Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of cyclized green fluorescent protein

Andreas Hofmann,[a]*† Hideo Iwai,[b] Sonja Hess,[c] Andreas Plückthun[b] and Alexander Wlodawer[a]

[a]Macromolecular Crystallography Laboratory, NCI at Frederick, Frederick, MD 21702, USA, [b]Biochemisches Institut der Universität Zürich, 8057 Zürich, Switzerland, and [c]Structural Mass Spectrometry Facility, Laboratory of Bioorganic Chemistry, NIDDK, NIH, Bethesda, MD 20892, USA

† Present address: Institute of Cell & Molecular Biology, The University of Edinburgh, Edinburgh EH9 3JR, Scotland.

Correspondence e-mail: hofmanna@ncifcrf.gov

Received 15 February 2002
Accepted 11 June 2002

PDB Reference: cyclized green fluorescent protein, 1kp5, r1kp5sf.

Crystals of cyclic green fluorescent protein (cGFP) engineered by the previously reported split intein technology [Iwai *et al.* (2001), *J. Biol. Chem.* **276**, 16548–16554] were obtained and the structure was solved using molecular replacement. Although the core of the protein can unambiguously be fitted from the first to the last residue of the genuine sequence, the electron density in the region of the linker peptide is rather poor owing to the high water content of the crystals. Therefore, it is concluded that this part of the protein is highly disordered in the present structure and is very flexible. This is supported by the absence of crystal contacts in the linker-peptide region and the fact that the core of the protein exhibits a very similar conformation to that known from other GFP structures, thereby not implicating any constraints arising from the presence of the artificial linker. Nevertheless, the density is consistent with the loop being intact, as confirmed by mass spectroscopy of dissolved crystals. The present structure contains an antiparallel cGFP dimer where the dimer interface is clearly different from other crystal structures featuring two GFP molecules. This adds further support to the fact that the cylinder surface of GFP is rather versatile and can employ various polar and non-polar patches in protein–protein interactions.

1. Introduction

A landmark feature of naturally occurring proteins is their architecture comprised of linear chains of amino acids that fold into three-dimensional structures according to their sequences. So far, naturally occurring cyclic backbone structures have been found only in small peptides (Saether *et al.*, 1995; Craik *et al.*, 1999; Daly *et al.*, 1999; Gonzalez *et al.*, 2000; Felizmenio-Quimio *et al.*, 2001; Korsinczky *et al.*, 2001). Introduction of circular topology was originally performed using chemical synthesis (Creighton & Goldenberg, 1984; Goldenberg & Creighton, 1984). More recently, intein-catalyzed protein ligation has been used to circularize proteins (Iwai & Plückthun, 1999; Camarero & Muir, 1999; Scott *et al.*, 1999; Evans *et al.*, 1999; Iwai *et al.*, 2001; Camarero *et al.*, 2001; Siebold *et al.*, 2001; Williams *et al.*, 2002). With this technology, circular forms even of large proteins could be prepared. For a recent review on circular proteins, see Trabi & Craik (2002).

According to polymer theory (Flory, 1956), a circular structure should possess increased stability compared with its linear pendant owing to the reduced conformational entropy in the denatured state. This has long been accepted as the basis for studies on cyclic proteins, but takes only the unfolded state of the folding reaction into consideration. In order to stabilize a protein (compared with its linear form) one also has to keep

the free energy of the product (the folded protein) at the same level. With several cyclized proteins such as, for example, the cyclic bovine pancreatic trypsin inhibitor (BPTI; Creighton & Goldenberg, 1984; Goldenberg & Creighton, 1984) and the cyclized Src homology 3 (SH3) domain of the murine c-Crk adapter protein (Camarero *et al.*, 2001), no stabilization has been found, since the cyclic product was destabilized by strained conformations. Dihydrofolate reductase (Scott *et al.*, 1999) and cyclized β-lactamase (Iwai & Plückthun, 1999) exhibit increased thermal stability when compared with their linear forms. The latter protein was also found to be protected against degradation by exopeptidases (Iwai & Plückthun, 1999). Cyclization can stabilize a protein with an otherwise unstable fold, as reported for a truncation mutant of the SH3 domain of the c-Crk adapter (Camarero *et al.*, 2001).

