Cyclic Green Fluorescent Protein Produced *in Vivo* Using an Artificially Split PI-*Pfu*I Intein from *Pyrococcus furiosus**

Hideo Iwai, Andreas Lingel, and Andreas Plückthun‡

From the Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

A cyclic protein was produced *in vivo* using the intein from *Pyrococcus furiosus* PI-*Pfu*I in a novel approach to create a circular permutation of the precursor protein by introducing new termini in the intein domain. Green fluorescent protein (GFP) was cyclized with this method *in vivo* on milligram scales. There was no by-product of linear or polymerized species isolated, unlike with other *in vitro* or *in vivo* cyclization methods utilizing inteins. Cyclized GFP unfolded at half the rate of the linear form upon chemical denaturation and required >2 days in 7 M guanidine hydrochloride until a residual fast folding phase (consistent with a persistent *cis*-proline) had disappeared. Cyclic GFP might become a novel tool for studying the role of termini and backbone topology in various biological processes such as protein degradation and translocation *in vivo* as well as *in vitro*.

All natural proteins known so far are linear chains of amino acids that fold into a unique three-dimensional structure dictated by the sequence of amino acids. A cyclic backbone structure has been found and synthetically introduced in small peptides, but is an almost unexplored topic in protein chemistry and protein structural research, in particular for larger proteins. The circular topology is also expected to lead to improved stability due to the reduced conformational entropy in the denatured state, according to polymer theory (1). It has indeed been shown experimentally that cyclized β-lactamase was stabilized against heat precipitation and exopeptidase degradation (2) and that the cyclization also improved *in vitro* thermal stability of dihydrofolate reductase (3). Although there was no significant stabilization observed in the pioneering work on a cyclic bovine pancreatic trypsin inhibitor prepared by chemical modification, several effects may have canceled out in this example (4).

A recent development in protein chemistry, the use of self-splicing proteins (often called inteins), has opened a general avenue to create a circular backbone topology as well as to ligate proteins and peptides *in vitro*. This procedure has been called expressed protein ligation or intein-mediated protein ligation (Refs. 5 and 6 and, for review, see Ref. 7). In this approach, an intein with an asparagine-to-alanine mutation in the active site is fused to one fusion partner. This mutation stops the enzymatic reaction at the stage of a C-terminal thio-ester of the fusion partner, selectively cleavable by thiols. The second peptide carrying an N-terminal cysteine acts as an *S*-nucleophile at the thioester group, forming a new peptide bond after S–N rearrangement. However, this approach has the disadvantage that it requires the nucleophilic thiol group of cysteine at the N terminus of one partner as well as a C-terminal thioester modification of the other partner, whose formation is catalyzed by the intein (5, 6). This reaction can be used to make cyclic peptides and proteins, but the intramolecular cyclization reaction always has to compete with other intermolecular reactions such as polymerization and hydrolysis of the thioester group (2, 8, 9), which reduce the cyclization efficiency and complicate the purification procedure (2).

An *in vivo* cyclization without by-products would be of importance for the large-scale production of cyclic proteins and peptides *in vivo*. This may also provide the possibility for creating a large library of cyclic peptides and proteins *in vivo* for functional selections as well as for biophysical characterization of cyclic proteins. Two independent groups have demonstrated that the naturally split intein DnaE from *Synechocystis* sp. PCC6803 (*Ssp* DnaE) can be used for cyclization of proteins and peptides *in vivo* by arranging the naturally separated two fragments of the intein, DnaE$_N$ (the N-terminal part) and DnaE$_C$ (the C-terminal part), in the order DnaE$_C$-extein-DnaE$_N$ (3, 10). In theory, this strategy should be feasible not only with the naturally split intein, but also with any intein by artificially creating a similar arrangement. This corresponds essentially to a circular permutation of a precursor protein that introduces new N and C termini into the intein domain. The end products should include a circular protein if the precursor protein folds into a functional structure (see Fig. 1).

In this report, we have tested our idea of cyclization by circular permutation of precursor proteins within intein domains using the intein PI-*Pfu*I[1] from *Pyrococcus furiosus*. We chose the green fluorescent protein (GFP) as our cyclization target because cyclic GFP might become a useful tool for studies on the roles of N and C termini in cellular environments.

EXPERIMENTAL PROCEDURES

Plasmid Construction for Circular GFPuv Expression—The C-terminal part (PI-*Pfu*I$_C$, residues 161–454) of the intein PI-*Pfu*I was amplified from pHisInα (11) using oligonucleotides 5′-CTTTAAGAAGGAGA-TATA-3′ and 5′-CTCAGTAAGAGATTTTTTTCTAGATCCGGTGTTGT-GGACG-3′, followed by digestion with *Nco*I and *Xba*I. The N-terminal part (PI-*Pfu*I$_N$, residues 1–160) of the intein was amplified from pCBDαIn (11) using oligonucleotides 5′-GTTCGGTACCGGATGCATA-GACGGAAAGG-3′ and 5′-TGGAAGCTTACTTAACATGTGAGTGG-3′,

[1] The abbreviations used are: PI, protein insert; GFP, green fluorescent protein; GFPuv, improved green fluorescent protein with high solubility; NTA, nitrilotriacetic acid; DTT, dithiothreitol; GdnHCl, guanidine hydrochloride; PcP, polymerase chain reaction.

followed by digestion with *Kpn*I and *Hin*dIII. The plasmids pHisInα and pCBDαIn were kindly provided by Dr. T. Yamazaki (Osaka University). The gene of GFP, improved by DNA shuffling (GFPuv, also called the Cycle3 mutant) and bearing the mutations F99S, M153T, and V163A (12), was obtained from pBAD/AC2 by polymerase chain reaction (PCR) amplification with primers 5′-CCTCTAGACATCATCACCACCATCAC-TCTAGAAAAGGAGAAGAACTYTTC-3′ and 5′-TAGGTACCGCGTGG-CACCAACCCAGCAGCWGTTAC-3′. The polymerase chain reaction product was digested with *Xba*I and *Kpn*I. These three PCR products were then ligated to form a fusion protein under the control of the T7 promoter in the expression vector pTFT74 (13) between the *Nco*I and *Hin*dIII sites in a stepwise manner, yielding the plasmid pIWT5563his.

