BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Toward Computationally Designed Self-Reporting Biosensors Using Leave-One-Out Green Fluorescent Protein

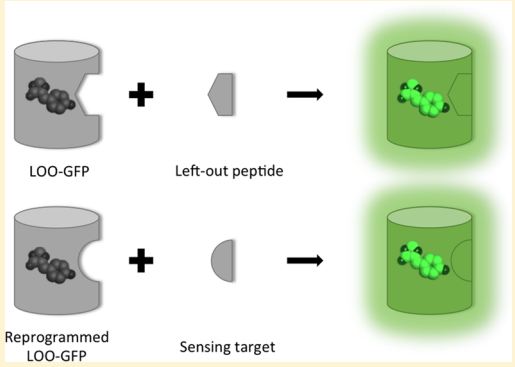
Yao-ming Huang,[†,‡] Shounak Banerjee,[‡] Donna E. Crone,[‡] Christian D. Schenkelberg,[‡] Derek J. Pitman,[‡] Patrick M. Buck,[‡] and Christopher Bystroff[*,‡,§]

[†]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, California 94158, United States

[‡]Department of Biological Sciences, Center for Biotechnology and Interdisciplinary Studies, and [§]Department of Computer Science, Rensselaer Polytechnic Institute, Troy, New York 12180, United States

Supporting Information

ABSTRACT: Leave-one-out green fluorescent protein (LOO*n*-GFP) is a circularly permuted and truncated GFP lacking the *n*th β-strand element. LOO7-GFP derived from the wild-type sequence (LOO7-WT) folds and reconstitutes fluorescence upon addition of β-strand 7 (S7) as an exogenous peptide. Computational protein design may be used to modify the sequence of LOO7-GFP to fit a different peptide sequence, while retaining the reconstitution activity. Here we present a computationally designed leave-one-out GFP in which wild-type strand 7 has been replaced by a 12-residue peptide (HA) from the H5 antigenic region of the Thailand strain of H5N1 influenza virus hemagglutinin. The DEEdesign software was used to generate a sequence library with mutations at 13 positions around the peptide, coding for approximately 3×10^5 sequence combinations. The library was coexpressed with the HA peptide in *E. coli* and colonies were screened for in vivo fluorescence. Glowing colonies were sequenced, and one (LOO7-HA4) with 7 mutations was purified and characterized. LOO7-HA4 folds, fluoresces in vivo and in vitro, and binds HA. However, binding results in a decrease in fluorescence instead of the expected increase, caused by the peptide-induced dissociation of a novel, glowing oligomeric complex instead of the reconstitution of the native structure. Efforts to improve binding and recover reconstitution using in vitro evolution produced colonies that glowed brighter and matured faster. Two of these were characterized. One lost all affinity for the HA peptide but glowed more brightly in the unbound oligomeric state. The other increased in affinity to the HA peptide but still did not reconstitute the fully folded state. Despite failing to fold completely, peptide binding by computational design was observed and was improved by directed evolution. The ratio of HA to S7 binding increased from 0.0 for the wild-type sequence (no binding) to 0.01 after computational design (weak binding) and to 0.48 (comparable binding) after in vitro evolution. The novel oligomeric state is composed of an open barrel.



Green fluorescent protein (GFP) has been shown to reconstitute its structure and fluorescent function after being split in a variety of ways, including short truncations,[1] two pieces of roughly equal size,[2] and truncated circular permutants.[3,4] This ability has found utilities in screening for solubility,[1] in protein complementation assays,[2] and in signaling conformational changes.[5] And since GFP requires no cofactors and may be expressed in diverse organisms, either as one piece or two, split-GFP is bound to find many additional applications.

GFP is an 11 stranded, closed β barrel enclosing a distorted α helix. A three-residue segment of the helix spontaneously matures via cyclization of the backbone, dehydration, and oxidation[6,7] to form a *p*-hydroxybenzylidene-imidazolidone moiety, the chromophore (CRO). The strand order, clockwise as seen from the termini end of the barrel, is 3−2−1−6−5−4−9−8−7−10−11, with all β pairings antiparallel except for strands 1−6. A consensus of evidence, including hydrogen/deuterium exchange,[8] in vivo solubilities,[9] folding kinetics,[3] circular permutants,[10,11] and noncircular (rewired) permutants,[11] suggests that the N-terminal segment, strands 1−6, and the central helix fold to their native state first. GFP folding kinetics are multiphasic, having three to five folding rates, ranging from 10 to 0.01 s^{-1}.[12] Not included among these phases is an observed, non-native trapped state that is related to the conformation of CRO[13] and a slow cis/trans backbone isomerization event at Pro89.[12,14] Once folded, GFP is extremely kinetically stable, unfolding with an estimated half-life greater than 1 month in water at neutral pH, standard temperature and pressure.[8] However, GFP unfolds rapidly at low pH. CRO maturation has a half-life in the range 19−83 min,[15,16] depending on the variant and the conditions. The chromophore retains its chemical structure after chemical,

Received: July 14, 2015
Revised: September 18, 2015
Published: September 23, 2015

DOI: 10.1021/acs.biochem.5b00786
Biochemistry 2015, 54, 6263−6273

thermal, or pH-induced unfolding. Fluorescence is quenched in the unfolded state via hula-twist motions in the excited state of the liberated CRO.[17]

Truncated or split GFPs exist in an unfolded or partially folded state until reconstitution, and are therefore weakly fluorescent or nonfluorescent. The rate of binding of circularly permuted GFP with strand 7 left out (LOO7-WT) to the exogenous strand 7 peptide was found to be limited by the rate of folding,[3] showing that LOO7-WT indeed exists in the unfolded state. The high solubility of LOO7-WT, as well as LOO-GFPs missing strands 4, 6, and 8 through 11,[9] suggests that it is not fully unfolded but exists in a near native intermediate conformation. Truncated superfolder GFP OPT with strand 11 removed was optimized for solubility in the truncated state,[1] and is also highly soluble in the unbound state. On the other hand, LOO-GFPs missing one of the β strands 1, 2, 3, or 5 are highly insoluble.

Since GFP folding is thousands of times faster than CRO maturation, the onset of fluorescence after reconstitution is much faster when CRO is already mature. Allowing the chromophore to mature before splitting would generate a faster-acting biosensor. Two methods have been demonstrated for separating the native state into split pieces after CRO maturation. (1) The two parts may be coexpressed as separate genes. The pieces spontaneously assemble upon folding and are then separated by unfolding, chromatographic separation, and refolding.[9] (2) The construct may be expressed as a single chain, with one strategically placed protease cut site. The two pieces are then separated by protease digestion followed by unfolding, chromatographic separation, and refolding.[18] Strategy (1) was used in the present work.

Having established that leave-one-out green fluorescent proteins (LOO-GFPs) bind their missing pieces and reconstitute fluorescence, a next goal is to design LOO-GFP so that it binds a different peptide, especially a peptide that is unique to a desired target, such as a virus. In this paper, we use computational methods to design a LOO-GFP to bind a peptide from H5N1 influenza virus hemagglutinin (HA) into its structure (Figure 1). The successful design should bind the HA peptide and signal the binding by green fluorescence.

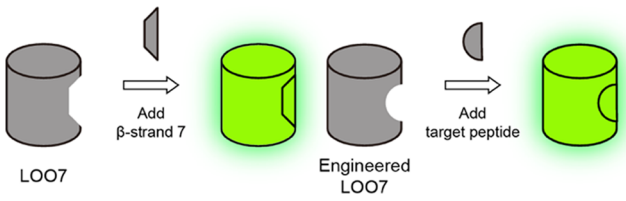


Figure 1. Leave-one-out method for biosensor design. Omitting a secondary structural element eliminates ths signal. Adding back the left-out piece recovers the signal. Designing the site to bind a different sequence creates a biosensor.