Pioneering work on cyclized proteins was carried out with BPTI, which was prepared by treatment of the native protein with carbodiimide (Goldenberg, 1985). The cyclic nature of this constructed protein was proven by NMR, but the crystal structure of cyclic BPTI, this time prepared by total chemical synthesis, has been reported only recently (Botos *et al.*, 2001). A cyclized form of the α-defensin NP-1 was shown to be less cytotoxic, thus making it a more convenient antimicrobial drug (Yu *et al.*, 2000). Structural information for small cyclic proteins is available from nuclear magnetic resonance (NMR) for members of the cyclotide family of peptides, including kalata B1 (Saether *et al.*, 1995), circulin A (Daly *et al.*, 1999), cycloviolacin O1 (Craik *et al.*, 1999) and the macrocyclic trypsin inhibitor from *Momordica cochinchinensis* MCoTI-II (Felizmenio-Quimio *et al.*, 2001). Other NMR structures of peptides with circular topology, including bacteriocin AS-48, a microbial relative of the mammalian NK-lysin (Gonzalez *et al.*, 2000), and the cyclic trypsin inhibitor SFTI-1 from sunflower (Korsinczky *et al.*, 2001), have also been elucidated and suggest that circular backbone structures are used in these small peptides together with disulfide bridges.

The green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* possesses a chromophore derived from the tripeptide Ser-Tyr-Gly located within its primary structure, without requiring any exogenous cofactor for its fluorescence at long wavelengths (Prendergast & Mann, 1978). The extraordinary potential of GFP as a reporter of gene expression, protein trafficking and interactions has been reviewed extensively (see, for example, Cubitt *et al.*, 1995). Recently, a cyclic form of GFP was reported (Iwai *et al.*, 2001) which might provide unique opportunities to study the importance of termini and/or circular backbone topology in various cellular processes, such as protein degradation and protein trafficking. The experimental approach used protein splicing of a circular permutated precursor protein that contained an artificially split intein from *Pyrococcus furiosus* (PI-*Pfu*I). *In vivo*, cyclization has been reported earlier with the help of the naturally split intein DnaE from *Synechocystis* sp. PCC6803 (Scott *et al.*, 1999; Evans *et al.*, 2000); however, in contrast to DnaE, the protocol using PI-*Pfu*I produces only the cyclic form, which can be purified rapidly from the bacterial cell extract, thus being more efficient.

GFP has been subject of extensive structural investigation, which can easily be seen from the total of 22 structures available in the PDB. The first crystal structures of wild-type GFP and the S65T mutant were determined almost simultaneously in 1996 (Ormö *et al.*, 1996; Yang *et al.*, 1996). While the structural properties as assessed with CD spectroscopy as well as the fluorescence characteristics of the cyclic form were indistinguishable from the linear form, the cyclized protein proved to be much more stable in unfolding experiments (Iwai *et al.*, 2001). As suggested by the rationale of polymer theory mentioned earlier, backbone cyclization increased the stability of GFP. It was thus anticipated that the protein core of the cyclized protein would not be in a conformation with additional strain imposed by the artificial linker. Nevertheless, as the rate of unfolding was measured to compare the stability of linear and cyclic GFP, since the protein never fully reaches equilibrium in unfolding transitions, the explanation of the higher stability cannot be limited to a consideration of the unfolded state.

We show in the present study that the core conformation of the cyclic protein is the same as that observed in linear GFP. The artificial peptide sequence covalently linking the N- and C-terminal residues of the precursor protein with a peptide bond adopts a flexible conformation as observed from the high degree of disorder.

2. Materials and methods

2.1. Protein expression, purification and crystallization

Generation of the construct of circular green fluorescent protein as well as its expression and purification has been described previously (Iwai *et al.*, 2001). Briefly, the plasmid pIWT5563his containing the DNA of GFP Cycle3 mutant and the split intein in the expression vector pTFT74 under the control of the T7 promoter was transformed into *Escherichia coli* BL21-CodonPlus(DE3)-RIL cells, which overexpresses tRNA for rare codons, a large number of which occur in the PI-*Pfu*I gene. The protein was expressed overnight at a slow rate exploiting the leakage of the expression vector and for an additional 6 h after induction with 0.1 m*M* isopropyl-β-D-thiogalactoside. Purification of the protein was performed by affinity chromatography with an Ni^{2+}–NTA column and chromatography on heparin Sepharose.

Green-coloured polyhedric crystals of cyclic GFP were obtained from 1.1 *M* $(NH_4)_2SO_4$, 0.1 *M* glycine pH 10.0 at 285 K after 12 weeks by the hanging-drop vapor-diffusion method.

