Article

Efficient isotope editing of proteins for site directed vibrational spectroscopy

Sebastian Peuker, Hanna Andersson, Emil Gustavsson, Kiran Sankar Maiti, Rafal Kania, Alavi Karim, Stephan Niebling, Anders Pedersen, Mate Erdelyi, and Sebastian Westenhoff

J. Am. Chem. Soc., **Just Accepted Manuscript** • DOI: 10.1021/jacs.5b12680 • Publication Date (Web): 21 Jan 2016

Efficient isotope editing of proteins for site directed vibrational spectroscopy

Sebastian Peuker[1*], Hanna Andersson[1], Emil Gustavsson[1], Kiran Sankar Maiti[1], Rafal Kania[1], Alavi Karim[1], Stephan Niebling[1], Anders Pedersen[2], Mate Erdelyi[1], Sebastian Westenhoff[1*]

*Corresponding authors: peukers@gmail.com, westenho@chem.gu.se

Affiliations

1. *Department of Chemistry and Molecular Biology, University of Gothenburg, 40530 Gothenburg, Sweden*
2. *Swedish NMR Centre at the University of Gothenburg, PO Box 465, SE-405 30 Gothenburg*

Keywords: cell-free protein synthesis, infrared spectroscopy

Abstract

Vibrational spectra contain unique information on protein structure and dynamics. However, this information is often obscured by spectral congestions and site-selective information is not available. In principle, sites of interest can be spectrally identified by isotope shifts, but site-specific isotope labeling of proteins is today only possible for favorable amino acids or with prohibitively low yields. Here, we present an efficient cell-free expression system for the site-specific incorporation of any isotope labelled amino acid into proteins. We synthesized 1.6 mg of green fluorescent protein with an isotope-labelled tyrosine from 100 ml cell-free reaction extract. We unambiguously identify spectral features of the tyrosine in the fingerprint region of the time-resolved infrared absorption spectra. Kinetic analysis confirms the existence of an intermediate state between photoexcitation and proton transfer, which lives for 3 ps. Our method lifts vibrational spectroscopy of proteins to a higher level of structural specificity.

Introduction

Proteins are in constant motion and their diverse functions are ultimately governed by this dynamic character[1,2]. Methods to study protein structural dynamics are in demand. Vibrational spectroscopy is a very powerful mean to probe the conformational change of chemical entities[3]. A number of time-resolved and multidimensional infrared spectroscopic techniques have been developed and they can be used to directly monitor vibrational coupling of chemical groups in a time-resolved manner[4–8]. Unfortunately, vibrational spectra of proteins are crowded. The assignment of spectral features to specific sites is at best tedious, but most often impossible. This seriously restricts the application of vibrational spectroscopy to study proteins dynamics.

A strategy to overcome the crowdedness of the spectra is to label specific sites of interest in the protein. It is often possible to label cofactors or ligands[8,9]. Moreover infrared-active reporter groups can be attached to the proteins after protein expression[10], by introducing unnatural amino acids with infrared-active functional groups during proteins synthesis[11–16] or by introducing isotope-edited amino acids[17–22]. Using infrared-active reporter groups has the advantage that they are typically easily identified in the spectra. These approaches have yielded remarkable insight into the rearrangements of secondary structure[13] and the solubilization of proteins[10], for example. The latter method, labelling with isotopes, ensures that the chemical properties of the proteins remain unaltered. This gives the method a distinct advantage over labelling with reporter groups

as it allows monitoring active sites that directly participate in the enzymatic activity of the protein.

The increased mass of the isotopes causes chemical bonds to oscillate at lower frequencies, which is observed as red-shifted peaks in the vibrational spectra[23]. Infrared spectroscopy of isotope labelled proteins has yielded unique insights into the folding dynamics of model peptides[19,20] and amyloid fibres[17], it was used to solve the three-dimensional dimer arrangement of a membrane-bound alpha helical protein[24], and it has helped to elucidate the photocycle of membrane proton pumps[25].

Today, the wider-spread application of site-specific isotope-labelling is restricted by protein synthesis. Most of the studies mentioned were proteins, which were produced by chemical peptide synthesis[17,19,20,24]. This is a versatile method for site-specific insertion of isotope labelled amino acids into proteins; however, it is also restricted to short peptide chains. Nevertheless, some proteins can be labeled site-selectively by connecting a chemically produced peptide chain with a truncated version of the protein[21]. In *in vivo* expression systems, site-specific isotope labelling is possible when an amino acid occurs only a single time in a polypeptide chain and an expression host auxotrophic for that amino acid is available[25,26].

With *in vitro* expression systems[26], tyrosines can be labelled due to the existence of an orthogonal tRNA/amino-acyl transferase pair in *E. coli*[25,27]. The most general *in vitro* method is to introduce

unnatural amino acids in response to an amber stop codon with chemically aminoacylated suppressor tRNA[28,29]. This strategy makes possible the incorporation of any amino acid, but it has not yet been used for infrared spectroscopic studies[30]. This is presumably because of the moderate protein production yields. Fundamentally limiting the yield is the release factor 1 (RF1) which competes with the suppressor tRNA for the amber stop codon. This problem has recently been addressed in a modified *E. coli* strain, developed by the Otting group[31]. The reaction mixture from this strain can be reliably depleted of RF1 with affinity chromatography.

Here, we combine the two approaches for *in vitro* expression. We explore two strategies for the synthesis of chemically aminoacylated suppressor tRNA and outline routes for the incorporation of all native amino acids. Finally, we develop a simplified procedure for $^{18}O^{13}C$-labelled amino acids from ^{13}C precursors, which is of particular practical importance for infrared spectroscopy. Combined, we were able to produce milligram quantities of site-specifically isotope labelled proteins. Using this strategy, we assigned vibrations of a key tyrosine in the active site of green fluorescent protein (GFP) to peaks in the infrared fingerprint spectral region.