Both computational protein design and in vitro evolution were used in this study. Computational protein design finds the lowest energy amino acid sequence given the backbone conformation and other structural constraints. The design algorithm searches sequence space within a defined range of allowed side chain substitutions, using a rotamer (side chain conformation) library.[19,20] The best energy score identifies the sequence most likely to fold into the targeted structure with the desired function. Computational protein design has been implemented in several laboratories and has produced, among many results, a hyper-thermophilic protein,[21] two small molecule biosensor proteins,[22] two novel enzymes,[23,24] and a novel protein fold.[25] In this work, we used DEEdesign, which uses a parallel and distributed computational strategy.[26] The energy function for DEEdesign was optimized for rotamer recovery in a large data set of high resolution structures.[27]

This paper describes progress toward making LOO-GFP biosensors "programmable." LOO7-WT has been altered computationally so that it reconstitutes its structure and its function in response to the presence of the new targeted peptide analyte. The results show that the designed biosensors fold and glow in the presence of the targeted peptide, both in vivo and in vitro. However, problems were also encountered including autofluorescence in the unbound state, weak peptide binding, and low specificity. With improvements in these areas, LOO-GFP biosensors could be computationally designed to bind a wide variety of short peptides, enabling the creation of customizable molecular sensors that do not depend on antibodies.

MATERIALS AND METHODS

Protein Design. To engineer customized LOO-GFP biosensors, we developed a software suite called DEEdesign. DEEdesign uses a combination of physics-based functions and knowledge-based functions to energetically evaluate the fitness of side chain rotamers in protein structures. The parameters used in the fitness scoring system were trained by a machine learning technique to reproduce the true side chain conformations in high-resolution crystal structures.[27]

The space of all possible sequences may be determined manually or by surveying the sequences in a deep multiple sequence alignment. Sequence space is searched using either Monte Carlo (MC)[48] or the dead-end elimination algorithm (DEE).[49] DEEdesign employs an array of DEE criteria for rotamer elimination including Original Singles,[49] Goldstein Singles,[50] Simple Split Singles,[51] Magic Bullet Split Singles,[52] Original Doubles,[49] Full Goldstein Doubles,[50] and Magic Bullet Goldstein Doubles.[52] "Singles" refers to eliminations of single rotamers, while "Doubles" refers to eliminations of rotamer pairs.

Protein design using DEEdesign consists of three stages. First, the rotamer search space is reduced by running DEE using only the van der Waals repulsion energy term. Second, DEE is applied using the full energy function.[27] Third, the set of remaining rotamers is augmented to include slightly perturbed side chain chi angles, or "child" rotamers, and a MC search is carried out using the full energy function. The sequence upon convergence of this third step is deemed the minimum energy configuration (MEC). Starting with the MEC, a library of similar low-energy sequences is generated by running MC with a gradually increasing temperature, a process called "simulated melting", and keeping the desired number of unique sequences, typically 2000.[26]

Water Rotamers. Crystallographic waters were included as additional rotamers to model the unique buried polar environment of the GFP chromophore. Dummy atoms were placed around a central oxygen atom at points corresponding to the ideal tetrahedral arrangement of hydrogen bond donor and acceptor atoms. One hundred evenly spaced points on a sphere were generated by gradient descent maximization of nearest neighbor distances as described in ref 53. Three hundred and six tetrahedrals were randomly sampled from these 100 points, corresponding to three points of ideal H-bond geometry per Å^2

DOI: 10.1021/acs.biochem.5b00786
Biochemistry 2015, 54, 6263−6273

at 2.8 Å. One of the 307 water rotamers is the null set, consisting of no atoms. Template-based hydrogen bonding and non-hydrogen bonding energy terms were applied as previously described.[27]

Gene Library Assembly, Screening, and in Vitro Evolution. Using in-house software, the designed sequence library was encoded to a gene library having degenerate codons for the variable positions. Bases were selected to avoid stop codons and to favor high frequency codons. The sequence was divided into overlapping forward and reverse 60bp degenerate oligonucleotides using DNAWorks.[36] Oligos were synthesized by IDT. Two rounds of PCR were performed to assemble the designed gene library as described previously.[3] Amplified LOO7 gene variants were cloned into NcoI and *Eco*RI sites of pCDFDuet-1 (Novagen) also carrying the gene for the HA or S7 peptide fused to Ssp-DnaB mini-intein (GI: 222435775). Ligated plasmids were transformed into Acella cells (Edge BioSystems). Transformed cells were plated on nitrocellulose membranes over selective LB agar plates and cultured at 37 °C overnight. Protein expression was induced by transferring membranes to plates containing 0.5 mM IPTG at 25 °C and green fluorescence was monitored for 24 h.

Clones that showed the highest in vivo fluorescence were selected and subjected to affinity maturation by four rounds of iterative error-prone PCR[39] and one round of DNA shuffling.[38] Randomly mutated and shuffled LOO7-HA variants were cloned and plate-screened as described above. After each round of error-prone PCR, the clone that showed the highest fluorescence intensity was used as the parent PCR template for the next round. The resulting sequences were labeled EP*n*-*m*, for which *n* is the round of error-prone PCR and *m* is the clone number. See Figure S1.

Purified plasmids from all EP clones were pooled in equimolar amounts and subjected to partial digestion using 0.15 U/μL DNaseI at 25 °C for 5 min. Partially digested gene fragments ranging from 150 to 300 bp were recovered and used for gene assembly as described above. The resulting sequences were labeled DS*n*, where *n* is clone number.

Expression, Purification, and Monomerization. Proteins used in this study were expressed and purified as previously described.[3]

The following protocol was adopted for monomerizing LOO7s. N-Terminal polyhistidine tagged protein was denatured by buffer exchange into TN buffer (50 mM Tris-Cl, 100 mM NaCl at pH 8.0) with 6 M guanidinium chloride (GuHCl). A centrifugal filter with a 3 kDa cutoff was used for this purpose. The denatured protein solution was then batch-adsorbed onto pre-equilibrated Ni-NTA agarose beads stored in 1.7 mL microtubes.

The guanidinium chloride was gradually diluted out in 0.6 M steps and two washes per step, with 10 bed volumes of buffer, to refold the immobilized protein. The protein-coated beads were then imaged by fluorescence microscopy using a Nikon inverted microscope and a SPOT Imaging solutions camera. A positive control, with a circularly permuted, disulfide-engineered GFP[42] was tested for successful refolding and associated reconstitution of fluorescence. Uncoated beads were used as a negative control.

Size Exclusion Chromatography. LOO7-HA4 was equilibrated with equimolar amounts of target peptide overnight and concentrated by diafiltration using a 3 kDa filter. The product was then run on a Superose 12 GL300 size exclusion column, with a bed volume of 24 mL at a flow rate of 0.5 mL/min using TN buffer as the mobile phase. Elusion was monitored by absorbance at 280 nm. Protein that was not equilibrated with the target peptide was run using the same procedure, for comparison. Molecular weights for elusion peaks were estimated by retention times, comparing to Biorad's Gel Filtration standard (catalog number: 151-1901).

Binding Affinity. Peptide binding affinities were determined by measuring the timecourse of green fluorescence intensity upon manual mixing of LOO-GFP with synthetic target peptide (>95% pure, Genscript) over a range of concentrations, using up to 100-fold molar excess. Triplicate-based averages of the signal amplitude of the least-squares fit of the fluorescence time traces were plotted as a function of peptide concentration. Single exponential fits gave low residual, and therefore higher order kinetic fits were not tried. The dissociation constant K_d was determined by fitting data to (a) a Langmuir isotherm or (b) an Eadie−Hofstee plot.

Modeling. An open barrel model for LOO7-HA4 (Figure S12) was generated for demonstrative purposes only, using the molecular modeling suite MOE (CCG, Montreal). The LOO7-HA4 structure derived from DEEdesign was energy minimized with β7 removed and no distance restraints (atoms did not move significantly from their template positions), and then with "pushing" restraints applied between α carbons of β8 and β10. CαCα distances could be increased from template distances of approximately 9 Å to approximately 18 Å without introducing Ramachandran outliers and without disturbing R96 and E22 positions with respect to the chromophore. Coordinates for the LOO7-HA4 model and its "open barrel" form are available in the Supporting Information.

