind to a six base pair cognate sequence by a combination of interactions: the N-terminus makes interactions in the minor groove, while the carboxy-terminal helix-turn-helix docks in the major groove [110].

We aligned the carboxy-terminal helix of the ZF with the amino-terminal helix of Enh such that the core residues of one fold coincide with surface residues of the other, choosing only the core residues in the overlapping sequence for the final construct (Fig. A.
Figure A.2.1  Structural modeling. A: A zinc finger (PDB ID 1MEY [109], green) bound to Zn$^{2+}$ (gold). B: Engrailed homeodomain (PDB ID 3HDD [110]). C: Zen, in the proposed Zn$^{2+}$-bound conformation. Colors are the same as in A and B. D: Two alternate conformations of Zen. The model is rotated 180° from C. The conserved Trp and Phe residues are shown as gray sticks. In the proposed Zn$^{2+}$-bound conformation, the Phe is packed in the zinc finger core. In the proposed Zn$^{2+}$-free conformation, the Phe is packed in the homeodomain core.
**Figure A.2.2** Sequence alignment. The proposed secondary structure of the Zn$^{2+}$-bound state is shown schematically. The residues which directly bind Zn$^{2+}$ are denoted by yellow stars, and the mobile Phe is denoted by a $\phi$. 
2.2. We hypothesized that the resulting protein, Zen should be capable of binding to Zn$^{2+}$ and DNA, and that the two binding interfaces are allosterically coupled. It is not immediately clear from the design whether Zn$^{2+}$-binding should increase or reduce DNA affinity.

We submitted the wild-type ZF and Enh sequences and the fusion sequence to analysis by AGADIR to predict the local secondary structure propensities (Fig. A.2.3) [90]. We found that the ZF has very little intrinsic helical propensity, whereas Enh has a 20-40% probability of forming a helix in residues 16-25 (α1, the amino-terminal α helix). AGADIR also predicted Enh to be ~20% helical in the region of α2 and ~2-3% helical in the region of α3. Surprisingly, the helical propensity of the Zen-α1 was nearly identical to that of the α helix of the ZF, despite 40% sequence identity with Enh-α1, suggesting that many of residues with high helical propensity were eliminated in the design of the fusion protein. Thus the Enh domain may be globally destabilized by sequence changes resulting from the fusion. Folding of the ZF domain, induced by Zn$^{2+}$-binding, could potentially stabilize folding of this helix, which in turn could stabilize the ENH domain. In this case, Zn$^{2+}$-binding would be expected to increase DNA-binding affinity.

An alternative model involves mutually-exclusive folding of the two domains. If, for some reason, the two domains could not fold simultaneously, then Zn$^{2+}$-binding would be expected to decrease DNA-binding affinity. This situation could arise because of a steric clash that is created when both domains are in a folded conformation, or because one or more residues is required for proper folding of both domains, but cannot be a part of both domains simultaneously. Both the ZF domain and the homeodomain have a conserved Phe residue that is required for proper folding. Modeling of the expected Zn$^{2+}$-bound and Zn$^{2+}$-free structures shows this residue in different positions in the two configurations
Figure A.2.3  AGADIR analysis of helical propensity. The predicted helical propensity for the α1 helix of Zen (black) matches the low propensity of the ZF domain (red), not the high propensity of the homeodomain (blue).
(Fig. A.2.1, A.2.2). From this observation, we predict that Zn\(^{2+}\)-binding decreases folding of the homeodomain.

A.3 Experimental validation

A.3.1 Co\(^{2+}\) and Zn\(^{2+}\) binding

In order to test whether Zen binds Zn\(^{2+}\) tetrahedrally to the ZF domain, we examined the absorbance spectrum in the presence of Co\(^{2+}\) (Fig. A.3.1). Most Zn\(^{2+}\) binding proteins also bind Co\(^{2+}\), but do so more weakly because the latter transition metal prefers an octahedral geometry. An enthalpic penalty is paid in tetrahedral binding, which is manifest in color around 640 nm [114]. Indeed, we observed distinct absorbance bands in this region of the spectrum when we added Co\(^{2+}\) to the protein, and we were able to bleach this color by the addition of stoichiometric amounts of Zn\(^{2+}\) (Fig. A.3.1, inset).