2.2. Mass spectrometry

Crystals of the cyclic protein were dissolved in water and checked with liquid chromatography–electrospray ionization mass spectrometry (LC–MS) on an Agilent 1100 LC-MSD instrument (Agilent, Palo Alto, CA, USA). A 100 µl (50 pmol) aliquot of the sample was loaded onto a 150 × 2.1 mm ID reversed-phase Zorbax SC-C3 column equipped with a C3 guard column at a flow rate of 200 µl min^{-1}. The protein was

Table 1
Data-collection statistics.

Values in parentheses refer to the last resolution shell.

Data set	CG3
Space group	$P6_522$
Unit-cell parameters (Å)	a = 142.5, b = 142.5, c = 183.8
Resolution (Å)	25–2.6
No. of measurements†	917968
No. of independent reflections	34481
Completeness (%)	97.8 (99.3)
$R_{\rm merge}$ ‡	0.078 (0.503)

† Including partial reflections. ‡ $R_{\rm merge} = \sum |I - \langle I \rangle| / \sum I$.

eluted after an initial wash of 25 min at 5% organic phase with a gradient from 5 to 100% organic phase (the aqueous phase contained 5% acetic acid in water; the organic phase was acetonitrile). MS data were acquired and analyzed with the ChemStation software.

The obtained mass of 27 498 g mol^{-1} shows that the protein has formed the chromophore and is indeed cyclized (molecular mass of linear GFP with chromophore: 27 518 g mol^{-1}).

2.3. Data collection and structure solution

In-house diffraction was limited to 3.5 Å, whereas the maximal resolution achieved at beamline X9B (Brookhaven National Laboratory) was 2.6 Å. Data processing was carried out with the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997) and the statistics are summarized in Table 1. The diffraction pattern indicated a hexagonal space group, with approximate unit-cell parameters $a = b = 142$, $c = 184$ Å. The self-rotation function calculated with *GLRF* (Tong & Rossmann, 1990) showed the presence of twofold axes at [100], [010], [110], [210] and [120], thus narrowing the space group to $P622$. ($00l$) reflections other than $l = 6n$ were systematically absent, which left $P6_122$ or $P6_522$ as final possibilities. The structure was solved by molecular replacement with the program *AMoRe* (Navaza, 1994) and the model 1emm (Palm *et al.*, 1997) as deposited in the PDB but omitting the chromophore. A clear solution was found in space group $P6_522$, with correlation coefficients of 32.4 and 67.5 for the first and second molecule, respectively (R factors of 0.443 and 0.316, respectively). The next highest peak appeared with correlation coefficients 32.4 and 34.4, respectively. The rather large unit-cell parameters and two molecules in the asymmetric unit are the reason for the Matthews coefficient (Matthews, 1968) of 5.0 Å^3 Da^{-1}, corresponding to 75% water content. This is reflective of the packing within the crystal, which is constituted of a ring-like arrangement of molecules with the protein dimer being the repetition unit. The diameter of these rings is approximately 100 Å.

2.4. Model building and refinement

The monomers in the asymmetric unit are related by proper non-crystallographic symmetry (NCS). The twofold axis runs almost parallel to [130]. Since this is very close to the crystallographic twofold axis at [120] it explains why the NCS axis was not observed in the self-rotation function.

Table 2
Refinement statistics.

Values in parentheses refer to the last resolution shell.

	Total	Molecule #1	Molecule #2
Refinement			
Resolution (Å)	20–2.6		
No. of reflections used for refinement ($F > 0$)	31869		
No. of non-H atoms	4072		
No. of non-H protein atoms	3801	1898	1903
Visible residues		11–243	11–243
R factor			
No. of reflections in working set	28,676		
No. of reflections in test set	3193		
R†	0.208 (0.258)		
$R_{\rm free}$‡	0.256 (0.319)		
Temperature factors			
Scaling	Anisotropic		
Average B factor (Å^2)	38.3	38.8	35.7
R.m.s. deviation for bonded atoms (Å^2)	3.662	3.612	3.711
Geometry, r.m.s. deviations			
Bond lengths (Å)	0.006		
Bond angles (°)	1.3		
Dihedral angles (°)	26.2		
Improper angles (°)	0.87		
Ramachandran plot, residues in			
Most favored regions (%)	88.0	88.4	88.2
Additionally allowed regions (%)	11.5	11.1	11.3
Generously allowed regions (%)	0.5	0.5	0.5
Disallowed regions (%)	0	0	0
Solvent statistics			
No. of water molecules	94		
No. of sulfate ions	8	4	4

† R factor = $\sum ||F_o| - |F_c|| / \sum |F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively. ‡ $R_{\rm free}$ defined in Brünger (1992).

