Introduction to Protein Structure

Function, evolution & experimental methods

Thomas Holberg Blicher
NNF Center for Protein Research
University of Copenhagen
Learning Objectives

- Outline the basic levels of protein structure.

- Outline key differences between X-ray crystallography and NMR spectroscopy.

- Identify relevant parameters for evaluating the quality of protein structures determined by X-ray crystallography and NMR spectroscopy.
Outline

- Protein structure, evolution and function
  - Inferring function from structure.
  - Modifying function

- Experimental techniques
  - X-ray crystallography
  - NMR spectroscopy

- Structure validation
"We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest....

...It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."

Once Upon a Time…

“Could the search for ultimate truth really have revealed so hideous and visceral-looking an object?” Max Perutz, 1964, on protein structure

John Kendrew, 1959, with myoglobin model
Why are Protein Structures so Interesting?

- They provide a detailed picture of interesting biological features, such as active site, substrate specificity, allosteric regulation etc.

- They aid in rational drug design and protein engineering.

- They can elucidate evolutionary relationships undetectable by sequence comparisons.
Protein Synthesis

- newly born protein
- amino acids
- large subunit
- tRNA
- mRNA
- small subunit
- P site
- A site
Proteins Are Polypeptides

- A polypeptide chain
- Hydrophobic collapse
Protein Folding

- Initially formed structure is in molten globule state (ensemble).

- Molten globule condenses to native fold via transition state.
Protein Folding

- Hydrophobic collapse
  - Hydrophobic residues cluster to “escape” interactions with water.

Myoglobin

Surface

Interior
Hydrophobic vs. Hydrophilic

- Globular protein (in solution)
  - Myoglobin

- Membrane protein (in membrane)
  - Aquaporin
Hydrophobic vs. Hydrophilic

- Globular protein (in solution)
- Membrane protein (in membrane)

Cross-section
Myoglobin

Cross-section
Aquaporin
Backbone Problems?

- Polar backbone groups form regular secondary structure to satisfy hydrogen bonding donors and acceptors.
Characteristics of Helices

- Aligned peptide units $\rightarrow$ Dipolar moment
- Ion/ligand binding
- Secondary and quaternary structure packing
- Capping residues
- The $\alpha$ helix ($i \rightarrow i + 4$)
- Other helix types! ($3_{10}$, $\pi$)
**β-Sheets**

- Multiple strands → sheet
  - Parallel vs. antiparallel
  - Twist

- Flexibility
  - Vs. helices
  - Folding
  - Structure propagation (amyloids)
  - Other…

Thioredoxin
β-Sheets

- Multiple strands → sheet
  - Parallel vs. antiparallel
  - Twist

- Flexibility
  - Vs. helices
  - Folding
  - Structure propagation (amyloids)
  - Other…
**β-Sheets**

- Multiple strands → sheet
  - Parallel vs. antiparallel
  - Twist

- Flexibility
  - Vs. helices
  - Folding
  - Structure propagation (amyloids)
  - Other...
β-Sheets

- Multiple strands → sheet
  - Parallel vs. antiparallel
  - Twist

- Flexibility
  - Vs. helices
  - Folding
  - Structure propagation (amyloids)
  - Other…
- Multiple strands $\rightarrow$ sheet
  - Parallel vs. antiparallel
  - Twist

- Strand interactions are non-local

- Flexibility
  - Vs. helices
  - Folding
Turns, Loops & Bends

- Between helices and sheets
- On protein surface
- Intrinsically “unstructured” proteins
Residue Patterns

- **Helices**
  - Helix capping
  - Amphiphilic residue patterns

- **Sheets**
  - Amphiphilic residue patterns
  - Residue preferences at edges vs. middle

- **Special residues**
  - Proline
    - Helix breaker
  - Glycine
    - In turns/loops/bends
Structure Levels

- Primary structure = Sequence
- Secondary Structure = Helix, sheets/strands, loops & turns
- Structural Motif = Small, recurrent arrangement of secondary structure, e.g.
  - Helix-loop-helix
  - Beta hairpins
  - EF hand (calcium binding motif)
  - Etc.
- Tertiary structure = Arrangement of Secondary structure elements
Quaternary Structure

- Assembly of monomers/subunits into protein complex
  - Backbone-backbone, backbone-side-chain & side-chain-side-chain interactions:
    - Intramolecular vs. intermolecular contacts.
    - For ligand binding side chains may or may not contribute. For the latter, mutations have little effect.

