

Article

Not peer-reviewed version

---

# Challenges and Solutions for Leave-One-Out Biosensor Design in the Context of a Rugged Fitness Landscape

---

[Shounak Banerjee](#) , [Keith Fraser](#) , Donna E Crone , Jinal C Patel , Sarah Bondos , [Christopher Bystroff](#) \*

Posted Date: 24 September 2024

doi: 10.20944/preprints202409.1923.v1

Keywords: green fluorescent protein; biosensors; folding; library screening; biosensor materials; computational design; folding energy landscape; chromophore; ultrabithorax fibers



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Article

# Challenges and Solutions for Leave-One-Out Biosensor Design in the Context of a Rugged Fitness Landscape

Shounak Banerjee <sup>1,2</sup>, Keith Fraser <sup>1</sup>, Donna E. Crone <sup>1</sup>, Jinal C. Patel <sup>1</sup>, Sarah Bondos <sup>3</sup> and Christopher Bystroff <sup>1,4,\*</sup>

<sup>1</sup> Biological Sciences, Computer Science, Rensselaer Polytechnic Institute, Troy NY 12180, Rensselaer Polytechnic Institute, Troy NY 12180

<sup>2</sup> Los Alamos National Laboratory, Los Alamos, NM 87545

<sup>3</sup> Medical Physiology, Texas A&M University, College Station, TX 77843

<sup>4</sup> Computer Science, Rensselaer Polytechnic Institute, Troy NY 12180, Rensselaer Polytechnic Institute, Troy NY 12180

\* Correspondence: bystrc@rpi.edu; Tel.: 518-276-3185

**Abstract:** The Leave-One-Out (LOO) Green Fluorescent Protein (GFP) approach to biosensor design combines computational protein design with split protein reconstitution. LOO-GFPs reversibly fold and gain fluorescence upon encountering its target peptide, which can be redefined by computational design of the LOO site. Such an approach can be used to create reusable biosensors for the early detection of emerging biological threats. Enlightening biophysical inferences for nine LOO-GFP biosensor libraries are presented, with target sequences from dengue, influenza, or HIV, replacing beta strands 7, 8 or 11. An initially low hit rate was traced to components of the energy function, manifesting in the over-rewarding of over-tight side chain packing. Also, screening by colony picking required a low library complexity, but designing a biosensor against a peptide of at least 12 residues requires a high complexity library. This double-bind was solved using a "piecemeal" iterative design strategy. Also, designed LOO-GFPs fluoresced in the unbound state due to unwanted dimerization, but this was solved by fusing a fully functional prototype LOO-GFP to a fiber-forming protein, *Drosophila* ultrabithorax, creating a biosensor fiber. One influenza hemagglutinin biosensor is characterized here in detail, showing a shifted excitation/emission spectrum, a micromolar affinity for the target peptide, and an unexpected photo-switching ability.

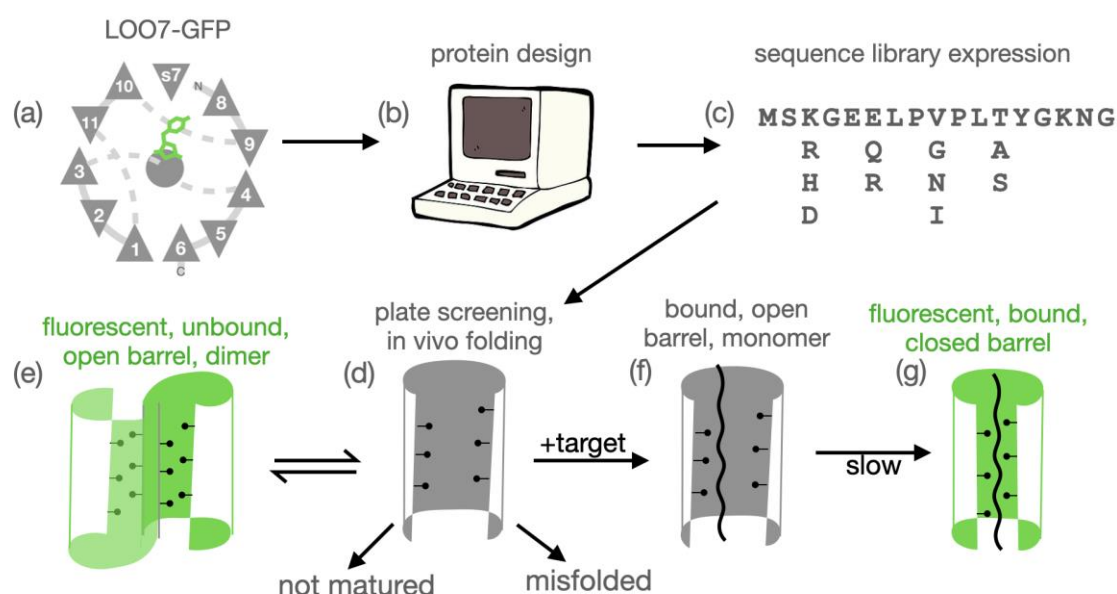
**Keywords:** green fluorescent protein; biosensors; folding; library screening; biosensor materials; computational design; folding energy landscape; chromophore; ultrabithorax fibers

## 1. Introduction

In the absence of effective vaccines for viral diseases, early detection represents the most effective strategy for mitigating disease spread. Insect-borne viruses provide an easily collected detection target where the virus resides in high concentration. If we envision large-scale viral detection systems on mosquito traps or in sewer lines<sup>1</sup>, economic concerns force us to avoid the reagents and equipment needed for *in-situ* PCR-based monitoring<sup>2</sup>. An alternative is to detect virus-specific proteins using a computationally designed fluorescent biosensor, which requires simple reagents, can be repeatedly re-used by unfolding and refolding, and can be monitored *in-situ* using a blue light and a camera.

With this aim we have developed leave-one-out green fluorescent protein (LOO-GFP), a platform for protein design that enables sequence-specific binding to short peptides producing a fluorescent signal<sup>3-5</sup>. The platform (Figure 1) was inspired by many known instances of splitting and reconstituting protein structure and function<sup>6-11</sup>. LOO-GFP is green fluorescent protein that has been circularly permuted and split to remove one of the eleven beta strands that make up the GFP barrel. There are eleven possible LOO-GFPs. Six of these (removal of beta strands 4, 5, 7, 8, 9, 10 or

11, See Figure 1(a)) are known to reconstitute fluorescence *in vivo*<sup>12</sup>. These are denoted LOO $n$ -<target>, where  $n$  is the number of the strand left-out and <target> is the source of the target sequence (or "GFP" if the target is the wild-type left-out strand). LOO $n$ -<target> is also used to refer to a library of LOO $n$  sequences designed and screened against <target>. LOO $n$ -<target>:X is a specific clone from that library, and LOO $n$ -<target>:X•Y is that protein bound to peptide Y. Algorithms and tools for designing the empty strand site to specifically accept a different exogenous peptide sequence have been developed<sup>4</sup>, along with high throughput screening methods for identifying good biosensor sequences in a library<sup>4</sup>.



**Figure 1.** (a) TIPS<sup>57</sup> diagram of LOO7-GFP with bound target peptide s7 replacing beta strand 7. (b) Computational protein design of the surrounding residues using Rosetta or DEEdesign accommodates the target peptide. (c) Design followed by simulated melting generates a sequence expression of desired complexity. The physical library is synthesized by degenerate-oligonucleotide assembly PCR and transformed into bacteria. (d) Individual clones coexpressed with target may fold. Or, they may misfold and/or not mature. (e) Correctly folded, unbound LOO-GFP, an open barrel, may form a dimeric state which matures and fluoresces. (f) Binding of target peptide shifts equilibrium to the bound state. (g) Target-bound LOO-GFP may or may not undergo barrel closure, the kinetic slow step, leading to maturation/fluorescence.

We have succeeded in computationally designing a LOO7-GFP to recognize a specific sequence from H5N1 influenza (Thailand 2004 strain) haemagglutinin (HA)<sup>4</sup>. When co-expressed with a 12-residue HA peptide (SSHEVSLGVSSA) in *E. coli*, a library of biosensor sequences (LOO7-HA4) produced dozens of fluorescent clones, and selected clones were studied in detail. But the *in vitro* complementation of clone LOO7-HA4:DS2 and the peptide HA revealed a surprising drop in fluorescence, opposite of the expected<sup>1</sup>. This drop was accompanied by a transition from dimer to a monomer. To our surprise, the dimeric unbound biosensor was fluorescent but the peptide-bound monomer was not! These results and challenges set the stage for the current work.

Immobilization of LOO-GFP should prevent dimerization and its consequent unbound autofluorescence. To test this hypothesis, we fused LOO-GFPs N-terminally to *Drosophila* ultrabithorax (Ubx), which forms fibers through dityrosine bonds<sup>13</sup>. LOO8-GFP fused to Ubx (LOO8-GFP-Ubx) forms fibers identical to unadulterated Ubx, and fluoresces when co-expressed with GFP strand 8 (s8)<sup>39</sup>. Background fluorescence was observed in the polymerized fibers, but was eliminated by diluting LOO8-GFP-Ubx 1:10 with wildtype Ubx. However, to avoid a permanent loss of fluorescence upon refolding, we needed to add 1.25M sodium chloride, presumably to mask the highly charged Ubx domain. The result was re-usable and programmable biosensor fibers<sup>39</sup>. In the

fiber construct reported here, the highly charged Ubx is blocked by a bulky spacer, maltose binding protein (MBP) inbetween the LOO-GFP and Ubx domains. This construct successfully recovers fluorescence in the presence of its peptide, and not in its absence, when the biosensor is regenerated by refolding.

Having eliminated all false positives of dimerization, we nonetheless found a very low rate of *true* positives in library screens. Our *native* format library screens described herein has the target peptide fused genetically to the biosensor, therefore the biosensor cannot exist in the unbound state and cannot dimerize. In the *native* screen all fluorescent colonies are true positives. But we saw very few fluorescent colonies, even when the library complexity was low enough to screen efficiently (1000-3000 sequences). We traced the problem to a combination of a prohibitively large design space (28 to 38 positions were designable) and a rough energy landscape of side chain packing in the core of the barrel. The rough energy landscape means that there is a "domino effect" in side chain packing, making the design space highly covariant and difficult to sample. Reducing the number of designable positions to 7 to 13 positions, reproducibly produced fluorescent colonies (Table 1), but this required using a "piecemeal" strategy to design the complete peptide binding site in pieces<sup>14</sup>.

Confusingly, tight packing in protein design has been variously credited for rigidity and thermostability<sup>15-20</sup>, but also cited for loss of native-like structure<sup>21</sup>. Perhaps the outcome depends on the protein architecture. GFP is a beta barrel protein with a fixed and immutable contained volume. As such, it may be more sensitive than other architectures to overpacking in the core as a result of designed mutations<sup>22-24</sup>. In our analysis of the library screening results from a structural perspective, we attribute the low hit rate to the tight packing.

