BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 10773–10779 10773
DOI: 10.1021/bi100975z

A Rewired Green Fluorescent Protein: Folding and Function in a Nonsequential, Noncircular GFP Permutant†

Philippa J. Reeder,‡ Yao-Ming Huang,§ Jonathan S. Dordick,*,§,∥ and Christopher Bystroff*,§

‡*Department of Chemical and Biological Engineering, University of Colorado, Boulder, Colorado 80309, United States,* §*Department of Biology, and* ∥*Department of Chemical and Biological Engineering, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York 12180, United States*

Received June 16, 2010; Revised Manuscript Received November 12, 2010

ABSTRACT: The sequential order of secondary structural elements in proteins affects the folding and activity to an unknown extent. To test the dependence on sequential connectivity, we reconnected secondary structural elements by their solvent-exposed ends, permuting their sequential order, called "rewiring". This new protein design strategy changes the topology of the backbone without changing the core side chain packing arrangement. While circular and noncircular permutations have been observed in protein structures that are not related by sequence homology, to date no one has attempted to rationally design and construct a protein with a sequence that is noncircularly permuted while conserving three-dimensional structure. Herein, we show that green fluorescent protein can be rewired, still functionally fold, and exhibit wild-type fluorescence excitation and emission spectra.

The topology of a protein, defined as the path of the protein backbone through space, is thought to be a primary determinant of the folding pathway (*1*). However, it has been shown that multiple backbone topologies can conserve the same core packing arrangement of secondary structure elements, where core packing arrangement refers to the spatial arrangement of contacting secondary structure elements (*2*). Yuan and Bystroff (*3*) have catalogued many of these cases using nonsequential structure-based alignment, determining, for example, that the nonhomologous proteins alkaline phosphatase [Protein Data Bank (PDB) entry 1ALK] and glycogen phosphorylase (PDB entry 1C8K) conserve a 111-residue structural core after a triple permutation of the sequence order. Agrawal and Kishan (*4*) identified a natural nonsequentially permuted structure between two functionally similar folds (five-strand β-barrels), oligonucleotide/oligosaccharide-binding (OB) and Src homology 3 (SH3) folds, with distinct topologies. Abyzov and Ilyin (*5*) noted that some recurrent packing arrangements in proteins conserve only the packing and not the sequence of secondary structural elements. The overall impression is that tertiary structure is dictated more by specific packing interactions than by topology, a theory also held by others (*6*). If this is true, then it should be possible to vary the topology of a protein in the loop regions, conserving the core side chains and retaining the native three-dimensional structure.

Along these lines, Regan (*7*) documented several studies involving loop redesign and circular permutations, observing that protein structures tolerate a remarkable number of insertions and other changes made at the topological level yet are still able to fold to stable and active structures. For example, Nagi et al. inserted loops of increasing size between two α-helices of the four-α-helix bundle protein, Rop, and observed a delay in the association of adjacent helices in the ordered packing of the folded state. These results highlight the relative importance of specific packing interactions in defining the fold of a protein. This is in opposition to the generally accepted model (*8*) in which the fold is defined by nonspecific forces of hydrophobic collapse combined with the topological constraints imposed by the ordering of secondary structural elements along the chain.

Many studies have shown that protein structural subdomains can associate in vivo and in vitro without the aid of peptidic bonds (*9*): for example, β-galactosidase and LacZ, used in blue-white screening (*10*, *11*), and the complementation of green fluorescent protein (GFP) β-strand 11 to its fully folded counterpart, β-strands 1−10, used for protein tagging and folding applications (*12*−*14*). In addition, Huang et al. (*15*) have previously shown that circularly permuted GFP can undergo complementation with β-strand 7, folding and binding simultaneously.

On the basis of this observed robustness of protein self-association, especially in GFP, we pose the following questions. (1) Can we change the topology of a protein without changing its core? (2) If so, how will such topological changes affect protein activity and stability? GFP has been chosen for the model system for this study having a high-resolution crystal structure (*16*), a well-studied folding pathway (*17*−*19*), and a chromophore maturation reaction (*20*), and because its intrinsic fluorescence is a convenient reporter of its folded state (*15*, *18*−*23*).

The model protein, GFP, is a highly stable β-barrel structure consisting of 11 β-strands connected by loops (*24*, *25*) that harbors a distorted α-helix containing the fluorescent chromophore (Figure 1A,B). GFP variants with improved folding properties have been generated: for example, "cycle3" GFP with three mutations, "folding reporter" with five mutations, "superfolder" with 11, and "OPT" with 16 (*22*, *26*). As a starting point for

†This work was supported by grants from the National Science Foundation (DBI 0448072, C.B.), the National Institutes of Health (GM88838, C.B.; GM66712, J.S.D.) and the New York State Foundation for Science, Technology and Innovation (NYSTAR, J.S.D.).

*To whom correspondence should be addressed. J.S.D.: Rensselaer Polytechnic Institute, 2213 Biotechnology, 110 8th St., Troy, NY 12180; phone, (518) 276-2899; fax, (518) 276-2207; e-mail, dordick@rpi.edu. C.B.: Rensselaer Polytechnic Institute, Science Center, 1st Floor, 110 8th St., Troy, NY 12180; phone, (518) 276-3185; fax, (518) 276-2344; e-mail, bystrc@rpi.edu.

