PROTEIN ENGINEERING

09/12/2010 - WS 10/11
Computational Systems Biology

Marc Offman
Systems Biology & Protein Engineering

• Where is the connection?
• Systems Biology
  • Systems Biology is Bioinformatics, but just more hip!
  • It’s all about money (funding)…
  • Since 2000 we want to look at the bigger pictures instead of the detail
• But as important the big picture, as important the detail remains
• Protein Engineering
  • In the last ten years a lot of development
  • Manipulate or design proteins to do what you want
  • Important for industry, pharma etc…
  • Engineering antibodies and overcoming aggregation
Systems Biology & Protein Engineering

Identify targets

Systems Biology

Protein Engineering

Use engineered proteins

Doyon et al. 2008, Nat Biotech
Overview

• Introduction
• Protein engineering: What has been done…
• Genetic Algorithms
• Molecular Dynamics

-------------------------------------------------------------------------

• Example: Protein Engineering of L-Asparaginase
• Summary
Protein Engineering – manipulate proteins

A

Structural

change conformation

L

find sequence
Protein Engineering – Examples

- Proteins are adapted (amino acids are mutated) to:
  - Increase stability
    - Loop stability: L-Asn (offman et al., 2010)
    - Interface stability: L-Asn
    - Overall stability: Cytosine Deaminase, half-life/melting temperature (Korkegian et al., 2005)
  - Change function: change endonuclease I-MsoI to identify new cleavage site (Ashworth et al., 2006)
  - Increase activity: L-Asn
  - Increase or change recognition of binding partner: Zink finger design (Doyon et al., 2008)
Protein Engineering – Examples

- Novel folds (protein design)
  - Top7 (Kuhlman et al., 2003)
Protein Engineering – Examples

- Novel functions
  - Design of retro-aldolase (Jiang et al., 2008)
Protein Engineering - Methods

• Sampling
  • Monte Carlo
  • Biased Monte Carlo / Replica exchange
  • Monte Carlo simulated annealing
  • Genetic Algorithms
  • Dead-end elimination
  • Molecular Dynamics
  • Directed evolution (experimental)
  • and of course combinations

http://www.ks.uiuc.edu/Research/vmd/
Protein Engineering - Methods

- Scoring
  - Pair potentials
  - Molecular mechanics forcefields (empirical)

\[ E(r^N) = w_1 \cdot bond(r^N) + w_2 \cdot angle(r^N) + w_3 \cdot tors(r^N) + w_4 \cdot non\_bond(r^N) \]

- Vacuum
- Implicit solvent
- Explicit solvent
- Quantum mechanics force fields

- Other scoring functions
Genetic Algorithms

- Mimics evolution
- Powerful search and optimization technique
- Previously used for modelling techniques
- Protein modelling/engineering involves huge space to be searched
Flow diagram

Offman et al., 2006
Recombination

- Single crossover
- Double crossover
Flow diagram

- Input population
- Grow population
  - H, C, P
- Cutdown population
- Converged? Yes/No
- Final population
Mutations

- Coil mutation
- Helix mutation
Flow diagram
Protein Mutation
Flow diagram

1. Input population
2. Grow population
3. Cutdown population
4. Converged?
   - No
     - [X] [X] [X] [X] [X]
   - Yes
     - Final population
Flow diagram
Molecular Dynamics

- Atoms move
- Approximation of known physics
- Based on statistical mechanics
- Statistical ensemble averages are equal to time averages
- Simulates interactions of protein and solvent over short time (10ns)
MD - Principles

time 0.0041 ps

- Give atoms initial positions $r^{(t=0)}$, choose short $\Delta t$
- Get forces $F = -\nabla V(r^{(t)})$ and $a = F/m$
- Move atoms: $r^{(t+1)} = r^{(t)} + v^{(t)} \Delta t + \frac{1}{2} a \Delta t^2 + \ldots$
- Move time forward: $t = t + \Delta t$
- Repeat as long as you need
MD Simulation

Offman et al., 2010
Protein Engineering of L-Asparaginase

- Overcoming drug resistance
- Engineering AS lid-loop
- Engineering tetramer interface
Acute lymphatic leukaemia

• Important component of first line treatment for acute lymphatic leukaemia
• Most common in children (2-4 years)
• Hydrolyses ASN to ASP and starves tumour
• Side effects:
  • Rapid degradation (drug resistance) & antigenicity
  • Liver, neural toxicity and others
Identify primary cleavage site
L-Asparaginase AIMS

• Identify primary cleavage site
• Increase activity and overcome drug resistance
• Reduce side effects
• Learn about secondary glutaminase activity

