Supplementary Information - Möglich *et al.*

Materials and methods

Molecular Biology

The gene encoding full-length YtvA from *B. subtilis* (residues 1-261) was PCR-amplified from genomic DNA and cloned into vector pET28c (Novagen, Madison, WI, USA), using restriction sites *NdeI* and *SacI*. The nucleotide sequence encoding full-length *B. japonicum* FixL (1-505) was amplified from plasmid pRJ7354[1]. Using restriction enzymes *NdeI* and *SalI*, the DNA fragment was cloned into pET28c. Protein fusions between the YtvA LOV and the FixL histidine kinase domain were generated by fusion PCR[2] with the above plasmids encoding full-length YtvA and FixL proteins as templates. Different variants of fusion proteins were obtained by site-directed mutagenesis and overlap-extension PCR[3]. Fusion kinases were cloned into pET28c using *NdeI* and *SalI* restriction sites. The DNA sequence for the *B. japonicum* FixJ protein (1-205) was amplified by PCR from plasmid pRJ7354 and cloned into pET28c using *NdeI* and *SacI*. All constructs were confirmed by DNA sequencing.

Protein purification

Expression constructs encoding fusion kinase variants were transformed into *E. coli* BL21(DE3) cells. Cells were grown at 37 ºC to optical densities between 0.3 and 0.4 at 600 nm, at which point expression was induced by adding 1 mM isopropyl-ß-D-

thiogalactopyranoside (IPTG). Following induction, cells were shaken at 16 ºC overnight. Cell lysis was carried out by sonication and the lysate was purified by metal ion affinity chromatography using Talon resin (Clontech, Mountain View, CA, USA). His_6-tagged protein was eluted from the resin with 200 mM imidazole and reconstituted with its cofactor FMN. The His_6 affinity tag was cleaved by digesting with biotinylated thrombin (Novabiochem, San Diego, CA, USA) overnight at 4 °C. Protease was quantitatively removed using streptavidin agarose and the protein was again passed over a Talon column. Further purification was achieved by either gel filtration chromatography on a Sephacryl 200 HR column or ion exchange chromatography on MonoQ resin. Fractions containing pure protein were pooled and concentrated by spin filtration (Amicon Ultra, Millipore, Billerica, MA). Concentrated samples were dialyzed into storage buffer (20 mM Tris/HCl pH 8.0, 20 mM NaCl, 25% (v/v) ethylene glycol). Protein concentration was determined based on the FMN extinction coefficient of 12500 M^{-1} cm^{-1} at 450 nm[4]. Protein samples were stored at -20 ºC and retained enzymatic activity for several months.

Full-length FixL was purified according to the above protocol with the following deviations. Prior to induction, 0.15 g hemin was added per litre of cell culture. After elution from the Talon column, the protein was reconstituted with its cofactor heme instead of FMN. Due to proteolytic degradation, the His_6-affinity tag was not cleaved off. FixL protein concentration was determined based on the absorption of the heme cofactor with an extinction coefficient of $1.26 \cdot 10^5$ M^{-1} cm^{-1} at 395 nm[5].

For the purification of FixJ, the corresponding expression plasmid (FixJ 1-205) was transformed into *E. coli* BL21(DE3) cells. Cells were grown at 37 ºC to optical densities of 0.3-0.4 at 600 nm, and protein expression was induced by adding 1 mM IPTG. After

induction, cells were grown overnight at 16 ºC. Cells were lysed by sonication and the lysate was purified by metal ion affinity chromatography on Talon resin. Bound protein was eluted with 200 mM imidazole and digested overnight at 4 °C with biotinylated thrombin to cleave off the His_6 affinity tag. Thrombin was removed with streptavidin agarose. The protein was further purified via metal ion affinity chromatography using Talon resin and concentrated by spin filtration. Concentrated samples were dialyzed into storage buffer (20 mM Tris/HCl pH 8.0, 200 mM NaCl, 25% (v/v) ethylene glycol) and kept at -20 ºC. Protein concentration was calculated using an extinction coefficient of 4860 M^{-1} cm^{-1} at 280 nm, as determined by the method of Edelhoch[6].

