Computational design of protein structures, functions, and assemblies
Protein Design Work Flow

• Computer calculation of optimal sequence for desired structure or function
• Read off amino acid sequence of designed protein
• Back translate to DNA sequence, and make gene
• Make protein and assay
Protein Design: find lowest energy sequence for desired structure and/or function
Designed cyclic peptides with stable backbone conformations

Design 10helix1

Design 8helix1

Design 5loop1

10helix1 NMR ensemble (undergoing refinement)

8helix1 NMR ensemble (undergoing refinement)

5loop1 NMR ensemble

Vikram Mulligan
Foldit Symmetric Designs

Brian Koepnick
De Novo active site design

I. Model reaction transition states and intermediates

II. Design disembodied ideal active site around transition states and intermediates

III. Design protein containing ideal active site
Kemp elimination reaction

Design Process:

choose catalytic motif

1) match in scaffold

2) design pocket

3) rank designs
De novo design process

1) choose catalytic motif
de novo design process

1) choose catalytic motif

2) match in scaffold

3) design pocket

4) rank designs
**de novo design process**

1) choose catalytic motif

2) match in scaffold

3) design pocket

4) rank designs
Examples of design models

Daniela Roethlisberger, Andrew Wollacott
Catalytic residue dependent activity!

- KE59
- KE59 E230A
- KE59 E230Q
- KE70
- KE70 D45N
- KE70 H17A

Graph showing product formation over time (sec) for different conditions.
De novo enzyme design--
Successes thus far

• General acid-base catalysis: Kemp elimination (Nature 2008)
• Covalent catalysis: novel aldol catalysts (Science 2008)
• Bimolecular reactions: Diels Alder (Science 2010), Baylis Hillman
• Polar transition state stabilization: ester hydrolysis
De novo enzyme design

• Can design active enzymes from scratch!
• Starting activities low, but can be increased readily by directed evolution (evolved Kemp $k_{cat}/K_m \sim 5 \times 10^5$)
• Need more precise positioning of catalytic groups, elimination of competing reactions, dynamics (?), etc.
• Enzymes are masters of art of compromise--have to do everything well!

<table>
<thead>
<tr>
<th></th>
<th>Activity</th>
<th>Development Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designed Enzymes</td>
<td>+++</td>
<td>&lt; 5 years</td>
</tr>
<tr>
<td>+directed evolution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catalytic Antibodies</td>
<td>+++</td>
<td>~25 years</td>
</tr>
<tr>
<td>Natural Enzymes</td>
<td>++++++++</td>
<td>~$10^8$ years</td>
</tr>
</tbody>
</table>
Design of Binding
“Hot-spot” centered approach to de novo protein-protein interface design

• Generate a disembodied hot-spot residue map of the target surface patch
• Dock large set of scaffolds against the surface patch, favoring configurations that support multiple hotspots.
• Build hot spots onto docked scaffolds:
  – Superimpose scaffold on hot-spot interaction requiring highest precision
  – Build on additional hot-spots by minimizing the scaffold rigid body, sidechain, and backbone degrees of freedom
• Optimize interface for binding affinity
• Filter designs on computed binding energy and shape complementarity
Design of binders to conserved epitope on Spanish Flu hemagglutinin

Structure of neutralizing Ab binding to stalk region of influenza HA
Ekiert, Wilson et al. Science 324:246
First, dock disembodied residues against HA surface
Second, find/build scaffold supporting the interacting residues.
Crystal structures of designed binders bound to HA closely match design models.
The method is far from perfect-only 2/80 designs bind the virus, and even these bind weakly.

How do we improve the energy function used in the design calculations?
How do we make tighter binders?

Library selection plus next gen sequencing
(Doug Fowler, Stan Fields)
Use next-gen sequencing to comprehensively map optimality of designed sequence (HB80)

1. Create library of all point mutants
2. Select for binding
3. Deep sequence (Illumina PE-76)
4. Compute ratio for each mutant of population in selected and unselected pool
5. Hotspot residues are largely invariant; opportunities for improving designs are revealed

Positions colored by Shannon entropy

Aaron Chevalier
Tim Whitehead
Binders to multiple sites on the Influenza HA enable unique readout of HA identity

<table>
<thead>
<tr>
<th>Group I</th>
<th>HB80.4</th>
<th>HB80.3_H5</th>
<th>HB36.5</th>
<th>HB36.5_H2</th>
<th>S24*</th>
<th>S52*</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>H2</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>H5</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>H6</td>
<td>yes</td>
<td>n.d.</td>
<td>yes</td>
<td>yes</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>H9</td>
<td>yes</td>
<td>n.d.</td>
<td>yes</td>
<td>yes</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>H12</td>
<td>yes</td>
<td>n.d.</td>
<td>no</td>
<td>no</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>H13</td>
<td>yes</td>
<td>n.d.</td>
<td>yes</td>
<td>yes</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group II</th>
<th>HB80.4 binding profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3</td>
<td>no</td>
</tr>
</tbody>
</table>

Eva Strauch  
David La  
Aaron Chevalier
Engineered inhibitor of a viral Bcl-2 protein (BHRF1) associated with lymphoma