Published on Web 11/23/2010

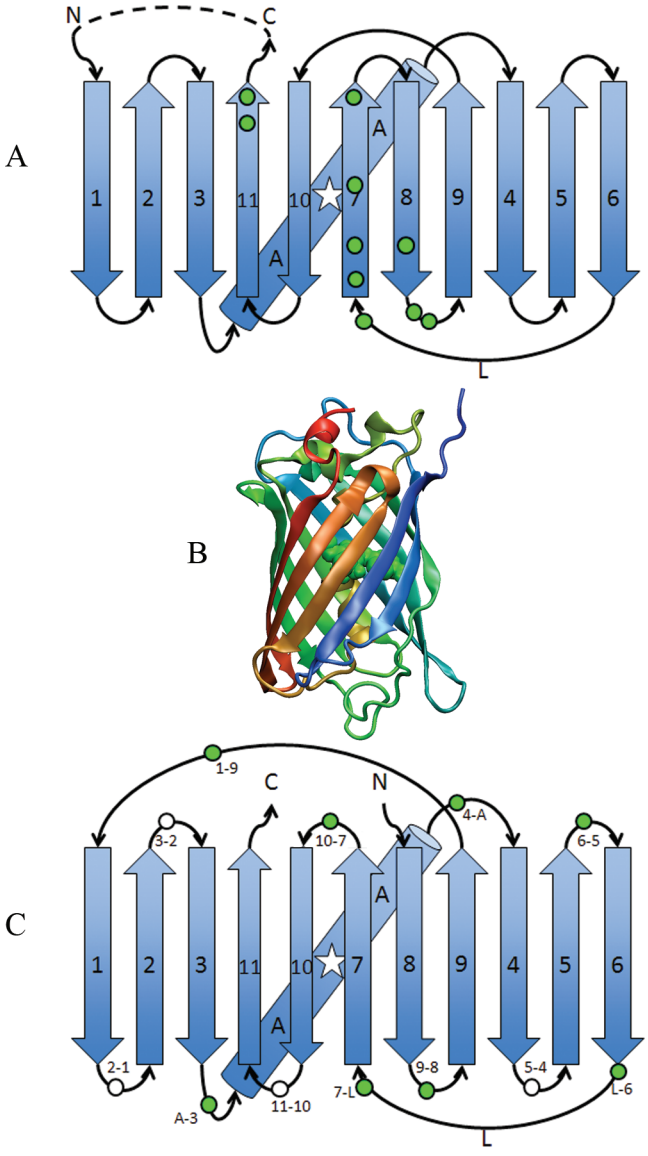

FIGURE 1: Structural makeup of GFP, rGFP3, and CP-rGFP3. (A) Strand order of GFP. β-Strands are drawn as arrows and numbered in N to C order, and the α-helices are drawn as cylinders. The study of Baird et al. is summarized by the green circles that identify locations of allowed circular permutations in wild-type GFP; the dotted line indicates the circular permutation introduced during this study (GGTGGS) (*21*). The peptidic chromophore is represented by a star. (B) Crystal structure of GFP (PDB entry 2b3q) in N (blue) to C (red) rainbow coloring with the chromophore shown in green space-filling mode. Rendered in VMD (*56*). (C) Strand order of rGFP3 with strand numbering maintained from GFP. Circles indicate the location of circular permutants analyzed in this study, colored white for nonfluorescent or weakly fluorescent constructs and green for highly fluorescent constructs. Circles are labeled to indicate the name of that construct, in the format CP N-C from Table 2.

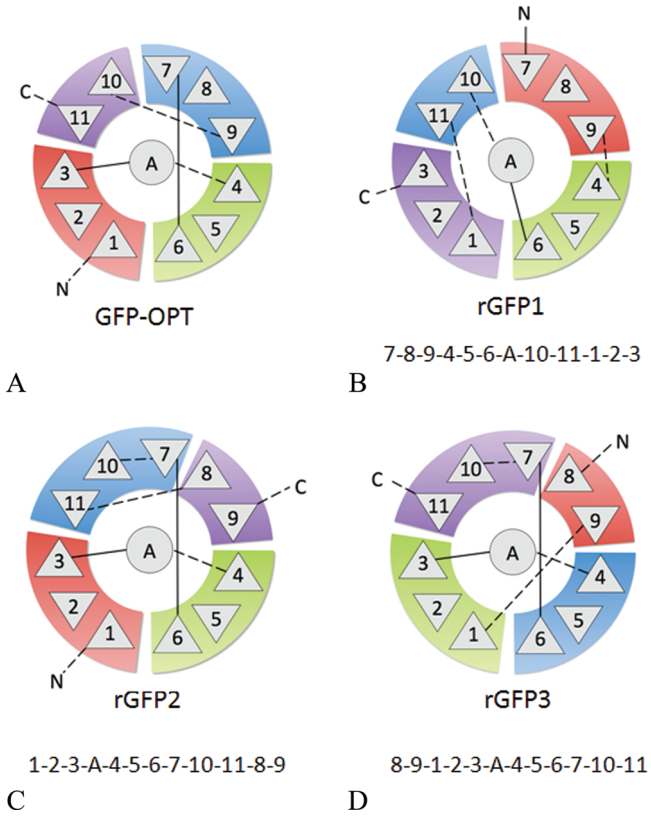

FIGURE 2: TOPS (*57*) cartoons of wild-type GFP and three rewired versions. β-Strands are shown as triangles, with the orientation indicating chain direction into (▽) or out of (△) the page. β-Strands are grouped as three β-meanders and a hairpin turn in colored blocks, ordered from N to C (red, green, blue, and then purple). The α-helix is shown as a circle in the center.

modeling, we have used the latter. GFP-OPT (herein simply GFP) emits green light (λ_{max} = 508 nm) under the excitation of cyan light (λ_{max} = 485 nm). The chromophore in GFP is derived from posttranslational, intramolecular cyclization and oxidation of the tripeptide motif Thr65 or Ser65-Tyr66-Gly67 (*27*) by a proposed autocatalytic mechanism (*28*) consisting of four distinct steps: folding, cyclization, oxidation, and dehydration. Positionally conserved, well-ordered intramolecular water molecules are located in the interior cavity (*28*, *29*), held in place by a proton relay system involving Asn146, His148, Arg168, Thr203, Ser205, and Glu222, all of which are part of the β-barrel. Mutations are tolerated at most positions in this network with regard to protein folding, but any small change in this well-organized network usually results in a shift of the GFP excitation or emission wavelengths (*17*, *23*). The chromophore, once formed, does not revert to a tripeptide or undergo any other covalent changes upon denaturation, using urea or if the pH is lowered to 2. It is nonetheless completely quenched in the unfolded state, presumably through nonradiative decay of the excited state through the solvent. The return of fluorescence upon refolding is immediate and has been used as a probe of the refolding pathway (*15*, *17*).