Expression and Purification—For protein expression, *Escherichia coli* strain BL21-CodonPlus(DE3)-RIL (Stratagene) was transformed with plasmid pIWT5563his. The use of this strain, which overexpresses *E. coli* tRNAs for rare codons, was necessary due to the large number of rare codons in the PI-*Pfu*I gene. The cells were grown at 30 °C in 1 liter of LB medium containing 50 μg/ml ampicillin and 30 μg/ml chloramphenicol until $A_{550} \sim 0.5$ was reached. The protein was then expressed at a slow rate by incubation at room temperature (~25 °C) overnight without induction, taking advantage of the leaky expression vector, followed by addition of isopropyl-β-D-thiogalactopyranoside to a final concentration of 0.1 mM and incubation for another 6 h. The cells were harvested by centrifugation at $4000 \times g$ for 10 min at 4 °C. The cell pellets, which showed bright green fluorescence, were resuspended in 50 mM sodium phosphate buffer (pH 8.0) and 300 mM NaCl. The cells were lysed by incubation with lysozyme (1 mg/ml) for 30 min and by sonication on ice. The cell lysate was loaded on a 10-ml Ni²⁺-NTA column (QIAGEN Inc.) after centrifugation at $20,000 \times g$ for 1 h at 4 °C. The column was washed with 50 mM sodium phosphate (pH 8.0), 300 mM NaCl, and 20 mM imidazole. The hexahistidine-tagged proteins including GFPuv were eluted with 300 mM imidazole (see Fig. 3*B*, *lane 2*). The eluted sample was further purified by organic extraction as described by Yakhnin *et al.* (14). Briefly, the sample was saturated with ammonium sulfate and extracted with ethanol. The aqueous phase was subsequently removed from the organic phase after addition of 1-butanol. The extracted aqueous fraction contained only a single band of ~28 kDa (see Fig. 3*B*, *lane 3*). The linear form of GFPuv was prepared by digestion of the circular form with thrombin. For this purpose, 1.5 ml of 10 μM cyclic GFPuv was incubated with 0.2 units of thrombin (Roche Molecular Biochemicals) for 2.5 h at room temperature, followed by purification on a 10-ml Ni²⁺-NTA column. The single band of the purified linear protein migrated more slowly than the circular form of GFPuv (see Fig. 3*B*, *lane 4*).

To examine the splicing reaction *in vivo*, the *E. coli* cells containing plasmid pIWT5563his were grown overnight at room temperature in LB medium, followed by induction with a final concentration of 0.1 mM isopropyl-β-D-thiogalactopyranoside for 4 h, and then either directly mixed with SDS loading buffer and boiled at 95 °C for 5 min or alternatively first sonicated and then resuspended in SDS loading buffer. The cell lysates were separated by 12% SDS-polyacrylamide gel electrophoresis. The proteins were then transferred to a polyvinylidene difluoride membrane for Western blot analysis. Western blotting was performed using the Tetra·His Antibody™ (QIAGEN Inc.) and an anti-GFP antibody (Roche Diagnostics) and visualized by colorimetric detection using alkaline phosphatase-conjugated anti-mouse IgG.

The structure of cyclic GFP was modeled based on the crystal structure of GFP (Protein Data Bank code 1EMC) (15) as follows. The modeled structure was obtained by simulated annealing using artificial distance constraints created from the crystal structure based on hydrogen bonds as well as proton-proton distances within 5 Å. The distance constraints between hydrogen atoms for the model building were artificially created as upper distance constraints based on proton-proton distances of <5 Å of all the hydrogen atoms calculated in the crystal structure with the program MOLMOL (16). All the methyl and methylene groups were treated as pseudo atoms. The hydrogen bonds found in the crystal structure were used as upper and lower distance constraints for the structure calculation. The chromophore structure was treated as the original unmodified amino acids during the calculation. The covalent bond between the N and C termini was introduced as upper and lower distance constraints. For the linker connecting both termini, there was no constraint applied besides the covalent bond connecting the termini. With a total of the 9188 distance constraints, obtained from both proton-proton distances and hydrogen bonds, three rounds of simulated annealing calculation were performed to obtain the model structure depicted in Fig. 2*B*. All calculations were performed
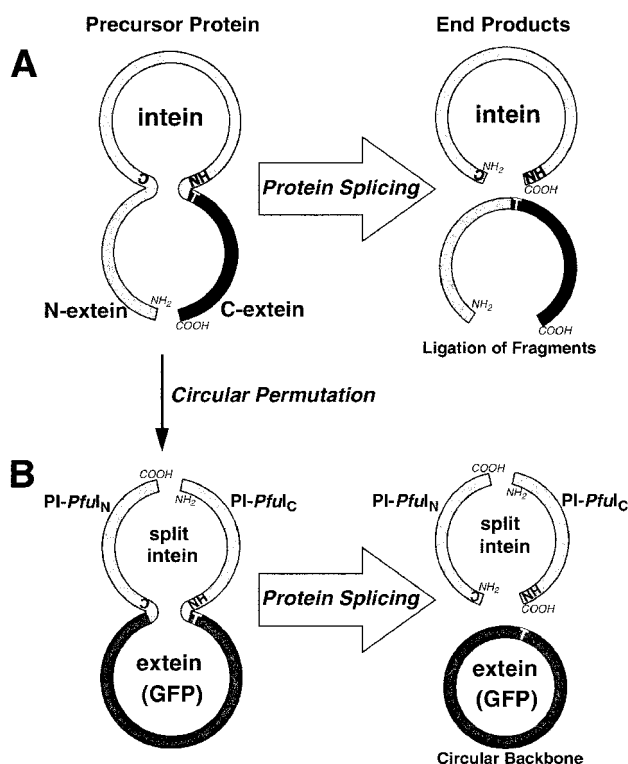

FIG. 1. **Schematic representation of the circular permutation approach for backbone cyclization.** *A*, ligation of two extein domains by protein splicing; *B*, circular permutation of the precursor protein containing an intein domain. Protein splicing results in three fragments, including an extein with a circular peptide backbone.

with the program DYANA (17) on a SGI Octane.

Circular Dichroism Measurements—CD wavelength scans were recorded with a Jasco J-715 spectropolarimeter. The protein samples were prepared in 100 mM potassium phosphate (pH 8.0), 50 mM NaCl, and 1 mM EDTA; the protein concentrations were 0.12 mg/ml (linear) and 0.15 mg/ml (circular). All spectra were recorded at 25 °C with 10 scans. The data were normalized to molar ellipticity with a path length of 0.1 cm.

Fluorescence Spectroscopy of the Cyclized and Linear Forms of GFPuv—All fluorescence spectra were recorded with a PTI Alpha Scan spectrofluorometer at 25 °C. A cuvette with 1-cm path length was used. The samples for the emission and excitation scans were prepared in 50 mM potassium phosphate (pH 6.5, 7, 7.5, or 8), 1 mM EDTA, and 1 mM DTT or in 50 mM Tris (pH 8.5), 1 mM EDTA and 1 mM DTT. The final protein concentration of all the samples was 1 μg/ml. For the excitation spectrum, the fluorescence intensity was monitored at 508 nm, whereas the emission spectrum was recorded with an excitation at 397 nm.

