Library-aided probing of linker determinants in hybrid photoreceptors

Robert Ohlendorf[1], Charlotte Helene Schumacher[1], Florian Richter[1] and Andreas Möglich[1,2,*]

[1]Humboldt-Universität zu Berlin, Institut für Biologie, Biophysikalische Chemie, Invalidenstraße 42, 10115 Berlin, Germany

[2]Universität Bayreuth, Lehrstuhl für Biochemie, Universitätsstraße 30, Geb. NW III, 95447 Bayreuth, Germany

[*]to whom correspondence should be addressed: andreas.moeglich@uni-bayreuth.de, robert.ohlendorf@gmail.com

Copyright Information

Abstract

Signaling proteins comprise interaction and effector modules connected by linkers. Throughout evolution, these recurring modules have multiply been recombined to produce the present-day plethora of signaling proteins. Likewise, modular recombination lends itself to the engineering of hybrid signal receptors, whose functionality hinges on linker topology, sequence and length. Often, numerous linkers must be assessed to obtain functional receptors. To expedite linker optimization, we devised the PATCHY strategy for the facile construction of hybrid gene libraries with defined linker distributions. Empowered by PATCHY, we engineered

photoreceptors, whose signal response was governed by linker length: whereas blue-light-repressed variants possessed linkers of 7*n* or 7*n*+5 residues, variants 7*n*+1 residues were blue-light-activated. Related natural receptors predominantly displayed linker lengths of 7*n* and 7*n*+5 residues but rarely of 7*n*+1 residues. PATCHY efficiently explores linker sequence space to yield functional hybrid proteins including variants transcending the natural repertoire.

Keywords

DNA library, Light-oxygen-voltage, Protein engineering, Sensor histidine kinase, Sensory photoreceptor, Signal transduction

Introduction

As one hallmark of life, organisms adapt their lifestyle and physiology in reaction to endogenous and exogenous signals. Precisely orchestrated networks of signaling proteins underpin information processing and adequate physiological responses. Signaling proteins are generally composed of several modules which distribute into two classes: interaction and effector modules[1]. Whereas the former class mediate interactions with other proteins or signals (sensor module), the latter class possess biological activity (e.g., enzymatic) for generating physiological output (effector module). Throughout evolution, these modules have been recombined and covalently connected via linkers numerous times, thus giving rise to the present-day diversity of signaling proteins. The recurrence of modules in multiple protein architectures indicates the versatility and adaptability of these building blocks.

These general aspects of signal transduction are exemplified by sensory photoreceptors which govern diverse processes of light adaptation including phototaxis in microorganisms[2] and flagellate algae[3], phototropism in higher plants[4], and visual perception in the vertebrate eye[5]. Despite bewildering variety of responses, light absorption is achieved by a surprisingly

small set of around ten different photosensor classes[6–8]. Each photosensor unit harbors an organic chromophore with a conjugated π electron system that absorbs photons in the near-UV/visible range. Light-induced photochemical reactions within the chromophore, e.g., bond isomerization or formation, are coupled to the photosensor scaffold where they promote conformational and dynamic transitions[6]. Initially confined to the immediate vicinity of the chromophore, these perturbations propagate to the effector module and thereby modulate its biological activity. Similar to other signal transduction proteins, the photosensor and effector modules of photoreceptors often localize to distinct protein domains, which are covalently connected by linker segments. This arrangement lends itself to the engineering of novel photoreceptors via rewiring of photosensor and effector modules[9], thus essentially recapitulating recombination events during evolution. The generation of such hybrid photoreceptors[9] not only yields mechanistic insight into signal transduction, but also it provides novel light-regulated actuators for optogenetics[10], i.e. the non-invasive, reversible and spatiotemporally precise manipulation of cellular events by light.

Engineered photoreceptors distribute into associating variants that undergo changes in oligomeric state during signal transduction and into non-associating variants that do not[9]. Associating photoreceptors are based on light-dependent recruitment of effector modules to cellular compartments or other proteins; hence, structural requirements on the linkers connecting modules of associating photoreceptors are often minimal. By contrast, in non-associating photoreceptors the linkers usually meet stringent structural requirements to enable signal transmission from photosensor to effector, primarily in form of order-disorder transitions and other tertiary and quaternary structural transitions[9]. Changes to the often α-helical linker as small as inclusion, omission or exchange of single residues can drastically affect receptor activity and regulation[11–15]. As a case in point, we have previously engineered the blue-light-repressed histidine kinase YF1 by exchanging the two Per-ARNT-Sim (PAS)[16] domains of *Bradyrhizobium japonicum* FixL (*Bj*FixL) for the light-oxygen-voltage (LOV)[4,17] domain of

Bacillus subtilis YtvA (*Bs*YtvA) (Fig. 1a)[12]. LOV domains, first discovered in higher plants[4], form a subclass of the PAS superfamily. Within the homodimeric YF1, the *Bs*YtvA LOV photosensor is connected to the *Bj*FixL effector unit, which comprises the DHp (dimerization/phospho-histidine) and CA (catalytic) domains, via an α-helical coiled-coil linker[18]; incremental elongation and shortening of this linker profoundly modulated light-dependent activity[12].

Even when structural information on the parental photosensor and effector units[13–15] or on the hybrid photoreceptor[18] itself is available, rational modification of the linker remains challenging, thus complicating photoreceptor engineering and optimization. Often, multiple linker variants are tested in trial-and-error manner before a suitable candidate is identified[13–15,19]. To expedite the sampling of variants and to thus facilitate the engineering of photo- and signal receptors, we have developed the PATCHY strategy (primer-aided truncation for the creation of hybrid proteins). In a one-pot reaction, PATCHY generates defined libraries of receptor variants that differ in length and composition of the linkers between their sensor and effector modules. We demonstrate the utility of PATCHY by applying it to the above fusion between the *Bs*YtvA LOV domain and the *Bj*FixL histidine kinase that had yielded YF1. Thereby, we identify multiple light-regulated receptor variants whose properties are primarily governed by linker length. Insertion or deletion of a defined number of residues suffices to convert YF1 from a light-repressed to a light-activated photoreceptor. Sequence analyses suggest that the underlying mechanistic principles evidenced in YF1 are widely shared among natural proteins, thus rendering PATCHY a generally applicable approach for the engineering and optimization of diverse proteins including signal receptors, light-regulated actuators and fluorescent reporters.

Results and Discussion

Primer-aided truncation for the creation of hybrid proteins (PATCHY)

In numerous signal receptors, including sensory photoreceptors[11–15,19], sensor histidine kinases[12,20], adenylate cyclases[21] and methyl-accepting chemoreceptors[22], the identity of the linker connecting sensor and effector modules crucially governs the properties of the composite protein. Using the engineered photoreceptor YF1 as a model system, we sought to systematically interrogate and evaluate the effects of linker length and sequence on activity and regulation. Of particular advantage, the three-dimensional structure of YF1 is known[18], and efficient functional assays for measuring light-regulated activity are in place[23] (Fig. 1b). In the parental receptors *Bs*YtvA and *Bj*FixL that YF1 derives from, the respective sensor and effector modules are connected by linkers of 23 and 27 residues (Fig. 1a). As a corollary, if one restricts fusion to sites within these linkers, there are $24 \cdot 28 = 672$ different ways to generate hybrid receptors that connect the *Bs*YtvA LOV photosensor to the *Bj*FixL histidine-kinase effector. The linkers of these hybrid receptors comprise between 1 and 51 residues. Notably, only four of these variants, denoted YF1 through YF4, have previously been made and characterized[12]. A comprehensive interrogation of all possible linker combinations stands to provide additional insight into signal transduction mechanisms.

We scoured the literature for methods for the generation of hybrid DNA libraries that encode all desired linker variants. Ideally, a library approach for rewiring two gene fragments A and B to generate hybrid receptors should satisfy several criteria: i) each library member should only contain a single fusion site, i.e. it should recombine exactly two fragments; ii) gene fusions between A and B should occur in a sequence- and homology-independent manner; iii) a mechanism should be provided by which fusions are restricted to defined sequence parts (here, the linkers of *Bs*YtvA and *Bj*FixL); and iv) the protocol should be facile in implementation. Apart from methods that only apply to hybrids between highly homologous genes[24,25], the SHIPREC (sequence homology-independent protein recombination)[26] and ITCHY (incremental

truncation for the creation of hybrid proteins)[27–29] approaches meet most of the above criteria (Fig. 2). Briefly, in SHIPREC the parental genes A and B are cloned in series and are then truncated at both termini via partial DNase I and S1 nuclease digest. Blunt-end ligation into a plasmid backbone yields libraries of single-fusion hybrid genes[26]. Conceptually similar, the ITCHY method employs a plasmid in which the two parental genes A and B are cloned in series and which can be linearized using a unique restriction site at the junction between A and B. In contrast to SHIPREC, truncations at both termini of the linearized plasmid are achieved via incremental digest with DNA exonuclease III (Exo-III) followed by blunt-end re-circularization. Varying reaction times of the Exo-III digest yield hybrid genes in which the two termini of the fragments A and B have been recessed to different extent[27,28]. In the THIO-ITCHY variation of the protocol, sulfur-containing nucleotide analogues are incorporated at random sites in the starting linearized plasmid; as a consequence, the incremental Exo-III digest stalls at these sites, and the time-point sampling of conventional ITCHY is rendered obsolete[30].

