DNA/Protein structure-function analysis and prediction

- Protein Structure Determination:
  - X-ray Diffraction  
    (Titia Sixma, NKI)
  - Electron Microscopy/Diffraction  
    (Titia Sixma, NKI)
  - NMR Spectroscopy  
    (Lorna Smith, Oxford)
  - Other Spectroscopic methods
Spectroscopy: The whole spectrum

| Near infrared | 12,800–3333 cm⁻¹ |
| Mid infrared  | 333–3333 cm⁻¹    |
| Far infrared  | 33–333 cm⁻¹      |

- **Radiofrequency**
- **Microwave**
- **Infrared**
- **Ultra-violet**
- **Vacuum ultraviolet**
- **X-rays, γ-rays**

<table>
<thead>
<tr>
<th>log(ν/Hz)</th>
<th>ν (cm⁻¹)</th>
<th>λ (μm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>5.0 × 10⁻³</td>
<td>3 km</td>
</tr>
<tr>
<td>6</td>
<td>5.0 × 10⁻⁴</td>
<td>3 m</td>
</tr>
<tr>
<td>7</td>
<td>5.0 × 10⁻⁵</td>
<td>30 cm</td>
</tr>
<tr>
<td>8</td>
<td>5.0 × 10⁻⁶</td>
<td>3 mm</td>
</tr>
<tr>
<td>9</td>
<td>5.0 × 10⁻⁷</td>
<td>0.03 mm</td>
</tr>
<tr>
<td>10</td>
<td>5.0 × 10⁻⁸</td>
<td>300 nm</td>
</tr>
<tr>
<td>11</td>
<td>5.0 × 10⁻⁹</td>
<td>3 nm</td>
</tr>
<tr>
<td>12</td>
<td>5.0 × 10⁻¹⁰</td>
<td>3 pm</td>
</tr>
<tr>
<td>13</td>
<td>5.0 × 10⁻¹¹</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>5.0 × 10⁻¹²</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>5.0 × 10⁻¹³</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>5.0 × 10⁻¹⁴</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>5.0 × 10⁻¹⁵</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>5.0 × 10⁻¹⁶</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>5.0 × 10⁻¹⁷</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>5.0 × 10⁻¹⁸</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>5.0 × 10⁻¹⁹</td>
<td></td>
</tr>
</tbody>
</table>

- **Nuclear magnetism**
- **Rotation**
- **Vibration**
- **Electronic**
- **Nuclear**
DNA/Protein structure-function analysis and prediction

- Protein Structure Determination:
  - X-ray Diffraction
  - Electron Microscopy/Diffraction
  - NMR Spectroscopy
  - Other Spectroscopic methods
Structure determination method X-ray crystallography

- Crystallization
- Purified protein
- Crystal
- X-ray Diffraction
- Electron density
- Phase problem
- 3D structure
- Biological interpretation
Protein crystals

- Regular arrays of protein molecules
  - ‘Wet’: 20-80% solvent
  - Few crystal contacts
- Protein crystals contain active protein
  - Enzyme turnover
  - Ligand binding

Example of crystal packing
Examples of crystal packing

Acetylcholinesterase
~68% solvent

β2 Glycoprotein I
~90% solvent
(exremely high!)
Problematic proteins

- Multiple domains

- Similarly, floppy ends may hamper crystallization: change construct

- Membrane proteins

- Glycoproteins
Experimental set-up

- Options for wavelength:
  - monochromatic, polychromatic
  - variable wavelength
Structure-Function Analysis

17 Jan 2006

Diffraction image

Diffuse scattering (from the fibre loop)

Water ring

Direct beam

Beam stop

reciprocal lattice (this case hexagonal)

Reflections \((h,k,l)\) with \(I(h,k,l)\)

Increasing resolution
Scattered X-rays reinforce each other only when Bragg’s law holds:

Bragg’s law: \( 2d_{hkl} \sin \theta = n\lambda \)
Phase Problem

- If phases $\alpha_{hkl}$ and structure factor $F(hkl)$ known:

$$\rho(x, y, z) = \frac{1}{V} \sum_{hkl} \tilde{F}(hkl) \exp\{ -2\pi i (hx + ky + lz) \}$$

$$= \frac{1}{V} \sum_{hkl} |\tilde{F}(hkl)| \exp\{ i \alpha_{hkl} \} \exp\{ -2\pi i (hx + ky + lz) \}$$

- compute the electron density $\rho(x,y,z)$
  - In the electron density build the atomic 3D model

- However, the phases $\alpha_{hkl}$ are **unknown**!
How important are these phases??

- Fourier transform photo’s of Karle (top left) and Hauptman (top right) (two crystallography pioneers)

- Combine amplitudes $F_K$ with phase $\alpha_H$ and inverse-fourier transform

- Combine amplitudes $F_H$ with phase $\alpha_K$ and inverse-fourier transform

(Taken from: Randy J. Read)
How can we solve the Phase Problem?

