## Module 1 overview

**lecture**

1. Introduction to the module
2. Rational protein design
3. Fluorescence and sensors

**lab**

1. Start-up protein eng.
2. Site-directed mutagenesis
3. DNA amplification

### PRESIDENT’S DAY

4. Review & gene analysis
5. Protein expression
6. Purification and protein analysis
7. Binding & affinity measurements
8. High throughput engineering

4. Prepare expression system
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. De novo protein engineering
   C. Knowledge needed for high-resolution design
   D. Computational modeling
Rational protein design:
Knowldege-based, deterministic engineering of proteins with novel characteristics

“Irrational” high throughput protein engineering:
Selection for desired properties from libraries of random variants

design/modeling
(often computer-aided)

generate required DNA constructs

express proteins

purify proteins

assess proteins 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:

Engineered antibodies as therapeutic agents:

**single-chain “nanobodies”**

![Diagram of heavy-chain antibody and pentavalent V₃H](image1)

**Heavy-chain antibody**
- CDRs
- Hinge
- C₃₂
- C₃₃
- 95 kDa; NA

**Nanobody (V₃H)**
- 15 kDa; 90 min

**Divalent V₃H**
- 33 kDa; NA

**Pentavalent V₃H (Pentabody)**
- 128 kDa; NA

**bispecific antibodies**

![Diagram of bispecific antibody and Leucine zippers](image2)

**Dimerization and docking domain**
- V₃A
- V₃B
- Anchoring domain

**Dock and lock**
- Fab A
- Fab B
- C-jun
- C-fos
- 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

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?
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

\[
E + \text{tyrosine} + \text{ATP} \rightarrow E\cdot\text{Tyr-AMP} + \text{PP}_i
\]

\[
E\cdot\text{Tyr-AMP} + \text{tRNA}^{\text{Tyr}} \rightarrow E + \text{Tyr-tRNA}^{\text{Tyr}} + \text{AMP}
\]

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) α-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.

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° 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>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Related task is to predict 2° structure from sequence

De novo design can be extended to 3° and 4° structure. Example is design of a functional enzyme from so-called coil-coil peptides:

7th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction

<table>
<thead>
<tr>
<th>Results Home</th>
<th>Groups</th>
<th>Table Browser</th>
<th>Refinement Results</th>
<th>Quality Assessment Results</th>
</tr>
</thead>
</table>

3D structure evaluation
First Model: ALL Models | Category filter: ALL TBM FM | Sort targets by: name, target size | Help

<table>
<thead>
<tr>
<th>T0283 (TBM) TABLES (1-97)</th>
<th>T0284 (TBM) TABLES (4-253)</th>
<th>T0285 (TBM) TABLES (7-105)</th>
<th>T0286 (TBM) TABLES (3-204)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Structure Image" /></td>
<td><img src="image2" alt="Structure Image" /></td>
<td><img src="image3" alt="Structure Image" /></td>
<td><img src="image4" alt="Structure Image" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T0287 (FM) TABLES (21-181)</th>
<th>T0288 (TBM) TABLES (1-26, 29-88)</th>
<th>T0289 (TBM) TABLES (4-310)</th>
<th>T0289_D1 (TBM) TABLES (4-223, 298-310)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image5" alt="Structure Image" /></td>
<td><img src="image6" alt="Structure Image" /></td>
<td><img src="image7" alt="Structure Image" /></td>
<td><img src="image8" alt="Structure Image" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T0289_D2 (TBM) TABLES (224-297)</th>
<th>T0290 (TBM) TABLES (1-173)</th>
<th>T0291 (TBM) TABLES (1-5, 15-175, 193-295, 305-310)</th>
<th>T0292 (TBM) TABLES (1-43, 49-74, 79-129, 137-164, 176-277)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image9" alt="Structure Image" /></td>
<td><img src="image10" alt="Structure Image" /></td>
<td><img src="image11" alt="Structure Image" /></td>
<td><img src="image12" alt="Structure Image" /></td>
</tr>
</tbody>
</table>

Stopped
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:

$$E_{tot} = E_{bond} + E_{angl} + E_{dihe} + E_{impr} + E_{VDW} + E_{elec} + E_{Hbond} + ...$$

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
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. De novo protein engineering
   C. Knowledge needed for high-resolution design
   D. Computational modeling