Although both SHIPREC and ITCHY have successfully been used for the generation of functional hybrid proteins[26–28], demanding experimental protocols and biased distributions of fusion constructs can hamper the wider application of these methods[31]. In particular, we note that neither of the two protocols offers a means of precisely confining hybrid fusions to defined segments of genes A and B, cf. above criterion iii). Moreover, as DNA fragmentation - be it by DNase I/S1 nuclease, be it by Exo-III - occurs essentially randomly, two thirds of all hybrid constructs will contain a frame shift at the fusion site and will hence be dysfunctional. Finally, in our hands it has proved challenging to titrate the activity and processivity of Exo-III in the ITCHY protocol such as to achieve the desired degree of DNA recession, cf. below.

To overcome these severe drawbacks, we devised the PATCHY method for the generation of defined, single-fusion, hybrid-gene libraries. Use of the specific gene fusion between *Bs*YtvA and *Bj*FixL as a test case allowed us to conveniently implement and optimize the method. On the basis of the reference construct YF1, we had previously developed the pDusk plasmid[23] which

affords efficient screening of the light-dependent activity of YF1 variants directly in *E. coli*[18,32–34], thereby obviating laborious protein expression and purification. Briefly, pDusk bears YF1 and its cognate response regulator *Bj*FixJ which in unison control the expression of a *Ds*Red[35] fluorescent reporter in blue-light-repressed fashion. For the original YF1 construct, reporter fluorescence is repressed by a factor of around 10- to 15-fold under blue light compared to darkness. To enable rapid screening of PATCHY linker libraries, we assembled within the pDusk background a starting construct, denoted YF_ori, that replaces YF1 with a tandem fusion of *Bs*YtvA residues 1-147 (N-terminal LOV photosensor plus entire linker) and *Bj*FixL residues 255-505 (entire linker plus C-terminal DHp/CA effector) (Fig. 1a). Notably, the *Bs*YtvA and *Bj*FixL gene fragments are interleaved with a short nucleotide stretch that deliberately introduces a frame-shift and a unique *Nhe*I restriction site for plasmid linearization. Due to this frame-shift, the pDusk-YF_ori construct encodes a dysfunctional receptor, hence displays constitutively low reporter fluorescence and does not respond to light in any detectable manner.

PATCHY hybrid libraries were generated in a single PCR reaction using the *Nhe*I-linearized pDusk-YF_ori as template and forward and reverse sets of staggered primers at equimolar concentrations (cf. Figs. 1b, 2). Notably, the fwd primers were staggered by increments of nucleotide triplets and anneal to one of the desired 28 specific fusion sites within the *Bj*FixL linker (corresponding to amino acids 255-282); likewise, the staggered rev primers annealed to one of the desired 24 fusion sites within *Bs*YtvA (corresponding to amino acids 124-147). Primers were designed with custom Python scripts such that they have largely uniform annealing temperatures. The PCR reaction amplified the entire plasmid and yielded a library of linearized constructs with the desired terminal truncations in both the *Bs*YtvA and *Bj*FixL linker regions, as determined by which primer pair was used for the amplification of a specific library member. Subsequent work-up, 5'-phosphorylation and re-ligation produced circular plasmids that were transformed and further analyzed, cf. below. To predispose the library against the

precursor pDusk-YF_ori construct, the plasmid library was digested with *Nhe*I prior to transformation.

To benchmark the PATCHY protocol, we analyzed the naïve (i.e. unselected) library by next-generation sequencing (NGS). We thus obtained around 5,350 sequences for in-frame fusions between *Bs*YtvA and *Bj*FixL, corresponding to an approximately 8-fold coverage of the theoretical library with $24 \cdot 28 = 672$ members (Fig. 3). Out of the expected variants, the NGS data included sequences for 578, corresponding to a fraction of approximately 86.0 %. Pleasingly, the NGS data indicated that each fwd and rev primer was used albeit to different extent. As no systematic bias towards specific lengths of fusion constructs was apparent, we assume that the differences in primer usage were due to varying annealing/elongation efficiencies during the PCR reaction. Based on the NGS results, the relative primer amounts in the PCR reaction might be varied to compensate for differing efficiencies. Alternatively, the PATCHY reaction could be repeated with a subset of certain fwd and rev primers to specifically generate linker variants that were initially underrepresented. In a similar vein, a non-uniform primer distribution might be chosen right from the start to deliberately bias the resultant PATCHY library towards certain constructs or groups of constructs.

For comparison, we also applied the Exo-III-based ITCHY method[29], that was reported to produce the most uniform construct distributions among several ITCHY protocols[31], to generate corresponding linker libraries. In our hands, speed and processivity of the Exo-III digest proved difficult to control, resulting in significant fractions of constructs in which not only the linker but also the ensuing photosensor and effector domains were truncated. Moreover, owing to the processivity of Exo-III, digest frequently occurred in asymmetric fashion where one of the two linkers was completely recessed but the other linker not at all. Finally, as pointed out above, two thirds of all constructs in ITCHY and SHIPREC libraries are frame-shifted. By contrast, PATCHY uses sets of oligonucleotide primers to precisely define the

theoretical composition of the library and to concomitantly minimize the number of aberrant byproducts (i.e. over-truncated and frame-shifted constructs).

Linker length governs activity and signal response in light-regulated histidine kinases

ITCHY and PATCHY libraries of *Bs*YtvA-*Bj*FixL hybrids within the pDusk background were transformed into *E. coli* cells, which were grown on agar plate, either in darkness or under constant blue light. Clones harboring hybrid constructs with an active histidine kinase were readily identified and isolated based on *Ds*Red reporter fluorescence (Fig. 1b). Following growth in 96-well format liquid culture in the dark and under blue light, *Ds*Red fluorescence of the clones was quantified. The nucleotide sequence of select clones was determined by DNA sequencing. Results for sequence-verified variants from both the ITCHY and PATCHY libraries are summarized in Fig. 4 and Tables 1 and 2. On a whole, the isolated variants fell into three classes: i) variants for which reporter gene expression was repressed by blue light, thus qualitatively corresponding to the YF1 construct; ii) variants for which reporter gene expression was enhanced by blue light; and iii) variants which showed constitutive reporter gene expression but little light regulation. Due to their impaired signal transduction, group iii) was not considered any further; rather, we focused on the light-switchable constructs and deduce several observations:

First, the group of light-repressed constructs contained the variants YF1, YF2 and YF4, consistent with our previous *in vitro* assays on purified proteins that showed repression of histidine-kinase activity by blue light for these constructs[12]. By contrast, the YF3 construct was not among the isolated constructs, again consistent with our previous work that revealed this variant to be devoid of histidine kinase activity. Notably, the light-repressed constructs YF1, YF2 and YF4 differ in the sequence of the linker between LOV photosensor and histidine-kinase effector but not in its length. To facilitate comparison of the isolated constructs, we counted the linker length as the number of residues between the C terminus of the *Bs*YtvA LOV domain

(residue 125 in *Bs*YtvA) and the N terminus of the *Bj*FixL DHp domain (residue 282 in *Bj*FixL) (Fig. 4). In this counting scheme, YF1 and *Bj*FixL both possess a linker length of 28. (Note that in a previous counting scheme[12] with somewhat different reference points the linker lengths of YF1 and *Bj*FixL amounted to 37 residues.) In the following, we refer to individual linker variants by their linker lengths (cf. Tables 1 and 2); for example, LL22.1 and LL22.2 denote two variants with linkers of a common length of 22 residues yet different sequences.

Second, functional, i.e. light-regulated, constructs varied in their linker length between 4 and 50 residues, thus almost covering the entire span of linker lengths expected for the PATCHY library (1 to 51 residues). Previous functional[12] and structural studies[18] provided clear-cut evidence that the linker in the homodimeric LOV histidine kinase YF1 and, by inference, in PAS-linked histidine kinases adopts continuous α-helical coiled-coil conformation. We assume that this also pertains to the majority of the linker variants identified presently, with only a few notable exceptions discussed below. As the rise per residue in canonical α-helical coiled coils amounts to ~ 1.5 Å[36], the observed linker lengths in the light-regulated variants translate into distances of separation between ~ 6 and 75 Å. Thus, in sensor histidine kinases signal clearly acts at a distance, and the absolute separation of sensor and effector modules appears of secondary importance at most. Signal transduction between remote sites is presumably enabled by the coiled-coil structure of the intervening linker[18,37]; with a calculated persistence length of about 150 nm, α-helical coiled coils arguably serve as rigid conduits suitable for transmitting signals over long molecular distances[16,38]. The fact that combinations of a single photosensor-effector pair retained light-dependent switching despite drastically varying linker length demonstrates the robustness of the underlying signal-transduction mechanism. We note that a linker length of 75 Å is enough to traverse a lipid bilayer; while the current variants are almost certainly soluble proteins, this mode of signal transmission along α-helical coiled coils would also be compatible with transmembrane signal receptors. Indeed, many bacterial signal

receptors and the major fraction of sensor histidine kinases span the plasma membrane, predominantly via α-helical and coiled coil transmembrane segments[39].

Third, sequence analyses had identified a conserved DIT consensus motif at the very C terminus of many PAS domains[12], corresponding to residues 125-127 in *Bs*YtvA. Structural studies revealed that these three residues mediate hydrogen-bond and salt-bridge contacts between the PAS core and C-terminal α-helical linkers[18,37,40,41]. Although PATCHY allowed the truncation of this motif, and such variants were indeed frequently observed in the naïve library (cf. Fig. 3), with a single exception all light-regulated variants retained the conserved DIT motif. In the LL6 construct with a linker length of 6 residues (i.e. 22 less than YF1) these residues were replaced with SRL deriving from the linker of *Bj*FixL. Although this construct evidently lacks the highly conserved DIT triplet, it still displayed robust, light-repressed reporter-gene expression, similar to YF1. The SRL residues may well form compensatory contacts that provide a rigid connector between PAS core and coiled coil, but in the absence of structural information this remains speculative.

