ChemPubSoc Europe

DOI: 10.1002/chem.201503908

CHEMISTRY A European Journal Full Paper

Structural Biology

Azidoethoxyphenylalanine as a Vibrational Reporter and Click Chemistry Partner in Proteins

Elise M. Tookmanian, Christine M. Phillips-Piro,* Edward E. Fenlon,* and Scott H. Brewer*[a]

Abstract: An unnatural amino acid, 4-(2-azidoethoxy)-L-phenylalanine (AePhe, **1**), was designed and synthesized in three steps from known compounds in 54% overall yield. The sensitivity of the IR absorption of the azide of AePhe was established by comparison of the frequency of the azide asymmetric stretch vibration in water and dimethyl sulfoxide. AePhe was successfully incorporated into superfolder green fluorescent protein (sfGFP) at the 133 and 149 sites by using the amber codon suppression method. The IR spectra of these sfGFP constructs indicated that the azide group at the 149 site was not fully solvated despite the location in sfGFP and the three-atom linker between the azido group and the aromatic ring of AePhe. An X-ray crystal structure of sfGFP-149-AePhe was solved at 1.45 Å resolution and provides an explanation for the IR data as the flexible linker adopts a conformation which partially buries the azide on the protein surface. Both sfGFP-AePhe constructs efficiently undergo a bioorthogonal strain-promoted click cycloaddition with a dibenzocyclooctyne derivative.

Introduction

The 20 canonical amino acids are genetically encoded and primarily responsible for the amazing diversity found in the hundreds of proteins that enable life. In addition to these, there are numerous uncommon amino acids that are typically formed by post-translational modifications (PTM), such as oxidization, methylation, acylation, or phosphorylation, of the canonical amino acids. In the past 25 years the ability to incorporate unnatural amino acids (UAA) into proteins has been developed and refined.[1–6] To date over 150 UAA have been incorporated into proteins genetically by codon reassignment and suppression—only one of several methods of incorporation.[1–6] UAA offer several advantages for the study of protein structure and function in that their steric and electronic properties can be rationally designed for the task at hand. UAA have been employed for many diverse purposes such as bioconjugation, reporters, and/or photoreactive groups.[6]

In the case of reporters, a unique spectroscopic signature that can be observed by fluorescence,[7–10] EPR,[11,12] NMR,[13,14] or vibrational spectroscopy[15–17] is required. In recent years vibrational UAA have been employed to study the local hydration and the electrostatic environments of proteins by IR[18–26] and Raman spectroscopy.[26,27] Nitriles and azides are arguably the two most important functional groups for vibrational UAA because of their small size, relative stability, sensitivity, and absorptions that occur in an open window in the IR.[18,28] We recently reported a direct comparison of these two vibrational reporters and demonstrated that for many applications azides are superior because they have a significantly larger extinction coefficient while maintaining excellent sensitivity to their local environment in terms of frequency shifts.[29] Azides also have the advantage of being a conduit to other functionalities through bioorthogonal click cycloaddition reactions with alkynes.[30,31] In fact, it is the bioconjugation applications[32–36] of azido-UAA that are most commonly employed, whereas their ability to also act as an IR reporter has been underutilized, with some notable exceptions.[20,23,37–40] Azido-UAA might be more widely employed as both IR reporters and bioconjugation conduits if an efficient and scalable synthesis was developed from readily available and relatively inexpensive starting materials. Here we report such a synthesis of an azido-UAA, 4-(2-azidoethoxy)-L-phenylalanine (AePhe, **1**; Figure 1 A), and demonstrate that it can be site-selectively incorporated into a protein to act as an IR reporter of local hydration and as a click partner for a bioorthogonal conjugation reaction.

[a] *E. M. Tookmanian, Prof. Dr. C. M. Phillips-Piro, Prof. Dr. E. E. Fenlon, Prof. Dr. S. H. Brewer*
Department of Chemistry Franklin & Marshall College
P.O. Box 3003, Lancaster, PA 17604 (USA)
E-mail: cpiro@fandm.edu
efenlon@fandm.edu
sbrewer@fandm.edu

Supporting information and ORCID(S) for this article is available on the WWW under http://dx.doi.org/10.1002/chem.201503908. It contains additional synthetic procedures, sequence of wt-sfGFP pre- and post-trypsin digest, additional protein mass spectral data, and sfGFP-149-AePhe X-ray crystal structure refinement progression.

Results and Discussion

Design and synthesis of AePhe

AePhe (**1**) was designed with an efficient synthesis in mind as protected tyrosine **2** is commercially available and relatively inexpensive. It was also rationalized that the flexible three-atom linker between the azido group and the aromatic ring would

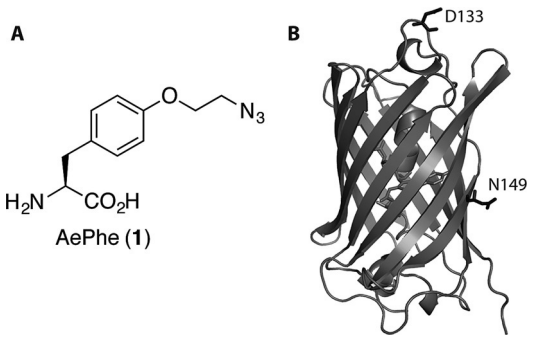


Figure 1. A) Structure of 4-(2-azidoethoxy)-L-phenylalanine. B) Structure of wild-type sfGFP (PDB ID 2B3P) with the 133 and 149 sites labeled and shown in dark sticks. The chromophore is shown in stick for reference.

increase the azide's availability for bioorthogonal click cycloaddition reactions relative to recently published azido-UAA, which had a rigid alkynyl linker.[40] The synthesis of AePhe was accomplished in a three-step process from known materials. Thus, tyrosine **2** was alkylated with 2-azidoethyl mesylate by using potassium carbonate as a base to provide azide **3** in 66% yield (Scheme 1). Deprotection of **3** was accomplished by saponification with lithium hydroxide followed by acid treatment to remove the *tert*-butoxycarbonyl group and provide **1**·HCl in 82% yield for the two steps. The streamlined synthesis requires only one chromatography purification and readily provided gram-quantities of AePhe in three steps and 54% overall yield. This is in contrast to the literature syntheses of other azido-UAA such as 4-azidomethyl-L-Phe (six steps, 33% yield),[23] 4-(3-azido-1-propynyl)-L-Phe (six steps, 42% yield),[40] and 4-(4-azido-1-butynyl)-L-Phe (six steps, 34% yield).[40] The commercially available azido-UAA 4-azido-L-Phe is unstable and undergoes photochemical decomposition.[23] This reactivity is useful for cross-linking experiments[41–44] but the utility of 4-azido-L-Phe as both a click partner[35,36,44] and a vibrational reporter[20,37] requires some precautions because of this unwanted photoreactivity. Numerous additional azido-UAA have been prepared[45] but few of these have been incorporated into proteins and fewer still have been incorporated for use as vibrational probes.[20,23,37–40] After submission of this manuscript recent articles were found that describe the syntheses of AePhe[46–48] and its subsequent incorporation into GFP for bioorthogonal reactions at sites other than those studied here.[46,47] However, these studies did not use the azide as a vibrational reporter nor provide X-ray crystallographic analysis.

