# 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 8: High throughput engineering

I. Paper discussion

II. General requirements for HT engineering
   A. High throughput vs. rational design
   B. Generating libraries

III. Selection techniques
   A. Phage display and related techniques
   B. Selection for properties other than affinity
Module 2 assignment

Protein engineering research article
  1. Abstract
  2. Introduction
  3. Materials and Methods
  4. Results
  5. Discussion
     Begin with brief summary of rationale and results
     Pick out a few high level topics to discuss (3-4 pp. DS)
     Relate to the major themes/questions from your introduction
     How reliable were your results, and what are caveats?
        (but do not dwell on minor technical problems)
     Did results confirm hypothesis and why/why not?
        (be specific and relate to protein structure)
     How do results relate to literature or other class findings?
     What follow-up experiments could be performed?
     What is the overall significance/impact of your results?
  6. References
  7. Figures
class mutants
2013

TDQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQD
E Q
3x 2x

MINEVDADGNSTIYFPFLETFMKMARKDSDKDEEIREREAHRVDKDNGYRT
R E R
2x S 2x

SAAQLRHMWMTNLGEKLTDDEEVDDEMIREADIDDGQVNYEEFVQMMTP
P R

to helix F

from helix E
1
TDQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVRSLGQNPEAELOD
  P  WP  P  D  ID
  R  HW  R  R
  R
51
MINEVDADGNGTIYFPEFLTMARKMKDTDSSEEEIREAFRVFDKDGNGYI
  H  HPE  D  K  P  G  PHHD  D
  I  K  E  K  A  K  P
  H
  K
  D
101
SAAQLRHVMTNLGEKLTDEEVDMEIREADIDGDGQVINYEVEFVQMTAK
  L  RS  S  K  P  G  R  F  D
  T
  L

class mutants
2008-10
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
Methods for generating mutant genetic libraries:
- site-directed mutagenesis with degenerate primers
- error-prone PCR
- gene shuffling

Degenerate primers

\[
\begin{align*}
gat & \quad aag & \quad gac & \quad ggc & \quad gat & \quad gcc & \quad acg & \quad att & \quad acc & \quad acc \\
D & \quad K & \quad D & \quad G & \quad D & \quad G & \quad T & \quad I & \quad T & \quad T \\
gac(c/g) & \quad xcc & \quad xxx \\
D/E & \quad S/P/T/A & \quad X \\
\end{align*}
\]

- not all combinations of AA’s possible at each position
- number of combinations expands exponentially
- degenerate primers synthesized by split-pool method
- standard primer design criteria must be considered
PCR polymerase and conditions may be chosen to promote mutations.

<table>
<thead>
<tr>
<th>Polymerase</th>
<th>Template doublings ($d^b$)</th>
<th>$lacI^-$ plaques ($% \pm SD$)</th>
<th>Mutation load$^c$ (per kilobase) ($\pm SD$)</th>
<th>Error rate$^d$ (per base) ($\times 10^{-6} \pm SD$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pfu-Pol (exo$^+$)</td>
<td>12.3</td>
<td>0.61 $\pm$ 0.09</td>
<td>0.017 $\pm$ 0.002</td>
<td>1.4 $\pm$ 0.2</td>
</tr>
<tr>
<td>Pfu-Pol (exo$^-$)</td>
<td>11.8</td>
<td>20 $\pm$ 1.7</td>
<td>0.58 $\pm$ 0.05</td>
<td>49 $\pm$ 4</td>
</tr>
<tr>
<td>Taq-Pol</td>
<td>11.6</td>
<td>3.9 $\pm$ 0.16</td>
<td>0.12 $\pm$ 0.006</td>
<td>10 $\pm$ 0.5</td>
</tr>
</tbody>
</table>

error rate = mutation load ÷ template doublings

normal PCR

- Taq
- dCTP, dTTP
- dGTP, dATP
- Mg$^{2+}$

error-prone PCR

- Taq
- dCTP, dTTP $\uparrow$
- dGTP, dATP $\downarrow$
- Mg$^{2+}$ $\uparrow$
- Mn$^{2+}$