From this result, we conclude that Zen binds zinc tetrahedrally in a ZF-like conformation.

A.3.2 Conformational switching

We used circular dichroism (CD) spectroscopy to determine that Zen is folded and predominantly helical in solution. Apo-Zen exhibits a far-UV CD spectrum that is similar to wild-type Enh (Fig. A.3.1). The addition of ZnCl\(_2\) to the buffer induces a loss of CD intensity at all wavelengths. This loss of intensity is fully reversible by the addition of EDTA to the buffer.

Homeodomains have a conserved Trp-Phe NH-π interaction in the native protein that quenches Trp fluorescence; thus, unfolding is marked by increased quantum yield and redshifted emission [115]. In our model, the conserved Phe packs against the Trp only in
**Figure A.3.1** Spectroscopic analysis of Zn\(^{2+}\)-binding and conformational changes. *A:* Visible absorbance spectrum taken in excess CoCl\(_2\). *A,* inset: Co\(^{2+}\) absorbance is bleached by titration with ZnCl\(_2\). *B:* CD wavelength scans. *C:* Fluorescence emission spectra (280 nm excitation) of Zen.
the apoprotein—in the Zn$^{2+}$-bound protein, the Phe switches to the ZF core and no longer interacts with the Trp (Fig. A.2.1) We tested this model by using steady state fluorescence spectroscopy. The single Trp in apo-Zen is fluorescent with a $\lambda_{\text{max}}$ at 321nm, indicative of a buried sidechain (Fig. A.3.1). 1mM Zn$^{2+}$ shifts the $\lambda_{\text{max}}$ to 342nm and increases the quantum yield 1.7-fold. 100mM NaCl instead of ZnCl$_2$ shifts the $\lambda_{\text{max}}$ to 350 nm with a 2.6-fold quantum yield increase. While the effect of NaCl remains puzzling (Enh fluorescence is nearly insensitive to NaCl), the effect of ZnCl$_2$ is consistent with our model and previous observations of homeodomains.

We tested whether Zen is cooperatively folded by denaturation with guanidinium chloride (Fig. A.3.2 and Table 4). Apo Zen unfolds in an apparently 2-state manner at ~1.5 M GdmCl, with a $\Delta G_{\text{N-U}}$ of 1.47 ± 0.08 kcal·mol$^{-1}$ and an m-value of 1.16 ± 0.05 kcal·mol$^{-1}$·M$^{-1}$. The addition of 100 µM Zn$^{2+}$ decreases the free energy of unfolding to 1.28 ± 0.08 kcal·mol$^{-1}$ and the m-value to 0.93 ± 0.04 kcal·mol$^{-1}$·M$^{-1}$, suggesting that zinc binds at the expense of homeodomain stability. Increasing the zinc concentration to 500 µM increases the free energy of unfolding to 1.5 ± 0.1 kcal·mol$^{-1}$, but the m-value remains the same.

Based on spectroscopy and chemical denaturation, we concluded that Zen is cooperatively folded, predominantly helical and capable of binding Zn$^{2+}$ tetrahedrally. While this data is compatible with our structural model (Fig. A.2.1), it does not demonstrate conclusively that the fold of Zen is similar to either of the parent proteins. Furthermore, it is unclear from these experiments how much of the homeodomain structure, if any, is retained in the Zn$^{2+}$-bound state. In order to answer these questions directly, we proceeded to NMR spectroscopy.
Figure A.3.2  Chemical denaturation. Red lines are least-squares fits to the data.
Table 4  Chemical denaturation results.