Nonsequential, noncircular sequence permutation by reconnecting loops, or "rewiring", is a new protein engineering approach that potentially would allow one to explore the effects of topological changes on protein folding, stability, and activity without changing the core interactions. In this study, rewiring GFP is shown to conserve the detailed core packing interactions as reported by the fluorescence spectrum. Other protein engineering and design techniques include rational single- and multiple-site mutations (*30*−*34*), a combinatorial method (*35*, *36*), in vitro evolution (*37*−*40*), and computational optimization of side chain packing (*41*), but none of these methods has the capacity to permute the protein sequence. We expect our method to be widely applicable, because sequence rearrangements are observed in known protein structures (*2*, *3*, *5*, *42*). Potentially, rewiring may be used to engineer the stability and folding kinetics of proteins, because topological properties are known to influence the rates of folding (*43*) and possibly unfolding (*44*).

Table 1: Summary of Loop Design

construct	loop location (β-strands)[a]	loop sequence	design comments
GFP-OPT	1-2-3-A-4-5-6-L-7-8-9-10-11		native
rGFP1	7-8-**9-4**-5-6-L-A-10-11-1-2-3	FEGDTL	same sequence as the 5−6 loop[b]
	7-8-9-4-5-6-**L-A**-10-11-1-2-3	MPGDG	I-sites[c] motif, proline α-C-cap
	7-8-9-4-5-6-L-**A-10**-11-1-2-3	EKG	design based on local interactions[d]
	7-8-9-4-5-6-L-A-10-**11-1**-2-3	GGTGGS	used in ref *21*
rGFP2	1-2-3-A-4-5-6-L-**7-10**-11-8-9	DGGV	design based on local interactions[d]
	1-2-3-A-4-5-6-L-7-10-**11-8**-9	G-GDGPKLVPD-S	9−10 loop with changes[e]
rGFP3	8-**9-1**-2-3-A-4-5-6-L-7-10-11	SGTGSG	loose, flexible loop[f]
	8-9-1-2-3-A-4-5-6-L-**7-10**-11	GGSGGT	loose, flexible loop[f]
rGFP2b (CP 1-9)	1-2-3-A-4-5-6-L-**7-10**-11-8-9	GGSGGT	loose, flexible loop[f]
	1-2-3-A-4-5-6-L-7-10-**11-8**-9	GGTGS	loose, flexible loop[f]
rGFP1b	7-8-**9-4**-5-6-L-A-10-11-1-2-3	GSGTGSG	loose, flexible loop[f]
	7-8-9-4-5-6-**L-A**-10-11-1-2-3	GGG	loose, flexible loop[f]
	7-8-9-4-5-6-L-**A-10**-11-1-2-3	GGDGG	loose, flexible loop[f]
	7-8-9-4-5-6-L-A-10-**11-1**-2-3	GGTGGS	loose, flexible loop[f]

[a]β-Strands are labeled according to the native ordering of GFP and highlighted in bold in the location of the loop sequence. [b]The FEGDTL loop was reused as the atom−atom distance between the final two amino acids on the associated β-strands was < 1 Å. Codon changes were introduced to enable gene assembly by assembly polymerase chain reaction. [c]I-Sites protein structure library (http://www.bioinfo.rpi.edu/bystrc/Isites2/) chosen on the basis of atom−atom distances at the associated β-strands. [d]These short linkages were chosen by analysis in MOE on the basis of the criteria that any loop should be as short as possible and closely associated with local structure. Thus, for EKG, Glu and Lys interact on the basis of charge, negative and positive, respectively, and the Gly faces outward and is small and polar. For DGGV, two Gly residues, which are small and flexible, interact with Asp and Val, which are negatively charged and hydrophobic. [e]The GDGPVLLPDN loop was reused with changes to GGDGPKLVPDS based on local structural interactions in the new location as follows. (1) The Lys residue replaces the Val as a larger residue (by Van de Waals volume, in cubic angstroms) as well as positively charged rather than hydrophobic (it faces outward in the new configuration). (2) The Val replaces the Leu as a smaller hydrophobic residue for the fit. (3) The Ser replaces the Asn as a smaller residue. (4) The Gly is inserted into the sequence to encourage the close fit of the new loop based on length. [f]For the final set of designs, we found that the sequence did not matter if the loop was adequately long to inhibit interaction with the protein structure (> 3 Å from the surface after MOE energy minimization) and made of small, hydrophilic, and flexible residues (Ser, Thr, and Gly) to encourage solubility.

RESULTS AND DISCUSSION

Characterization of rGFP1 and rGFP2. Rewiring GFP was initiated with a conservative design strategy in which changes were restricted to short reconnections, without dividing the three β-meanders (β-meanders are three-strand antiparallel sheets) or the 10−11 hairpin turn, to preserve the wild-type supersecondary structure elements as much as possible (Figure 2A). The direction of the β-strands was also preserved. Given these constraints, only a single nonsequential permutation was geometrically possible (Figure 2B). rGFP1 was generated by removing four existing loops and adding four de novo linker sequences (Table 1). Modeling was conducted using the Molecular Operating Environment (MOE, CCG, Montreal, QC) (*45*). rGFP1 exhibited no fluorescence and no mature chromophore absorption peak at 380 nm when acid-denatured (data not shown). Circular dichroism (CD) of this construct shows ~50% less β-strand signal (218 nm) relative to GFP (Figure 3), suggesting a fast dynamic equilibrium between folded and unfolded states or the misfolded state. A fast dynamic equilibrium would prevent efficient formation of the chromophore, a slow autocatalyzed reaction with a half-life on the order of minutes to hours. However, its expression as a soluble protein suggests that rGFP1 folds to a compact structure.