Fourth, in addition to the previously characterized YF1, YF2 and YF4 constructs all of which have linkers of 28 residues, several variants with other linker lengths also displayed light-repressed reporter-gene expression. Strikingly, the majority of these variants were offset in their linker lengths by multiples of seven residues (Fig. 4); for example, the variants LL14, LL21, LL35.1/2 and LL49.1/2 all showed less reporter expression in blue light than in darkness. Put another way, the linker lengths of the major group of light-repressed variants conformed to 7*n* residues. A second, smaller group of light-repressed variants featured linker lengths that are offset by additional -2 residues; in this group the linker length corresponded to 7*n*+5 residues, e.g., in the variants LL19, LL33 and LL47. This seven-residue or heptad periodicity of activity and regulation by light is indicative of the coiled-coil nature of the linker between the sensor and effector modules of the functional homodimer, as proposed before[12,16] and as later borne out in the crystal structure of YF1[18]. To emphasize this striking aspect, we evaluated the

number of identified light-repressed variants possessing a linker length of $7n+m$, where m is the helical register ($m \in \{0, 1, 2, 3, 4, 5, 6\}$) (Fig. 5a). Evidently, light-repressed variants were mainly obtained if m equaled 0 or 5. Note that within a canonical coiled coil seven residues make up two complete helical turns; accordingly, each increment of m by 1 corresponds to a change in angular orientation of 720° / 7 ≈ 103° (Fig. 5b).

Fifth, certain linker variants (LL22.1/2, LL36, LL43 and LL50) gave rise to a blue-light-induced elevation of reporter-gene expression, as opposed to the reduction of gene expression seen for YF1. Intriguingly, light-activated variants showed a heptad periodicity similar to the light-repressed variants, but the predominant helical register was $7n+1$, i.e. it was shifted by +1 relative to the $7n$ register of YF1 and *Bj*FixL (Fig. 5). Apparently, a shift in helical register from $7n$ to $7n+1$, corresponding to a change in angular orientation of +103°, promoted inversion of signal polarity. These observations suggest that the light signal manifests as a structural change with a rotational component. We previously posited that light absorption within dimeric YF1 leads to a rotation of the two LOV domains relative to each other that then translates through the C-terminal α-helical coiled coil as a torque movement[12,33]. Such motions would induce a supertwist in the coiled coil and concomitant change in angular orientation, which is well suited for propagation over long molecular distances.

Sixth, a few linker variants neither fell into the clades $7n$, $7n+1$ or $7n+5$, yet still showed light-regulated reporter-gene expression, which at least for certain cases can be rationalized. In particular, in the L44 variant that derived from the ITCHY library, by accident a proline mutation was introduced at the junction between the *Bs*YtvA and *Bj*FixL fragments. As the imino acid proline possesses a secondary amine group instead to the primary amine in other proteinogenic amino acids, it is expected to locally disrupt α helices and coiled coils. Resultant deviations from canonical helical structure arguably account for the light-repressed phenotype of the L44 variant although it fell into the class $7n+2$. Similarly, the above-mentioned LL6 variant in which the conserved DIT motif is replaced by SRL displayed light-repressed reporter-

gene expression despite belonging to the 7*n*+6 class of linker variants. This deviation could be due to the different nature of the connector between the *Bs*YtvA LOV domain and the linker in LL6.

Seventh, the PATCHY library theoretically comprised a total of 672 different hybrid receptors, if one disregards additional variants arising from mutations. However, our screen only identified 29 different variants with clear-cut light-regulated reporter-gene expression; even the predominantly light-repressed 7*n* and 7*n*+5, and the light-activated 7*n*+1 classes comprise only a fraction of the theoretically expected constructs of these particular linker lengths. Certainly, the absence of specific hybrid proteins might merely reflect that they were not represented in the original naïve library and could hence not be selected for. However, as the NGS data (cf. Fig. 3) indicated that all PCR primers were used in the PATCHY procedure and that the majority of expected constructs was indeed represented in the naïve library, this notion is not tenable. We rather conclude that the vast majority of linker variants not selected from the PATCHY libraries were impaired in activity and regulation. Apparently, linker length and resultant helical register play the dominant roles in governing activity and light regulation (cf. above), but there is an additional contribution of primary structure. In particular, α-helical coiled coils are stabilized by periodically spaced hydrophobic residues at the helical interface and by salt bridges and polar interaction along the helices (Fig. 5b). If these interactions are disrupted by joining two linker fragments that are out of sync, structure and signal transduction could be impaired. In support of this notion, we previously found that single mutations in the coiled-coil linker of YF1 that disrupted such interactions abrogated proper response to light[32].

Synthetic biology recapitulates and transcends natural systems

As a blue-light-regulated histidine kinase, YF1 is emblematic for many naturally occurring signal-receptor proteins, in particular sensory photoreceptors and sensor histidine kinases. To assess to which extent the presently observed influence of linker composition on activity and

regulation is reflected in these receptors, we analyzed a large body of proteins with domain architecture similar to YF1, i.e. proteins that contain a PAS domain succeeded by a DHp domain. By scouring the Pfam and Uniprot databases, we identified around 30,000 proteins showing this architecture and aligned the sequences of the linkers between PAS and DHp domains. Similar to our previous analyses on a much smaller dataset[12], we found that the linker lengths in natural PAS-histidine kinases, again calculated as the number of residues between the C terminus of the PAS domain and the N terminus of the DHp domain, are not uniformly distributed but rather favor certain discrete values. Heptad analysis of the linker length distribution revealed two predominant groups with helical registers of 7*n* and 7*n*+5. Strikingly, these are exactly the two registers into which the majority of the presently generated light-repressed hybrid proteins fell. We had previously assigned the 7*n* and 7*n*+5 groups evident among natural PAS-histidine kinases as two distinct subclasses related to another, yet employing somewhat different signaling mechanisms[12]. However, our present data indicate that even a single pair of sensor and effector modules can effectively communicate along linkers in both the 7*n* and 7*n*+5 registers. These findings raise the tantalizing possibility that the 7*n* and 7*n*+5 subgroups of PAS-histidine kinases are merely manifestations of the same overarching family with a shared, canonical signaling mechanism.

The signal-receptor variants generated by PATCHY not only recapitulated architectures evidenced in natural signal receptors but also transcended them in at least two ways. First, we presently identified the light-regulated histidine kinase LL4 in which the linker between the C terminus of the LOV/PAS domain and the N terminus of the DHp domain amounted to as little as 4 amino acids. Based on inspection of the YF1 crystal structure[18], such a linker would put the photosensor and effector modules flush against another. Astonishingly, the minimum linker length found among 30,000 natural PAS-histidine kinases was 5. Moreover, out of a total of 29 light-regulated variants, the PATCHY linker libraries contained 4 light-regulated variants with linkers of 10 or fewer residues, which starkly contrasts with the natural PAS-histidine kinase

sequences of which less than 2 ‰ have this short linkers. Second, the PATCHY approach identified several light-activated variants that predominantly fell into the 7*n*+1 class of linker lengths. Strikingly, corresponding linker lengths are only sparsely populated among natural PAS-histidine kinases. Based on this observation, it appears that this particular mechanism of tuning the signal response of receptors is not widely used in nature. Taken together, these findings illustrate that synthetic biology, here implemented via the PATCHY approach, is well suited to not only recapitulate 'what is' but to also explore 'what could be'.

Applications of PATCHY

The highly discretized distribution and heptad periodicities of linker lengths in PAS-histidine kinases argue for continuous coiled-coil linkers and shared aspects of signal transduction in this large protein family. Other histidine-kinase architectures, e.g., those connected to GAF domains show similar periodicities and hence presumably utilize coiled-coil connectors as well[42]. Consequently, deliberate changes of the helical register of coiled-coil linkers may be a widely applicable route towards reengineering signal receptors. Beyond histidine kinases, coiled coils recur as important connectors and signal transducers in numerous receptors[42,43]; for example, this is strikingly evidenced by the stringent heptad periodicity of linker lengths in PAS-GGDEF proteins[16] and tandem-GAF domains[44]. Similar signaling mechanisms and, by that token, means of modulating the signal response might hence apply to these classes of signal receptors, too.

Library approaches like PATCHY are well suited to generate and test in parallel sizeable numbers of linker variants of receptors. Provided a rapid functional readout is available, preferably one with high throughput such as fluorescence-activated cell sorting, these libraries are screened in facile manner to help unraveling the impact of linker properties on receptor activity and regulation. We note that rewiring of sensor and effector modules and the variation of linker composition arguably mimics natural recombination processes during evolution.

Specifically, homologous and non-homologous recombination effectively shuffle sensor and effector modules; such recombination events, in particular in case of non-homologous recombination, are often accompanied by insertion and deletion of a few nucleotide bases (so-called indel mutations) that in turn would lead to linker modifications much in the same manner as the ones probed by PATCHY.

Compared to other approaches suitable for generating hybrid gene libraries, PATCHY offers a facile protocol and the means of precisely determining which part of linker sequence space is sampled. Given the reduced price of oligonucleotide synthesis, PATCHY can also be cost-effective. Subsequent (or, even simultaneous) mutation of the sensor, linker and effector via error-prone PCR[45] further derivatizes the linker libraries. Moreover, PATCHY facilitates the combination of several sensor modules via variable linkers, for example for integration of several input signals from multiple sensors[46]. Deliberate variation of linker length and sequence via PATCHY not only applies to the *de novo* engineering of signal receptors but also to other use cases. On the one hand, existing receptors - be they of natural origin, be they previously engineered - can be modified in their linkers, such as to improve or modulate their activity and signal response. On the other hand, PATCHY also applies to the engineering of fluorescent sensors and reporters[47,48]. In summary, PATCHY thus efficiently generates defined hybrid gene libraries for diverse scenarios and thereby complements rational design approaches.