Results

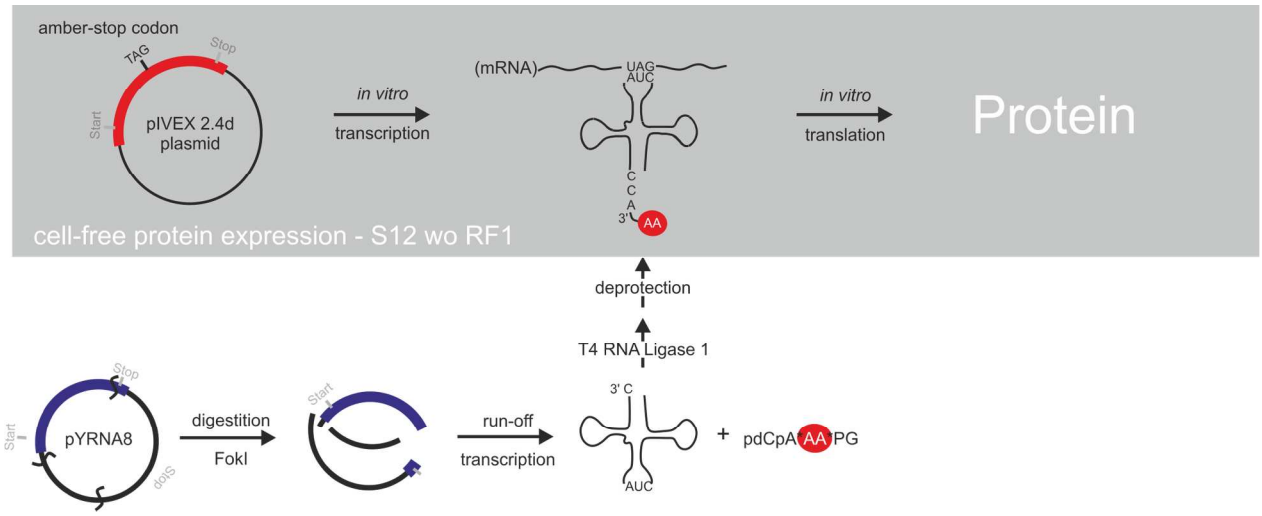


Figure 1: Site-directed introduction of isotope labelled amino acids into proteins.

Scheme of the experimental principle applied to synthesize isotope labelled proteins. Details are given in the main text. (AA amino acid; RF1 release factor 1; PG protection group, pdCpA 5'-O-phosphoryl-2'-deoxycytidylyl-(3'--> 5')adenosine).

In the following sections the three major components of the method are described: The cell-free expression system, the synthesis of the chemically aminoacylated suppressor $tRNA^{AA}_{CUA}$, and the synthesis of isotope labelled and appropriately protected amino acids. The final section describes how isotope labelled GFP was used to assign chromophore vibrations in UV pump - infrared probe spectra. Figure 1 is an overview of the synthesis.

Cell free expression system

We used a cell-free protein expression system in batch mode[32,33]. In order to maximize the incorporation efficiency of the modified amino acids, the S12 extract was produced from a

genomically modified *E. coli* strain BL21 Star (DE3):RF1-CBD3 developed by Otting and coworkers[31]. This strain allows the removal of release factor 1 protein (RF1) by means of affinity chromatography. With this, the termination of the translation is avoided when the stop codon is reached[34]. The complete removal of RF1 from the S12 was confirmed by Western blotting (Fig. 2a), and the expression yield of the modified extract was 75 µg protein per mL reaction mixture.

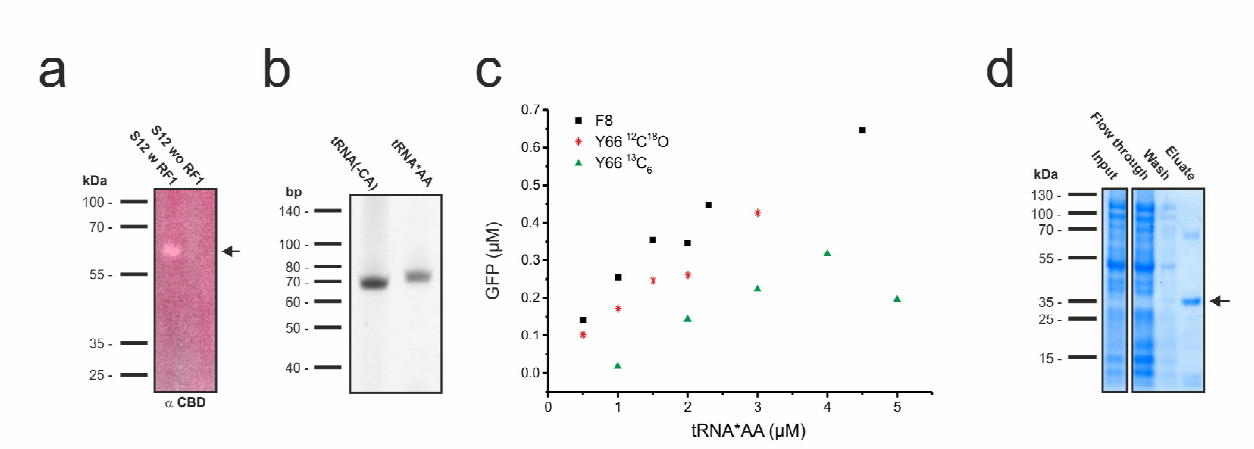


Figure 2: Characterization of the site-directed labelling procedure.

(a) Removal of RF1 verified by Western blotting. RF1 was detected by an anti-chitin binding domain antibody (α CBD). The apparent molecular weight of RF1-$(CBD)_3$ is 69.9 kDa. (b) Products of the run-off transcription and ligation verified by a 10 % Novex® TBE-Urea Gel stained with SybrSafe. The length of tRNA(-CA) and tRNA*AA is 74 and 76 bases, respectively. (c) Test of suppression efficacy of $tRNA^{Phe}_{CUA}$, $tRNA^{Tyr^{12}C^{18}O}_{CUA}$, and $tRNA^{Tyr^{13}C_6}_{CUA}$ by cell-free expression of $GFP^{Phe8TAG}$ and $GFP^{Tyr66TAG}$ respectively. Expression of GFP was quantified by means of fluorescence. (d) SDS-PAGE of the purification of GFP using $tRNA^{Tyr^{13}C_6}_{CUA}$ by Ni-affinity chromatography. The apparent molecular weight of GFP is 29.8 kDa.