The loss of chromophore maturation in rGFP1 effectively refuted the assumption that retaining the supersecondary structures was a "conservative" design strategy. It suggested that larger structural units must be topologically conserved for folding and function. This was not unexpected, because previous studies have implicated connections and/or ordering of the first six strands (including the central helix), as required for foldability (*21*, *46*). Demidov et al. showed that a fragment of GFP containing only strands 1−6 is capable of forming a mature chromophore when expressed with the remaining strands as a separate chain (*47*). Furthermore, Baird et al. (*21*, *48*) have conducted exhaustive circular permutations of wild-type GFP, effectively locating all

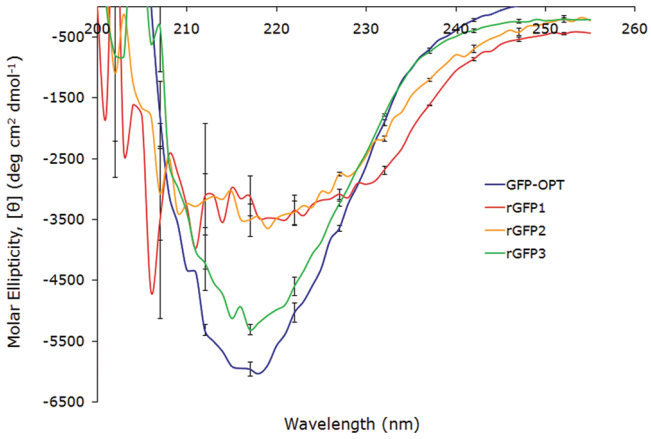

FIGURE 3: Circular dichroism of GFP and all rewire constructs, obtained in 2 mM phosphate buffer (pH 7.4).

covalent linkages that can be cleaved without a loss of foldability or fluorescence, and they found no locations in strands 1−6 that could serve as the new termini for a functional circular permutant. A similar analysis by Pedelacq et al., using "folding reporter" GFP, found that placement of the termini only in positions located after strand 6 (specifically, between β-strands 6 and 7 and β-strands 8 and 9) yielded whole cell fluorescence that was greater than 10% of that of the native protein (*22*). However, more recently, Kent et al. (*49*) have shown that the helix, located between strands 3 and 4 in the sequence, can be left out entirely, and folding will occur when it is added exogenously. In rGFP1, we linked the interior helix to strand 10 instead of strand 4, breaking up the first six strands, resulting in a non-native state. This supports the requirement of a larger unpermuted unit, which may or may not include the helix.

In our next design, we focused on permuting only strands 7−11. If strands 1−6 and the helix are conserved and all strand

directions are retained, there are a total of 12 possible topologies, including the wild type. The most conservative of these, rGFP2 (Figure 2C), is an approximate mirror image of the wild-type topology (Figure 2A), with two new linkers (see Table 1 for loop details and Table 1S of the Supporting Information for all sequences) and a minimal change in the sequential strand order. However, like rGFP1, rGFP2 was soluble but not fluorescent (data not shown), consistent with a non-native or misfolded state, because a natively folded state would have fluorescence. On the basis of the CD spectra, rGFP1 and rGFP2 contain less β-secondary structure than the native state.

To ask whether the rGFP2 topology was intrinsically unstable or whether the designed loops led to instability, we took the approach of Nagi et al. (*8*), increasing the linker loop lengths and making them more flexible. Rationally designed linker sequences were replaced with glycine-rich sequences, such as those used previously to link the termini in circular permutants of GFP (*21*, *22*). In the new construct, rGFP3, the sequence order of rGFP2 was modified by circular permutation in which the terminal strands, 9 and 1, were connected by a flexible linker (SGTGSG) and the linker connecting strand 11 and 8 was cleaved to make the new termini (Figure 2D). Additionally, the designed three-residue tight turn between strands 7 and 10 in rGFP2 was relaxed into a more flexible six-residue loop (GGSGGT) in rGFP3 (Table 1). In this design, the linkers are assumed to be more unstructured overall than they were in rGFP1 and rGFP2. This construct was soluble and fluorescent. We conclude from this experiment that the rGFP2 topology is not intrinsically unstable but that the designed loops in rGFP2 favored uncharacterizable non-native states in the structure.

Characterization of rGFP3. The rewired permutant rGFP3 exhibits fluorescence excitation (485 nm) and emission (508 nm) maxima that are indistinguishable from those of GFP (Figure 4). The similar Stokes shift suggests that the barrel in rGFP3 is formed in a manner identical to that of the native GFP. Both unfold at low pH and renature upon returning to neutral pH (Figure 1S). rGFP3 has approximately 88% of the CD signal at 218 nm [β-sheet (Figure 3)], much higher than that of either rGFP1 or rGFP2. The relative quantum yield, calculated as the ratio of emission at 508 nm to absorbance at 485 nm at pH 7.5, is essentially the same for GFP and rGFP3, as is the ratio of chromophore absorbance at 485 nm (pH 7.5, folded) to absorbance at 380 nm (pH 2, unfolded). However, rGFP3 has between 80 and 90% of the GFP fluorescence when corrected for protein concentration (Figure 4). Combining these observations, we conclude that a small fraction (10−20%) of the rGFP3 molecules failed to form the chromophore. This uncharacterized misfolded state may be homo- or heterogeneous in nature. Andrews et al. (*50*) have described such a monomeric, misfolded state of GFP that is trapped during refolding and is kinetically stable. It is possible that we are seeing a similar trapped state in rGFP3, preventing chromophore maturation in a small fraction. Alternately, the missing chromophore signal could be due to chemical modifications, such as spontaneuos side chain cleavage at Y66 (*51*).

Thermal deactivation of GFP and rGFP3 was also indistinguishable (Figure 5), and given that it depends on the rate of unfolding, this suggests that there is no change in the hydrogen bonding within the β-barrel and no changes in core packing, either of which would affect the kinetic stability.

Characterization of Circular Permutants of rGFP3. To study further the folding pathway of rGFP3 and GFP, we created and characterized comprehensive circular permutants (CPs) of

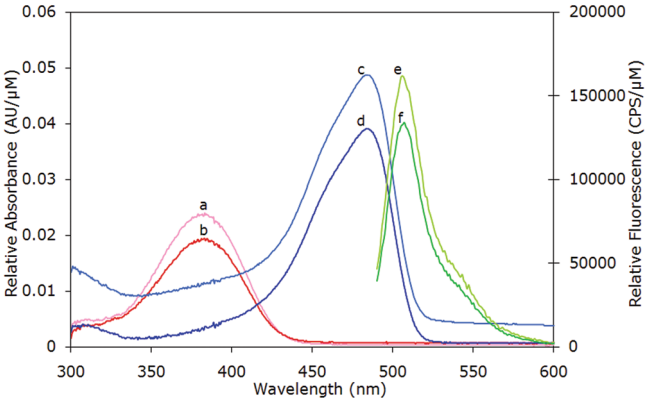

FIGURE 4: Absorbance and fluorescence characterization of rGFP3 and GFP. Absorbance (300−600 nm, 1 nm intervals, 2 nm slit width) and fluorescence emission (excitation at 485 nm, emission at 490−600 nm, 1 nm intervals, 3 and 1 nm excitation and emission slit widths, respectively) profiles of rGFP3 and GFP, normalized to concentration (10 or 0.1 μM for absorbance or fluorescence, respectively). (a and b) Absorbance of the denatured chromophore at pH 2: GFP and rGFP3, respectively. (c and d) Absorbance of the chromophore at pH 7.5: GFP and rGFP3, respectively. (e and f) Fluorescence emission by excitation at 485 nm at pH 7.5: GFP and rGFP3, respectively.