Unfolding and Refolding Kinetic Measurements—The kinetic unfolding and refolding experiments were performed in solutions containing 50 mM potassium phosphate (pH 8), 50 mM sodium chloride, 1 mM EDTA, 1 mM DTT, and various concentration of GdnHCl. In the unfolding experiments, the protein solution was diluted 1:320 in the unfolding solution containing GdnHCl, resulting in a final protein concentration of 0.5 μg/ml at 25 °C. The change in the chromophore fluorescence intensity was fitted with a single exponential, $A \cdot \exp(-k \cdot t)$, where A, k, and t are amplitude, kinetic constant, and time, respectively, using SigmaPlot software (SPSS Inc., Chicago, IL). This equation assumes no residual fluorescence at infinite time. In the refolding experiments, the proteins were first denatured in 6 M GdnHCl for 3 h at room temperature (see Fig. 6*B*) (18, 19) or denatured in 7 M GdnHCl for various times at room temperature (see Fig. 7), followed by 1:25–40 dilution in the refolding buffer. The final protein concentration was ~0.5 μg/ml for all measurements. The change in the chromophore fluorescence intensity was fitted with a double exponential, $A \cdot (1 - \exp(-k_1 \cdot t)) + B \cdot (1 - \exp(-k_2 \cdot t)) + C$, where k_1 and k_2 are the kinetic constants for the fast and the slow phases, respectively. The fraction of the fast phase in Fig. 7*C* was determined by calculating $A/(A + B)$ from the fitted values. The concentrations of GdnHCl were determined after the kinetic measurements by measuring the refractive index.

FIG. 2. **The construct for *in vivo* cyclization of GFPuv.** *A*, the detailed construct of the precursor protein containing an artificially split intein (PI-*Pfu*I) and an extein (GFPuv). The original amino acid sequence of GFPuv is *underlined*. The hexahistidine tag in front of GFPuv (in *boldface*) and the thrombin recognition site are *shaded*. The thrombin cleavage site is indicated by an *arrowhead*. The original terminal residue numbers of GFPuv, between which three-dimensional structure is visible, are labeled on top of the primary sequence. The first N-terminal residue of the extein domain is in *italics*. *B*, model of cyclized GFPuv from the crystal structure of a GFP calculated by the program DYANA using artificial constraints obtained from the x-ray structure (Protein Data Bank code 1EMC) (see "Experimental Procedures"). The original N- and C-terminal locations are indicated by *N* and *C*, respectively. The first N-terminal residue of the extein part is displayed in a ball-and-stick model. The chromophore is shown in a Corey-Pauling-Koltun model. The figure was created with the program MOLMOL (16).
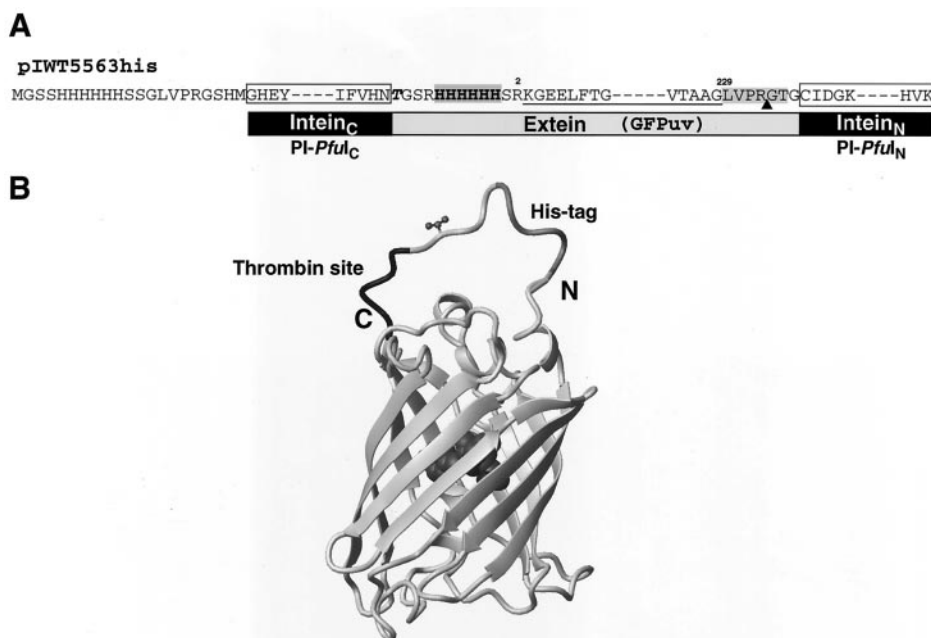
RESULTS

Design of the Construct for in Vivo Cyclization—Our concept to produce a circular protein *in vivo* is simple. The conventional and natural protein splicing activity of an intein is to connect the C terminus of the N-terminal fragment (N-extein) and the N terminus of the C-terminal fragment (C-extein) (Fig. 1*A*). These two interacting termini can also be within one extein domain. Topologically, this corresponds to joining the original N and C termini of the extein domains and, at the same time, introducing new termini into the intein domain (Fig. 1*B*). This is nothing but a circular permutation of the precursor protein.

After the protein splicing reaction, the precursor protein gives rise to three protein fragments, *viz.* a cyclized extein and the N- and C-terminal parts of the intein. For our study, we chose the intein PI-*Pfu*I from *P. furiosus* because it has been demonstrated that this intein can be artificially split in the extended loop between residues 160 and 161 to perform *trans*-ligations *in vitro* (11, 20, 21).

As the target for cyclization (*i.e.* the extein), we chose the improved green fluorescent protein GFPuv (also called Cycle3 mutant) (12) since cyclic GFP might constitute a new tool for studying the roles of N or C termini or protein backbone topology in cellular processes such as protein degradation and translocation *in vivo*. Moreover, the properly folded end product can be conveniently monitored by its green fluorescence. Furthermore, GFP had been successfully circularly permuted by joining the N terminus (residue 1) and the C terminus (residue 238) with a linker of 6 amino acids (22, 23), suggesting that a new connecting loop would likely lead to functional GFP.