Methods

Generation of ITCHY and PATCHY linker libraries

The starting construct YF_ori was created via overlap-extension PCR in the pDusk (GenBank JN579120) and encoded residues 1-147 of *Bs*YtvA and residues 255-505 of *Bj*FixL. A unique *Nhe*I site and a frame shift were introduced at the junction to yield a strictly inactive starting construct. ITCHY libraries were generated using a modified version of the published protocol[29].

Unless stated otherwise all enzymes were purchased from Thermo Fisher Scientific Inc. (St. Leon-Rot, Germany). The YF_ori-pDusk template was digested with *Nhe*I, and the linearized plasmid was isolated via gel extraction. The open ends were incrementally digested using 100 U exonuclease III per µg DNA at 4°C. Aliquots were taken at 10, 30, 60, 120, 300 s, and the reaction was immediately stopped with S1 nuclease reaction mix. Samples were then incubated at 22°C for 30 min to allow blunt-end generation by S1 nuclease, pooled and purified via PCR clean-up (Machery-Nagel GmbH & Co. KG, Düren, Germany). Plasmid termini were blunted and phosphorylated by a mixture of T4 DNA polymerase, Klenow fragment and T4 polynucleotide kinase (End-repair kit). After another PCR clean-up the plasmids were ligated with high-concentration T4 DNA Ligase (30 U μl^{-1}). Remaining template plasmids were removed by *Nhe*I digestion and dephosphoralytion (alkaline phosphatase [FastAP]), and the ligated plasmids were again purified via PCR clean-up.

PATCHY libraries were generated using forward and reverse sets of staggered primers annealing in the *Bs*YtvA or *Bj*FixL linker regions of YF_ori. The primers were designed with custom Python (https://www.python.org/) scripts available at https://github.com/vrylr/PATCHY.git. Forward and reverse primers were pooled and applied at a final concentration of 0.5 µM each in a PCR with the YF_ori-pDusk template. The PCR product was purified via PCR clean-up and resuspended in T4 DNA ligase buffer. For phosphorylation and ligation the linear plasmids were incubated with T4 polynucleotide kinase for 30 min at 37°C; PEG4000 and T4 DNA ligase were added, and incubation continued for 1 h at 22°C. Remaining template plasmids were removed by addition of *Nhe*I and FastAP. As all these enzymes have full activity in T4 DNA ligase buffer, DNA purification is not required between steps.

Naïve PATCHY libraries were analyzed by next-generation sequencing (NGS, GATC Biotech AG). Plasmid libraries were randomly fragmented by ultrasound and sequenced on an Illumina platform that yields paired-end reads of 125 base pairs length. In total, about $2 \cdot 5{,}350 = 10{,}700$

reads were obtained that correspond to in-frame fusions of *Bs*YtvA and *Bj*FixL. Given a theoretical library size of 24 · 28 = 672 linker variants, this amounted to about 8-fold library coverage. Evaluation of sequence data was performed with Python and Numpy (http://www.numpy.org/), and graphs were prepared with Matplotlib (http://matplotlib.org/).

Library screening

ITCHY or PATCHY libraries were transformed into *E.coli* DH10b, and kinase activity was evaluated using pDusk as described before[23,32]. Briefly, the transformed cells were grown on agar plates in the dark or under blue light (470 nm, 40 µW cm^{-2}). Clones harboring active kinase fusions were then selected based on their *Ds*Red fluorescence (excitation 470 nm) using a 590 nm high-pass filter and sorted into 96-deep-well plates containing 600 µl LB supplemented with 50 µg ml^{-1} kanamycin (LB/Kan). The plates were sealed with gas-permeable film and incubated for 22 h at 37°C and 800 rpm. 6 µl from each well were used to inoculate two deep-well plates with 600 µl LB/Kan per well, which were then incubated for 22 h at 37°C and 800 rpm in the dark or under blue light (470 nm, 60 µW cm^{-2}). Absorbance at 600 nm (OD_{600}) and *Ds*Red fluorescence (excitation 554 ± 9 nm, emission 591 ± 20 nm) were measured in black-walled 96-well µClear plates (Greiner BioOne, Frickenhausen, Germany) using a Tecan Infinite M200 PRO plate reader (Tecan Group Ltd., Männedorf, Switzerland). Data were normalized to the fluorescence per OD_{600} of YF1 in pDusk under dark conditions. Light-regulated constructs were identified by comparing *Ds*Red levels under dark and light conditions and were submitted for DNA sequencing (GATC Biotech). Sequencing data were analyzed with custom Python scripts and ClustalX. All verified clones were again tested in the 96-well setup; data represent average values of two biological replicates ± standard deviation.

Multiple sequence alignment

Protein entries comprising PAS or DHp domains were retrieved from the Pfam database[42] (version 28.0, May 2015) and saved locally. For PAS, this corresponded to the Pfam families PF00989, PF13426, PF08447, PF08448, PF13188, PF08348, PF12860, PF13596, PF16527, PF08446, PF14598, PF16736, PF07310 and PF08670; the DHp dataset comprised the Pfam families PF00512, PF07536, PF07568 and PF07730. Within these two datasets, entries were identified that contain consecutive PAS and DHp domains separated by no more than 100 residues (according to the Pfam annotation). Full sequences of these entries were retrieved from Uniprot[43] and saved locally. Linkers between PAS and DHp domains were extracted from the Uniprot dataset and aligned with MAFFT[49]. Linker sequence and length within the resultant multiple sequence alignment were evaluated. Sequence retrieval, processing and analysis were carried out with custom Perl scripts (http://www.perl.org).

Acknowledgements

We thank our group for discussion and comments on the manuscript. Financial support by Boehringer-Ingelheim Fonds (R.O.), by Deutsche Forschungsgemeinschaft (DFG) grant RI2468/1-1 (F.R.), by the DFG Cluster of Excellence 'Unifying Concepts in Catalysis' (A.M.) and by a Sofja-Kovalevskaya Award by the Alexander-von-Humboldt Foundation (A.M.) is gratefully acknowledged.

Abbreviations

PAS: Per-ARNT-Sim, LOV: Light-Oxygen-Voltage, HisK: Histidine kinase, PATCHY: Primer-aided truncation for the creation of hybrid proteins, ITCHY: Incremental truncation for the creation of hybrid proteins, PCR: Polymerase Chain Reaction, SHIPREC (Sequence homology-independent protein recombination), NGS: next-generation sequencing

References

1. Pawson, T., and Nash, P. (2003) Assembly of cell regulatory systems through protein interaction domains. *Science 300*, 445–452.
2. Engelmann, T. W. (1883) Bacterium photometricum. *Pflüg. Arch. Eur. J. Physiol. 30*, 95–124.
3. Nagel, G., Ollig, D., Fuhrmann, M., Kateriya, S., Musti, A. M., Bamberg, E., and Hegemann, P. (2002) Channelrhodopsin-1: a light-gated proton channel in green algae. *Science 296*, 2395–2398.
4. Christie, J. M., Reymond, P., Powell, G. K., Bernasconi, P., Raibekas, A. A., Liscum, E., and Briggs, W. R. (1998) Arabidopsis NPH1: a flavoprotein with the properties of a photoreceptor for phototropism. *Science 282*, 1698–1701.
5. Ernst, O. P., Lodowski, D. T., Elstner, M., Hegemann, P., Brown, L. S., and Kandori, H. (2014) Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chem. Rev. 114*, 126–163.
6. Möglich, A., Yang, X., Ayers, R. A., and Moffat, K. (2010) Structure and function of plant photoreceptors. *Annu. Rev. Plant Biol. 61*, 21–47.
7. Brown, B. A., Cloix, C., Jiang, G. H., Kaiserli, E., Herzyk, P., Kliebenstein, D. J., and Jenkins, G. I. (2005) A UV-B-specific signaling component orchestrates plant UV protection. *Proc. Natl. Acad. Sci. U. S. A. 102*, 18225–18230.
8. Ortiz-Guerrero, J. M., Polanco, M. C., Murillo, F. J., Padmanabhan, S., and Elías-Arnanz, M. (2011) Light-dependent gene regulation by a coenzyme B12-based photoreceptor. *Proc. Natl. Acad. Sci. U. S. A. 108*, 7565–7570.
9. Ziegler, T., and Möglich, A. (2015) Photoreceptor engineering. *Front. Mol. Biosci. 2*.
10. Deisseroth, K., Feng, G., Majewska, A. K., Miesenböck, G., Ting, A., and Schnitzer, M. J. (2006) Next-generation optical technologies for illuminating genetically targeted brain circuits. *J. Neurosci. Off. J. Soc. Neurosci. 26*, 10380–10386.
11. Levskaya, A., Chevalier, A. A., Tabor, J. J., Simpson, Z. B., Lavery, L. A., Levy, M., Davidson, E. A., Scouras, A., Ellington, A. D., Marcotte, E. M., and Voigt, C. A. (2005) Synthetic biology: engineering Escherichia coli to see light. *Nature 438*, 441–442.
12. Möglich, A., Ayers, R. A., and Moffat, K. (2009) Design and signaling mechanism of light-regulated histidine kinases. *J. Mol. Biol. 385*, 1433–1444.
13. Wu, Y. I., Frey, D., Lungu, O. I., Jaehrig, A., Schlichting, I., Kuhlman, B., and Hahn, K. M. (2009) A genetically encoded photoactivatable Rac controls the motility of living cells. *Nature 461*, 104–108.
14. Gasser, C., Taiber, S., Yeh, C.-M., Wittig, C. H., Hegemann, P., Ryu, S., Wunder, F., and Moglich, A. (2014) Engineering of a red-light-activated human cAMP/cGMP-specific phosphodiesterase. *Proc. Natl. Acad. Sci. 111*, 8803–8808.
15. Ryu, M.-H., Kang, I.-H., Nelson, M. D., Jensen, T. M., Lyuksyutova, A. I., Siltberg-Liberles, J., Raizen, D. M., and Gomelsky, M. (2014) Engineering adenylate cyclases regulated by near-infrared window light. *Proc. Natl. Acad. Sci. 111*, 10167–10172.
16. Möglich, A., Ayers, R. A., and Moffat, K. (2009) Structure and Signaling Mechanism of Per-ARNT-Sim Domains. *Structure 17*, 1282–1294.
17. Conrad, K. S., Manahan, C. C., and Crane, B. R. (2014) Photochemistry of flavoprotein light sensors. *Nat. Chem. Biol. 10*, 801–809.
18. Diensthuber, R. P., Bommer, M., Gleichmann, T., and Möglich, A. (2013) Full-Length Structure of a Sensor Histidine Kinase Pinpoints Coaxial Coiled Coils as Signal Transducers and Modulators. *Structure 21*, 1127–1136.
19. Strickland, D., Moffat, K., and Sosnick, T. R. (2008) Light-activated DNA binding in a designed allosteric protein. *Proc. Natl. Acad. Sci. U. S. A. 105*, 10709–10714.