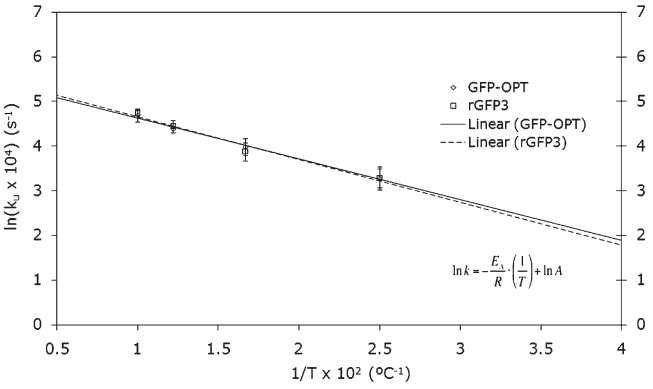

FIGURE 5: Thermal deactivation of rGFP3 vs GFP. Initial rates of fluorescence quenching were measured for four temperatures (40, 60, 80, and 100 °C). The Arrhenius relationship, shown in the figure, was plotted to determine the difference in the activation energy term between the two species extrapolated to room temperature ($1/T \times 10^2$ equaling 4.0 $°C^{-1}$).

rGFP3, with new N- and C-termini placed at the intersection of each secondary structure in the chain, including breaks on either end of the long 6−7 loop (L in Table 2 and Figure 1C). This resulted in 12 new constructs, including one, CP 1-9, that mirrors the topology of rGFP2 (Table 1S of the Supporting Information). Table 2 details the in vivo solubility and fluorescence for each construct. Two constructs with 0% solubility were solubilized with urea (Materials and Methods) and subsequently observed to be fluorescent. It is interesting to note that we found viable chain terminus locations within the first seven secondary structure elements, strands 1−6, whereas Pedalacq et al., working in "folding reporter" GFP, did not (*21*, *22*). Our template, GFP, includes a few stabilizing mutations beyond those in the folding reporter (Table 1S of the Supporting Information), perhaps slowing the rate of aggregation for slow folding permutants. Alternatively, the nonsequential permutations and flexible linkers in rGFP3 may have altered off-pathway intermediates and misfolded states. Many of the permutants whose termini are within strands 1−6 are indeed marginally fluorescent or completely dark, including CPs 2-1, 3-2, and 5-4; however, several, CPs A-3, 4-A,

Table 2: Characterization of rGFP3 Circular Permutants[a]

construct	sequence oqrder	fluorescent	% solubility
GFP-OPT	1-2-3-A-4-5-6-L-7-8-9-10-11	yes	80
rGFP3	8-9-1-2-3-A-4-5-6-L-7-10-11	yes	80
CP 9-8	9-1-2-3-A-4-5-6-L-7-10-11-8	yes	75
CP 1-9	1-2-3-A-4-5-6-L-7-10-11-8-9	yes	67
CP 2-1	2-3-A-4-5-6-L-7-10-11-8-9-1	yes	20
CP 3-2	3-A-4-5-6-L-7-10-11-8-9-1-2	yes	0
CP A-3	A-4-5-6-L-7-10-11-8-9-1-2-3	yes	67
CP 4-A	4-5-6-L-7-10-11-8-9-1-2-3-A	yes	50
CP 5-4	5-6-L-7-10-11-8-9-1-2-3-A-4	no	0
CP 6-5	6-L-7-10-11-8-9-1-2-3-A-4-5	yes	67
CP 6-L	L-7-10-11-8-9-1-2-3-A-4-5-6	yes	100
CP L-7	7-10-11-8-9-1-2-3-A-4-5-6-L	yes	40
CP 10-7	10-11-8-9-1-2-3-A-4-5-6-L-7	yes	80
CP 11-10	11-8-9-1-2-3-A-4-5-6-L-7-10	no	20

[a]Summary of circular permutant expression data based on fluorescence data of purified constructs and visual inspection of sodium dodecyl sulfate−polyacrylamide gel electrophoresis of lysates (soluble) and pellets (insoluble) following French press (% soluble is defined by relative amount soluble to insoluble). In two cases, CP 3-2 and CP 5-4, the insoluble pellet protein was denatured and then allowed to refold on Ni-agarose column media (4 h at room temperature while the His tag immobilized on the column) before the fluorescence was reassessed. Sequence order refers to the secondary structure naming, based on the sequence of GFP, where numbers indicate β-strands going from the N- to C-terminus, A is the α-helix, and L is the long unstructured loop connecting strands 6 and 7.

and 6-5, were strongly fluorescent (Figure 1C). Further study of the folding pathways of these constructs may reveal alternative folding pathways with altered kinetic phases.

Surprisingly, the location that Cabantous et al. utilized with great success in GFP-superfolder to create a complementation system (*26*), CP 11-10, yielded very little soluble protein in the context of rGFP3. Its misfolding signals a change in the folding pathway that is still not characterized. Possibly the nonspecific association of the inserted 7−10 and 11−8 loops, both loose and long and on the same end of the barrel, confounds folding in this permutation.

Also surprising was the fact that circular permutants involving the central α-helix formed soluble and fluorescent products (CP A-3 and 4-A), contrary to the circular permutations in "folding reporter" GFP reported by Pedelacq et al. (*22*), who saw marginal fluorescence in whole cells with analogous CP-GFP constructs, but consistent with the results of Kent et al. (*49*), who showed that the helix can be left out and added exogenously reconstituting the fluorescent form.