Acting on the hypothesis that overpacking led to false positives, we designed a new influenza biosensor library called LOO7-HA5. Rational choices were made to the design palette to reduce overpacking. As intended, the library produced a sequence showing a gain in fluorescence upon binding. The increased QY is shown to be the result of a closed, native-like barrel.

These experiments, the challenges and their solutions, have broadened our understanding of the design space of GFP, painting a picture of a rough fitness landscape of design<sup>23,25</sup>, and leading to a new strategy for design -- a fiber-based method for LOO-GFP biosensor library screening, and a piecemeal approach for iterative sampling of the design space.



Table 1.

[illegible]

## 2. Materials and Methods

### 2.1. Target Selection

Target peptides were chosen based on sequence similarity to pertinent LOO strands. A manually curated structure-based sequence alignment of all fluorescent proteins with known structures (ca. 2012) was used to create a PROSITE<sup>58</sup> format regular expression for input to an in-house pattern search tool. Square brackets [] indicate allowed, curly brackets {} indicate disallowed. Sequences that matched the pattern were further excluded to avoid N-glycosylation sites, where applicable.

## 2.2. Computational Protein Library Design

In this discussion, "design space" refers to the number of residue positions that are variable, while "design palette" refers to set of amino acids at each position of a design space. In computational protein design, the input is a design palette and the output is a "library", which is then encoded using degenerate codons, and then screened experimentally. "Complexity" is the number of sequence

permutations encoded in the library. The "template" consists of backbone atoms and all non-designed side chain atoms, including chromophore and waters. Residue numbering for all sequences discussed here is taken from superfolder GFP (sfGFP, PDB:2B3P), regardless of any truncations, extensions, or circular permutations.

The LOO7-HA4 library was calculated using DEEdesign as previously described<sup>4</sup>. The input design palette was set by mutating the LOO strand to the target sequence, then creating the palette of each nearest neighbor side chain manually by inspection of the structure in the MOE molecular modeling suite (Chemical Computing Group, Canada) and choosing amino acids. The palette and the template are input to the design program DEEdesign. The palette (or library) is encoded as a plain-text character string with brackets for designed residues. Within DEEdesign, the library expression was narrowed to the desired complexity using the Dead-end Elimination algorithm<sup>52</sup> (DEE) followed by Simulated Melting<sup>4</sup> (SM). Table 1 shows the output libraries.

All other libraries were calculated using the Rosetta Macromolecular Modeling suite<sup>26,27</sup> (Rosetta) using a parallel computing script called "*plastic protein design*"<sup>53</sup> (PPD). PPD attempts to capture the dynamics of backbone and sidechain atoms during design. To start, we generate a template ensemble using Rosetta Backrub (BR) tool<sup>28</sup>, where parameters were tuned as described previously<sup>29</sup>. Then, in parallel calculations, individual compute nodes proceeded with Rosetta protein design independently, using the Metropolis Monte Carlo algorithm to select side chain rotamers from the Dunbrack backbone-dependent rotamer library<sup>30</sup> subject to the *talaris2013* scoring function<sup>27</sup>. In PPD, compute nodes share their progress with each other periodically. If a side chain rotamer is accepted by all the members of the template ensemble then it is kept, even if each ensemble member accepts a slightly different conformation, called a sub-rotamer. This method was meant to select for side chain rotamers that move in concert with the backbone. The improvement or lack thereof of the PPD approach as compared to static template methods is difficult to experimentally verify, but several protein designers tout the importance of modeling motion when designing proteins<sup>31–33</sup>.

Rational redesign of the LOO7-HA5 library was performed using MOE to evaluate a library that was generated as follows. Briefly, homology models of LOO7-HA4<sup>4</sup> in the pre- and post-chromophore maturation states were generated using Rosetta, with the Cycle3 R96M mutant (PDB:2AWJ) as the template for the pre-maturation state and superfolder GFP (PDB: 2B3P) as the template for the post-maturation state. Background mutations were simulated for transitions to the Superfolder OPT<sup>8</sup> template. For both models, iterations of mutation and side chain packing were performed in torsion space, using the Rosetta PackRotamersMover function and subsequent energy minimization using the MinMover function, both using the *talaris2013* scoring function<sup>27</sup>. This protocol was used for subsequent determination of minimum energy configurations (MEC). Rotamers compatible with both states of the protein were selected manually. Starting from the MEC, SM was performed to generate libraries of the desired complexity. Buried waters were added to MECs using DOWSER<sup>34</sup>.

Manual intervention in library design was carried out as follows. The design selections after SM were compared to a manually curated sequence profile representing several GFP homologs, based on structural alignment. The library was pruned or otherwise modified based on visual assessment of side chain's environment in MOE, along with the presence or absence of the amino acid in the sequence profile of fluorescent proteins, while considering the placement of internal waters. The resulting library (LOO7-HA5) after computational and manual design before experimental screening had a complexity of 276480 sequences.

"Piecemeal protein design" is a strategy to reduce the design space, thereby increasing the likelihood of success in finding a fluorescent colony (hit). The target, usually a 12 - 14 residue peptide, is divided into pieces which may be as small as two residues. Each piece is used to define a design space as above. After computational design, experimental screening and sequence analysis of the hits, neighboring pieces are joined, and the hit sequences are used to define the design palette for another round of experimental screening (no additional computational screening was done). The screened libraries of the pieces were coalesced to make a library for the full-length of the target peptide.

### 2.3. Library Construction and Plate Screening

Gene libraries were generated by encoding all mutations from the computationally-designed library from DEEdesign or Rosetta at each position with degenerate codons, then preparing the gene for assembly PCR using DNAworks<sup>36</sup>. The output of DNAworks are the sequences of overlapping, single-stranded, forward and reverse 60-base oligonucleotides optimized for a 60°C annealing temperature. Assembly PCR<sup>37</sup> was performed to link all overlapping oligos, and in doing so incorporating variable codons randomly. Inclusion bias at this step was tested for as described in Supplementary Data and found to be absent. The first forward oligo contains a NcoI site and the last reverse oligo an EcoRI site. To make the oligo pool, 1 picomole of each oligo was added to a standard 50 µL Phusion DNA polymerase PCR mixture (New England Biolabs). 25 cycles of amplification were performed at an annealing temperature of 60°C and with an extension time of 1 min. 1 µL of the product from assembly PCR was used as the template for standard amplification PCR using the terminal oligonucleotides as primers.

Most of the libraries were expressed in the "native" library format with the target sequence replacing the LOO segment and without circular permutation. The rest of the gene outside of the target sequence contained the designed mutations as degenerate codons. The purified DNA library was cut with NcoI and EcoRI and cloned into the pET28a+ vector for expression in BL21(DE3) *E. coli*. Transformed bacteria were plated on nitrocellulose membranes over selective media. Protein expression was then induced by transferring membranes to plates containing 0.5 mM IPTG and antibiotics. Fluorescent colonies were grown up and sequenced (MCLAB, S. San Francisco). Table 1 shows the libraries, with highlighting (black background) for hits, meaning amino acids found in the fluorescent colonies, if any. The sequences were modeled onto the template structure and inspected using MOE.

Libraries LOO7-HA4, LOO7-HA5 and LOO7-EDIII were screened in Leave-One-Out (LOO) format, constructed by permuting, and truncating the coding sequence to remove the target strand. A glycine-rich linker (GGSGGT) was used to bridge the original N- and C- termini. A degenerate codon-containing oligonucleotide set, with 5' NcoI and 3' EcoRI restriction sites, was used to assemble each LOO library. After cutting with NcoI/EcoRI this library was inserted into the pCDFDuet-1 vector at multiple cloning site 1. The target peptide was fused C-terminally to *Synechocystis sp.* intein, as described previously<sup>38</sup> and was inserted into multiple cloning site 2. Constructs were confirmed by restriction digest and sequencing.

LOO7-EDIII and LOO8-NS1#4 were screened in the Leave-It-In format (LII) format. Termini were linked as for LOO format, where the target peptide serving as the new N-terminus. The purified DNA library was cut with NcoI and EcoRI and cloned into the pET28a+ vector for expression in BL21 *E. coli*. Plate screening was done as for the native libraries. If a library was screened in more than one format, then *native* format was screened first, followed by LII, followed by LOO.

#### 2.4. LOO7-HA5:ES1 Expression and Purification

One colony of LOO7-HA5, called "ES1", was selected for detailed analysis. LOO7-HA5:ES1 was expressed in the target peptide-bound state and purified, as previously described for LOO7-HA4<sup>4</sup>. The only exception is that expression was induced at 18°C instead of 37°C. The purified and dialyzed biosensor-peptide complex LOO7-HA5:ES1•HA5 was buffer exchanged into denaturing buffer containing 6M guanidinium hydrochloride (GndHCl) and incubated at room temperature for 2 hours to remove the HA5 peptide. The denatured protein solution was passed through a 1 cm column of pre-equilibrated HisPur resin (Thermo Scientific) by gravity flow. Three column volumes of denaturing buffer were used to wash away any remaining peptide. Slow refolding was then performed by using a gradient to gradually remove GndHCl. This was followed by the addition of 6 mL of elution buffer to elute the protein from the column. The refolded LOO protein, now devoid of the "priming" target peptide, was dialyzed into a buffer containing 10mM sodium phosphate and 5mM sodium chloride at pH 8.0 overnight at 4°C.

#### 2.5. LOO7-HA5:ES1 Size Exclusion Chromatography

LOO7-HA5:ES1 protein was equilibrated with 10-fold molar excess of target HA5 peptide (conc. approx. 1mM) in P buffer (10mM sodium phosphate and 5mM sodium chloride at pH 8.0) 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 as the mobile phase using P buffer. The eluted protein was monitored by measuring the absorbance at 280nm. Protein that was not equilibrated with the target peptide was run using the same procedure, for comparison. Molecular weights for elution peaks were estimated by retention times, comparing to Biorad's Gel Filtration Standard.

## 2.6. LOO7-HA5:ES1 Peptide Binding Affinity

Fluorescence based manual mixing experiments were used to measure the binding kinetics of LOO7-HA5:ES1 for the HA5 peptide. Synthetic and >95% pure peptide (GenScript) was used for all binding experiments. 300  $\mu$ L of peptide at varying concentrations were mixed with 2700  $\mu$ L of 0.1  $\mu$ M LOO-GFP biosensor in P buffer. The peptide concentrations used were all super-stoichiometric, from 0.1  $\mu$ M to 100 $\mu$ M. The upper and lower limits for signal amplitudes were determined by starting with 100  $\mu$ M peptide and lowering the concentration in a 2-fold dilution series. The upper limit of peptide concentration is defined here as the concentration of peptide at which the signal amplitude was maximum. The lower limit was obtained by dilution with 300  $\mu$ L buffer. Fluorescence trajectories were recorded for 300 seconds with a time step of 0.1 sec on a Fluorolog Tau 3A spectrofluorometer (Jobin-Yvon Horiba), in a 3.5 mL quartz cuvette (Starna). Excitation and emission wavelengths were set to 472 nm and 499 nm, with 3nm and 5nm slit widths respectively.