Incorporation of AePhe at two sites in sfGFP

Using the amber codon suppression technology, AePhe was site-specifically incorporated at two unique sites in our model protein system, superfolder green fluorescent protein (sfGFP), a 247-residue β-barrel protein (Figure 1B). This system is an ideal model as it produces robust protein yields, has a readily visible confirmation of protein expression, and allows the incorporation a number of UAA at various sites.[23,40,46,47,49] AePhe was incorporated at the 133 site (sfGFP-133-AePhe) on a flexible loop and the 149 site (sfGFP-149-AePhe) on the side of the β-barrel with the side chain facing solvent. A protein gel indicates protein expression only when the unnatural amino acid AePhe is added to the media (Figure 2), which indicates

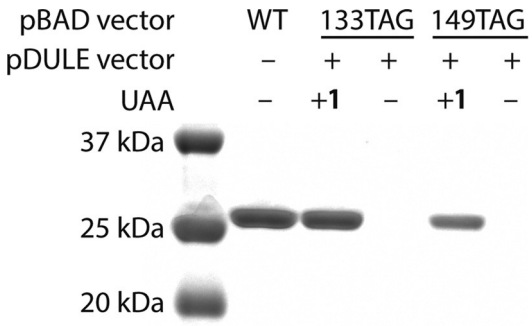
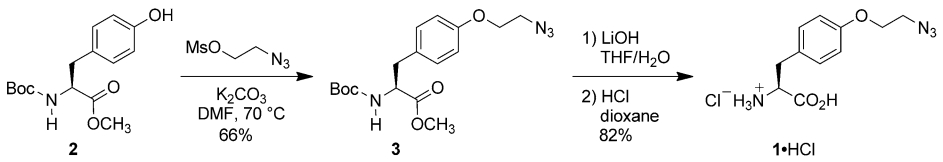


Figure 2. Coomassie stained 12% tris-glycine SDS-PAGE illustrating efficient, site-specific incorporation of AePhe with high fidelity at either the 133 or 149 sites. The protein constructs were expressed from pBAD-sfGFP (WT in lane 2); pBAD-sfGFP-133TAG and pDULE-AzidoPhe in the presence (lane 3) or absence (lane 4) of **1**; pBAD-sfGFP-149TAG and pDULE-AzidoPhe in the presence (lane 5) or absence (lane 6) of **1**. The proteins were purified as described in the Experimental Section prior to running the gel.

the selective nature of the synthetase/tRNA system to incorporate only the UAA. Furthermore, we have confirmed the incorporation of AePhe into sfGFP by mass spectral analysis. Comparing the observed mass of wt-sfGFP with the observed mass of sfGFP-133-AePhe or sfGFP-149-AePhe produced the expected mass differences for the replacement of the 133-aspartic acid with AePhe ($\Delta = +117$ Da) and the 149-asparagine with AePhe ($\Delta = +118$ Da), see Table 1.

Sensitivity of the azide asymmetric stretch of AePhe to the local environment

The dependence of the azide asymmetric stretching frequency of AePhe to solvent was measured to examine the potential of



Scheme 1.

Table 1. sfGFP mass spectrometry data.

Protein construct	Exptl.[a] MW (Da)	Exptl. Δ wt [Da]	Calcd Δ wt [Da]
wt sfGFP	26,863	–	–
sfGFP-133-AePhe	26,979	116	117
sfGFP-149-AePhe	26,981	118	118
sfGFP-133-AePhe-DBCO	27,370	507	507
sfGFP-149-AePhe-DBCO	27,369	507	508

[a] Experimental mass spectrometry data error ± 1 Da.

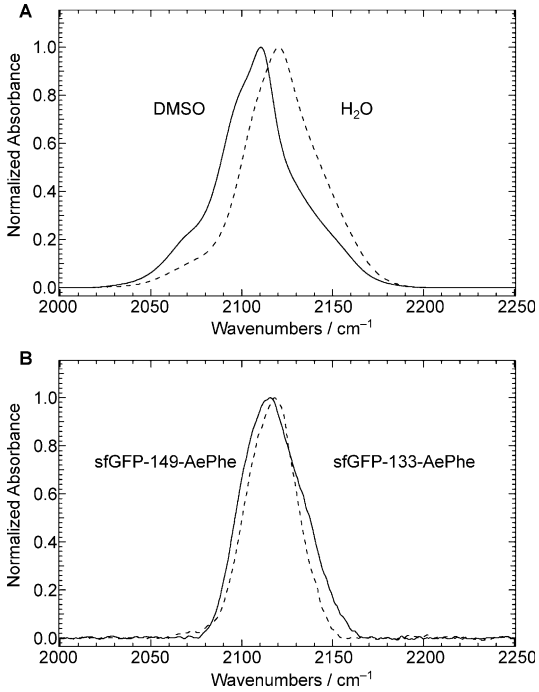


Figure 3. Transmission FTIR absorbance spectra of AePhe dissolved in either DMSO or a basic aqueous solution to a concentration of ~50 mM (A) or sfGFP constructs containing AePhe at either site 149 (solid curve) or 133 (dashed curve) dissolved in a pH 7.3 aqueous buffer composed of 50 mM sodium phosphate and 150 mM sodium chloride to a concentration of ~1–3 mM (B). The spectra were recorded at 25 °C, baseline corrected, and intensity normalized.