The crystal structure of GFP indicates that this protein has a cylinder-like β-barrel structure containing the chromophore in the distorted central helix surrounded by the β-barrel (Fig. 2*B*) (15, 24, 25). The visible N and C termini are not in very close proximity (~19 Å between residues 2 and 229), but are located on the same side of the cylinder. The N-terminal residue and the C-terminal fragment after residue 229 in the crystal structures of GFP are often invisible and probably disordered (15, 24, 25). The original sequence of GFPuv from residues 3 to 228 is unchanged in our construct (Fig. 2*A*, *underlined*), and a hexahistidine tag was added in front of the GFPuv gene, whereas a thrombin-specific proteolytic site (LVPRGT, similar to the thrombin-specific sequence LVPRGS,

with the Thr residue being introduced to generate a *Kpn*I site) was introduced behind the GFPuv gene (Fig. 2*A*). The hexahistidine tag and its flanking sequence together constitute an 11-residue linker between the N-terminal intein domain (PI-*Pfu*I$_C$) and the second residue of GFPuv. The thrombin recognition sequence is part of a 6-residue linker between Leu[229] of GFPuv and the C-terminal intein domain (PI-*Pfu*I$_N$).

Fig. 2*B* provides an overview of the modeled cyclic GFPuv structure, and the first residue (threonine) after the N-terminal intein domain (PI-*Pfu*I$_C$) is depicted as a ball-and-stick model. The distance between residues 2 and 229 is ~19 Å. The linker of 17 amino acids, resulting from the spliced N- and C-terminal extension, should be more than sufficient to connect residues 2 and 229. Even an extended linker of 5 or 6 residues might be sufficiently long to span this distance. However, we have observed that the removal of 8 residues including the hexahistidine tag resulted in weaker fluorescence of the cells (data not shown). This is probably a consequence of the linker between PI-*Pfu*I$_C$ and GFPuv being too short to allow efficient folding into the correct structure of both GFPuv and the intein domains due to some steric hindrance. In this study, we have focused on the GFPuv construct of 245 amino acids shown in Fig. 2*A*. The presence of residue 229 (isoleucine) was found to be essential for functional structure formation of GFP in a truncation experiment, but it could be replaced by leucine with no measurable loss of fluorescence (26). Therefore, we replaced this residue with leucine to introduce a thrombin-specific proteolytic site.

GFPuv-extein and the artificially split inteins together form a precursor protein with a total length of 720 amino acids (83 kDa). Upon protein splicing, it is expected that three fragments are released: the N-terminal fragment PI-*Pfu*I$_C$ (315 amino acids, 37 kDa) with an N-terminal hexahistidine tag, the C-terminal fragment containing PI-*Pfu*I$_N$ (160 amino acids, 18.7 kDa) without any tag, and cyclized GFPuv (245 amino acids, 27.5 kDa) with an internal hexahistidine tag and a thrombin cleavage site.

In Vivo Cyclization of GFPuv and Purification of Cyclic GFPuv—The precursor protein consisting of the intein domains and GFPuv was expressed in LB medium at room temperature as described under "Experimental Procedures." The cells were harvested and showed bright green fluorescence under a UV
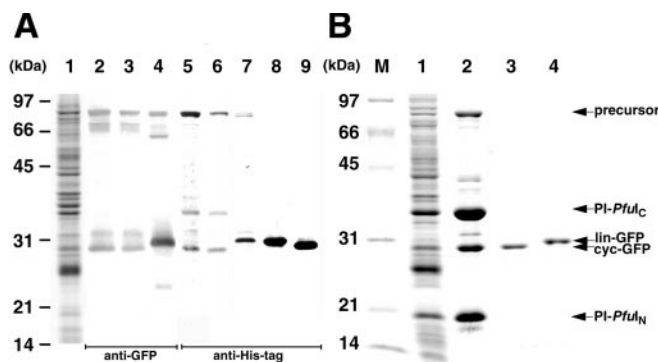
FIG. 3. **SDS-polyacrylamide gel electrophoresis analysis of *in vivo* cyclization in *E. coli.* A**, Western blot analysis. *Lanes 2–9* are Western blots using either an anti-GFP antibody (*lanes 2–4*) or an anti-His tag antibody (*lanes 5–9*). *Lane 1*, total cell lysate prepared by boiling and stained with Coomassie Blue; *lane 2*, total cell lysate prepared by boiling; *lane 3*, total cell lysate prepared by sonication; *lane 4*, total cell lysate prepared by sonication and digested with thrombin; *lane 5*, total cell lysate prepared by boiling; *lane 6*, total cell lysate prepared by sonication; *lane 7*, total cell lysate prepared by sonication and digested with thrombin; *lane 8*, purified linear GFPuv; *lane 9*, purified cyclic GFPuv. *B*, purification. *Lane 1*, supernatant of the cell lysate; *lane 2*, elution from a Ni^{2+}-NTA column as described under "Experimental Procedures"; *lane 3*, after organic extraction; *lane 4*, after thrombin digestion. All samples in *B* were analyzed on Coomassie Blue-stained 12% SDS-polyacrylamide gels. *lin-GFP*, linear GFPuv; *cyc-GFP*, cyclic GFPuv.

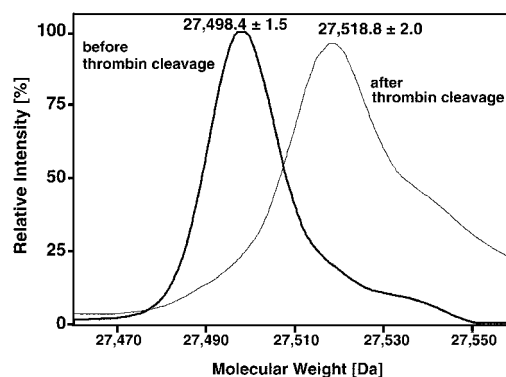FIG. 4. **Electron spray ionization mass spectrometry of purified GFPuv.** The *thick line* is the spectrum of purified GFPuv. The *thin line* is the spectrum of the sample after proteolytic digestion by thrombin. Linear GFPuv with the properly formed chromophore is expected to have a molecular mass of 27,517.8 Da.

lamp of 365-nm wavelength. Total cell lysates were prepared either by resuspension in SDS loading buffer containing 1 mM DTT and boiling at 95 °C for 10 min (Fig. 3*A*, *lanes 2* and *5*) or by sonication in 20 mM Tris (pH 8.0) and resuspension in SDS loading buffer containing 1 mM DTT, followed by boiling at 95 °C for 5 min (*lanes 3* and *6*). The cell lysates were analyzed by Western blotting with an anti-GFP antibody (Fig. 3*A*, *lanes 2–4*) and an anti-His tag antibody (*lanes 5–9*). Two stronger bands of ~80 and 28 kDa were detected with the anti-His tag antibody as well as with the anti-GFP antibody, corresponding to the precursor protein and GFPuv. The two samples prepared by sonication or by boiling showed the same pattern on the Western blots, indicating that the ~30-kDa bands were not produced during cell disruption (Fig. 3*A*). The 28-kDa band migrated more slowly after thrombin digestion and at the same position as the linear form, as explained in detail below, strongly suggesting that this band is the circular form of GFPuv, which is present *in vivo* already before purification (Fig. 3*A*, *lanes 4* and *7*). It is noteworthy that there was no detectable band on Western blots at the position of purified linear GFPuv, but we do not know the nature of the ~33-kDa band, which might be a proteolytically truncated precursor (Fig. 3*A*, *lanes 2* and *3*).