Phosphorylation Assays

Kinase activity assays were conducted using procedures closely similar to those described for wild-type FixL[5,7]. All experiments were performed in reaction buffer (50 mM Tris/HCl pH 8.0, 50 mM KCl, 100 μM $MnCl_2$, 5% (v/v) ethylene glycol) at (23 ± 1) ºC. For the fusion kinases 5 mM β-mercaptoethanol was added to the reaction mix. Phosphorylation reactions were started by adding 1 mM γ-^{32}P ATP/$MgCl_2$. γ-^{32}P ATP was purchased from Perkin Elmer (Waltham, MA, USA) and mixed with excess cold ATP/$MgCl_2$ (Sigma, St. Louis, MO, USA) to obtain a specific activity of 0.25 Ci/mmol. At desired times aliquots were removed from the reaction, mixed with half the volume stop buffer (0.5 M Tris/HCl pH 6.8, 0.2 M NaCl, 40 mM EDTA, 30% (v/v) glycerol, 4% (v/v) SDS, 2% (v/v) β-mercaptoethanol) and kept on ice. Samples were run on a 14% (w/v) SDS-PAGE. Polyacrylamide gels were dried and radioactivity was quantitated using a phosphorimager (Molecular Imager FX, BioRad, Hercules, CA, USA).

The standard phosphorylation reaction mixture contained 2 μM fusion kinase or FixL and 50 μM FixJ. To determine the effect of light on kinase activity, reactions were performed in the dark (i. e. dim levels of red light) or under constant white light illumination from a fiber optics illuminator (model 9745-00, Cole-Parmer Instrument Co., Chicago, IL, USA) 30 s prior to and constantly during the experiment (136 mW power). Absorption spectroscopy confirms that under these conditions the LOV domain is fully photobleached. To minimize experimental errors, reactions in dark and light were performed in parallel. For subsaturating illumination a (430 ± 10) nm interference filter (Andover, Salem, NH, USA) was employed which reduced the light output to 240 μW. The concentration of the fusion kinase in autophosphorylation reactions was 30 μM. All experiments on FixL were done with its ferric form (Fe^{III}), as confirmed by absorption spectroscopy. The cyanomet form was obtained by adding 10 mM potassium cyanide to the reaction mix.

Data analysis was carried out with ProFit (QuantumSoft, Uetikon, Switzerland). Phosphorylation velocities were determined by linear fits to the initial time points of the reaction. Recovery kinetics of YF1 kinase activity after photobleaching were evaluated according to the reaction scheme shown in Fig. 3C. Observable phosphorylation activities, A, were fitted to the analytical solution of the recovery reaction:

$$A(t) = A_{\max} \cdot \left[1 + \exp(-2t/\tau) - 2\exp(-t/\tau)\right]$$

Spectroscopic Measurements

All measurements were conducted at (23 ± 1) ºC. Absorption spectra were recorded on a Shimadzu UV-1650 PC spectrophotometer (Columbia, MD, USA). Circular dichroism

measurements were carried out on a Jasco J-715 spectropolarimeter (Easton, MD, USA) in a 0.1 cm cuvette. Samples were diluted into 10 mM Na phosphate pH 8.0, 10 mM NaCl and degassed prior to measurement.

Analytical Ultracentrifugation

Sedimentation equilibrium centrifugation was carried out on a Beckman Optima XL-A centrifuge (Palo Alto, CA, USA) using a 60Ti rotor. Protein samples were diluted into 10 mM Na phosphate pH 8.0, 10 mM NaCl to final concentrations of 7.1, 1.8 and 0.45 μM. Experiments were conducted in six-sector cells at (23 ± 0.3) °C and revolution speeds of 9000 and 12000 rpm. Using MATLAB (MathWorks, Natick, MA, USA), data were globally fitted to a single species model. A value of 0.73 ml/g was used for the partial specific volume υ[8]. Confidence intervals of fitted parameters were estimated by rigorous error analysis[9].

