PROTEIN STRUCTURE DETERMINATION

Determination of protein secondary structures

Often times in a protein, one would like a simple method to determine the presence of secondary structures, and also if there are any changes in the secondary structures in a protein in response to some external stimuli. Two methods that have been used for this are circular dichroism spectroscopy and infrared spectroscopy.

Proteins in solution are asymmetric, due to the fact that they are comprised of chiral molecules in a single stereoisomeric form. As a result, these molecules in solution will absorb differently light which is circularly polarized in the clockwise direction and light that is circularly polarized in the counterclockwise direction. Achiral molecules would absorb the two types of light equally at a given wavelength. In circular dichroism, macromolecular structures give the strongest absorption characteristics in the molecule. In particular, the absorbance of the amide bonds are measured, since the orientation and alignment of the peptide bonds in a repeating secondary structure would provide multiple identical contributions to the CD spectrum.

Here are examples of three CD spectra of proteins that form a single secondary structure. Poly-L-glutamic acid forms an α-helix in solution at pH 4.5. The CD spectrum of this α-helical protein is characterized by two minima, one at about 207 nm and the second at about 222 nm. Poly-L-(lysine-leucine) folds into an antiparallel β-sheet; this spectrum is best characterized by a minimum at about 218 nm. Alanine-glycine forms a continuous β-turn; this has a maximum at 206 nm and a minimum at about 224 nm.

These characteristics can be used in a real protein to determine the relative amounts of the different secondary structures in a protein. Figure 10.16 shows the CD spectra of hemoglobin, EcoR1 endonuclease, and tumor necrosis

factor alpha. Hemoglobin is a protein that is almost entirely α-helix, and this shows in the CD spectrum by the strong double mimima specta. Tumor necrosis factor contains no α-helix and a large amount of β-sheet, resulting in the spectrum shown on the overhead. EcoR1 endonuclease has both α-helix and β-strand, and as a result the CD spectrum has contributions from both structures.

In 1981, Hennessey and Johnson measured the CD spectra of sixteen different proteins whose three dimensional structure was known by x-ray diffraction. They developed an algorithm from this collection of spectra to estimate the relative amounts of helix, beta strand, and beta turn from a CD spectrum. These are available on line for free; one example can be found at:

http://www.embl-heidelberg.de/~andrade/k2d.html

Another method that has proven very useful in the determination of protein secondary structure is infrared spectroscopy. The amide bonds that comprised the connections between the amino acids in proteins may have different vibrational frequencies depending upon their structure. The two bands of interest are the C=O stretch (also called the Amide I band), and a band generated by combining the N-H bend with the C-N stretch (also called the Amide II band). The amide I band has a vibrational frequency of 1653 wavenumbers for an α-helix and 1640 wavenumbers for a β-strand. The amide II band has a vibrational frequency of 1545 for an α-helix and 1525 for a β-strand.

Byler and Susi have developed the most popular method of analyzing FTIR spectra of proteins for secondary structure. They measured the IR spectra for 21 proteins in the amide I region between 1600 – 1700 cm-1. They compared the spectra for the amide I band with the secondary structure for the proteins determined by x-ray crystallography. By doing a Fourier deconvolution (a special mathy kind of thing), they were able to determine that the amide I band for any particular protein is comprised of eleven possible components, where not all

proteins contain every component. The eleven components are:

α-helix: 1654 cm-1

β-strand: 1624, 1631, 1637, 1675 cm-1

β-turns: 1663, 1670, 1683, 1688, 1694 cm-1

other structures: 1645 cm-1

Here is an example of how the data can be used to estimate secondary structures. The solid line is a spectrum of ribonuclease S, and it clearly has a blobby like appearance. Using a deconvolution technique, six bands are defined with varying intensities. Analysis of these bands estimates the secondary structure of ribonuclease as 50% α-helix and 25% β-strand, in excellent agreement with the x-ray structure.

These methods are relatively easy to perform, requiring small amounts of protein and somewhat affordable instrumentation. To determine the precise global three dimensional structure of a protein, more elaborate methodology is necessary.