AePhe to serve as a vibrational reporter of local environments. Figure 3A shows the equilibrium FTIR spectra of AePhe in the 2000 to 2250 cm^{-1} region dissolved in either DMSO or water, which were selected as mimics of the hydrophobic and hydrophilic environments present in proteins, respectively. These spectra show that the IR absorbance profile of AePhe in the azide region in both solvents is complicated and not symmetrical—likely the result of the presence of multiple sub-components and/or anharmonic effects.[50,51] The highest intensity frequency component of the profile was assigned to the azide asymmetric stretch frequency of AePhe similar to previous studies of azides.[50] This vibrational frequency shifts from 2110.7 cm^{-1} in DMSO to 2120.4 cm^{-1} in water. This 9.7 cm^{-1} blue shift is principally the result of hydrogen bonding between water molecules and the azido group of AePhe,[23,40,52] which is absent in the DMSO solution because DMSO lacks a hydrogen-bond-donating proton. The direction and magnitude of this shift are similar to other azido-modified UAA[40] and indicates the potential of AePhe to serve as a reporter of local hydration in proteins in spite of the complicated IR absorbance profile in the azide region.

Figure 3B shows the equilibrium FTIR spectra of AePhe genetically incorporated at either site 149 or 133 in sfGFP dissolved in an aqueous pH 7.3 buffer composed of 50 mM sodium phosphate and 150 mM sodium chloride in the region 2000–2250 cm^{-1} to probe the local protein environment at these two sites. The buffer was selected for comparison with previous FTIR studies of sfGFP constructs containing vibrational reporter UAA.[23,40,49] The absorption profile in the region of the azide asymmetric stretch vibration in each of these protein constructs appears less complicated compared to the free AePhe dissolved in DMSO or water (Figure 3A). This modulation of the absorption profile suggests that the incorporation of AePhe into the protein alters the anharmonic coupling and/or the frequency of the potential sub-components of the observed absorption band relative to the free AePhe.

The azide asymmetric stretching frequency of sfGFP-149-AePhe and sfGFP-133-AePhe were observed to be 2115.5 and 2118.0 cm^{-1}, respectively, as shown in Figure 3B. The position of the azide asymmetric stretch of sfGFP-133-AePhe indicates that the azido group is hydrated at this site because of the similarity of this frequency with that of AePhe dissolved in water (2120.4 cm^{-1}). This hydration state is not surprising since sfGFP crystal structures[44,53–55] show that this site is a fully solvated loop region of the protein and the side chain does not appear able to fold back onto the protein surface.

Site 149 is located on one of the β-strands of the β-barrel of sfGFP in which the side chain of the amino acid is directed away from the protein into the solvent in numerous crystal structures.[44,53–55] The azido group of 4-(4-azido-1-butynyl)-L-Phe, which has a rigid spacer group, at this site was recently found to be fully solvated.[40] Thus, the azido group of AePhe has the potential to be fully hydrated depending on the conformation of the linker between the azido and the aromatic ring of AePhe. However, the position of the azide asymmetric stretch vibration of sfGFP-149-AePhe suggests that the azido group at this site is partially buried and not fully solvated. This assignment is based on the red shift of the azide asymmetric stretching frequency compared to sfGFP-133-AePhe, which was fully hydrated. On the other hand, the magnitude of the red shift between the two sites (2.5 cm^{-1}) compared to the magnitude of the red shift of AePhe dissolved in DMSO compared to water (9.7 cm^{-1}) suggests that the azido group of sfGFP-149-AePhe is only partially solvated. To further explore the conformation of the side chain of AePhe incorporated at site 149, the crystal structure of this construct was determined.

X-ray crystal structure of sfGFP-149-AePhe

The X-ray crystal structure of sfGFP-149-AePhe was solved (Figure 4A) to confirm the fidelity of the UAA incorporation and explore the difference in the IR spectra of the asymmetric stretch vibration of AePhe at the 149 site versus the 133 site.

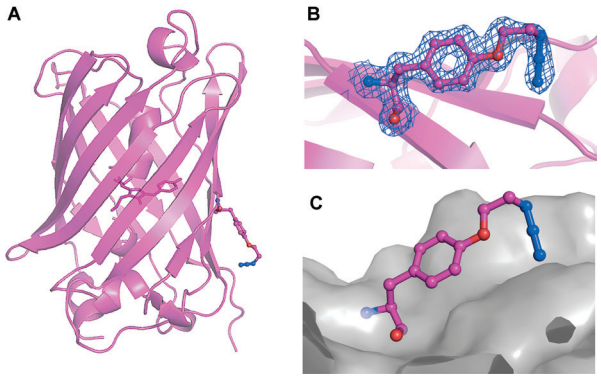


Figure 4. X-ray crystal structure of sfGFP-149-AePhe. A) Overall structure with chromophore shown in sticks for reference and AePhe shown in ball and stick and colored by atom type. B) $2F_o-F_c$ electron density at 1.0 σ illustrating incorporation of AePhe at the 149 site. C) Surface representation surrounding the 149 site emphasizing the orientation of the azido group towards the protein surface and partially buried from solvent by neighboring residues. The structure of sfGFP-149-AePhe is deposited in the Protein Databank and assigned the PDB ID 5EHU.

