Author Manuscript

Title: Reversible and Tunable Photoswitching of Protein Function through Genetic Encoding of Azobenzene Amino Acids in Mammalian Cells

Authors: Ji Luo; Subhas Samanta; Marino Convertino; Nikolay Dokholyan; Alexander Deiters

To be cited as: ChemBioChem 10.1002/cbic.201800226

Link to VoR: https://doi.org/10.1002/cbic.201800226

Reversible and Tunable Photoswitching of Protein Function through Genetic Encoding of Azobenzene Amino Acids in Mammalian Cells

Ji Luo,[a] Subhas Samanta,[a] Marino Convertino,[b] Nikolay V. Dokholyan,[b] and Alexander Deiters*[a]

Abstract: The genetic encoding of three different azobenzene phenylalanines with different photochemical properties was achieved in human cells using an engineered pyrrolysyl tRNA/tRNA synthetase pair. In order to demonstrate reversible light-control of protein function, azobenzenes were site-specifically introduced into firefly luciferase. Computational strategies were applied to guide selection of potential photoswitchable sites that lead to a reversibly controlled luciferase enzyme. Direct reversible photoswitching of enzymatic function in live cells has been accomplished through genetically encoded photoswitchable amino acids. In addition, the new azobenzene analogs provide enhanced thermal stability, high photoconversion, and responsiveness to visible light. These small molecule photoswitches can reversibly photocontrol protein function with excellent spatiotemporal resolution and preferred sites for incorporation can be computationally determined, thus providing a new tool for investigating biological processes.

Introduction

As a classical light-switchable motif, azobenzene has found extensive biological applications.[1] It undergoes light-induced, wavelength-selective *cis*/*trans* isomerizations with a large quantum yield and resulting significant structural differences between the two isomers. Azobenzene derivatives have been shown to enable reversible photocontol of the function of nucleic acids, peptides, enzymes, receptors, and ion channels in cells and in animals.[2] In order to improve the azobenzene photoswitches that can be isomerized upon irradiation with visible light, substitutions on both benzene rings have been introduced. One approach is to red-shift the π → π* band of the *cis*/*trans*-isomer, resulting in good switching properties.[3] Another possibility is to enhance the thermal stability of the *trans*-isomer through lowering the energy of the n orbital and thereby increasing the energy of the n → π* transition.[4]

The Schultz and Wang groups both demonstrated site-specific incorporation of azobenzene photoswitches into proteins via genetic code expansion. Schultz *et al.* used an orthogonal *Methanococcus jannaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS)/tRNA^{Tyr} pair to incorporate the azophenylphenylalanine **AzoF** into the bacterial catabolite activator protein (CAP) to control its promoter-binding activity in biochemical assays.[5] The Wang group developed photoswitchable amino acids capable of bioconjugation to nearby cysteine residues using an evolved *Methanosarcina mazei* pyrrolysyl-tRNA synthetase (*Mm*PylRS) and demonstrated reversible alteration of calmodulin conformation and of NMDA receptor function.[6] Using the same *Mm*PylRS, the Lin lab reported a series of red-shifted azobenzene amino acids and incorporated the azobenzene derivatives into sfGFP in *E. coli*.[7]

Here, we report the ability to apply the parent **AzoF** amino acid to reversibly control protein function in mammalian cells (Figure 1). Furthermore, the two new azobenzene derivatives **F$_2$AzoF** and **F$_4$AzoF** were genetically encoded in bacterial and mammalian cells and provide improved optical properties and greatly increased thermal stability of the *cis* isomer. Their design was inspired by observations[3a, 8] that fluoro substitutions at the *ortho*-position significantly increased the stability of the *cis* form of azobenzenes without increasing the overall size of the molecule. Site-specific installation of **AzoF**, **F$_2$AzoF**, and **F$_4$AzoF** into a target protein was guided by computational studies and was applied to the reversible photocontrol of protein function with spatiotemporal resolution in live cells.

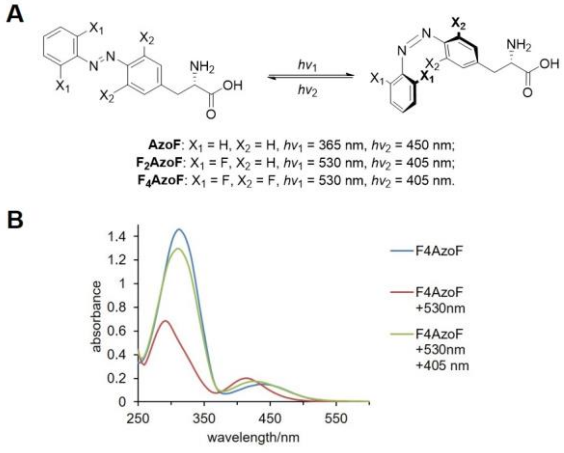


Figure 1. A) Photoisomerization of azobenzene analogs **AzoF**, **F$_2$AzoF**, and **F$_4$AzoF**. B) UV/Vis spectra of the *trans*-photoisomer (>99% *trans*, blue, in the dark-adapted state), *cis*-photoisomer (91% *cis*, red, 530 nm), and *trans*-photoisomer (84% *trans*, green, 405 nm) of **F$_4$AzoF** in PBS (pH 7.4).

Results and Discussion

To comprehensively determine the photoswitching efficiency of the three azobenzene amino acids, absorption spectra were recorded and *cis*/*trans* ratios of the corresponding

[a] Ji Luo, Dr. Subhas Samanta, Prof. Dr. Alexander Deiters
University of Pittsburgh, Department of Chemistry, Pittsburgh, PA 15260, United States
E-mail: deiters@pitt.edu
[b] Marino Convertino, Prof. Dr. Nikolay V. Dokholyan
University of North Carolina at Chapel Hill, Department of Biochemistry and Biophysics, Chapel Hill, NC 27599, United States

Supporting information for this article is given via a link at the end of the document.

photostationary states were determined by HPLC analysis (Table 1 and Supporting Figures S1–3).[1b, 3a] Introducing electron-withdrawing fluorines *ortho* to the azo group effectively separates the n → π* bands of the *trans* and *cis* isomers, allowing for reversible isomerization using visible light. Substitution with fluorine also provides more favorable photostationary states and enhances thermodynamic stability of the *cis* isomer. Notably, tetrafluoro substitution (as in **F_4AzoF**) leads to distinct separation in excitation wavelength of both π → π* and n → π* transition bands between the *trans* and *cis* isomer (Supporting Figure S4).