The synthesis of the chemically aminoacylated suppressor $tRNA^{AA}_{CUA}$

Chemically aminoacylated tRNA(-CA)$^{AA}_{CUA}$ was synthesized by enzymatically ligating 3' truncated tRNA(-CA)$_{CUA}$ with aminoacyl-pdCpA (synthesis see below) *in vitro* using T4 RNA Ligase 1 (Fig. 2b) [35,36]. The precursor, 3' truncated tRNA(-CA)$_{CUA}$, was prepared by run-off transcription from FokI digested template DNA (pYRNA8) (Fig. 2b, left lane) and complete digestion was observed using 0.25 U FokI per µg pYRNA8 in 15 min. As judged by a 10% TBE Urea gel, the ligation of aminoacyl-pdCpA to tRNA(-CA)$_{CUA}$ was quantitative (Fig. 2b, right lane).

We tested the efficiency of this system by the expression of green fluorescent protein (GFP). The level of fluorescence is directly proportional to the level of functional full length protein. We incorporated a phenylalanine at position 8 and a tyrosine at position 66 by expression of GFP from GFPcyc3^{Phe8TAG} and GFPcyc3^{Tyr66TAG} templates, respectively. Fig. 2c shows that the level of expressed GFP scales with the concentration of the respective tRNA$^{AA}_{CUA}$. The level of expressed GFP scales equally for the F8 and Y66 labelled with $\mathrm{Tyr}^{12}\mathrm{C}^{18}\mathrm{O}$, but the incorporation was lower for Y66 labelled with $\mathrm{Tyr}^{13}\mathrm{C}_6$. We attribute this difference to non-optimal deprotection of the particular batch of tRNA$^{\mathrm{Tyr}^{13}\mathrm{C}_6}_{CUA}$. Up to 30% of the tRNA$^{AA}_{CUA}$ was translated into proteins.

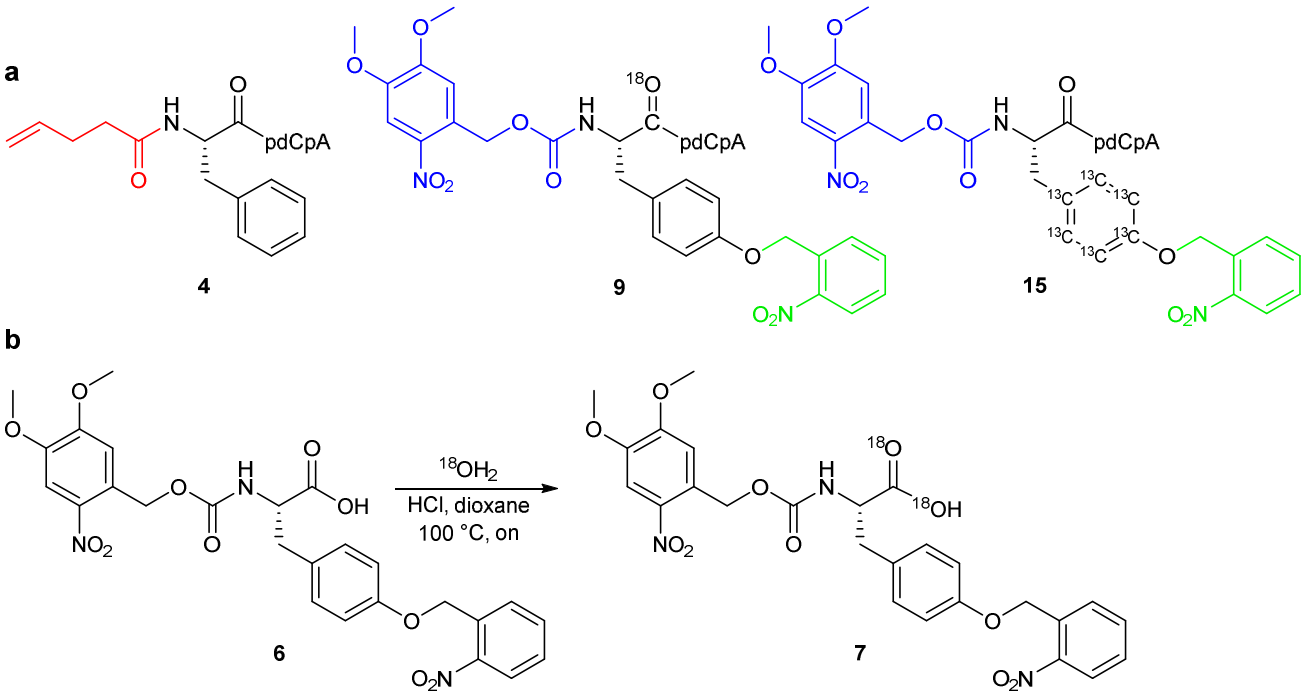
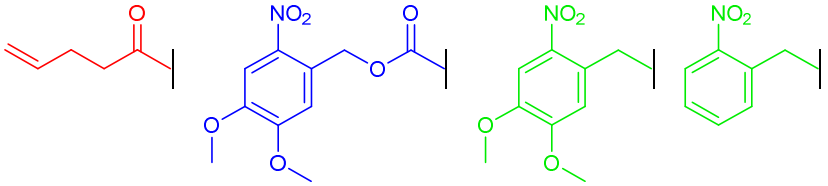


Figure 3: Chemical intermediates in the preparation of aminoacyl-tRNAs

(a) Aminoacylated dinucleotide derivatives 4, 9 and 15 were synthesized (Supplementary Scheme 1–3) and used in the preparation of tRNAs for site-specific incorporation into GFP. The iodine labile 4-PO protecting group in red was used for phenylalanine derivative 4, while the photolabile NVOC and NB protecting groups in blue and green were used for tyrosine derivatives 9 and 15, respectively. (b) A simplified procedure for ^{18}O labelling of carboxylic acids, such as that of amino acid derivative 6 above, has been developed utilizing ^{18}O-enriched water, 1,4-dioxane, and commercially available 4 M HCl in 1,4-dioxane as acid source (Supplementary Scheme 2).