RESULTS

Programming of Leave-One-Out GFP Binding. Several researchers have shown that truncated GFP binds the missing secondary elements and reconstitutes fluorescence.[1,3,9,18,28−30] The left-out parts can range from single or two elements[30] to half of the β barrel.[31] For example, the term "LOO7-WT" refers to the circularly permuted sequence of the template OPT-GFP[1] truncated to remove strand 7, also containing an N-to-C linker and an N-terminal His-tag (Figure S1). LOO7-WT binds to strand 7 with submicromolar affinity.[3] We will use the shortened term "LOO7" to refer to all GFP constructs with strand 7 removed, generically.

Knowing that the reconstitution can occur with most of the individually left-out β strands[9] and can be optimized by directed evolution,[1] we postulate that for a given peptide there should exist leave-one-out GFP sequences that bind to the peptide and reconstitute fluorescence. However, the massive search space will pose a challenge to finding those sequences in an efficient way.

Here we propose a hybrid strategy to program LOO7 by first computationally designing its sequences to accommodate a selected peptide and then in vitro evolving the sequences to improve the binding. For our proof-of-concept design, the target peptide is derived from H5N1 influenza hemagglutinin. The abbreviation HA is used herein to refer to the 12-residue peptide from hemagglutinin. LOO7-HA refers to all truncated, permuted, computationally designed GFP sequences that bind HA in the strand 7 position. LOO7-HA4 is one of these. LOO7-DS1 and LOO7-DS2 are two in vitro evolved variants of LOO7-HA4.

To preserve the native residue numbering for circular permuted and/or split GFP constructs, we used the sequence

DOI: 10.1021/acs.biochem.5b00786
Biochemistry 2015, 54, 6263−6273

Table 1. Results of Design and Selection for Designable Positions of LOO7-GFP[a]

		DEEdesign				
position	wild-type residue	input	output	degenerate codon	library	LOO7-HA4
83	F	AFILMVW	FW	TKS	FLCW	**W**
84	F	AFILMVW	FM	WTK	MILF	F
161	I	AFILMVW	ILV	VTA	ILV	I
163	A	AFILMVW	I	ATC	I	**I**
164	N	NST	NST	AVY	NST	**T**
165	F	AFILMVW	F	TTT	F	F
167	V	AFILMVW	IMV	RTN	IMV	V
168	R	RNST	NST	AVY	NST	**T**
200	Y	YKHR	HKR	MRS	HKR	**H**
201	L	LFMW	FM	WTK	MILF	L
202	S	SKHR	HKR	MRS	HKR	**K**
204	Q	QNST	NST	AVY	NST	**T**
224	V	AFILMVW	IV	RTK	IV	V

[a]Position numbers refer to the unpermuted superfolder GFP OPT sequence as shown in Figure S1. LOO7-HA4, a specific design with 7 mutations (bold), was one of approximately 75 glowing colonies. Degenerate codons are shown using IUPAC nucleotide notation.

numbering of the unpermuted superfolder GFP structure (PDB ID 2B3P). Therefore, regardless of permutation, the chromophore is residue 66, replacing residues 65−67 in the translated sequence. In this paper residues 65−67 refer to TYG sequence numbering before chromophore maturation. The LOO7 sequences additionally contain an N-terminal His-tag (MTHHHHHHSSG) and a linker connecting the original C and N-termini (GGTGGS). The numbering for all GFP variants discussed here is provided in Figure S1.

Selection of the Target Sequence for Sensing. A signature sequence pattern was created that defined the minimum sequence requirements for a peptide to replace the S7 peptide in the OPT-GFP structure. The pattern describes the periodicity of hydrophobic/hydrophilic residues in β strand 7, residues 146−157, as follows: 146-{C}{PC}[GANDTEHKSVW]{C}[EAYFILMWV]{PC}[AYFILMWV]{PC}{PC}-{PC}{PC}{PC}-157, where {} indicates excluded and [] allowed amino acids. The periodicity of polar and nonpolar groups is maintained so that the peptide in solution will tend toward a beta strand structure.[32] Note that cysteine was excluded sequence because its presence might cause complications in the purification process. Proline was prohibited because it would interfere with β sheet hydrogen-bond formation. Otherwise, only three inward facing side chain positions (residues 148, 150, and 152) were regarded as constrained by hydrophobicity or by proximity to the chromophore.

Hemagglutinin was selected as the detection target because it is the most abundant surface protein in the influenza viral coat.[33] The hemagglutinin sequence had 38 matches to the signature pattern above. Of these, we selected the target sequence SSHEVSLGVSSA (HA) because it occurs only in the H5 subtype of influenza strains and is unique to the two Thailand strains (A/Thailand/16/2004, GI:145284476). In most other strains of the H5 subtype, the valine at position 5 of the selected peptide is an alanine. In LOO7-WT, position 5 is an inward facing position, making it a potential specificity determining position. Thus, by choosing this target sequence, a successful biosensor would distinguish the two Thailand strains from other H5 strains.

Computational Design of LOO7-GFP. To design a LOO7 to receive HA in the S7 site, we first modeled the HA sequence in place of the wild type sequence S7, changing 146-NSHNVYITADKQ-157 (S7) to 146-SSHEVSLGVSSA-157 (HA), a change of 9 out of 12 sequence positions. On visual inspection it appeared that all of the changes could be accommodated by rotameric changes in neighboring side chains and none would require large backbone shifts. Of the 12 S7 side chains, only V150 and I152 are completely buried in the core. V150 is conserved in HA, while I152 is changed to a leucine. Most of the sequence changes are located on the surface and are expected to contribute less to the binding specificity. Despite this similarity, the HA has no detectable binding to LOO7-WT.

Computational designs were based on the coordinates of OPT-GFP, which was modeled on superfolder GFP[34] (PDB ID 2B3P). Thirteen side chains within 7 Å of S7 were allowed to explore different amino acids from a predefined design space, as shown in Table 1. The predefined design space expresses the hydrophobic and hydrophilic patterning observed in a multiple sequence alignments of GFP homologues and it helps to ensure the generation of GFP-like sequences. In addition, ordered water molecules surrounding the chromophore are thought to be involved in the proton relay that facilitates the chromophore maturation.[16,35] To preserve this property, water molecules adjacent to the chromophore were included in the design process using water rotamers (Figure S2). Computational design, as described in Materials and Methods, resulted in a library of candidate sequences for LOO7-HA.

Gene Library Synthesis and Plate Screening. The sequence library was generated by creating degenerate codons to cover sequence variable positions in the expression (Table 1). The unavoidable inclusion of unwanted amino acids in three positions increased the theoretical number of sequences in the library by 8-fold, giving a total complexity of $\sim 3 \times 10^5$. The nucleotide sequence expression, containing degenerate codons, was divided into overlapping forward and reverse 60-based oligonucleotides using the DNAworks software.[36] Degenerate oligos were purchased (IDT DNA) and assembled using assembly PCR[37] to form the gene library.

The library was cloned into MCS1 of vector pCDFDuet-1(Novagen), while the gene for the HA peptide N-terminally fused to Ssp-DnaB mini-intein (GI: 222435775) was cloned into MCS2. The two genes were coexpressed anticipating that reconstitution of structure and chromophore maturation would occur in vivo. Transformed *E. coli* were first plated on

DOI: 10.1021/acs.biochem.5b00786
Biochemistry 2015, 54, 6263−6273

nitrocellulose membranes over selective media. Protein expression was then induced by transferring membranes to plates containing 0.5 mM IPTG and antibiotics.

About 10% of the 3×10^5 possible sequences were visually screened for fluorescence on plates, 24 h after induction of the coexpressed genes. Approximately 75 out of the 2500 total colonies showed green fluorescence upon blue light excitation (Figure S3). One colony that exhibited a high fluorescence intensity was selected for the further studies, herein called "LOO7-HA4". Sequencing confirmed that LOO7-HA4 contained seven mutations near the binding site of the HA peptide, while the other six designed positions retained the wild type amino acid. All of the mutations were derived from the computationally designed library.

Inspection of the modeled mutations revealed new potential sources of binding affinity and specificity for HA. Buried mutation I152L is accommodated by A163I, which blocks the position of the I152's second gamma carbon but allows L152 (Figure S4). On the surface, redesigning 199-HYLSTQ-204 to 199-HHLKTT-204 accommodates outward facing peptide side chain changes N149E, Y151S and T153G by removing hydrophobic packing, adding a salt bridge, and shape complementarity (Figure S5).

