CMB311 Fall 2017, Fri Sep 22

Lecture 8 Notes

Protein Purification, Protein Sequencing and Protein Structure Determination II

In this lecture we discussed protein structure determination. There are actually two aspects to protein structure. One is the protein's amino acid sequence, otherwise known as primary (1°) structure. This is necessary before trying to interpret the experimentally determined 3D structure. Here we discuss both protein sequencing and 3D structure determination.

Protein Sequencing

There are several aspects to protein sequencing:

- Total hydrolysis of the protein by acid, followed by HPLC (high-performance liquid chromatography- just a method for working with small amounts of small molecules). This identifies the amino acid composition of the protein but does not provide any sequence information.
- Cleave at specific sites, determine sequence of individual smaller polypeptides
- Order smaller peptides

Peptide Sequencing by Edman Degradation (This is pretty important)

- Phenylisothiocyanate (PITC) reacts with the N-terminus of a peptide under mildly alkaline conditions to form a phenylthiocarbamoyl substitution. The derivative is cyclized with TFA (trifluoroacetic acid), to release the N-terminal amino acid residue as a thiazolinone derivative. Organic extraction and treatment with aqueous acid yield the N-terminal amino acid as a phenylthiohydantoin (PTH) derivative.
- The process is repeated with the remainder of the peptide chain to determine the N-terminus exposed at each stage until the entire peptide is sequenced.
- *Just know that this method leads to the sequential removal of the N-terminal amino acid of a peptide.*
- This method is limited in the number of amino acids that can be sequenced (usually 20-40), due to the accumulation of partially reacted products. So, what to do?

Protein Cleavage (This is pretty important)

Protein cleaved at specific sites by:

Enzymes:

- **Trypsin** cuts after (+) charged side chains (Arg and Lys)
- **Chymotrypsin** cuts after aromatic residues (Phe, Tyr, Trp)

Chemical reagent:

- **Cyanogen Bromide** (CNBr); cuts after internal Met

After a series of cleavage products is sequenced, the overlapping polypeptide sequences can be aligned, allowing the complete protein sequence to be determined. Given a series of peptide sequences, you

should be able to align them and determine the whole protein sequence. Alternatively, given the protein sequence, you should be able to predict the peptides generated by CNBr, trypsin and chymotrypsin (not for a multiple choice exam, but for problem sets). Here are a couple of examples:

1. A protein is cleaved in three separate experiments with CNBr, Trypsin or Chymotrypsin, yielding the following peptides whose sequence was determined by Edman Degradation:

CNBr:

ELVISFSTREM
MEYWDEM
LIYLEG
KGAGM

Trypsin:

GAGMELVISFSTR
MEYWDEMK
EMLIYLEG

Chymotrypsin:

DEMKGAGMELVISF
STREMLIY
LEG
MEYW

How do you solve this problem? It's actually a lot easier than it looks. First of all, note that for each cleavage reaction, all the peptides but one ends with the amino acid that is recognized by the cleavage agent. For instance, three of the peptides generated with Chymotrypsin end with aromatic amino acids, but one, LEG, does not. Note also that LEG is at the end of peptides in all three reactions (LIYLEG in the CNBr reaction and EMLIYLEG in the Trypsin reaction). This tells you right away that this is the C-terminal peptide as it has an end that was not generated by the cleavage. Second, note that many of the peptides from different reactions have overlapping sequences (for instance STREMLIY from the Chymotrypsin reaction overlaps two peptides in the Trypsin reaction, GAGMELVISFSTR and EMLIYLEG, indicating that these two peptides are linked in the protein). By finding the overlaps, it's quite easy to establish the protein sequence as:

MEYWDEMKGAGMELVISFSTREMLIYLEG

You should check that this protein sequence can be made from each of the collection of peptides.

2. A protein is cleaved with Trypsin or Chymotrypsin and yields the following products. What is the protein sequence?

Trypsin:

VGAHAGEYGAEATE

AAWGK

VLSPAK

TNVK

Chymotrypsin:

GAEATE

GKVGAHAGEY

VLSPAKTNVKAAW

Summary

- Proteins sequenced by a combination of methods
- Digestion with enzymes or chemicals produces short overlapping peptides
- Individual peptides sequenced using Phenylisothiocyanate (Edman Degradation)
- Complete protein sequence determined by aligning overlapping peptide sequences

Protein Crystallization and Structure Determination

Although our focus thus far has been on protein structure, these same methods are used to determine the structures of nucleic acids, protein-nucleic acid complexes, and so on. We covered two approaches; X-ray crystallography and cryo-electron microscopy (cryo-EM).

- Like non-biological molecules, proteins can be crystallized.
- As with ordinary organic compounds, crystallization was originally used as a method of protein purification.
- It was later realized that crystals of proteins, like crystals of small molecules, could be used for structure determination by X-ray diffraction.

Protein molecules are arranged in a repeating array within a crystal, as with crystals of NaCl or anything else. Note that the repeating unit may consist of more than one protein molecule, and in some instance may contain many. Crystals of proteins diffract X-rays, usually obtained at synchrotron radiation sources, like Oak Ridge National Laboratory. The diffraction pattern is mathematically related to the positions of atoms in the crystal, allowing their coordinates in three-dimensional space to be determined. In reality, an electron density map is generated, and the protein must be modeled into the electron density to create the protein structure model. What is one way to validate a protein structure model?

Summary

- High energy X-rays are diffracted by atoms in protein crystals