Protein Structure Analysis

Iosif Vaisman

2004

Protein Structure Hierarchy

- Primary - the sequence of amino acid residues
- Secondary - ordered regions of primary sequence (helices, beta-sheets, turns)
- Tertiary - the three-dimensional fold of a protein subunit
- Quaternary - the arrangement of subunits in oligomers.

Adopted from Branden and Tooze

Anfinsen's Dogma

Three-dimensional structure of a protein is determined solely by its amino-acid sequence.

Native conformation of the protein is the global-minimum free energy conformation.

Complexity of protein structure (Levinthal paradox)

100 residue protein
3 conformations per residue

Number of distinct conformations:
$$3^{100} \cong 10^{48}$$

Sampling time \(\cong 10^{30}\) years

Levinthal paradox

3 conformations per residue is a very conservative estimate

Complexity

P (Polynomial)
complexity class of decision problems for which execution time of a computation is no more than a polynomial function of the problem size

NP (Nondeterministic Polynomial)
complexity class of decision problems for which answers can be checked by an algorithm whose run time is polynomial in the size of the input
Protein Folding Problem

Given: sequence
Find: structure

The problem is NP-complete

Protein Folding Problem

Problem for us, not for proteins.
They just fold...

(Ken Dill)

Protein Structure Determination

X-ray crystallography
NMR spectroscopy
Neutron diffraction
Electron microscopy
Atomic force microscopy

X-ray crystallography

Bragg's Law
\[ n\lambda = 2d \sin \theta \]

Phase determination: MIR and MAD
(Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction)

Fourier Transforms
**X-ray crystallography**

Electron density map created from multi-wavelength data (Arg)

**X-ray crystallography**

Experimental electron density map and model fitting (apoE four helix bundle)

**Dynamics of Database Growth**

![Graph showing database growth over time](image)

**PDB Holdings List: 28-Sep-2004**

<table>
<thead>
<tr>
<th>Prot</th>
<th>Pr/DNA</th>
<th>DNA</th>
<th>Carb total</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray</td>
<td>21606</td>
<td>1067</td>
<td>748</td>
</tr>
<tr>
<td>NMR</td>
<td>3276</td>
<td>103</td>
<td>610</td>
</tr>
<tr>
<td>Total</td>
<td>24882</td>
<td>1170</td>
<td>1358</td>
</tr>
</tbody>
</table>
PDB Growth

Growth of New Folds in PDB

Protein structure models

Adapted from D.Baker and A.Sali, Science, 2001, 294, 93-96