Table 1. Measured absorption maxima (λ_{max}, orange), initial dark state (green), and photostationary states (blue) of azobenzene amino acids in DMSO/PBS (10/90, pH 7.4) at the indicated wavelengths.

UAA	**AzoF**	**F_2AzoF**	**F_4AzoF**
λ_{max} *trans* isomer	335 nm (π→π*)	435 nm (n→π*)	447 nm (n→π*)
λ_{max} *cis* isomer	300 nm (π→π*)	420 nm (n→π*)	417 nm (n→π*)
trans:*cis* *(dark state)*	99:1	97:3	99:1
trans:*cis* (hv_1)	22:78 (365 nm)	18:82 (530 nm)	9:91 (530 nm)
trans:*cis* (hv_2)	76:24 (450 nm)	70:30 (405 nm)	84:16 (405 nm)

Genetic code expansion utilizes an orthogonal aminoacyl-tRNA synthetase/tRNA pair to selectively incorporate an unnatural amino acid (UAA) into proteins, in response to an unassigned UAG amber codon introduced at a desired site into a gene of interest.[9] It enables site-specific incorporation of diverse UAAs into proteins in cells and animals by engineering the translational machinery.[10] In our lab, we utilize *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase / $tRNA_{CUA}$ (*Mb*PylRS/$tRNA_{CUA}$) pairs, which are functional and orthogonal in both bacteria and eukaryotic cells. To genetically incorporate **AzoF** into proteins in mammalian cells, a screen was carried out of a small panel of ten *Mb*PylRS mutants, which were created through introduction of select mutations at six positions (L270, Y271, L274, N311, C313, and Y349) and have previously been shown to accept structurally diverse phenylalanine derivatives as substrates.[11-12] The synthetase panel was screened in mammalian cells using a plasmid encoding a dual fluorescent reporter gene: mCherry-EGFP containing a TAG mutation on the linker between mCherry and EGFP.[13] Two PylRS mutants led to EGFP expression in the presence of **AzoF** (Figure 2A). The first PylRS, named AzoFRS1, has five mutations: Y271M, L274A, N311A, C313A, and Y349F. The second one, termed AzoFRS2, contains L270F, L274M, N311G, C313G, and Y349F mutations (Supporting Figure S5). Fluorescence imaging revealed that mCherry expressions were observed as expected and that AzoFRS1/2 showed no EGFP expression in the absence of **AzoF**. Strong EGFP expression was detected when **AzoF** (0.25 mM) was added to the media. This confirms the fidelity of AzoFRS1/2 for incorporation of **AzoF** in mammalian cells. Because fluorine has a similar size to hydrogen and the symmetric fluoro-substituents minimally change the planar geometry of the *trans* isomer,[3a, 14] **F_2AzoF** and **F_4AzoF** should not significantly increase the steric demand on the synthetase. Thus, incorporation of **F_2AzoF** and **F_4AzoF** into proteins in mammalian cells was first tested using the same *Mb*AzoFRS1/tRNA pair. Unexpectedly, very low EGFP expression was observed, suggesting that AzoFRS1 did not allow for incorporation of **F_2AzoF** and **F_4AzoF** in mammalian cells. Gratifyingly, the AzoFRS2 synthetase was found to incorporate not only **AzoF**, but also **F_2AzoF** and **F_4AzoF** with high efficiency. Overall, mCherry and EGFP expression was observed in the presence of **AzoF**, **F_2AzoF** and **F_4AzoF** (0.25 mM), while no EGFP-expressing cells were observed in the absence of the UAAs (Figure 2A). In addition, full-length mCherry-EGFP protein was detected in HEK293T cells using an HA-tag antibody, confirming full-length protein expression at good levels (Figure 2B). Subsequently, incorporation of **AzoF**, **F_2AzoF**, and **F_4AzoF** into protein in *E. coli* was investigated using *Mb*AzoFRS2. As expected, it showed robust levels of sfGFP-Y151TAG expression in the presence of the three amino acids (1 mM), while no sfGFP expression in the absence of the azobenzene amino acids was observed (Figure 2C). The expression yields of sfGFP-**AzoF**, -**F_2AzoF**, and -**F_4AzoF** are 8.0 mg/L, 3.2 mg/L, and 1.7 mg/L, respectively. The incorporation was further confirmed by ESI-MS analysis of purified proteins, revealing a mass of [sfGFP-**AzoF**] and [sfGFP-**AzoF** + Na^+] of 28350.137 Da and 28373.461 Da (expected MS: 28350.20 Da with Na^+), a mass of [sfGFP-**F_2AzoF**] and [sfGFP-**F_2AzoF** + Na^+] of 28386.805 Da and 28410.496 Da (expected MS: 28386.20 Da with Na^+), and a mass of 28423.633 Da (calculated MS: 28422.80 Da) for incorporation of sfGFP-**F_4AzoF** (Supporting Figure S6).

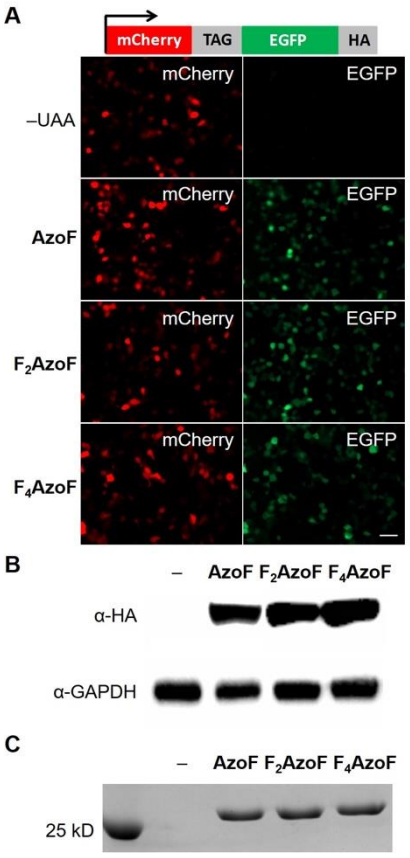


Figure 2. Genetic encoding of azobenzene amino acids in pro- and eukaryotic cells. A) Fluorescence micrographs of HEK293T cells expressing mCherry-EGFP in the presence of p*Mb*AzoFRS2-mCherry-TAG-EGFP-HA and PyltRNA$_{CUA}$ incubated in the absence or presence of **AzoF**, **F$_2$AzoF**, and **F$_4$AzoF**. B) Western blots of mCherry-TAG-EGFP-HA with an anti-HA tag antibody and an anti-GAPDH antibody proving the fidelity of the incorporation of the three amino acids in response to a TAG codon in mammalian cells. C) Genetic incorporation of **AzoF**, **F$_2$AzoF**, and **F$_4$AzoF** using *Mb*AzoFRS2 in *E. coli*. Coomassie stained gel of Ni-NTA purified sfGFP-Y151TAG expressed in the absence and presence of **AzoF**, **F$_2$AzoF**, and **F$_4$AzoF**.