- Myoglobin

- Hemoglobin
Structures of Unstructured Regions

- Estimate: 20% of all proteins contain unstructured regions.
  - 1% of structures in PDB contain unstructured regions.

- Structural genomics
  - Special structural genomics projects
  - Selection and modification of targets
  - Prediction of crystallisable domains

Iakoucheva & Dunker, *Structure* 2003
Degrees of Structure

- **Unstructured (conformational ensemble)**
  - For example, ACTR (no NCBD)

- **Molten globule (conformational ensemble)**
  - For example, NCBD (no ACTR)

- **Linked folded domains (beads on a string)**
  - For example, zinc fingers (no DNA)

- **Mostly folded, local disorder**
  - For example, eIF4E (N terminus is unfolded)

**Folding on target binding**

- **ACTR-NCBD complex**
- **Zinc-finger1-3-DNA complex**
- **eIF4E-eIF4G complex**
What’s the Fuss About?

- Properties of Disordered Regions
  - Flexible, i.e. adaptable
  - Accessible
    - Contain Extended Linear Motifs (ELM)
  - Different behaviour in interaction interfaces
    - Very adaptable
    - Many hydrophobic interactions (close packing)
  - No fixed structure without interaction partner
  - Folding upon binding
Proteins Are Polypeptides

- The peptide bond
- A polypeptide chain
Ramachandran Plot

- Allowed backbone torsion angles in proteins

The Ramachandran Plot:

- Beta-sheet
- Left handed alpha-helix
- Right handed alpha-helix

Peptide torsion angles:

N, phi, psi, omega

C-alpha

H, R, H
Torsion Angles

\[ \phi = 180^\circ \]
\[ \phi = 0^\circ \]
\[ \phi = +90^\circ \]
\[ \phi = -90^\circ \]
Ramachandran Plots
The Amino Acids

Acidic and amide side chains
- Aspartate
- Asparagine
- Glutamate
- Glutamine

Basic side chains
- Lysine
- Histidine
- Arginine

Aliphatic side chains
- Valine
- Isoleucine
- Leucine
- Glycine
- Alanine

Aromatic side chains
- Tryptophan
- Phenylalanine
- Tyrosine

Hydroxyl or sulfur-containing side chains
- Serine
- Methionine
- Threonine
- Cysteine

Cyclic side chain
- Proline

http://www.ch.cam.ac.uk/magnus/molecules/амино/
Grouping Amino Acids

Aliphatic

Sulphur Containing

Aromatic

Hydrophobic

Hydroxylic

Tiny

Small

Acidic

Amino Acids

A alanine (ala)
R arginine (arg)
N asparagine (asn)
D aspartic acid (asp)
C cysteine (cys)
Q glutamine (gln)
E glutamic acid (glu)
G glycine (gly)
H histidine (his)
I isoleucine (ile)
L leucine (leu)
K lysine (lys)
M methionine (met)
F phenylalanine (phe)
P proline (pro)
S serine (ser)
T threonine (thr)
W tryptophan (trp)
Y tyrosine (tyr)

http://www.dreamingintechnicolor.com/InfoAndIdeas/AminoAcids.gif
The Evolution Way

- Based on Blosum62 matrix
- Measure of evolutionary substitution probability
In evolution **structure** is conserved longer than both **function** and **sequence**.
Form vs. Function

- Divergent evolution
  - Common ancestor
  - New function

- Convergent evolution
  - Different ancestor
  - Same function
Sequence vs. Function – I

- **Trypsin**
  - positive

- **Chymotrypsin**
  - large hydrophobic

- **Elastase**
  - Small hydrophobic

- Divergent evolution
  - Same fold
  - Different specificities
  - Small changes in binding pocket
Sequence vs. Function – II

- Trypsin
- Subtilisin

Convergent evolution
Engineering & Design

- **Protein engineering**
  - Overpacking
  - Buried polar groups
  - Cavities

- **Drug design**
  - Target specificity/selectivity
  - Function
  - Mutations

**COX-1/COX-2**
- Arthritis
- Designed to prevent drug side effects

http://publications.nigms.nih.gov/structlife/chapter4.html

**HIV protease**

Blundell et al. (2002), *High-throughput crystallography for lead discovery in drug design*, Nature Reviews Drug Discovery 1, 45-54.