The initial model was rebuilt with the program *O* (Jones *et al.*, 1991) and subjected to extensive cycles of computational refinement interspersed with visual inspection and manual adjustments. Refinement was carried out using the conjugate-gradient method with *CNS* v.1.0 (Brünger *et al.*, 1998) and in the early stages moderate NCS restraints were applied on the core of both molecules. The maximum-likelihood target function was employed and typical protocols consisted of a positional refinement followed by simulated annealing, grouped and individual B-factor refinement and a final positional refinement. A bulk-solvent model and overall anisotropic B-factor correction were applied throughout the procedure. Since tracing of the genuine N- and C-termini as well as the artificial linker region in the electron density was very difficult, we used cyclic NCS averaging of the electron-density maps after each completed refinement cycle. Averaging was carried out with the program *DM* from the *CCP*4 package (Collaborative Computational Project Number 4, 1994). In contrast, conventional as well as composite omit procedures did not improve the quality of the density. The structure was refined to a final R factor of 0.208 ($R_{\rm free}$ = 0.256) and reasonable overall geometry as monitored with the program *PROCHECK* (Laskowski *et al.*, 1993); the refinement statistics are summarized in Table 2.

?th > The position of some isolated residues within the artificial linker (243–10) is supported by the electron density, which was obtained after map-averaging procedures. Other parts of the linker region are not defined and were modeled following strictly geometrical considerations. Therefore, we decided to set the occupancy to zero for most of these residues in both molecules.

3. Results

3.1. Overall structure and comparison with linear GFP

Wild-type GFP was engineered as described previously (Iwai *et al.*, 2001) and the artificial linker introduced into the protein was comprised of ten amino acids (TGSRHHHHHH) at the N-terminus (Fig. 1). Additionally, Ala1 and Ser2 of the genuine sequence were replaced by Ser11 and Arg12, respectively. At the C-terminal end, residue Gly228 is the last original amino acid and positions 229–235 were replaced by the sequence LVPRGTG. Residues 236–238 of the wild-type sequence were removed since their electron density was invisible in many earlier GFP crystal structures. The final cGFP is thus composed of 245 residues, seven residues more than the wild-type protein. Crystals obtained with cGFP were

```
  1  tgsrhhhhhh SrKGEELFTG VVPILVELDG DVNGHKFSVS GEGEGDATYG
                ASKGEELFTG VVPILVELDG DVNGHKFSVS GEGEGDATYG
                   GEELFTG VVPILVELDG DVNGHKFSVS GEGEGDATYG
                   GEELFTG VVPILVELDG DVNGHKFSVS GEGEGDATYG

 51  KLTLKFICTT GKLPVPWPTL VTTFSYGVQC FSRYPDHMKR HDFFKSAMPE
     KLTLKFICTT GKLPVPWPTL VTTFSYGVQC FSRYPDHMKR HDFFKSAMPE
     KLTLKFICTT GKLPVPWPTL VTTLTYGVQC FSRYPDHMKR HDFFKSAMPE
     KLTLKFICTT GKLPVPWPTL VTTFGYGVQC FSRYPDHMKQ HDFFKSAMPE

101  GYVQERTISF KDDGNYKTRA EVKFEGDTLV NRIELKGIDF KEDGNILGHK
     GYVQERTIFF KDDGNYKTRA EVKFEGDTLV NRIELKGIDF KEDGNILGHK
     GYVQERTIFF KDDGNYRTRA EVKFEGDTLV NRIELKGIDF KEDGNILGHK
     GYVQERTISF KDDGNYKTRA EVKFEGDTLV NRIELKGIDF KEDGNILGHK

151  LEYNYNSHNV YITADKQKNG IKANFKIRHN IEDGSVQLAD HYQQNTPIGD
     LEYNYNSHNV YIMADKQKNG IKVNFKIRHN IEDGSVQLAD HYQQNTPIGD
     LEYNYNSHNV YIMADKQKNG IKVNFKTRHN IEDGSVQLAD HYQQNTPIGD
     LEYNYNSHNV YITADKQKNG IKANFKIRHN IEDGSVQLAD HYQQNTPIGD

201  GPVLLPDNHY LSTQSALSKD PNEKRDHMVL LEFVTAAGLV PRGTG
     GPVLLPDNHY LSTQSALSKD PNEKRDHMVL LEFVTAAGIT
     GPVLLPDNHY LSTQSALSKD PNEKRDHMVL LEFVTAAGIT
     GPVLLPDNHY LSTQSALSKD PNEKRDHMVL LEFVTAAGIT
```