<table>
<thead>
<tr>
<th>[Zn²⁺]</th>
<th>$\Delta G_u$ / kcal·mol⁻¹</th>
<th>$m$ / kcal·mol⁻¹·M⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 µM</td>
<td>1.47 ± 0.08</td>
<td>1.16 ± 0.05</td>
</tr>
<tr>
<td>100 µM</td>
<td>1.28 ± 0.08</td>
<td>0.93 ± 0.04</td>
</tr>
<tr>
<td>500 µM</td>
<td>1.54 ± 0.1</td>
<td>0.95 ± 0.06</td>
</tr>
</tbody>
</table>
The $^1\text{H}^{15}\text{N}$ HSQC spectrum of Zen, taken at pH 6.1 with 4 mM Zn$^{2+}$, features some well-dispersed peaks characteristic of a folded protein (Fig. A.3.3). However, there is also a subset of peaks that are broader and less dispersed. In the absence of Zn$^{2+}$ the spectrum is less dispersed, and almost all peaks shift position. These results suggest that Zen is at least partially folded in both the Zn$^{2+}$-bound and apo states. However, comparison of our spectra with previously published Enh spectra was not helpful in assigning any resonances [116].

We proceeded to triple-resonance experiments using $^{13}\text{C}/^{15}\text{N}$ labeled Zen. Although the data were collected over the course of a week, comparison of $^1\text{H}^{15}\text{N}$ HSQC spectra taken at various time points throughout the process show consistent line shape and chemical shift, indicating that the Zn$^{2+}$-bound protein is stable in solution at high concentration (data not shown). However, the apoprotein is notably less soluble, consistent with the notion that unfolding of the ZF motif exposes some hydrophobic surface.

We assigned approximately two-thirds of the resonances for the Zn$^{2+}$-bound protein. In general, residues that are part of the ZF domain have more dispersed proton chemical shifts than residues in the homeodomain, again indicating that only the ZF domain is well-folded. C\text{\textalpha} and C\text{\textbeta} secondary chemical shift analysis shows that nearly all of the assigned residues are in the expected secondary-structure conformation, including the $\alpha_2$ and $\alpha_3$ helices of the homeodomain (Fig. A.3.4) [117]. This seems inconsistent with the idea that the homeodomain is unfolded, however, similar chemical shifts have been reported for the helices in a mutationally-denatured variant of Enh [116].

We were unable to assign any more of the protein due to weak resonances and spectral overlap. Comparison of the HNCACB spectrum of Zen with that of Protein G illustrates
Figure A.3.3  NMR analysis of Zn\(^{2+}\)-induced conformational changes. HSQC spectra of Zen in the Zn\(^{2+}\)-bound (black contours) and the Zn\(^{2+}\)-free state (red contours) are shown. Resonances derived from the zinc-finger region are labeled in black, while those from the homeodomain region are labeled in blue.
Figure A.3.4  Chemical shift analysis of secondary structure. $C_\alpha$ and $C_\beta^{13}C$ secondary chemical shifts ($\delta\Delta$) are plotted for assigned residues. A diagram of the proposed secondary structure of the Zn$^{2+}$-bound conformation is also shown.
that the Zen spectrum is poorly-resolved compared with other small proteins (Fig. A.3.5).

A.3.3 DNA binding

In order to test whether DNA binding by Zen is allosterically controlled by Zn\(^{2+}\) binding, we measured DNA binding affinity by surface plasmon resonance (SPR, Biacore) in the presence and absence of Zn\(^{2+}\). Using SPR for DNA binding analysis presents significant hurdles, the most important being the tendency of the strongly basic DNA binding proteins to interact nonspecifically with the surface of the sensor chip. We largely alleviated this problem by using a sensor chip with very short conjugating linkers (as opposed to the longer and more negatively charged dextran linkers) and by using neutravidin, a less negatively charged variant of streptavidin, to capture the biotinylated oligonucleotide ligands. We used hairpin oligonucleotides to achieve a DNA duplex that is not susceptible to loss of activity because of the conjugate strand washing away during regeneration. Finally, we found that a nonsense oligonucleotide is a better reference surface than bare neutravidin [118].