Finally, CP 1-9, topologically identical to rGFP2, is made fluorescent and soluble by the addition of loose linkers (Table 1), again pointing to overly constrained loops, not topology, as the main reason for its misfolding. A loose linker introduced between strands 7 and 10 replaced the short DGGV loop in rGFP2. The short linker may have introduced an early folding event, the formation of a 7−10 hairpin, that should have been late. With increases in the loop length and the corresponding entropy of loop closure, the association between strands 7 and 10 would be slowed, as shown in other systems (*8*), making it a later folding event and more consistent with the proposed native sequence of folding events.

After finding that unstructured loops overcame the folding difficulties of rGFP2 (CP 1-9), we similarly revisited rGFP1 by increasing the four loop lengths and by introducing flexible glycines into the loops (Table 1). This looser construct was again soluble but not fluorescent, indicating that loop inflexibility was not the cause of misfolding in this case. Instead, some aspect of the rGFP1 topology must have led to a compact and stable non-native state.

It should be noted that full loop modeling algorithms [such as MODELLER (*52*)] were not used in this study. As a result, some of the initial failures (rGFP2, not rGFP1) may have been due to the use of loops with the wrong structural propensities. Using glycine-rich flexible loops, having no specific structural tendencies, is naturally better than using loops of the wrong structure, but loops with the correct structural propensities would have been better still. Further, we base our major conclusions on the presence of green fluorescence, which requires the autocatalyzed synthesis of the intrinsic chromophore to take place as in the native GFP. Our assumption is that the natively packed core residues are sufficient for this chromophore maturation reaction and that no aspect of the kinetics of folding, or the series of events that we call a folding pathway, is required for this maturation to occur.

CONCLUSIONS

In these case studies, we have explored rewired GFPs in which the connectivity of secondary structure elements was altered but the core packing arrangement was not disturbed. In one case (rGFP3 and some of its circular permutants), we saw that the native packing could be attained during folding, as signaled by green fluorescence, the CD spectrum, and other biophysical similarities. The same topology failed to fold if we used inflexible (or poorly chosen) linkers (rGFP2). Finally, a more severely altered topology did not fold correctly even if flexible linkers were used (rGFP1), as signaled by the lack of fluorescence and significant differences in the CD spectrum. The latter result suggests that for GFP's particular spatial arrangement of secondary structure elements, there exists at least one rewired connectivity that has no pathway to the natively packed state.

This rewiring method can likely be generalized to other proteins. We propose that by studying which sequential rearrangements are possible and which are not, it is possible to deduce the folding pathways of proteins and, further, to engineer them to have increased stability and improved folding kinetics.

MATERIALS AND METHODS

Computational Modeling of rGFP. New structural models were constructed using Molecular Operating Environment (MOE), including point mutations, sequence permutations, loop structure, and loop sequence design (*45*). Loop lengths were determined by estimating the distance to be spanned. Loop sequences were either selected on the basis of a motif library (*53*) or set to a "loose linker" sequence such as GGSGGT (see the Supporting Information for details). Database searches were used to build loose linker loop conformations. Loop structures, but not the conserved secondary structure elements, were energy-minimized under vacuum using a molecular mechanics force field. Buried surface areas were calculated using MASKER (*54*).

Construction of Genes and Plasmids. Genes of rewired constructs were constructed by assembly polymerase chain reaction (PCR) (*55*), described here briefly. Overlapping oligomers for each gene were designed using DNAworks version 2.0 (http://mcl1.ncifcrf.gov/dnaworks/dnaworks2.html) with the following parameters: 40-nucleotide oligo, codon frequency threshold of 20, 50 mM Na^+/K^+, oligo concentration of 200 nM, annealing temperature of 64 °C, one solution, and 2 mM Mg^{2+}.

Assembly PCR was performed first with each oligomer at 0.2 ng/mL (www.idtdna.com), Pfu Turbo (Stratagene), an annealing temperature of 59 °C (5 °C below the set point temperature in DNAworks), and 35 cycles. Of this material, 1 μL was used as the template for amplification PCR with forward and reverse primers (the outermost oligomers on each DNA strand) each at 0.2 μM, annealing temperature of 62 °C, and 35 cycles.

Genes of circular permutants of rGFP3 were constructed by blunt end ligation (T4 kinase reaction to provide phosphate groups for ligation followed by T4 ligase reaction, NEB enzymes) of the rGFP3 gene to form a circular template for PCR. New 5′ and 3′ primers were designed for each circular permutant to include new EcoRI and NheI sites as well as stop codon followed by PCR performed as described above (Pfu Turbo from Stratagene).

Completed genes with included NheI and EcoRI [enzymes from New England Biosciences (NEB)] restriction sites were inserted into pET28a (Novagen) using T4 ligase per the manufacturer's instructions (NEB) and transformed into XL-1Blue or BL21 for DNA sequencing (www.mclab.com) and BL21 for expression (Invitrogen). DNA was collected using a Sigma GeneElute Plasmid Miniprep kit and buffer exchanged using a Promega Wizard SV Gel and PCR-cleanup System kit.

Protein Expression and Purification. Sequenced plasmids were transformed into BL21 cells and grown in LB medium (Qbiogene) with 50 μg/mL kanamycin (Sigma) to an OD_{600} of 0.8−1.0 at 37 °C with 200 rpm shaking before induction with 1 mM IPTG for further incubation at room temperature (20−22 °C) with 200 rpm shaking for 18−24 h. Cells were collected by centrifugation (10 min at 7700*g*), washed, and resuspended in 50 mM Na_2HPO_4, 50 mM Tris, 300 mM NaCl, and 10 mM imidazole (pH 7.5) (NBB, 8 mL/g cells, chemicals from Sigma). A French press was used to lyse the cells (with the addition of 1 μg/mL DNase I, Sigma) in a 30 mL cell, 1800 psi with two passes. Debris was removed by centrifugation at 4000 rpm for 15 min at 4 °C. Supernatants were added and bound to a Ni-NTA column (2 mL of resin/g of cells, Invitrogen) equilibrated with NBB, washed with NWB (NBB with 20 mM imidazole) for 4 column volumes, and eluted into NEB (NBB with 200 mM imidazole). Insoluble pellets were resuspended in DBB (NBB with 8 M urea, no imidazole), bound to a pre-equilibrated Ni-NTA column, slowly washed with step dilutions of DWB (NWB with 8 M urea, no imidazole) into NWB at room temperature, and finally eluted into NEB. Fluorescence was observed in some cases following these washes after at least 90 min (chomophore maturation time). Protein was dialyzed into 50 mM Tris and 100 mM NaCl (pH 7.5) (TN, Sigma) using Pierce 1 mL, 10000 MWCO Slide-dialyzer cassettes in 3.5 L of buffer overnight at 4 °C with gentle stirring.

