Computational design and testing of leave-one-out GFP biosensors

Christopher Bystroff
Biological Science, Computer Science
Rensselaer Polytechnic Institute
Leave-one-out Green Fluorescent Protein

**computational protein design**

LOO GFP + Missing Peptide → Reconstituted GFP

Designed LOO GFP + Target Peptide → Target Peptide-GFP
Simple, low-tech point of care diagnostics:
Not antibody-based.
May be grown in bacteria.
Requires no added reagents.
Little training needed.
Built-in reporting.
Reusable.
Retains fluorescence when frozen, dried.
GFP tolerates protein engineering

**Split GFP studies**

**trans** truncated GFP  \quad **cis** truncated GFP  \quad **split GFP complex**

- **add peptide**
- **add peptide**


**GFP** tolerates protein engineering

- **GFP**\textsubscript{1-10} + **GFP**\textsubscript{11}
- **Linker**
- **Protein X**
- **GFP**\textsubscript{1-10}**GFP**\textsubscript{11}

Summary slide: GFP tolerates protein engineering

- LOO
- Circular permutation
- Non-circular permutation
- Deletion


How does GFP fold?

Some guiding notions:
• A folding pathway is an ensemble. Many ways exist
• Topology determines the allowed pathways.
• Perturbation of kinetics tells us what parts fold early or late.
Kinetic perturbation as a tool for folding pathways

- Make a mutation, permutation, or truncation
- See how it affects the rate of folding, misfolding, binding, solubility, CD, fluorescence, HD exchange, etc.
Multiphase kinetics

\[ F(t) = \sum_{i=1, 2, 3} (A_i(1-e^{-t_{ki}})) \]

Since peptide binding kinetics = Folding kinetics, slow step in binding is folding. Unbound state is (partially) unfolded.
# 3-phase folding

## Table 1. Characteristics of GFP variants

<table>
<thead>
<tr>
<th>Variants</th>
<th>$k_1$ (1/s)</th>
<th>$k_2$ (1/s)</th>
<th>$k_3$ (1/s)</th>
<th>$A_1$ (%)</th>
<th>$A_2$ (%)</th>
<th>$A_3$ (%)</th>
<th>b Quantum yield (%)</th>
<th>c pKa</th>
</tr>
</thead>
<tbody>
<tr>
<td>t7SP</td>
<td>0.19997</td>
<td>0.01403</td>
<td>0.00137</td>
<td>35.17</td>
<td>43.87</td>
<td>20.96</td>
<td>17.99 (± 1.40)</td>
<td>8.08</td>
</tr>
<tr>
<td></td>
<td>(± 0.05186)</td>
<td>(± 0.00122)</td>
<td>(± 0.00040)</td>
<td>(± 0.95)</td>
<td>(± 0.71)</td>
<td>(± 1.32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t7SPm</td>
<td>0.15993</td>
<td>0.01327</td>
<td>0.00122</td>
<td>35.37</td>
<td>43.70</td>
<td>20.92</td>
<td>39.50 (± 0.27)</td>
<td>8.22</td>
</tr>
<tr>
<td></td>
<td>(± 0.00133)</td>
<td>(± 0.00050)</td>
<td>(± 0.00005)</td>
<td>(± 1.28)</td>
<td>(± 1.00)</td>
<td>(± 0.76)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPT</td>
<td>0.34307</td>
<td>0.02357</td>
<td>0.00289</td>
<td>56.96</td>
<td>34.92</td>
<td>8.12</td>
<td>100 (± 0.17)</td>
<td>5.00</td>
</tr>
<tr>
<td></td>
<td>(± 0.03827)</td>
<td>(± 0.00080)</td>
<td>(± 0.00001)</td>
<td>(± 0.23)</td>
<td>(± 0.46)</td>
<td>(± 0.37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c7SP</td>
<td>0.36067</td>
<td>0.04897</td>
<td>0.01213</td>
<td>44.18</td>
<td>17.87</td>
<td>37.95</td>
<td>78.63 (± 2.37)</td>
<td>7.87</td>
</tr>
<tr>
<td></td>
<td>(± 0.06859)</td>
<td>(± 0.00225)</td>
<td>(± 0.00012)</td>
<td>(± 1.95)</td>
<td>(± 1.90)</td>
<td>(± 0.57)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Refolding of each acid-denatured protein was performed three times independently, and the time dependent fluorescence recovery $F$ was fitted to the triple exponential equation, $F=A_1(1-e^{-t k_1})+A_2(1-e^{-t k_2})+A_3(1-e^{-t k_3})$, with six fitting parameters. Amplitudes of exponential terms ($A_1$, $A_2$, $A_3$) are shown in relative percentage (± S.D.) and rate constants ($k_1$, $k_2$, $k_3$) are provided in 1/s (± S.D.).