• Patel et al., 2009 JCI
• Offman et al., 2010 Blood
L-Asn cleavage site

Ecoli
ATK- SNYTVGK

Erwinia
GTQTTTGYKAGA

Ecoli
TSMSADGPFLNL

Erwinia
TAISADGPMNL

Ecoli
DKASANRGVLYV

Erwinia
DKQSRGGRGVMV

N24

N143
L-Asn flexible loop

closed loop

open loop

D143

N24
## Experimental data on cleavage

<table>
<thead>
<tr>
<th>Protein</th>
<th>IU/mg</th>
<th>Ratio to WT</th>
<th>AEP cleavage</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>1012</td>
<td>1.00</td>
<td>Yes</td>
</tr>
<tr>
<td>N24G</td>
<td>453</td>
<td>0.45</td>
<td>NO</td>
</tr>
<tr>
<td>D124G</td>
<td>28</td>
<td>0.03</td>
<td>Yes</td>
</tr>
<tr>
<td>N143G</td>
<td>1106</td>
<td>1.09</td>
<td>Yes</td>
</tr>
<tr>
<td>N24G D124G</td>
<td>36</td>
<td>0.04</td>
<td>Yes</td>
</tr>
<tr>
<td>D124G N143G</td>
<td>53</td>
<td>0.05</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Overcoming drug resistance and controlling activities
Role of N24
Loop flexibility

![Graph showing loop flexibility with RMSF in Å as a function of AA # for WT and N24G.](image-url)
Orientation of THR12

WT

N24G

hydroxyl

methyl

time
Flow diagram
Lid-loop sampling

<table>
<thead>
<tr>
<th>MUTANT</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>N24T</td>
<td>473</td>
</tr>
<tr>
<td>N24H</td>
<td>88</td>
</tr>
<tr>
<td>N24A</td>
<td>53</td>
</tr>
<tr>
<td>N24S</td>
<td>22</td>
</tr>
<tr>
<td>D281E</td>
<td>22</td>
</tr>
<tr>
<td>N24M</td>
<td>14</td>
</tr>
<tr>
<td>N24L</td>
<td>11</td>
</tr>
<tr>
<td>N24F</td>
<td>9</td>
</tr>
<tr>
<td>N24W</td>
<td>8</td>
</tr>
<tr>
<td>N24G</td>
<td>8</td>
</tr>
<tr>
<td>N24V</td>
<td>3</td>
</tr>
</tbody>
</table>

2,000 samples
Flexibility

A

![Graph showing flexibility analysis. The graph plots RMSF (Å) against residue number for different variants: WT, N24G, N24A, N24T, N24H, and N24S[D281E]. The x-axis represents residue number, and the y-axis represents RMSF (Å).]
Internal flexibility
T12 orientation
Important HB during MD
Hydrogen Bonds

C

Shortest distance

Don-Acc distance

Average Distance [Å]

<table>
<thead>
<tr>
<th>Average Distance</th>
<th>N24G</th>
<th>WT</th>
<th>N24A</th>
<th>N24T</th>
</tr>
</thead>
<tbody>
<tr>
<td>20:367</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21:367</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24:281</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24:283</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25:283</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Hydrogen Bonds

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>N24G</th>
<th></th>
<th>WT</th>
<th>N24G</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>ALL</td>
</tr>
<tr>
<td>12-25</td>
<td>20.05%</td>
<td>9.30%</td>
<td>19.90%</td>
<td>5.20%</td>
<td><strong>13.61%</strong></td>
</tr>
<tr>
<td>24-25</td>
<td>1.05%</td>
<td>3.20%</td>
<td>3.90%</td>
<td>0.55%</td>
<td><strong>2.18%</strong></td>
</tr>
<tr>
<td>24-281</td>
<td>33.75%</td>
<td>45.00%</td>
<td>37.95%</td>
<td>29.50%</td>
<td>36.55%</td>
</tr>
<tr>
<td>24-283</td>
<td>0.45%</td>
<td>0.05%</td>
<td>0.30%</td>
<td>0.05%</td>
<td>0.21%</td>
</tr>
<tr>
<td>25-283</td>
<td>20.65%</td>
<td>15.30%</td>
<td>48.70%</td>
<td>11.15%</td>
<td><strong>23.95%</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>N24A</th>
<th>N24T</th>
<th></th>
<th>N24A</th>
<th>N24T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>ALL</td>
</tr>
<tr>
<td>12-25</td>
<td>16.85%</td>
<td>4.15%</td>
<td>10.35%</td>
<td>10.15%</td>
<td>10.38%</td>
</tr>
<tr>
<td>24-25</td>
<td>0.00%</td>
<td>0.25%</td>
<td>0.00%</td>
<td>0.05%</td>
<td>0.08%</td>
</tr>
<tr>
<td>24-281</td>
<td>2.65%</td>
<td>6.40%</td>
<td>0.70%</td>
<td>4.05%</td>
<td>3.45%</td>
</tr>
<tr>
<td>24-283</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td><strong>0.00%</strong></td>
</tr>
<tr>
<td>25-283</td>
<td>81.60%</td>
<td>7.75%</td>
<td>16.15%</td>
<td>8.45%</td>
<td><strong>28.49%</strong></td>
</tr>
</tbody>
</table>
Overcoming drug resistance and controlling activities
Interface re-sampling