20. Wang, B., Zhao, A., Novick, R. P., and Muir, T. W. (2014) Activation and inhibition of the receptor histidine kinase AgrC occurs through opposite helical transduction motions. *Mol. Cell 53*, 929–940.
21. Winkler, K., Schultz, A., and Schultz, J. E. (2012) The S-Helix Determines the Signal in a Tsr Receptor/Adenylyl Cyclase Reporter. *J. Biol. Chem. 287*, 15479–15488.
22. Ferris, H. U., Zeth, K., Hulko, M., Dunin-Horkawicz, S., and Lupas, A. N. (2014) Axial helix rotation as a mechanism for signal regulation inferred from the crystallographic analysis of the E. coli serine chemoreceptor. *J. Struct. Biol. 186*, 349–356.
23. Ohlendorf, R., Vidavski, R. R., Eldar, A., Moffat, K., and Möglich, A. (2012) From Dusk till Dawn: One-Plasmid Systems for Light-Regulated Gene Expression. *J. Mol. Biol. 416*, 534–542.
24. Stemmer, W. P. (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature 370*, 389–391.
25. Zhao, H., Giver, L., Shao, Z., Affholter, J. A., and Arnold, F. H. (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol. 16*, 258–261.
26. Sieber, V., Martinez, C. A., and Arnold, F. H. (2001) Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol. 19*, 456–460.
27. Ostermeier, M., Shim, J. H., and Benkovic, S. J. (1999) A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol. 17*, 1205–1209.
28. Ostermeier, M., Nixon, A. E., Shim, J. H., and Benkovic, S. J. (1999) Combinatorial protein engineering by incremental truncation. *Proc. Natl. Acad. Sci. 96*, 3562–3567.
29. Ostermeier, M., and Lutz, S. (2003) The creation of ITCHY hybrid protein libraries. *Methods Mol. Biol. Clifton NJ 231*, 129–141.
30. Stefan Lutz, Marc Ostermeier, and Stephen J Benkovic. Rapid generation of incremental truncation libraries for protein engineering using α-phosphothioate nucleotides.
31. Ostermeier, M. (2003) Theoretical distribution of truncation lengths in incremental truncation libraries. *Biotechnol. Bioeng. 82*, 564–577.
32. Gleichmann, T., Diensthuber, R. P., and Möglich, A. (2013) Charting the Signal Trajectory in a Light-Oxygen-Voltage Photoreceptor by Random Mutagenesis and Covariance Analysis. *J. Biol. Chem. 288*, 29345–29355.
33. Diensthuber, R. P., Engelhard, C., Lemke, N., Gleichmann, T., Ohlendorf, R., Bittl, R., and Möglich, A. (2014) Biophysical, mutational, and functional investigation of the chromophore-binding pocket of light-oxygen-voltage photoreceptors. *ACS Synth. Biol. 3*, 811–819.
34. Yee, E. F., Diensthuber, R. P., Vaidya, A. T., Borbat, P. P., Engelhard, C., Freed, J. H., Bittl, R., Möglich, A., and Crane, B. R. (2015) Signal transduction in light-oxygen-voltage receptors lacking the adduct-forming cysteine residue. *Nat. Commun. 6*, 10079.
35. Strack, R. L., Strongin, D. E., Bhattacharyya, D., Tao, W., Berman, A., Broxmeyer, H. E., Keenan, R. J., and Glick, B. S. (2008) A noncytotoxic DsRed variant for whole-cell labeling. *Nat. Methods 5*, 955–957.
36. Lupas, A. N., and Gruber, M. (2005) The structure of alpha-helical coiled coils. *Adv Protein Chem 70*, 37–78.
37. Wang, C., Sang, J., Wang, J., Su, M., Downey, J. S., Wu, Q., Wang, S., Cai, Y., Xu, X., Wu, J., Senadheera, D. B., Cvitkovitch, D. G., Chen, L., Goodman, S. D., and Han, A. (2013) Mechanistic insights revealed by the crystal structure of a histidine kinase with signal transducer and sensor domains. *PLoS Biol. 11*, e1001493.
38. Wolgemuth, C. W., and Sun, S. X. (2006) Elasticity of α -Helical Coiled Coils. *Phys. Rev. Lett. 97*.
39. Gao, R., and Stock, A. M. (2009) Biological insights from structures of two-component proteins. *Annu. Rev. Microbiol. 63*, 133–154.
40. Circolone, F., Granzin, J., Jentzsch, K., Drepper, T., Jaeger, K.-E., Willbold, D., Krauss, U., and Batra-Safferling, R. (2012) Structural Basis for the Slow Dark Recovery of a Full-Length LOV Protein from Pseudomonas putida. *J. Mol. Biol. 417*, 362–374.

41. Herrou, J., and Crosson, S. (2011) Function, structure and mechanism of bacterial photosensory LOV proteins. *Nat. Rev. Microbiol. 9*, 713–723.
42. Finn, R. D., Mistry, J., Schuster-Böckler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L., and Bateman, A. (2006) Pfam: clans, web tools and services. *Nucleic Acids Res 34*, D247–D251.
43. Consortium, T. U. (2008) The universal protein resource (UniProt). *Nucleic Acids Res 36*, D190–5.
44. Rockwell, N. C., Ohlendorf, R., and Möglich, A. (2013) Cyanobacteriochromes in full color and three dimensions. *Proc. Natl. Acad. Sci. U. S. A. 110*, 806–807.
45. Leung, D. W., Chen, E., and Goeddel, D. V. (1989) A method for random mutagenesis of a de_fined DNA segment using a modi_ed polymerase chain reaction. *Technique 1*, 11–15.
46. Möglich, A., Ayers, R. A., and Moffat, K. (2010) Addition at the molecular level: signal integration in designed Per-ARNT-Sim receptor proteins. *J. Mol. Biol. 400*, 477–486.
47. St-Pierre, F., Chavarha, M., and Lin, M. Z. (2015) Designs and sensing mechanisms of genetically encoded fluorescent voltage indicators. *Curr. Opin. Chem. Biol. 27*, 31–38.
48. Ibraheem, A., and Campbell, R. E. (2010) Designs and applications of fluorescent protein-based biosensors. *Curr. Opin. Chem. Biol. 14*, 30–36.
49. Katoh, K., and Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol. 30*, 772–780.

Figure legends

Figure 1

(a) The blue-light-repressed histidine kinase YF1 comprises the *Bs*YtvA-LOV domain (residues 1-127) and the *Bj*FixL-histidine kinase including its linker (L^{F2}) (residues 258-505). The YF_ori construct additionally contains the entire linker from the *Bs*YtvA protein (L^{Y}). (b) *Bs*YtvA-*Bj*FixL variants were screened in the pDusk background, where catalytically active hybrids promote phosphorylation of the response regulator *Bj*FixJ and subsequent expression of the red-fluorescent reporter *Ds*Red.

Figure 2

Methods for generating hybrid gene libraries. SHIPREC starts from a fusion of two parental genes A and B in reverse order, separated by an interjacent unique restriction site (RE). DNaseI randomly nicks the DNA followed by S1 nuclease digest to yield randomly truncated fusion genes. Blunt-end ligation and subsequent re-linearization using the RE site generates fusion variants with the desired order of genes A and B that are ligated into a plasmid to create the final libraries. ITCHY and PATCHY employ a common template plasmid that comprises genes A

and B in tandem, separated by a unique restriction site. ITCHY involves linearization of the template plasmid using the RE site, followed by nuclease-mediated truncation of both termini and blunt-end ligation. Parallel reactions with different incubation times yield hybrid genes in which the termini of genes A and B are recessed to varying extent. PATCHY uses PCR with staggered sets of forward and reverse primers to create linear plasmids with genes A and B truncated at positions corresponding to the primer annealing sites. Phosphorylation and blunt-end ligation yields a single plasmid library harboring all desired gene hybrids.