b. Fluorescence intensity of 508nm under the excitation of 485nm was measured three times independently and normalized by the O.D.485 to present the quantum yield of each protein. The relative quantum yield compared to OPT is provided in percentage (± S.D.).

c. Fluorescence intensity of 508nm under the excitation of 485nm and under different pH buffers was measured three times independently, and the transition curve was fitted to the Henderson-Hasselbalch equation to obtain the pKa (± S.D.).
3-phase folding

U

Fast

Medium

Slow

F

?
Proline cis-trans isomerization is the slow phase.

Refolding GFP in the presence of CyPA removes the slow phase. CyPA catalyzes X-Pro cis-trans isomerization.

Which Proline?
Fast-folding variant by design: AT-GFP

Suspected slow folding proline replaced by loop design.

AT-GFP folds faster and more efficiently, recovers 50% more fluorescence upon refolding.

GFP refolds into a trapped state.

Andrews et al, JMB 2009
AT-GFP lacks trapped state.

AT-GFP

AT-GFP lacks trapped state.

AT-GFP

Guanidinium

pH jump

OPT-GFP

Gnd, pH jump

Trapped state must be the trans P89 state
New model

- **cis-P89**: $U_{cis}$
  - Fast
  - Medium
  - **F**

- **trans-P89**: $U_{trans}$
  - Slow
  - very slow
  - **M**
  - trapped

- **Fast**
- **Medium**
Strand 7 has fast HD-exchange

LOO7-GFP binds strand 7 and glows brighter.

Strand 7 has fast HD-exchange

peptide binding kinetics = Folding kinetics


Survey of all possible leave-one-outs

In vivo solubility

Do LOO-GFP solubilities tell us about the folding pathway of GFP?

When we leave out a SSE, we block the folding pathway.

The incompletely folded structure is soluble to the extent that its folding is complete.

GFP folding pathway based on LOO-GFP solubilities

The assumption is that the earlier it folds, the less soluble when you leave it out.
Can we computationally design a biosensor based on GFP?

Find two columns (rotamers) within the same residue, where one is always better than the other. Eliminate the rotamer that can always be beat. (repeat until only 1 rotamer per residue)
“Piecemeal” protein design

Dividing up a design job reduces **memory** and **CPU time**.

“Edge effects”, errors caused by piecemealing the search space, are avoided by doing a **second pass**.

---

Power-law runtime order

---

One-pass, Two-pass, Three-pass:

5. for i in central {
   6.   piece[i] = Split(master_space, residue[i], range)
   7.   if (Three-pass) {
   8.       N = nfold * Numsplit(piece[i], rot_cutoff)
   9.       for j in [1..N] {
   10.          split_piece[j] = Split(piece[i], rot_cutoff)
   11.          MEC = Design(template, split_piece[j])
   12.          split_profile[j] = Expand(MEC, template)
   13.     }
   14.   }
   15.   subspace[i] = Union(split_profile[1..N], union_cutoff)
   16.   = piece[i]
   17.   template, subspace[i])
   18.   Expand(MEC, template)
   19.   Profile[1..size(central), union_cutoff]
   20.   if-three-pass) {
   21.       template, subspace)
   22.       Expand(MEC, template)
   23.   }
   24. }

---

Derek Pitman
Interactive ROSETTA

Christian Schenkelberg
GeoFold
all geometrically possible rotations/translations

GeoFOLD uses *pivots* and *hinges* to unfold the native state, then uses *buried surface area* and *entropy* to simulate unfolding kinetics.