RADIATION TECHNIQUES FOR EXAMINING PROTEIN STRUCTURE

A number of other physical techniques can be used to characterize and understand proteins. One type of physical measurements can generally be listed as RADIATION TECHNIQUES. Radiation techniques involve exposing proteins, usually in solution, sometimes in crystalline form, to radiation of some energy. A measurement is made on how the protein either absorbs or scatters the applied energy. The methods I'd like to discuss in brief are:

NMR and

X-ray crystallography

1. Nuclear magnetic resonance - I could take a semester to attempt to explain the concepts and applications of NMR and not do an adequate job, but I will go ahead

and attempt to introduce NMR relative to protein structure. Nuclear magnetic resonance can be used to measure nuclear transitions in atoms that are capable of undergoing nuclear transitions. The isotopes that are most useful in protein NMR are H-1, C-13, and N-15. H-1 is 99.9% naturally abundant, and if you learned anything about NMR in the past, this is typically where you began. C-13 is about 1.1% naturally abundant, which is a fairly small amount but can still be used. N-15 is about 0.37% abundant, small but again useful. For the NMR-savvy among us, all of these nuclei are spin ½, meaning that a single transition is measured and barring environmental effects, a single peak is observed.

If one takes a nucleus such as H-1, this nucleus has a single proton which can exist in one of two different spin states. These spin states are slightly different in energy. If you add energy to the system at the same frequency as the energy difference, you will cause the nuclei to "excite" from the lower spin state to the higher spin state. If you stop applying the energy, you can follow the relaxation of the excited spins back down to the ground state. This, in a nutshell, is the NMR phenomenon.

The difference in energy between the two spin states is very small at room temperature under normal conditions, small enough that there is sufficient thermal energy available to allow the transition to occur. To increase the energy difference between the two spin states, we put the nuclei into a very strong magnetic field. This prevents transition between the two spin states until radiofrequency energy is applied.

Again, to review, the frequency of the energy needed to make the transition depends upon the environment that the nuclei find themselves in. On an organic molecule, slightly more energy will be needed to cause the transition in a hydrogen bound to an aliphatic carbon than to a carbon bound to an aromatic carbon. Neighboring oxygens and nitrogens will also affect the energy needed for

the transition. In an organic molecule containing many protons, a distribution of energies will be needed to excite all of the protons, and these are displayed as a series of peaks in an NMR spectrum. By convention, the frequencies needed to carry out the transition are divided by the strength of the applied magnetic field, and are reported as ratio units called parts per million (ppm).

Structural information is derived from NMR spectra because nuclei interact with each other both through covalent bonds and through space. Through-bond interactions are derived using a method called Correlation Spectroscopy (or COSY for short), and through space interactions are measured using Nuclear Overhauser effect spectroscopy (NOESY).

COSY -

This type of experiments describe through-bond relationships in proteins - that is, that protons from an arg are adjacent on the protein from protons from a his. This method takes advantage of spin-spin coupling that occurs between protons on adjacent atoms in a protein molecule. For example, on a simple organic molecule such as ethanol, the proton spectrum would consist of three peaks, one for the methyl protons, one for the ethyl protons, and one for the hydroxyl proton. However, the fact that each proton exists in two spin states of approximately equal population means that the electronic environment experienced by each individual proton causes the resonance to be split into two peaks.

Coupling constants can be used effectively to determine atom connectivities in a protein. It can also be used to determine the orientation of hydrogens near each other in the sequence. A simple example of this is the coupling constants associated with ring systems such as cyclohexanol. If you look at the protons adjacent to the H on the hydroxyl group, these will couple in different ways with different magnitudes of coupling constants. The coupling

constant is dependent upon the bond angle between the two interacting hydrogens, as seen on the board. This relationship is called the Karplus relationship, and can be used to determine the angles of hydrogens connected through bonds.

A more powerful NMR method in terms of structural determination is taking advantage of the Nuclear Overhauser Effect (NOE). In NOE, nuclei that are excited have the potential to influence the excitation of nuclei that are nearby in space, by either increasing or decreasing their apparent absorbance strength. The magnitude of the effect is a function of distance between the two interacting nuclei. NOESY thus provides spatial information, and can be used to determine which amino acids in a protein are close to each other in space (while not necessarily close to each other in sequence.

Deconvolution of through-bond and through-space relationships of protons in a protein will, after a great deal of analysis, lead to a 3-dimensional structure. This used to be a very time consuming and computer intensive process, but has evolved now to require a simple linux cluster and a relatively unsophisticated computer system.

