Module 2 overview

lecture
1. Introduction to the module
2. Rational protein design
3. Fluorescence and sensors
4. Protein expression

lab
1. Start-up protein eng.
2. Site-directed mutagenesis
3. DNA amplification
4. Prepare expression system

SPRING BREAK
5. Review & gene analysis
6. Purification and protein analysis
7. Binding & affinity measurements
8. High throughput engineering

5. Gene analysis & induction
6. Characterize expression
7. Assay protein behavior
8. Data analysis
Lecture 2: Rational protein design

I. “Blob-level” protein design
   A. Engineered fusion proteins
   B. Knowledge required for blob-level engineering

II. Protein engineering at high resolution
   A. Modifying existing proteins
   B. \textit{De novo} protein engineering
   C. Knowledge needed for high-resolution design
   D. Computational modeling
Rational protein design:

Knowledge-based, deterministic engineering of proteins with novel characteristics

- design/modeling (often computer-aided)
- generate required DNA constructs
- express proteins
- purify proteins
- assess proteins for desired characteristics

“Irrational” high throughput protein engineering:

Selection for desired properties from libraries of random variants

- design/modeling (often computer-aided)
- generate library of DNA constructs
- express proteins
- screen proteins in high throughput assay
- assess “hits” for desired characteristics
“Blob-level” protein design

- Basic idea is to combine protein units of defined function (domains) to engineer a fusion protein with novel functionality
- Examples include sensors, signal transduction components, transcription factors, therapeutics, etc.
GFP-based approaches extend to other sensors:


Can you think of other sensors one could construct based on this design strategy?

An early “synthetic biology” project—signal transduction triggered by a small molecule dimerizing agent:

Image from Spencer, D. M., et al. "Controlling Signal Transduction with Synthetic Ligands." *Science* 262, no. 5136 (1993): 1019-24. DOI: 10.1126/science.7694365. © AAAS. All rights reserved. This content is excluded from our Creative Commons license. For more information, see http://ocw.mit.edu/fairuse.
Engineered antibodies as therapeutic agents:

**single-chain “nanobodies”**

- Heavy-chain antibody: 95 kDa; NA
- Divalent $V_{H}H$: 33 kDa; NA
- Nanobody ($V_{H}H$): 15 kDa; 90 min
- Pentavalent $V_{H}H$ (Pentabody): 128 kDa; NA

**bispecific antibodies**

- $V_{H}A$ and $V_{H}B$ with dimerization and docking domains
- Anchoring domain
- Dock and lock
- Fab A and Fab B
- Leucine zippers


What knowledge is required for “blob-level” protein engineering?:
- rough geometry of protein domains (low resolution structure)
- secondary structure, if insertions or disruptions are planned
- desired linker properties (length, flexibility, hydrophilicity)

Example: CaM-based calcium sensors

Courtesy of National Academy of Sciences, U. S. A. Used with permission.
What we’ve called “blob-level” design is useful for combining functionalities associated with individual protein domains—but what if we want to create new functionalities or make subtle manipulations?

Image removed due to copyright restrictions. Cartoon: A business meeting in which the presenter, using an overhead transparency projector, puts up the slide "Attention to Detail" and the slide is projecting upside down. The presenter says, "Which Brings me to my Next Point."
See http://www.cartoonstock.com/cartoonview.asp?search=site&catref=mban740
Protein engineering at high resolution

• Alter/tune properties of proteins by making structurally or computationally informed changes at the amino acid level
• In some cases, produce entirely new proteins based on predictions of structure and function from amino acid sequence
• Can be “rational” when combined with structural information and/or computational modeling approaches
• Can be “irrational” when combined with high throughput screening and random mutagenesis (to be discussed later in the module)

This is what we are doing in the lab for this module!
1. We looked at the CaM & GFP structures and made predictions about which point mutations would shift the calcium affinity of pericam.
2. We are now going to produce the mutant genes and proteins, and assay purified molecules for desired properties.
3. If we had more time, we might then go on and make a new round of predictions/mutant proteins, to continue the process of tuning the calcium affinity.
Classic example: tyrosyl-tRNA synthetase, engineered to study mechanism of catalysis

\[ E + \text{tyrosine} + \text{ATP} \rightarrow E\cdot\text{Tyr}\cdot\text{AMP} + PP_i \]
\[ E\cdot\text{Tyr}\cdot\text{AMP} + t\text{RNA}^{\text{Tyr}} \rightarrow E + \text{Tyr}\cdot t\text{RNA}^{\text{Tyr}} + \text{AMP} \]

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>( k_3 ), s(^{-1} )</th>
<th>( K_S ) for tyrosine, ( \mu\text{M} )</th>
<th>( K_S ) for ATP, mM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyrosyl-tRNA synthetase(^+)</td>
<td>38</td>
<td>12</td>
<td>4.7</td>
</tr>
<tr>
<td>Tyrosyl-tRNA synthetase(\text{His-45} \rightarrow \text{Gly-45})</td>
<td>0.16</td>
<td>10</td>
<td>1.2</td>
</tr>
<tr>
<td>Tyrosyl-tRNA synthetase(\text{Thr-40} \rightarrow \text{Ala-40})</td>
<td>0.0055</td>
<td>8.0</td>
<td>3.8</td>
</tr>
<tr>
<td>Tyrosyl-tRNA synthetase(\text{Thr-40} \rightarrow \text{Ala-40}; \text{His-45} \rightarrow \text{Gly-45})</td>
<td>0.00012</td>
<td>4.5</td>
<td>1.1</td>
</tr>
</tbody>
</table>


Courtesy of Robin Leatherbarrow. Used with permission.
Rational design can also be used to stabilize proteins—general route to improvement of function/utility.

