BE.342/442 Thursday, October 6, 2005
Topic: Practical Aspects of Single Crystal X-ray Crystallography, Part 2

X-ray diffraction was key to solving structures such as RNA and collagen. However, X-ray crystallography of fibers is extremely difficult, because the fibers entangle rather than aligning in a crystalline structure.

Important developments in structural biology in the last decade include: expression of proteins via gene coding (rather than purification from tissue! one exception is rhodopsin, which is sometimes purified from cow's eye), availability of beamlines at synchrotron radiation sources, and computing hardware and integrated software packages.

How is synchrotron radiation produced? (E.g, CERN in Geneva, Switzerland.) Charged particles are bent around a ring, releasing high-energy light tangential to the curve of their path. The highest available energies have been climbing steadily since the 1960's with X-ray tubes, until modern third-generation light sources such as ESRF, which have a billion times the brilliance of x-ray tubes.

At the frontier of structural biology is the structural determination of

- the nucleosome core, where DNA is wrapped around histones (1997)
  - 8 proteins
  - 146 DNA base pairs
  - 52 X-ray reflections
- the ribosome subunit 30S (2000)
  - 21 proteins
  - 1542 nucleotides
  - 254 reflections
- the ribosome subunit 50S (2000)
  - 35 proteins
  - 3020 nucleotides
  - 666,000 reflections
- and the bluetongue virus core (1998)
  - 900 proteins
  - 19,000 RNA bases
  - 3,300,000 reflections

Without synchrotron radiation, enormous structures such as the ribosome could not have been solved.

Structural genomics aims to provide a structure or a model for every protein in all sequenced genomes. Enourmous funds have gone into structural genomics initiatives, especially from venture capitalist sources. However, the field has many challenges. For example, of the proteins that can be cloned from a typical genome, only approximately one-third are soluble (membrane proteins—forget it!), and only four-fifths of these soluble proteins can be crystallized.

Other light sources:

Linear accelerator (e.g., SLAC at Stanford, CA) provides a highly focused beam of higher energies than a synchrotron.
X-ray free electron laser (F-EEL): an undulator focuses electrons into microbunches over several hundred meters.

Free electron lasers would allow amazing studies, such as
structural studies on a single special of macromolecular assembles,
virus structures and the structure of viral genomes,
imaging of 2-D crystals of membrane proteins,
time-resolved experiments,
even X-ray tomography of whole cells!

Challenges include:
Sample handling (container-free methods, mass spectroscopy, electrospray ionization)
Phase determination (oversampling, holographic methods)
Reduction of pulse length (currently 200 femtoseconds, optimal is10 femtoseconds).

Ron Ruth, et al. of Lyncean Technologies is developing a tabletop x-ray diffractometer that can fit into a 3-meter-long space (discussed last time).

Collagen

About 30% of all human proteins in the human body are collagen, as with most animals. Collagen allows multicellular organisms to hold their cells together.

There are 27 types of collagen: Type I, II, II, IV, …. XXVII.

Collagen is an important biological material for tissue culture, tissue engineering, and regenerative medicine. It is also a useful biomaterial for cosmetics and reconstructive surgery. However, there are concerns with using xenografted (animal) collagen: animal diseases can enter the human population this way.

Collagen self-assembly in a cell:
After translation in ribosomes, collagen is modified (hydroxylated and glycosylated) to increase hydration properties. Collagen self-assembles into trimers (triple helices) inside the cell. Then it is secreted into the extracellular space, where it can further self-assemble into fibers and bundles.

Collagen structures:
Electron microscopy images of collagen showed banding patterns. An important repeating unit in collagen is the sequence $(\text{Gly-Pro-Pro})_2$. In general, the sequence $(\text{Gly-X-Y})_n$ is important for free turns in a protein. The pattern of glysine in every third position of the chain enables the triple-helical structure. A 24-residue segment of the collagen triple helix was determined by x-ray diffraction, initially by Alexander Rich and Francis Crick. They build a simple triple helical mall-and-stick model. The three strands winding around each other are held together by intermolecular hydrogen bonds.

Forty years after Crick and Rich, in 1994, a space-filling model of the collagen structure confirmed their model.

Compare the three well-understood helical structures in biology: the alpha-helix, beta-helix, and collagen triple helix.

Collagen fibers will align so that hydrophobic alanine groups match up across adjacent strands, to maximize hydrophobic interactions. This controls the crystallization of collagen. A single Gly→Ala substitution at the position $(POG)_4$-PO**A**-$(POG)_5$ allowed a 30-residue peptide chain to form a single crystal for high-resolution x-ray diffraction.

A key advancement in the understanding of collagen was the 1.9-angstrom resolution structure, which essentially solved the collagen structure. This advancement provided detailed structural information for the low-resolution collagen model that had previously been intractable. It also provided information relevant to all collagen-like helices, especially those bearing glysine substitutions. The new model showed that formation of inter-chain hydrogen bonds keep the three helices in a specific register.

Experiments showed that supercoiling of 3 individual polyproline helices formed a left-handed triple helical structure. Triple helices can change their helical twists through small variations of the main chain torsion angles without an appreciable change in the unit height or inter-chain H-bond pattern. Local twist variations may result from twist sequence variations, such as PYG, XOP, or XYG instead of POG.

Hydration of collagen is key to its structure. (This explains why collagen hydrates so well, and is such a great material for cosmetics!) Water molecules mediate inter-chain H-bonds wherever there is a point disruption of the optimal sequences. The outside of the triple helix is surrounded by a hydration shell in aqueous solution, trapping an enormous number of water molecules. The cylindrical and semi-clathrate-like hydration shell determines the lateral separation of the triple helices in macromolecular assembly by forming water “bridges” that are sensitive to the individual residues in the collagen triple helix. For example, hydroxyprolines interact with the hydration cylinder to provide more hydrogen bonds.

These point mutations can cause changes in the behavior of collagen through then hydration shell. This accounts for the mechanical differences between the different collagen types found in different tissues of the body, and also accounts for the effects of certain genetic diseases involving collagen.