Firefly luciferase was selected as a proof-of-principle target in order to demonstrate the utility of the photoswitchable amino acid to reversibly photocontrol protein function in live mammalian cells. Photoactivatable luciferases have been developed through the introduction of photocleavable caging groups by us and by the Chen lab.[15] Yet, reversible photocontrol of a bioluminescence enzyme has not been achieved. A genetically encoded photoswitchable luciferase provides an additional tool for the cellular detection of ATP concentrations at different locations and different time points.[16]

In order to efficiently identify amino acids in potentially allosteric regions of FLuc useful for the incorporation of **AzoF** without the need for extensive mutational analysis, computational structural modeling and protein stability calculations were carried out for the three different conformational states of firefly luciferase in the two-step chemiluminescence reaction: the open inactive state (state A), the closed active state for the generation of luciferyl-AMP (state B), and the closed active state for the generation of oxyluciferin (state C).

We evaluated the conformational flexibility of FLuc residues by using Eris, an in house-developed software that efficiently and accurately computes the stability changes of proteins upon mutations and/or conformational exploration (i.e., local repacking) of the three-dimensional orientation of amino acid side chains.[17] FLuc residues characterized by a high degree of local repacking are likely to constitute an allosteric site of the protein,[18] and, thus, potential candidate for the incorporation of **AzoF**. The magnitude of the amino acid local repacking can be quantified by determining the standard deviation (σΔG) of the protein conformations' free energy computed as estimated by Eris. The higher is the number of conformational states that theoretically accessible to a specific amino acid, the higher is its estimated σΔG value. For a given residue i, the statistical significance of its σΔG value is determined by its $Z-score_i$ defined as:

$$Z-score_i = \left(\sigma\Delta G_i - \overline{\overline{\sigma\Delta G}}\right)/S$$

where $\sigma\,\Delta\,G_i$ is the magnitude of the local repacking around the residue i, $\sigma\,\Delta\,G$ is the average magnitude of local repacking over all protein residues, and S is the relative standard deviation. FLuc residues for which the Z-score value is greater than two have less than 97.7% chance to be randomly identified as highly flexible (i.e., high local repacking), therefore they are likely part of an allosteric site of the protein.[19] Consequently, they are excellent candidates for substitution with **AzoF** whose light-mediated activation would translate the configurational changes upon light-switching into changes in protein/enzyme function. Our calculations, which require less than 48 hours to be completed (see experimental section), predict a series of allosteric residues in luciferase, and we targeted aromatic residues at those sites for **AzoF** substitution in order to improve our chances for obtaining a photoswitchable luciferase enzyme. Within the list of statistically significant potential candidates (Table 2), F294 and F432 both have high Zscore values (3.06 and 3.76, respectively); however, both are followed by sterically demanding F295 and F433 residues, and based on the protein structure we expect severe steric hindrance in incorporating sterically demanding azophenylphenylalanine analogs. Y340 (Z-score 3.46) is an essential catalytic residue in the ATP binding pocket of FLuc and thus was not modified.[20] W426 (Z-score 3.14) is very close to the flexible N-terminal of FLuc, thus suggesting that it is not a good candidate to achieve optical switching of enzyme function. Of the remaining possible sites, Y280TAG, F294TAG, Y340TAG, W417TAG, W426TAG, F432TAG, and F433TAG were finally selected for experimental validation.

Table 2. Potential residue candidates for **AzoF** incorporation in luciferase (Z-score higher than two).

Residue	$\sigma_{\Delta G}$ (kcal/mol)	Z-score
F268	1.98	2.15
F273	2.17	2.48

Y280	2.44	2.95
F292	2.39	2.85
F294	2.51	3.06
Y304	2.12	2.40
Y340	2.74	3.46
W417	2.40	2.88
W426	2.55	3.14
F432	2.91	3.76
F433	2.46	2.97
Y444	1.96	2.13
Y447	2.37	2.83
F465	2.25	2.62

Light-switching luciferase assays in HEK293T cells were carried out by incorporation of **AzoF** into FLuc at seven different sites: Y280TAG, F294TAG, Y340TAG, W417TAG, W426TAG, F432TAG, and F433TAG (Figure 3). To genetically incorporate **AzoF** into FLuc, cells were co-transfected with the p*Mb*PylAzoFRS2/PyltRNA$_{CUA}$ expression plasmid and the corresponding pFLuc-TAG construct in the absence or presence of **AzoF**. After 24 h incubation, the transfected cells were treated with UV light (5 min, 20 mW/cm^2) or kept in the dark and FLuc activities were subsequently measured. Not surprisingly, most mutations either showed activity regardless of UV illumination or showed complete loss of activity, except for FLuc-W280→**AzoF** and FLuc-W417→**AzoF**, which appear to be allosteric sites that allow for photoswitching of FLuc activity. The activity of wild-type Fluc was not affected by UV exposure (Supporting Figure S8). The experimental identification of two out of seven predicted sites supports the use of the described computational protocol for the rapid detection of potential amino acid candidates fort he incorporation of photoswitchable motifs. Subsequent efforts were focused on the W417 mutation due to its highest overall luciferase activity.