Figure 1
Sequence of cyclic GFP in comparison with the structurally visible parts of different crystallized variants of linear GFP. Residues engaged in protein–protein contacts as observed in different crystal structures are highlighted in different colors for each structure. The corresponding sequences are shown in different rows: first row, cyclic GFP (1kp5); second row, 1gfl (Cubitt *et al.*, 1995); third row, 1emc (Palm *et al.*, 1997); fourth row, 1b9c (Battistutta *et al.*, 2000). Lower case indicates residues modeled in the crystal structure as alanines owing to poor electron density.

Table 3
Structural alignment.

Calculations were performed with the program *ALIGN* (Hofmann & Wlodawer, 2002).

Molecule #1 *versus*	Residue range, molecule #1	Positional r.m.s. (Å)	R.m.s. ΔB (Å^2)
Molecule #2 (312–543)	12–243	0.289	8.864
Linear GFP (5–229)†	15–239	0.495	9.143

† PDB code 1emm.

subjected to mass spectrometry and yielded a molecular mass of 27 498 g mol^{-1}, in agreement with a previous report (Iwai *et al.*, 2001).

Each molecule of the dimer is comprised of a regular β-barrel with 11 strands on the outside of the cylinder; this topology was also termed β-can (see Fig. 2). The barrels have a diameter of about 24 Å and a length of approximately 42 Å. A single helix runs along the axis of the cylinder and contains the chromophore almost in the center. This shielding by the β-barrel protects the chromophore against bulk solvent. The structure of the core of the circular protein is identical to linear GFP as reported earlier (Ormö *et al.*, 1996; Yang *et al.*, 1996); the r.m.s. deviation between the cyclic GFP and the structure used as a model for molecular replacement (1emm) is 0.4 Å (see Table 3 and Fig. 3). Even the chromophore conformation shows only slight conformational differences in both structures. The two molecules of the current structure differ only in the conformation of the artificial linker constituted by amino acids 241–245 and 1–15.

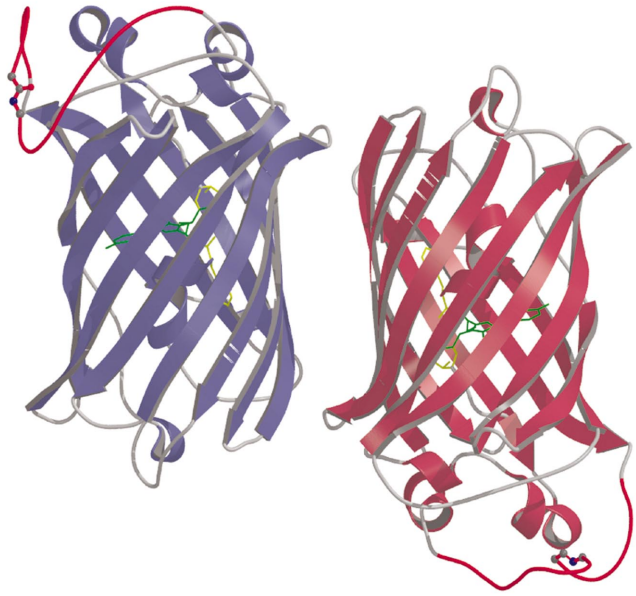

Figure 2
The cGFP dimer. Molecule #1 is colored in blue and molecule #2 in dark red. For each molecule the artificial linker is shown in red and the chromophore as well as the preceding and the following residues are drawn explicitly (green and yellow). The figure was created with *MOLSCRIPT*/*BOBSCRIPT* (Kraulis, 1991; Esnouf, 1999) and rendered with *POVRay* (Persistence of Vision Development Team, 1999).