We measured \(K_d\) values for Enh binding to a high affinity cognate (TAATTA) site and a lower affinity (TAATCC) site, and confirmed that the measurements were the same as published values (data not shown) [119]. We then proceeded to measure DNA binding affinity of Zen. We found that Zn\(^{2+}\) apparently increases the affinity of Zen for cognate DNA by 50-fold (Fig A.3.6 and Table 5). In contrast, an identical concentration of Zn\(^{2+}\) decreases the affinity of Enh for cognate DNA by \(~1.5\) fold, probably due to electrostatic shielding. However, the kinetic traces show multiple phases, especially in the highest concentration (50nM) protein samples (Fig. A.3.6). This appears to be due to nonspecific binding of Zen to the surface of the SPR chip. Furthermore, the apparent result of zinc
Figure A.3.5  Comparison of spectral quality in protein G and Zen. HNCACB spectra of Protein G (left) compared with that of Zen (right).
Figure A.3.6  DNA binding studies. A: Baseline-subtracted kinetic SPR response traces for Enh and Zen binding to cognate DNA. B: Equilibrium binding analysis of the same data.
<table>
<thead>
<tr>
<th>Protein</th>
<th>Zn(^{2+})</th>
<th>(K_d / \text{nM})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enh</td>
<td>+</td>
<td>56 ± 5</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>34 ± 3</td>
</tr>
<tr>
<td>Zen</td>
<td>+</td>
<td>113 ± 10</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>&gt; 5 \times 10^3</td>
</tr>
</tbody>
</table>
increasing the DNA affinity conflicts with structural evidence that zinc destabilizes the homeodomain.

We attempted to corroborate the SPR results using an electrophoretic mobility shift assay (EMSA). We were unable to measure any DNA binding using this method. In contrast, Enh bound DNA with an affinity too high to measure under our assay conditions (data not shown). Upon reexamination of our design, we believe that the loss of DNA affinity in Zen is a result of the fusion. Homeodomains bind DNA through two interactions—docking of the carboxy-terminal helix into the major groove is supplemented by interactions between the amino-terminal tail and the minor groove. The ZF-Enh fusion eliminated basic residues in this tail region, perhaps causing the loss of DNA binding.

A.4 Conclusions

The design of Zen is a limited success. Zn$^{2+}$ binding causes a large conformational change, destabilizing the homeodomain in favor of ZF folding. However, there is no change in DNA binding affinity and, indeed, DNA binding seems to be lost in both states.

A.5 Materials and Methods

Cloning, expression and purification

Urs Schmidt-Ott generously provided a clone of *Drosophila* engrailed. A gene encoding the fusion protein was constructed by splicing a synthetic oligonucleotide (IDT) corresponding to the planned amino-terminal zinc-finger motif into wild-type Enh. To facilitate splicing, Arg 24 of Enh was mutated to Ser, thereby introducing an EcoRI site.
We did not correct this mutation in the final construct because it appeared unlikely to significantly destabilize the protein or interfere with function.

The resulting construct was subcloned into pET-21 (Novagen) and overexpressed in the *E. coli* strain Rosetta 2 (DE3) pLysS (Novagen). The protein was purified from inclusion bodies as follows. A cell pellet from 1 L of culture was resuspended in 20 mL buffer (50 mM MOPS, pH 7, 250 mM NaCl, 20 mM DTT). Lysozyme was added to ~10 mg/L and the mixture incubated at room temperature for 30 minutes. Triton X-100 was added to 1% while stirring, and then DNAse I was added to 20 mg/L and the mixture incubated 37 °C for 30 minutes. After incubation, the inclusion bodies were pelleted by centrifugation at 30,000 × g for 30 minutes. The pellet was resuspended in 5 M Urea, 50 mM MES, pH 6.5, 20 mM DTT using a homogenizer, and the mixture was clarified by centrifugation at 30,000 × g for 30 minutes. The supernatant was loaded onto a 5 mL HiTrap SP column (GE Healthcare), washed with the resuspension buffer and eluted with 2M NaCl. The protein was further purified by HPLC and stored as lyophilized powder, which dissolved easily in low salt buffer to concentrations exceeding 10 mg/mL.

For NMR experiments, the protein was expressed in *E. coli* grown in M9 media, with 15NH4Cl as the sole nitrogen source and 13C glucose as the sole carbon source.