Table 1. Overview of most common amino acid protecting group strategies applicable for the preparation of chemically aminoacylated tRNAs.

Protecting group strategy α-Amine	Side-chain	Examples of natural amino acids
4-PO	-	Non-polar; Gly, Ala, Val, Leu, Ile, Met, Pro, Phe
		Polar basic; Arg
4-PO	NV/NB	Polar; Ser, Thr, Asn, Gln
		Polar acidic; Asp, Glu
4-PO	NVOC	Polar basic; Lys
NVOC	-	Non-polar; Gly, Ala, Val, Leu, Ile, Met, Pro, Phe, Trp
		Polar basic; Arg
NVOC	NV/NB	Polar; Ser, Thr, Cys, Asn, Gln, Tyr
		Polar acidic; Asp, Glu
		Polar basic; His
NVOC	NVOC	Polar basic; Lys



4-PO NVOC NV NB

Synthesis of aminoacyl-pdCpAs

Aminoacylated dinucleotide derivatives (aminoacyl-pdCpAs) were synthesized by aminoacylation of pdCpA with the desired amino acid (Fig. 1)[35]. To prevent side-reactions and to increase the stability of the amino acid-nucleotide ester linkage, the α-amino and side-chain functional groups were protected. We explored two protecting group strategies[37]. The iodine labile 4-pentenoyl (4-PO) group was used as an α-amino protecting group in the preparation of phenylalanyl-pdCpA **4** (Fig. 3 and Supplementary Scheme 1). It is ideal for amino acids without side-chain functionality (Table 1)[38], but can also be applied in orthogonal protecting strategies, e.g. together with photolabile side-chain protecting groups. However, it is unsuitable for amino acids having oxidizable side-chains. Tyrosine falls into this category and therefore the photolabile 6-nitroveratryloxycarbonyl (NVOC) and 2-nitrobenzyl (NB) groups were used as α-amino and side-chain protecting groups in the synthesis of tyrosinyl-pdCpA **9** and **15**, respectively (Fig. 3, Supplementary Schemes 2 and 3). 6-Nitroveratryl (NV) is a possible substitute for NB and Table 1 summarizes suitable protection strategies for all natural amino acids. Compound **4** was produced with an overall yield of 21% starting from L-phenylalanine, and **9** and **15** were produced with overall yields of 4% and 6% starting from the commercially available L-tyrosine and L-[*phenyl*-$^{13}C_6$]tyrosine, respectively.

A simple method for ^{18}O labelling of amino acids

Labelling the backbone carbonyl function as $^{13}C=^{18}O$ is particularly important for probing the amide I region of infrared spectra[39]. The amide I vibrations gives rise to strong absorption peaks at approximately 1640 cm^{-1} and the introduction of $^{13}C=^{18}O$ labels results in a ~60 cm^{-1} red shift which enables spectral isolation[39]. Unfortunately, $^{13}C=^{18}O$-labelled amino acids are generally not commercially available. Synthesis may be achieved from ^{13}C-labelled precursors by refluxing the compound in a water-free dioxane/$H_2{}^{18}O$ mixture at acidic pH[40]. We applied a simplified procedure for ^{18}O labelling of the backbone carbonyl in the synthesis of compound **9**. The exchange was performed under acidic conditions, adding 4 M HCl/1,4-dioxane to the amino acid derivative that was premixed with ^{18}O-enriched water in 1,4-dioxane under nitrogen atmosphere (Fig. 3 and Supplementary Scheme 2). The reaction mixture was heated to 100°C using an oil bath in a sealed microwave vial overnight. LC-MS confirmed quantitative ^{18}O labelling within 20 h (See the Supplementary Information for analytical data). In contrast to earlier reports, using this method neither *in situ* generation of HCl-gas[18,40], nor repetition of the reaction is necessary for full conversion.