- Re-sampling of interface to change glutaminase activity
- Shown before mutation residue 248
- Using pseudo metropolis criterion
- 24,000 new complexes by GA

\[ P(\Delta s) = \begin{cases} 
  \frac{1}{\Delta s \ln 2} & \Delta s \leq 0 \\
  e^{\frac{\Delta s}{\sigma}} & \Delta s > 0 
\end{cases} \]

where

\[ \Delta s = \frac{S_{\text{new}}}{S_{\text{old}}} - 1 \]
Results

|        | A | R | N | D | C | Q | E | G | H | I | L | K | M | F | P | S | T | W | Y | V |
| N24    | 8%| 2%| 8%| 7%| 0%| 2%| 2%| 49%| 1%| 2%| 3%| 1%| 0%| 1%| 1%| 1%| 7%| 3%| 0%| 1%| 2%|
| R195   | 4%| 26%| 2%| 1%| 1%| 7%| 8%| 3%| 3%| 1%| 5%| 16%| 0%| 0%| 9%| 4%| 3%| 0%| 2%| 3%|
| Y250   | 5%| 0%| 6%| 1%| 0%| 0%| 0%| 7%| 9%| 0%| 0%| 0%| 0%| 0%| 0%| 42%| 21%| 5%| 0%| 3%| 0%|
| D281   | 2%| 3%| 15%| 10%| 1%| 1%| 13%| 39%| 1%| 0%| 1%| 2%| 0%| 0%| 2%| 7%| 2%| 0%| 0%| 1%|

The diagram and table above show the distribution of different mutant positions across a set of samples. Each column represents a different position, and the colors indicate the proportion of counts for each amino acid. The last column, labeled "Σ(AA < 10%)", sums the counts for all amino acids with less than 10% occurrence.
Free energy of binding

A

\[ \Delta G_{MM-PBSA} \text{ [kcal/mol]} \]

-140
-120
-100
-80
-60
-40
-20
0

Cavity volume and Compactness


Graph C: Cavity size [Å²] for WT, N24G, N24A, N24T, N24AY250L, and N24AR196S.
Y250L & R195S
Asparaginase activity

[Bar chart showing Asparaginase activity [U/mg] for different variants: N24G, N24A, N24T, N24A R195S, N24A Y250L, N24T R195S, N24T Y250L, and WT. The variants are compared against a control line at 1,200 U/mg.]
Glutaminase activity
Cell kill assay

![Graph showing relative IC₅₀ (WT/mut) vs. relative glutaminase activity (mut/WT) weighted by asparaginase activity for different cell lines and mutations.]

- SupB15
- MV 4:11
- REH

Key mutations:
- N24A
- Y250L
- N24A R195S
- N24T

Significance levels:
- * p < 0.05
- ** p < 0.01
Correlation predictors/experiments

- T12 hydroxyl group orientation to asparaginase activity: 0.96

- Compactness and glutaminase activity: 0.83

- Correlation glutaminase activity and cell toxicity: avg 0.73 (MV4:11 0.99, SupB15 0.7, REH 0.5)
Conclusions

- Recovered activity, overcome resistance
- N24A
  - 17% more asparaginase activity
  - 30% lower IC50
  - Reduce dosage and immunogenicity
- N24A R195S
  - As active as WT
  - Less glutaminase activity (50%)
Conclusions

• N24T
  • Double Glutaminase activity
  • No major impact on cell kill

• N24A Y250L
  • No Glutaminase activity
  • IC50 250% higher

• Glutaminase activity needed
Summary: Protein Engineering

• ?????
Acknowledgements

• Paul Bates & BMM
• Vaskar Saha & CRUK Children’s Cancer Group

• Marcin Krol
• Paul Fitzjohn

• Burkhard Rost