The Use of Circular Dichroism in the Investigation of Protein Structure and Function

Sharon M. Kelly* and Nicholas C. Price

Division of Biochemistry and Molecular Biology, Institute of Biomedical and Life Sciences, University of Glasgow, G12 8QQ, Scotland, UK

Abstract: Circular Dichroism (CD) relies on the differential absorption of left and right circularly polarised radiation by chromophores which either possess intrinsic chirality or are placed in chiral environments. Proteins possess a number of chromophores which can give rise to CD signals. In the far UV region (240-180 nm), which corresponds to peptide bond absorption, the CD spectrum can be analysed to give the content of regular secondary structural features such as α-helix and β-sheet. The CD spectrum in the near UV region (320-260 nm) reflects the environments of the aromatic amino acid side chains and thus gives information about the tertiary structure of the protein. Other non-protein chromophores such as flavin and haem moieties can give rise to CD signals which depend on the precise environment of the chromophore concerned. Because of its relatively modest resource demands, CD has been used extensively to give useful information about protein structure, the extent and rate of structural changes and ligand binding. In the protein design field, CD is used to assess the structure and stability of the designed protein fragments. Studies of protein folding make extensive use of CD to examine the folding pathway; the technique has been especially important in characterising "molten globule" intermediates which may be involved in the folding process. CD is an extremely useful technique for assessing the structural integrity of membrane proteins during extraction and characterisation procedures. The interactions between chromophores can give rise to characteristic CD signals. This is well illustrated by the case of the light harvesting complex from photosynthetic bacteria, where the CD spectra can be analysed to indicate the extent of orbital overlap between the rings of bacteriochlorophyll molecules. It is therefore evident that CD is a versatile technique in structural biology, with an increasingly wide range of applications.

1. INTRODUCTION: BASIS OF CD TECHNIQUE

CD refers to the differential absorption of the left and right circularly polarised components of plane-polarised radiation. This effect will occur when a chromophore is chiral (optically active) either (a) intrinsically by reason of its structure, or (b) by being covalently linked to a chiral centre, or (c) by being placed in an asymmetric environment. In practice the plane-polarised radiation is split into its two circularly polarised components by passage through a *modulator* subjected to an alternating (50 kHz) electric field. The modulator usually consists of a piezoelectric quartz crystal

and a thin plate of isotropic material (e.g. quartz) tightly coupled to the crystal. The alternating electric field induces structural changes in the quartz crystal which make the plate transmit circularly polarised light at the extremes of the field. If, after passage through the sample, the left and right circularly polarised components are not absorbed (or are absorbed to the same extent), combination of the components would regenerate radiation polarised in the original plane. However, if one of the components is absorbed by the sample to a greater extent than the other, the resultant radiation (combined components) would now be *elliptically polarised*, i.e. the resultant would trace out an ellipse (Fig. **1**). In practice, the CD instrument (spectropolarimeter) does not recombine the components but detects the two components separately; it will then display the dichroism at a given wavelength of radiation

---

*Address correspondence to this author at Division of Biochemistry and Molecular Biology, Institute of Biomedical and Life Sciences, Room B2-35, Joseph Black Building, University of Glasgow, G12 8QQ, Scotland, UK, Tel./Fax: (+44)-141-330-6447; e-mail: Sharon.Kelly@bio.gla.ac.uk
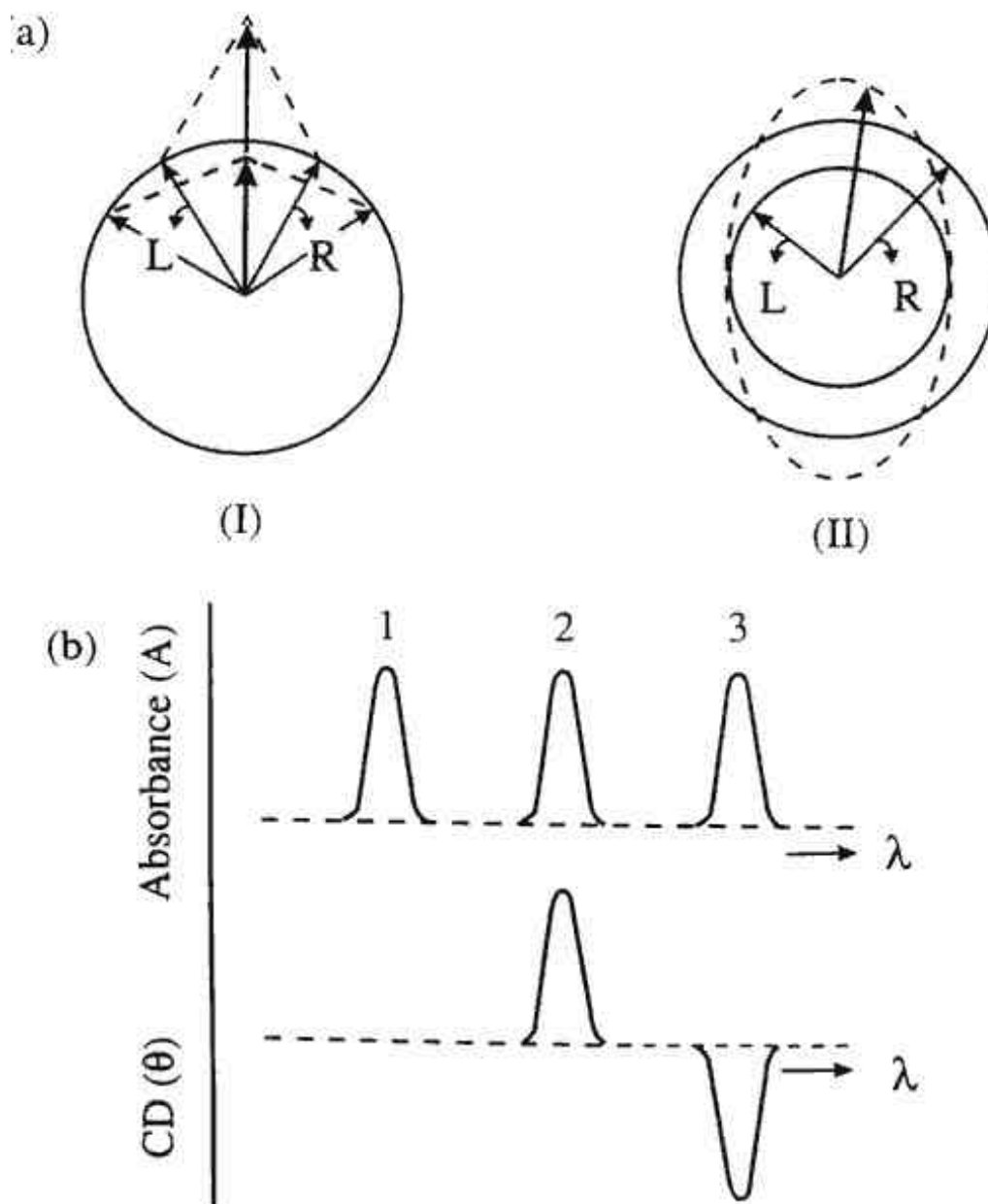
Fig. (1). Origin of the CD effect. (a) The left (L) and right (R) circularly polarised components of plane polarised radiation: (I) the two components have the same amplitude and when combined generate plane polarised radiation; (II) the components are of different magnitude and the resultant (dashed line) is elliptically polarised. (b) The relationship between absorption and CD spectra. Band 1 is not chiral; band 2 has a positive CD spectrum with L absorbed more than R; band 3 has a negative CD spectrum.