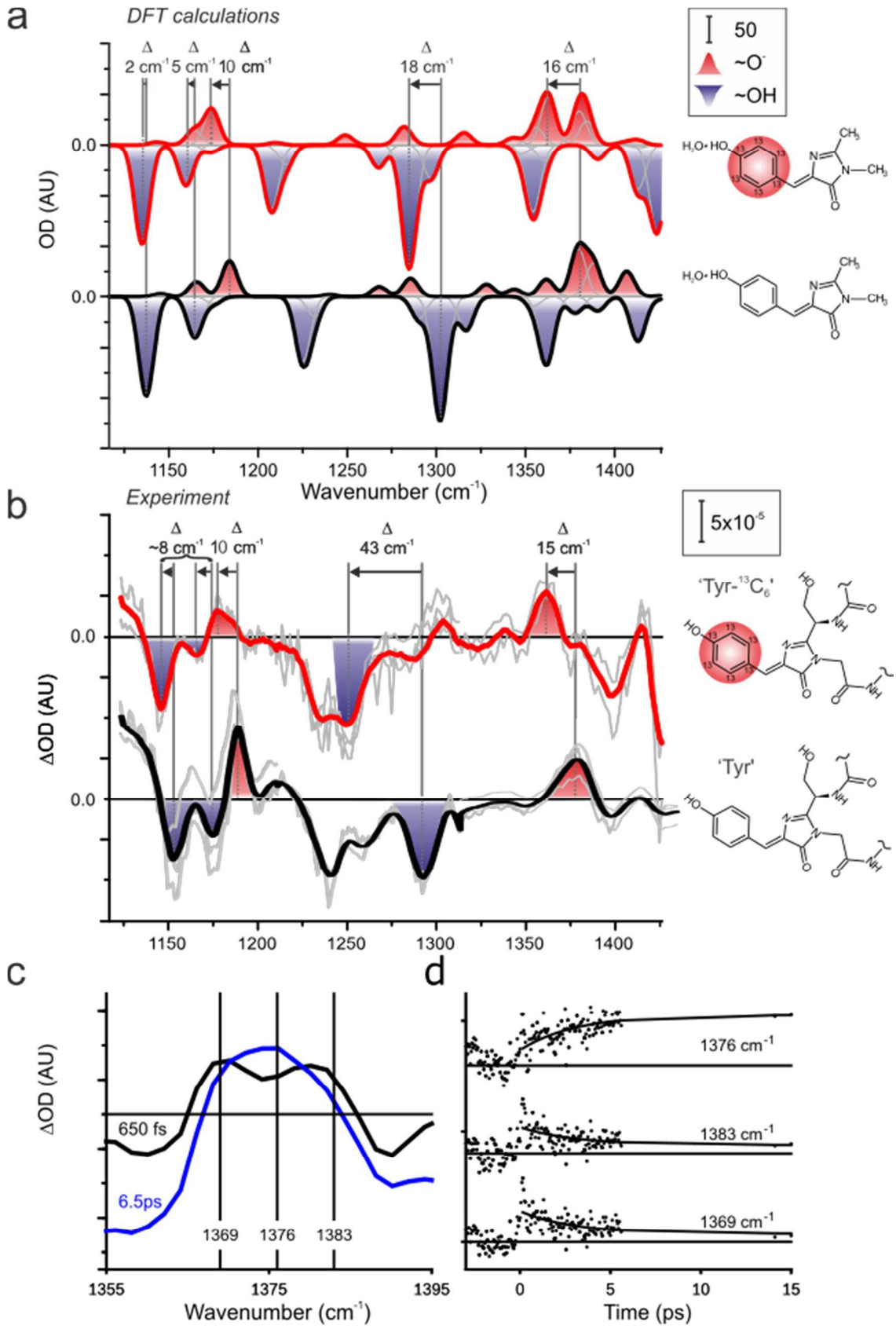

a
DFT calculations
Δ 2 cm-1
Δ 5 cm-1
Δ 10 cm-1
Δ 18 cm-1
Δ 16 cm-1
50
~O-
~OH
OD (AU)
0.0
0.0
1150
1200
1250
1300
1350
1400
Wavenumber (cm-1)
b
Experiment
Δ ~8 cm-1
Δ 10 cm-1
Δ 43 cm-1
Δ 15 cm-1
5x10-5
'Tyr-13C6'
'Tyr'
ΔOD (AU)
0.0
0.0
1150
1200
1250
1300
1350
1400
c
d
ΔOD (AU)
650 fs
6.5ps
1369
1376
1383
1355
1375
1395
Wavenumber (cm-1)
1376 cm-1
1383 cm-1
1369 cm-1
0
5
10
15
Time (ps)

Figure 4: UV pump infrared probe spectra of $GFP^{Tyr66\text{-}^{13}C6}$.

(a) DFT calculations of infrared absorption spectra of model compounds representing the chromophore of GFP as illustrated. The isotope labelled part of the molecule is highlighted. The single modes underlying the spectra shown in grey are plotted as Gaussian functions with a full width at half maximum of 10 cm^{-1}. (b) UV pump - infrared probe spectra of *wild type* (black) and $GFP^{Tyr66\text{-}^{13}C6}$ (red). Three spectral windows were recorded with center photon energies of 1370 cm^{-1}, 1259 cm^{-1}, and 1165 cm^{-1}. The averaged and aligned spectra drawn in bold lines and the underlying data sets are drawn in grey. (c) Difference spectra at two delay times after photoexcitation. The spectrum denoted as '650 fs' is the average of all spectra from 500 fs to 800 fs and the spectrum denoted as '6.5 ps' is integrated from 3 ps to 10 ps. (d) Kinetics at the wavenumbers as indicated. The kinetics are integrated over +/- 3 cm^{-1}.

Isotope edited infrared spectroscopy

We used our isotope labelling method to unambiguously assign the fingerprint vibrations in difference infrared spectra of tyrosine 66 in GFP. Tyr66 is part of the chromophore and deprotonates upon photoexcitation within a few picoseconds. The proton is transferred past a water molecule and a serine side-chain to glutamate 222[8,41,42]. We target the 1100 cm^{-1} to 1400 cm^{-1} fingerprint region of the infrared spectrum. While this spectral range contains a wealth of information on chemical structure and dynamics, the features are also notoriously difficult to assign[3].

To guide the protein synthesis, we calculated the spectral shifts of different isotope labelling patterns on tyrosine using density functional theory (DFT) calculations. While DFT calculations are usually not very reliable in predicting the peak position of fingerprint vibrations, the computation of isotope shifts should be much more robust. Replacing all carbons in the phenyl ring by ^{13}C ($Tyr^{13}C_6$) produced the largest predicted shift (Fig. 4a). We therefore synthesized $GFPcyc3^{Tyr66Tyr13C6}$ in a preparative scale and purified it by affinity chromatography (Fig. 2d). From 100 ml cell-free reaction, 1.6 mg purified and concentrated $GFPcyc3^{Tyr66Tyr13C6}$ was obtained, which relates to about 13% of the $tRNA^{Tyr13C6}_{CUA}$ incorporated into the purified protein.

This sample was subjected to UV pump - infrared probe difference spectrometry using our home built setup (see methods for details). GFP was photo-excited using laser pulses centered at 400

nm. Infrared difference spectra from 1426 cm^{-1} to 1117 cm^{-1} are shown in Fig. 4b for the labelled GFP (red line) and for a wt GFP sample (black line). Inspection of these spectra reveals that the position of the positive peaks at 1375 cm^{-1} and 1188 cm^{-1}, and the negative peaks at 1292 cm^{-1}, 1174 cm^{-1}, and 1152 cm^{-1} are red-shifted in the isotope labelled sample. We conclude that these peaks must be due to side-chain vibrations of Tyr66. The assignment also allows us to conclude that the negative peak at 1240 cm^{-1} does not originate from the Tyr66 residue. This peak is likely caused by a different amino acid participating in the photoreactions of GFP.

