Protein Structure Analysis

Iosif Vaisman

2012

Protein Engineering

Increase catalytic activity
Change substrate binding site to increase specificity
Change the thermal stability
Increase proteins resistance to proteases
Change codon composition

Computational Mutagenesis

Assumption: the structural differences between each mutant and the wild-type protein are usually minor and, therefore, their tessellations are similar

Approach: a single tessellation of either the wild-type or mutant protein structure can be used to develop environmental descriptors for quantitative evaluation of changes in mutant properties

Residue and mutant score

\[ q_{\text{res}} = \sum q_i \]
\[ Q_{\text{mut}} = \sum (q_{i}^{\text{res}} - q_{i}^{\text{mut}}) \]

Dealunay simplices classification
Delaunay Tessellation of Protein Structure

Abstract each amino acid to a point
Atomic coordinates - Protein Data Bank (PDB)

Delaunay tessellation: 3D “tiling” of space into non-overlapping, irregular tetrahedral simplices. Each simplex objectively defines a quadruplet of nearest-neighbor amino acids at its vertices.

**Compositional propensities of Delaunay simplices**

\[
q_{ijkl} = \log \frac{f_{ijkl}}{p_{ijkl}}
\]

- \( f_{ijkl} \): observed quadruplet frequency,
- \( p_{ijkl} = C_{a_i a_j a_k a_l} \): residue frequency

**Examples:**

- AAAA: \( C = \frac{4!}{4!} = 1 \)
- AAAV: \( C = \frac{4!}{(3! \times 1!)} = 4 \)
- AAVV: \( C = \frac{4!}{(2! \times 2!)} = 6 \)
- AVVR: \( C = \frac{4!}{(1! \times 1! \times 1! \times 1!)} = 12 \)
- AVSR: \( C = \frac{4!}{(1! \times 1! \times 1! \times 1!)} = 24 \)

**Log-likelihood of amino acid quadruplets with different compositions**

\[
C = \frac{4!}{\prod_i (t_i !)}
\]

**Counting Quadruplets**

- assuming order independence among residues comprising Delaunay simplices, the maximum number of all possible combinations of quadruplets forming such simplices is 8855

\[
\begin{align*}
C & \quad D & \quad E & \quad F & \quad 20 \choose 4 & \quad 4845 \\
C & \quad C & \quad D & \quad E & \quad 20 \choose 2 & \quad 3420 \\
C & \quad C & \quad D & \quad D & \quad 20 \choose 2 & \quad 190 \\
C & \quad C & \quad C & \quad D & \quad 20 \times 19 & \quad 380 \\
C & \quad C & \quad C & \quad C & \quad 20 & \quad 20 \\
\end{align*}
\]

**Total:** 8855

**Reversibility Analysis**

Forward Mutation

SLE1 ‘reference’ PDB

SLE2 Calculated Mutant

Reverse Mutation

SLE1 Calculated ‘reference’

SLE2 ‘reference’ PDB

**Structural Analysis**

Reference Difference

S1,E1 ‘reference’ PDB

Mutant Difference

S1,E2 Calculated Mutant

S1,E1 Calculated ‘reference’

S1,E2 ‘reference’ PDB

S2,E1 Calculated ‘reference’

S2,E2 Calculated Mutant

S2,E2 Mutant PDB
Computational mutagenesis of T4 lysozyme
Reversibility of mutations

<table>
<thead>
<tr>
<th>Protein</th>
<th>Mutation</th>
<th>Score change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1l63</td>
<td>T26E</td>
<td>-2.49</td>
</tr>
<tr>
<td>1801</td>
<td>E26T</td>
<td>2.01</td>
</tr>
<tr>
<td>1l63</td>
<td>A82S</td>
<td>1.49</td>
</tr>
<tr>
<td>1231</td>
<td>S82A</td>
<td>-1.49</td>
</tr>
<tr>
<td>1l63</td>
<td>V87M</td>
<td>-0.28</td>
</tr>
<tr>
<td>1cu3</td>
<td>M87V</td>
<td>0.22</td>
</tr>
<tr>
<td>1l63</td>
<td>A93C</td>
<td>-1.98</td>
</tr>
<tr>
<td>138l</td>
<td>C93A</td>
<td>1.78</td>
</tr>
<tr>
<td>1l63</td>
<td>T152S</td>
<td>-1.08</td>
</tr>
<tr>
<td>1goj</td>
<td>S152T</td>
<td>1.12</td>
</tr>
</tbody>
</table>

Protein-protein and protein-DNA interfaces (HMG-D)

DNA binding residues in HMG1

Coordinate file 1ckt: Ohndorf U M et al. Nature 399:708

Universal Model Approach:
980 Experimental Mutants from 20 Proteins

Protein Engineering