250 *Biochemistry* **2011,** *50,* 250–257
DOI: 10.1021/bi1016183

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Crystallographic Study of a Site-Specifically Cross-Linked Protein Complex with a Genetically Incorporated Photoreactive Amino Acid[†,‡]

Shin Sato,[§,⊥] Shinya Mimasu,[§,⊥,∥] Aya Sato,[§,⊥] Nobumasa Hino,[§] Kensaku Sakamoto,[§] Takashi Umehara,[§] and Shigeyuki Yokoyama*[,§,∥]

[§]*RIKEN Systems and Structural Biology Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan, and* [∥]*Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.* [⊥]*These authors contributed equally to this work.*

Received October 7, 2010; Revised Manuscript Received November 18, 2010

ABSTRACT: The benzophenone photophore is widely used to photo-cross-link macromolecules. Recent developments in genetic code expansion have allowed the biosynthesis of proteins with *p*-benzoyl-L-phenylalanine (*p*Bpa) at defined sites, for covalent bonding with interacting proteins. However, the structure of a photo-cross-linked protein complex had not been revealed, and thus neither the actual structure of the "photobridge" in a complex nor the influence of this covalent bridge on the overall complex structure was known. In this study, we determine the crystal structure of the cross-linked complex of the liver oncoprotein gankyrin and the C-terminal domain of S6 proteasomal protein (S6C), at 2.05 Å resolution. First, the photoreactive amino acid was separately incorporated into gankyrin at 16 sites on the protein surface, and two variants that efficiently formed a covalent bond with S6C were found. The yield of one of the cross-linked products, with *p*Bpa in place of Arg85 in gankyrin, was maximized for crystallization via optimization of the duration of complex exposure to 365 nm light. The structure revealed that the carbonyl group of the benzophenone of *p*Bpa85 formed a covalent bond exclusively with the Cγ atom of Glu356 in S6C, showing the high selectivity of formation of cross-links by *p*Bpa. In addition, the cross-linked structure exhibited little structural distortion from the native complex structure. Our results demonstrated that cross-linking with site-specifically incorporated *p*Bpa preserves the native binding mode and is useful for probing protein−protein interactions.

The benzophenone photophore has been conjugated with biomolecules, including lipids, nucleotides, and sugars, for use in a broad range of biological analyses, such as labeling and detecting proteins of interest, identifying interacting partners, and mapping binding domains in cells or in vitro (*1*−*7*). The development of *p*-benzoyl-L-phenylalanine (*p*Bpa), a photoreactive amino acid, has allowed the incorporation of benzophenones into peptides at any desired sites by chemical synthesis and allowed the photolabeling of biomolecules of interest. These applications take advantage of the characteristics of the benzophenone photophore, which is chemically stable and activated by 365 nm light with little damage to biomolecules and preferentially reacts with C−H bonds (*3*). The activated benzophenones, which generate radicals, do not react with solvent water, and the yield of the cross-linked products increases with longer exposures to light. This property is particularly beneficial for detecting weak and transient interactions in biological systems.

Non-natural amino acids can be site-specifically incorporated into proteins by using cell-free and cell-based protein synthesis systems with expanded genetic codes. The photoreactive amino acid, *p*Bpa, is thus genetically encoded with the amber codon in *Escherichia coli*, yeast, and mammalian cells, by expressing an exogenous pair of an amber suppressor tRNA and an engineered tyrosyl-tRNA synthetase specific to *p*Bpa (*8*−*11*). The first demonstration of protein complex cross-linking was performed in *E. coli* cells, with a glutathione *S*-transferase dimer as a model system (*8*). In mammalian cells, the epidermal growth factor (EGF) receptor (EGFR) and adaptor protein GRB2 were covalently linked with each other, via incorporation of *p*Bpa into GRB2 in the proximity of the site involved in binding to the phosphorylated peptide tail of EGFR (*9*). The cross-linking reaction between EGFR and GRB2 depended on both the EGF stimulus to the cell and the position at which *p*Bpa was incorporated into GRB2, which showed that the cross-link thus formed reflects the proper interaction between the molecules and that *p*Bpa can react only with amino acid residues within its proximity (*9*). In addition, the G protein-coupled receptor (GPCR) Ste2p and its ligand, α factor, were covalently linked by incorporating *p*Bpa into Ste2p in yeast *Saccharomyces cerevisiae* (*12*), confirming the ability of a *p*Bpa-containing GPCR protein to capture a natural ligand in its native environment.

Some important questions still remained unanswered, concerning (i) the detailed requirements for *p*Bpa to form a cross-link in a protein complex, (ii) the structure of the cross-linked site in the complex, (iii) the influence of cross-linking on the overall complex structure, and (iv) the procedure for preparing an

[†]This work was supported by the Targeted Proteins Research Program (TPRP) of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (to S.Y.) and by Grants-in-Aid to A.S. (21770195) and T.U. (19790083 and 22790101) from MEXT. S.M. is a research fellow of the Japan Society for the Promotion of Science (JSPS).

[‡]The structural coordinates of the gankyrin−S6C cross-linked complex have been deposited in the Protein Data Bank as entry 3AJI.

*To whom correspondence should be addressed. E-mail: yokoyama@biochem.s.u-tokyo.ac.jp. Phone: +81-45-503-9196. Fax: +81-45-503-9195.

 Published on Web 12/03/2010

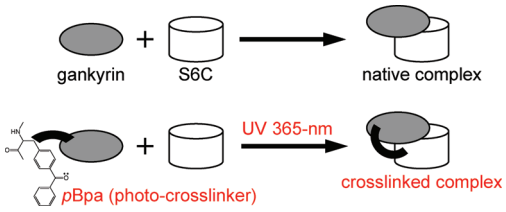

FIGURE 1: Schematic representation of the photo-cross-linking of a protein complex for structure determination. The target proteins, the oncoprotein gankyrin and the C-terminal domain of S6 (S6C), form a native complex (top). In this study, the residues composing the surface located in several different positions of gankyrin were replaced with *p*Bpa (bottom). After being exposed to 365 nm UV light, the complex of the two proteins is expected to be photo-cross-linked.

adequate amount of the complex with sufficient purity for structural studies. In this study, we cross-linked the liver oncoprotein gankyrin with a binding partner, the C-terminal domain of S6 ATPase (S6C) from the 26S proteasome (Figure 1), because the crystal structure of the native complex of these molecules was available (*13*). These two proteins were coexpressed in *E. coli* cells, to allow S6C to remain soluble, because this domain was insoluble in the absence of coexpressed gankyrin (*13*). *p*Bpa was incorporated into gankyrin at several different sites, selected on the basis of the complex structure, to find a position that would achieve efficient cross-linking between gankyrin and S6C. We determined the crystal structure of the cross-linked complex between gankyrin containing *p*Bpa at a specific site and the partner protein S6C at 2.05 Å resolution and found that the benzophenone of *p*Bpa generated a unique, covalent bridge to the target partner protein, with minimal distortion of the native complex structure.