We also performed DFT computations of the truncated chromophore. DFT calculations are often used to aid assignment of infrared spectra, however, the computed absolute spectral positions usually carry a significant uncertainty. This was also the case for our calculations and we adjusted the wavenumber scales by a constant factor to match the wild-type spectrum. When it comes to predicting isotope-induced spectral shifts, DFT should be much more robust, because the influence of the increased mass of the isotopes on the reduced mass of the vibrations can be estimated with precision. Indeed, the predicted isotope shifts agree with the observed shifts (Fig 4a and b).

With the tyrosine peaks assigned based on isotope-induced spectral shifts, we now inspect the peaks at 1375 cm^{-1} closer. The peak converts from a double peak at early times after photoexcitation (650 fs, Fig. 4c) into a single peak (6.5 ps, Fig 4c). The conversion has a

characteristic time constant of 3 +/- 1 ps (Fig 4d) and the final spectral shape is stable for > 1 ns (data not shown). We interpret this conversion to be a sign of deprotonation of the chromophore[43]. Directly after photoexcitation, the chromophore relaxes in its electronically excited, protonated state and after 3 ps the proton departs from the phenol-OH group, leaving the tyrosinate behind[8].

Discussion

We have demonstrated that the active site of a protein can be site-specifically isotope-labelled in sufficient yields to perform time-resolved infrared spectroscopy. The production yield in our cell-free system (75 µg protein per mL reaction mixture) was limited by the concentration of aminoacylated tRNA (Fig S1 and 2c). This compound has to be chemically synthesized for each amino acid and is also the most expensive reactant. The overall material cost of a labelled protein was approximately 6000 Dollar/mg protein. The T4 RNA Ligase 1 (New England Biolabs) and the pdCpa (ATDBio), which are needed in the synthesis of the aminoacylated tRNA, are the main cost-drivers. It would therefore be interesting to explore alternative methods for more efficient production of the aminoacylated tRNA.

With the label in place, we were readily able to assign peaks in the infrared difference absorption spectra to specific amino acid vibrations, while conserving the functionality of the wild-type

protein. Previously these assignments were not possible, despite the abundance of knowledge about the GFP photocycle[8,41,42], spectroscopic experience[3], and highly developed computational methods. Specific isotope labelling, as demonstrated here, enabled the spectral interpretation. If the three-dimensional structure of a protein is known, such definite assignments add a spatial dimension to infrared spectroscopy.

The fingerprint region of infrared spectra yields highly specific information. This is illustrated by the spectral evolution of the peak at 1375 cm^{-1} (Fig 4c and 4d). From this peak we find that the deprotonation of Tyr66 occurs after a few picoseconds of dwell time, which confirms earlier proposals[8,22]. The deprotonation could be limited by additional structural rearrangements of the protein or the chromophore. However, the dwell time could also reflect the tunneling barrier of the protons. The latter suggestion is supported by the lengthening of the proton relay time when the protons are exchanged for deuterium[8,41]. It is conceivable that our method can be used to deduct specific information about transient states in other photoactive proteins, such as photoreceptor proteins or photosynthetic proteins.

Here, we used isotope-induced shifts to assign absorption peaks in *difference* infrared spectra. However, isotope shifts can also be used to isolate peaks in *absolute* spectra. This has, for example, been exploited to study the aggregation behavior of amyloids[17] and unfolding of peptides[19] with multidimensional infrared spectroscopy. With the method presented in this paper,

larger proteins could in principle be studied in a similar way. Whether the desired peaks can be isolated requires further investigation, because spectral crowding becomes increasingly restricting with protein size. Also, it is often not easy to predict how the excitonic coupling of amide groups would changed by isotope labelling of one specific site. Nevertheless, experiments that require site-specific isotope labelling of larger proteins have been suggested. One intriguing idea is to determine the ion occupancy[44] in the selection filter of the ion channel KcsA with two-dimensional spectroscopy[35].. This proposed experiment has not yet been performed because site-selective isotope labelling of the protein was not possible. The method presented here can open this bottleneck.

Methods

Wild-type GFP

Wild type GFP was produced by cloning GFPcyc3[45] into the vector pIVEX2.4d[33] (Roche Applied Science) and heterologous expression in *E. coli* BL21 Star (DE3) (Invitrogen). Purification was by Ni^{2+}-affinity chromatography (Ni-NTA Agarose, Quiagen).

Extract preparation, and RF1 depletion

S12 extract was prepared from BL21 (DE3):RFI-CBD_3 [31] as described by Pedersen[33]. To remove RF1, the S12 extract was passed over chitin resin (New England Biolabs) directly after the incubation and removal was confirmed by Western blotting (Fig. 2a). After addition of 1 mM

DTT the S12 extract was dialyzed twice against extract buffer supplemented with β-mercaptoethanol (1 ml/l), flash frozen, and stored at -80 °C.