some mutations are more likely than others

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Pfu-Pol(exo$^-$) D473G$^a$</th>
<th>Taq-Pol (Mn$^{2+}$/unbalanced dNTPs)$^b$</th>
<th>Taq-Pol (Mn$^{2+}$/unbalanced dNTPs)$^c$</th>
<th>Taq-Pol (unnatural mutagenic bases)$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A→T/T→A</td>
<td>28</td>
<td>40.9</td>
<td>11.4</td>
<td>0.2</td>
</tr>
<tr>
<td>A→C/T→G</td>
<td>7.4</td>
<td>7.3</td>
<td>3.3</td>
<td>8.4</td>
</tr>
<tr>
<td>A→G/T→C</td>
<td>19.2</td>
<td>27.6</td>
<td>60.9</td>
<td>78.3</td>
</tr>
<tr>
<td>G→A/C→T</td>
<td>22</td>
<td>13.6</td>
<td>18.1</td>
<td>13.2</td>
</tr>
<tr>
<td>G→C/C→G</td>
<td>7.3</td>
<td>1.4</td>
<td>4.3</td>
<td>0.7</td>
</tr>
<tr>
<td>G→T/C→A</td>
<td>10.3</td>
<td>4.5</td>
<td>1.8</td>
<td>0.0</td>
</tr>
<tr>
<td>Insertion</td>
<td>2.9</td>
<td>0.3</td>
<td>Not given</td>
<td>~0</td>
</tr>
<tr>
<td>Deletion</td>
<td>2.9</td>
<td>4.2</td>
<td>Not given</td>
<td>~0</td>
</tr>
</tbody>
</table>
Gene shuffling techniques mimic diversity due to meiotic recombination:
- fragments of homologous genes combined using “sexual PCR”
- diversity may arise from error prone PCR or multiple genes

How are libraries of mutant proteins screened?

All major methods include a strategy to keep DNA sequence info associated with the proteins that are being screened.

Phage display is a versatile high throughput method to do this:

protein “displayed” on the coat of a bacteriophage, by fusing to a natural phage coat protein

M13 phage (contains DNA)
Application: phage-displayed peptides that bind to GaAs

selected sequences

phages patterned on target substrate

Yeast display: similar to phage display, but with proteins fused to a *Saccharomyces* cell wall protein (DNA in yeast)

What would you expect advantages to be, compared with phage display?

In this example, a population of scFvs was screened for binding to an antigen

left: selection criterion for FACS assay

right: comparison of wt (blue) and selected (red) scFv binding

Why not more stringent selection?  

What about properties other than affinity?

A simple example: screen for DsRed variants with different excitation and emission wavelengths—how could this be done?

another example: neurotransmitter sensor for MRI

This screen only involved ~500 variants/round; under what circumstances would you expect this level of throughput to be successful?

Which type of screening method to use?

<table>
<thead>
<tr>
<th>screen method</th>
<th>throughput</th>
<th>other notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELEX</td>
<td>$10^{15}$</td>
<td>selection of DNA/RNA</td>
</tr>
<tr>
<td>ribosome display</td>
<td>$10^{15}$</td>
<td><em>in vitro</em> protein synthesis</td>
</tr>
<tr>
<td>phage display</td>
<td>$10^{11}$</td>
<td>best for small proteins/peptides</td>
</tr>
<tr>
<td>yeast display</td>
<td>$10^{8}$</td>
<td>compatible w/eukar. proteins</td>
</tr>
<tr>
<td>plate assays</td>
<td>$&lt; 10^5$</td>
<td>versatile but more complex</td>
</tr>
</tbody>
</table>

**Number of variants in a protein library**

\[
\text{x residues} = 20^x \text{ possible variants} \\
12 \text{ residues} = 4 \times 10^{15} \text{ variants}
\]

Lesson: impossible to cover *sequence space* except with short sequences (or few positions) and only the most high throughput techniques.
Good luck with your papers!

http://www.youtube.com/watch?v=qbtgmuyo1B8