Fluorescence Decreases upon Peptide Binding. LOO7-HA4 was coexpressed with intein-HA and purified using a His-tag and Ni-NTA agarose beads. Under the conditions used, the HA peptide would be cleaved from the intein-HA fusion by the natural action of intein, leaving the free peptide, HA. The bound HA was removed by denaturing and refolding the protein on Ni-NTA agarose beads followed by elution using imidazole, as previously described.[9] Purified, unbound LOO7-HA4 was soluble and contained a mature chromophore, as determined by UV absorption (Figure S6). The chromophore maturation efficiency, defined as

$$\frac{A_{385}(\text{L007-HA4})}{A_{280}(\text{L007-HA4})} \Big/ \frac{A_{385}(\text{OPT-GFP})}{A_{280}(\text{OPT-GFP})}$$

was 70%, much higher than corresponding efficiency for LOO7-WT at 40%.

After elution from Ni-NTA agarose beads, LOO7-HA4 was autofluorescent, having approximately 3% of the molar quantum yield of OPT-GFP. This is qualitatively similar to the level of autofluorescence in unbound LOO7-WT.[3] Autofluorescence also occurs in all five other unbound LOO*n*-GFPs that glow and reconstitute fluorescence, where n = 4, 7, 8, 9, 10, and 11, corresponding to the number of the β-strand omitted (ref 9 and unpublished data).

Although evidence for fluorescence reconstitution was observed when LOO7-HA4 was coexpressed with HA peptide in vivo, the addition of synthetic HA peptide (Genscript) to purified LOO7-HA4 in solution produced a 50% decrease in the fluorescent signal instead of the expected increase (Figure 2 shows the kinetic trace for LOO7-DS2, which behaves similarly to LOO7-HA4). In the control experiment, the addition of the strand 7 peptide (S7) to purified LOO7-WT produced the expected 1.7-fold increase in fluorescence as previously shown.[3] Concentrations of HA up to 1 mM had no effect on the fluorescence of LOO7-WT. However, submicromolar concentrations of S7 produced a 50% decrease in the signal of LOO7-HA4. To explain the peptide-induced decrease in fluorescence, we hypothesized that purified unbound LOO7-HA4 was forming a fluorescent oligomeric complex.

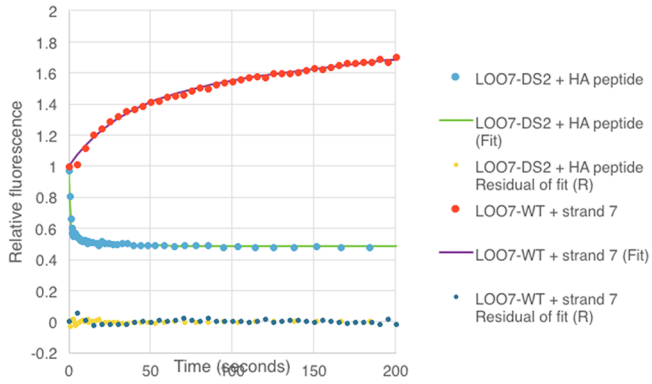


Figure 2. Fluorescence trajectory upon adding peptide in solution to LOO7-WT and LOO7-DS2. Lines are least-squares fits to the data. Flat residual (lower lines) shows the quality of the fit.

Peptide Binding Shifts Oligomer Equilibrium. The presence of approximately equal amounts of monomers and low order oligomers (dimers, trimers, tetramers) in solutions of pure, unbound LOO7-HA4 was observed in size exclusion chromatograms (SEC) (Figure 3). Higher molecular weight

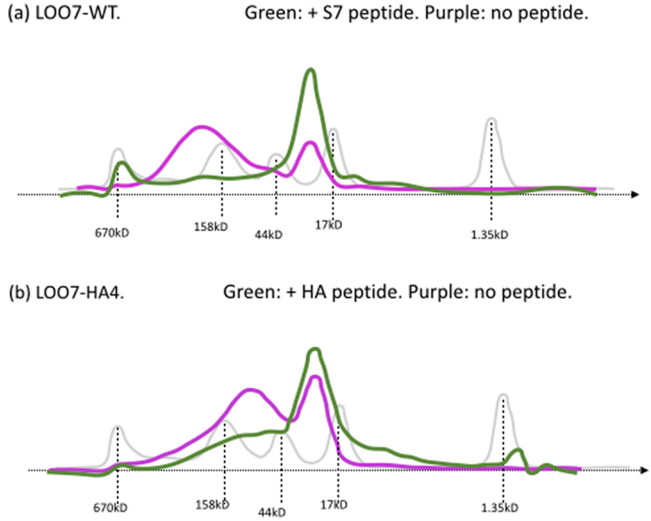


Figure 3. SEC traces for LOO7-WT and LOO7-HA4. (a) Purified LOO7-WT with (green) and without (purple) the addition the S7 peptide. Gray trace are the molecular weight standards, marked by kD. (b) Purified LOO7-HA4 with (green) and without (purple) addition of synthetic HA peptide. The sharp peak at approximate 27kD corresponds to monomers. The broad peaks at higher MW correspond to nonmonodisperse mixtures of oligomeric states which are in slow equilibrium with monomer. LOO7-WT forms higher order and a greater fraction of oligomers than LOO7-HA4.

fractions were fluorescent while monomer-sized fractions were not initially fluorescent. However, the low molecular weight fraction recovered fluorescence upon incubation at RT. Addition of the synthetic HA peptide to LOO7-HA4 shifted the ratio of oligomers to monomers from approximately 1:1 to approximately 1:3 in the SEC. The results were different for the parent construct LOO7-WT, which exhibited a broader spectrum of somewhat higher order oligomers. Adding S7 to LOO7-WT shifted the oligomer/monomer ratio from around 4:1 to approximately 1:10. Since the addition of peptide clearly shifts the equilibrium away from oligomeric species and toward the monomer, and since the shift correlates with shifts in