- **Direct Methods**
  - small molecules and small proteins
  - needs atomic resolution data $(d < 1.2 \text{ Å})$

- **Difference method using heavy atoms**
  - multiple isomorphous replacement (MIR)
  - anomalous scattering (AS)
  - combinations (SIRAS, MIRAS)
    - Difference method using variable wavelength
      - multiple-wavelength anomalous diffraction (MAD)

- **Using a homologous structure**
  - molecular replacement
Building a protein model

- Find structural elements:
  - $\alpha$-helices, $\beta$-strands
- Fit amino-acid sequence
Building a protein model

- Find structural elements:
  - $\alpha$-helices, $\beta$-strands
- Fit amino-acid sequence
Effects of resolution on electron density

Note: map calculated with perfect phases

\[ d = 4 \text{ Å} \]
Effects of resolution on electron density

Note: map calculated with perfect phases

\[ d = 3 \text{ Å} \]
Effects of resolution on electron density

\[ d = 2 \text{ Å} \]

Note: map calculated with perfect phases
Effects of resolution on electron density

\[ d = 1 \text{ Å} \]

Note: map calculated with perfect phases
Refinement process

- Bad phases
  \[ \rightarrow \text{poor electron density map} \]
  \[ \rightarrow \text{errors in the protein model} \]

- Interpretation of the electron density map
  \[ \rightarrow \text{improved model} \]
  \[ \rightarrow \text{improved phases} \]
  \[ \rightarrow \text{improved map} \]
  \[ \rightarrow \text{even better model} \]

... iterative process of refinement
Validation

- Free R-factor (cross validation)
  - Number of parameters/observations
- Ramachandran plot
- Chemically likely (WhatCheck)
  - Hydrophobic inside, hydrophilic outside
  - Binding sites of ligands, metals, ions
  - Hydrogen-bonds satisfied
  - Chemistry in order
- Final B-factor values
DNA/Protein structure-function analysis and prediction

- Protein Structure Determination:
  - X-ray Diffraction
  - Electron Microscopy/Diffraction
  - NMR Spectroscopy
  - Other Spectroscopic methods
Electron microscopy

- Single particle
  - Low resolution, not really atomic
  - Less purity of protein, more transient state analysis
- Two-dimensional crystals
  - Suited to membrane proteins
- Fibres
  - Acetylcholine receptor
  - Muscles, kinesins and tubulin
- Preserve protein by
  - Negative stain (envelope only)
  - Freezing in vitreous ice (Cryo-EM, true density maps)
- High resolution possible but difficult to achieve
  - For large complexes: Combine with X-ray models
Electron diffraction from 2D crystals: Nicotinic Acetylcholine Receptor

G-protein coupled receptors

Unwin et al, 2005
Electron diffraction: near atomic resolution

- Structure of the alpha beta tubulin dimer by electron crystallography.
  Nogales E, Wolf SG, Downing KH.
  Nature 1998 391 199-203
Single particle Electron Microscopy

Select particles

Sort into classes

Average

Reconstruct 3D image
Cryo EM reconstruction: Tail of bacteriophage T4

DNA/Protein structure-function analysis and prediction

• Protein Structure Determination:
  – X-ray Diffraction
  – Electron Microscopy/Diffraction
  – NMR Spectroscopy
  – Other Spectroscopic methods
1D NMR spectrum of hen lysozyme (129 residues)

- Too much overlap in 1D: → 2D
- $^1H-^1H$
- $^1H-^{15}N$
- $^1H-^{13}C$
Step 1: Identification of amino acid spin systems
2D COSY spectrum of peptide in D$_2$O

Val $\beta\gamma$

Val $\alpha\beta$

Thr $\beta\gamma$

D1 (ppm)

D2 (ppm)
Step 2: Sequential assignment

\[ \alpha H(i)-NH(i+1) \text{ NOE} \]

\[ NH(i)-NH(i+1) \text{ NOE} \]

\[ \alpha H(i)-NH(i+1) \text{ NOE} \]
2D NOESY spectrum

- Peptide sequence (N-terminal NH not observed)
- Arg-Gly-Asp-Val-Asn-Ser-Leu-Phe-Asp-Thr-Gly
Nitroreductase Dimer: 217 residues

- Too much overlap in 2D:
  → 3D
- $^1\text{H} - ^1\text{H} - ^{15}\text{N}$
- $^1\text{H} - ^{13}\text{C} - ^{15}\text{N}$
Structural information from NMR: NOEs

• For macromolecules such as proteins:
  – Initial build up of NOE intensity $\propto \frac{1}{r^6}$
  – Between protons that are < 5Å apart
NOEs in $\alpha$-helices

- NH-NH(i,i+1) 2.8Å
- $\alpha$H-NH(i,i+3) 3.4Å
- $\alpha$H-$\beta$H(i,i+3) 2.5-4.4Å