MATERIALS AND METHODS

Materials. *p*-Benzoyl-L-phenylalanine (*p*Bpa) was purchased from Bachem AG. The anti-FLAG M2 antibody was purchased from Sigma-Aldrich. The anti-hemagglutinin antibody was purchased from Roche Diagnostics. The horseradish peroxidase-conjugated anti-mouse IgG and anti-rabbit IgG antibodies were both purchased from Amersham Biosciences.

Incorporation of pBpa into Gankyrin. The *p*Bpa was incorporated into gankyrin at specific sites, essentially according to the method described previously (*8*). The gene encoding a variant of the archaeal tyrosyl-tRNA synthetase specific to *p*Bpa, MjBpaRS-1 (*8*), with the Asp286Arg substitution enhancing the recognition of the amber anticodon (*14*), was placed under the control of the *tyrS* promoter and cloned in vector pACYC184, which also carried three copies of the gene encoding an amber suppressor tRNA derived from the *Methanococcus jannaschii* $tRNA^{Tyr}$. This suppressor tRNA was described previously (*15*) but has G and U at positions 45 and 47, respectively. The three copies of the tRNA gene were transcribed from a single promoter, the *lpp* promoter, to the *rrnC* terminator. This plasmid, designated *p*BpaRS3tRNA, was used to transform *E. coli* cells to produce proteins containing *p*Bpa at the amber position.

Construction of the Gankyrin−S6C Expression Plasmids. The *gankyrin* gene in the pETDuet1-gankyrin-S6C expression plasmid (*13*), encoding full-length mouse gankyrin (residues 1−231) and the C-terminal region of human S6 ATPase (residues 337−418; S6C), was mutagenized using a QuikChange site-directed mutagenesis kit (Stratagene) to generate a series of clones with an amber codon at positions N7, Y15, Q38, A50, D71, G73, G84, R85, Q104, K116, K149, D169, T170, E171, E183, and R184. For detection of the cross-linked complexes, a FLAG tag sequence (DYKDDDDK) was added to the C-terminus of S6C.

Protein Expression, Photo-Cross-Linking, and Purification. *E. coli* BL21(DE3) (Novagen) was used for the coexpression of the gankyrin and S6C proteins. For all protein expression, the bacteria were grown at 37 °C in LB medium, containing 1% glucose, 50 mg/mL ampicillin, and 34 mg/mL chloramphenicol, until the OD_{600} reached a value of 0.8. Protein expression was then induced by the addition of LB medium, containing the following final concentrations: 10 mM HEPES-Na (pH 7.2), 1 mM *p*Bpa (dissolved in HCl immediately before use), and 0.4 mM IPTG. After a 6 h induction of the 2 L culture at 37 °C, the cells were harvested by centrifugation and resuspended in 20 mM Tris-HCl (pH 8.0), containing 20 mM imidazole and 100 mM NaCl. The cells were then homogenized with a sonicator (TOMY), and the homogenate was centrifuged at 5800*g* for 30 min. The resultant supernatants containing the gankyrin (*p*Bpa85)−S6C complex were then placed on ice in a cold room (4 °C) and exposed to 365 nm light from an 8 W lamp (UVP) for 10 h. The samples were centrifuged at 5800*g* for 30 min. The supernatants were filtered through a MILLEX-HV PVDF 0.45 μm membrane (Millipore) and then applied to a HisTrap HP column (GE Healthcare). The complex was eluted with 20 mM Tris-HCl (pH 8.0) buffer, containing 500 mM imidazole and 100 mM NaCl, and was then loaded onto a HiPrep desalting column (GE Healthcare). The samples containing the cross-linked complex were fractionated in 20 mM Tris-HCl (pH 8.0) buffer with 100 mM NaCl. The photo-cross-linked complex was loaded onto a Mono Q 5/50 GL anion-exchange column (GE Healthcare), which was eluted with a linear NaCl gradient from 0.1 to 1.0 M. The fractions containing the target complex were combined and then loaded onto a HiPrep desalting column, which was eluted with 20 mM Tris-HCl (pH 8.0) buffer containing 100 mM NaCl. For further purification of the cross-linked complex, the protein solution was concentrated to less than 5 mL with an Amicon Ultra filter [10000 molecular weight cutoff (MWCO) (Millipore)] and was then fractionated on a HiLoad 16/60 Superdex 200 gel-filtration column (GE Healthcare), which was eluted with 20 mM Tris-HCl (pH 8.0) buffer, containing 150 mM NaCl and 2 mM DTT. The cross-linked complex was concentrated and finally fractionated on a HiLoad 16/60 Superdex 200 column. The pooled cross-linked gankyrin (*p*Bpa85)−S6C complex was concentrated to 7 mg/mL with an Amicon Ultra filter (10000 MWCO). Photo-cross-linking and purification of the gankyrin (*p*Bpa149)−S6C cross-linked complex were performed essentially according to the same procedure.

Western Blotting. Bacterial cells that coexpressed gankyrin and S6C were lysed in buffer A [30 mM Tris-HCl buffer (pH 7.4), 10% glycerol, 1% Triton X-100, 5 mM EDTA, and 0.05% sodium deoxycholate, with a 1:100 dilution of protease inhibitor cocktail (Nacalai Tesque)]. The proteins were subjected to the photo-cross-linking procedure, fractionated by sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE), and transferred to polyvinylidene difluoride (PVDF) membranes (Millipore). The proteins on the membrane were probed with antibodies and then detected with the ECL plus immunodetection system (Amersham Biosciences). The membranes were stripped in 62.5 mM Tris-HCl buffer (pH 6.8), containing 2% SDS and 0.7% 2-mercaptoethanol, for 30 min at 55 °C in preparation for being probed with the second antibody. The band intensity was measured with an image analyzer, LAS-1000plus (FUJIFILM).