After the protein was expressed on a large scale (1 liter), the cell lysate was centrifuged, and the soluble fraction was applied to a Ni^{2+}-NTA column. Bound proteins were eluted with 300 mM imidazole (Fig. 3*B*, *lane 2*). There were four major bands of 80, 35, 28, and 19 kDa in the eluted fraction. The band with the highest molecular mass corresponds to the unspliced precursor protein. The product of 28 kDa was identified as GFPuv by a second step of purification and mass spectrometry. The 35- and 19-kDa fragments are likely to be the spliced and noncovalently associated N- and C-terminal intein domains PI-*Pfu*I$_N$ and PI-*Pfu*I$_C$, respectively, because they could associate under nondenaturing conditions to form the whole functional intein, and then both were purified via the N-terminal hexahistidine tag of PI-*Pfu*I$_C$ (*cf.* the 35-kDa band in the anti-His tag blot) (Fig. 3*A*, *lanes 5* and *6*). It is worth mentioning that elution from the Ni^{2+}-NTA column resulted in a bright green fluorescent frac-

tion, indicating the presence of properly folded GFPuv. To purify GFPuv further, the elution fraction was subjected to organic solvent extraction (14). After the extraction, there was only a single band of 28 kDa (Fig. 3*B*, *lane 3*), and the yield was ~3 mg from 1 liter of bacterial culture. The solution had a bright green color, and the ratio of absorbance at 397 nm to that at 280 nm was ~1.2, which is a good indication of the purity of GFP (27). Furthermore, there were no detectable polymeric forms found.

We have shown that this single band is indeed the cyclized form of GFPuv (see "Confirmation of Cyclization of GFPuv" below). It should be emphasized that there was no detectable contaminant at the position of the purified linear form, neither in the elution of the Ni^{2+}-NTA column nor in the organic extraction fraction. This could be an advantage over other *in vitro* cyclization methods, which always involve a mixture of linear and circular forms (2, 8, 9). Even the *in vivo* cyclization using the naturally split intein *Ssp* DnaE yielded a fraction of the linear form (3, 10). We have observed the unspliced full-length precursor protein in the cell extracts, but no significant linear form of GFPuv was present in the purified fraction. To regenerate the splicing activity, the precursor protein has to fold correctly. Once properly folded into the correct structure, the splicing efficiency is probably high. Since it might be more difficult to form the functional structure of the artificially split intein than that of the naturally split intein, the precursor might be more aggregation-prone, and only correctly spliced molecules might escape precipitation.

Confirmation of Cyclization of GFPuv—The cyclization was confirmed by electron spray ionization mass spectrometry and Edman degradation after linearization by thrombin cleavage. The target protein was designed to have a unique and specific proteolytic site (LVPRGT, similar to the normal thrombin recognition sequence LVPRGS) following the GFPuv sequence and in front of the PI-*Pfu*I$_N$ sequence (Fig. 2*A*). Purified GFPuv was treated with thrombin, followed by amino acid sequencing. We obtained the sequence GTGTG for the digested linear protein, but no sequence was detected for the purified GFPuv sample without thrombin digestion, confirming that the cyclization was successful. The precursor would have given the sequence GTGCI (Fig. 2*A*). The purified protein migrated faster than the linear protein upon SDS-polyacrylamide gel electrophoresis, which is also expected for circular proteins because the circular form would have a smaller radius of gyration than the linear form in the denatured state.

The molecular mass of purified GFPuv before and after the thrombin treatment was analyzed by electron spray ionization mass spectrometry (Fig. 4). From the amino acid sequence, the
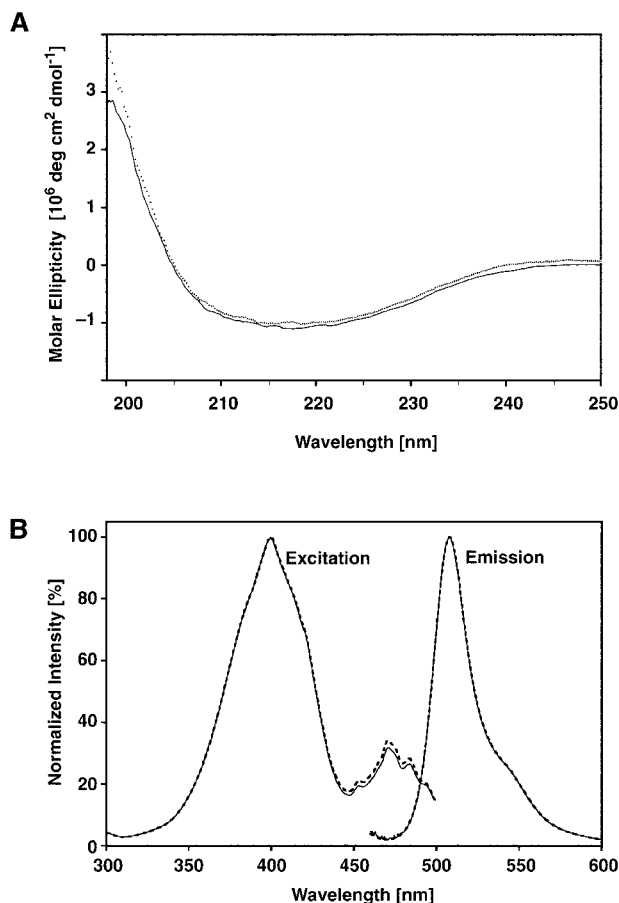
In Vivo Cyclization of Green Fluorescent Protein

A

B

FIG. 5. **Spectroscopic characterization of the linear and circular forms of GFPuv.** *A*, CD spectra of the linear and circular forms of GFPuv. The *dotted line* represent the data points of the linear form. The *solid line* is the spectrum of the cyclic form. Both spectra were measured at 25 °C and pH 8.0. *B*, emission and excitation spectra of the cyclic and linear forms of GFPuv. Fluorescence emission scans were obtained by excitation at 397 nm at pH 7.5. Fluorescence excitation spectra were recorded at an emission wavelength of 508 nm at pH 7.5. The *dashed* and *solid lines* indicate the linear and cyclic forms, respectively. The intensities were normalized to the peak maximum. *deg*, degrees.

unmodified linear form is expected to have a molecular mass of 27,537.8 Da. Chromophore formation is due to an autocyclodehydration ($-H_2O$) and an autooxidation ($-2H$); therefore, linear GFPuv with the properly formed chromophore is expected to lose 20 Da and to have a molecular mass of 27,517.8 Da (28). The observed molecular mass of the non-thrombin-treated sample was, however, $27,498.4 \pm 1.5$ Da, which is 19 Da smaller than the expected value for the linear molecule, indicating loss of one water molecule. This is consistent with cyclization (dehydration) (Fig. 4). Subsequent specific proteolysis with thrombin *increased* the mass to $27,518.8 \pm 2.0$ Da, suggesting hydration (18 Da), *viz.* hydrolysis of a peptide bond. These observations fit perfectly with the cyclization of this molecule and, at the same time, with the proper formation of the chromophore, which can also be observed by its green fluorescence.