In vivo Experiments

All *in vivo* experiments were conducted in *E. coli* Tuner cells (Novagen, Madison, WI, USA) which are derived from BL21 cells and are Δ*lacZY*. The FixJ gene was cloned in tandem with the fusion kinases into above pET28c-based expression plasmids using *SacII* and *HindIII*. The fusion kinases and FixJ are transcribed as a single operon from these plasmids. For the reporter construct, a fusion between the *B. japonicum* $FixK_2$ promoter, comprising the region 250 nucleotides directly upstream of the $FixK_2$ start codon, and the LacZ gene was cloned into a pRK290-derived plasmid using *KpnI* and *HindIII*[10].

Plasmids encoding the fusion kinases and LacZ reporter construct respectively were transformed into Tuner cells. Cells were grown at 37 ºC in the presence of 0.5 mM IPTG to optical densities between 0.4 and 0.7 at 600 nm. Aliquots of the cells were lysed and β-galactosidase activity was quantitated as described[11,12]. Cell cultures were either kept in the dark or constantly illuminated (white light, ca. 1 mW power) and otherwise treated identically. Reported β-galactosidase activities represent the average of at least five independent cell cultures.

Multiple Sequence Alignment

A search in the SMART database[13] identified 5938 entries that are annotated as containing both PAS and DHp (SMART name HisKA) domains. This list was filtered for protein sequences in which the DHp domain is immediately preceded by a PAS domain, i. e. there are at most 50 residues between the annotated C-terminus of the PAS domain and the N-terminus of the DHp domain. The resulting 3194 sequences were aligned using CLUSTAL W[14]. The alignment was manually adjusted in regions of poor quality. Average hydrophobicities were calculated according to [15]. To prepare Fig. 5, sequences were grouped by their linker length and from each group a subset was chosen at random such that each major branch of the CLUSTAL guide tree is represented. Conservation and average hydropathy were calculated over the entire sequence alignment. Preparation of the sequences and evaluation of the multiple sequence alignment were carried out with in-house Perl scripts. Hydrophobic sequence positions within the linker region are labeled with letters a and d, according to standard coiled-coil nomenclature[16,17].

Additional Table

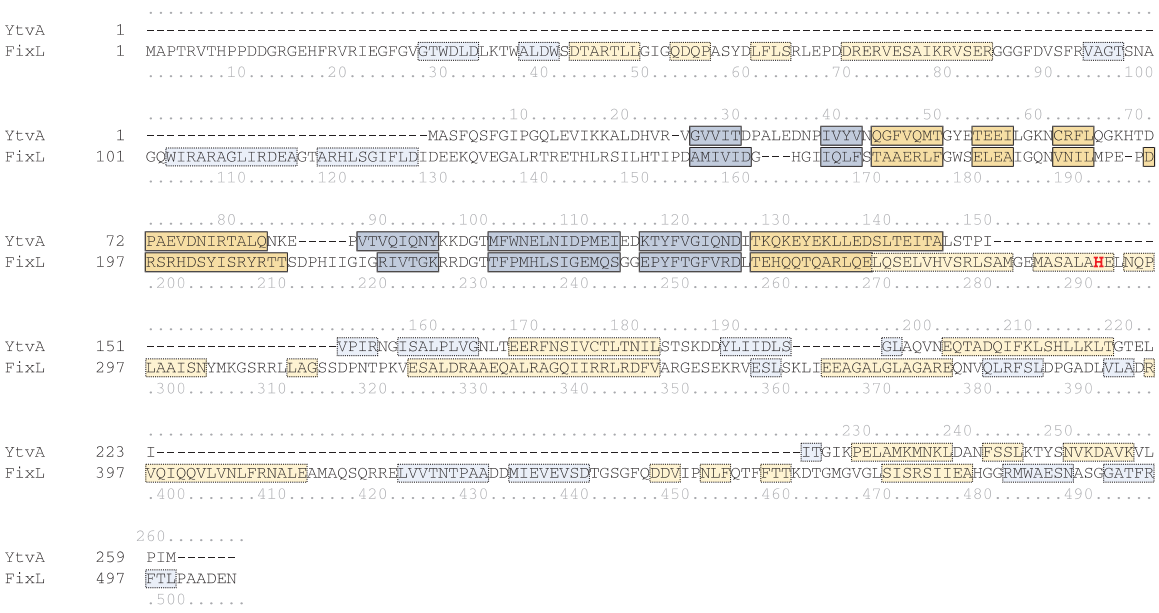
Supplementary Table I - Kinetic Parameters of YF1 and FixL