Table 1: Summary of Cross-Linking Efficiencies

gankyrin residue	S6C target	Cα−Cα distance (Å)[a]	cross-link efficiency	Cζ[b]−nearest C distance (Å)[a]	Cγ[b]−Cζ[b]−nearest C angle (deg)[c]
N7	E389	9.13	−	10.60	108.25
Y15	I393	11.23	−	5.09	132.39
Q38	E389	9.12	−	4.91	116.20
A50	E356	7.01	−	8.23	100.73
D71	K397	9.23	−	2.23	119.71
G73	E400	12.33	−	13.22	124.18
G84	E356	7.87	−	5.23	110.59
R85	E356	9.37	+++	3.43	128.29
Q104	E400	10.11	−	3.42	122.16
K116	D359	7.69	+	3.97	124.55
K149	R342	9.24	+++	3.06	124.57
D169	R338	14.36	−	6.98	118.68
T170	P367	8.36	−	1.83	108.58
E171	P367	7.33	+	2.36	115.46
E183	R339	9.42	−	4.26	98.33
R184	R339	10.75	+	4.11	104.10

[a]The interatomic distance involved in the cross-linking reaction is shown as the linear distance. [b]The imaginary phenylalanine at the indicated gankyrin residue. [c]Angle formed by the line from the Cζ atom to the Cγ atom of the imaginary phenylalanine and the line from the Cζ atom to the target carbon.

Crystallization, Data Collection, and Structure Determination. Crystallization of the cross-linked gankyrin (*p*Bpa85)−S6C complex was performed at 20 °C, by the sitting-drop vapor diffusion method. Small crystals of the complex were obtained with the precipitant containing 0.2 M $MgCl_2$, 30% polyethylene glycol 4000, and 0.1 M Tris-HCl (pH 8.5). To obtain larger crystals for X-ray analysis, the crystallization conditions were further optimized. The best crystals were obtained with a precipitant containing 31% polyethylene glycol 4000 (pH 8.5) and 0.26 M $MgCl_2$, at 20 °C. X-ray diffraction data sets were collected at beamlines BL26B2 and BL41XU (SPring-8, Hyogo, Japan). The data were integrated and scaled with HKL2000 (*16*). The structure of the gankyrin−S6C complex was determined by molecular replacement, employing the structure of the native gankyrin−S6C complex [Protein Data Bank (PDB) entry 2DVW] as a search model, using Molrep from the CCP4 suite (*17*). The model was refined with CNS (*18*), and structural modeling was performed using Coot, with careful inspection of the $2F_o - F_c$, $F_o - F_c$, and omitted electron density maps. The current refined model consists of 302 residues and 285 water molecules, with final R_{work} and R_{free} values of 17.1 and 22.9%, respectively, at 2.05 Å resolution. The Ramachandran plot of the gankyrin−S6C complex structure is excellent, with 97.3% of the residues in the favored region, 2.7% in the allowed region, and no residues in the outlier region, as determined with Rampage (*17*).

Examination of the pBpa-Targeted Amino Acids. The glutamate codon at position 356 of the S6C gene was mutagenized to an aspartate or methionine codon by using a QuikChange site-directed mutagenesis kit, to create E356D or E356M, respectively. The photo-cross-linked gankyrin (*p*Bpa85)−S6C (E356D) and gankyrin (*p*Bpa85)−S6C (E356M) complexes were expressed and purified as described above. To examine the target amino acid selectivity of *p*Bpa, we fractionated the cross-linked complex in each mutant by SDS−PAGE and stained it with CBB.

RESULTS

Identifying the Positions for Introducing pBpa into Gankyrin for Efficient Cross-Linking. For the crystallographic analysis of a cross-linked protein complex, it is necessary to identify the position(s) for *p*Bpa incorporation that will allow efficient cross-linking. The *E. coli* cell-based amber suppression system was employed to incorporate *p*Bpa into gankyrin. The amber suppression produces the full-length protein with *p*Bpa at the amber position and a truncated product, because the amber codon either directs the incorporation of *p*Bpa or terminates translation. Therefore, we introduced *p*Bpa into gankyrin, which can be produced in excess relative to S6C, when they are coexpressed in *E. coli* cells (*13*). We selected 16 amino acid residues in gankyrin to be replaced with *p*Bpa (Table 1), on the basis of the previously reported native complex structure of gankyrin and S6C (PDB entry 2DVW). These residues are on the surface of gankyrin, and their Cα atoms are either close to or reasonably oriented toward the nearest Cα atom of S6C, with Cα−Cα distances ranging from 7.01 to 14.36 Å (Table 1). Some residues were also included as putative negative controls: D71, which forms a hydrogen bond with K397 of S6C, was included to investigate the cross-linking efficiency at the protein−protein binding interface, and D169, which has its side chain oriented toward, but relatively distant from, the side chain of an S6C residue (i.e., K338), was tested to evaluate the distance limit of the cross-linking reaction (Table 1). Each of the 16 *p*Bpa-containing gankyrin variants was exposed, together with S6C, to 365 nm light for 30 min (Figures 1 and 2). However, some mutants (N7, Y15, and D169) were barely expressed, probably because of poor amber suppression in the sequence context, or the instability of the generated mutants. Western blotting detected the bands corresponding to the cross-linked complexes of the gankyrin variants and S6C, when *p*Bpa was substituted for R85, K116, K149, E171, and R184 (Figure 2). These residues do not directly interact with S6C residues in the complex structure, and all of the Cα−Cα distances were within a certain range (7.33−10.75 Å) (Table 1). The yield of the cross-linked product was the highest for the gankyrin variants with *p*Bpa at positions 85 and 149 (gankyrin *p*Bpa85 and gankyrin *p*Bpa149, respectively), although the proportion of S6C molecules involved in the cross-linked complex could not be quantitatively determined from the band intensities in the Western blotting pattern, which thus does not precisely reflect the relative amounts of the free and cross-linked molecules. These variants were used for the subsequent large-scale preparation of the highly purified, cross-linked complexes.