Concentration and Fluorescence Measurements. Protein concentrations were determined using the BCA assay (Pierce). Fluorescence was measured using a Jorbin Yvon-Horiba Fluorlog-3 machine. Samples were diluted to 0.1 μM in TN, and the excitation wavelength was determined by scanning in 1 nm intervals between 200 and 500 nm with slit widths set at 3 and 1 nm for excitation and emission, respectively, monitoring emission at 508 nm. After the dominant excitation wavelength had been determined, the profile of emission was measured by excitation at 485 nm and emission scanning at 490−600 nm, in 1 nm intervals, with slit widths set at 3 and 1 nm for excitation and emission, respectively. Absorbance (both absorbance of the denatured chromophore at 380 nm and absorbance of the active chromophore at 485 nm) of rGFP3 and control GFP was characterized by dilution of the protein to 5.5 μM in TN buffer and measurement of absorbance from 300 to 600 nm on a Shimadzu UV−vis spectrometer (2 nm slit width, 1 cm path length). HCl was used to lower the pH to 2 for mature chromophore determination over the same absorbance wavelengths with the same sample.

Circular Dichroism. Samples for CD were prepared in 2 mM sodium phosphate buffer (pH 7.5) and analyzed on a spectrometer (OLIS DSM-10 CD, Online Instruments) in a 10 mm path length quartz cuvette at concentrations between 30 and 300 μg/mL. Ellipticity was measured from 255 to 180 nm at room temperature. All measurements were performed in triplicate and normalized to concentration to determine molar ellipticity (degrees square centimeters per decimole).

Thermal Unfolding. Thermal unfolding was monitored by fluorescence. The initial rate of fluorescence quenching was measured for four temperatures (40, 60, 80, and 100 °C) over 20 s in triplicate [485 and 508 nm for excitation and emission, respectively; 1.5 nm excitation and emission slit widths; 0.1 μM protein in 25 mM Tris-HCl and 25 mM NaCl (pH 7), Sigma].

SUPPORTING INFORMATION AVAILABLE

Sequence information for all constructs used in this study (Table 1S). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Baker, D. (2000) A surprising simplicity to protein folding. *Nature 405*, 39–42.
2. Efimov, A. V. (1997) Structural trees for protein superfamilies. *Proteins 28*, 241–260.
3. Yuan, X., and Bystroff, C. (2005) Non-sequential structure-based alignments reveal topology-independent core packing arrangements in proteins. *Bioinformatics 21*, 1010–1019.
4. Agrawal, V., and Kishan, R. K. (2001) Functional evolution of two subtly different (similar) folds. *BMC Struct. Biol. 1*, 5.
5. Abyzov, A., and Ilyin, V. A. (2007) A comprehensive analysis of non-sequential alignments between all protein structures. *BMC Struct. Biol. 7*, 78.
6. Grantcharova, V., Alm, E. J., Baker, D., and Horwich, A. L. (2001) Mechanisms of protein folding. *Curr. Opin. Struct. Biol. 11*, 70–82.
7. Regan, L. (1999) Protein redesign. *Curr. Opin. Struct. Biol. 9*, 494–499.
8. Nagi, A. D., and Regan, L. (1997) An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold Des. 2*, 67–75.
9. Morell, M., Ventura, S., and Aviles, F. X. (2009) Protein complementation assays: Approaches for the in vivo analysis of protein interactions. *FEBS Lett. 583*, 1684–1691.
10. Ruther, U. (1980) Construction and properties of a new cloning vehicle, allowing direct screening for recombinant plasmids. *Mol. Gen. Genet. 178*, 475–477.
11. Ullmann, A., Jacob, F., and Monod, J. (1967) Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the β-galactosidase structural gene of *Escherichia coli*. *J. Mol. Biol. 24*, 339–343.
12. Cabantous, S., and Waldo, G. S. (2006) In vivo and in vitro protein solubility assays using split GFP. *Nat. Methods 3*, 845–854.
13. Ghosh, I., Hamilton, A. D., and Regan, L. (2000) Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein. *J. Am. Chem. Soc. 122*, 5658–5659.
14. Hu, C. D., Chinenov, Y., and Kerppola, T. K. (2002) Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol. Cell 9*, 789–798.
15. Huang, Y. M., and Bystroff, C. (2009) Complementation and reconstitution of fluorescence from truncated circularly permuted green fluorescent protein. *Biochemistry 48*, 929–940.
16. Wood, T. I., Barondeau, D. P., Hitomi, C., Kassmann, C. J., Tainer, J. A., and Getzoff, E. D. (2005) Defining the role of arginine 96 in green fluorescent protein fluorophore biosynthesis. *Biochemistry 44*, 16211–16220.
17. Enoki, S., Saeki, K., Maki, K., and Kuwajima, K. (2004) Acid denaturation and refolding of green fluorescent protein. *Biochemistry 43*, 14238–14248.