Comparison of Fluorescence and CD Spectra of the Cyclic and Linear Forms of GFPuv—The fluorescence spectrum of GFP is a good indicator of the correctly folded structure of GFP because the unique fluorescence of GFP is a result of both its chromophore formation and properly folded three-dimensional structure. In Fig. 5*B*, the fluorescence properties of cyclized GFPuv and thrombin-linearized GFPuv are compared. Both the emission and excitation profiles of the cyclic and linear forms of GFPuv are essentially identical, which strongly sug-

gests that the three-dimensional structures of both forms are also very similar because the fluorescence of the chromophore is sensitive to its structural environment (28). There were no differences in the fluorescence properties observed between the circular and linear forms of GFPuv at pH 6.0–8.5 (data not shown). The emission maximum of 508 nm and the excitation maximum of 399 nm for our cyclic and linear forms of GFPuv are also very close to the reported values for both wild-type GFP (excitation maximum of 395–397 nm and emission maximum of 504 nm) (28) and GFPuv (excitation maximum of 397 nm and emission maximum of 506 nm) (28), suggesting that there is no significant structural change. Since cyclization merely adds another loop outside the β-barrel and since linearization with thrombin converts this loop into two presumably unstructured tails, this similarity was anticipated.

The secondary structures of the linear and cyclic forms of GFPuv were compared using CD spectroscopy as shown in Fig. 5*A*. Both spectra from the linear and circular forms of GFPuv show very similar profiles, suggesting that there is no significant difference in the secondary structure between the linear and circular forms.

Unfolding and Refolding of the Cyclic and Linear Forms of GFPuv—The entropically most expensive process during protein folding is to bring the residues that are distant in the primary sequence close to each other. Therefore, the cyclization of the protein would be expected to reduce such entropic cost during folding. In addition, because the cyclization prevents "peeling" of terminal secondary structure elements, it might make a protein more difficult to unfold. Hence, we were interested in elucidating if the cyclization of GFPuv improves stability, as observed in other cyclized proteins (2, 3), and how the folding or unfolding kinetics might be affected. However, GFP itself is already a very stable protein once its structure has formed. The unfolding of GFPuv seems to be not fully reversible; and due to very slow processes both in unfolding and refolding, it is not possible to study the equilibrium precisely (18). Therefore, we have investigated whether there is a difference in the rates of unfolding and refolding between the circular and linear forms. These processes were monitored by the change in the fluorescence intensity at 508 nm with an excitation of the chromophore at 397 nm. The time course of the unfolding reactions could be fitted with a single exponential, and we assumed that the fluorescence would be zero at infinite time (total loss of chromophore fluorescence). In Fig. 6*A*, unfolding of the linear and circular forms in 6.7 M GdnHCl is followed by fluorescence intensity *versus* time. It is clearly visible that the circular form unfolded more slowly than the linear form. At 7 M GdnHCl, the unfolding rate of the circular form was about half of the linear form (Fig. 6*B*). The unfolding rates are plotted against the final GdnHCl concentration in Fig. 6*B*. GFPuv is no more fully unfolded in 5.5 M GdnHCl (18), preventing a comparison at low GdnHCl concentrations.

In Fig. 7, we examined the refolding process by monitoring the changes in the fluorescence intensity at 508 nm upon excitation at 397 nm. Folding was initiated by dilution in the refolding buffer after denaturation either in 6 M GdnHCl for 3 h (18) (Fig. 6*B*) or in 7 M GdnHCl for various lengths of time. Refolding was biphasic for both molecules, consistent with previous measurements (18), and the relative amplitude of the slow phase was greater in linear GFP than in the circular molecule (Fig. 7). The recovery of fluorescence appeared to be faster for the circular form than for the linear form (Figs. 6*B* and 7, *A* and *B*), but this was due largely to changes in amplitudes and not in rate constants. We found that the rate constants for the slow phase in 0.5 M GdnHCl, refolded from denaturation in 6 M GdnHCl, are similar for both forms (($4.0 \pm$
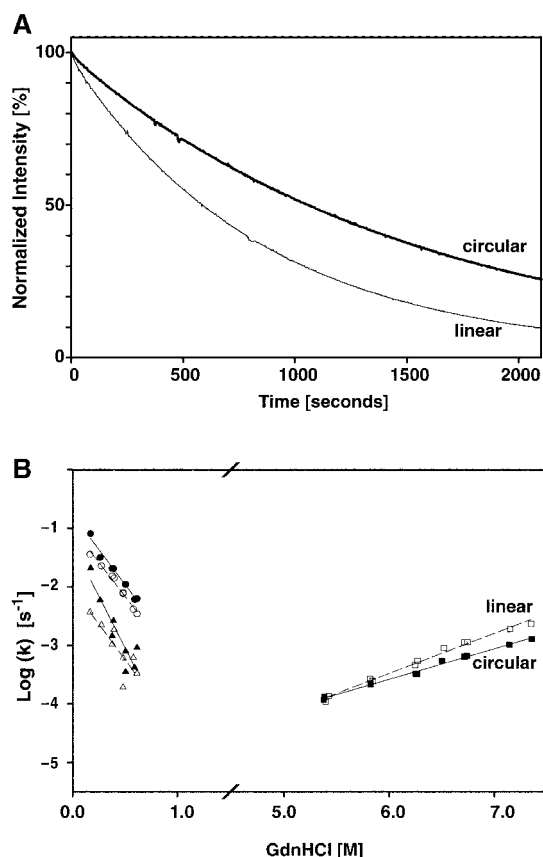
FIG. 6. **Unfolding kinetics of the linear and cyclic forms of GFPuv.** *A*, typical unfolding kinetics of the linear and circular forms of GFPuv. The measurements were performed at pH 8.0 with a final concentration of 6.7 M GdnHCl. *B*, unfolding and refolding kinetics of the linear and circular forms of GFPuv *versus* GdnHCl concentrations. The unfolding rate constants were determined by fitting the change in the fluorescence intensity at 25 °C and pH 8 with a single exponential (see "Experimental Procedures"). □ and ■, unfolding rates for the linear and cyclic forms, respectively. The refolding rate constants were obtained by fitting the change in the fluorescence intensity with a double exponential after denaturation in 6 M GdnHCl for 3 h (see "Experimental Procedures"). The fast and slow kinetic constants are represented by *circles* and *triangles*, respectively. The *open* and *closed symbols* represent the linear and circular forms, respectively. The data points were fitted by linear regression. The *dashed* and *solid lines* represent the linear and cyclic forms, respectively.