Fitting the fluorescent traces was complicated by the rapid light-induced first order decay. To factor out this decay, the fluorescence trace was fit as follows. (1) Counts were normalized to the starting value ( $4.0e5 \pm 0.5e5$  counts/sec) for each run. (2) Light-induced first order fluorescence decay was fit to the times before ( $k=0.002$  s<sup>-1</sup>) and after ( $k=0.001$  s<sup>-1</sup>) addition of the peptide. (3) The difference between these two baselines, after accounting for the dilution factor (0.9 or 0.95), was taken as the peptide-induced fluorescence gain. The rate of peptide-induced fluorescence gain was fit ( $k_b = 0.37$  s<sup>-1</sup>  $\pm 25\%$ ) by global, non-linear least squares. Fits were validated by a flat, zero residual.

## 2.7. Fluorescence Quantum Yield

For all variants discuss in this paper, the relative fluorescence quantum yield (QY) was calculated by measuring counts-per-second (cps) at 508nm with 485nm excitation on a Fluorolog Tau 3A spectrofluorometer using a 3.5 mL quartz cuvette with 1 cm pathlength, using slit widths 2nm and 2nm for excitation and emission respectively. Measured cps was compared to that of 1.0  $\mu$ M sfGFP-OPT in P buffer at pH 8 measured using the same settings. The ratio was scaled by the  $\mu$ M concentration of the variant to get the QY.

## 2.8. SGMU Fiber Preparation

Cloning, purification, and fiber formation for fusion proteins eGFP-Ubx, mCherry-Ubx, and LOO8-GFP-Ubx are described elsewhere<sup>39</sup>. The new construct presented here is a multi-part fusion composed of Sortase A (SrtA), LOO-GFP, maltose binding protein (MBP) and ultrabithorax (Ubx). The construct SrtA-LOO11-GFP-MBP-Ubx is named as "LOO11-SGMU" (see Supporting Data). The construct includes an N-terminal histidine tag followed by the target (priming) peptide, followed by SrtA, followed by the SrtA cut site, followed by LOO-GFP, then MBP, then Ubx.

E coli BL21(DE3) cells expressing SGMU constructs in the pET28a+ vector were induced with 1mM IPTG and allowed to grow overnight at 28°C. The cell pellets were resuspended in Binding Buffer (50mM Na<sub>2</sub>HPO<sub>4</sub>, 500 mM NaCl, pH 8) with Protease Inhibitor Cocktail and DNase (Thermo Fisher Scientific, USA) added. Cells were lysed by sonication. Cell debris (pelleted for 30 minutes at 12,000 rpm) was discarded, and the supernatant was filter sterilized using a 0.45  $\mu$ m filter. LOO11-SGMU protein was purified from lysate using a His-Trap 5mL affinity column washed with wash buffer (50mM Na<sub>2</sub>HPO<sub>4</sub>, 500 mM NaCl, 100 mM Imidazole, pH 8), eluted with elution buffer (50mM



Na<sub>2</sub>HPO<sub>4</sub>, 500 mM NaCl, 500 mM Imidazole, pH 8). SDS-PAGE/Western showed the expected 130kD band, along with several degradation products.

A shallow metal tray was filled to the brim with Fiber Buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 500 mM NaCl, 5 % Glucose, pH 8) with the buffer domed slightly above the edges of the tray. 2-3 mL of purified SGMU at a concentration of 3 mg/mL was layered over the buffer, drop by drop, moving the pipette around the tray while dropping and without stirring. Tray was protected overnight with an aluminum foil tent, not touching the tray or the liquid. Tray was left at room temperature from 6h to 2 days, typically 18 -20 hr. A plastic block (clean, 80-well microcentrifuge tube rack) was gently placed along the far rim of the tray. Buffer was gently added below the surface using a serological pipette, so as not to disrupt any biomaterial that had formed on the surface, as needed to make the buffer surface contact with the plastic block. Block was pulled slowly across the tray until 2–5cm remained. This procedure collected all the floating biofilm into a smaller area. By placing the tip of a paper clip just below the surface and slowly pulling, a fiber formed, which was wrapped around the paperclip. All experiments with LOO11-SGMU were carried out by soaking the fibers on the paper clip in various solutions and imaging in an inverted fluorescence microscope using excitation wavelength 450-490nm and emission wavelength is 520-560, using band-pass filters.

*SGMU Experiment 1.* A fiber was imaged. The fiber was washed with pH 2 10 mM GlycineHCl buffer for 10 minutes (fluorescence disappears). The fiber was washed with pH 8 50 mM sodium phosphate buffer for 10 minutes.

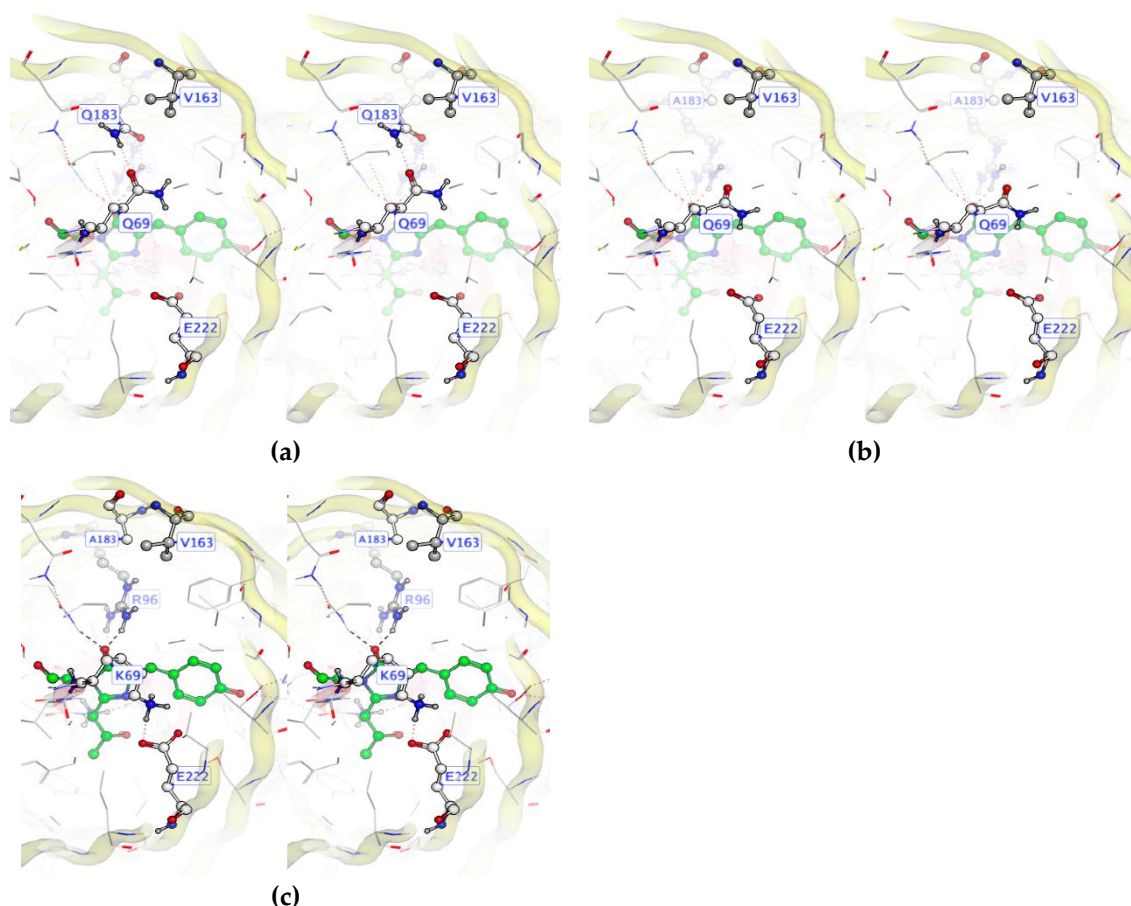
*SGMU Experiment 2.* A fiber was imaged. The fiber was soaked in 100mM CaCl<sub>2</sub> for 10 minutes. Experiment 1 follows. Fluorescence is not expected to reappear.

*SGMU Experiment 3.* Experiment 2 is carried out. Then fiber was washed with 1mM target peptide in pH 8 50 mM sodium phosphate buffer for 10 minutes, then reimaged. Alternatively, fiber was soaked in a solution containing a heat denatured protein containing the target sequence.

### 3. Results

#### 3.1. Library Screens

A full and detailed analysis of each of the eight computationally designed and experimentally screened libraries is provided as **Supplementary Data**. In summary, we found that there were two issues with the Rosetta force field that explained the results. Close packing of sidechains is rewarded instead of penalized. Also, unsatisfied hydrogen bonds are not penalized enough. Along with these issues, we also observed GFP-specific challenges in protein design, namely a rugged fitness landscape. An example of ruggedness in a fitness landscape is given in Figure 2, which shows a "domino effect" of design choices. Ruggedness means that there is no simple mutational pathway from one folded and functional state to another. Instead mutations are interdependent. A sequence may be unfit in several ways. It may not fold, or not mature the chromophore, or not bind the peptide, or it could dimerize in the unbound state. We hypothesize that the rugged landscape arises from the beta barrel topology, which is like a closed container with no room to expand or contract in the core.



**Figure 2.** Domino effect in the LOO8-NS1#2 library. (a) Target mutation A163V creates clash with Q183. (b) Q183A alleviates clash but leaves Q69 unsatisfied, compromising protein folding. (c) Q69K makes salt bridge with E222. CRO loses catalytic E222 positioning, preventing chromophore formation.