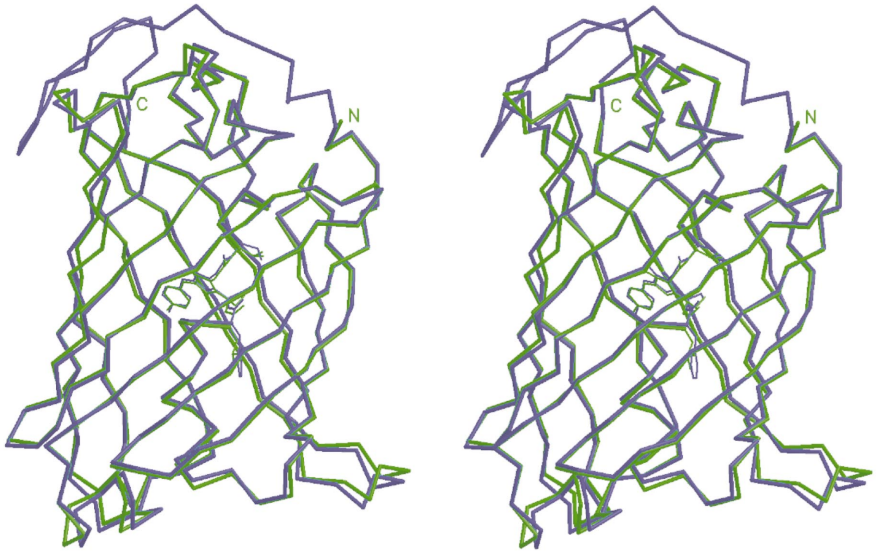

Figure 3
Comparison of cyclic GFP with linear GFP. Superposition of molecule #1 of cGFP (blue) with linear GFP (green) as deposited in the PDB with the accession code 1emm (Palm *et al.*, 1997). The figure was created with *MOLSCRIPT* (Kraulis, 1991) and rendered with *Raster3D* (Merritt & Bacon, 1997).

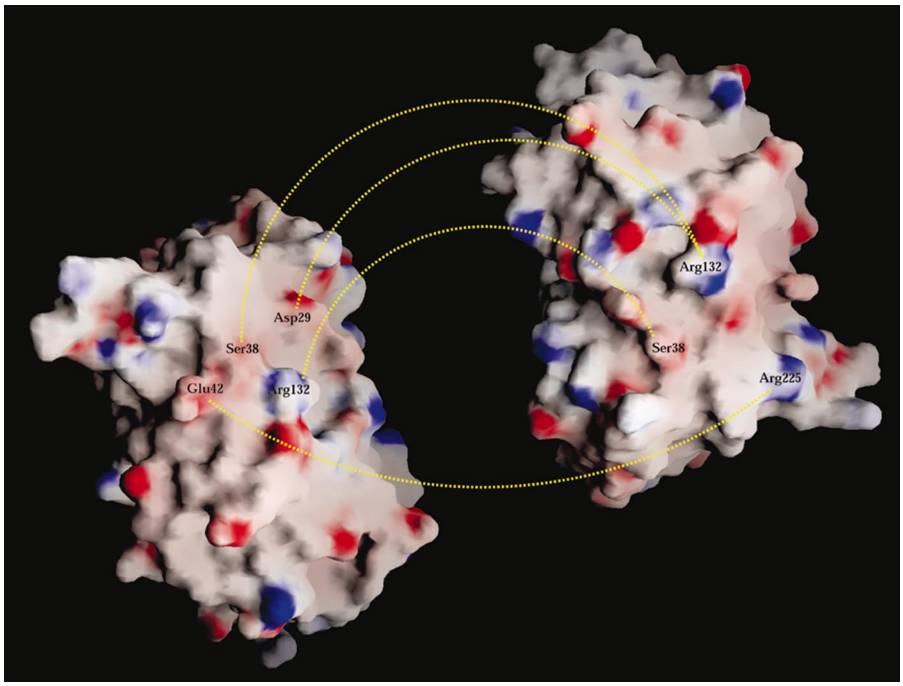

Figure 4
The dimer interface. The explosion figure of the dimer interface shows molecule #1 on the left and molecule #2 on the right. The figure was created with *GRASP* (Nicholls, 1992).

While the density from the first (Lys12) to the last (Gly238) residue of the genuine GFP sequence is excellent, the density in the linker region suffers from rather poor quality. Therefore, we employed composite omit-map procedures as implemented in the *CNS* package (Brünger *et al.*, 1998) and cyclic NCS averaging of electron-density maps after each refinement round. While the omit-map protocols did not improve the density at all, map averaging allowed placement of certain residues and a subsequent tracing of the linker. Parts of the artificial linker were modeled as polyalanine and the occupancy was set to zero for several residues in that region. The linker residues cap the cylinder in a rather loose manner and the apparent absence of any conformational strain allows flexible conformations in that region. In summary, we believe that this part of the structure is highly disordered which prevents us from tracing the backbone of the linker unambiguously.

Nevertheless, even with the model-like character of the linker in mind, it is quite obvious that the conformations in this region are different for both molecules of the present structure. Taking the first and last residues of the genuine GFP sequence as fixed, the linker peptide is easily able to span the distance without any strain, allowing flexible conformations as well as the possibility of annealing along the surface of the core.