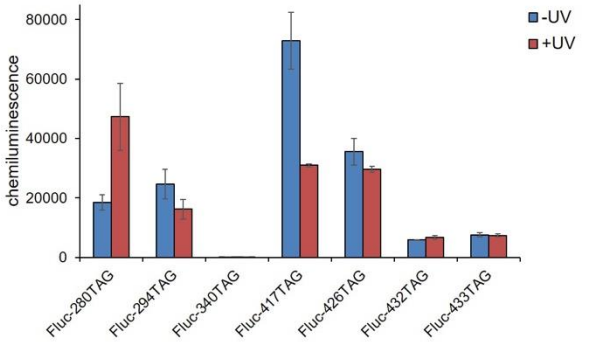


Figure 3. Mutant screening for reversible photoswitching of firefly luciferase containing **AzoF** in live mammalian cells. Error bars represent standard deviations from three independent experiments.

As revealed by Eris calculations, the introduction of the azophenylphenylalanine at position W417 may cause a decrease in the stabilization of the luciferase, particularly pronounced in state B and for the **AzoF** *cis*-isomer (Supporting Figure S7), which may affects the conformational change from state B to C or from state B to A. Thus, the light-triggered *cis*/*trans*-isomerization of azobenzene can affect the stability of the luciferase-luciferyl-AMP complex, which is formed in the first step of luciferin oxidation, thereby interfering with the oxidation process. Furthermore, the binding energy of D-luciferyl-AMP, as evaluated by MedusaScore,[21] an accurate force-field based scoring function which estimates binding energies by modeling physical interactions between proteins and small molecules, is slightly less favorable in the case of *cis*-FLuc-W417→**AzoF** in both the B and C states, making the formation of the protein-ligand complex less favorable for the bioluminescence reaction. Specifically, FLuc-W417 → *cis*-**AzoF** shows higher (i.e., less favorable) interaction energy in states B and C (–30.9 kcal/mol and –36 kcal/mol, respectively) compared to wild-type FLuc (–31.9 kcal/mol and –37.0 kcal/mol, respectively) and FLuc-W417 → *trans*-**AzoF** (–33.5 kcal/mol and –41.9 kcal/mol; Table 3). Based on the computational calculations, when W417→*trans*-**AzoF** isomerizes to the *cis* conformation, we speculate that the amino acids within the luciferase binding site rearrange, resulting in overall destabilization of the protein, leading to higher (less favorable) interaction energy with the natural substrate. This may explain why the *cis*- and *trans*-W417→**AzoF** luciferase mutants display successive off-and-on switching behaviors upon light illumination.

Table 3. Binding energy of D-luciferyl-AMP upon repacking at W417 for the wild-type enzyme in the AzoF mutant in the *trans* and *cis* geometry.

State	Geometry	MedusaScore (kcal/mol)
B	wild-type	–31.9
	cis-AzoF	–30.9
	trans-AzoF	–33.5
C	wild-type	–37.0
	cis-AzoF	–36.0
	trans-AzoF	–41.9

In order to demonstrate reversible photoswitching of protein function, HEK293T cells were subjected to alternating UV and white light exposures for five individual steps (5 min each). The activity of FLuc-W417→**AzoF** showed the expected reversible change with each step of light exposure, as activity decreased after UV-induced *trans* to *cis* photoswitching and was restored after *cis* to *trans* switching through visible light exposure (Figure 4A). The two different light treatments did not affect the activity of FLuc-TAG in the absence of **AzoF** (negative control; Supporting Figure S7). These results support that **AzoF** is functioning as a reversible light-switch for protein activity in live

human cells. The activity of the Fluc-W417→**AzoF** mutant was reduced in the case of the *cis*-isomer, while reversion to the *trans*-isomer showed recovery of activity (Figure 4A).

To investigate whether the protein activity could be reversibly photoswitched for spatial control, two culture dishes of HEK293T cells expressing FLuc-W417→**AzoF** were irradiated alternatingly through a mask that covered half of each dish. In one dish (Figure 4B, left), the left side was illuminated with visible white light (>400 nm), while the right one was irradiated with UV (365 nm). In the other dish (Figure 4B, right), the left half was illuminated with UV (365 nm) followed by white light (>400 nm), while the right one was treated sequentially by UV (365 nm), white (>400 nm), and UV (365 nm) light. Thus, the half-dishes from left to right represent areas with increasing numbers of localized, reversible switching events. Cells that were exposed to visible white light (>400 nm) exhibited increased luminescence (*trans*-**AzoF** activates Fluc), while UV-exposed cells showed diminished luminescence (*cis*-**AzoF** inactivates Fluc), resulting in a visible "half-moon" pattern (Figure 4B). These spatial-control results are in agreement with the Bright-Glo luciferase assay results (Figure 4A) and demonstrate that **AzoF** can be efficiently and reversibly switched *in vivo*, allowing reversible control of firefly luciferase activity with temporal and spatial resolution.

The drawbacks of **AzoF** include: 1) photoisomerization requires UV light, and 2) the *cis*-isomer is thermally unstable and reverts back to *trans* even in the absence of irradiation. To address these shortcomings, fluorine atoms were installed at the four *ortho* positions to induce a bathochromic shift that makes **F_4AzoF** responsive to visible light, and also greatly improves the thermal stability of its *cis*-isomer and the isomer ratios in the photostationary states.[3a] In order to demonstrate the thermostability of the *cis*-isomer of **F_4AzoF** after incorporation into protein, the luciferase assays in live cells were either performed immediately or one day after light exposure to 530 nm (*trans* → *cis* switching). Importantly, the enzymatic activity for FLuc-W417→*cis*-**F_4AzoF** remains low due to stability of the *cis* isomer – even after 24 hours in live cells in a 37 °C incubator (Figure 5B). This is in contrast to *cis*-**AzoF**, which is fully converted back to *trans*-**AzoF** after 24 h incubation at 37 ºC in the dark (Figure 5A). The ability to choose light-switchable amino acids with fast (**AzoF**) and slow (**F_4AzoF**) thermal reversion rates enables switching experiments on short and long timescales, thereby further expanding the optochemical toolbox of light-controlled amino acids. In addition, the genetic encoding of **F_4AzoF** provides more suitable switching wavelengths, red-shifted from the UV-A range.