Maximization of the Yield of the Cross-Linked Products. Cell extracts were prepared from 2 L cultures of *E. coli* cells

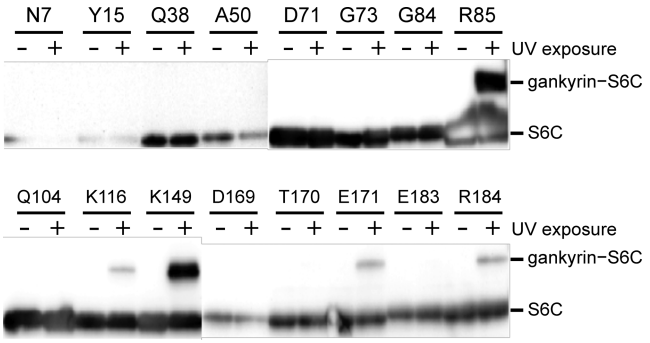

FIGURE 2: Screening the cross-linking efficiency. The cross-linking efficiency was analyzed by Western blotting, using a FLAG antibody. *p*-Benzoyl-L-phenylalanine (*p*Bpa) was incorporated at 16 different sites in gankyrin. The positions of the amber (Am) mutations in gankyrin are indicated at the top of the gel. Minus signs and plus signs indicate no irradiation and UV irradiation at 365 nm for 30 min, respectively. Note that the antibody reacted with the FLAG tag at the C-terminus of S6C, and that photo-cross-linked complexes were detected when residues R85, K116, K149, E171, and R184 of gankyrin were each substituted with *p*Bpa. Strong cross-linking efficiency was observed when the gankyrin proteins with the *p*Bpa substitutions at R85 (*p*Bpa85) and K149 (*p*Bpa149) were used.

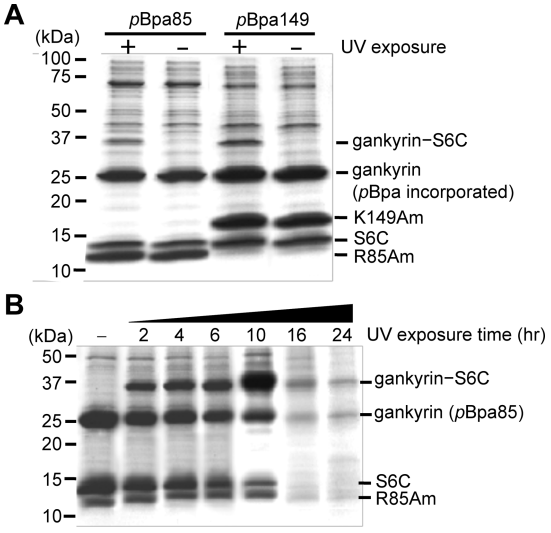

FIGURE 3: Optimization of the photo-cross-linking conditions. (A) SDS−PAGE profiles of affinity-purified gankyrin (*p*Bpa85)−S6C and gankyrin (*p*Bpa149)−S6C complexes. Minus signs and plus signs indicate no irradiation and UV irradiation at 365 nm for 30 min, respectively. The HisTrap HP column affinity-purified fractions were electrophoresed. The positions of the size markers are indicated at the left. The gel was stained with Coomassie Brilliant Blue R250. The molecular masses of the gankyrin−S6C cross-linked complex, the *p*Bpa-incorporated six-His gankyrin, and the FLAG-tagged S6C are 37.4, 26.7, and 10.8 kDa, respectively. The sizes of the truncated gankyrin products at the R85 and K149 amber codons are 10.6 and 17.5 kDa, respectively. Note that the 30 min UV irradiation at 365 nm generated a small, clear band of the cross-linked protein complex, indicated by gankyrin−S6C. (B) Optimization of the UV irradiation time for the cross-linking reaction. Note that the amount of the cross-linked complex (i.e., gankyrin−S6C) was maximal when the sample was UV-irradiated at 365 nm for 10 h. The amount of the cross-linked complex decreased when the sample was further UV-irradiated for 16 or 24 h.

coexpressing S6C and either gankyrin *p*Bpa85 or *p*Bpa149 and then were exposed to 365 nm light for 30 min. The cross-linked complexes were affinity-purified using nickel-chelating columns (Figure 3A). With the 30 min exposure, only small amounts of the cross-linked complexes were obtained, as shown on an SDS

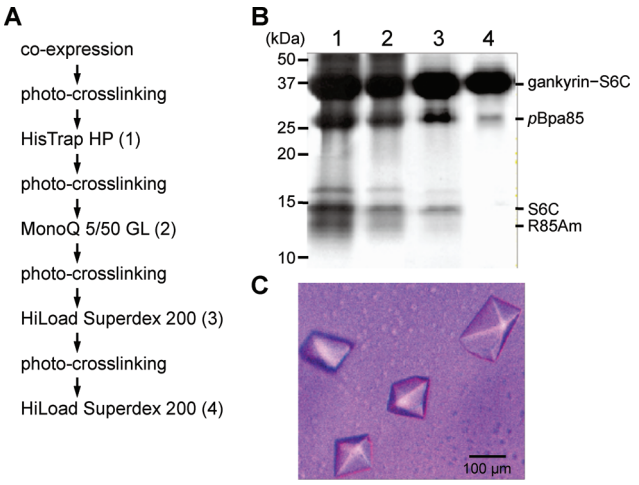

FIGURE 4: Purification and crystallization of the gankyrin (*p*Bpa85)−S6C cross-linked complex. (A) Purification scheme. The numbers in parentheses correspond to the lanes in panel B. (B) Purification of the gankyrin (*p*Bpa85)−S6C cross-linked complex. The purity of the cross-linked complex at each step was analyzed by SDS−PAGE. Gel filtration chromatography was performed twice (lanes 3 and 4). The final purity of the cross-linked complex is estimated to be >90%, although a small amount of unreacted gankyrin protein (i.e., *p*Bpa85) still remained in the final fraction (lane 4). (C) Crystallization of the purified gankyrin (*p*Bpa85)−S6C cross-linked complex. The best single crystals were obtained using precipitant conditions of 0.1 M Tris-HCl (pH 8.5) and 0.26 M $MgCl_2$, containing 31% polyethylene glycol 4000, by the hanging drop vapor diffusion method.

gel stained with CBB, whereas large amounts of noncovalently bound gankyrin and S6C remained in the extract, together with the truncated gankyrin molecules (Figure 3A); less than 5% of the affinity-purified proteins were cross-linked. Using gankyrin *p*Bpa85, we then examined if the yield of the cross-linked product could be increased by a longer duration of light exposure. The yield increased with light exposure for up to 10 h on ice and decreased with exposures longer than 16 h, making a smeared pattern (Figure 3B). Gankyrin *p*Bpa149 also displayed a similar time course-dependent cross-linking efficiency (data not shown). We thus set the duration of light exposure to 10 h for each step in the following experiments.