DOI: 10.1021/acs.biochem.5b00786
Biochemistry 2015, 54, 6263−6273

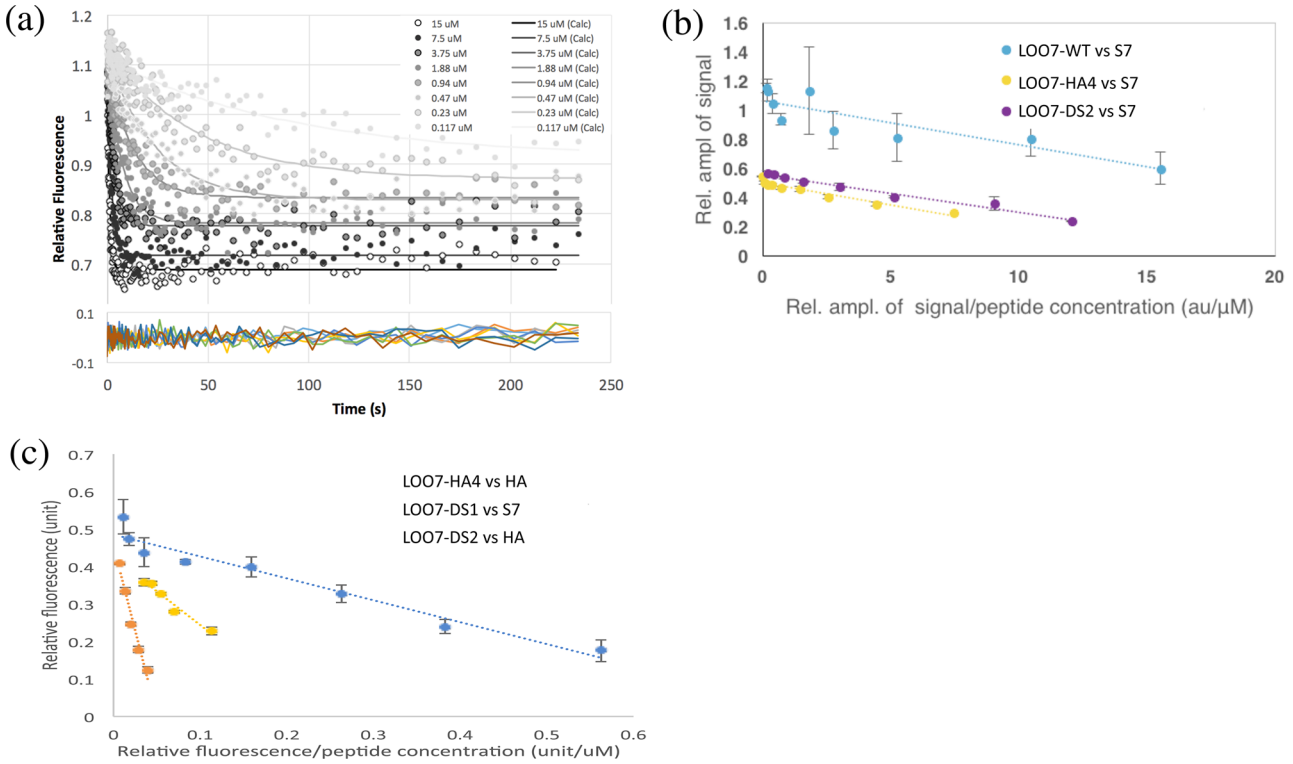


Figure 4. Peptide binding assay. (a) Example of fluorescent traces for varying concentrations of S7 peptide (0.12−15 μM) added at time zero to LOO7-DS2. Kinetic traces all fit single exponential decays. Residuals of the Excel/Solver fits are shown along the bottom of the plot. Data were logarithmically downsampled for better behavior in Solver. (b, c) Eadie−Hofstee plots for S7 peptide binding to each of the three variants discussed. Data plotted are relative amplitude of fluorescence signal (positive or negative) over peptide concentration, versus relative amplitude of signal in arbitrary units. Error bars are standard deviations for triplicate measurements. Linear fits are shown. Slopes represent dissociation constants, K_d.

fluorescence intensity of opposite sign, we conclude that the monomeric state binds the peptide more tightly than the oligomeric state, and that the bound, monomeric state of LOO7-WT is natively folded while that of LOO7-HA4 is not.

To confirm the relative lack of fluorescence of unbound monomers relative to oligomers, we created a persistent monomeric state by refolding the protein in the immobilized state. His-tagged LOO7-WT and LOO7-HA4 were each refolded on Ni-NTA agarose beads as described in Materials and Methods. The fluorescence was lost upon unfolding in 6 M GuHCl, and did not return upon bringing the beads back to folding conditions by washing with pH 7.5 buffer, confirming that oligomers glow and unbound monomers do not. Note that GuHCl unfolding causes no chemical change to the chromophore structure.

Unfortunately, addition of up to 1 mM HA peptide to monomerized LOO7-HA4 on beads did not produce a detectable increase in fluorescence (Figure S7a−d), whereas stepwise addition of S7 peptide to monomerized LOO7-WT produced a positive signal with the expected saturation behavior and peptide concentration-independent (i.e., folding) kinetics (Figure S8).

To rationalize the structural state of the unbound LOO7-HA4 in solution, we considered the rapid kinetics of HA peptide binding, contrasting it to the slow binding of S7 to LOO7-WT (Figure 2). The fast peptide binding signifies that folding, a much slower process, is not occurring. As expected, the rate of peptide binding was found to be peptide concentration dependent (Figure 4), consistent with binding being the rate limiting step. The HA peptide binding kinetics were least-squares fit to a single exponential for all concentrations studied (0.18−15.0 μM), suggesting a single binding site and binding mode. We conclude that LOO7-HA4 exists in a partially unfolded state that binds the HA peptide.

Increased in Vivo Fluorescence by in Vitro Evolution. An effort was made to improve the binding affinity and recover fluorescence reconstitution by random mutagenesis and high-throughput screening. DNA shuffling[38] and error-prone PCR[39] were performed on the LOO7-HA4 gene to generate mutants. The mutated sequences were cloned into *E. coli* that coexpress the intein-HA gene and plate-screened for green fluorescence. Plates were visualized using a Dark Reader transilluminator with long-pass filter (Clare Chemical). Green colonies were selected and sequenced. Qualitatively brighter clones were used as the source of template DNA for further rounds of in vitro evolution. Figure S1 shows the sequences of randomly selected glowing variants of LOO7-HA4.

Several glowing colonies were picked and streaked on a plate for side-by-side comparison (Figure S9). After coinduction of protein and target peptide, fluorescence of the streaks was monitored for 24 h. Differences in the rate of growth of the fluorescence intensity were interpreted as differences in the stability and peptide binding ability of the mutants. Two clones were selected for further study because they produced earlier chromophore maturation (LOO7-DS1) or elevated fluorescence intensity (LOO7-DS2) when coexpressed with the HA peptide (Figure S10). LOO7-DS1 and LOO7-DS2 had nine and 16 mutations, respectively. All three biosensor candidates were further characterized, as described below.

Peptide-Dependent in Vivo Chromophore Maturation Rate. To assess chromophore maturation rates, we imaged Petri plates at time points after induction, and compared the

DOI: 10.1021/acs.biochem.5b00786
Biochemistry 2015, 54, 6263−6273

Table 2. Chromophore Maturation Efficiency (CRO_m) and Dissociation Constants (K_d) for LOO7-WT (WT) and LOO7-HA4 after Computational Design (HA4) and after Additional Optimization by in Vitro Evolution (DS1, DS2) for Peptides from Hemagglutinin, SSHEVSLGVSSA (HA), and from Strand 7 of GFP, NSHNVYITADKQ (S7)[a]

		K_d						
LOO7-GFP variant	CRO_m	HA	SD	S7	SD	$S7_{ext}$	SD	$K_d(S7)/K_d(HA)$
WT	40%	>1 mM		570 nM ↑	50	30.5 nM ↑	12	$<6 \times 10^{-7}$
HA4	70%	9.0 μM ↓	1.26	85 nM ↓	20.4	35.1 nM ↓	18	0.0094
DS1	40%	>1 mM		2.8 μM ↓	1.23	1.80 μM ↓	0.233	$<3 \times 10^{-6}$
DS2	80%	600 nM ↓	299	289 nM ↓	63.7	25.8 nM ↓	2.71	0.48

[a]The S7 peptide was extended by two residues to NFNSHNVYITADKQ to make $S7_{ext}$, and WT for this peptide is truncated by four C-terminal residues (EYNF). ↓: Signal decreases upon binding. ↑: Signal increases upon binding. Specificity for HA versus S7 peptides is expressed as the ratio of K_d's, $K_d(S7)/K_d(HA)$. SD: standard deviation.

dual-expressed (pCDFDuet-1) construct to another construct with the peptide target gene in MCS2 removed by cloning. Chromophore maturation levels were assessed qualitatively by green fluorescence. LOO7-DS1 showed fluorescence within 4 h of induction in the presence of the HA peptide, but almost no fluorescence at 4 h in the absence of the HA peptide. However, after 24 h LOO7-DS1 showed fluorescence with and without the HA peptide (Figure S10). LOO7-DS2 exhibited maturation of chromophore within 24 h in the presence of the coexpressed HA peptide, while the same sequence in the absence of the peptide showed barely detectable fluorescence after 24 h. This suggests a specific interaction between LOO7-DS2 mutant and the HA peptide in vivo leading to faster chromophore maturation.

LOO7-WT also exhibits very slow maturation of the chromophore in the absence of S7, on the time scale of weeks to months for purified protein stored at 4 °C. LOO7-DS1 catalyzes chromophore formation in the absence of a complete and native β barrel structure, possibly aided by the stabilizing forces of oligomerization. Chromophore formation in LOO7-DS2 showed a greater dependency on the presence of the HA peptide than LOO7-DS1, implying that the oligomeric form of unbound LOO7-DS2 is less well folded than that of LOO7-DS1. All LOO7 constructs were less efficient than OPT-GFP at chromophore maturation (Table 2), but LOO7-DS2 achieved 80% of the wild-type rate.