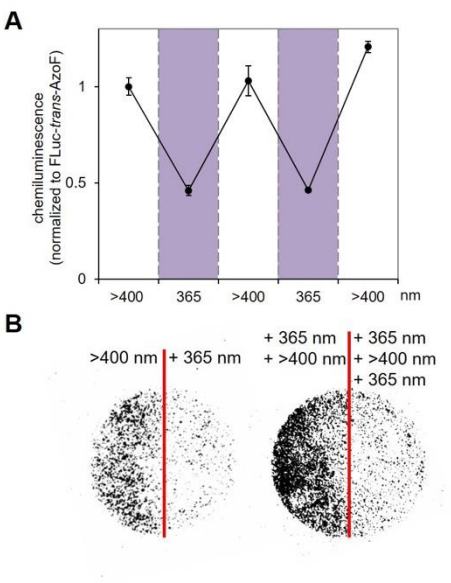


Figure 4. Reversible photoswitching of Firefly luciferase containing a W417**AzoF** mutation in live mammalian cells. A) HEK293T cells underwent five light-switching steps in total: >400 nm (visible light) → 365 nm (UV light) → >400 nm → 365 nm → >400 nm. Chemiluminescence was normalized to the activity of FLuc-W417→*trans*-**AzoF** that was kept in the dark. Error bars represent standard deviations from three independent experiments. B) Spatial control of reversible photoswitching of FLuc activity in live cells by exposing two halves of culture dishes containing transfected HEK293T cells to increasing numbers of switching events as indicated by the different wavelengths for visible light (>400 nm) and UV (365 nm) light.

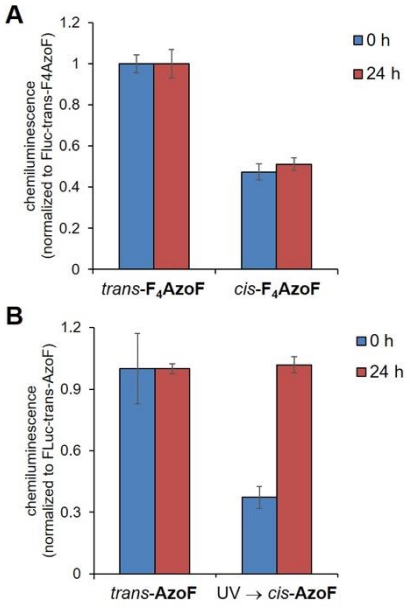


Figure 5. Switching experiment confirming that (B) *cis*-**AzoF** is not, but that (A) *cis*-**F_4AzoF** is thermally stable when incorporated into protein in HEK293T cells. Cells were kept in the dark before light irradiation. After 24 h post-transfection, cells were irradiated by 530 nm LED (A) or 365 nm UV light (B), and then incubated for 24 h at 37 ºC in the dark. Chemiluminescence was normalized to non-irradiated cells. Error bars represent standard deviations from three independent experiments.

The activity of FLuc-W417→**F_4AzoF** was reversibly controlled through a *trans* → *cis* and *cis* → *trans* configurational change of **F_4AzoF** upon 530 nm and 405 nm irradiation, respectively.

Furthermore, photoswitching of FLuc activity was reproducible over five cycles (Figure 6), demonstrating that **F_4AzoF** was functioning as a reversible photoswitch for FLuc activity in live cells. **F_4AzoF** was irradiated with blue light (405 nm) in order to obtain a PSS of 84% *trans*-**F_4AzoF** through photoisomerization, as reported in Table 1. Thus, in the first step, the dark-adapted *trans*- **F_4AzoF** was exposed to 405 nm light so that all cells contained active FLuc at the same level. Cells were then exposed to green light (530 nm) followed by blue light (405 nm) two successive times, inducing isomerization of **F_4AzoF**, effectively turning "off and on" FLuc function, allowing photoswitchable control of the enzymatic activity (Figure 6).

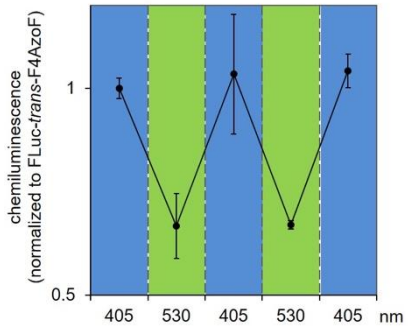


Figure 6. Reversible photoswitching of FLuc-W417→**F_4AzoF** activity in live mammalian cells. Cells underwent five light-switching steps in total: 405 nm → 530 nm → 405 nm → 530 nm → 405 nm. Error bars represent standard deviations from three independent experiments.

Conclusions

In summary, the three reversibly photoswitchable azobenzene amino acids **AzoF**, **F_2AzoF**, and **F_4AzoF** were site-specifically incorporated into proteins in both pro- and eukaryotic cells through genetic code expansion. The parent, unsubstituted **AzoF** requires UV irradiation for *trans* to *cis* photoswitching and spontaneously ($t_{1/2}$ at 23 °C is 12.8 h) reverts back to the thermodynamically more stable *trans* isomer under physiological conditions, even when kept in the dark. This can be a useful property when biological processes on a similar timescale are being optically controlled; however, it may impact control of protein function at longer time scales and may complicate experimental designs and interpretation of experimental outcomes. In contrast, the half-life of *cis*-**F_4AzoF** ($t_{1/2}$ at 25 °C is 2 years)[4] is dramatically increased, reducing thermal relaxation and showing complete stability for at least 24 h in proteins in live human cells. Due to the enhanced thermal stability of the *cis*-**F_4AzoF**, it could be used to investigate long term biological events as well as for kinetics experiments in live animals. Moreover, the **F_4AzoF** amino acid shows a red-shifted absorption spectrum enabling rapid photochemical *trans* to *cis* and *cis* to *trans* switching with visible light at 530 nm and 405 nm, respectively. This is the result of the electron-withdrawing effect of the fluorine substituents, which adjusts the energetics of the n → π* transition in **F_4AzoF**, enabling photoswitching with visible light. The responsiveness of **F_4AzoF** to the visible light avoids negative effects induced by UV light and makes it more suitable for *in vivo* applications. Moreover, **F_4AzoF** displays excellent *trans* → *cis* isomerization (up to 91%). The two n → π* bands of *trans*- and *cis*-**F_4AzoF** are well separated, enabling the selective isomerization of both isomers with enhanced photoconversions, compared to the overlap of the n → π* bands of the **AzoF** isomers (78% of *trans* → *cis* isomerization).