$2.8) \times 10^{-4}$ s^{-1} for the linear form and $(2.6 \pm 3.5) \times 10^{-4}$ s^{-1} for the circular form) and that the time constants of the fast phase seem to be only slightly larger for the circular form than for the linear form $((6.5 \pm 0.7) \times 10^{-3}$ s^{-1} for the linear form and $(11.4 \pm 1.6) \times 10^{-3}$ s^{-1} for the circular form). The values were the same within experimental errors whether denaturation was carried out in 6 or 7 M GdnHCl. The rate constants did depend on final GdnHCl concentration, however (Fig. 6*B*). Nevertheless, what changed dramatically between the circular and linear forms and as a function of denaturation time was the amplitude of both phases.

We consider that this is due to a slow disappearance of residual structure in GFP, which could conceivably permit the survival of the single *cis*-proline (Pro[89]) (15, 24, 25). The fast phase corresponds to the folding of molecules with Pro[89] in *cis*, and the slow phase with Pro[89] in *trans*. In fact, the longer denaturation time and higher concentration of GdnHCl reduced the amplitude of the fast phase, probably indicating the slow loss of some residual structure in the denatured state (Fig. 7*C*). The reduction of this amplitude was clearly slower for the circular form (Fig. 7*C*), indicating that it maintains this residual structure more strongly. This observation is also supported
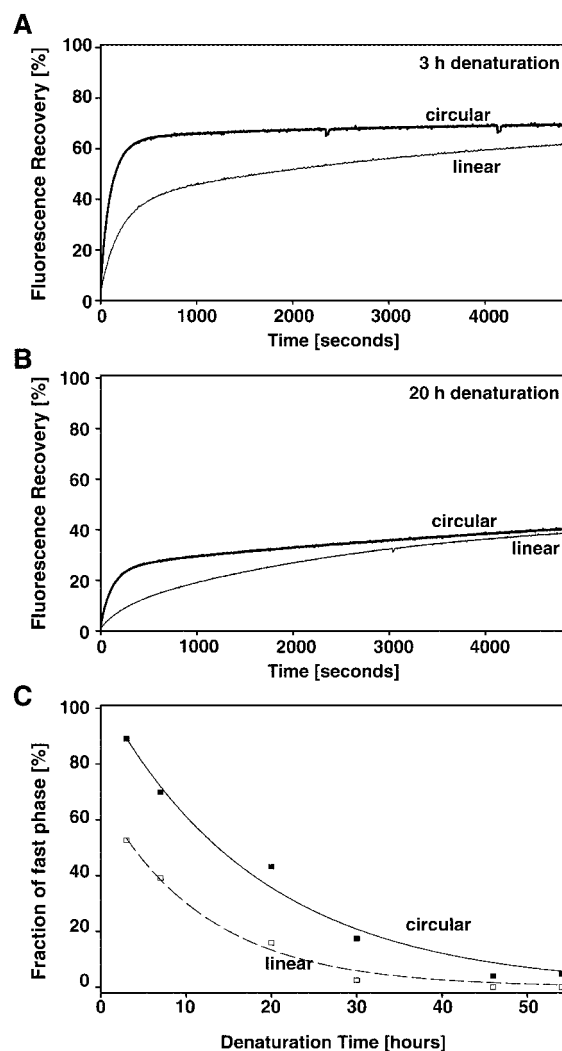

FIG. 7. **Refolding kinetics of the linear and cyclic forms of GFPuv.** *A*, refolding kinetics of the linear and circular forms of GFPuv after 3 h of denaturation in 7 M GdnHCl (pH 8.0). The reaction was started by dilution to a final concentration of 0.5 M GdnHCl. *B*, same as *A*, but after 20 h of denaturation in 7 M GdnHCl. In *A* and *B*, *thin* and *thick lines* represent the linear and circular forms, respectively. All traces were obtained by following the intensity at 508 nm upon excitation at 397 nm. *C*, fraction of the fast phase in the recovered total fluorescence *versus* denaturation time. The points were fitted with an exponential function. ■, circular form; □, linear form.

by the residual secondary structure detected with CD spectra after denaturation in 6 and 7 M GdnHCl for 3 h (data not shown).

The analysis is complicated by the fact that longer denaturation times dramatically reduce the recovery of the chromophore fluorescence. Therefore, it is difficult to compare the refolding kinetics of the two forms from an unambiguously fully denatured state. We assume that the entropic effect of circularization, *i.e.* the faster initial collapse of the chain to a productive topology, has already taken place long before the beginning of the manual mixing measurements and would be invisible by following the recovery of the chromophore fluorescence. Undoubtedly, further studies will be needed to untangle the refolding process of GFP, and the effect of circularization on any putative stopped-flow phases or even stopped-flow burst phases will have to be evaluated. Nevertheless, from the data presented here, there is clear evidence that the circular molecule is kinetically protected against chemical denaturation. This can be seen in both the rate of loss of chromophore fluorescence (Fig. 6*A*) and the rate of loss of the fast phase during

refolding (Fig. 7*C*), indicating loss of residual structure in the denatured state.

DISCUSSION

In this report, it is demonstrated that a cyclized protein can be produced *in vivo* by making a circular permutation of a precursor protein by artificially splitting a naturally occurring intein domain. We used the PI-*Pfu*I intein from *P. furiosus*, which has been successfully used for protein ligations (11, 20). This permutation approach to produce a circular backbone topology has several advantages over other *in vitro* as well as *in vivo* cyclization approaches. Since circular permutations of proteins have been found in nature, our experiments indicate at least the possibility that some circular peptides might exist that do use an intein mechanism of cyclization.