Affinity and Specificity. LOO7-HA4, LOO7-DS1, and LOO7-DS2 were analyzed to determine the affinity for synthetic HA and S7 peptide (Table 2). Affinity measurements were done by plotting the equilibrium amplitude of the signal change as a function of peptide concentration, using either the gain or loss of fluorescence as appropriate (Figure 4).

No change in fluorescence was observed upon addition of HA to LOO7-DS1. This was surprising, since the in vivo fluorescence was perceptibly increased relative to LOO7-HA4. From this, we conclude that the increased fluorescence was due to a greater tendency to glow without the peptide present, a negative but nonetheless informative result. Increased autofluorescence due to an increased tendency to dimerize in the unbound state may have fooled our screen.

Fluorescence intensity loss of 50% was observed upon addition of HA to LOO7-DS2, similar in magnitude to the effect of the peptide on LOO7-HA4. The kinetics of peptide binding to LOO7-DS2 was single-exponential and peptide concentration dependent.

The binding of HA to LOO7-WT was undetectable, while HA binding to LOO7-HA4 was measured to be 9.0 μM, representing the appearance of a binding site where there had been none. However, binding of S7 was paradoxically tighter in LOO7-HA4 than in the LOO7-WT. We propose a measure of success in design to be the binding affinity ratio, $K_d(S7)/K_d(HA)$, which measures the improvement in binding of the target peptide over the original, native ligand. By this measure, LOO7-HA4 shows improvement at $K_d(S7)/K_d(HA) = 0.0094$ as compared to $<6 \times 10^{-7}$ for LOO7-WT. LOO7-DS2 showed further improvement, with 15-fold tighter binding to HA and 3-fold weaker binding to S7 to give $K_d(S7)/K_d(HA) = 0.48$.

The reason for the lack of fluorescence reconstitution in the LOO7-HA variants, even in the face of submicromolar peptide binding, may have been that the energetic impetus provided by binding was insufficient to overcome the barrier to the closing the β barrel to form the fully native state. This would explain why fluorescence was observed in vivo, where the proteins had hours to associate and fold, but not in vitro where experiments were cut off after 250 s. To follow up on this hypothesis, we extended the S7 peptide N-terminally by 4 residues and truncated the C-terminus by 2 residues (EYNF added at N terminus and KQ truncated at C-terminus; target is now 142–155 instead of 146–157). Synthetic 14-residue "$S7_{ext}$" peptide binding was assayed to each construct, LOO7-WT, LOO7-HA4, LOO7-DS1 and LOO7-DS2. The extension, which adds a single buried hydrophobic side chain, did increase the affinity by 10-fold for all constructs except LOO7-HA4, where no difference was seen. But once again only fluorescence quenching was observed for each of the three LOO7-HA variants. No increase in brightness was observed despite a K_d of 25.8 nM in the case of LOO7-DS2. Only a slight increase in fluorescence at long time-scales (800 s) suggests that the high folding barrier theory has any merit. More likely, folding is impaired by one or more of the mutations.

■ DISCUSSION

Our view of the nature of the fluorescent state of GFP has changed radically. Previously, we thought that leaving out the seventh strand would quench fluorescence by exposing the chromophore to bulk solvent[40] or altering the ionic state of the chromophore.[41] However, our new view and the broader understanding of GFP quenching is that solvent exposure is irrelevant; rather, folding is required to stabilize the planar conformation of the excited state of the chromophore by blocking the "asynchronous hula twist" rotation of the *p*-hydroxybenzylidene moiety that leads to overlap of the ground state and excited state energy landscapes and triggers nonradiative decay.[17] We note in retrospect that computationally designed sequence libraries were screened for fluorescence, not directly for peptide binding, on the assumption that fluorescence would correlate with peptide binding. That assumption turned out to be wrong. Instead we were screening

DOI: 10.1021/acs.biochem.5b00786
Biochemistry 2015, 54, 6263–6273

for a more ordered chromophore microenvironment, attainable by any means including oligomerization. A better screen for a peptide biosensor library would be one that detected binding directly, along with and independent of fluorescence.

A Model for LOO7-GFP Dynamics. The results show that computational design coupled with in vitro evolution can produce a protein-peptide pair that bind with submicromolar affinity. However, the results fall short of demonstrating the reconstitution of the native, functional state upon binding the peptide, as originally intended. To better understand the sources of this failure, we here build a working model of the system that is consistent with all of the data (Figure 5). In the context of this model, we hope to explain the differences between the expected and the observed behavior of the designed biosensor.

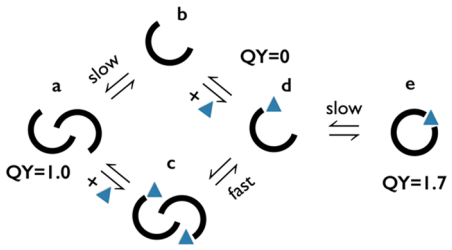


Figure 5. Proposed working model for LOO7-GFP oligomerization, binding, barrel closing, and nonproductive binding. Open arcs represent LOO7-GFPs in open β barrel configuration. Triangles are peptides (S7 or HA). QY = quantum yield relative to (a) the unbound, oligomeric state. Peptide binding (a–c, b–d) is followed by (e) barrel closing to form a more fluorescent state.

The decreased fluorescence of LOO7-HA4 in the presence of the HA peptide suggests that the bound state chromophore environment is less ordered than the unbound state chromophore environment, contrary to expectations. At the same time, no folding or unfolding transitions are observed in the peptide binding kinetics. Therefore, the increased order that leads to increased fluorescence must be provided by oligomerization of a partially folded state.

The less folded state in all LOO7-GFPs must be near-native because the protein binds the peptide. If it were not near-native, then peptide binding would not have been observed. A near-native folding intermediate of GFP has been previously described, consisting of an open barrel[42] whose kinetics of closing are sensitive to specific engineered disulfide bonds. The proposed openings in the barrel are at the β7/ β8 cleft and at the β4/β9 cleft. The β7/ β8 location is the most likely site of barrel opening for LOO7-HA4, since the omission of β7 eliminates several of the interactions that close the barrel at that site. The existence of an opening between β7 and β8 in wild-type GFP is supported by a high degree of solvent exposure in H/D exchange NMR studies[43] and relatively high mobility of the atoms in this region in molecular dynamics simulations.[44] It therefore seems reasonable to model unbound, monomeric LOO7 as an open barrel having some degree of additional space between β strands β8 and β10, such that binding a peptide to either the β8 or the β10 surface would not be sufficient to complete and close the barrel.

We have established that oligomerization correlates with increased autofluorescence, and that peptide binding results in rapid loss of fluorescence. But formation and dissociation of oligomeric LOO7-HA4 is slow on the time scale of the SEC experiment (Figure 3). We hypothesize, therefore, that a thermodynamic box exists which has two pathways to the monomeric bound state, a slow pathway that is consistent with the time scale of the SEC experiment, and a fast pathway in the presence of the peptide that is consistent with the binding kinetics (Figure 5). We know that the monomeric open barrel state has negligible fluorescence, therefore peptide-induced dissociation leads to quenching. Peptide binding could trigger dissociation through a mechanism such a strand displacement, where a tight binding strand replaces a weak, intermolecular strand–strand pairing one hydrogen bond at a time, like a zipper. The strand displacement mechanism has been observed in pilus assembly in Gram-negative bacteria.[45]

The observation of a single exponential phase of HA peptide binding to LOO7-HA4 is consistent with a single binding site on the open barrel. Two possible events follow the binding of the peptide: the barrel may close to form the native fluorescent state, or a peptide could bind at the second site, preventing barrel closure. We have no evidence for the doubly bound state.

In the working model (Figure 5), all variants of leave-one-out GFP exist in the oligomeric state "a" before adding peptide. Upon S7 peptide addition, LOO7-WT shifts to the native state "e", but adding HA peptide to LOO7-WT has no consequence. LOO7-HA4 and LOO7-DS2 both shift to state "d" upon binding of either HA, S7, or $S7_{ext}$, but neither protein proceeds spontaneously to state "e". LOO7-DS1 does not proceed to state "d" in the presence of HA, but does so in the presence of S7 or $S7_{ext}$.

