

## CHECKLIST OF KEY CONCEPTS

### Introductory concepts

1. There have been dramatic changes in the tools available for investigating proteins recently.
2. Mass spectrometry is now applicable to proteins.
3. High speed DNA sequencing allows the determination of protein sequences since the genetic code is known.
4. Data bases and bioinformatics techniques allow the large quantities of information to be accessed, searched and manipulated.

### Purification of proteins

5. The first step in the purification of proteins usually involves isolation of the cell fraction in which the protein occurs.
6. A method for estimating the amount of the protein in question is required. This normally involves estimation of its activity.
7. Electrophoresis can be used if only very small amounts of a protein are needed eg. for mass spectrometry.
8. There is a range of techniques where larger amounts of protein are necessary. The first step often involves selective precipitation.

### Column chromatography

9. The material used to separate proteins is packed into a column and the protein mixture flows through it.
10. Gel filtration separates molecules on the basis of size (Fig. 3.22). Molecules that are too large to enter pores in the matrix pass through whereas smaller molecules enter the beads and are slowed.
11. The effluent is collected using a fraction collector so that differently sized molecules are collected into different tubes.
12. Ion exchange chromatography separates molecules on the basis of their relative affinity for a charged polymer.
13. Affinity chromatography involves a specific interaction between a compound covalently attached to the matrix and the protein in question.
14. Reverse phase chromatography uses a matrix with hydrophobic groups and elution with increasing hydrophobicity or changing ionic strength.
15. The speed of separations using column chromatography can be increased by using high pressures (HPLC).

### SDS polyacrylamide gel electrophoresis

16. This is used primarily as an analytical tool.
17. The gel normally contains a detergent such as SDS that inserts itself into the proteins via its hydrophobic tail. The large charge from the SDS swamps the charge on the protein.
18. Proteins then move and are separated by molecular sieving with small proteins moving more rapidly than large ones.

19. Dyes such as Coomassie blue are used to visualise the proteins after separation.
20. Marker proteins of known size are used to calibrate the gel.
21. A typical apparatus is shown in Fig. 5.2 and a typical result in Fig. 5.3.
22. Gels where the protein has not been denatured in this way are also used. Separation depends partly on charge and partly on size.
23. In isoelectric focussing a pH gradient is established using polymers called ampholines. Proteins migrate to the pH at which they have zero charge (isoelectric point).
24. Two dimensional gel electrophoresis separates proteins by isoelectric focusing in the first dimension and SDS electrophoresis in the second. It can separate many proteins in a single run.

### **Immunological detection of proteins**

25. Specific antibodies can be prepared against a given protein and can detect very small amounts of that protein.
26. In Western blotting proteins that have been separated on a gel are transferred to plastic and the antibody bound. The antibody can be labeled with either radioactivity or fluorescence.

### **Methods of protein sequencing**

27. Automated sequencing of proteins involves labelling the N-terminal amino acid with phenyl isothiocyanate then detaching the labelled derivative. The next amino acid is then treated in the same way and so on.
28. The protein is linked to a column by its C terminal end.
29. Each amino acid is identified by use of a calibrated separating column.
30. This technique becomes unreliable after about 30 such steps but will work with 0.01  $\mu\text{g}$  of protein.
31. Larger proteins are sequenced by degradation into specific fragments by two different methods then identifying overlaps to reconstruct the original sequence.
32. Fragments can be used to generate sequence information, especially for smaller peptides. Generally large proteins are treated with proteases or cyanogens bromide to produce fragments small enough for analysis in the mass spectrometer.

### **The role of data bases in protein sequencing**

33. If a partial sequence is available it can be matched to a data base and the protein often can be identified.

### **Deduction of amino acid sequences of proteins from the base sequence of genes**

34. Given the speed with which DNA can be sequenced it is usually easier to deduce the amino acid sequence for a protein from its gene than to determine the sequence directly.

35. The complete genome has been sequenced for a number of species, including humans, and if the location of a gene can be found then the protein sequence can be determined.

### **Determination of the three-dimensional structure of proteins**

36. The primary sequence does not yet provide information about how a protein is folded.  
37. This information is required for almost all investigations of the molecular processes that involve proteins.

### **X-ray diffraction**

38. In this process protein crystals are bombarded with X-rays and the resulting diffraction pattern examined. By rotating the crystal a series of patterns are obtained.  
39. This allows the localization of atoms within the crystal.  
40. The analysis is aided if heavy metals atom can be placed in the crystal at a small number of defined places without changing the structure.

### **Nuclear magnetic resonance (NMR)**

41. Protons have a spin and this can be used to generate a signal. The signal is affected by the environment of the nucleus where the proton is located.  
42. In particular it allows an investigator to find atoms that are close to each other even though they are not close in the primary structure.  
43. NMR can be used on proteins in solution.

### **Analysis of proteins by mass spectrometry**

44. This technique requires very small amounts of protein and provides an accurate estimate of molecular weight.  
45. This information can allow identification of a polypeptide via a data base.  
46. Information of protein sequences and post-translational modifications can also be obtained.  
47. Mass spectrometers consist of an ion source, a mass analyzer and an ion detector.

### **Ionisation methods for protein and peptide mass spectrometry**

48. In matrix assisted laser desorption of ions (MALDI) the protein or peptide is mixed with UV absorbing material on a solid surface. A pulse of UV light then causes rapid desorption and the production of positively charged ions.  
49. In electrospray ionization (ESI) the protein is introduced into the apparatus at a high potential in a fine spray. The droplets lose the solvent and enter the high vacuum system as ions.

### **Types of mass analysers**

50. These include quadrupole, time of flight (TOF) and ion-trap.

51. Apparatus using combinations of ionization methods and mass analysers is available eg. ESI/quadrupole, MALDI/TOF.

### **Identification of proteins using mass spectrometry for peptide mass analysis and database searching**

52. If a protein has been identified previously and entered into a database then it can be identified rapidly.
53. A small amount of the protein is digested with trypsin and the mixture analysed using a MALDI mass spectrometer. The pattern of the products then allows identification.

### **Identification of proteins by limited sequencing and database searching**

54. If the peak generated by a peptide in a mass spectrometer is selected and introduced into a second phase mass spectrometer it will fragment and the fragment pattern can be used to find the sequence of the peptide.

### **Sequencing a protein by mass spectrometry**

55. A new protein can be fully sequenced by determination of the digestion products using two different cleavage techniques (eg. two different enzymes) and sequencing each peptide. Overlaps allow total sequence elucidation.

### **Molecular weight determination of proteins**

56. Mass spectrometry achieves this with a high degree of accuracy in around 1 minute.

### **Analysis of post-translational modification of proteins**

57. A variety of techniques are available, including analysis before and after enzymic removal of the modifying group.

### **Proteomics and mass spectrometry**

58. The proteome is the total protein complement of a cell or organism. While the genome is constant within a given organism the proteome varies from cell to cell.
59. Two dimensional electrophoresis combined with mass spectrometry is widely used to investigate the proteome.
60. It is also possible to determine the proteome of individual protein complexes in the cell if these can be separated.

### **Bioinformatics and databases**

61. The efficient storage of data and its rapid retrieval is critical in bioinformatics.
62. A range of publicly accessible databases exist.
63. Search programmes usually seek homologies.

64. Information obtained is used to investigate evolutionary relationships and the function of domains.