The most unique feature of our system is that mainly the cyclic form of GFPuv was produced in the cell, as demonstrated by Western blotting, and the pure cyclic form could be purified from the *E. coli* cell extract. No by-product of the linear form was purified, which both have often been obtained in other cyclization methods due to the hydrolysis of the thioester bond at the C terminus of an extein before peptide bond formation (2, 3, 9, 10). Although an unspliced precursor remained in our system, due to much larger molecular mass, it appeared to be mostly insoluble and thus could easily be separated from the spliced products. Therefore, we could obtain the pure circular form without any additional procedures to remove a by-product of the linear form, as was necessary in our previous work (2). This separation is not easy due to the very similar physicochemical properties of the circular and linear forms. The absence of the linear form in the isolated product also makes our new approach advantageous compared with a similar *in vivo* cyclization approach using the naturally split intein *Ssp* DnaE, which produces by-products of linear forms even *in vivo*. The *in vivo* cyclization could be very useful for generating intracellular cyclic peptide and protein libraries (3, 10).

We assume that the reason for the lack of contaminating linear forms is mostly the intein used here. The products of *cis*-splicing with DnaE were contaminated with a product in which only the N-terminal extein piece of *Ssp* DnaE was cleaved off (10, 29). If the same single cleavage occurred when it was used for the cyclization reaction, this would result in a linear form. In contrast, *trans*-splicing with PI-*Pfu*I does not give rise to single cleavage at the junction between the first extein and the intein domain (11). This could result from a difference between PI-*Pfu*I and *Ssp* DnaE with regard to different rates of N–S acyl migration. Nevertheless, this point requires further investigation, as does the influence of the extein used.

Cyclized GFPuv shows some characteristic differences in molecular properties compared with the linear form, despite the very similar structural properties shown by CD and fluorescence spectra of both forms. In particular, the unfolding process was clearly slower at high concentrations of GdnHCl (Figs. 6 and 7*C*). These data suggest that cyclized GFPuv is stabilized compared with the linear form against chemical denaturation. As has been suggested previously, the backbone cyclization could be a general strategy for stabilizing proteins following the rationale of polymer theory (2). The detailed mechanism of the stabilization effect is now under investigation with other proteins in our laboratory.

In conclusion, the cyclization approach by circular permutation of the precursor protein has been very efficient in producing a large quantity of a circular protein, and it could be potentially useful for generating other circular proteins. In addition, the cyclization efficiency seems to be very high, and there are no significant amounts of by-products of linear or polymerized forms isolated. Our *in vivo* cyclization approach with its high efficiency of cyclization may provide possibilities for several new applications of cyclic peptides and proteins *in vivo* as well as for biophysical characterization of circular proteins. Selection of biologically active cyclic proteins or peptides using a combinatorial approach with *in vivo* functional selections, *in vivo* expression of the cyclization-stabilized proteins without changing their primary structures, and new tools to study the biological functions of the topology or termini in cellular environments are only some of the interesting future studies. In particular, cyclized GFP may be potentially useful for such studies since the unique fluorescence of GFP has made it a convenient indicator of expression, processing, and cellular environments.

REFERENCES

1. Flory, P. J. (1956) *J. Am. Chem. Soc.* **78,** 5222–5235
2. Iwai, H., and Plückthun, A. (1999) *FEBS Lett.* **459,** 166–172
3. Scott, C. P., Abel-Santos, E., Wall, M., Wahnon, D. C., and Benkovic, S. J. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 13638–13643
4. Goldenberg, D. P. (1985) *J. Cell. Biochem.* **29,** 321–335
5. Muir, T. W., Sondhi, D., and Cole, P. A. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 6705–6710
6. Evans, T. C., Jr., Benner, J., and Xu, M. Q. (1998) *Protein Sci.* **7,** 2256–2264
7. Perler, F. B., and Adam, E. (2000) *Curr. Opin. Biotechnol.* **11,** 377–383
8. Camarero, J. A., and Muir, T. W. (1999) *J. Am. Chem. Soc.* **121,** 5597–5598
9. Evans, T. C., Jr., Benner, J., and Xu, M. Q. (1999) *J. Biol. Chem.* **274,** 18359–18363
10. Evans, T. C., Jr., Martin, D., Kolly, R., Panne, D., Sun, L., Ghosh, I., Chen, L., Benner, J., Liu, X. Q., and Xu, M. Q. (2000) *J. Biol. Chem.* **275,** 9091–9094
11. Otomo, T., Teruya, K., Uegaki, K., Yamazaki, T., and Kyogoku, Y. (1999) *J. Biomol. NMR* **14,** 105–114
12. Crameri, A., Whitehorn, E. A., Tate, E., and Stemmer, W. P. (1996) *Nat. Biotechnol.* **14,** 315–319
13. Ge, L., Knappik, A., Pack, P., Freund, C., and Plückthun, A. (1995) in *Antibody Engineering* (Borrebaeck, C. A. K., ed) pp. 229–266, Oxford University Press, New York
14. Yakhnin, A. V., Vinokurov, L. M., Surin, A. K., and Alakhov, Y. B. (1998) *Protein Expression Purif.* **14,** 382–386
15. Palm, G. J., Zdanov, A., Gaitanaris, G. A., Stauber, R., Pavlakis, G. N., and Wlodawer, A. (1997) *Nat. Struct. Biol.* **4,** 361–365
16. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics* **14,** 52–55
17. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) *J. Mol. Biol.* **273,** 283–298
18. Fukuda, H., Arai, M., and Kuwajima, K. (2000) *Biochemistry* **39,** 12025–12032
19. Reid, B. G., and Flynn, G. C. (1997) *Biochemistry* **36,** 6786–6791
20. Yamazaki, T., Otomo, T., Oda, N., Kyogoku, Y., Uegaki, K., Ito, N., Ishino, Y., and Nakamura, H. (1998) *J. Am. Chem. Soc.* **120,** 5591–5592
21. Ichiyanagi, K., Ishino, Y., Ariyoshi, M., Komori, K., and Morikawa, K. (2000) *J. Mol. Biol.* **300,** 889–901
22. Baird, G. S., Zacharias, D. A., and Tsien, R. Y. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 11241–11246
23. Topell, S., Hennecke, J., and Glockshuber, R. (1999) *FEBS Lett.* **457,** 283–289
24. Ormö, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996) *Science* **273,** 1392–1395
25. Yang, F., Moss, L. G., and Phillips, G. N., Jr. (1996) *Nat. Biotechnol.* **14,** 1246–1251
26. Li, X., Zhang, G., Ngo, N., Zhao, X., Kain, S. R., and Huang, C. C. (1997) *J. Biol. Chem.* **272,** 28545–28549
27. Ward, W. W., Prentice, H. J., Roth, A. F., Cody, C. W., and Reeves, S. C. (1982) *Photochem. Photobiol.* **35,** 803–808
28. Tsien, R. Y. (1998) *Annu. Rev. Biochem.* **67,** 509–544
29. Kolly, R. (1999) Protein cyclization using two-intein vectors. M.Sc. thesis, Ecole Superieure de Biotechnologie de Strasbourg