### 3.2. LOO7-HA5:ES1 Peptide Binding

Colony ES1 (see Supporting Data), the only positive in the LOO-format LOO7-HA5 library screen, underwent further study. LOO7-HA5:ES1 was expressed and purified as described above, and peptide-induced fluorescence kinetics was measured. After correcting for light-induced fluorescence decay, amplitudes of peptide-induced fluorescence gain were accurately measured and fit to peptide concentrations using the Langmuir equation for single-site isothermal binding<sup>47</sup> (Figure 4). The result  $K_d=30\ \mu\text{M}$  is slightly weaker than the  $K_d=9.0\ \mu\text{M}$  affinity previously reported for LOO7-HA4:DS2 using the shorter HA4 peptide<sup>4</sup> (Figure 3). The kinetics of peptide-induced fluorescence gain was fit to a first-order inverse exponential decay equation, giving rate  $k_b = 0.4\ \text{s}^{-1}$  ( $t_{1/2}=1.7\text{s}$ ) which is consistent with protein folding being the rate limiting step<sup>4</sup>. SEC was performed with and without the HA5 peptide, showing a transition from a multimeric state to a monomeric state when peptide is added (Figure 5). Similar SEC results were seen for LOO7-HA4:DS2<sup>4</sup> and for LOO7-GFP<sup>3</sup>. The major difference is that LOO7-HA4:DS2 rapidly lost fluorescence upon addition of the HA4 peptide, while LOO7-HA5:ES1 gains about 27% in QY relative to the unbound, multimeric state, when the peptide HA5 is added.

142

157

|

|

(a) wild-type

HKLEYNFNSHNVIYITADKQKNG

(b) LOO7-HA4<sup>6</sup>

HKLEYNFSSHEVSLGVSSAKNG

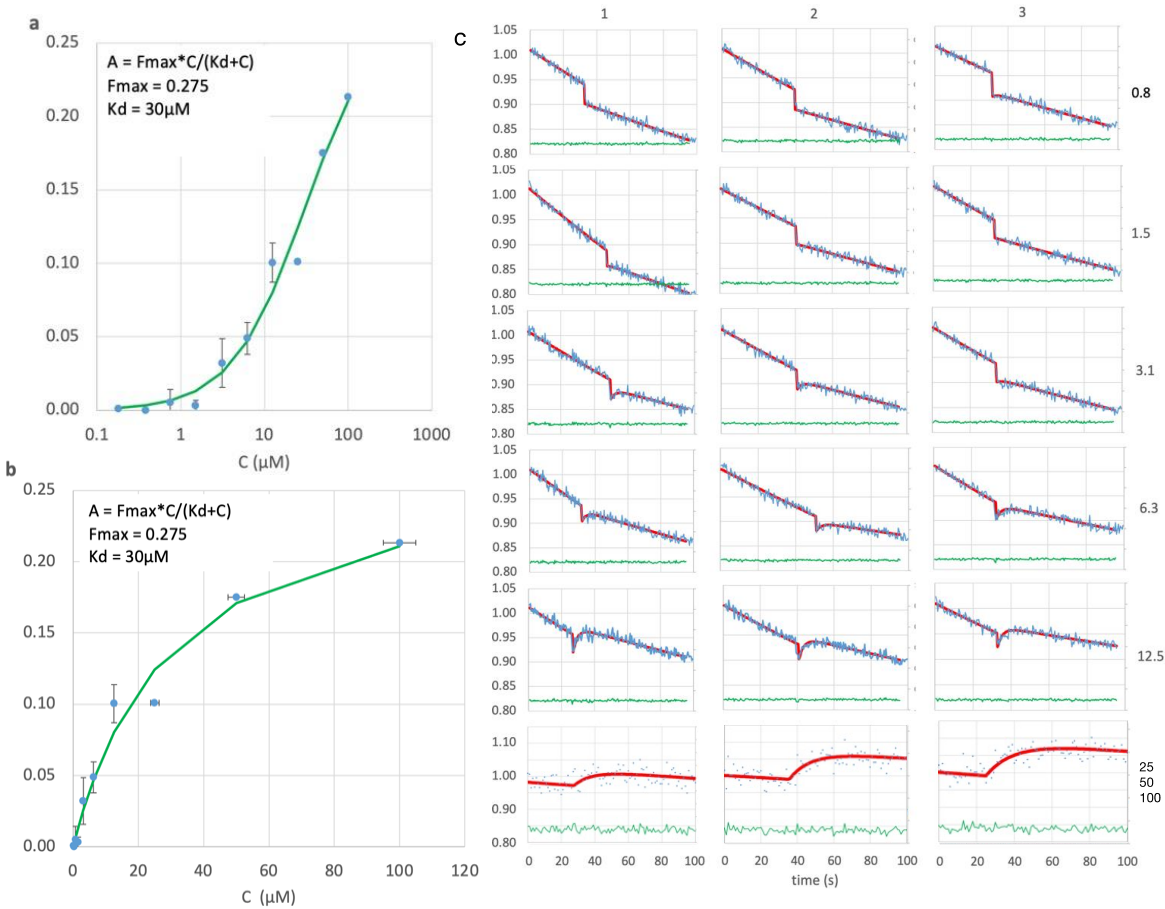
(c) LOO7-HA5

HKLKSSWSSHEVSLGVSKQKNG

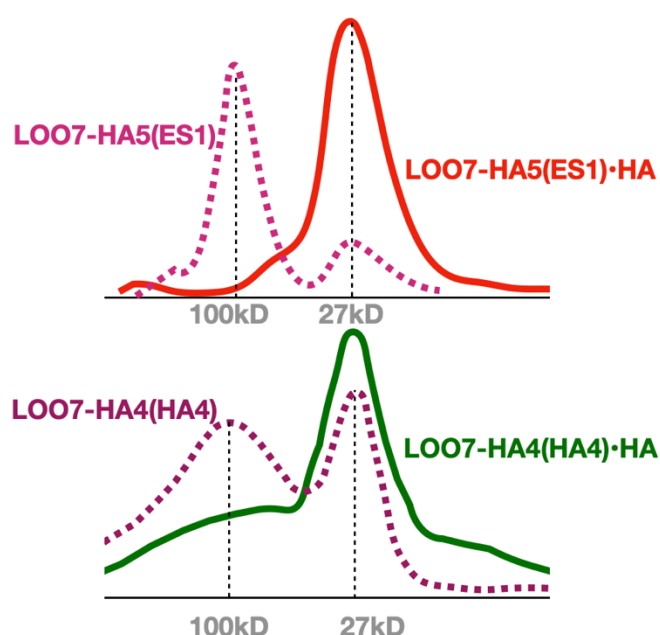
(d) LOO7 expr.

142-ZXZ[FYWLM]ZZHZ[LIVA]Z[LIVA]ZXXXX-157

**Figure 3.** (a) Wild-type sequence of strand 7 (residues 139-160) from sfGFP-OPT<sup>8</sup>. (b) LOO7-HA4 strand 7 sequence with target peptide HA4 (red) derived from residues 139-150 of Influenza A virus (A/Thailand/16/2004(H5N1)) hemagglutinin. (c) LOO7-HA5 strand 7 sequence with target peptide HA5 (red) derived from residues 135-148 of Influenza A virus. (d) PROSITE expression<sup>58</sup> used to screen for candidate LOO7 target sequences, where X={C} , Z={CP} .

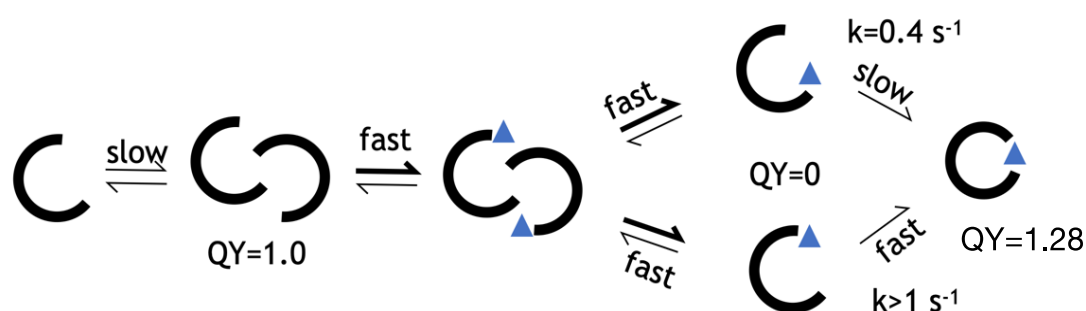


**Figure 4.** (a) Log and (b) linear scale plots for peptide concentration, C, versus observed signal amplitude, showing fit (green) to the Langmuir equation to give  $K_D=30\mu\text{M}$ .(c) Fluorescence (blue, ex485/em508 ) versus time after opening the shutter for all peptide concentrations (right side,  $\mu\text{M}$ ), in triplicate (top margin) except for 25-100 $\mu\text{M}$ . Counts (vertical axis, blue lines) were normalized to the starting value for each run. Concentrations 25-100 $\mu\text{M}$  were run at a 2x higher dilution and smaller slit width. Red : least squares fit. Green: residual of the fit.



**Figure 5.** Size exclusion chromatography of HA peptide-bound and unbound LOO7-HA constructs. Dominant multi-meric state is apparently tetramer.

Having shown that folding is the rate limiting step for peptide-induced 27% increase in the QY, we can assert that the final bound state of LOO7-HA5:ES1 is a *closed barrel*, while that of LOO7-HA4:DS2 (whose rate of fluorescence signal change was much too fast to be due to folding) is an *open barrel*. Our working model for the LOO-GFP system (Figure 6) includes a dimeric, open-barrel, unbound state, and a monomeric, closed-barrel, bound state. The lack of barrel closing is attributed to overpacking, specifically at W83 in LOO7-HA4:DS2. In energy-minimized and rotamer optimized models, W83 makes close contacts with N185 on one side and L195 on another side. On the other hand, F83 in LOO7-HA5:ES1 adopts the same rotamer but has no such clashes. A tighter interaction implies a higher entropic barrier to barrel closing.



**Figure 6.** Kinetic model for HA5. From left to right: unbound monomer, fluorescent unbound multimer, peptide-bound multimer, peptide-bound monomer with open barrel, fluorescent peptide-bound monomer with closed barrel. Values for QY and rate determined in this work.