Site-specifically inserting azobenzene analogs into proteins in cells with an expanded genetic code enables the optical activation and deactivation of protein function in a reversible fashion using light illumination with two different wavelengths. However, selecting an amino acid residue for substitution is not trivial and usually involves extensive trial-and-error mutagenesis. In order to place this process on a firmer, rational foundation, a computational structural modeling approach was implemented. In order to narrow the set of potential installation sites on a given protein, we determined areas that provide a high degree of local repacking (i.e., potential allosteric sites) and thus may have an impact on protein function when perturbed through azobenzene photoisomerization. Specifically, seven predicted (and suitable) sites on firefly luciferase were experimentally tested, and one (W417) provided reversible light-switching of enzymatic activity when used for azobenzene incorporation. FLuc activity can be controlled through the isomerization of azobenzene, incorporated at W417, upon illumination. The *cis* conformation of azobenzene reduces FLuc function, while the *trans* conformation fully restores FLuc activity, providing reversible, photoswitchable control over enzymatic activity. Reversible on and off switching of luciferase activity for up to five cycles was demonstrated. In addition to reversible temporal control of protein function, reversible spatial control of protein activation and deactivation was demonstrated as well. This approach of combining genetic code expansion with synthetic chromophore engineering on protein residues identified through computational structural modeling adds a highly versatile system to the optogenetic toolbox for the light-regulation of protein function in cells and animals.

Experimental Section

Incorporation of azobenzene analogs in mammalian cells. HEK 293T cells were seeded at ~50,000 cells per well and grown in Dulbecco's Modified Eagle's Medium (200 µL, DMEM, Gibco) supplemented with FBS (Sigma, 10%), Pen-Strep (Corning Cellgro, 1%) and L-glutamine (Alfa Aesar, 2 mM) in 96-well plates (Greiner) under a humidified atmosphere containing CO_2 (5%) at 37 °C. Cells were transiently transfected with p*Mb*AzoFRS2-mCherry-TAG-EGFP-HA and p4CMVE-U6-PylT (100 ng of each plasmid) at ~80% confluency using linear polyethylenimine (LPEI; 1.5 µL, 0.323 mg/mL) in DMEM (200 µL) without any unnatural amino acid or with **AzoF**, **F_2AzoF**, and **F_4AzoF** (0.25 mM). After a 24 h incubation at 37 °C, the media were replaced with PBS and the cells were imaged with a Zeiss Axio Observer Z1 Microscope (10× objective) using EGFP (38HE: Ex 470/40; Em 525/50) and mCherry (43HE: Ex 550/25; Em 605/70) filter cubes. To further confirm the fidelity of incorporation and the expression of the fusion protein, Western blots were performed. HEK 293T cells were co-transfected with p*Mb*AzoFRS2-

mCherry-TAG-EGFP-HA and p4CMVE-U6-PylT (1.5 µg of each plasmid) using LPEI (15 µL, 0.323 mg/mL) in the presence or in the absence of **AzoF**, **F_2AzoF**, and **F_4AzoF** (0.25 mM) in 6-well plates. After a 24 h incubation, the cells were washed with chilled PBS (2 mL) and lysed in mammalian protein extraction buffer (200 µL, GE Healthcare) with complete protease inhibitor cocktail (Sigma) through shaking at 4 °C for 15 min, and the cell lysates were cleared through centrifugation (14,000 g, 4 °C, 20 min). The protein lysates were boiled with loading buffer and analyzed by 10% SDS-PAGE (60 V for 20 min, and then 150 V for 1 h). After gel electrophoresis and transferring to a polyvinylidene difluoride (PVDF) membrane (80 V for 1.5 h, GE Healthcare), the membrane was blocked in TBS (5 mL) with Tween 20 (Fisher Scientific, 0.1%) and milk powder (5%) for 1 h. The membranes were probed and incubated with the primary α-HA antibody (Y-11) rabbit monoclonal IgG (sc-805, Santa Cruz Biotech) overnight at 4 °C, followed by a goat anti-rabbit IgG-HRP secondary antibody (sc-2031, Santa Cruz Biotech) for 1 h at room temperature. The membranes were washed with TBS with Tween 20 (0.1 %) three times, and then incubated in working solution (Thermo Scientific SuperSignal West Pico Chemiluminescent substrate) for 5 min. The Western blots were imaged using the "chemi" autofocus settings on a ChemiDoc (BioRad).

Photoswitching of firefly luciferase. HEK 293T cells (~50,000 cells per well) were cultured in DMEM (Gibco) supplemented with FBS (Sigma, 10%), Pen-Strep (Gibco, 1%) and L-glutamine (Alfa Aesar, 2 mM) in 96-well plates (Greiner) under a humidified atmosphere containing CO_2 (5%) at 37 °C. At ~80% confluency, cells seeded on plates were co-transfected with p*Mb*AzoFRS2-4PylT and pFLuc-W417TAG (100 ng of each plasmid) using LPEI (1.5 µL, 0.323 mg/mL), and media was changed to fresh DMEM supplemented without or with **AzoF** or **F_4AzoF** (0.25 mM, 200 µL). After double transfection and incubation (24 h), the media was changed to DMEM without unnatural amino acids. Subsequently, cells were irradiated alternating between two light sources consecutively five times – alternating between 365 nm UV light for 5 min (Dual UV transilluminator) and visible white light for 5 min (JUST Normlicht) for **AzoF**, or between 530 nm light (LUMILEDS LXML-PM01-0100) for 30 s per well and 405 nm light (LEDENGIN LZ1-10UA00-00U8) 30 s per well for **F_4AzoF**. After light irradiation, cells were lysed by addition of 100 µL of substrate solution (Promega) in a 96-well plate (BD Falcon), and luminescence was measured using a Bright-Glo luciferase assay kit (Promega) and a microplate reader (Tecan M1000) with an integration time of 1 s.

Spatial control of firefly luciferase. HEK293T cells at ~80% confluency were co-transfected with pMbAzoFRS2-4PylT and pFLuc-W417TAG (2 µg of each plasmid) using LPEI (15 µL, 0.323 mg/mL) in the presence of **AzoF** (0.25 mM) in 35-mm dishes (WillCo-Dish). After a 24 h incubation, the transfected cells were washed three times using PBS (pH 7.4) to remove excess **AzoF**. Plates were exposed to either UV (365 nm UV transilluminator, VWR) or white light (JUST Normlicht, RPimaging) irradiation for 5 min, while the other half of the plate was protected using aluminum foil, and then the plates were imaged using a ChemiDoc system (Bio-Rad).