3.2. The dimer of GFP

The structure of cyclic GFP in this study features a dimer within the asymmetric unit where both cylinders are aligned antiparallel to each other. The two monomers are held together by polar interactions of residues Asp29, Ser38, Arg132 and Arg225 (Fig. 4). The buried surface area in this dimer is 1345 Å^2, which corresponds to about 6% per molecule. According to Janin (1997), this value already indicates a significant probability that this is not a random but rather a specific dimer interaction. It was indeed reported previously that GFP can form homodimers in crystals as well as in solution (Phillips, 1997). This author also determined an apparent dissociation constant of about 100 μM, which indicates a rather weak dimer formation. The self-association of GFP does have a functional impact, since fluorescence changes have been observed upon dimerization (Wu *et al.*, 1997) and fluorescence resonance energy transfer can occur when the chromophores of two molecules are in spatial vicinity owing to self-association (Mitra *et al.*, 1996).

The comparison of different crystal structures that feature a homodimer shows that diverse dimerization principles are used (see Fig. 1). In at least three structures, the homodimer is

established in a similar fashion involving a core of hydrophobic side chains and several polar interactions (Yang *et al.*, 1996; Wu *et al.*, 1997). However, other homodimeric crystal structures are known where dimer formation is based on different residues or polar interactions only. The current structure of cGFP introduces yet another set of residues responsible for formation of the homodimer. Thus, it is obvious that different surface patches around the cylinder of the β-can of GFP can serve as dimer interfaces. The possibility for GFP to dimerize by invoking different surface areas and the reported weak dissociation constant indicate a rather unspecific dimerization behavior of this protein which, on the other hand, makes the protein an ideal candidate for the formation of large aggregates, as no special orientation is required in order to facilitate protein–protein contacts.

4. Conclusion

As anticipated from previous work (Iwai *et al.*, 2001), the crystal structure of cGFP proves that cyclization of GFP as carried out with the present construct does not impact on the overall structure of this very unique fold. While the cyclic protein is stabilized against denaturation compared with the linear form, this stabilization is not reflected in any changes of the β-can structure and the core conformation of the cyclic protein is the same as observed in crystal structures of linear GFP. The linker peptide caps one side of the β-can and seems to be highly flexible. The lack of particular crystal contacts in that region of the molecule which would fixate the linker residues into a certain position leads to a high degree of disorder and thus makes it difficult to build an accurate model. These observations also explain why the linker does not seem to induce strain into the structure and thus stabilizes GFP even further (Iwai *et al.*, 2001) and the unchanged core structure further strengthens this interpretation. The disorder within the linker region suggests that the linker can be shortened or elongated without serious structural effects.

The current structure was obtained from a space group which is observed for the first time with GFP despite the large number of known crystal structures. The remarkable feature is certainly the ring-like arrangement of protein molecules leading to extremely high water content within the unit cell. As observed in other crystal structures before, cGFP is found to be dimeric with two molecules oriented in an antiparallel fashion. However, a more detailed analysis reveals that the dimer interfaces for different crystal structures are quite variable. This renders GFP self-association quite unspecific and indicates a broad versatility of the cylinder surface for protein–protein contacts.

We thank Zbigniew Dauter (NCI and NSLS, Brookhaven National Laboratory) for help with data collection on beamline X9B and Lewis Pannell (Structural Mass Spectrometry Facility, Laboratory of Bioorganic Chemistry, NIDDK) for his support.