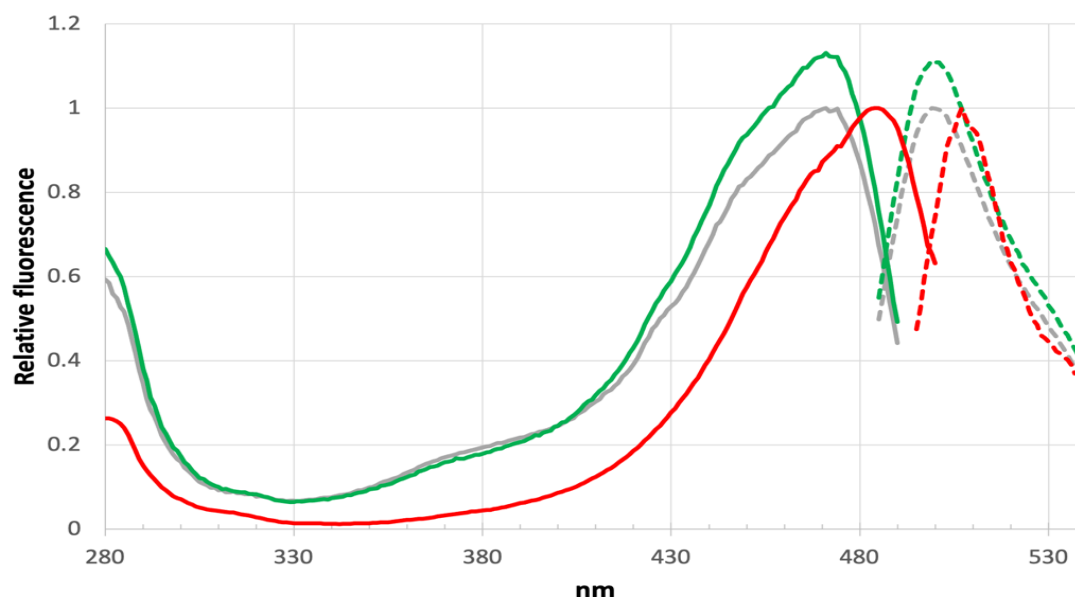
### 3.3. LOO7-HA5:ES1 Photo-Switching and Blue Shift

Peptide-bound LOO7-HA5:ES1•HA5 shows a first order decay in fluorescence upon excitation with 471 nm light. One of the other fluorescent members of the LOO7-HA4 library, a clone called LOO7-HA4:DS2<sup>6</sup>, also carried an AYG chromophore (heterogroup PIA in 5DTZ<sup>46</sup>) showing a similar albeit less pronounced light-induced first-order decay. This decay is due to the intrinsic properties of the PIA chromophore, which is known to confer reversible photo-switching ability<sup>46</sup>. Upon



irradiation, the chromophore switches to the *trans* stereoisomer, quenching fluorescence. In some cases the stereoisomer switches back spontaneously and 471/499 excitation/emission returns. In other cases, the reversal back to *cis* requires irradiation at 405 nm<sup>41</sup>. Since we were not interested in photoswitching, we made the reversion mutant LOO7-HA5:ES1/A65T to recover the wild-type TYG chromophore (heterogroup CRO), however, this resulted in no fluorescence. The appearance of colonies of the A65T revertant lacked the tint we have come to associate with a mature chromophore, therefore we speculate that the A65T revertant mutation blocked chromophore maturation in the context of the other mutations in the complex LOO7-HA5:ES1/A65T•HA5.

A significant blue shift in excitation and emission are found for LOO7-HA5:ES1•HA5 versus LOO7-HA4:DS2•HA4 or wild-type GFP, with peak excitation at 471nm and peak emission at 499nm, shifted by -13 and -9 nm respectively (Figure 7). T65A removes a hydrogen bond donor from to the N2 imidazolidone ring nitrogen, destabilizing the anionic form of the PIA chromophore. By removing that stabilizing interaction, the PIA chromophore exists more in the neutral form and its less extensive conjugation absorbs higher energy photons. In fact, comparison of the excitation/emission spectra with wild-type GFP shows not only the blue shift but also a significant increase in excitation at lower wavelengths (350-400nm). The neutral form of the chromophore absorbs at the lower end of the 350-490 nm excitation range<sup>42</sup>. The slightly increased Stokes shift in LOO7-HA5:ES1 could be due to the loss of T65 OG, which is one of the atoms in the excited state proton transfer pathway<sup>43-44</sup>.



**Figure 7.** Excitation/emission spectra. Excitation peaks (solid lines) for sfGFP-OPT in 484nm, emission peak (dashed lines) 507nm. Excitation peak for LOO7-HA5•ES1, 471nm, emission peak 499nm.

### 3.4. Experiments on Fibers

The previous fiber-based LOO-GFP construct<sup>39</sup>, called LOO8-GFP-Ubx, enhanced peptide-induced fluorescence, and eliminated the background fluorescence from dimerization. Therefore deploying LOO-GFP biosensors on Ubx fibers may be strategy to improve biosensor function. But it revealed an electrostatic interference problem between LOO-GFP and the highly-charged fiber forming protein Ubx. Quantum mechanical simulations suggest that GFP quantum yield is very sensitive to the electrostatic environment<sup>54</sup>. We sought to improve the deficiencies with the Ubx fiber system noted in our pilot studies. First, in the new SGMU construct we have inserted a 43 kD spacer, MBP, between LOO-GFP (27 kD) and Ubx (40 kD). This insertion should reduce intramolecular LOO-GFP:Ubx interactions in the fiber. Second, not all of the LOO8-GFP-Ubx monomers incorporated the co-expressed target peptide, therefore we included the target peptide in the same chain, separated by (and removable by) a calcium-activated self-cleaving protease, SrtA. Third, we

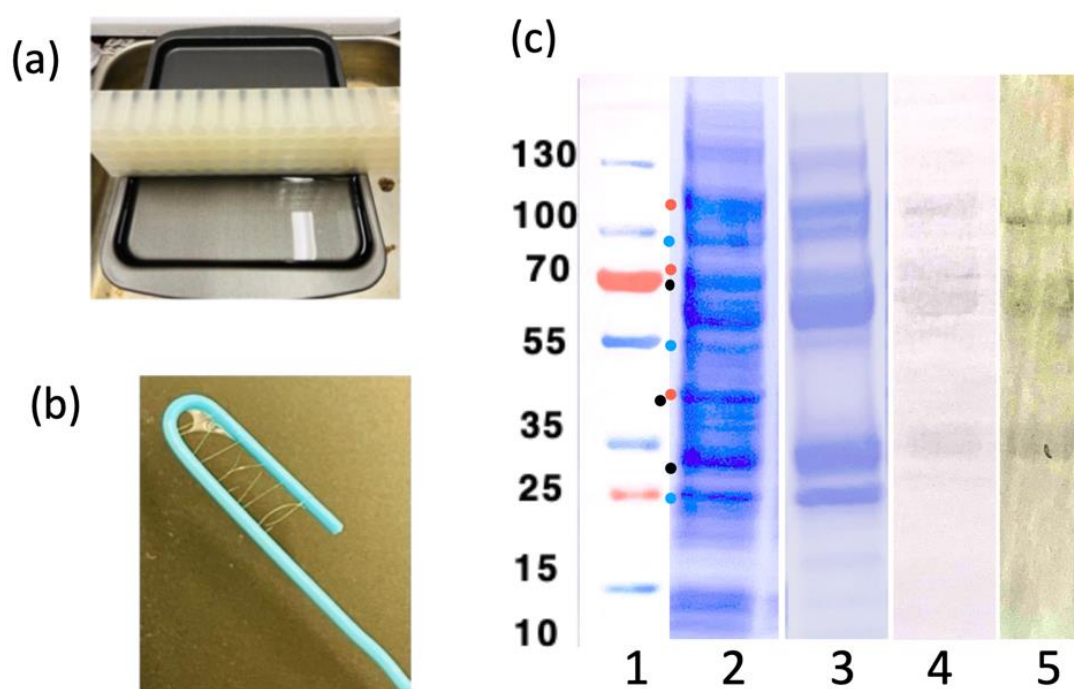
used LOO11 instead of LOO8 to avoid a patch of negatively-charged side chains near strand 8 that we hypothesized allowed LOO8:Ubx ionic interactions to block peptide binding.

Figure 8 shows the purification and the fiber-making setup. LOO11-SGMU is estimated to be at least 90% pure. The extra bands are degradation products since they react with either an anti-His tag antibody or an anti-Ubx antibody. Floppy linkers holding the component domains together are likely to be the sites of protease cleavage. It is possible that proteolytic cleavage is off-target activity of SrtA, a component of SGMU. Fortunately, the amount of full-length protein in the mix was sufficient to make fibers and characterize them. Ubx self-purifies as it forms fibers<sup>55</sup>.

To test whether the altered SGMU construct could unfold and refold in fibers, we denatured and renatured as described in SGMU Experiment 1. Unfolding and refolding LOO11-SGMU recovered approximately 50% of the original fluorescence in the presence of the covalently linked target peptide. The incomplete recovery was expected, due to the well-known "trapped state" of refolded GFP, in which approximately half of the chromophore takes on the non-fluorescent *trans* state<sup>45</sup>. Encouragingly, the partial recovery of fluorescence on refolding required no added salt or detergent, in contrast to the previous construct, LOO8-GFP-Ubx, which needed high salt to prevent irreversible electrostatic interference from Ubx. We conclude that inserting MBP between LOO-GFP and Ubx diminished the electrostatic effect.

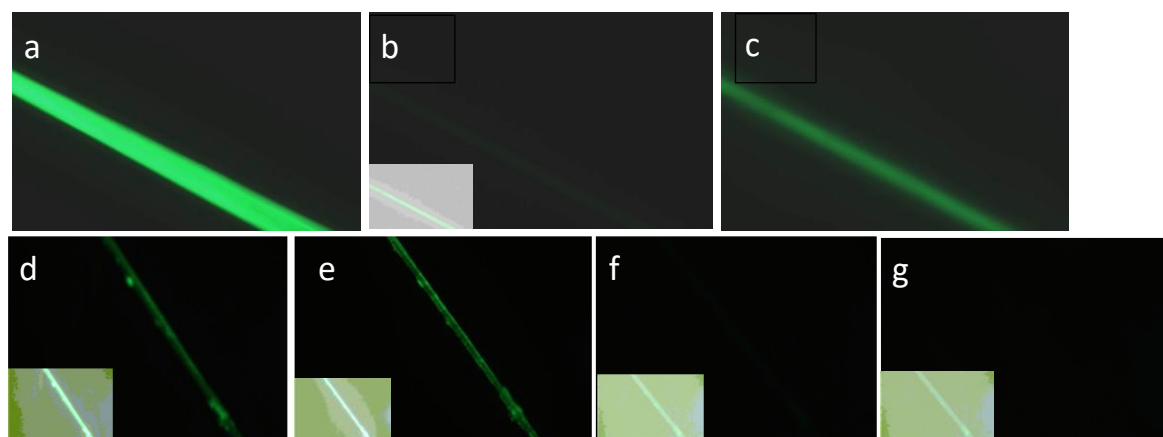
SGMU Experiment 2 produced complete loss of fluorescence after the target peptide was removed by activating the SrtA with CaCl<sub>2</sub> (Figure 9). This shows that the unbound state of LOO11-GFP is not forming dimers on the fiber as is does in solution, and that only the peptide-bound state is fluorescent. This is remarkable because the fibers were composed of 100% LOO11-SGMU, whereas in the pilot study the concentration of the sensor fusion had to be reduced to 10% (10% LOO8-Ubx, 90% plain Ubx) in order to remove the background fluorescence<sup>39</sup>.