Experimental Methods

X-ray crystallography
&
NMR spectroscopy
Methods for Structure Determination

- X-ray crystallography
- Nuclear Magnetic Resonance (NMR)
- Modelling techniques

- More exotic techniques
  - Cryo electron microscopy (Cryo EM)
  - Small angle X-ray scattering (SAXS)
  - Neutron scattering
X-ray Crystallography

- No size limitation.
- Protein molecules are "stuck" in a crystal lattice.
- Some proteins seem to be uncrystallizable.
- Slow.

- Especially suited for studying structural details.

- Lattice and unit cell
X-rays

Fourier transform
The Importance of Resolution

4 Å
3 Å
2 Å
1 Å

low

high
Key Parameters

- **Resolution**
- **R values**
  - Agreement between data and model.
  - Usually between 0.15 and 0.25, should not exceed 0.30.
- **Ramachandran plot**
- **B factors**
  - Contributions from static and dynamic disorder
    - Well determined ~10-20 Å², intermediate ~20-30 Å², flexible 30-50 Å², invisible >60 Å².
NMR Spectroscopy

- Upper limit for structure determination currently ~50 kDa.
- Protein molecules are in solution.
- Dynamics, protein folding.
- Slow.

- Especially suited for studies of protein dynamics of small to medium size proteins.
NMR Basics

- NMR is nuclear magnetic resonance
- NMR spectroscopy is done on proteins IN SOLUTION
- Only atoms $^1\text{H}$, $^{13}\text{C}$, $^{15}\text{N}$ (and $^{31}\text{P}$) can be detected in NMR experiments
- Proteins up to 30 kDa
- Proteins stable at high concentration (0.5-1mM), preferably at room temperature
NMR Spectroscopy

Time Domain Spectrum

Structure of a molecule
Evalutation of NMR Structures

- Atomic backbone RMSD:

\[
RMSD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \langle x_i \rangle)^2}{n}}
\]

Well-defined structures
RMSDs < 0.6 Å

Less well-defined structures
RMSDs > 0.6 Å

1T1H, Andersen et al. JBC, 2004

3GF1, Cooke et al. Biochemistry, 1991
Evaluation of NMR Structures

What regions in the structure are most well-defined?

Look at the pdb ensembles to see which regions are well-defined

1RJH
Nielbo et al, Biochemistry, 2003
Which Structural Model?

- Normally NMR structure models are listed according to the total energy and the number of violations.
- Model 1 in the PDB file is often the one with lowest energy and fewest violations.
- Use that model as template for modelling.
NMR versus X-ray Crystallography

- Hydrogen atoms are observed!
- Only $^{13}\text{C},^{15}\text{N}$ and $^1\text{H}$ are observed
- Study of proteins in solution
- Only proteins up to 30-40 kDa
- No total “map” of the structure
- Information used is incomplete and used as restraints
- An ensemble of structures is submitted to PDB
- The solved structure can be used for further dynamics characterization with NMR
Summary I – Protein Structure

- Proteins consist of amino acids.
- Polypeptide chains fold into specific 3D structures.
- **Function** is performed by the **folded** protein.
- Proteins are dynamic and only marginally stable.

*Image adapted from: National Human Genome Research Institute.*
Summary

- In evolution structure is conserved longer than both function and sequence.

- X-ray crystallography
  - Proteins in crystal lattice
  - Many details – one model
  - Resolution, R-values, Ramachandran plot

- NMR spectroscopy
  - Proteins in solution
  - Fewer details – many models
  - Violations, RMSD, Ramachandran plot