*AGADIR analysis*

The sequences were submitted to AGADIR (http://www.embl-heidelberg.de/Services/serrano/agadir/agadir-start.html) and scored with the following settings: pH = 7, temp. = 278, ionic strength = 0.1, Nterm = acetylated, Cterm = amidated. The scores were plotted and analyzed using Igor Pro.
CD Spectroscopy

Lyophilized Zen was resuspended in degassed 10 mM HEPES, pH 7.2, 0.1 mM DTT, and stored under nitrogen. The sample was diluted to 25 µM protein and 1 mM HEPES and 0.01 mM EDTA, and a wavelength scan was taken. ZnCl$_2$ was added to 0.1 mM and another scan taken. Finally, EDTA was added to 1 mM and a third scan taken. A wavelength scan of wild-type Enh was taken under similar conditions. Spectra were acquired on a Jasco J-715 spectropolarimeter equipped with a temperature controlled cuvet holder.

Co$^{2+}$ binding

166 µM CoCl$_2$ was added to 120 µM Zen in 10 mM Tris, pH 8.0, 1 mM TCEP, and an absorbance spectrum was taken using a Hewlett Packard 8452A diode array spectrometer. The extinction coefficient at 642 nm was calculated to be 673 M$^{-1}$. ZnCl$_2$ was titrated from a stock solution.

Tryptophan Fluorescence

Zen was dissolved to 1 µM in 10 mM Tris, pH 8.0, 0.01 mM EDTA, and 1 mM TCEP. Emission spectra (280 nm excitation) were acquired on a FluoroMax-3 spectrofluorometer (Horiba Jobin Yvon).

Chemical Denaturation

CD was monitored on a Jasco J-715 spectropolarimeter. Guanidinium chloride was titrated using a Microlab 500 series titrator. ΔG and m were determined using the linear extrapolation method [120].
NMR Spectroscopy

$^{1}$H-$^{13}$C-$^{15}$N triple resonance spectra were acquired on a Varian Unity Inova 600 MHz spectrometer. nmrPipe and nmrViewJ were used for data processing and analysis [87, 88].

DNA binding

Surface plasmon resonance experiments were performed on a Biacore 3000, using a C1 sensor chip coated with neutravidin. Three 5' biotinylated hairpin oligos, two containing Enh binding sites (TAATTA and TAATCC) and one containing a nonsense site (TATATA) were bound to the chip. A seven-step 1:2 dilution series starting at 500nM was performed, in HEPES-buffered saline (25 mM HEPES, pH 7.4, 125 mM NaCl, 0.005% Tween, 0.1 mM TCEP, 0.1 mM ZnCl$_2$). For zinc-free binding, 1 mM EDTA was added to ligand free zinc. The data were analyzed with independent kinetic and equilibrium models using Clamp and Scrubber (http://www.cores.utah.edu/interaction).
References cited


Title: New approaches to the design of allsteric proteins

Author: Lloyd Devin Strickland

Advisor: Tobin R. Sosnick

Additional committee members: Sean Crosson, Phoebe A. Rice, Shohei Koide

Abstract

Understanding how allostery, the conformational coupling of distant functional sites, arises in highly evolvable systems is of considerable interest in areas ranging from cell biology to protein design and signalling networks. We rationally designed 12 fusions between the naturally photoactive LOV2 domain from *A. sativa* phototropin 1 and the *E. coli* trp repressor. When illuminated, one of the fusions selectively binds operator DNA and protects it from nuclease digestion. We demonstrated that signal transduction between the two domains is mediated by the local unfolding of a shared, linking helix. The degree of switching in this protein is modest, ~6-fold, due to incomplete dark-state suppression of DNA binding activity. We discovered two mutations in this helix that increase the degree of switching to >70-fold. We show that these mutations work by stabilizing the helix-LOV domain association, and hence, improve the dark-state suppression without decreasing lit-state activation. We propose an analytical model accounting for this effect. Natural protein photosensors are promising tools for engineering optical control into biological systems. We are just beginning to understand how these genetically encoded light detectors can be easily coupled to arbitrary effectors. Because these mutations are in the LOV domain itself, our mutations can potentially improve the performance of all LOV2-based photoswitches.