SGMU Experiment 3, the addition of exogenous peptide, did not deliver the expected fluorescence recovery. A saturating amount of s11 peptide was added to the fiber for 10 minutes, but no appreciable increase in fluorescence was observed. We also added an unfolded GFP experimental variant (itself non-fluorescent) containing the s11 sequence, but this too did not reconstitute fluorescence. This suggests that there may still be electrostatic interference, despite the MBP spacer. Therefore the SGMU design solves several of the issues encountered in the pilot study. Work is ongoing to solve this electrostatic problem.



**Figure 8.** (a) Buffer reservoir setup, used for the making fibers. (b) A LOO11-SGMU fibers, approx. 100  $\mu\text{m}$  thick, wrapped around a paper clip. (c) SDS-PAGE analysis of the LOO11-SGMU lysate

before/after HisTrap (Ni-NTA) column chromatography. (Lane 1) Precision plus protein standards (Bio-Rad) marked in kilodaltons. (Lane 2) Lysate. (Lane 3) HisTrap elution. (Lane 4) Western blot of elution using anti-Ubx. (Lane 5) Western blot of elution using anti-Histag. All gel images are aligned using standards. Dots on lane 1 are possible degradation product sizes based on locations of linkers. Blue dots contain the N-terminal Histag. Red dots contain the C-terminal Ubx. Black dots have neither.



**Figure 9.** Results of SGMU Experiments 1, 2. (a) LOO11-SGMU fiber in fluorescence microscope using Nikon Blue filter. (b) Same fiber after glycine pH 2 buffer added. Inset in each case is an enhanced image optimized for contrast, showing presence and location of fiber. (c) Same fiber after phosphate pH 8 buffer added, observed for 5 minutes. 35% of the original fluorescence returned. (d) Different LOO11-SGMU fiber under microscope using Nikon Blue filter. (e) Same fiber after  $\text{CaCl}_2$  buffer added, observed for 10 minutes. 40% loss in brightness. (f) Same fiber after glycine pH 2 buffer added, observed for 5 minutes. Less than 1% of brightness remains. (g) Same fiber after PBS pH 8 buffer added, observed for 5 minutes. No fluorescence is recovered. Signal-to-noise is panel (c) divided by panel (g).

## 4. Discussion

### 4.1. Biosensor Library Screening

Creating LOO-GFP biosensors by computational design and plate screening for fluorescence encountered several obstacles, and each obstacle led to a solution or at least a partial solution, and each came with a lesson. First, we found that libraries in which many residues are designed, failed to produce a functional protein sequence whether the designable positions were sampled at high complexity (many sequence permutations) or at low complexity. High complexity screening relies more on the number of colonies screened, while low complexity depends more on the computational design accuracy. The beta barrel structure of GFP is unusually stable<sup>48</sup>, and has been described as having a “badlands” sequence fitness landscape<sup>25</sup>, characterized by steep energy differences and highly covariant energetic effects of mutation. We described one such covariant or “domino effect” in Figure 2. Sampling a badlands energy landscape is difficult because optimal residue choices depend on other residue choices, in a chain of dependency. The search space paradox was resolved by focusing on a smaller design space, using the Piecemeal approach<sup>14</sup>. Functional, fluorescent sequences were found when the number of designed positions was 15 or less, whether the library complexity was high, low, or medium. This sweet-spot design space size corresponds to the nearest neighbor residues of 4 to 5 consecutive positions of the target peptide. Thus, a full-length biosensor design can be accomplished in three “pieces”, using the piecemeal design approach<sup>14</sup>.

But underlying the search space problem were at least two defects in the energy function of design. Too often side chains were packed together tightly instead of finding loosely packed solutions. Loose side chain packing provides less of an entropic barrier to folding. In the spirit of flexible packing, we experimented with a design protocol called Plastic Protein Design (PPD)<sup>53</sup>, used

with Rosetta. This method favors amino acid choices that are compatible with backbone motions. But PPD did not prevent overpacking, and several examples were found where overpacking was the most likely cause of the failure of fold, bind, or catalyze chromophore maturation. We hereby assert that *an energy function for design should reward loose packing instead of rewarding close contacts*. The entropy gained should more than compensate for the loss of the short-ranged van der Waals attractive term. Alternatively, the attractive vdW force could be extended to a carbon-carbon distance of 5.8Å to capture the attractive force of desolvation<sup>50</sup>. Another shortcoming of the Rosetta scoring function, until recently, was its lack of a penalty on hydrogen bonding groups that are unsatisfied and buried in the core<sup>51</sup>.

The three-step screening process (native, LII, LOO) was intended to gradually increase the selection pressure. Our strategy was to first express all combinations of all mutations in the native format, then pick fluorescent individuals and sequence them. Second, we would express the selected sequences as a circular permutant with the target sequence at the new N or C terminus, then pick fluorescent individuals again and sequence them. Finally, we would screen the selected mutations (all combinations) as LOO-GFPs with the target expressed exogenously as a fusion to a separate protein. It was believed that many mutations would be tolerated in the native format, but that circular permutation would impose an additional energetic penalty, allowing the selection to identify the most stable designs. The LOO format imposes an additional energetic penalty by making it two pieces, intensifying the selection pressure. Though fluorescent colonies were obtained up to the LII stage, none of the selected individuals shortlisted at this stage were fluorescent as LOO-GFPs. These results and those published previously<sup>15-17,23</sup> suggest a rough fitness landscapes of intact GFPs and their self-assembling split counterparts.

#### 4.2. LOO-GFP Biosensors on Fibers

Because the stability of the LOO-GFP variants appears to limit the successful identification of active variants in a library, a universal approach to stabilize these proteins was needed. Proteins genetically fused to Ubx and incorporated into materials are remarkably stabilized. For instance, enhanced GFP (EGFP) in Ubx materials can withstand boiling, submersion in organic solvents, and even autoclaving conditions<sup>56</sup>. Thus embedding LOO-GFPs in Ubx materials may impart function in variants incapable of folding to monomers. This approach has the added benefit of immobilizing the LOO-GFP, facilitating analysis.

LOO8-GFP was added N-terminally to Ubx, and the purified construct fluoresced and formed fibers as described<sup>39</sup>. In that work, the priming strand could be removed by dropping the pH to 2. Binding of exogenous s8 peptide was demonstrated, and coincided with a gain of fluorescence, culminating in the proof of concept of a biosensor screening system with almost zero background fluorescence and almost no possibility of dimerization or aggregation. But this success was achieved only after solving a mysterious and seemingly permanent loss of function after unfolding and refolding. The loss was attributed to electrostatic interactions between LOO-GFP and Ubx, because adding high salt (1.25M NaCl) prevented the loss of biosensing<sup>39</sup>.

Another solution to the electrostatic interference problem was to introduce MBP in between LOO-GFP and Ubx. The construct s11-SrtA-LOO11-GFP-MBP-Ubx (see Supporting Data) formed fibers, fluoresced, and was able to recover fluorescence (about 50%) after unfolding and refolding using pH shift in the presence of the covalently linked target peptide (s11). No salt or detergent was required in this case. There was no fluorescence if the target peptide was removed by activating the sortase A (SrtA) followed by washing, solving the fluorescent multimer problem. The recovery of 50% of the fluorescence was expected since even full-length, wild-type GFP recovers only about half of the fluorescence upon unfolding and refolding<sup>45</sup>.

#### 4.3. Influenza Biosensor Design

In this work we re-engineered an existing LOO7-GFP biosensor for H5N1 hemagglutinin. The key objectives were to obtain a gain of fluorescence upon binding and a higher specificity, using the wild-type strand 7 peptide as the control. While both goals were met, the current prototype still binds



to the wild-type strand 7. Even the dramatic change caused by Y173S was not sufficient to find a design that excluded the original peptide from binding it. This points to an inherent conflict in this type of design problem. We want to make the fewest and most conservative changes to be assured of finding a successful sequence, but on the other hand we want to include dramatic changes so that binding specificity is maximally shifted to the new target. That goal remains..

## 5. Conclusions

The LOO-GFP format provides a promising, self-reporting, protein-only biosensor system that can be computationally designed and expressed in bacteria. It can even be immobilized on fibers to make biosensor materials. The challenges encountered in creating this capability were both computational and GFP-specific. A LOO-GFP biosensor needs to recognize and bind at least a 12-residue peptide, requiring changes at an estimated 28 to 38 positions. This creates a double bind, a high enough library complexity is both necessary for success and out of reach of experimental screening. We solved this problem by dividing the design space into manageable pieces, a method called "piecemeal". Three rounds of piecemeal design are needed to design a biosensor. Underlying the library complexity problem was the ruggedness of the fitness landscape. Observed domino effects in design choices means that functional sequences are sparse, with highly interdependent mutations. Underlying the energy landscape is the energy function, in which we discovered two critical defects. The first is the rewarding of overpacking of sidechains in the core. An energy function for design should reward loose packing instead of tight packing. The second is the absence of a penalty for missing hydrogen bonds. Also, the LOO-GFP method itself has an inherent weakness in that the unbound state is an open barrel that can form soluble homo-multimers. Our past efforts, including the ones described in this work, were constrained to solve two often conflicting optimization problems of (a) increasing binding affinity and (b) preventing homo-multimerization. Therefore most importantly, we show that immobilization of LOO-GFP on Ubx fibers simplifies the optimization to only one goal, which is increased binding affinity. It also vastly improves the signal to noise ratio, providing higher dynamic range and sensitivity for detection of pathogens at low titers.

**Supplementary Materials:** The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

**Author Contributions:** Conceptualization, C.B., K.F., and S.Ba.; methodology, S.Ba., K.F., S. Bo.; software, S.Ba. and C.B; validation, J.C.P.; formal analysis, C.B, S.Ba.; investigation, D.E.C., J.C.P., S.Ba., K.F.; writing—original draft preparation, C.B., S. Ba. and S.Bo.; writing—review and editing, C.B., S. Ba., K.F., S.Bo., D.E.C.; visualization, C.B., J.C.P., ; supervision, C.B., D.E.C., S.Bo.; project administration, C.B.; funding acquisition, C.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** NIGMS grant R01 GM099827 to C.B. and S.Bo.

**Data Availability Statement:** Sequences and structures referenced in this work are available from NCBI using the accession codes in Appendix A or found in Supplementary Data.

**Acknowledgments:** Special thanks to Kripa Nand, Thomas Jordan, Christian Schenkelberg, Ravinder Kumar, and Kacey Kilpatrick for experimental and computational contributions. Thanks to RPI's Center for Computational Innovations (CCI) for the use of the Bluegene Q Supercomputer.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A

### Accession Codes

- Sortase A. partial [*Staphylococcus aureus*]. GenBank: WP\_084935205.1, PDB: 1T2W.
- Maltose binding protein maltose/maltodextrin ABC transporter substrate-binding protein MalE [*Escherichia coli*] GenBank: EFB2378405.1. PDB: 7MQ7.
- Superfolder green fluorescent protein (sfGFP), precursor to sfGFP-OPT. GenBank: UFQ89826.1. PDB: 2B3P.