References

Battistutta, R., Negro, A. & Zanotti, G. (2000). *Proteins*, **41**, 429–437.
Botos, I., Wu, Z., Lu, W. & Wlodawer, A. (2001). *FEBS Lett.* **509**, 90–94.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.
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. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Camarero, J. A., Fushman, D., Sato, S., Giriat, I., Cowburn, D., Raleigh, D. P. & Muir, T. W. (2001). *J. Mol. Biol.* **308**, 1045–1062.
Camarero, J. A. & Muir, T. W. (1999). *J. Am. Chem. Soc.* **121**, 5597–5598.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Craik, D. J., Daly, N. L., Bond, T. & Waine, C. (1999). *J. Mol. Biol.* **294**, 1327–1336.
Creighton, T. E. & Goldenberg, D. P. (1984). *J. Mol. Biol.* **179**, 497–526.
Cubitt, A. B., Heim, R., Adams, S. R., Boyd, A. E., Gross, L. A. & Tsien, R. Y. (1995). *Trends Biochem. Sci.* **20**, 448–455.
Daly, N. L., Koltay, A., Gustafson, K. R., Boyd, M. R., Casas-Finet, J. R. & Craik, D. J. (1999). *J. Mol. Biol.* **285**, 333–345.
Esnouf, R. M. (1999). *Acta Cryst.* D**55**, 938–940.
Evans, T. C., Benner, J. & Xu, M. Q. (1999). *J. Biol. Chem.* **274**, 18359–18363.
Evans, T. C., Martin, D., Kolly, R., Panne, D., Sun, L., Ghosh, I., Chen, L., Benner, J., Liu, X. Q. & Xu, M. Q. (2000). *J. Biol. Chem.* **275**, 9091–9094.
Felizmenio-Quimio, M. E., Daly, N. L. & Craik, D. J. (2001). *J. Biol. Chem.* **276**, 22875–22882.
Flory, P. (1956). *J. Am. Chem. Soc.* **78**, 5222–5235.
Goldenberg, D. P. (1985). *J. Cell. Biochem.* **29**, 321–335.
Goldenberg, D. P. & Creighton, T. E. (1984). *J. Mol. Biol.* **179**, 527–545.
Gonzalez, C., Langdon, G. M., Bruix, M., Galvez, A., Valdivia, E., Maqueda, M. & Rico, M. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 11221–11226.
Hofmann, A. & Wlodawer, A. (2002). *Bioinformatics*, **18**, 209–210.
Iwai, H., Lingel, A. & Plückthun, A. (2001). *J. Biol. Chem.* **276**, 16548–16554.
Iwai, H. & Plückthun, A. (1999). *FEBS Lett.* **459**, 166–172.
Janin, J. (1997). *Nature Struct. Biol.* **4**, 973–974.
Jones, T. A., Zou, J. Y., Cowan, S. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Korsinczky, M. L., Schirra, H. J., Rosengren, K. J., West, J., Condie, B. A., Otvos, L., Anderson, M. A. & Craik, D. J. (2001). *J. Mol. Biol.* **311**, 579–591.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Mitra, R. D., Silva, C. M. & Youvan, D. C. (1996). *Gene*, **173**, 13–17.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Nicholls, A. (1992). *GRASP: Graphical Representation and Analysis of Surface Properties*. Columbia University, New York, USA.
Ormö, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y. & Remington, S. J. (1996). *Science*, **273**, 1392–1395.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Palm, G. J., Zdanov, A., Gaitanaris, G. A., Stauber, R., Pavlakis, G. N. & Wlodawer, A. (1997). *Nature Struct. Biol.* **4**, 361–365.
Persistence of Vision Development Team (1999). *POVRay - The Persistence of Vision Raytracer*, http://www.povray.org.
Phillips, G. N. (1997). *Curr. Opin. Struct. Biol.* **7**, 821–827.

Prendergast, F. P. & Mann, K. G. (1978). *Biochemistry*, **17**, 3448–3453.
Saether, O., Craik, D. J., Campbell, I. D., Sletten, K., Juul, J. & Norman, D. G. (1995). *Biochemistry*, **34**, 4147–4158.
Scott, C. P., Abel-Santos, E., Wall, M., Wahnon, D. C. & Benkovic, S. J. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 13638–13643.
Siebold, C., Flukiger, K., Beutler, R. & Erni, B. (2001). *FEBS Lett.* **504**, 104–111.
Tong, L. & Rossmann, M. (1990). *Acta Cryst.* A**46**, 783–792.
Trabi, M. & Craik, D. J. (2002). *Trends Biochem. Sci.* **27**, 132–138.
Williams, N. K., Prosselkov, P., Liepinsh, E., Line, I., Sharipo, A., Littler, D. R., Curmi, P. M., Otting, G. & Dixon, N. E. (2002). *J. Biol. Chem.* **277**, 7790–7798.
Wu, C. K., Liu, Z. J., Rose, J. P., Inouye, S., Tsuji, F., Tsien, R. Y., Remington, S. J. & Wang, B. C. (1997). *Bioluminescence and Chemoluminescence*, edited by J. W. Hastings, L. J. Kricka & P. E. Stanley, pp. 399–402. Chichester: Wiley.
Yang, F., Moss, L. G. & Phillips, G. N. (1996). *Nature Biotechnol.* **14**, 1246–1252.
Yu, Q., Lehrer, R. I. & Tam, J. P. (2000). *J. Biol. Chem.* **275**, 3943–3949.