Cell-free protein expression

Cell-free protein expression was performed in batch mode as described[33]. For every batch of S12 extract the optimal Mg^{2+} was determined by maximizing GFP fluorescence at 520 nm. To preprare tRNA(-CA)$^{\mathrm{AA*NVOC/NB}}_{\mathrm{CUA}}$, aminoacylated pdCpA (**1–15**) was synthesized as described in the extended methods second (SI) and coupled to tRNA(-CA)$_{\mathrm{CUA}}$, prepared by run-off transcription[46,47], by using T4 RNA Ligase 1[48] (New England Biolabs). Deprotection of ~25 µM tRNA(-CA)$^{\mathrm{AA*NVOC/NB}}_{\mathrm{CUA}}$ was done at acidic pH, 2 ºC at a volume of 400 µl in a 1.5 ml in a Protein LoBind tube (Eppendorf, serves as an additional HP filter >320 nm). The sample was illuminated for 10 min using a 500 W Xe lamp equipped with a water filter and a 320 nm HP filter (Schott). After exposure, samples were immediately flash frozen, stored at -80 ºC, and used without further purification. Deprotection of ~25 µM tRNA(-CA)$^{\mathrm{AA*pentenyl}}_{\mathrm{CUA}}$ was achieved by treating the suppressor tRNA with 5 mM iodine (10% (v/v) tetrahydrofuran) for 45 min at 25 ºC. After deprotection 0.3 M NaOAc (pH 5.2) was added, tRNA(-CA)$^{\mathrm{AA}}_{\mathrm{CUA}}$ precipitated by isopropanol, resuspended in 1 mM NaOAc (pH 4.5) and stored at -80 ºC.

Isotope-labelling

Details of the synthesis are provided in the SI

DFT calculations

Density functional theory (DFT) calculations were performed with a system comprising the GFP chromophore and a coordinating water molecule in the gas phase with the software package TURBOMOLE 6.6. More details are found in the SI.

Pump-probe spectroscopy

For time-resolved pump probe infrared spectroscopy the sample was excited by the second harmonic (400 nm, diameter of ~120 µm, 250 nJ/pulse) of a mode-locked femtosecond Ti:sapphire amplifier at 5 kHz repetition rate (Spitfire Ace, Spectra-Physics) and probed by infrared pulse generated by a Topas Twin (Light conversion) and non-collinear difference frequency generation (approximately 3 nJ/pulse and a diameter of ~180 µm at the sample position). In a first set of experiments (Fig 4b), the excitation beam was passed through a 10 cm long glass rod (15 mm aperture, 100 mm L High NA, Hexagonal Light Pipe) to stretch the pulse to a duration of approximately 2 ps. A second set of experiments (Fig 4c and 4d) was performed with femtosecond pulses with a duration of approximately 100fs. Reference probe pulses, which were note guided through the sample, were used to correct for shot-to-shot laser fluctuations. A iHR 320 spectrometer (Horiba) equipped with a grating (5 grooves/mm, 8000 nm blaze wavelength, Horiba) and two 64 element mercury cadmium telluride detector arrays (MCT-8-

2x64, Infrared Associates) with associated digitizing software (Infrared System Development Corporation, FPAS 0144) were used for detection. To limit photo degradation the sample was moved perpendicular to the probe beam direction over a total area of ~1.1 cm^2 using two linear stages in xy-configuration (4218S-04-01R0, Newmark).

Acknowledgement

We thank Jean-Michel Betton and Cordula Jany for providing the GFP cycle 3 vector, Sydney Hecht for the pYRNA8 vector, and Gabriela Kovácsová for providing the blot in Fig. 2d. Ashley Hughes and Matthijs Panman are acknowledged for critically reading the manuscript. We thank Linnea Isaksson for help with the additional assay (SI Fig 1).

Supporting Information. Extended Methods, Reaction Schemes, ^{1}H and ^{13}C NMR Spectra, CD spectra, and HPLC (ESI^{+}/ESI^{-}/UV). This material is available free of charge via the Internet at http://pubs.acs.org.

(1) Henzler-Wildman, K.; Kern, D. *Nature* **2007**, *450* (7172), 964.

(2) Vinson, V. J. *Science (80-.).* **2009**, *324* (5924), 197.

(3) Barth, A. *Biochim. Biophys. Acta* **2007**, *1767* (9), 1073.

(4) Hamm, P.; Lim, M. H.; Hochstrasser, R. M. *J. Phys. Chem. B* **1998**, *102* (31), 6123.

(5) Kolano, C.; Helbing, J.; Kozinski, M.; Sander, W.; Hamm, P. *Nature* **2006**, *444* (7118), 469.

(6) Zheng, J.; Kwak, K.; Asbury, J.; Chen, X.; Piletic, I. R.; Fayer, M. D. *Science (80-.).* **2005**, *309* (5739), 1338.

(7) Zhao, W.; Wright, J. C. *Phys. Rev. Lett.* **1999**, *83* (10), 1950.

(8) Fang, C.; Frontiera, R. R.; Tran, R.; Mathies, R. A. *Nature* **2009**, *462* (7270), 200.

(9) Deng, H.; Callender, R. In *Infrared and Raman Spectroscopy of Biological Materials, eds Gremlich and Yan*; 2001; pp 477–515.

(10) King, J. T.; Kubarych, K. J. *J. Am. Chem. Soc.* **2012**, *134* (45), 18705.

(11) Waegele, M. M.; Culik, R. M.; Gai, F. *J Phys Chem Lett* **2011**, *2*, 2598.

(12) Liu, C. C.; Schultz, P. G. *Annu. Rev. Biochem. Vol 79* **2010**, *79*, 413.

(13) Ye, S. X.; Zaitseva, E.; Caltabiano, G.; Schertler, G. F. X.; Sakmar, T. P.; Deupi, X.; Vogel, R. *Nature* **2010**, *464* (7293), 1386.

(14) Bazewicz, C. G.; Lipkin, J. S.; Smith, E. E.; Liskov, M. T.; Brewer, S. H. *J. Phys. Chem. B* **2012**, *116* (35), 10824.

(15) Schultz, K. C.; Supekova, L.; Ryu, Y.; Xie, J.; Perera, R.; Schultz, P. G. *J. Am. Chem. Soc.* **2006**, *128* (43), 13984.

(16) Getahun, Z.; Huang, C.-Y.; Wang, T.; De León, B.; DeGrado, W. F.; Gai, F. *J. Am. Chem. Soc.* **2003**, *125* (2), 405.

(17) Middleton, C. T.; Marek, P.; Cao, P.; Chiu, C.; Singh, S.; Woys, A. M.; de Pablo, J. J.; Raleigh, D. P.; Zanni, M. T. *Nat Chem* **2012**, *4* (5), 355.