Purification of Gankyrin−S6C Cross-Linked Complexes. To obtain a large quantity of the purified gankyrin−S6C cross-linked complex for crystallization, optimizations of both the photo-cross-linking procedures and the column chromatography steps were required. In the previous section, we found that the cross-linking efficiency was ~50% after a 10 h UV irradiation, and a longer duration time damaged the cross-linked product, thus reducing its yield (Figure 3B). Although a 40 h UV illumination would theoretically produce a >90% yield of the cross-linked product, this was not possible, because of the radiation damage. Therefore, the UV illumination period was divided into four 10 h sections, each performed before the four following purification steps: (1) nickel-chelating affinity chromatography, (2) anion-exchange chromatography, (3) gel filtration chromatography, and (4) another step of gel filtration chromatography (Figure 4A). The noncovalent gankyrin−S6C complex does not dissociate during chromatography, thus making separate rounds of UV illumination feasible. We purified each of the cross-linked complexes, gankyrin *p*Bpa85−S6C and *p*Bpa149−S6C. The fractions containing the gankyrin *p*Bpa85−S6C cross-linked complex after the chromatography steps were analyzed, as shown in Figure 4B. The final purity of the cross-linked complex

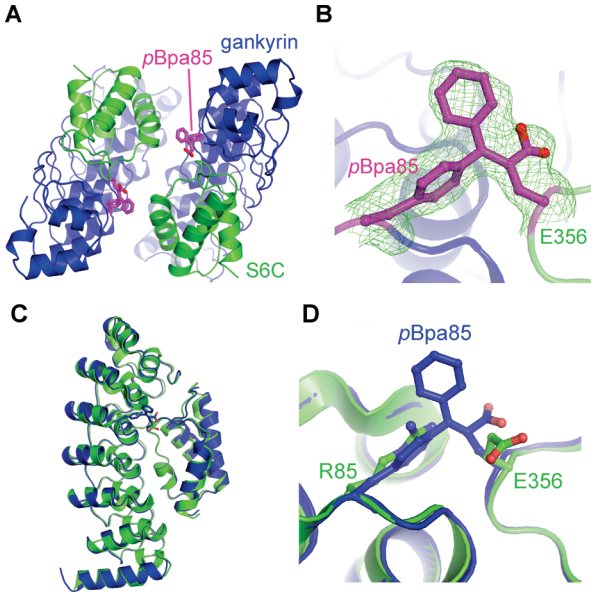

FIGURE 5: Structure of the gankyrin (*p*Bpa85)−S6C cross-linked complex. (A) Overall structure. Two complexes in the asymmetric unit of primitive rhombohedral space group *R*3 are shown. Gankyrin is colored blue, S6C green, and *p*Bpa magenta. (B) Close-up view of the cross-linked site. The $F_o - F_c$ omit map, contoured at 3.4σ, is colored green. *p*Bpa, cross-linked to the E356 side chain of S6C, is colored magenta. (C) Superposition of the overall structures of the cross-linked and native complexes. The cross-linked complex is colored blue and the native complex green. (D) Superposition of the cross-linked site. Colors are the same as in panel C.

was >90%, with an only marginal amount of free gankyrin *p*Bpa85 remaining (Figure 4B). The gankyrin *p*Bpa149−S6C cross-linked complex was also purified to a similar degree of homogeneity (data not shown).

Next, we tried to crystallize the two purified gankyrin−S6C photo-cross-linked complexes. Because the original native complex was crystallized in a precipitant containing 23% polyethylene glycol 6000 (pH 7.5) and 0.2 M $MgSO_4$, we initially tried similar crystallization conditions but could not obtain good single crystals (data not shown). We therefore newly screened the crystallization conditions and finally obtained the best single crystals of the *p*Bpa85−S6C cross-linked complex using a precipitant containing 31% polyethylene glycol 4000 (pH 8.5) and 0.26 M $MgCl_2$ (Figure 4C).

Crystal Structure of the Gankyrin pBpa85−S6C Cross-Linked Complex. Using the crystals of the gankyrin *p*Bpa85−S6C cross-linked complex, we collected X-ray diffraction data sets at BL26B2 and BL41XU at the SPring-8 synchrotron facility. Interestingly, the gankyrin *p*Bpa85−S6C cross-linked complex crystal was in primitive rhombohedral space group *R*3, with two complexes in the asymmetric unit, in contrast to that of the native complex crystal, which was in primitive orthorhombic space group $P2_12_12_1$, with one complex in the asymmetric unit (*13*). We determined the crystal structure of the gankyrin *p*Bpa85−S6C cross-linked complex at 2.05 Å by molecular replacement, using the native complex structure (PDB entry 2DVW) as a search model (Figure 5A). The refinement statistics are summarized in Table 2. Remarkably, the electron density corresponding to the cross-linked site was clearly detected in both complexes in the asymmetric unit (Figure 5B). From the $F_o - F_c$ omit map, the electron density unambiguously indicated that the *p*Bpa residue, which was substituted for the R85 residue in gankyrin *p*Bpa85, was uniquely bonded to the E356 side chain of S6C (Figure 5B).

Table 2: Data Collection and Refinement Statistics for the Gankyrin−S6C Cross-Linked Complex

Data Collection	
space group	*R*3
cell dimensions	
a, *b*, *c* (Å)	103.16, 103.16, 154.99
α, *β*, *γ* (deg)	90, 90, 120
resolution (Å)	50.0−2.05 (2.18−2.05)[a]
R_{sym} or R_{merge}	9.7 (49.6)
$I/I\sigma$	19.8 (2.93)
completeness (%)	100.0 (100.0)
redundancy	4.2 (4.2)
Refinement	
resolution (Å)	42.93−2.05
no. of reflections	38604
R_{work}/R_{free}	17.1/22.9
no. of atoms	4996
protein	4636
ligand/ion[b]	34
water	326
B factor	
protein	30.63
ligand/ion[b]	51.28
water	41.92
root-mean-square deviation	
bond lengths (Å)	0.032
bond angles (deg)	2.8

[a]Values in parentheses are for the highest-resolution shell. [b]*p*Bpa is counted as a ligand.