protein	- signal			+ signal	factor
	v_0 (h^{-1})	K_M^{ATP} (µM)	K_M^{FixJ} (µM)	v_0 (h^{-1})	inhibition
YF1	56.4 ± 2.8	33 ± 2	1.4 ± 0.4	< 0.04	> 1400
FixL	80.2 ± 3.6	206 ± 13	1.6 ± 0.4	0.086 ± 0.014	~ 1000

Additional Figures

Supplementary Figure 1

Sequence alignment between YtvA and FixL. The respective PAS sensor domains were aligned based on their three-dimensional structures and the other parts by sequence similarity. α-helices and β-sheets are indicated as yellow and blue boxes; lighter shaded secondary structure elements are based on homology predictions.

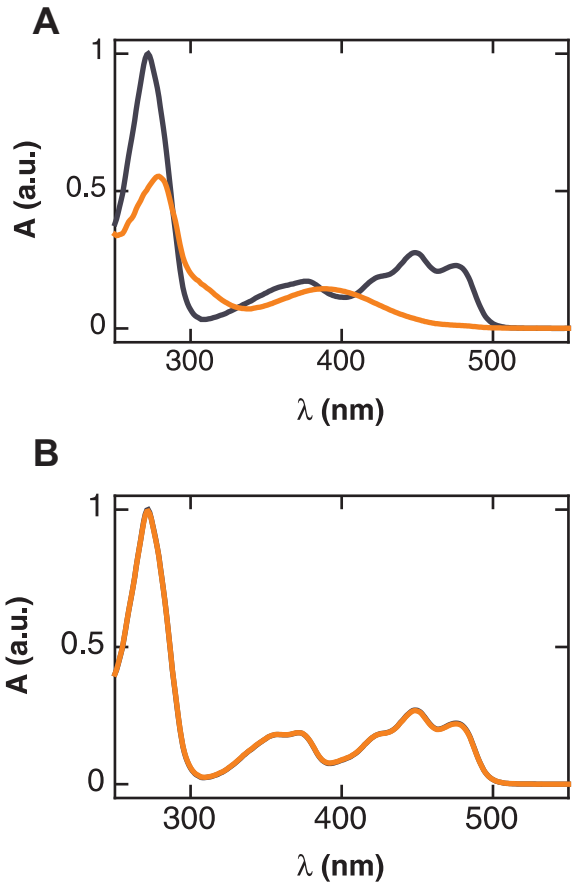


Supplementary Figure 2

Normal LOV photochemistry is abolished in the C62A mutant of YF1.