Figure 3

Next-generation sequencing of naïve PATCHY libraries. Squares in the central matrix indicate that a given hybrid gene was observed in the naïve PATCHY library with the color denoting the frequency of occurrence within a total of 5,350 sequences; the YF1 construct is marked with a cross. Numbers on the *x*-axis represent the 28 forward primers annealing within the *Bj*FixL linker (residues 255-282). Numbers on the *y*-axis represent the corresponding 24 reverse primers for the *Bs*YtvA linker (residues 124-147). Ergo the shortest possible construct (*Bs*YtvA 1-124 + *Bj*FixL 282-505) is situated in the bottom left corner and the longest (*Bs*YtvA 1-147 + *Bj*FixL 255-505) in the top right corner of the matrix. Line graphs indicate usage of the forward and reverse primers.

Figure 4

Light-regulated *Bs*YtvA-*Bj*FixL hybrid variants. (a) Sequence alignment of the linker regions Linker lengths were calculated as the number of residues between N124 of the *Bs*YtvA-LOV domain and G283 of *Bj*FixL DHp. The DIT motif (blue), the phosphor-accepting histidine (grey) and proline residues (red) are highlighted. The position of the *Nhe*I site in YF_ori is denoted with an asterisk. (b) Histidine-kinase activity in the pDusk assay of selected *Bs*YtvA-*Bj*FixL fusions in the dark (black bars) or under blue light (white bars).

Figure 5

Linker lengths of light-regulated *Bs*YtvA-*Bj*FixL hybrids. (a) Whereas light-repressed (filled circles) *Bs*YtvA-*Bj*FixL hybrids showed periodic linker lengths of *7n* or *7n*+5 residues, light-induced variants (open circles) primarily occured at linker lengths of *7n*+1. (b) Angular orientation in a coiled coil changes by 103° per residue. Predominantly hydrophobic interactions between non-polar residues at positions *a* and *d* as well as often ionic interactions between charged residues at positions *e* and *g* stabilize the dimeric structure.

Figure 6

Linker-length distributions (a) and heptad periodicities (*7n*+*m*) (b) of natural PAS-coupled histidine kinases. The helical registers *m* and corresponding linker lengths are marked in blue (*7n*), red (*7n*+5*)*, or green (7*n*+1), respectively.

Tables

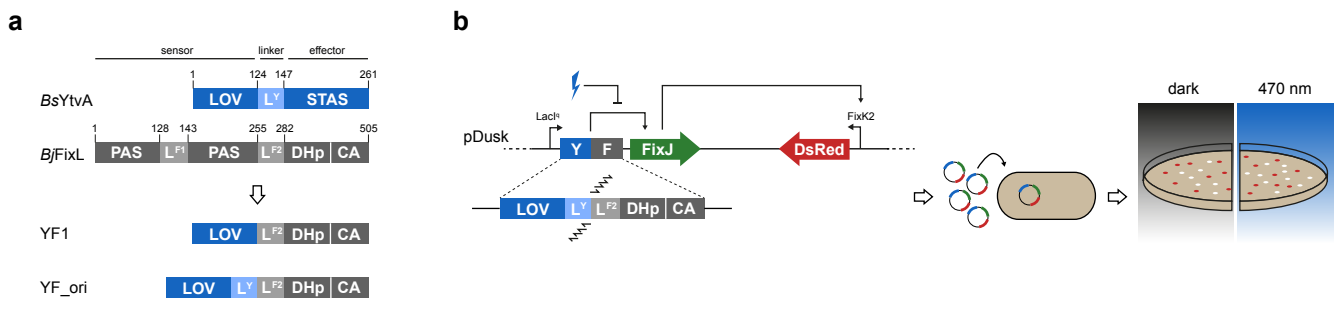
Table 1 - Light-repressed *Bs*YtvA-*Bj*FixL hybrids from ITCHY and PATCHY libraries.

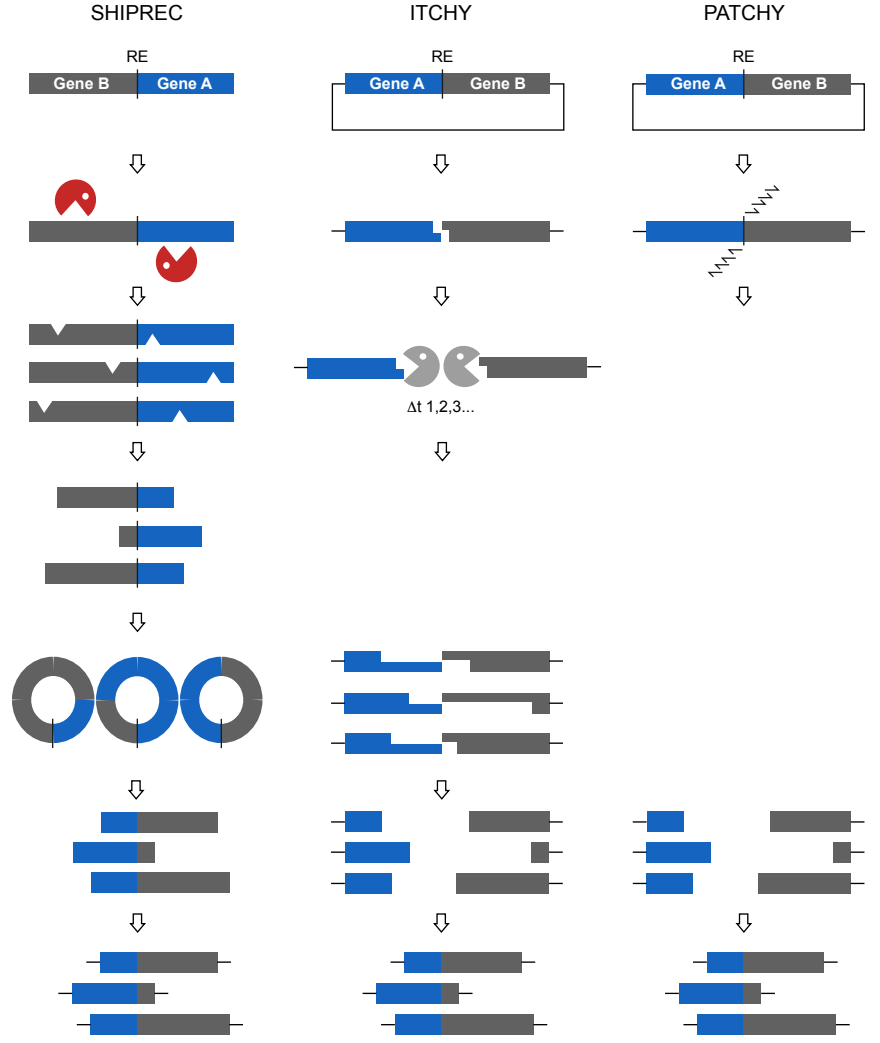
Construct	Linker length	Helical register	F/OD_{600} dark	F/OD_{600} 470 nm	Ratio
LL6	6	6	0.41 ± 0.05	0.05 ± 0.01	8 ± 3
LL7	7	0	0.17 ± 0.02	0.03 ± 0.01	5 ± 2
LL10	10	3	1.5 ± 0.1	0.54 ± 0.06	2.8 ± 0.5
LL14	14	0	0.89 ± 0.09	0.16 ± 0.02	6 ± 1
LL19	19	5	0.76 ± 0.07	0.13 ± 0.02	6 ± 1
LL21	21	0	0.94 ± 0.06	0.39 ± 0.03	2.4 ± 0.3
LL25	25	4	0.18 ± 0.02	0.057 ± 0.007	3.2 ± 0.7
YF1	28	0	1.0 ± 0.1	0.021 ± 0.006	50 ± 20
YF2	28	0	1.5 ± 0.1	0.19 ± 0.03	8 ± 2
YF4	28	0	1.5 ± 0.1	0.043 ± 0.005	34 ± 7
LL29	29	1	0.34 ± 0.03	0.08 ± 0.01	4.3 ± 0.9