When the benzophenone photophore in *p*Bpa absorbs a photon at 365 nm, the carbonyl group oxygen of *p*Bpa becomes electrophilic and interacts with weak C−H bonds (*3*). The E356 side chain of S6C contains two carbons that covalently bond to hydrogen, at C*β* and C*γ*. In the cross-linked site of the complex, we observed two protrusions of electron density: a slightly larger one that corresponded well to the phenyl group of *p*Bpa and a slightly smaller one, which is thought to correspond to the carboxyl group of the E356 side chain (Figure 5B). From the position of this carboxyl group, the cross-linked site in the E356 side chain of S6C was sterically determined to be C*γ* (Figure 5B). The electron density showed no other form of bonding between S6C and gankyrin. Therefore, on the basis of the 2.05 Å resolution data, we conclude that the carbonyl group of *p*Bpa85 in gankyrin bound covalently to the C*γ* atom of E356 in S6C, and the cross-linking bridged by *p*Bpa occurred uniquely and site-specifically (Figure 5B).

Importantly, the photo-cross-linked structure exhibited minimal distortion (Figure 5C), as compared to the native gankyrin−S6C complex structure, with root-mean-square deviations of 0.5 Å for gankyrin and 0.6 Å for S6C. A slight structural difference was observed at the cross-linked site (Figure 5D), in which the C*γ* atom in the side chain of E356 in S6C was relatively straighter, to form the covalent bond with *p*Bpa85 in gankyrin.

Proposed Model for the Photo-Cross-Linking Reaction between Gankyrin pBpa85 and S6C. To demonstrate the important role of the C*γ* atom in the side chain of E356 in S6C, E356 was mutated to either aspartate (E356D) or methionine (E356M), and the formation of cross-linking was tested. S6C E356D could not be cross-linked with gankyrin *p*Bpa85, whereas S6C E356M was efficiently cross-linked (Figure 6A). This result clearly indicates that *p*Bpa targeted the C*γ* atom of the methylene

group of E356 or M356 and could not react with either the Cγ atom of the carboxyl group or the Cβ atom of the methylene group of D356, as shown in Figure 6B. On the basis of this finding and the previously proposed reaction scheme (*3*), we propose a cross-linking reaction mechanism (Figure 7). In this model, the final adduct should form a double bond at Cγ of E356 in S6C, which is consistent with the observation that the six carbon atoms involved in or next to this double bond were coplanar in the $F_o - F_c$ omit map (Figure 5B). *p*Bpa is likely to react with the C−H bond of an amino acid residue and thus can potentially covalently bond with any residue through its main chain Cα atom (*3*). However, gankyrin *p*Bpa85 formed a unique, covalent bond with only the Cγ atom in the side chain of the targeted residue, not with the C−H bond at the Cβ position of D356 (Figure 6B), showing that both the distance from and the orientation toward the C−H bond in the protein complex are crucial factors for *p*Bpa to form the cross-link.

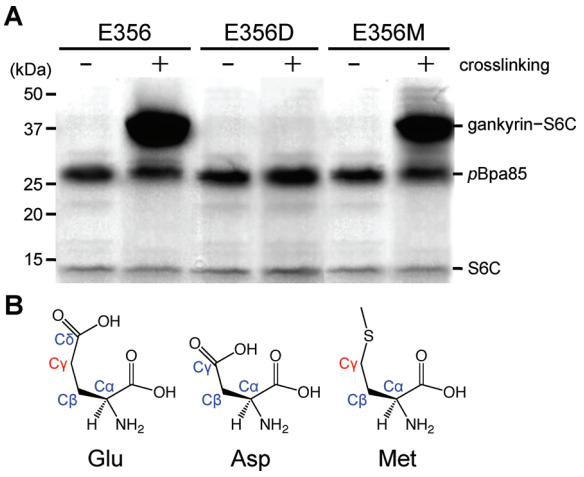

FIGURE 6: Cross-linking reactivity of *p*Bpa in gankyrin to E356 of S6C. (A) Cross-linking reactions using S6 mutants. E356 of S6C, which reacts with *p*Bpa in gankyrin, was replaced with aspartate (E356D) or methionine (E356M) for side chain comparisons. (B) Structures of the three amino acids used in the mutant assay. Carbon atoms possibly involved in the cross-linking reaction are colored red, while other side chain carbons are colored blue.

Distances between Gankyrin and S6C Residues Allowing Photo-Cross-Linking. Among the 16 tested *p*Bpa positions in gankyrin, five positions (85, 116, 149, 171, and 184) allowed cross-linking with S6C. To identify the distance requirement for cross-link formation, we built structural models of the native gankyrin−S6C complex, in which each residue was replaced with phenylalanine as a mimic of *p*Bpa, and determined the distance between the Cζ atom of these phenylalanine residues and the carbon atom of the nearest C−H bond in the bound S6C (the putative target carbon) (Table 1). The orientation of the side chain of Phe was adjusted to place the Cζ atom closest to the nearby carbon, without compromising the $\varphi-\psi$ angle. These analyses revealed that the allowed distances for cross-linking span a narrow range, from 2.36 to 4.11 Å, and that the Cζ−Cγ line of Phe and the line connecting the Cζ atom and the putative target carbon form an angle ranging from 104.10° to 128.29° (Table 1). Because the distance between the *p*Bpa carbon corresponding to Cζ of Phe and the oxygen of the carbonyl group is approximately 2.4 Å, the photo-cross-linking occurred when the reactive carbonyl oxygen of *p*Bpa was located within 1.7 Å of the targeted S6C carbon. The only exception is position 104, where *p*Bpa did not form a cross-link, although the Cζ atom of the Phe placed at this position is within cross-linking distance of the nearest carbon in E400 of S6C (Table 1). This exception probably existed because Q104 of gankyrin forms a hydrogen bond to E400 of S6C in the native complex, and the substitution of *p*Bpa for the residue either weakened or prevented the interaction between gankyrin and S6C.