Mutations That Affect Barrel Opening. The question arises of what sequence changes may explain the evidence. If we can identify such mutations, we may be able to rescue fluorescence reconstitution. F83W, a computationally designed mutation that was conserved in all evolved sequences including LOO7-DS1 and LOO7-DS2, creates a crowded corner around β7 (Figure 6). W83 is on the central helix, bordering residues

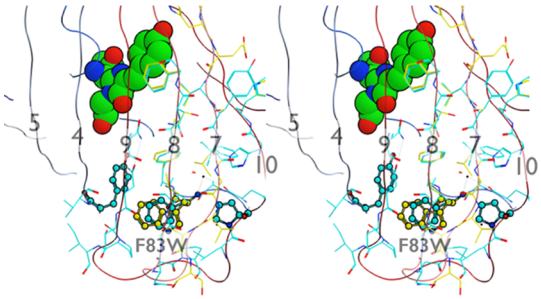


Figure 6. Computationally designed mutation F83W creates clashes with residues on β strand 4 and 10. Opening the β barrel would probably relieve these clashes. Reverting W83 to F might result in a closed barrel.

on β4 and β10. Opening the barrel would likely relieve the clashes, and thus W83 favors the open barrel state. Reverting to F83 would test this hypothesis. A163I adds three carbon atoms to the core. The Ile side chain interacts with the target residue L149, providing a source of specificity against the wild type isoleucine at that position. A reversion, I163 to A or an I163G mutation, combined with mutations of L201 to M or F and of I161 to A or V may improve packing.

Mutations that conserve or improve in vivo fluorescence occurred all over the protein, not just near the active site (Figure S11). Nine of the 19 mutations are on the protein

DOI: 10.1021/acs.biochem.5b00786
Biochemistry 2015, 54, 6263–6273

surface (e.g., E33D, K51N, K98T, T229S), and are not expected to affect the HA peptide binding directly but may affect oligomer formation. Seven are buried in the core (e.g., V15A, T64A, A84T, N118S), and three are in an unstructured loop (e.g., M230L, E232D). One, T65A, was within the chromophore itself and may have contributed to the slower chromophore maturation observed in LOO7-DS2, relative to LOO7-DS1. Only one randomly derived mutation was found in a previously computationally designed position (R168T to A). It is apparent that the process of in vitro evolution improved the autofluorescent state in the case of LOO7-DS1, rather than improving the energy of the bound state. The numerous surface mutations may have had an effect on the oligomerization tendencies. Note that LOO7-DS1 was found to be relatively agnostic to the presence of the HA peptide in vivo in chromophore maturation rate studies, which is consistent with the absence of binding affinity.

Others have accelerated the folding of GFP by designing a loop connecting the central helix to β4 to remove a problematic proline at position 89.[14] The cis−trans isomerization of this residue is known to be the slowest phase in GFP's folding. Thus, it is interesting to note that an A87T mutation is observed in both LOO7-DS1 and LOO7-DS2. This residue is one before the pre-Pro position of this cis-peptide, and a local structural distortion caused by this residue could favor the cis over the trans state, which could explain the increase in fluorescence of the LOO7-DS1 and LOO7-DS2 with respect to the original LOO7-HA4.

The observed high degree of chromophore maturation seems to contradict the open barrel model for the LOO7 designs, but this is not conflicting evidence. A rigid chromophore microenvironment is required to stabilize the excited state, but there is no evidence that rigidity is required for the catalysis of the backbone cyclization and dehydration reactions that generate the chromophore. On the contrary, the large atomic motions observed between the immature state of the protein and the mature form suggest the need for flexibility in the core.[46,47] Localization of R96 and E222 has been shown to be key to catalysis of the backbone cyclization,[7] but a mechanical compression model for the catalysis remains unproven.[47] We have explored the open barrel structure using molecular dynamics (MOE, CCG, Montreal, Canada) and find that R96 and E222 can retain their native conformations relative to the chromophore, even as the β8/β10 cleft doubles in width (Figure S12). The target peptide for this experiment is on an exposed edge of an exposed domain of hemagglutinin, lending promise that binding to an open LOO-HA4 barrel may be possible on the surface of the virus (Figure S13).

CONCLUSION

An attempt to make a fluorescent biosensor for the influenza virus using computational design and in vitro evolution led to some important lessons with regard to the dynamic behavior of GFP in the leave-one-out state. We find that the truncated protein can adopt an open barrel conformation, and that this conformation catalyzes chromophore maturation and fluoresces without requiring the structural reconstitution that comes with binding the left-out peptide. The selection for green fluorescence rather than a direct measure of peptide binding in the high-throughput screening process set up the discovery of a novel fluorescent species consisting of low order oligomers of a partially folded state of GFP. The new species detects the target peptide, although not in the intended way, by catalyzing the dissociation of the fluorescent complex and the concomitant loss of the fluorescent signal.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.5b00786.

Coordinates for the LOO7-HA4 model and its "open-barrel" form (PDB)

Sequences derived from in vitro evolution; water rotamers; plate screening; before and after computational design; binding pocket, wild-type and DS2; UV absorption spectrum of denatured LOO7-HA4; immobilized LOO7-HA4; peptide binding to immobilized LOO7-WT; secondary plate screen for in vivo fluorecence; time course of chromophore maturation with and without peptide; distribution of designed and evolved mutations; open-barrel conformation of LOO-GFP; hemagglutinin showing HA peptide (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: bystrc@rpi.edu. Telephone: 518-276-3185.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

This research was supported by NIH Grant GM099827 to C.B.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Protein design calculations were carried out at the Rensselaer Polytechnic Insitute Center for Computational Innovations.

ABBREVIATIONS

GFP, green fluorescent protein; LOO, leave-one-out; WT, wild-type; S7, β-strand 7 of green fluorescence protein; HA, hemagglutinin, hemagglutinin peptide 123−134; CRO, chromophore; DEE, dead-end elimination algorithm; MC, Monte Carlo algorithm; MEC, minimum energy configuration; GuHCl, guanidinium chloride; SEC, size exclusion chromatogram

REFERENCES

(1) 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.

(2) Barnard, E., McFerran, N. V., Trudgett, A., Nelson, J., and Timson, D. J. (2008) Development and implementation of split-GFP-based bimolecular fluorescence complementation (BiFC) assays in yeast. *Biochem. Soc. Trans. 36*, 479−482.

(3) Huang, Y. M., and Bystroff, C. (2009) Complementation and reconstitution of fluorescence from circularly permuted and truncated green fluorescent protein. *Biochemistry 48*, 929−940.

(4) Kent, K. P., Oltrogge, L. M., and Boxer, S. G. (2009) Synthetic control of green fluorescent protein. *J. Am. Chem. Soc. 131*, 15988−15989.

(5) Jeong, J., Kim, S. K., Ahn, J., Park, K., Jeong, E. J., Kim, M., and Chung, B. H. (2006) Monitoring of conformational change in maltose