18. Sanders, J. K., and Jackson, S. E. (2009) The discovery and development of the green fluorescent protein, GFP. *Chem. Soc. Rev. 38*, 2821–2822.
19. Zimmer, M. (2002) Green fluorescent protein (GFP): Applications, structure, and related photophysical behavior. *Chem. Rev. 102*, 759–781.
20. Craggs, T. D. (2009) Green fluorescent protein: Structure, folding and chromophore maturation. *Chem. Soc. Rev. 38*, 2865–2875.
21. Baird, G. S., Zacharias, D. A., and Tsien, R. Y. (1999) Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl. Acad. Sci. U.S.A. 96*, 11241–11246.
22. Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol. 24*, 79–88.
23. Tsien, R. Y. (1998) The green fluorescent protein. *Annu. Rev. Biochem. 67*, 509–544.
24. Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996) Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science 273*, 1392–1395.
25. Yang, F., Moss, L. G., and Phillips, G. N. (1996) The molecular structure of green fluorescent protein. *Nat. Biotechnol. 14*, 1246–1251.
26. Cabantous, S., Terwilliger, T. C., and Waldo, G. S. (2005) Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotechnol. 23*, 102–107.
27. Cody, C. W., Prasher, D. C., Westler, W. M., Prendergast, F. G., and Ward, W. W. (1993) Chemical Structure of the Hexapeptide Chromophore of the *Aequorea* Green-Fluorescent Protein. *Biochemistry 32*, 1212–1218.
28. Rosenow, M. A., Huffman, H. A., Phail, M. E., and Wachter, R. M. (2004) The crystal structure of the Y66L variant of green fluorescent protein supports a cyclization-oxidation-dehydration mechanism for chromophore maturation. *Biochemistry 43*, 4464–4472.
29. Wachter, R. M. (2007) Chromogenic cross-link formation in green fluorescent protein. *Acc. Chem. Res. 40*, 120–127.
30. Bujnicki, J. M. (2003) Crystallographic and bioinformatic studies on restriction endonucleases: Inference of evolutionary relationships in the "midnight zone" of homology. *Curr. Protein Pept. Sci. 4*, 327–337.
31. Eijsink, V. G. H., Bjork, A., Gaseidnes, S., Sirevag, R., Synstad, B., van den Burg, B., and Vriend, G. (2004) Rational engineering of enzyme stability. *J. Biotechnol. 113*, 105–120.
32. Hansson, M. D., Karlberg, T., Rahardja, M. A., Al-Karadaghi, S., and Hansson, M. (2007) Amino acid residues His183 and Glu264 in *Bacillus subtilis* ferrochelatase direct and facilitate the insertion of metal ion into protoporphyrin IX. *Biochemistry 46*, 87–94.
33. McKinney, M. K., and Cravatt, B. F. (2006) Structure-based design of a FAAH variant that discriminates between the N-acyl ethanolamine and taurine families of signaling lipids. *Biochemistry 45*, 9016–9022.
34. Sankpal, U. T., and Rao, D. N. (2002) Mutational analysis of conserved residues in HhaI DNA methyltransferase. *Nucleic Acids Res. 30*, 2628–2638.
35. Lutz, S., and Patrick, W. M. (2004) Novel methods for directed evolution of enzymes: Quality, not quantity. *Curr. Opin. Biotechnol. 15*, 291–297.
36. Woycechowsky, K. J., Vamvaca, K., and Hilvert, D. (2007) Novel enzymes through design and evolution. *Adv. Enzymol. Relat. Areas Mol. Biol. 75*, 241–297, xiii.
37. Eijsink, V. G. H., Gaseidnes, S., Borchert, T. V., and van den Burg, B. (2005) Directed evolution of enzyme stability. *Biomol. Eng. 22*, 21–30.
38. Farinas, E. T., Butler, T., and Arnold, F. H. (2001) Directed enzyme evolution. *Curr. Opin. Biotechnol. 12*, 545–551.
39. Kaur, J., and Sharma, R. (2006) Directed evolution: An approach to engineer enzymes. *Crit. Rev. Biotechnol. 26*, 165–199.
40. Rubin-Pitel, S. B., and Zhao, H. M. (2006) Recent advances in biocatalysis by directed enzyme evolution. *Comb. Chem. High Throughput Screening 9*, 247–257.
41. Dahiyat, B. I., and Mayo, S. L. (1996) Protein design automation. *Protein Sci. 5*, 895–903.
42. Bujnicki, J. M. (2002) Sequence permutations in the molecular evolution of DNA methyltransferases. *BMC Evol. Biol. 2*, 3.
43. Plaxco, K. W., Simons, K. T., and Baker, D. (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol. 277*, 985–994.
44. Xia, K., Manning, M., Hesham, H., Lin, Q., Bystroff, C., and Colon, W. (2007) Identifying the subproteome of kinetically stable proteins via diagonal 2D SDS/PAGE. *Proc. Natl. Acad. Sci. U.S.A. 104*, 17329–17334.
45. Molecular Operating Environment software (MOE) (2008) Chemical Computing Group, Inc., Montreal, QC.
46. Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol. 24*, 79–88.
47. Demidov, V. V., Dokholyan, N. V., Witte-Hoffmann, C., Chalasani, P., Yiu, H. W., Ding, F., Yu, Y., Cantor, C. R., and Broude, N. E. (2006) Fast complementation of split fluorescent protein triggered by DNA hybridization. *Proc. Natl. Acad. Sci. U.S.A. 103*, 2052–2056.
48. Topell, S., Hennecke, J., and Glockshuber, R. (1999) Circularly permuted variants of the green fluorescent protein. *FEBS Lett. 457*, 283–289.
49. Kent, K. P., Oltrogge, L. M., and Boxer, S. G. (2009) Synthetic control of green fluorescent protein. *J. Am. Chem. Soc. 131*, 15988–15989.
50. Andrews, B. T., Roy, M., and Jennings, P. A. (2009) Chromophore packing leads to hysteresis in GFP. *J. Mol. Biol. 392*, 218–227.
51. Barondeau, D. P., Kassmann, C. J., Tainer, J. A., and Getzoff, E. D. (2007) The case of the missing ring: Radical cleavage of a carbon-carbon bond and implications for GFP chromophore biosynthesis. *J. Am. Chem. Soc. 129*, 3118–3126.
52. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., and Sali, A. (2006) Comparative protein structure modeling using Modeller. Current Protocols in Bioinformatics, Chapter 5, Unit 5, 6, Wiley, New York.
53. Bystroff, C., and Baker, D. (1998) Prediction of local structure in proteins using a library of sequence-structure motifs. *J. Mol. Biol. 281*, 565–577.
54. Bystroff, C. (2002) MASKER: Improved solvent-excluded molecular surface area estimations using Boolean masks. *Protein Eng. 15*, 959–965.
55. Stemmer, W. P. C., Crameri, A., Ha, K. D., Brennan, T. M., and Heyneker, H. L. (1995) Single-Step Assembly of a Gene and Entire Plasmid from Large Numbers of Oligodeoxyribonucleotides. *Gene 164*, 49–53.
56. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graphics 14*, 33–38.
57. Westhead, D. R., Slidel, T. W., Flores, T. P., and Thornton, J. M. (1999) Protein structural topology: Automated analysis and diagrammatic representation. *Protein Sci. 8*, 897–904.