Our (non-communicating) multi-template designs looked crazy…
Adding backbone flexibility

“Plastic” protein design

1. Generate multiple templates*
2. Randomly place a rotamer on all templates, calculate combined energy**.
   
   ** Combined energy may be **Boltzmann-weighted** and/or **local**. Rotamer may adjust to each template independently.
3. Accept or reject.
4. Repeat steps 2, 3 until convergence.

Predicting rotamers on known target/template pairs

LOO7-GFP-based H5N1 Influenza biosensor

HA peptide modeled in silico

SSHEVSLGVS

LOO7-GFP

WT strand 7 sequence removed

NSHVNYITADKQ

## Design process for LOO7-HA4

<table>
<thead>
<tr>
<th>13 Designable Positions</th>
<th>Wild-type residue</th>
<th>Computational design</th>
<th>Degenerate codon</th>
<th>Library</th>
<th>Colony HA4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Input</td>
<td>Output</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>83 F</td>
<td>AFILMVW</td>
<td>FW</td>
<td>TKS</td>
<td>FLCW</td>
<td>W</td>
</tr>
<tr>
<td>84 F</td>
<td>AFILMVW</td>
<td>FM</td>
<td>WTK</td>
<td>MILF</td>
<td>F</td>
</tr>
<tr>
<td>161 I</td>
<td>AFILMVW</td>
<td>ILV</td>
<td>VTA</td>
<td>ILV</td>
<td>I</td>
</tr>
<tr>
<td>163 A</td>
<td>AFILMVW</td>
<td>I</td>
<td>ATC</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>164 N</td>
<td>NST</td>
<td>NST</td>
<td>AVY</td>
<td>NST</td>
<td>T</td>
</tr>
<tr>
<td>165 F</td>
<td>AFILMVW</td>
<td>F</td>
<td>TTT</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>167 V</td>
<td>AFILMVW</td>
<td>IMV</td>
<td>RTN</td>
<td>IMV</td>
<td>V</td>
</tr>
<tr>
<td>168 R</td>
<td>RNST</td>
<td>NST</td>
<td>AVY</td>
<td>NST</td>
<td>T</td>
</tr>
<tr>
<td>200 Y</td>
<td>YKHR</td>
<td>HKR</td>
<td>MRS</td>
<td>HKR</td>
<td>H</td>
</tr>
<tr>
<td>201 L</td>
<td>LFMW</td>
<td>FM</td>
<td>WTK</td>
<td>MILF</td>
<td>L</td>
</tr>
<tr>
<td>202 S</td>
<td>SKHR</td>
<td>HKR</td>
<td>MRS</td>
<td>HKR</td>
<td>K</td>
</tr>
<tr>
<td>204 Q</td>
<td>QNST</td>
<td>NST</td>
<td>AVY</td>
<td>NST</td>
<td>T</td>
</tr>
<tr>
<td>224 V</td>
<td>AFILMVW</td>
<td>IV</td>
<td>RTK</td>
<td>IV</td>
<td>V</td>
</tr>
</tbody>
</table>

Overnight colonies of LOO7-HA library

7 mutations
In vitro evolution

Error-prone PCR

DNA shuffling

red: computationally designed
white: random
Chromophore maturation is faster in the presence of target peptide
LOO7-GFP:HA senses HA peptide, but…

...Fluorescence decreases, instead of increasing, upon peptide binding!
Peptide binding dissolves oligomers

(a) LOO7-

Green: + S7 peptide. Purple: no

(b) LOO7-

Green: + HA peptide. Purple: no

Size-exclusion chromatography
Fast binding kinetics: peptide doesn’t close the barrel

Why won't the barrel close???

- We selected for fluorescence, not binding.
- LOO-GFP found another way to glow.
- Energy calculations were off.

F83W mutation is actually overcrowded.
How much do we really know about GFP folding? Can we hope to design GFP-based biosensors without knowledge of its folding? Fluorescence? Chromophore maturation?
Non-circular permutations of GFP secondary structure elements.


rGFP3 folds. rGFP1 does not.