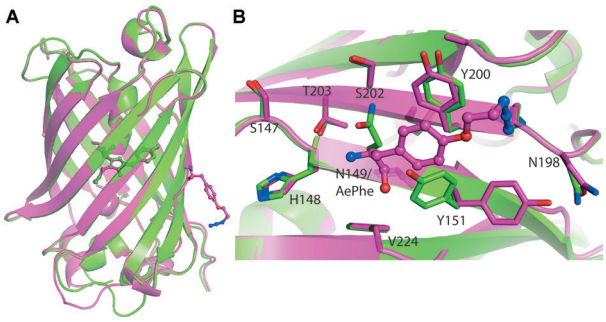


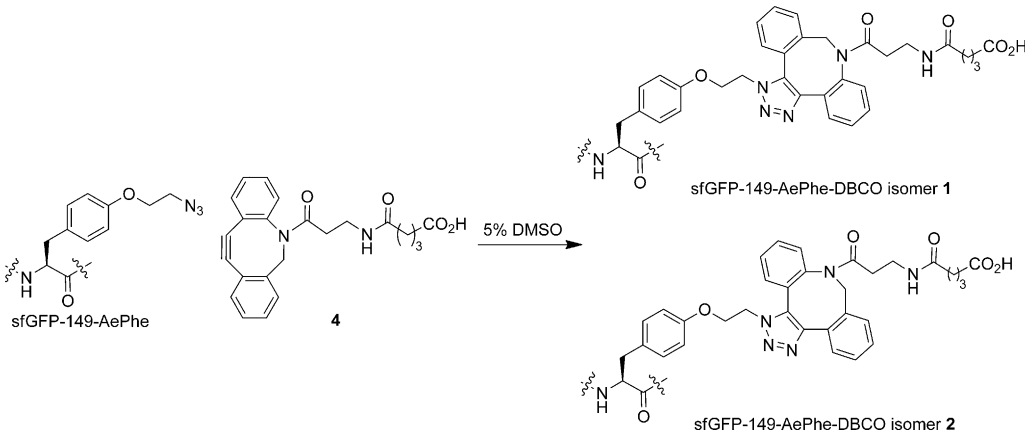
Figure 5. Alignment of sfGFP-149-AePhe (magenta) with wt-sfGFP (green) (A) overall and (B) in the vicinity of the 149 site. AePhe is shown in ball and stick and side chains of residues within 10 Å of the site are shown in sticks and labeled. All explicitly shown side chains are colored by atom type with carbon green (wt) or magenta (AePhe mutant), oxygen red, and nitrogen blue.

The 1.45 Å crystal structure illustrated electron density for the full AePhe amino acid at the 149 site (Figure 4B and Supporting Information). The azido group is orientated towards the protein surface (Figure 4C), supporting the interpretation that the IR shift described above was caused by a less solvated environment compared with the 133 site and the fully solvated 4-(4-azido-1-butynyl)-L-Phe previously studied.[40]

The X-ray crystal structure of sfGFP-149-AePhe also permits the observation of structural changes upon the incorporation of AePhe in sfGFP. As the 149 site is on the side of the β-barrel facing towards the solvent it is expected that there would be little disruption of the protein structure as indicated by the overall root-mean squared deviation (RMSD) between the wt-sfGFP and sfGFP-149-AePhe structures. Indeed, the RMSD was 1.25 Å for all-atom and 0.84 Å for C_α atoms only for the overall structural alignment shown in Figure 5A. An alignment of the residues within 10 Å of the 149 site indicates RMSDs of 1.45 and 0.25 Å for all-atoms and C_α atoms, respectively. A view of the environment within 10 Å of the 149 site illustrates a few side chain movements upon incorporation of AePhe, most notably the approximately 180° rotation of the 149 and 151 side chains when compared to wt-sfGFP (Figure 5B), accounting for the larger all-atom RMSD for the region around the 149 site compared to the overall structure. However, the side chain movements do not result in a large cascade of movements and the backbone of the protein remains mostly unchanged, as indicated by the low RMSD of C_α atoms in the area. Thus the crystal structure provides important structural insights into the nature of the hydration state, which was assigned based upon the observed azide asymmetric stretching frequency of AePhe in sfGFP-149-AePhe and sfGFP-133-AePhe thereby providing further support for the ability of AePhe to serve as a sensitive vibrational reporter UAA.

Click chemistry in sfGFP at incorporated AePhe residues

The ability of azido-UAA such as AePhe to act as vibrational reporters is valuable but they also have the ability to act as a conduit for PTM of proteins through a bioorthogonal click chemistry reaction with alkynes. To demonstrate the ability of AePhe to undergo PTM we explored click reactions under the two standard conditions, strain-promoted[31] and copper(I)-cat-



Scheme 2.

alyzed[30] azide–alkyne cycloadditions. The strain-promoted click reaction proceeded smoothly in duplicate experiments at both incorporation sites. Thus, when sfGFP-149-AePhe or sfGFP-133-AePhe were treated with dibenzocyclooctyne (DBCO) acid derivative **4** the click reaction cycloadduct (Scheme 2) was confirmed by mass spectral analysis (see Table 1). In each case the mass increased by ~390 Da (the mass of **4**) over the AePhe construct which corresponds to a gain of ~507 Da over wtGFP. As shown in Scheme 2, this reaction likely produced two isomeric cycloadducts.

On the other hand, repeated attempts at the copper(I)-catalyzed click reaction of sfGFP-149-AePhe or sfGFP-133-AePhe with propargyl alcohol using Hong et al.'s optimized conditions[30] for the catalyst, ligand, and additional reagents failed to produce the expected cycloadduct according to mass spectral analysis (see Supporting Information). The reasons for this failure are unclear. When the reaction was attempted with the polyhistidine-tag (His-tag) additional copper(I) was added, as recommended, and the reaction was also attempted after the His-tag had been cleaved by a trypsin digest. Propargyl alcohol is fully miscible with water, so the solubility of the click partner was not an issue.

Conclusion

An efficient and scalable synthesis of AePhe was accomplished in three steps from known compounds in 54% overall yield. The IR asymmetric stretch vibration of the azido group of AePhe was shown to be sensitive to its local environment through a solvent study. AePhe was successfully incorporated into sfGFP at the 133 and 149 sites. IR spectroscopy indicated that, as expected, the azido group at the 133 site was fully solvated, whereas, somewhat unexpectedly, the azido group at the 149 site was not. An X-ray crystal structure of sfGFP-149-AePhe provided an explanation for the IR data as the flexible linker adopts a conformation, which partially buries the azide on the protein surface. Both sfGFP-AePhe constructs readily underwent a bioorthogonal strain-promoted click cycloaddition with DBCO **4** demonstrating that AePhe is valuable both as a vibrational reporter and as a conduit for bioconjugation.

Experimental Section

General information

All chemical reagents were purchased from either Sigma–Aldrich or TCI America and used without further purification. Deuterated solvents ($CDCl_3$, 98.8% D enrichment and [D_6]DMSO, 99.5% D enrichment) were purchased from Cambridge Isotope Laboratories. DH10B cells and pBAD were purchased from Invitrogen. 2-Azidoethyl methanesulfonate was synthesized by the literature method of Demko and Sharpless,[56] except that methanesulfonyl chloride was utilized in place of toluenesulfonyl chloride (see the Supporting Information). Protected tyrosine **2** was purchased from TCI America. Dibenzocyclooctyne (DBCO) C6-Acid derivative **4** was purchased from Click Chemistry Tools. All aqueous solutions were prepared with 18 MΩ cm water.