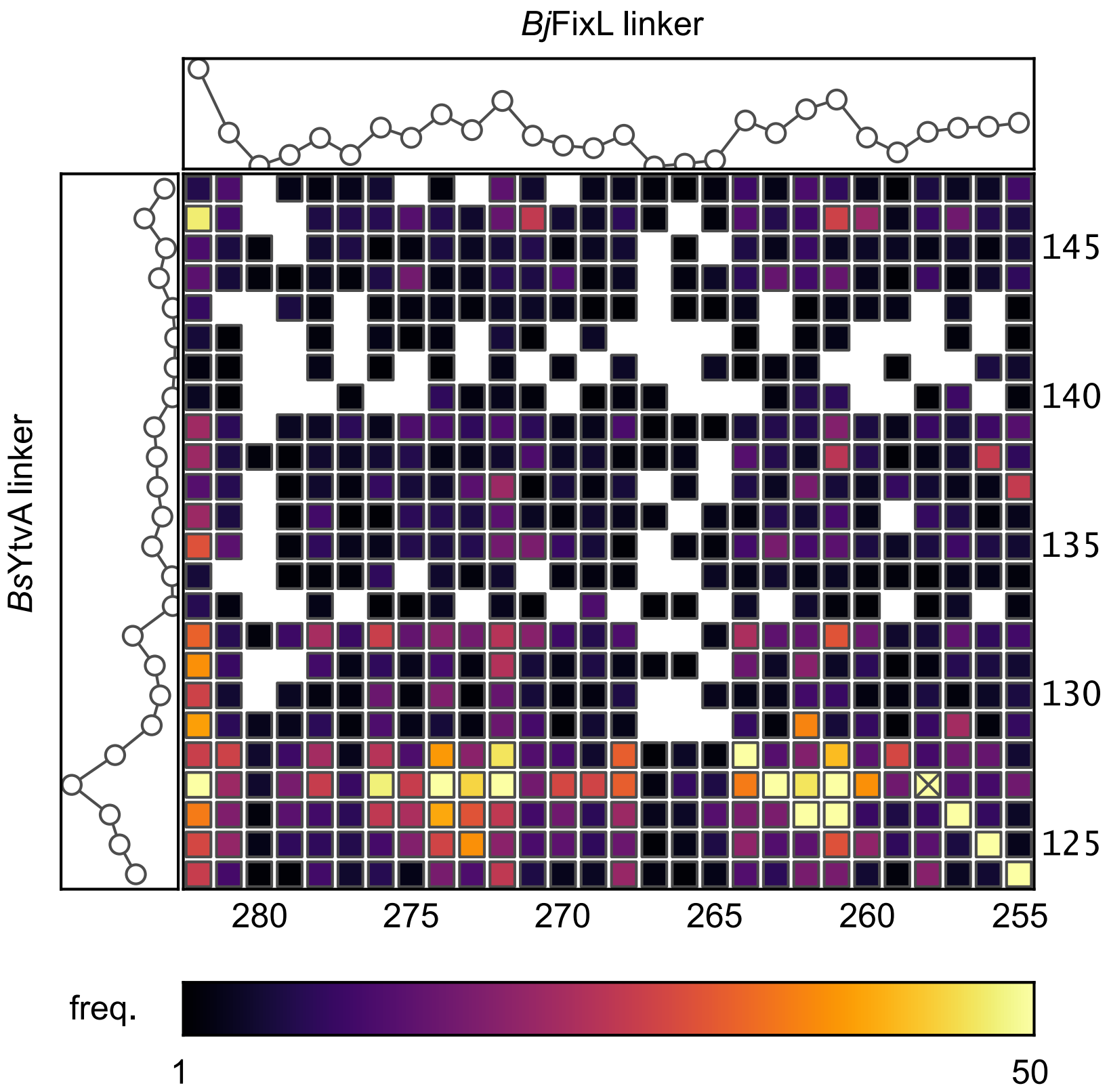
LL33	33	5	0.26 ± 0.03	0.032 ± 0.008	8 ± 3
LL35.1	35	0	0.8 ± 0.1	0.06 ± 0.01	13 ± 4
LL35.2	35	0	0.89 ± 0.06	0.22 ± 0.02	4.1 ± 0.6
LL42	42	0	1.7 ± 0.1	0.034 ± 0.005	50 ± 10
LL44	44	2	1.7 ± 0.1	0.084 ± 0.009	20 ± 4
LL47	45	5	0.27 ± 0.04	0.018 ± 0.004	15 ± 5
LL49.1	49	0	1.8 ± 0.2	0.040 ± 0.007	40 ± 10
LL49.2	49	0	1.3 ± 0.2	0.018 ± 0.005	80 ± 30

Table 2 - Light-activated *Bs*YtvA-*Bj*FixL hybrids from ITCHY and PATCHY libraries.

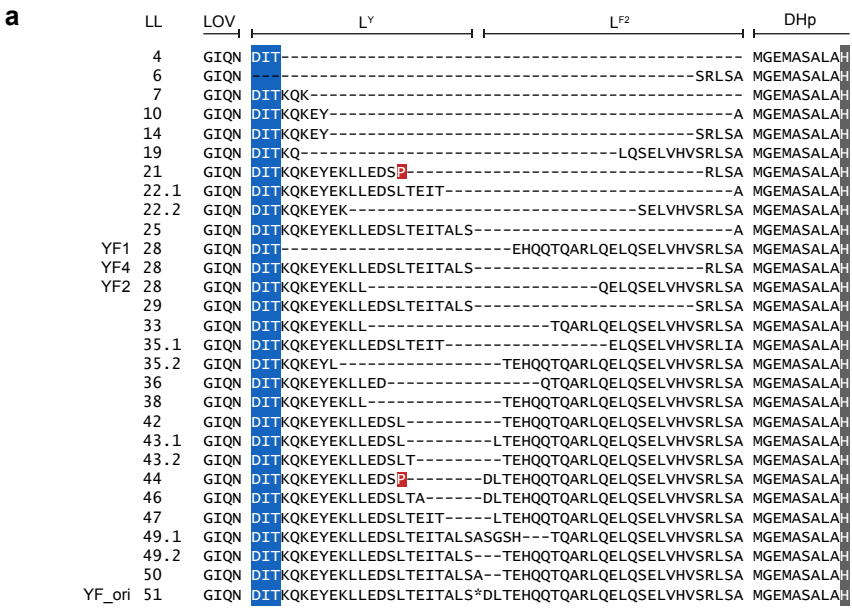
Construct	Linker length	Helical register	F/OD_{600} dark	F/OD_{600} 470 nm	Ratio
LL4	4	4	0.30 ± 0.04	0.60 ± 0.09	0.5 ± 0.1
LL22.1	22	1	0.014 ± 0.003	0.10 ± 0.02	0.14 ± 0.06
LL22.2	22	1	0.15 ± 0.02	0.65 ± 0.09	0.23 ± 0.07
LL36	36	1	0.15 ± 0.02	0.6 ± 0.1	0.24 ± 0.08
LL38	38	3	0.13 ± 0.01	1.5 ± 0.1	0.09 ± 0.02
LL43.1	43	1	0.016 ± 0.004	0.8 ± 0.1	0.02 ± 0.01
LL43.2	43	1	0.005 ± 0.003	0.21 ± 0.03	0.03 ± 0.02
LL46	45	4	0.067 ± 0.007	0.71 ± 0.07	0.09 ± 0.02
LL50	50	1	0.7 ± 0.1	1.6 ± 0.3	0.4 ± 0.1


a
sensor
linker
effector
1
124
147
261
BsYtvA
LOV
LY
STAS
1
128
143
255
282
505
BjFixL
PAS
LF1
PAS
LF2
DHp
CA
YF1
LOV
LF2
DHp
CA
YF_ori
LOV
LY
LF2
DHp
CA
b
pDusk
LacI^R
Y
F
FixJ
FixK2
DsRed
LOV
LY
LF2
DHp
CA
dark
470 nm

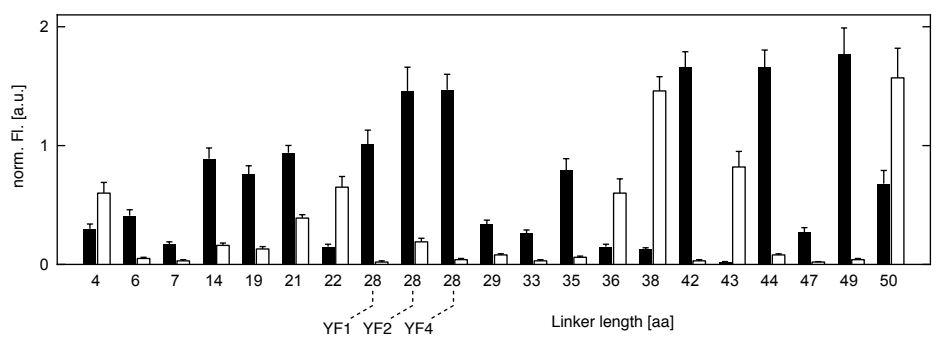

SHIPREC
ITCHY
PATCHY
RE
Gene B
Gene A
RE
Gene A
Gene B
RE
Gene A
Gene B
Δt 1,2,3...

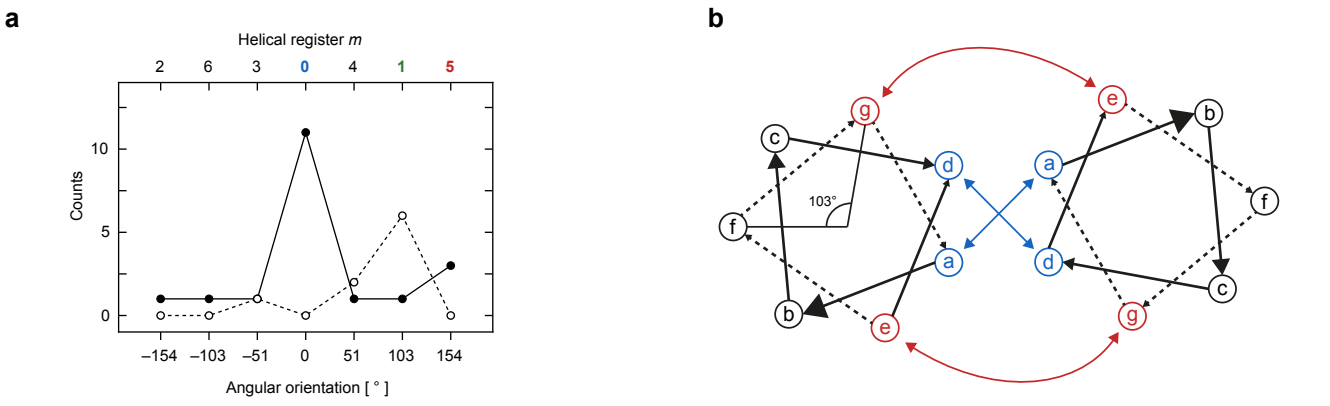

*Bj*FixL linker
*Bs*YtvA linker
145
140
135
130
125
280
275
270
265
260
255
freq.
1
50