Protein expression in *E. coli*. The plasmid pBAD-sfGFP-Y151TAG-pylT[22] was co-transformed with pBK-AzoFRS2 into chemically competent *E. coli* Top10 cells. A single colony was grown in LB media overnight and 400 µL of the overnight culture was added to 25 mL LB media supplemented with 1 mM of **AzoF**, **F_2AzoF**, and **F_4AzoF** and 25 µg/mL tetracycline and 50 µg/mL kanamycin. Cells were grown at 37 °C, 250 rpm, and protein expression was induced with 0.1% arabinose when OD_{600} reached ~0.5. After overnight expression at 37 °C, cells were harvested and washed with PBS. Cell pellets were re-suspended in 6 mL of phosphate lysis buffer (50 mM, pH 8.0), Triton X-100 (60 µL, 10%) and protease inhibitor (6 µL, Sigma), gently mixed, and incubated on ice for 1 h. Cell mixtures were sonicated (Fisher Scientific 550 Sonic Dismembrator) with six short bursts of 30 s on ice followed by 30 s intervals for cooling, and then the cell lysates were centrifuged at 4 °C and 13,000 g for 10 min. The supernatant was transferred to a 15 mL conical tube and 100 µL Ni-NTA resin (Qiagen) was added. The mixture was incubated at 4 °C for 2 h under mild shaking. The resin was then collected by centrifugation (1,000 g, 10 min) at 4 °C, washed three times with 300 µL of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0), followed by two washes with 300 µL of wash buffer (lysis buffer containing 20 mM imidazole). The protein was eluted twice with 200 µL of elution buffer (lysis buffer containing 250 mM imidazole). The purified proteins were analyzed by 10% SDS-PAGE (60 V for 20 min, and then 150 V for 1 h), and stained with Coomassie Blue. The protein mass was obtained by electrospray ionization mass spectrometry (Figure S6).

Computational calculations and modeling, *Protein structural model generation.* In order to predict potential residues for the incorporation of **AzoF** to generate a photoswitchable luciferase, we started from the crystallographic coordinates of the Ppy9 variant, in different conformational rearrangements as retrieved from the Protein Data Bank[23] (PDB): (*i*) open inactive state (state A, PDB ID: 1LCI[20]); (*ii*) closed active state for the generation of luciferyl-AMP (state B, PDB ID: 4G36[24]); and (*iii*) closed active state for the generation of oxyluciferin (state C, PDB ID: 4G37[24]). With the aim of obtaining the wild-type isoforms of each luciferase structure, we generated the following mutations: T214A, A215L, I232A, F295L, E354K, C81S, C216A, C258S, C391S, I108C, Y447C using our in-house developed software Eris.[25] The quality of our generated models was assessed using Gaia (http://troll.med.unc.edu/chiron/login.php,[26]), which compares their intrinsic structural properties to high-resolution crystal structures. No critical issues were found with the structural models, which were further used for further computational studies.

*Identification of residues for **AzoF** incorporation and effects on protein stability and substrate binding.* We evaluated the local conformational flexibility (repacking) of FLuc residues by using Eris, and estimated its magnitude by determining the standard deviation (σΔG) of the protein free energy values computed over all the generated protein conformations. σΔG values provide an estimate of the protein local conformational flexibility (i.e., ability to repack amino acids' side chains around each amino acid), which can be used to predict if specific residues are likely to comprise an allosteric site.[18] In order to identify amino acids in potentially allosteric luciferase regions useful for the incorporation of **AzoF**, we performed an exhaustive scan and repack along the entire primary sequence using Eris.[26] More specifically, we completed 2,000 Monte Carlo explorations of the conformational space within 10 Å from each single amino acid in the three crystallographic structures under investigation, and we computed the σΔG upon repacking (Table 2). The calculations required 45 hours to be completed on 64 computing nodes, each with 12-core, 2.93 GHz Intel processors, 12M L3 cache (Model X5670). Additionally, we computed the difference of free energy (ΔΔG) between wild-type- and **AzoF**-luciferase (both *cis*- and *trans*- isomers) in the three conformational states under investigation upon repacking at the amino acid sequence position 417. The structure of 4-azophenylphenylalanine in both *cis* and *trans* states was manually modeled in place, and subjected to 1,000 runs of minimization using the Medusa force field.[27] We observed a general destabilization of the mutants in the three different luciferase conformations, which affects more the *cis*-isomer (Supporting Figure S9) in both the D-luciferyl-AMP-bound and the apo-state. Additionally, we evaluated for each single conformational state, the binding energy of D-luciferyl-AMP, using MedusaScore,[21] an accurate force-field based scoring function which

provides an estimate of binding energy values by modelling physical interactions between proteins and small molecules (Table 3).

Acknowledgements

The work was supported by the National Science Foundation (CHE-1404836) and the Charles E. Kaufman Foundation of The Pittsburgh Foundation.

Keywords: azobenzene • photoswitching • unnatural amino acid • optochemical biology

[1] a) C. Brieke, F. Rohrbach, A. Gottschalk, G. Mayer, A. Heckel, *Angew Chem Int Ed Engl* **2012**, *51*, 8446-8476; b) A. A. Beharry, G. A. Woolley, *Chem Soc Rev* **2011**, *40*, 4422-4437; c) M. M. Russew, S. Hecht, *Adv Mater* **2010**, *22*, 3348-3360; d) R. J. Mart, R. K. Allemann, *Chem Commun* **2016**, *52*, 12262-12277; e) N. Ankenbruck, T. Courtney, Y. Naro, A. Deiters, *Angew Chem Int Ed Engl* **2018**, *57*, 2768-2798.