- Ultrabithorax (Ubx), isoform F [*Drosophila melanogaster*]. NCBI Reference Sequence: NP\_996219.1.
- Hemagglutinin [Influenza A virus (A/Thailand/2(SP-33)/2004(H5N1))]. GenBank: AAS65618.1.
- Dengue virus polyprotein [dengue virus type 2], includes cleavage products NS1 (776..1127) and E-protein (281..775). GenBank: WPF60237.1.
- Envelope glycoprotein [Human immunodeficiency virus 1]. GenBank: AEQ75975.1.
- Ssp dnaE intein [*Synechocystis* sp.]. UniProtKB/Swiss-Prot: P74750.2

## Appendix B

### Abbreviations

LOO, LOO7, LOO8, LOO11, Leave-One-Out protein or leave-one-out library. Number indicates GFP strand left out; LII, leave-it-in protein or leave-it-in library. Circular permutant of GFP; Native (format), library generated in the normal sequential GFP order with one strand mutated to the target sequence; GFP, green fluorescent protein; PCR, polymerase chain reaction; HA, hemagglutinin antigen; DNA, deoxyribonucleic acid Ubx, Ultrabithorax; MBP, maltose binding protein; SrtA, Sortase A; epPCR, error-prone PCR; DS, DNA shuffling; DS2, DNA shuffling clone 2; ES1, epPCR + DS clone 1; HA4, HA5, HA biosensor libraries; QY, quantum yield or relative quantum yield; vdW, van der Waals; PPD, plastic protein design; BR, Back-rub; CCI, Center for Computational Innovations; MOE, Molecular Operating Environment; MEC, Minimum energy configuration; SM, simulated melting; EDIII, dengue e-protein biosensor library; NS1, dengue non-structural protein 1 biosensor library; HIV, human immunodeficiency virus, envelope protein biosensor library; BL21, strain of *E. coli*; cps, counts per second; DEE, dead-end elimination; K<sub>d</sub>, dissociation constant; CRO, TYG chromophore; PIA, AYG chromophore; SGMU, fiber-based biosensor material consisting of SrtA, LOO-GFP, MBP and Ubx; s7, s8, s11, GFP strand 7, 8 or 11; SEC, size exclusion chromatography.

### References

1. Medema, G.; Heijnen, L.; Elsinga, G.; Italiaander, R.; Brouwer, A. Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands. *Environ. Sci. Technol. Lett.* 2020, 7 (7), 511–516. <https://doi.org/10.1021/acs.estlett.0c00357>.
2. Minhas, N.; Gurav, Y. K.; Sambhare, S.; Potdar, V.; Choudhary, M. L.; Bhardwaj, S. D.; Abraham, P. Cost-Analysis of Real Time RT-PCR Test Performed for COVID-19 Diagnosis at India's National Reference Laboratory during the Early Stages of Pandemic Mitigation. *PloS One* 2023, 18 (1), e0277867. <https://doi.org/10.1371/journal.pone.0277867>.
3. Huang, Y.; Bystroff, C. Complementation and Reconstitution of Fluorescence from Circularly Permuted and Truncated Green Fluorescent Protein†. *Biochemistry* 2009, 48 (5), 929–940. <https://doi.org/10.1021/bi802027g>.
4. Huang, Y.; Banerjee, S.; Crone, D. E.; Schenkelberg, C. D.; Pitman, D. J.; Buck, P. M.; Bystroff, C. Toward Computationally Designed Self-Reporting Biosensors Using Leave-One-Out Green Fluorescent Protein. *Biochemistry* 2015, 54 (40), 6263–6273. <https://doi.org/10.1021/acs.biochem.5b00786>.
5. Do, K.; Boxer, S. G. Thermodynamics, Kinetics, and Photochemistry of  $\beta$ -Strand Association and Dissociation in a Split-GFP System. *J. Am. Chem. Soc.* 2011, 133 (45), 18078–18081. <https://doi.org/10.1021/ja207985w>.
6. Ding, Y.; Xu, M.-Q.; Ghosh, I.; Chen, X.; Ferrandon, S.; Lesage, G.; Rao, Z. Crystal Structure of a Mini-Intein Reveals a Conserved Catalytic Module Involved in Side Chain Cyclization of Asparagine during Protein Splicing\*. *J. Biol. Chem.* 2003, 278 (40), 39133–39142. <https://doi.org/10.1074/jbc.M306197200>.
7. Pelletier, J. N.; Campbell-Valois, F.-X.; Michnick, S. W. Oligomerization Domain-Directed Reassembly of Active Dihydrofolate Reductase from Rationally Designed Fragments. *Proc. Natl. Acad. Sci.* 1998, 95 (21), 12141–12146. <https://doi.org/10.1073/pnas.95.21.12141>.
8. Cabantous, S.; Terwilliger, T. C.; Waldo, G. S. Protein Tagging and Detection with Engineered Self-Assembling Fragments of Green Fluorescent Protein. *Nat Biotech* 2005, 23 (1), 102–107.
9. Kent, K. P.; Childs, W.; Boxer, S. G. Deconstructing Green Fluorescent Protein. *J. Am. Chem. Soc.* 2008, 130 (30), 9664–9665. <https://doi.org/10.1021/ja803782x>.

10. Ozawa, T.; Kaihara, A.; Sato, M.; Tachihara, K.; Umezawa, Y. Split Luciferase as an Optical Probe for Detecting Protein-Protein Interactions in Mammalian Cells Based on Protein Splicing. *Anal. Chem.* 2001, 73 (11), 2516–2521. <https://doi.org/10.1021/ac0013296>.
11. Ullmann, A.; Jacob, F.; Monod, J. Characterization by in Vitro Complementation of a Peptide Corresponding to an Operator-Proximal Segment of the  $\beta$ -Galactosidase Structural Gene of *Escherichia Coli*. *J. Mol. Biol.* 1967, 24 (2), 339–343. [https://doi.org/10.1016/0022-2836\(67\)90341-5](https://doi.org/10.1016/0022-2836(67)90341-5).
12. Huang, Y.-M.; Nayak, S.; Bystroff, C. Quantitative in Vivo Solubility and Reconstitution of Truncated Circular Permutants of Green Fluorescent Protein. *Protein Sci.* 2011, 20 (11), 1775–1780. <https://doi.org/10.1002/pro.735>.
13. Howell, D. W.; Tsai, S.-P.; Churion, K.; Patterson, J.; Abbey, C.; Atkinson, J. T.; Porterpan, D.; You, Y.-H.; Meissner, K. E.; Bayless, K. J.; Bondos, S. E. Identification of Multiple Dityrosine Bonds in Materials Composed of the *Drosophila* Protein Ultrabithorax. *Adv. Funct. Mater.* 2015, 25 (37), 5988–5998. <https://doi.org/10.1002/adfm.201502852>.
14. Pitman, D. J.; Schenkelberg, C. D.; Huang, Y.-M.; Teets, F. D.; DiTursi, D.; Bystroff, C. Improving Computational Efficiency and Tractability of Protein Design Using a Piecemeal Approach. A Strategy for Parallel and Distributed Protein Design. *Bioinforma. Oxf. Engl.* 2014, 30 (8), 1138–1145. <https://doi.org/10.1093/bioinformatics/btt735>.
15. Goldenzweig, A.; Goldsmith, M.; Hill, S. E.; Gertman, O.; Laurino, P.; Ashani, Y.; Dym, O.; Unger, T.; Albeck, S.; Prilusky, J.; Lieberman, R. L.; Aharoni, A.; Silman, I.; Sussman, J. L.; Tawfik, D. S.; Fleishman, S. J. Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Mol. Cell* 2016, 63 (2), 337–346. <https://doi.org/10.1016/j.molcel.2016.06.012>.
16. Missimer, J. H.; Steinmetz, M. O.; Baron, R.; Winkler, F. K.; Kammerer, R. A.; Daura, X.; van Gunsteren, W. F. Configurational Entropy Elucidates the Role of Salt-Bridge Networks in Protein Thermostability. *Protein Sci. Publ. Protein Soc.* 2007, 16 (7), 1349–1359. <https://doi.org/10.1110/ps.062542907>.
17. Karpusas, M.; Baase, W. A.; Matsumura, M.; Matthews, B. W. Hydrophobic Packing in T4 Lysozyme Probed by Cavity-Filling Mutants. *Proc. Natl. Acad. Sci. U. S. A.* 1989, 86 (21), 8237–8241. <https://doi.org/10.1073/pnas.86.21.8237>.
18. Chen, J.; Stites, W. E. Replacement of Staphylococcal Nuclease Hydrophobic Core Residues with Those from Thermophilic Homologues Indicates Packing Is Improved in Some Thermostable Proteins. *J. Mol. Biol.* 2004, 344 (1), 271–280. <https://doi.org/10.1016/j.jmb.2004.09.008>.
19. Chen, J.; Lu, Z.; Sakon, J.; Stites, W. E. Increasing the Thermostability of Staphylococcal Nuclease: Implications for the Origin of Protein Thermostability. *J. Mol. Biol.* 2000, 303 (2), 125–130. <https://doi.org/10.1006/jmbi.2000.4140>.
20. Gerk, L. P.; Leven, O.; Müller-Hill, B. Strengthening the Dimerisation Interface of Lac Repressor Increases Its Thermostability by 40 Deg. *C. J. Mol. Biol.* 2000, 299 (3), 805–812. <https://doi.org/10.1006/jmbi.2000.3706>.
21. Munson, M.; Balasubramanian, S.; Fleming, K. G.; Nagi, A. D.; O'Brien, R.; Sturtevant, J. M.; Regan, L. What Makes a Protein a Protein? Hydrophobic Core Designs That Specify Stability and Structural Properties. *Protein Sci. Publ. Protein Soc.* 1996, 5 (8), 1584–1593. <https://doi.org/10.1002/pro.5560050813>.
22. Yadahalli, S.; Jayanthi, L. P.; Gosavi, S. A Method for Assessing the Robustness of Protein Structures by Randomizing Packing Interactions. *Front. Mol. Biosci.* 2022, 9, 849272. <https://doi.org/10.3389/fmolb.2022.849272>.
23. Banerjee, S.; Schenkelberg, C. D.; Jordan, T. B.; Reimertz, J. M.; Crone, E. E.; Crone, D. E.; Bystroff, C. Mispacking and the Fitness Landscape of the Green Fluorescent Protein Chromophore Milieu. *Biochemistry* 2017, 56 (5), 736–747. <https://doi.org/10.1021/acs.biochem.6b00800>.
24. Dahiyat, B. I.; Mayo, S. L. Probing the Role of Packing Specificity in Protein Design. *Proc. Natl. Acad. Sci.* 1997, 94 (19), 10172–10177. <https://doi.org/10.1073/pnas.94.19.10172>.
25. Sarkisyan, K. S.; Bolotin, D. A.; Meer, M. V.; Usmanova, D. R.; Mishin, A. S.; Sharonov, G. V.; Ivankov, D. N.; Bozhanova, N. G.; Baranov, M. S.; Soylemez, O.; Bogatyreva, N. S.; Vlasov, P. K.; Egorov, E. S.; Logacheva, M. D.; Kondrashov, A. S.; Chudakov, D. M.; Putintseva, E. V.; Mamedov, I. Z.; Tawfik, D. S.; Lukyanov, K. A.; Kondrashov, F. A. Local Fitness Landscape of the Green Fluorescent Protein. *Nature* 2016, 533 (7603), 397–401. <https://doi.org/10.1038/nature17995>.
26. Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; Davis, I. W.; Cooper, S.; Treuille, A.; Mandell, D. J.; Richter, F.; Ban, Y. E.; Fleishman, S. J.; Corn, J. E.; Kim, D. E.; Lyskov, S.; Berrondo, M.; Mentzer, S.; Popović, Z.; Havranek, J. J.; Karanicolas, J.; Das, R.; Meiler, J.; Kortemme, T.; Gray, J. J.; Kuhlman, B.; Baker, D.; Bradley, P. ROSETTA3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. *Methods Enzym.* 2011, 487, 545–574. <https://doi.org/10.1016/B978-0-12-381270-4.00019-6>.
27. O'Meara, M. J.; Leaver-Fay, A.; Tyka, M. D.; Stein, A.; Houlihan, K.; DiMaio, F.; Bradley, P.; Kortemme, T.; Baker, D.; Snoeyink, J.; Kuhlman, B. Combined Covalent-Electrostatic Model of Hydrogen Bonding Improves Structure Prediction with Rosetta. *J. Chem. Theory Comput.* 2015, 11 (2), 609–622. <https://doi.org/10.1021/ct500864r>.