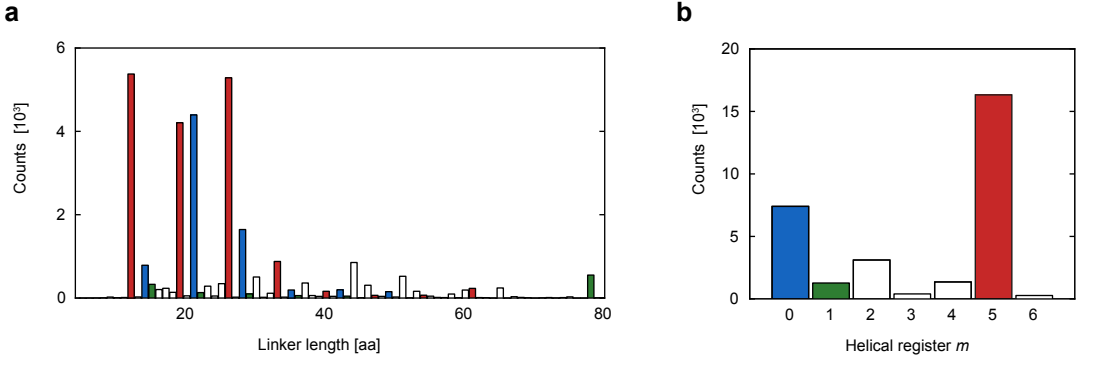
a

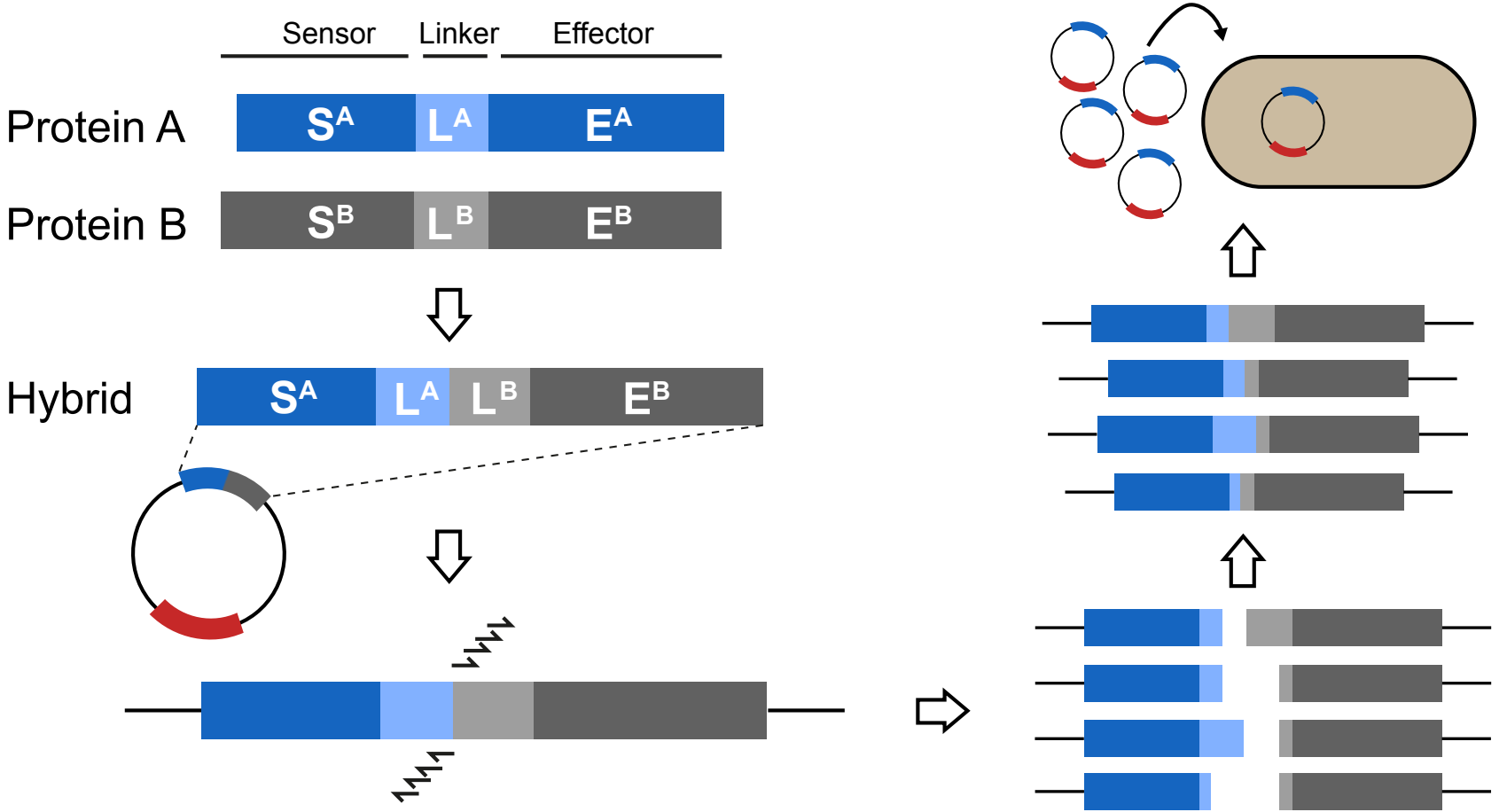

LL
LOV
L^Y
L^F2
DHp
4 GIQN DIT---------- MGEMASALAH
6 GIQN DIT---------- SRLSA MGEMASALAH
7 GIQN DITKQK------- MGEMASALAH
10 GIQN DITKQKEY---------- A MGEMASALAH
14 GIQN DITKQKEY---------- SRLSA MGEMASALAH
19 GIQN DITKQ---------- LQSELVHVSRLSA MGEMASALAH
21 GIQN DITKQKEYEKLLEDSP---------- RLSA MGEMASALAH
22.1 GIQN DITKQKEYEKLLEDSLTEIT---------- A MGEMASALAH
22.2 GIQN DITKQKEYEK---------- SELVHVSRLSA MGEMASALAH
25 GIQN DITKQKEYEKLLEDSLTEITALS---------- A MGEMASALAH
YF1 28 GIQN DIT---------- EHQQTQARLQELQSELVHVSRLSA MGEMASALAH
YF4 28 GIQN DITKQKEYEKLLEDSLTEITALS---------- RLSA MGEMASALAH
YF2 28 GIQN DITKQKEYEKLL---------- QELQSELVHVSRLSA MGEMASALAH
29 GIQN DITKQKEYEKLLEDSLTEITALS---------- SRLSA MGEMASALAH
33 GIQN DITKQKEYEKLL---------- TQARLQELQSELVHVSRLSA MGEMASALAH
35.1 GIQN DITKQKEYEKLLEDSLTEIT---------- ELQSELVHVSRLIA MGEMASALAH
35.2 GIQN DITKQKEYL---------- TEHQQTQARLQELQSELVHVSRLSA MGEMASALAH
36 GIQN DITKQKEYEKLLED---------- QTQARLQELQSELVHVSRLSA MGEMASALAH
38 GIQN DITKQKEYEKLL---------- TEHQQTQARLQELQSELVHVSRLSA MGEMASALAH
42 GIQN DITKQKEYEKLLEDSL---------- TEHQQTQARLQELQSELVHVSRLSA MGEMASALAH
43.1 GIQN DITKQKEYEKLLEDSL---------- LTEHQQTQARLQELQSELVHVSRLSA MGEMASALAH
43.2 GIQN DITKQKEYEKLLEDSLT---------- TEHQQTQARLQELQSELVHVSRLSA MGEMASALAH
44 GIQN DITKQKEYEKLLEDSP-------- DLTEHQQTQARLQELQSELVHVSRLSA MGEMASALAH
46 GIQN DITKQKEYEKLLEDSLTA------ DLTEHQQTQARLQELQSELVHVSRLSA MGEMASALAH
47 GIQN DITKQKEYEKLLEDSLTEIT----- LTEHQQTQARLQELQSELVHVSRLSA MGEMASALAH
49.1 GIQN DITKQKEYEKLLEDSLTEITALSASGSH--- TQARLQELQSELVHVSRLSA MGEMASALAH
49.2 GIQN DITKQKEYEKLLEDSLTEITALS--- TEHQQTQARLQELQSELVHVSRLSA MGEMASALAH
50 GIQN DITKQKEYEKLLEDSLTEITALSA-- TEHQQTQARLQELQSELVHVSRLSA MGEMASALAH
YF_ori 51 GIQN DITKQKEYEKLLEDSLTEITALS*DLTEHQQTQARLQELQSELVHVSRLSA MGEMASALAH

b


norm. Fl. [a.u.]
2
1
0
4 6 7 14 19 21 22 28 28 28 29 33 35 36 38 42 43 44 47 49 50
YF1 YF2 YF4
Linker length [aa]


a
Helical register m
2
6
3
0
4
1
5
Counts
10
5
0
–154
–103
–51
0
51
103
154
Angular orientation [°]
b
g
e
c
d
a
b
f
103°
f
a
d
b
e
g
c


a
Counts [10³]
6
4
2
0
20
40
60
80
Linker length [aa]
b
Counts [10³]
20
15
10
5
0
0
1
2
3
4
5
6
Helical register m


Sensor
Linker
Effector
Protein A
S^A
L^A
E^A
Protein B
S^B
L^B
E^B
Hybrid
S^A
L^A
L^B
E^B