Reactions were stirred with a magnetic stir bar and completed under a dry argon atmosphere. Reactions carried out above ambient temperature were conducted in an oil bath. Analytical TLC was conducted using 0.2 mm silica plastic coated sheets (Selecto Scientific) with F254 indicator and 230–400 mesh silica gel was used for flash column chromatography.

All NMR characterization was completed using a Varian INOVA 500 multinuclear Fourier transform NMR spectrometer (^{1}H NMR at 499.7 MHz and ^{13}C NMR at 125 MHz). Chemical shifts are reported in parts per million (ppm) and coupling constants are reported in Hertz (Hz). NMR spectra in $CDCl_3$ were referenced to the residual solvent peak ($CHCl_3$ = 7.26 ppm) for ^{1}H NMR and the solvent peak ($CDCl_3$ = 77.0 ppm) for ^{13}C NMR. NMR spectra in [D_6]DMSO were referenced to the residual solvent peak ([D_5]DMSO = 2.49 ppm) for ^{1}H NMR and the solvent peak ([D_6]DMSO = 39.5 ppm) for ^{13}C NMR. All IR characterization for synthetic purposes were carried out as ATR thin films with frequencies reported in cm^{-1}. MS analyses of the free UAA for synthetic purposes were carried out on an Agilent 1100 series LC/MSD SL ion trap mass spectrometer with electrospray ionization and MS/MS capabilities.

The ESI-Q-TOF mass analysis was performed on the same purified protein samples used for the linear IR measurements. This analysis was performed at the Mass Spectrometry Facility at the University of Illinois Urbana-Champaign under the direction of Dr. Furong Sun. The protein samples were desalted into a 20 mM ammonium acetate solution using PD10 gel filtration columns, lyophilized, and re-suspended in 1:1 H_2O/CH_3CN with 0.2% formic acid prior to analysis.

N-(*tert*-Butoxycarbonyl)-4-(2-azidoethoxy)-L-phenylalanine methyl ester (3)

To a solution of **2** (3.01 g, 10.1 mmol, 1 equiv) in dry DMF (100 mL) were added K_2CO_3 (4.26 g, 30.4 mmol, 3 equiv) and 2-azidoethyl methanesulfonate (4.80 g, 29.1 mmol, 2.9 equiv). The heterogeneous mixture stirred for 28 h at 70 °C. The reaction mixture was cooled and diluted with diethyl ether and water. The organic layer was washed with brine, dried over $MgSO_4$, filtered through a glass frit, and concentrated under reduced pressure. The crude product was purified by column chromatography (hexanes/ethyl acetate 3:1) to yield **3** (2.38 g, 6.55 mmol, 66%) as a light yellow oil: ^{1}H NMR ($CDCl_3$): δ = 7.05 (d, J = 8.6, 2H), 6.85 (d, J = 8.6, 2H), 4.97 (d, J = 7.8, 1H), 4.54 (m, 1H), 4.13 (t, J = 5.1, 2H), 3.71 (s, 3H), 3.58 (t, J = 5.1, 2H), 3.04 (m, 1H), 1.42 ppm (s, 9H); ^{13}C NMR ($CDCl_3$): δ = 172.37, 157.31, 155.07, 130.39, 128.74, 114.66, 79.91, 66.95, 54.50, 52.20, 50.16, 37.48, 28.29 ppm; IR: $\tilde{\nu}$ = 3357.6 (br, w), 2977.7 (w), 2109.5 (m), 1742.3 (m), 1709.3 (s), 1611.9 (w), 1511.1 (s), 1438.8 (m), 1365.4 (m), 1243.1 (s), 1163.4 (s), 1058.2 (m), 1017.4 (m), 838.9 cm^{-1} (m); MS (ESI) 387 [M+Na^+, 100], 265 (15), 422 (10).

4-(2-Azidoethoxy)-L-phenylalanine hydrochloride (1·HCl)

To a solution of **3** (2.35 g, 6.44 mmol) in THF/H_2O (3:1, 20 mL) was added LiOH monohydrate (0.415 g, 9.66 mmol, 1.5 equiv). The solution stirred at room temperature for 4 h and then the pH was adjusted to ~2.5 by using aq. $NaHSO_4$ (0.5 M). The reaction mixture was diluted with ethyl acetate and the organic layer was washed with water and brine, dried over $MgSO_4$, filtered through a glass frit, and concentrated under reduced pressure. The residue was dissolved in dry HCl in dioxane (2.5 M, 17 mL). After stirring for 4 h at room temperature the solution was concentrated under reduced pressure to about one-third of its initial volume and pentane was added. The white precipitate was collected by vacuum filtration to

provide **1**·HCl (1.51 g, 5.25 mmol, 82%) as a white powder: ^{1}H NMR ([D$_6$]DMSO): δ = 7.20 (d, J = 8.8, 2H), 6.90 (d, J = 8.8, 2H), 4.13 (t, J = 4.9, 2H), 4.05 (t, J = 6.3, 1H), 3.62 (t, J = 4.9, 2H), 3.09 ppm (d, J = 6.3, 2H); ^{13}C NMR ([D$_6$]DMSO): δ = 170.74, 157.65, 131.20, 127.73, 115.00, 67.26, 53.77, 50.02, 35.21 ppm; IR: $\tilde{\nu}$ = 3435.5 (br, w), 2970.6 (m), 2874.5 (m), 2109.0 (s), 2068.0 (w), 1737.2 (s), 1612.0 (w), 1513.8 (s), 1487.7 (s), 1445.2 (w), 1289.9 (m), 1247.2 (s), 1230.3 (s), 1204.0 (m), 1114.6 (m), 1061.8 (w), 837.7 (m), 803.3 cm^{-1} (m); MS (ESI) 251.0 [M^+, 100], 233.9 (25), 205 (5).