### Dissociation constants (nM) for prosurvival Bcl-2 family members

<table>
<thead>
<tr>
<th>Protein</th>
<th>BHRF1</th>
<th>Bcl-2</th>
<th>Bcl-W</th>
<th>Mcl-1</th>
<th>Bfl-1</th>
<th>Bcl-XL</th>
<th>Bcl-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bim-BH3</td>
<td>12 ± 4</td>
<td>2.02 ± 0.08</td>
<td>2.1 ± 0.1</td>
<td>0.6 ± 0.2</td>
<td>2.1 ± 0.3</td>
<td>3 ± 1</td>
<td>12.2 ± 0.1</td>
</tr>
<tr>
<td>BINDI</td>
<td>0.22 ± 0.05</td>
<td>2,100 ± 100</td>
<td>870 ± 40</td>
<td>40 ± 10</td>
<td>2,600 ± 800</td>
<td>810 ± 80</td>
<td>&gt; 10,000</td>
</tr>
</tbody>
</table>
Engineered inhibitor of a viral Bcl-2 protein (BHRF1) associated with lymphoma

Mitochondrial CytC Release

Intracellular Delivery and Cell Death
De novo design of small molecule binding proteins

Christy Tinberg
Jiayi Dou
Sagar Khare

Expression
Binding

+ 730 µM DIG
DIG10.2 Crystal Structure Confirms Model

$S_c = 0.67$

Backbone RMS = 0.460 Å  
All-atom RMS = 0.53 Å  
Ligand RMS = 1.00 Å
Rosetta MatDes: A general method for designing protein-based materials

Wild-type protein: *S. enterica* PduT
- 3 subunits, 60 kDa
- $C_3$ symmetry

Designed self-assembling protein
- 24 subunits (8 trimers), 480 kDa
- $O$ (octahedral) symmetry
Crystal structures closely match design models

O333
24 subunits (8 trimers), 480 kDa
O (octahedral) symmetry

T308
12 subunits (4 trimers), 276 kDa
T (tetrahedral) symmetry

Neil King, Will Sheffler Science 2012
Design of multi-component materials

Neil King Jacob Bale, Will Sheffler
Characterization of designed two component assemblies

T33-09
4 ▲ + 4 ▲
14 mutations

T33-21
4 ▲ + 4 ▲
22 mutations

T33-15
4 ▲ + 4 ▲
22 mutations

T33-28
4 ▲ + 4 ▲
18 mutations

T32-28
4 ▲ + 6 ◊
21 mutations

Normalized abs

Elution volume (mL)
9 12 15 18

Gonen/Gonen (Janelia)
Crystal structures very close to design models

<table>
<thead>
<tr>
<th>Structure</th>
<th>Resolution (Å)</th>
<th>R / R_free</th>
<th>RMSD (backbone, 24 chains)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T33-15</td>
<td>2.7</td>
<td>0.205/0.250</td>
<td>1.4</td>
</tr>
<tr>
<td>T33-21</td>
<td>2.6</td>
<td>0.232/0.242</td>
<td>1.5</td>
</tr>
<tr>
<td>T33-28</td>
<td>4.5</td>
<td>0.341/0.344</td>
<td>0.7</td>
</tr>
<tr>
<td>T32-28</td>
<td>4.0</td>
<td>0.274/0.301</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Neil King, Jacob Bale, Will Sheffler

McNamara/Yeates (UCLA)
Route to improved vaccines and targeted delivery?

**Nanocages** for targeted delivery, vaccine design, synthetic biology

**Nanolayers** for bioactive materials and diagnostics

**Nanowires** for molecular or electronic transport
Final Resolution = 14Å
Unit Cell Dimensions, $a=b=120\,\text{Å}$, $\gamma=120^\circ$
p6 symmetry
Towards Next Generation Vaccines

- Computationally designed stabilized epitopes that elicit broadly neutralizing antibodies
- Engineered into self assembling two-component virus-like nanoparticles.

SOSIP HIV epitope Trimer

2-Component self-assembling nanoparticle.
Can enlist the general public to solve design problems! (FoldIt)

- Protein structure determination
- Algorithm discovery
- Radical enzyme backbone redesign
**de novo** designed Diels-Alderase

**DA_20_10 Active Site View**, Catalytic Residues

**Diels-Alder Reaction Progress Curve**
(1x PBS, 298K, 0.1mM Diene, 3mM Dienophile, 20uM Protein)

```
y = 158.71x - 29.437
```

```
y = 3.8669x - 0.5669
```

**nM Product** vs **Hours**

- **DA_20_10 (Active)**
- **Blank**
- **DA_20_01 (Inactive Variant)**

Justin Siegel, Alex Zanghellini, Science 2010
Can we improve activity of designed Diels Alderase by remodeling active site loops? Let's ask the Fold.it community!
Helical hairpin insertion leads to 18-fold greater catalytic activity greater than DA_20_10

Chris Eiben
Justin Siegal
Foldit Players!
Crystal structure shows both helices in 24-amino acid designed loop are placed correctly.

DESIGN (GREEN/PURPLE) vs. CRYSTAL STRUCTURE (BROWN/GOLD)

Jacob Bale
Barry Stoddard