Fig. 1. Computer graphics simulation of T4 lysozyme (Ile$^3 \rightarrow$ Cys) $\alpha$-carbon chain, showing the amino- and carboxyl-chain termini (N and C, respectively), the three cysteines (●), and the active site (star). Cys$^3$ and Cys$^97$ are connected by a schematic disulfide.


© AAAS. All rights reserved. This content is excluded from our Creative Commons license. For more information, see [http://ocw.mit.edu/fairuse](http://ocw.mit.edu/fairuse)
The "holy grail" of rational engineering is to design entire proteins \textit{de novo} to fold into a defined shape (and ideally carry out a function).

Simplest task is to design peptides with defined 2\degree structure

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>$f_\alpha^b$</th>
<th>$P_\alpha^c$</th>
<th>$f_{\alpha i}^b$</th>
<th>$P_{\alpha i}^c$</th>
<th>$f_\beta^b$</th>
<th>$P_\beta^c$</th>
<th>$f_e^b$</th>
<th>$P_e^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>0.522</td>
<td>1.45</td>
<td>0.272</td>
<td>1.59</td>
<td>0.167</td>
<td>0.97</td>
<td>0.311</td>
<td>0.66</td>
</tr>
<tr>
<td>Arg</td>
<td>0.282</td>
<td>0.79</td>
<td>0.115</td>
<td>0.67</td>
<td>0.154</td>
<td>0.90</td>
<td>0.564</td>
<td>1.20</td>
</tr>
<tr>
<td>Asn</td>
<td>0.263</td>
<td>0.73</td>
<td>0.090</td>
<td>0.53</td>
<td>0.113</td>
<td>0.65</td>
<td>0.624</td>
<td>1.33</td>
</tr>
<tr>
<td>Asp</td>
<td>0.351</td>
<td>0.98</td>
<td>0.090</td>
<td>0.53</td>
<td>0.137</td>
<td>0.80</td>
<td>0.514</td>
<td>1.09</td>
</tr>
<tr>
<td>Cys</td>
<td>0.278</td>
<td>0.77</td>
<td>0.056</td>
<td>0.33</td>
<td>0.222</td>
<td>1.30</td>
<td>0.500</td>
<td>1.07</td>
</tr>
</tbody>
</table>


© American Chemical Society. All rights reserved. This content is excluded from our Creative Commons license. For more information, see \url{http://ocw.mit.edu/fairuse}.

Related task is to predict 2\degree structure from sequence

```
MADQLTEEQIAEFKEAFLFDDGGTITKELGTVMRSGLGQNTEAEQLDMINEVDADG
| helix | <----------------------> | <----------> | <-----------> |
| sheet | EEEEEEEE               | EEEEEEEE   | EEEEEEE   |
| turns | T  TT                  | T  TT      | T  TT     |

NGTIDFPEFTMMARKMDITSEEEEIREAFRVFDKLGNGYISAAELRVMTNLGEKLTE
| helix | <----------------------> | <----------> |
| sheet | EEEEEEEE               | EEEEEEEE   |
| turns | TT  T                  | T  T       | T  T      |

EVDEMIREAIDGQVNYEEFVQMTAK
| helix | <----------------------> |
| sheet | EEEEEEEE               |
| turns | T  T                   |
```
De novo design can be extended to 3° and 4° structure. Example is design of a functional enzyme from so-called coil-coil peptides:


Copyright © 2004 National Academy of Sciences, U.S.A.
7th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction

Courtesy of Protein Structure Prediction Center. Used with permission.
What knowledge is required for “high-resolution” protein engineering?:
• determination of 3D structure, for mutagenesis-based engineering
• knowledge of protein folding rules for de novo engineering
• computational modeling techniques usually required

Computational methods important for protein engineering:
• modeling & visualization
• energy/thermodynamic calculations
• searching conformation and sequence spaces
• comparison with known protein structures/sequences

The basis of more automated analysis of structural perturbations than our own “inspect and try” approach involves use of an energy function to evaluate plausibility of candidate structures:

\[ E_{tot} = E_{bond} + E_{angl} + E_{dihe} + E_{impr} + E_{VDW} + E_{elec} + E_{Hbond} + \ldots \]

This may be evaluated using a force field (e.g. CHARMM19) and atomic coordinates available from simulation or modified PDB file.
Computational techniques for investigation of specific structures:

- molecular dynamics: simulate physically plausible movements of a protein, with a “rule” that describes probability of motions in conjunction with the energy function at a given temperature
- energy minimization: gradually perturb a model protein structure to find a locally favorable structure (energy minimum) in the neighborhood of a starting structure
- both techniques can be applied after in silico mutagenesis, e.g. to anticipate the effect of mutation on stability or ligand binding

Simulation of anthrax toxin dissociating from its receptor

Video available at [http://www.ks.uiuc.edu/Gallery/Movies/](http://www.ks.uiuc.edu/Gallery/Movies/), see "Stretching the anthrax toxin-receptor complex"