DOI: 10.1021/acs.biochem.5b00786
Biochemistry 2015, 54, 6263−6273

binding protein using split green fluorescent protein. *Biochem. Biophys. Res. Commun. 339*, 647−651.
(6) Rosenow, M. A., Patel, H. N., and Wachter, R. M. (2005) Oxidative chemistry in the GFP active site leads to covalent cross-linking of a modified leucine side chain with a histidine imidazole: implications for the mechanism of chromophore formation. *Biochemistry 44*, 8303−8311.
(7) Sniegowski, J. A., Lappe, J. W., Patel, H. N., Huffman, H. A., and Wachter, R. M. (2005) Base catalysis of chromophore formation in Arg96 and Glu222 variants of green fluorescent protein. *J. Biol. Chem. 280*, 26248−26255.
(8) Huang, J. R., Craggs, T. D., Christodoulou, J., and Jackson, S. E. (2007) Stable intermediate states and high energy barriers in the unfolding of GFP. *J. Mol. Biol. 370*, 356−371.
(9) Huang, Y. M., Nayak, S., and Bystroff, C. (2011) Quantitative in vivo solubility and reconstitution of truncated circular permutants of green fluorescent protein. *Protein science: a publication of the Protein Society 20*, 1775−1780.
(10) 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.
(11) Reeder, P. J., Huang, Y. M., Dordick, J. S., and Bystroff, C. (2010) A rewired green fluorescent protein: folding and function in a nonsequential, noncircular GFP permutant. *Biochemistry 49*, 10773−10779.
(12) Enoki, S., Saeki, K., Maki, K., and Kuwajima, K. (2004) Acid denaturation and refolding of green fluorescent protein. *Biochemistry 43*, 14238−14248.
(13) Andrews, B. T., Roy, M., and Jennings, P. A. (2009) Chromophore packing leads to hysteresis in GFP. *J. Mol. Biol. 392*, 218−227.
(14) Rosenman, D. J., Huang, Y. M., Xia, K., Fraser, K., Jones, V. E., Lamberson, C. M., Van Roey, P., Colon, W., and Bystroff, C. (2014) Green-lighting green fluorescent protein: faster and more efficient folding by eliminating a cis-trans peptide isomerization event. *Protein science: a publication of the Protein Society 23*, 400−410.
(15) Reid, B. G., and Flynn, G. C. (1997) Chromophore formation in green fluorescent protein. *Biochemistry 36*, 6786−6791.
(16) Tsien, R. Y. (1998) The green fluorescent protein. *Annu. Rev. Biochem. 67*, 509−544.
(17) Zhang, Q., Chen, X., Cui, G., Fang, W. H., and Thiel, W. (2014) Concerted asynchronous hula-twist photoisomerization in the S65T/H148D mutant of green fluorescent protein. *Angew. Chem., Int. Ed. 53*, 8649−8653.
(18) Kent, K. P., Childs, W., and Boxer, S. G. (2008) Deconstructing green fluorescent protein. *J. Am. Chem. Soc. 130*, 9664−9665.
(19) Dunbrack, R. L., Jr., and Karplus, M. (1993) Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J. Mol. Biol. 230*, 543−574.
(20) Lovell, S. C., Word, J. M., Richardson, J. S., and Richardson, D. C. (2000) The penultimate rotamer library. *Proteins: Struct., Funct., Genet. 40*, 389−408.
(21) Malakauskas, S. M., and Mayo, S. L. (1998) Design, structure and stability of a hyperthermophilic protein variant. *Nat. Struct. Biol. 5*, 470−475.
(22) Looger, L. L., Dwyer, M. A., Smith, J. J., and Hellinga, H. W. (2003) Computational design of receptor and sensor proteins with novel functions. *Nature 423*, 185−190.
(23) Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Rothlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Barbas, C. F., 3rd, Hilvert, D., Houk, K. N., Stoddard, B. L., and Baker, D. (2008) De novo computational design of retro-aldol enzymes. *Science 319*, 1387−1391.
(24) Rothlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J. L., Althoff, E. A., Zanghellini, A., Dym, O., Albeck, S., Houk, K. N., Tawfik, D. S., and Baker, D. (2008) Kemp elimination catalysts by computational enzyme design. *Nature 453*, 190−195.
(25) Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science 302*, 1364−1368.
(26) Pitman, D. J., Schenkelberg, C. D., Huang, Y. M., Teets, F. D., Ditursi, D., and Bystroff, C. (2014) Improving computational efficiency and tractability of protein design using a piecemeal approach. A strategy for parallel and distributed protein design. *Bioinformatics 30*, 1138.
(27) Huang, Y. M., and Bystroff, C. (2013) Expanded explorations into the optimization of an energy function for protein design. *IEEE/ACM Trans. Comput. Biol. Bioinf. 10*, 1176−1187.
(28) Do, K., and Boxer, S. G. (2011) Thermodynamics, kinetics, and photochemistry of beta-strand association and dissociation in a split-GFP system. *J. Am. Chem. Soc. 133*, 18078−18081.
(29) Kent, K. P., and Boxer, S. G. (2011) Light-activated reassembly of split green fluorescent protein. *J. Am. Chem. Soc. 133*, 4046−4052.
(30) Cabantous, S., Nguyen, H. B., Pedelacq, J. D., Koraichi, F., Chaudhary, A., Ganguly, K., Lockard, M. A., Favre, G., Terwilliger, T. C., and Waldo, G. S. (2013) A new protein-protein interaction sensor based on tripartite split-GFP association. *Sci. Rep. 3*, 2854.
(31) 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.
(32) Xiong, H., Buckwalter, B. L., Shieh, H. M., and Hecht, M. H. (1995) Periodicity of polar and nonpolar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides. *Proc. Natl. Acad. Sci. U. S. A. 92*, 6349−6353.
(33) Lamb, R. A., and Krug, R. M. (2001) Orthomyxoviridae: the viruses and their replication. In *Fields Virology* (Fields, B. N., Knipe, D. M., Howley, P. M., and Griffin, D. E., Eds.), 4th ed., pp 1487−1531, Lippincott Williams & Wilkins, Philadelphia.
(34) 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.
(35) 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.
(36) Hoover, D. M., and Lubkowski, J. (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic acids research 30*, e43.
(37) Stemmer, W. P., 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.
(38) Joern, J. M. (2003) DNA shuffling. *Methods in molecular biology 231*, 85−89.
(39) Cirino, P. C., Mayer, K. M., and Umeno, D. (2003) Generating mutant libraries using error-prone PCR. *Methods in molecular biology 231*, 3−9.
(40) Orm, 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.
(41) Llopis, J., McCaffery, J. M., Miyawaki, A., Farquhar, M. G., and Tsien, R. Y. (1998) Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proc. Natl. Acad. Sci. U. S. A. 95*, 6803−6808.
(42) Pitman, D. J., Banerjee, S., Macari, S. J., Castaldi, C. A., Crone, D. E., and Bystroff, C. (2015) Exploring the folding pathway of green fluorescent protein through disulfide engineering. *Protein science: a publication of the Protein Society 24*, 341−353.
(43) Seifert, M. H., Georgescu, J., Ksiazek, D., Smialowski, P., Rehm, T., Steipe, B., and Holak, T. A. (2003) Backbone dynamics of green fluorescent protein and the effect of histidine 148 substitution. *Biochemistry 42*, 2500−2512.
(44) Helms, V., Straatsma, T. P., and McCammon, J. A. (1999) Internal Dynamics of Green Fluorescent Protein. *J. Phys. Chem. B 103*, 3263−3269.

DOI: 10.1021/acs.biochem.5b00786
Biochemistry 2015, 54, 6263−6273

(45) Remaut, H., Rose, R. J., Hannan, T. J., Hultgren, S. J., Radford, S. E., Ashcroft, A. E., and Waksman, G. (2006) Donor-strand exchange in chaperone-assisted pilus assembly proceeds through a concerted beta strand displacement mechanism. *Mol. Cell 22*, 831−842.
(46) 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.
(47) Barondeau, D. P., Putnam, C. D., Kassmann, C. J., Tainer, J. A., and Getzoff, E. D. (2003) Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proc. Natl. Acad. Sci. U. S. A. 100*, 12111−12116.
(48) Shakhnovich, E. I., and Gutin, A. M. (1993) A new approach to the design of stable proteins. *Protein Eng., Des. Sel. 6*, 793−800.
(49) Desmet, J., De Maeyer, M., Hazes, B., and Lasters, I. (1992) The dead-end elimination theorem and its use in protein side-chain positioning. *Nature 356*, 539−542.
(50) Goldstein, R. F. (1994) Efficient rotamer elimination applied to protein side-chains and related spin glasses. *Biophys. J. 66*, 1335−1340.
(51) Pierce, N. A., Spriet, J. A., Desmet, J., and Mayo, S. L. (2000) Conformational splitting: A more powerful criterion for dead-end elimination. *J. Comput. Chem. 21*, 999−1009.
(52) Gordon, D. B., and Mayo, S. L. (1998) Radical performance enhancements for combinatorial optimization algorithms based on the dead-end elimination theorem. *J. Comput. Chem. 19*, 1505−1514.
(53) Bystroff, C. (2002) MASKER: improved solvent-excluded molecular surface area estimations using Boolean masks. *Protein Eng., Des. Sel. 15*, 959−965.

DOI: 10.1021/acs.biochem.5b00786
Biochemistry 2015, 54, 6263−6273