Expression and purification of sfGFP constructs

The following plasmids were obtained from Dr. Ryan A. Mehl (Oregon State University) in which the numbering scheme employed here corresponds to the published wild-type sfGFP (wt-sfGFP) crystal structure (PDB ID 2B3P).[53] The plasmid pBAD-sfGFP was generated by inserting a codon-optimized gene coded with wt-sfGFP[53] containing a C-terminal 6-His affinity tag into a pBAD plasmid, generating pBAD-sfGFP.[57,58] Site-directed mutagenesis was used to individually replace the D133 or N149 codons with the amber stop codon (TAG) to generate the pBAD-sfGFP-133TAG and pBAD-sfGFP-149TAG plasmids, respectively. The engineered aminoacyl-tRNA synthetase for the incorporation of *para* azido-modified phenylalanine UAA was inserted into a pDULE plasmid, generating pDULE-AzidoPhe.[23,57] DH10B E. coli cells were transformed with pBAD-sfGFP for wt-sfGFP, while DH10B E. coli cells were co-transformed with pBAD-sfGFP-133TAG or pBAD-sfGFP-149TAG and pDULE-AzidoPhe for the sfGFP-133-UAA and sfGFP-149-UAA constructs, respectively. Transformed cells were used to inoculate 6 mL of noninducing media, which grew to saturation while shaking (250 rpm) at 37 °C. Autoinduction media (250 mL) was then inoculated with 2.5 mL of the cultured cells. A 1–2 mL solution of AePhe was added 1 h post-inoculation for a final concentration of 1 mM. Negative control experiments did not include AePhe in the autoinduction media (Figure 2). The autoinduction media cultures shook at 37 °C for 24–30 h when the cells were harvested by centrifugation and the expressed protein was purified using TALON cobalt ion-exchange chromatography (Clontech) similar to a previous literature procedure.[23] To verify site-specific incorporation of the UAA into site 133 and 149 with high efficiency and fidelity, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE, Figure 2) and electrospray ionization quadrupole time-of-flight (ESI-Q-TOF) mass analysis (Table 1) at the University of Illinois at Urbana-Champaign were carried out. The sfGFP constructs containing AePhe (**1**) generally yielded 5–10 mg of purified protein per liter of autoinduction media. Protein yields were calculated using the extinction coefficient of sfGFP at 488 nm.[53] The following protein constructs, sfGFP-133-AePhe and sfGFP-149-AePhe, were the result of incorporation of AePhe into sfGFP site 133 or site 149, respectively. PD10 gel filtration columns were used to desalt the purified protein constructs into an aqueous buffer solution (50 mM sodium phosphate, 300 mM sodium chloride) with a pH of 8. The C-terminal 6-His affinity tag was cleaved utilizing a catalytic amount of trypsin (1%). After 2 h of incubation at 37 °C, the 6-His affinity tag was fully cleaved, resulting in a 239 residue protein. To deactivate the trypsin, a ten-fold molar excess of phenylmethanesulfonyl fluoride (PMSF) compared to trypsin was added to the solution. Removal of sfGFP still containing a 6-His affinity tag sfGFP was accomplished by TALON cobalt ion-exchange chromatography. The resulting purified UAA containing sfGFP constructs were desalted into the appropriate buffer for further steps utilizing PD10 gel filtration columns or buffer exchange and concentration using 10 K MWCO Centricons (Millipore).

Equilibrium FTIR measurements

The equilibrium FTIR absorbance spectra were recorded using a Bruker Vertex 70 FTIR spectrometer, equipped with a globar source, KBr beamsplitter, and a liquid nitrogen cooled mercury cadmium telluride (MCT) detector. The spectra were measured using a dual-compartment, temperature-controlled transmission cell composed of two calcium fluoride windows with a path length of ~100 μm. The spectra were recorded at 25 °C with a resolution of 1.0 cm^{-1} and were the result of 1024 or 2048 scans. Analysis of the spectra was performed in Igor Pro (Wavemetrics, Inc).

Structure determination of sfGFP-149-AePhe

A 40 mg mL^{-1} solution of sfGFP-149-AePhe in a 20 mM Hepes buffer pH 7.5 was combined with a precipitation solution (20% PEG 8000, 100 mM Hepes pH 7.5) in a 1:1 ratio to form crystals in a sitting drop well at room temperature. Single green crystals were mounted on loops, cryoprotected by consecutive soaks of 8, 13, and 25% ethylene glycol-supplemented precipitant solution and frozen in liquid nitrogen. Diffraction data were collected at the NE-CAT 24-ID-E beamline at the Advanced Photon Source (APS) at Argonne National Lab (ANL) and processed in space group C2 to 1.45 Å using HKL2000.[59] A molecular replacement solution was determined using Phaser[60] (TFZ = 27.5, LLG = 7651) with the search model of wild-type sfGFP (PDB ID 2B3P)[53] in which residue 149 was replaced with an alanine. After a single round of refinement a phenylalanine was modeled into the $2F_o-F_c$ density at position 149; and following a few more rounds of refinement the azido-ethoxy group at the *para* position of the phenylalanine was modeled using the F_o-F_c difference density (see Supporting Information). Rounds of manual refinement in COOT[61] and automated refinement in Phenix[62] were continued to produce the reported structure with R/R_{free} of 17.5/20.8% (Table 2).

Table 2. X-ray data collection and processing statistics for the sfGFP-149-AePhe structure.

Space group	C2
Cell dimensions	
a [Å]	129.917
b [Å]	37.492
c [Å]	91.848
α, β, γ [°]	90, 106.307, 90
λ [Å]	0.97914
T [K]	100
unique reflections	73 332
resolution range [Å][a]	50–1.45 (1.48–1.45)
average Redundancy[a]	3.8 (3.9)
completeness [%][a]	96.4 (94.3)
$I/\sigma[I]$[a]	16.5 (1.9)
# sfGFP molecules per asu	2
R_{cryst} (R_{free}) [%][b]	17.5 (20.8)
rms deviations	
bond lengths [Å]	0.009
bond angles [°]	1.357
Ramachandran plot [%][c]	
preferred	97.6 (404)
allowed	1.9 (8)
outliers	0.5 (2)

[a] The number in parentheses is for the highest resolution shell. [b] $R_{cryst} = \Sigma hkl \; ||F_o(hkl)| - |F_c(hkl)|| / \Sigma hkl \; |F_o(hkl)|$. $R_{free} = R_{cryst}$ for a test set of reflections (5%) not included in refinement. [c] Numbers in parentheses are the number of residues in each category.