28. Smith, C. A.; Kortemme, T. Backrub-like Backbone Simulation Recapitulates Natural Protein Conformational Variability and Improves Mutant Side-Chain Prediction. *J. Mol. Biol.* 2008, 380 (4), 742–756. <https://doi.org/10.1016/j.jmb.2008.05.023>.
29. Schenkelberg, C. D.; Byströff, C. Protein Backbone Ensemble Generation Explores the Local Structural Space of Unseen Natural Homologs. *Bioinforma. Oxf. Engl.* 2016, 32 (10), 1454–1461. <https://doi.org/10.1093/bioinformatics/btw001>.
30. Shapovalov, M. V.; Dunbrack, R. L. A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. *Struct. Lond. Engl.* 1993 2011, 19 (6), 844–858. <https://doi.org/10.1016/j.str.2011.03.019>.
31. Mandell, D. J.; Kortemme, T. Backbone Flexibility in Computational Protein Design. *Curr. Opin. Biotechnol.* 2009, 20 (4), 420–428. <https://doi.org/10.1016/j.copbio.2009.07.006>.
32. Huang, P.-S.; Ban, Y.-E. A.; Richter, F.; Andre, I.; Vernon, R.; Schief, W. R.; Baker, D. RosettaRemodel: A Generalized Framework for Flexible Backbone Protein Design. *PLOS ONE* 2011, 6 (8), e24109. <https://doi.org/10.1371/journal.pone.0024109>.
33. Humphris, E. L.; Kortemme, T. Prediction of Protein-Protein Interface Sequence Diversity Using Flexible Backbone Computational Protein Design. *Structure* 2008, 16 (12), 1777–1788. <https://doi.org/10.1016/j.str.2008.09.012>.
34. Zhang, L., & Hermans, J. (1996). Hydrophilicity of cavities in proteins. *Proteins: Structure, Function, and Bioinformatics*, 24(4), 433–438.
35. Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *J. Chem. Theory Comput.* 2011, 7 (12), 4026–4037. <https://doi.org/10.1021/ct200196m>.
36. Hoover, D. M. DNAWorks: An Automated Method for Designing Oligonucleotides for PCR-Based Gene Synthesis. *Nucleic Acids Res.* 2002, 30 (10), 43e–443. <https://doi.org/10.1093/nar/30.10.e43>.
37. Stemmer, W. P.; Cramer, A.; Ha, K. D.; Brennan, T. M.; Heyneker, H. L. Single-Step Assembly of a Gene and Entire Plasmid from Large Numbers of Oligodeoxyribonucleotides. *Gene* 1995, 164 (1), 49–53. [https://doi.org/10.1016/0378-1119\(95\)00511-4](https://doi.org/10.1016/0378-1119(95)00511-4).
38. Sun, Z.; Chen, J.; Yao, H.; Liu, L.; Wang, J.; Zhang, J.; Liu, J.-N. Use of Ssp dnaB Derived Mini-Intein as a Fusion Partner for Production of Recombinant Human Brain Natriuretic Peptide in Escherichia Coli. *Protein Expr. Purif.* 2005, 43 (1), 26–32. <https://doi.org/10.1016/j.pep.2005.05.005>.
39. (39) Booth R, Jons A, Gong X, Banerjee S, Rye H, Byströff C, Bondos S. (2024) Immobilization and enhancement of a heterodimeric biosensor in fibrous protein biomaterials. *Biomacromolecules* (in review).
40. Sniegowski, J. A., Lappe, J. W., Patel, H. N., Huffman, H. A., & Wachter, R. M. (2005). Base catalysis of chromophore formation in Arg96 and Glu222 variants of green fluorescent protein. *Journal of Biological Chemistry*, 280(28), 26248–26255
41. Habuchi, S., Ando, R., Dedecker, P., Verheijen, W., Mizuno, H., Miyawaki, A., & Hofkens, J. (2005). Reversible single-molecule photoswitching in the GFP-like fluorescent protein Dronpa. *Proceedings of the National Academy of Sciences*, 102(27), 9511–9516.
42. Voityuk, A. A., Kummer, A. D., Michel-Beyerle, M. E., & Röscher, N. (2001). Absorption spectra of the GFP chromophore in solution: comparison of theoretical and experimental results. *Chemical Physics*, 269(1–3), 83–91
43. Chattoraj, M., King, B. A., Bublit, G. U., & Boxer, S. G. (1996). Ultra-fast excited state dynamics in green fluorescent protein: multiple states and proton transfer. *Proceedings of the National Academy of Sciences*, 93(16), 8362–8367
44. Stoner-Ma, D., Jaye, A. A., Ronayne, K. L., Nappa, J., Meech, S. R., & Tonge, P. J. (2008). An alternate proton acceptor for excited-state proton transfer in green fluorescent protein: rewiring GFP. *Journal of the American Chemical Society*, 130(4), 1227–1235.
45. Andrews, B. T., Roy, M., & Jennings, P. A. (2009). Chromophore packing leads to hysteresis in GFP. *Journal of molecular biology*, 392(1), 218–227.
46. El Khatib, M., Martins, A., Bourgeois, D., Colletier, J. P., & Adam, V. (2016). Rational design of ultrastable and reversibly photoswitchable fluorescent proteins for super-resolution imaging of the bacterial periplasm. *Scientific reports*, 6(1), 18459.
47. Misak, N. Z. (1993). Langmuir isotherm and its application in ion-exchange reactions. *Reactive Polymers*, 21(1–2), 53–64.
48. Huang, J. R., Craggs, T. D., Christodoulou, J., & Jackson, S. E. (2007). Stable intermediate states and high energy barriers in the unfolding of GFP. *Journal of molecular biology*, 370(2), 356–371.
49. Pedelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and Characterization of a Superfolder Green Fluorescent Protein. *Nat Biotech* 2006, 24 (1), 79–88.
50. Ghosh, T., García, A. E., & Garde, S. (2001). Molecular dynamics simulations of pressure effects on hydrophobic interactions. *Journal of the American Chemical Society*, 123(44), 10997–11003



51. Coventry, B., & Baker, D. (2021). Protein sequence optimization with a pairwise decomposable penalty for buried unsatisfied hydrogen bonds. *PLoS computational biology*, 17(3), e1008061.
52. Desmet, J., Maeyer, M. D., Hazes, B., & Lasters, I. (1992). The dead-end elimination theorem and its use in protein side-chain positioning. *Nature*, 356(6369), 539-542.
53. Schenkelberg, C. D. (2016). Novel Computational Methods for Modeling Backbone Flexibility and Improving Side-Chain Prediction for Protein Design Applications. Rensselaer Polytechnic Institute.
54. Park, J. W., & Rhee, Y. M. (2016). Electric field keeps chromophore planar and produces high yield fluorescence in green fluorescent protein. *Journal of the American Chemical Society*, 138(41), 13619-13629.
55. Greer, A. M., Huang, Z., Oriakhi, A., Lu, Y., Lou, J., Matthews, K. S., & Bondos, S. E. (2009). The *Drosophila* transcription factor ultrabithorax self-assembles into protein-based biomaterials with multiple morphologies. *Biomacromolecules*, 10(4), 829-837.
56. Tsai, S. P., Howell, D. W., Huang, Z., Hsiao, H. C., Lu, Y., Matthews, K. S., ... & Bondos, S. E. (2015). The Effect of Protein Fusions on the Production and Mechanical Properties of Protein-Based Materials. *Advanced Functional Materials*, 25(9), 1442-1450.
57. Michalopoulos, I., Torrance, G. M., Gilbert, D. R., & Westhead, D. R. (2004). TOPS: an enhanced database of protein structural topology. *Nucleic acids research*, 32(suppl\_1), D251-D254.
58. Bairoch, A. (1991). PROSITE: a dictionary of sites and patterns in proteins. *Nucleic acids research*, 19(Suppl), 2241.
59. Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O., & Clubb, R. T. (2001). Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences*, 98(11), 6056-6061.
60. Huang, Y. M., Nayak, S., & Bystroff, C. (2011). Quantitative in vivo solubility and reconstitution of truncated circular permutants of green fluorescent protein. *Protein Science*, 20(11), 1775-1780.]
61. Fox, J. D., & Waugh, D. S. (2003). Maltose-binding protein as a solubility enhancer. *E. coli Gene Expression Protocols*, 99-117.
62. Foos, N., Maurel-Zaffran, C., Mate, M. J., Vincentelli, R., Hainaut, M., Berenger, H., ... & Graba, Y. (2015). A flexible extension of the *Drosophila* ultrabithorax homeodomain defines a novel Hox/PBC interaction mode. *Structure*, 23(2), 270-279.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.