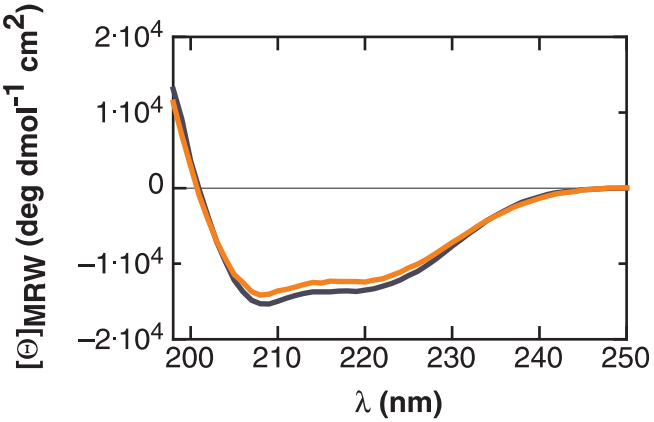
A. UV/vis spectra of YF1 in the dark (blue) and after light absorption (orange). The change in the spectra is indicative of thioether bond formation between cysteine residue 62 and the FMN cofactor.

B. Corresponding spectra of YF1 C62A indicate that this mutant does not undergo the normal LOV photocycle.



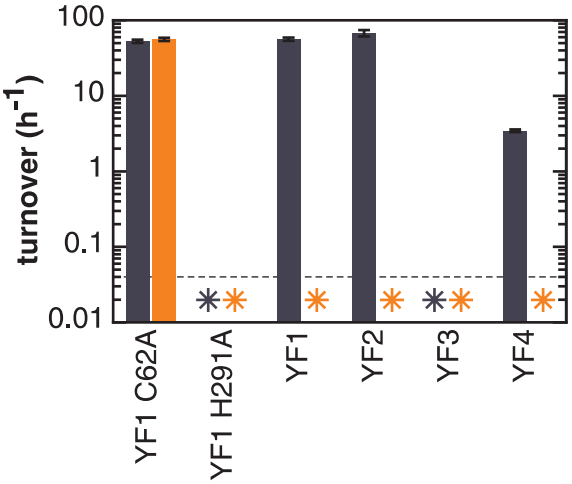
Supplementary Figure 3

Circular dichroism spectra of YF1 in the dark state (blue) and after light absorption (orange). The signal at 208 nm decreases by 5-10% following light absorption.



Supplementary Figure 4

Activity of fusion variants YF1-4 and YF1 mutants C62A and H291A in turnover assays. Blue bars denote activity in the dark and orange bars under constant illumination with white light (136 mW). Asterisks indicate activities below the detectable limit of 0.04 h^{-1} (dashed line). In the mutant C62A the LOV photocycle is abolished and kinase activity independent of light. Replacing the phosphoacceptor H291 with alanine abolishes kinase activity.



Supplementary Figure 5

Sequence alignment of the Jα linker region reveals a four amino acid insertion in the sequence of the *Caulobacter crescentus* histidine kinase LovK in comparison to FixL (yellow). The LovK residues HSDGLQQM were inserted in YF1 after residue 268 to obtain constructs YF1 I1-8.

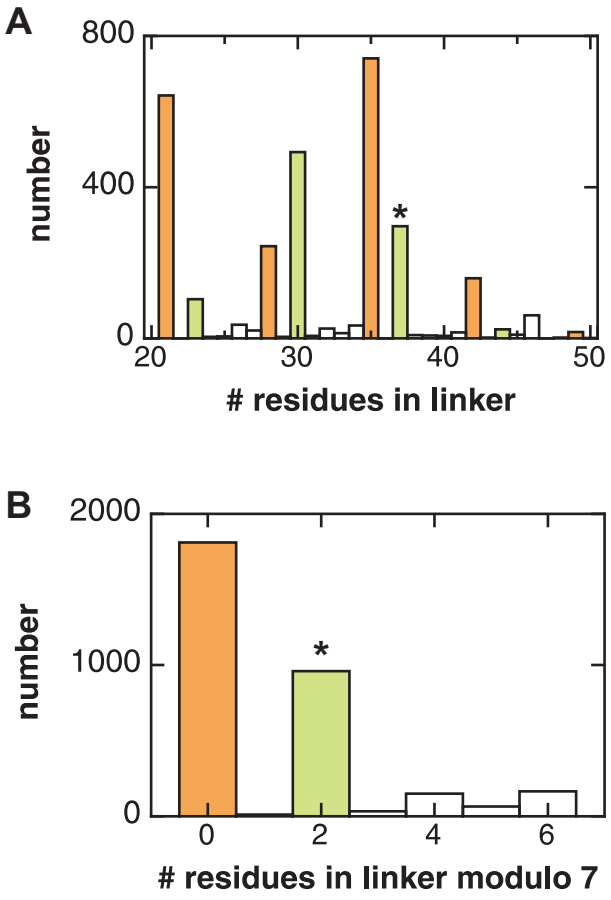
```
              ...140.......150.......160.......170.......180...
LovK    135   DVTDKKVVELRARDHSDGLQQMVEERTRELTEALKQKTALLHEVDHRVK
FixL    255   DLTEHQQTQARLQE----LQSELVHVSR-LSAMGEMASALAHELNQPLA
              ...260.........270........280.......290........
```