Re-wired GFP suggests late pathway open-barrel intermediates.

Non-circular permutations of GFP secondary structure elements.

Energy landscape view of GFP folding

Drawn based on kinetics data from
Andrews et al, JMB 2009;
Enoki et al, Biochem. 2004;
Huang & Bystroff, Biochem. 2009

Structures of late folding intermediates implied by re-wiring.

Biochemistry 49(51), 10773-10779.
C-link GFP

A more detailed folding pathway of GFP

Piecemeal approach: Influenza Haemagglutinin sensor

- Original target sequence: SSHEVSLGVSSA (12mer)
- Original strand 7: NSHNVYITADKQ (12mer)
- New strand 7: EYNFNSHNVYITAD (14mer)
- New target: KSSWSSSHEVSLGVS (14mer)
- Piecemeal approach:
  - design around SW
  - Use hits from SW design while designing around KS
Results: piece 1, **SW**

- **EYSWSSHEVSLGVS**

Native topology  
Circular permutant  
Leave-one-out

courtesy Shounak Banerjee
Dengue E-protein sensor, in 2 pieces

- Target sequence: NSVTNIELE
- Corresponding to FNSHNVYIT on strand 7
- Piecemeal approach:
  - Design NIELE first
  - Design NSVT
  - Merge hits
  - Design full target.

courtesy Shounak Banerjee
A Better chromophore environment in AT-GFP

Hypothesis: N96Q and N96Q-F94Y mutations should stabilize AT-GFP by restoring a H-bonding network and improve quantum yield.

Quantum yield improved; matches wild type.

Reference: Courtesy: Julia Maynard
Making guests cozy in GFP

**Pseudocode:**

1. Pick best environment-independent rotamer combination on sequence, given backbone
2. Optimize sequence and rotamers around this sequence by either:
   - Allowing target sidechains to move and penalizing high RMSD
   - Restraining target rotamers to 1 standard deviation of parent chi angle
3. Iterate over cycles of sequence and rotamer optimization and energy minimization
4. Stop when scores converge to 1-2 REU

---

### F145 in GFP

\[
\chi_1 = -142.7^\circ
\]

### What all F’s like to be

\[
\chi_1 = 300^\circ
\]

### Phe r1 Rotamer Probabilities

---

### Table: Conformational Probabilities and Average Angles

<table>
<thead>
<tr>
<th>Conformation</th>
<th>Prob (%)</th>
<th>Average Angles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chi1</td>
<td>Chi2</td>
</tr>
<tr>
<td>g'</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td>g'</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>t</td>
<td>0.36</td>
<td>0.36</td>
</tr>
<tr>
<td>t</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>g-</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td>g-</td>
<td>0.44</td>
<td>0.44</td>
</tr>
</tbody>
</table>

---

http://dunbrack.fccc.edu/bbdep2010/figures/fy_r1.gif

**10° Bin: Phi [-160,-150), Psi [130,140]**

---


---

courtesy Shounak Banerjee
Correlation between thermostability and quantum yield

Keith Fraser
Quantum mechanics studies point to hula twist quenching mechanism.

In the protein...

Excited state rotation of hula twist angles has a barrier.

In vacuo...

Excited state rotation of hula twist angles has no barrier.
Unfolding rate vs hula twist

Ground state, seams closed

CRO planar, fixed

Seam opening

CRO hula twisting

More stable, slow unfolding

Less stable, faster unfolding

ΔG

Hula twist coords
The Bystroff lab @ RPI

Pictured: Me, Olu Lawal, Shounak Banerjee, Katie Saporita, Alex McDonald, Emily Conre, Monica Miles, Dr. Donna Crone, Keith Fraser, Danielle Chan, Benjamin Walcott, Danielle Basore.

Not pictured: Christian Schenkelberg, Victoria Jones, Ben Wright, Kangning Zhang, Quadis Evans, Mason Gentner, Emilie Mausser, Kanthi Bommareddy, Julia Reimertz.