Strain-promoted click cycloaddition reactions

The sfGFP construct (either sfGFP-133-AePhe or sfGFP-149-AePhe) was initially desalted into a buffer (pH 7.3; 50 mM sodium phosphate; 150 mM sodium chloride) using PD10 columns and concentrated if necessary utilizing 10 K MWCO Centricons (Millipore). The strained-promoted cycloaddition reaction was carried out with a DBCO acid derivative (**4**, MW = 390 Da), which was initially dissolved in DMSO to make a stock solution at a concentration of 20 mM. The reactions were carried out in 1.5 mL tubes containing a final volume of 500 μL with final concentrations of 100 μM protein, 300 μM DBCO **4**, and 5% (v/v) DMSO. All click reactions were allowed to react for 15–18 h at 37 °C while turning in a Labnet mini labroller. For all click reactions, dialysis was used to remove the reaction reagents by dialyzing for 8–12 h with buffer (pH 7.3; 50 mM sodium phosphate; 150 mM sodium chloride) using Slide-A-Lyzer MINI Dialysis Devices. The samples were desalted using PD10 columns into 20 mM ammonium acetate and lyophilized for ESI-Q-TOF mass spectrometry analysis.

Acknowledgements

We thank R. Keyes for the synthesis of 2-azidoethyl methanesulfonate, K. Hess for mass spectral (ESI) analyses of small molecules, L. Mertzman for ordering supplies, and B. Buckwalter for acquiring NMR spectra. This work was supported by F&M Hackman funds, F&M Eyler Funds, Leser Scholars Grant, and Snavely Awards to E.M.T. NSF (CHE-1053946) to S.H.B. and NIH (R15M093330-02) to E.E.F. and S.H.B. X-ray diffraction data were collected at the Advanced Photon Source (APS) on the Northeastern Collaborative Access Team (NE-CAT) beamline 24-ID-E, which is supported by an NIH NIGMS grant P41 GM103403. APS is operated for the U.S. Department of Energy (DOE) Office of Science by Argonne National Laboratory and is supported by the U.S. DOE under Contract No. DE-AC02-06CH11357.

Keywords: azides · click chemistry · IR spectroscopy · structural biology · vibrational reporters

[1] C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **1989**, *244*, 182–188.
[2] M. W. Nowak, P. C. Kearney, M. E. Saks, C. G. Labarca, S. K. Silverman, W. Zhong, J. Thorson, J. N. Abelson, N. Davidson, P. G. Schultz, D. A. Dougherty, H. A. Lester, *Science* **1995**, *268*, 439–442.
[3] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413–444.
[4] H. Neumann, *FEBS Lett.* **2012**, *586*, 2057–2064.
[5] K. Wals, H. Ovaa, *Front Chem.* **2014**, *2*, 15.
[6] A. Dumas, L. Lercher, C. D. Spicer, B. G. Davis, *Chem. Sci.* **2015**, *6*, 50–69.
[7] M. J. Tucker, R. Oyola, F. Gai, *J Phys Chem. B* **2005**, *109*, 4788–4795.
[8] M.-L. Tsao, D. Summerer, Y. Ryu, P. G. Schultz, *J. Am. Chem. Soc.* **2006**, *128*, 4572–4573.
[9] P. Marek, R. Gupta, D. P. Raleigh, *ChemBioChem* **2008**, *9*, 1372–1374.
[10] E. J. Petersson, J. M. Goldberg, R. F. Wissner, *Phys. Chem. Chem. Phys.* **2014**, *16*, 6827–6837.
[11] M. R. Fleissner, E. M. Brustad, T. Kálai, C. Altenbach, D. Cascio, F. B. Peters, K. Hideg, S. Peuker, P. G. Schultz, W. L. Hubbell, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21637–21642.
[12] M. J. Schmidt, J. Borbas, M. Drescher, D. Summerer, *J. Am. Chem. Soc.* **2014**, *136*, 1238–1241.
[13] J. C. Jackson, A. J. T. Hammill, R. A. Mehl, *J. Am. Chem. Soc.* **2007**, *129*, 1160–1166.
[14] D. H. Jones, S. E. Cellitti, X. Hao, Q. Zhang, M. Jahnz, D. Summerer, P. G. Schultz, T. Uno, B. H. Geierstanger, *J. Biomol. NMR* **2010**, *46*, 89–100.
[15] B. A. Lindquist, K. E. Furse, S. A. Corcelli, *Phys. Chem. Chem. Phys.* **2009**, *11*, 8119–8132.
[16] H. Kim, M. Cho, *Chem. Rev.* **2013**, *113*, 5817–5847.
[17] J. Ma, I. M. Pazos, W. Zhang, R. M. Culik, F. Gai, *Annu. Rev. Phys. Chem.* **2015**, *66*, 357–377.
[18] Z. Getahun, C.-Y. Huang, T. Wang, B. De León, W. F. DeGrado, F. Gai, *J. Am. Chem. Soc.* **2003**, *125*, 405–411.
[19] K. C. Schultz, L. Supekova, Y. Ryu, J. Xie, R. Perera, P. G. Schultz, *J. Am. Chem. Soc.* **2006**, *128*, 13984–13985.
[20] S. Ye, T. Huber, R. Vogel, T. P. Sakmar, *Nat. Chem. Biol.* **2009**, *5*, 397–399.
[21] A. T. Fafarman, S. G. Boxer, *J Phys Chem. B* **2010**, *114*, 13536–13544.
[22] D. C. Urbanek, D. Y. Vorobyev, A. L. Serrano, F. Gai, R. M. Hochstrasser, *J. Phys. Chem. Lett.* **2010**, *1*, 3311–3315.
[23] C. G. Bazewicz, M. T. Liskov, K. J. Hines, S. H. Brewer, *J Phys Chem. B* **2013**, *117*, 8987–8993.
[24] D. M. Walker, E. C. Hayes, L. J. Webb, *Phys. Chem. Chem. Phys.* **2013**, *15*, 12241–12252.
[25] I. Peran, T. Oudenhoven, A. M. Woys, M. D. Watson, T. O. Zhang, I. Carrico, M. T. Zanni, D. P. Raleigh, *J Phys Chem. B* **2014**, *118*, 7946–7953.
[26] C. H. Londergan, R. Baskin, C. G. Bischak, K. W. Hoffman, D. M. Snead, C. Reynoso, *Biochemistry* **2015**, *54*, 83–95.
[27] C. L. Weeks, A. Polishchuk, Z. Getahun, W. F. DeGrado, T. G. Spiro, *J. Raman Spectrosc.* **2008**, *39*, 1606–1613.
[28] I. T. Suydam, S. G. Boxer, *Biochemistry* **2003**, *42*, 12050–12055.
[29] X. S. Gai, B. A. Coutifaris, S. H. Brewer, E. E. Fenlon, *Phys. Chem. Chem. Phys.* **2011**, *13*, 5926–5930.
[30] V. Hong, S. I. Presolski, C. Ma, M. G. Finn, *Angew. Chem. Int. Ed.* **2009**, *48*, 9879–9883; *Angew. Chem.* **2009**, *121*, 10063–10067.
[31] J. C. Jewett, C. R. Bertozzi, *Chem. Soc. Rev.* **2010**, *39*, 1272–1279.
[32] D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2009**, *131*, 8720–8721.
[33] S. C. Reddington, E. M. Tippmann, D. D. Jones, *Chem. Commun.* **2012**, *48*, 8419–8421.
[34] D. M. Abdeljabbar, F. J. Piscotta, S. Zhang, A. J. Link, *Chem. Commun.* **2014**, *50*, 14900–14903.
[35] S. Naganathan, S. Ray-Saha, M. Park, H. Tian, T. P. Sakmar, T. Huber, *Biochemistry* **2015**, *54*, 776–786.
[36] A. M. Hartley, A. J. Zaki, A. R. McGarrity, C. Robert-Ansart, A. V. Moskalenko, G. F. Jones, M. F. Craciun, S. Russo, M. Elliott, J. E. Macdonald, D. D. Jones, *Chem. Sci.* **2015**, *6*, 3712–3717.
[37] S. Ye, E. Zaitseva, G. Caltabiano, G. F. X. Schertler, T. P. Sakmar, X. Deupi, R. Vogel, *Nature* **2010**, *464*, 1386–1389.
[38] S. Nagarajan, H. Taskent-Sezgin, D. Parul, I. Carrico, D. P. Raleigh, R. B. Dyer, *J. Am. Chem. Soc.* **2011**, *133*, 20335–20340.
[39] J.-H. Choi, D. Raleigh, M. Cho, *J. Phys. Chem. Lett.* **2011**, *2*, 2158–2162.
[40] E. M. Tookmanian, E. E. Fenlon, S. H. Brewer, *RSC Adv.* **2015**, *5*, 1274–1281.
[41] J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 9026–9027.
[42] K. Zhang, M. R. Diehl, D. A. Tirrell, *J. Am. Chem. Soc.* **2005**, *127*, 10136–10137.
[43] I. S. Carrico, S. A. Maskarinec, S. C. Heilshorn, M. L. Mock, J. C. Liu, P. J. Nowatzki, C. Franck, G. Ravichandran, D. A. Tirrell, *J. Am. Chem. Soc.* **2007**, *129*, 4874–4875.
[44] S. C. Reddington, P. J. Rizkallah, P. D. watson, R. Pearson, E. M. Tippmann, D. D. Jones, *Angew. Chem. Int. Ed.* **2013**, *52*, 5974–5977; *Angew. Chem.* **2013**, *125*, 6090–6093.
[45] H. Johansson, D. S. Pedersen, *Eur. J. Org. Chem.* **2012**, 4267–4281.
[46] A. Tuley, Y.-S. Wang, X. Fang, Y. Kurra, Y. H. Rezenom, W. R. Liu, *Chem. Commun.* **2014**, *50*, 2673–2675.
[47] J. C. Maza, J. R. McKenna, B. K. Raliski, M. T. Freedman, D. D. Young, *Bioconjugate Chem.* **2015**, *26*, 1884–1889.
[48] B. A. Seigal, W. H. Connors, A. Fraley, R. M. Borzilleri, P. H. Carter, S. L. Emanuel, J. Fargnoli, K. Kim, M. Lei, J. G. Naglich, M. E. Pokross, S. L. Posy, H. Shen, N. Surti, R. Talbott, Y. Zhang, and N. K. Terrett, *J. Med. Chem.* **2015**, *58*, 2855–2861.
[49] C. G. Bazewicz, J. S. Lipkin, E. E. Smith, M. T. Liskov, S. H. Brewer, *J Phys Chem. B* **2012**, *116*, 10824–10831.
[50] M. W. Nydegger, S. Dutta, C. M. Cheatum, *J. Chem. Phys.* **2010**, *133*, 134506.