Supplementary Figure 6

Heptad periodicity of linker length in sensor histidine kinases.

A. Distribution of the number of residues between the C-terminus of PAS domain and the active-site histidine (indicated by blue arrows in Fig. 5) in 3194 sensor histidine kinase sequences. Note the heptad periodicity in the frequency distribution, highlighted in orange and green. The asterisk indicates the class *B. japonicum* FixL falls into.

B. Modulo 7 of the distribution shown in panel A. Together, the two groups shown in orange and green account for more than 85% of all sequences.



Additional References

1. Anthamatten, D. & Hennecke, H. (1991). The regulatory status of the fixL- and fixJ-like genes in Bradyrhizobium japonicum may be different from that in Rhizobium meliloti. *Mol Gen Genet* **225**, 38-48.
2. Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K. & Pease, L. R. (1989). Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77**, 61-8.
3. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51-9.
4. Losi, A., Ghiraldelli, E., Jansen, S. & Gärtner, W. (2005). Mutational effects on protein structural changes and interdomain interactions in the blue-light sensing LOV protein YtvA. *Photochem Photobiol* **81**, 1145-52.
5. Gilles-Gonzalez, M. A., Gonzalez, G., Perutz, M. F., Kiger, L., Marden, M. C. & Poyart, C. (1994). Heme-based sensors, exemplified by the kinase FixL, are a new class of heme protein with distinctive ligand binding and autoxidation. *Biochemistry* **33**, 8067-73.
6. Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6**, 1948-54.
7. Gilles-Gonzalez, M. A., Caceres, A. I., Sousa, E. H., Tomchick, D. R., Brautigam, C., Gonzalez, C. & Machius, M. (2006). A proximal arginine R206 participates in switching of the Bradyrhizobium japonicum FixL oxygen sensor. *J Mol Biol* **360**, 80-9.
8. Ralston, G. (1993). *Introduction to analytical ultracentrifugation*, Beckman Instruments, Fullerton, CA, USA.
9. Bevington, P. R. & Robinson, D. K. (2003). *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York.
10. Ditta, G., Schmidhauser, T., Yakobson, E., Lu, P., Liang, X. W., Finlay, D. R., Guiney, D. & Helinski, D. R. (1985). Plasmids related to the broad host range vector, pRK290, useful for gene cloning and for monitoring gene expression. *Plasmid* **13**, 149-53.
11. Miller, J. H. (1972). *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
12. Zhang, X. & Bremer, H. (1995). Control of the Escherichia coli rrnB P1 promoter strength by ppGpp. *J Biol Chem* **270**, 11181-9.
13. Letunic, I., Copley, R. R., Pils, B., Pinkert, S., Schultz, J. & Bork, P. (2006). SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res* **34**, D257-60.
14. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-80.

15. Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105-32.
16. McLachlan, A. D. & Stewart, M. (1975). Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. *J Mol Biol* **98**, 293-304.
17. Lupas, A. N. & Gruber, M. (2005). The structure of alpha-helical coiled coils. *Adv Protein Chem* **70**, 37-78.