DISCUSSION

This study revealed the first crystal structure of a protein complex cross-linked by a photoreactive group, which was site-specifically incorporated by genetic code expansion. The overall structure of the gankyrin−S6C complex, cross-linked by *p*Bpa, showed little distortion from the native complex structure (Figure 5C,D), although the crystals of these two complexes happened to belong to different space groups. The root-mean-square deviations between the two complex structures (0.5 and 0.6 Å for the gankyrin and S6C moieties, respectively) are mostly attributable to local structural changes at the cross-linking site

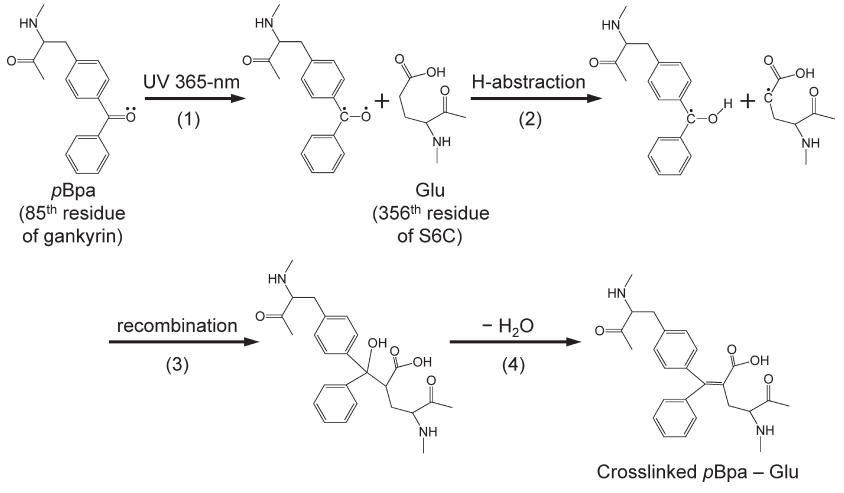

FIGURE 7: Model for the cross-linking reaction between *p*Bpa in gankyrin and glutamate 356 of S6C. The absorption of a photon at 365 nm results in the promotion of one electron from a nonbonding sp^2-like n-orbital on the carbonyl oxygen in *p*Bpa (1). The electron-deficient n-orbital of the carbonyl oxygen becomes electrophilic and interacts with the weak C−H bond of the neighboring glutamate 356 of S6C. Consequently, a hydrogen is abstracted to complete the half-filled n-orbital (2). The ketyl radical of *p*Bpa and the alkyl radical of glutamate 356 recombine to generate a new C−C bond, yielding a benzopinacol-type compound (3). Finally, the cross-linking reaction is completed by dehydration (4).

(Figure 5D). This finding illustrates that the site-specific cross-link formation preserves the proper binding mode, thus underscoring the usefulness of *p*Bpa for probing protein−protein interactions, when it is site-specifically incorporated into proteins. In addition, the *p*Bpa in gankyrin exhibited high selectivity in targeting a carbon atom in the binding protein and formed a covalent bond exclusively with the Cγ atom of E356 in S6C. This selectivity is not due to a preference for glutamate, because gankyrin *p*Bpa85 also reacted efficiently with the Met356 variant of S6C, consistent with previous studies reporting the reactivity of methionine with *p*Bpa (*19*, *20*). A model building study was performed to interpret the results of the cross-linking experiments with the 16 gankyrin variants and revealed that both the distance and geometric relationship between the carbonyl group of *p*Bpa and the reactive C−H bond were crucial factors. The two strongly cross-linked positions (i.e., positions 85 and 149) and the three weakly cross-linked positions (i.e., positions 116, 171, and 184) exhibited narrow ranges for the distance (i.e., 2.36−4.11 Å) and the angle (i.e., 104.10−128.29°) between *p*Bpa and the targeted C−H bond (Table 1). These strict requirements prevent *p*Bpa from bonding with various residues in the binding protein and from generating structural heterogeneity within the cross-linked complex. In addition, the side chains of each of the five cross-linked residues were not directly involved in the protein−protein interaction with each of the potential S6C residue targets. When a residue within the protein−protein binding interface, such as D71, was substituted with *p*Bpa, it could not cross-link with the partner protein, even though the carbonyl group of *p*Bpa and a potentially reactive C−H bond in S6C were very close (i.e., 2.23 Å). Because the substitution with the cross-linker should thus not disturb the native noncovalent bonding, a residue existing at the periphery of the protein−protein binding interface, in which the side chain is distant from a potential C−H bond of the partner molecule, at a distance relative to the reach of the cross-linker, would be a good candidate for cross-linking.

Genetic code expansion has enabled the site-specific incorporation of a variety of non-natural amino acids into proteins for various purposes. For structural studies, iodinated derivatives of L-phenylalanine and L-tyrosine have been incorporated into proteins, to facilitate single-wavelength anomalous dispersion (SAD) phasing for X-ray crystallography (*21*, *22*). This study demonstrated the feasibility of another application of genetic code expansion for protein structural studies. In this study, we described a procedure, including separate rounds of UV illumination, for preparing a sufficient amount of a cross-linked protein complex for crystallization. The *p*Bpa was remarkably stable during the preparation process; it maintained the reactivity during a 10 h exposure to UV light, and the yield of the cross-linked product steadily increased over this period. Although UV illumination durations of >10 h damaged the cross-linked product, we were able to prepare more than 90% pure cross-linked product with the current procedure, in which four separate 10 h UV illumination steps were included between the column chromatography steps (Figure 4A). This procedure was practically effective for several reasons: the total protein preparation time was reduced, by cross-linking the sample purified during the day at night, and the purity and quantity of the sample could be carefully monitored during the preparation. Proteins containing *p*Bpa can be produced in a variety of host systems, including *E. coli*, yeast, and mammalian cells (*8*−*11*), and probably also in insect cells (*23*), enhancing the utility of this cross-linkable amino acid.