(18) Middleton, C. T.; Woys, A. M.; Mukherjee, S. S.; Zanni, M. T. *Methods* **2010**, *52* (1), 12.

(19) Ihalainen, J. A.; Paoli, B.; Muff, S.; Backus, E. H. G.; Bredenbeck, J.; Woolley, G. A.; Caflisch, A.; Hamm, P. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (28), 9588.

(20) Smith, A. W.; Lessing, J.; Ganim, Z.; Peng, C. S.; Tokmakoff, A.; Roy, S.; Jansen, T. L. C.; Knoester, J. *J. Phys. Chem. B* **2010**, *114* (34), 10913.

(21) Schwans, J. P.; Sunden, F.; Gonzalez, A.; Tsai, Y.; Herschlag, D. *Biochemistry* **2013**, *52* (44), 7840.

(22) Stoner-Ma, D.; Jaye, A. A.; Matousek, P.; Towrie, M.; Meech, S. R.; Tonge, P. J. *J. Am. Chem. Soc.* **2005**, *127* (9), 2864.

(23) Decatur, S. M. *Acc. Chem. Res.* **2006**, *39* (3), 169.

(24) Remorino, A.; Korendovych, I. V; Wu, Y. B.; DeGrado, W. F.; Hochstrasser, R. M. *Science (80-.).* **2011**, *332* (6034), 1206.

(25) Sonar, S.; Lee, C. P.; Coleman, M.; Patel, N.; Liu, X. M.; Marti, T.; Khorana, H. G.; Rajbhandary, U. L.; Rothschild, K. J. *Nat. Struct. Biol.* **1994**, *1* (8), 512.

(26) Bernhard, F.; Tozawa, Y. *Curr Opin Struct Biol* **2013**, *23* (3), 374.

(27) Yabuki, T.; Kigawa, T.; Dohmae, N.; Takio, K.; Terada, T.; Ito, Y.; Laue, E.; Cooper, J.; Kainosho, M.; Yokoyama, S. *J. Biomol. NMR* **1998**, *11* (3), 295.

(28) Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science (80-.).* **1989**, *244* (4901), 182.

(29) Ellman, J. A.; Volkman, B. F.; Mendel, D.; Schultz, P. G.; Wemmer, D. E. *J Am Chem Soc* **1992**, *114* (20), 7959.

(30) Sobhanifar, S.; Reckel, S.; Junge, F.; Schwarz, D.; Kai, L.; Karbyshev, M.; Lohr, F.; Bernhard, F.; Dotsch, V. *J. Biomol. NMR* **2010**, *46* (1), 33.

(31) Loscha, K. V; Herlt, A. J.; Qi, R. H.; Huber, T.; Ozawa, K.; Otting, G. *Angew. Chemie-International Ed.* **2012**, *51* (9), 2243.

(32) Isaksson, L.; Enberg, J.; Neutze, R.; Göran Karlsson, B.; Pedersen, A. *Protein Expr. Purif.* **2012**, *82* (1), 218.

(33) Pedersen, A.; Hellberg, K.; Enberg, J.; Karlsson, B. G. *N Biotechnol* **2011**, *28* (3), 218.

(34) Park, H. S.; Hohn, M. J.; Umehara, T.; Guo, L. T.; Osborne, E. M.; Benner, J.; Noren, C. J.; Rinehart, J.; Soll, D. *Science (80-.).* **2011**, *333* (6046), 1151.

(35) Robertson, S. A.; Ellman, J. A.; Schultz, P. G. *J Am Chem Soc* **1991**, *113* (7), 2722.

(36) England, T. E.; Uhlenbeck, O. C. *Nature* **1978**, *275* (5680), 560.

(37) Hecht, S. M. *Protein Engineering: Nucleic Acids and Molecular Biology*; Springer, 2009; Vol. 22.

(38) Lodder, M.; Golovine, S.; Laikhter, A. L.; Karginov, V. A.; Hecht, S. M. *J Org Chem* **1998**, *63* (3), 794.

(39) Ihalainen, J. A.; Bredenbeck, J.; Pfister, R.; Helbing, J.; Chi, L.; van Stokkum, I. H. M.; Woolley, G. A.; Hamm, P. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (13), 5383.

(40) Marecek, J.; Song, B.; Brewer, S.; Belyea, J.; Dyer, R. B.; Raleigh, D. P. *Org Lett* **2007**, *9* (24), 4935.

(41) van Thor, J. J. *Chem. Soc. Rev.* **2009**, *38* (10), 2935.

(42) Di Donato, M.; van Wilderen, L. J. G. W.; Van Stokkum, I. H. M.; Stuart, T. C.; Kennis, J. T. M.; Hellingwerf, K. J.; van Grondelle, R.; Groot, M. L. *Phys. Chem. Chem. Phys.* **2011**, *13* (36), 16295.

(43) Chattoraj, M.; King, B. A.; Bublitz, G. U.; Boxer, S. G. *Proc. Natl. Acad. Sci.* **1996**, *93* (16), 8362.

(44) Köpfer, D. A.; Song, C.; Gruene, T.; Sheldrick, G. M.; Zachariae, U.; de Groot, B. L.; Kopfer, D. A. *Science (80-.).* **2014**, *346* (6207), 352.

(45) Crameri, A.; Whitehorn, E. A.; Tate, E.; Stemmer, W. P. *Nat Biotechnol* **1996**, *14* (3), 315.

(46) Gao, R.; Zhang, Y.; Choudhury, A. K.; Dedkova, L. M.; Hecht, S. M. *J Am Chem Soc* **2005**, *127* (10), 3321.

(47) Robertson, S. A.; Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Nucleic Acids Res.* **1989**, *17* (23), 9649.

(48) England, T. E.; Bruce, A. G.; Uhlenbeck, O. C. *Methods Enzym.* **1980**, *65* (1), 65.

Insert Table of Contents artwork here