[51] J. S. Lipkin, R. Song, E. E. Fenlon, S. H. Brewer, *J. Phys. Chem. Lett.* **2011**, *2*, 1672 – 1676.
[52] J.-H. Choi, K.-I. Oh, M. Cho, *J. Chem. Phys.* **2008**, *129*, 174512.
[53] J.-D. Pédelacq, S. Cabantous, T. Tran, T. C. Terwilliger, G. S. Waldo, *Nat. Biotechnol.* **2006**, *24*, 79 – 88.
[54] X. Liu, J. Li, C. Hu, Q. Zhou, W. Zhang, M. Hu, J. Zhou, J. Wang, *Angew. Chem. Int. Ed.* **2013**, *52*, 4805 – 4809; *Angew. Chem.* **2013**, *125*, 4905 – 4909.
[55] A. B. Dippel, G. M. Olenginski, N. Maurici, M. T. Liskov, S. H. Brewer, C. M. Phillips-Piro, *submitted* **2015**.
[56] Z. P. Demko, K. B. Sharpless, *Org. Lett.* **2001**, *3*, 4091 – 4094.
[57] S. J. Miyake-Stoner, C. A. Refakis, J. T. Hammill, H. Lusic, J. L. Hazen, A. Deiters, R. A. Mehl, *Biochemistry* **2010**, *49*, 1667 – 1677.
[58] E. E. Smith, B. Y. Linderman, A. C. Luskin, S. H. Brewer, *J Phys Chem. B* **2011**, *115*, 2380 – 2385.
[59] Z. Otwinowski, W. Minor, *Methods Enzymol.* **1997**, *276*, 307 – 326.
[60] L. C. Storoni, A. J. McCoy, R. J. Read, *Acta Crystallogr. Sect. D* **2004**, *60*, 432 – 438.
[61] P. Emsley, K. Cowtan, *Acta Crystallogr. Sect. D* **2004**, *60*, 2126 – 2132.
[62] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, T. C. Terwillinger, P. H. Zwart, *Acta Crystallogr. Sect. D* **2010**, *66*, 213 – 221.

Received: September 29, 2015
Published online on November 26, 2015