With these advantages, the immortalization of a protein complex by cross-linking may be applied to other X-ray crystallographic studies. Many protein complexes involve unstable or transient interactions between components, making it difficult to prepare these complexes in an intact form and in a sufficient amount for crystallography. To apply our approach to these cases, we must determine the appropriate positions for *p*Bpa incorporation that will achieve efficient cross-linking. On the basis of the structural information about a protein component in the free form, the residues on the protein surface or possibly interacting with the other protein may be identified and replaced with *p*Bpa, to achieve cross-linking. Without any structural information, the cross-linking must be attempted via replacement of many of the residues on the predicted protein surface separately with a cross-linkable amino acid. In such an attempt to identify appropriate incorporation sites in the proximity of the binding interface, *p*Bpa, with its relatively "short reach", may not be useful, and a non-natural amino acid with a long side chain and a photoreactive group near the end of the chain should be developed. A lysine derivative, as a candidate for such a "long amino acid", has been incorporated into proteins by engineering pyrrolysyl-tRNA synthetase (*24*). After the appropriate sites have been located, *p*Bpa may be incorporated, to identify the site with efficient and unique cross-linking. Thus, the site-specific photo-cross-linking method will facilitate structural studies of proteins, not only by stabilizing the protein complex but also by possibly defining the states of a complex that switches internally during its function.

ACKNOWLEDGMENT

We thank Drs. Naoki Umezawa (Nagoya City University, Nagoya, Japan) and Yoshihiro Nakamura for structural discussions, Dr. Yoshihide Hayashizaki (RIKEN OSC) for the FANTOM cDNA, the BL26B2 and BL41XU beamline staffs at SPring-8 for data collection, and Ms. Tomoko Nakayama and Ms. Azusa Ishii for clerical assistance.

REFERENCES

1. Bayley, H., and Knowles, J. R. (1977) Photoaffinity labeling. *Methods Enzymol. 46*, 69–114.
2. Chowdhry, V., and Westheimer, F. H. (1979) Photoaffinity labeling of biological systems. *Annu. Rev. Biochem. 48*, 293–325.
3. Dormán, G., and Prestwich, G. D. (1994) Benzophenone photophores in biochemistry. *Biochemistry 33*, 5661–5673.
4. Xie, J., and Schultz, P. G. (2006) A chemical toolkit for proteins: An expanded genetic code. *Nat. Rev. Mol. Cell Biol. 7*, 775–782.
5. Tanaka, Y., Bond, M. R., and Kohler, J. J. (2008) Photocrosslinkers illuminate interactions in living cells. *Mol. BioSyst. 4*, 473–480.
6. Kauer, J. C., Erickson-Viitanen, S., Wolfe, H. R., Jr., and DeGrado, W. F. (1986) *p*-Benzoyl-L-phenylalanine, a new photoreactive amino acid. Photolabeling of calmodulin with a synthetic calmodulin-binding peptide. *J. Biol. Chem. 261*, 10695–10700.
7. Brunner, J. (1993) New photolabeling and crosslinking methods. *Annu. Rev. Biochem. 62*, 483–514.
8. Chin, J. W., Martin, A. B., King, D. S., Wang, L., and Schulz, P. G. (2002) Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A. 99*, 11020–11024.
9. Hino, N., Okazaki, Y., Kobayashi, T., Hayashi, A., Sakamoto, K., and Yokoyama, S. (2005) Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat. Methods 2*, 201–206.
10. Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z., and Schultz, P. G. (2003) An expanded eukaryotic genetic code. *Science 301*, 964–967.
11. Liu, W., Brock, A., Chen, S., Chen, S., and Schultz, P. G. (2007) Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods 4*, 239–244.

12. Huang, L. Y., Umanah, G., Hauser, M., Son, C., Arshava, B., Naider, F., and Becker, J. M. (2008) Unnatural amino acid replacement in a yeast G protein-coupled receptor in its native environment. *Biochemistry 47*, 5638–5648.
13. Nakamura, Y., Nakano, K., Umehara, T., Kimura, M., Hayashizaki, Y., Tanaka, A., Horikoshi, M., Padmanabhan, B., and Yokoyama, S. (2007) Structure of the oncoprotein gankyrin in complex with S6 ATPase of the 26S proteasome. *Structure 15*, 179–189.
14. Kobayashi, T., Nureki, O., Ishitani, R., Yaremchuk, A., Tukalo, M., Cusack, S., Sakamoto, K., and Yokoyama, S. (2003) Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat. Struct. Biol. 10*, 425–432.
15. Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the genetic code of *Escherichia coli*. *Science 292*, 498–500.
16. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276*, 307–326.
17. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D50*, 760–763.
18. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D54*, 905–921.
19. Kage, R., Leeman, S. E., Krause, J. E., Costello, C. E., and Boyd, N. D. (1996) Identification of methionine as the site of covalent attachment of a *p*-benzoyl-phenylalanine-containing analogue of substance P on the substance P (NK-1) receptor. *J. Biol. Chem. 271*, 25797–25800.
20. Pérodin, J., Deraët, M., Auger-Messier, M., Boucard, A. A., Rihakova, L., Beaulieu, M. E., Lavigne, P., Parent, J. L., Guillemette, G., Leduc, R., and Escher, E. (2002) Residues 293 and 294 are ligand contact points of the human angiotensin type 1 receptor. *Biochemistry 41*, 14348–14356.
21. Xie, J., Wang, L., Wu, N., Brock, A., Spraggon, G., and Schultz, P. G. (2004) The site-specific incorporation of *p*-iodo-L-phenylalanine into proteins for structure determination. *Nat. Biotechnol. 22*, 1297–1301.
22. Sakamoto, K., Murayama, K., Oki, K., Iraha, F., Kato-Murayama, M., Takahashi, M., Ohtake, K., Kobayashi, T., Kuramitsu, S., Shirouzu, M., and Yokoyama, S. (2009) Genetic encoding of 3-iodo-L-tyrosine in *Escherichia coli* for single-wavelength anomalous dispersion phasing in protein crystallography. *Structure 17*, 335–344.
23. Mukai, T., Wakiyama, M., Sakamoto, K., and Yokoyama, S. (2010) Genetic encoding of non-natural amino acids in *Drosophila melanogaster* Schneider 2 cells. *Protein Sci. 19*, 440–448.
24. Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K., and Yokoyama, S. (2008) Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N^{ε}-(*o*-azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem. Biol. 15*, 1187–1197.