- Cytochrome c$_{552}$ 3D $^1$H-$^{15}$N NOESY
  - $\alpha$H-NH(i,i+3)
  - residues 38-47
  - form $\alpha$-helix
NOEs in β-strands

- $\alpha$H-NH(i,j) 2.3 Å
- $\alpha$H-NH(i,j) 3.2 Å
Long(er) range NOEs

- Provide information about
  - Packing of amino acid side chains
  - Fold of the protein
- NOEs observed for $\text{C}_\alpha\text{H}$ of Trp 28 in hen lysozyme
Spin-spin coupling constants

- Fine structure in COSY cross peaks
- For proteins $3J(HN,H_{\alpha})$ useful:
  - Probes main chain $\phi$ torsion angle
- Hen lysozyme:
Structural information from NMR

3. Hydrogen exchange rates

- Dissolve protein in D$_2$O and record series of spectra
  - Backbone NH ($^1$H) exchange with water ($^2$H)
- Slow exchange for NH groups
  - In hydrogen bonds
  - Buried in core of protein

\[ \text{After 20 mins} \quad \text{After 68 hours} \]

$^1$H-$^{15}$N HSQC spectra for SPH15 in D$_2$O
## Structural information from NMR

- **NOEs**
  - $1H_\alpha - 2H_\beta$: 1.8-2.5 Å (strong)
  - $2H_\alpha - 40HN$: 1.8-5.0 Å (weak)
  - $3H_\varepsilon - 88H_\gamma 2$: 1.8-5.0 Å (weak)
  - $3H_\delta - 55H_\beta$: 1.8-3.5 Å (medium)

- **Spin-spin coupling constants**
  - $1C-2N-2C_\alpha-2C$: $-120+/-40^\circ$
  - $4C-5N-5C_\alpha-5C$: $-60+/-30^\circ$
  - $5C-6N-6C_\alpha-6C$: $-60+/-30^\circ$

- **Hydrogen exchange rates**
  - $10HN-6CO$: 1.3-2.3 Å  
    $10N-6CO$: 2.3-3.3 Å
  - $11HN-7CO$: 1.3-2.3 Å  
    $11N-7CO$: 2.3-3.3 Å
NMR structure determination: hen lysozyme

- 129 residues
  - ~1000 heavy atoms
  - ~800 protons
- NMR data set
  - 1632 distance restraints
  - 110 torsion restraints
  - 60 H-bond restraints
- 80 structures calculated
- 30 low energy structures used
Solution Structure Ensemble

• Disorder in NMR ensemble
  – lack of data?
  – or protein dynamics?
DNA/Protein structure-function analysis and prediction

- Protein Structure Determination:
  - X-ray Diffraction
  - Electron Microscopy/Diffraction
  - NMR Spectroscopy
  - Other Spectroscopic methods
Ultrafast Protein Spectroscopy

- Structure-sensitive technique
- State of protein and substrate
  - Redox state
  - Protonation
  - Elektronen
- Follow reactions in real-time
- Why ultrafast spectroscopy?
  - Molecular movement:
    - Time scales of 10 fs – 1 ps (10^{-15} – 10^{-12} s)
  - O ↔ H stretching frequency: ~3500 cm^{-1} (~ 10^{-14} s)
- Transfer or movement of proton, electron or C-atom
Excited state difference spectra of Chl and Pheo

Keto and ester C=O in Chlorophyll a, Pheophytin a are redox and environmental probes
Why is vibrational spectroscopy sensitive to structure?

$X - C - O - H$

$O - X$

$\omega_1$  $\omega_1'$

$\omega_2$  $\omega_2'$
H-bond response during the photocycle

Replace Glu with Gln which donates weaker H-bond
## Summary

<table>
<thead>
<tr>
<th>Method</th>
<th>Pro</th>
<th>Con</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray crystallography</td>
<td>High resolution</td>
<td>Crystals required</td>
</tr>
<tr>
<td></td>
<td>No size limits</td>
<td>Phase problem</td>
</tr>
<tr>
<td></td>
<td>Easy addition of ligands</td>
<td></td>
</tr>
<tr>
<td>Electron microscopy / diffraction</td>
<td>Single particles possible</td>
<td>Labour intensive</td>
</tr>
<tr>
<td></td>
<td>Well suited to membrane proteins</td>
<td>High resolution difficult</td>
</tr>
<tr>
<td>NMR spectroscopy</td>
<td>In solution</td>
<td>Resolution variable</td>
</tr>
<tr>
<td></td>
<td>Dynamic information possible</td>
<td>Limited size (&lt; ~30kDa)</td>
</tr>
<tr>
<td>Other spectroscopy</td>
<td>(Very) high time resolution (ps!)</td>
<td>Technically extremely complex</td>
</tr>
<tr>
<td></td>
<td>Very sensitive</td>
<td>Limited applicability</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Limited information</td>
</tr>
</tbody>
</table>