X-RAY DIFFRACTION AND CRYSTALLOGRAPHY

X-ray diffraction is a powerful method for determining the 3-dimensional structure of a variety of proteins. The general procedure in crystallography is:

1. Purify large quantities of a protein of interest. Crystallography requires mg quantities of pure protein. Sometimes, multiple purification conditions are attempted; proteins purified using one type of purification will not crystallize, but proteins in a different environment may form crystals. Establishing crystallization conditions is still a trial and error process, and the amount of protein needed depends as much on luck as it does on proper technique.
2. Prepare a proper crystal - Crystal preparation is somewhat of an art.

Historically, a relatively large, very regular crystal is needed, with about a 0.5 mm diameter on one side. With higher power irradiation sources, the necessary crystal size is decreased, such that crystals with diameters of less than 0.1 mm are often suitable for analysis.

To make a crystal, you take a concentrated solution of pure protein (usually around 20 mg/mL or so) and mix it with another solution containing some combination of protein precipitants. These may be salts, organic molecules, pH changes, etc., that will induce the protein to form crystals. This mixture is then placed in a sealed chamber that will provide a means for the slow dehydration of the protein solution, gently inducing crystallization. A popular method is the "hanging drop" method, where 1 uL of protein solution and 1 uL of precipitant solution are mixed onto a glass cover slip. This slip is inverted over a chamber containing 1 mL of the precipitant solution. Since the solution in the chamber is more concentrated than the solution on the drop, the system will attempt to come to equilibrium by the slow transfer of water vapor from the drop to the precipitant reservoir. This often induces precipitation to occur.

Crystal preparation is now often carried out with robotic systems that will mix together solutions of protein and precipitant in an ordered manner. This eliminates the need for human manipulation of the solutions. Robots don't eat, sleep, or make pipetting errors; the same can't be said of graduate students.

3. After crystal preparation, irradiate the crystal with X-rays. Electrons in the protein crystals scatter the X-rays to form a diffraction pattern. This is detected on X-ray film as a pattern of dots of varying intensity. Detailed analysis of this X-RAY DIFFRACTION PATTERN yields a 3-dimensional structure. Without going into detail on the math, here is the general approach to converting diffraction data into a three dimensional structure. First, you enter the amino acid sequence of the protein into the computer; this provides the source of all of the

electron density needed to solve the structure. Next, you provide some starting point for the structure solution. If the structure of a similar protein has been determined, you can use this three dimensional structure as a starting point. These structures are recorded in a depository known as the PDB (Protein Data Bank) data file, where the position of each atom in three dimensional space is recorded. If you don't have a three dimensional structure in hand, finding a starting point is more difficult. One often derivatizes the protein with heavy atom reagents (such as mercury or lead) that will react with cys amino acid side chains or his amino acid side chains. The positions of the very electron rich lead or mercury will be obvious on the diffraction pattern, and by knowing the reactivity of the derivatizing reagent, you know the identity of the amino acid adjacent to the heavy metal. After acquiring data from several derivatives, you can usually get sufficient information to generate an initial structure.

After an initial structure is determined, one uses computer simulation to take the initial structure and determine the diffraction pattern that this structure would generate. You then compare the calculated diffraction pattern with the actual diffraction pattern, and determine statistically the goodness of the fit. You then tweak the structure, again using known computer algorithms, and continually monitor the fit. Once you get to the point where there is a good fit between the experimental and calculated structure, you're finished.

The determination of a three dimensional structure by X-ray crystallography is a long and tedious process - problems include

- Growing a good crystal - some proteins, especially glycoproteins, are resistant to crystallization. This has been largely overcome by the use of crystallization kits and crystallization robots. The kits contain a series of solutions that are used to test a wide variety of crystallization conditions. All you need is a large amount of protein.

- Image reconstruction - this historically has taken a great deal of time. This part has been immensely helped by increased computing speed and new minimization technology. Image reconstruction is largely done now on either Linux or Unix machines. Image viewing is done in three dimensions, with newer computers allowing for the viewing of proteins in three dimensions.

Here is the data provided by x-ray analysis. The blue blobs are the information provided by the x-ray analysis – these are the nuclei in the molecule and their connectivities. To do the minimization, you need to know the amino acid sequence of the protein. You do a fit of the amino acid sequence to the electron density. You then determine the mathematical "goodness of fit" of the structure to the electron density, and make some revisions in the protein and check to see if the fit has improved. You continue to iterate until the fit reaches an acceptable level.