[2] a) F. Bonardi, G. London, N. Nouwen, B. L. Feringa, A. J. Driessen, *Angew Chem Int Ed Engl* **2010**, *49*, 7234-7238; b) X. G. Liang, T. Mochizuki, H. Asanuma, *Small* **2009**, *5*, 1761-1768; c) M. Geoffroy, D. Faure, R. Oda, D. M. Bassani, D. Baigl, *Chembiochem* **2008**, *9*, 2382-2385; d) F. Z. Zhang, A. Zarrine-Afsar, M. S. Al-Abdul-Wahid, R. S. Prosser, A. R. Davidson, G. A. Woolley, *J Am Chem Soc* **2009**, *131*, 2283-2289; e) B. Schierling, A. J. Noel, W. Wende, L. T. Hien, E. Volkov, E. Kubareva, T. Oretskaya, M. Kokkinidis, A. Rompp, B. Spengler, A. Pingoud, *Proc Natl Acad Sci USA* **2010**, *107*, 1361-1366; f) M. R. Banghart, A. Mourot, D. L. Fortin, J. Z. Yao, R. H. Kramer, D. Trauner, *Angew Chem Int Ed Engl* **2009**, *48*, 9097-9101; g) A. A. Beharry, L. Wong, V. Tropepe, G. A. Woolley, *Angew Chem Int Ed Engl* **2011**, *50*, 1325-1327; h) P. Gorostiza, E. Y. Isacoff, *Physiology* **2008**, *23*, 238-247; i) Y.-H. Tsai, S. Essig, J. R. James, K. Lang, J. W. Chin, *Nat Chem* **2015**, *7*, 554–561.

[3] a) D. Bleger, J. Schwarz, A. M. Brouwer, S. Hecht, *J Am Chem Soc* **2012**, *134*, 20597-20600; b) S. Samanta, A. A. Beharry, O. Sadovski, T. M. McCormick, A. Babalhavaeji, V. Tropepe, G. A. Woolley, *J Am Chem Soc* **2013**, *135*, 9777-9784.

[4] C. Knie, M. Utecht, F. L. Zhao, H. Kulla, S. Kovalenko, A. M. Brouwer, P. Saalfrank, S. Hecht, D. Bleger, *Chem-Eur J* **2014**, *20*, 16492-16501.

[5] M. Bose, D. Groff, J. Xie, E. Brustad, P. G. Schultz, *J Am Chem Soc* **2006**, *128*, 388-389.

[6] a) C. Hoppmann, V. K. Lacey, G. V. Louie, J. Wei, J. P. Noel, L. Wang, *Angew Chem Int Edit* **2014**, *53*, 3932-3936; b) C. Hoppmann, I. Maslennikov, S. Choe, L. Wang, *J Am Chem Soc* **2015**, *137*, 11218-11221; c) V. Klippenstein, C. Hoppmann, S. Ye, L. Wang, P. Paoletti, *Elife* **2017**, *6*, e25808.

[7] A. A. John, C. P. Ramil, Y. Tian, G. Cheng, Q. Lin, *Org Lett* **2015**, *17*, 6258-6261.

[8] L. Stricker, E. C. Fritz, M. Peterlechner, N. L. Doltsinis, B. J. Ravoo, *J Am Chem Soc* **2016**, *138*, 4547-4554.

[9] C. C. Liu, P. G. Schultz, *Annu Rev Biochem* **2010**, *79*, 413-444.

[10] J. W. Chin, *Nature* **2017**, *550*, 53-60.

[11] Y. S. Wang, X. Fang, A. L. Wallace, B. Wu, W. R. Liu, *J Am Chem Soc* **2012**, *134*, 2950-2953.

[12] W. Wan, J. M. Tharp, W. R. Liu, *Biochimica et biophysica acta* **2014**, *1844*, 1059-1070.

[13] D. D. Young, T. S. Young, M. Jahnz, I. Ahmad, G. Spraggon, P. G. Schultz, *Biochemistry-Us* **2011**, *50*, 1894-1900.

[14] H. Fliegl, A. Kohn, C. Hattig, R. Ahlrichs, *J Am Chem Soc* **2003**, *125*, 9821-9827.

[15] a) J. Y. Zhao, S. X. Lin, Y. Huang, J. Zhao, P. R. Chen, *J Am Chem Soc* **2013**, *135*, 7410-7413; b) J. Luo, R. Uprety, Y. Naro, C. Chou, D. P. Nguyen, J. W. Chin, A. Deiters, *J Am Chem Soc* **2014**, *136*, 15551-15558; c) J. Luo, J. Torres-Kolbus, J. Liu, A. Deiters, *Chembiochem* **2017**, *18*, 1442-1447.

[16] M. S. Lee, W. S. Park, Y. H. Kim, W. G. Ahn, S. H. Kwon, S. Her, *Sensors-Basel* **2012**, *12*, 15628-15637.

[17] S. Yin, F. Ding, N. V. Dokholyan, *Structure* **2007**, *15*, 1567-1576.

[18] D. Tobi, I. Bahar, *Proc Natl Acad Sci USA* **2005**, *102*, 18908-18913.

[19] A. Panjkovich, X. Daura, *BMC Bioinformatics* **2012**, *13*, 273.

[20] E. Conti, N. P. Franks, P. Brick, *Structure* **1996**, *4*, 287-298.

[21] S. Yin, L. Biedermannova, J. Vondrasek, N. V. Dokholyan, *J Chem Inf Model* **2008**, *48*, 1656-1662.

[22] R. Uprety, J. Luo, J. Liu, Y. Naro, S. Samanta, A. Deiters, *Chembiochem* **2014**, *15*, 1793-1799.

[23] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res* **2000**, *28*, 235-242.

[24] J. A. Sundlov, D. M. Fontaine, T. L. Southworth, B. R. Branchini, A. M. Gulick, *Biochemistry* **2012**, *51*, 6493-6495.

[25] S. Yin, F. Ding, N. V. Dokholyan, *Nat Methods* **2007**, *4*, 466-467.

[26] P. Kota, F. Ding, S. Ramachandran, N. V. Dokholyan, *Bioinformatics* **2011**, *27*, 2209-2215.

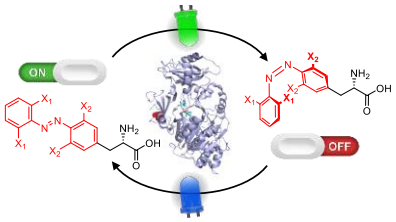
[27] N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, E. I. Shakhnovich, *Fold Des* **1998**, *3*, 577-587.

Entry for the Table of Contents (Please choose one layout)

Layout 1:

FULL PAPER

Text for Table of Contents



*Ji Luo, Subhas Samanta, Marino Convertino, Nikolay V. Dokholyan, Alexander Deiters**

Page No. – Page No.

Reversible and Tunable Photoswitching of Protein Function through Genetic Encoding of Azobenzene Amino Acids in Mammalian Cells