expressed as either the *difference in absorbance* of the two components ($\Delta A = A_L - A_R$) or as the *ellipticity* in degrees (θ) ($\theta = \tan^{-1}(b/a)$, where b and a are the minor and major axes of the resultant ellipse). There is a simple numerical relationship between ΔA and θ (θ in degrees), i.e. $\theta = 32.98 \Delta A$. A CD spectrum is obtained when the dichroism is measured as a function of wavelength (Fig. **1**).

If two or more identical chromophores are in close proximity, the interactions between them can give rise to a sigmoidal CD spectrum. The exciton coupling model [1] can be used to account for this behaviour. In proteins, such coupling can occur from close pairs of Trp side chains (see Section 3.2) or from more extended sets of chromophores. The latter situation is seen in the case of the ring of 18 B850 bacteriochlorophyll molecules in the LH2

light harvesting complex from photosynthetic bacteria (see Section 4.7).

In practice in most biological work, the observed ellipticities are of the order of 10 millidegrees (i.e. the difference in absorbance between the two circularly polarised components of the incident radiation is of the order of 3×10^{-4} absorbance units). It is therefore clear that careful attention must be paid to the experimental conditions to ensure that meaningful data are obtained (see Section 2).

The aim of this review article is to show how CD measurements (especially when used in combination with other structural techniques) can give valuable insights into several aspects of protein structure and function. Some earlier reviews of CD have covered general aspects of its application to biological systems [2], protein-ligand interactions and conformational changes [3], and protein folding and unfolding [4]. In addition, a comprehensive multi-author treatise on CD has been published [5].

2. EXPERIMENTAL ASPECTS OF CD

This section will give a brief outline of some of the important experimental aspects of obtaining CD data; this is important in appreciating the conditions under which reliable data can be obtained. Further details of these aspects can be found in a number of review articles [6-8].

2.1 CD Instrumentation

The light source for most conventional CD instruments is a xenon arc, which gives good output over the range of wavelengths (178 to 1000 nm) used for virtually all studies on proteins. It is necessary to flush the instrument with N_2 gas in order to remove O_2 from the lamp housing and the sample compartment so as (a) to prevent ozone formation and to minimise damage to the optical system, and (b) to allow measurements to be made below 200 nm. In a N_2 atmosphere, the lower wavelength limit is below 180 nm, and modern instruments using efficient N_2 purging will allow measurements to be made down to at least 175 nm. In order to record data below this wavelength, vacuum UV techniques or the use of high intensity synchrotron radiation sources [9,10] are necessary.

To obtain reliable CD data, it is important to pay attention to the instrument and to the sample. As far as the instrument is concerned, regular maintenance and calibration with a suitable chiral standard such as 1S-(+)-10-camphorsulphonic acid is essential. In view of the relatively small size of the signal a number of experimental parameters are adjusted in order to improve the quality of the data. These include:- the time constant, the scan rate, the number of scans and the bandwidth. In most cases, suitable "steady state" CD spectra in the near and far UV can be obtained using a time constant of 2 sec, a scan rate of 10 nm/min, accumulation of two to four scans and a bandwidth of 2 nm or less [7]. As a rule of thumb, the product of the time constant and the scan speed should be less than 0.5 nm in order to avoid distortion of the spectrum [7]. The acquisition of CD data using stopped flow mixing techniques and using synchrotron radiation are discussed in sections 2.5 and 2.6 respectively.

2.2 Sample Preparation for CD

It is important that the sample of protein should be homogeneous and should be freed of highly scattering particles by either centrifugation or passage through a suitable filter (e.g. 0.2 μm). The total absorbance of the sample at the wavelength in question should not exceed about one unit, otherwise the spectral noise will become excessive; above a certain point, an automatic cutoff may operate leading to an apparent decline of the CD signal to zero. It is essential to minimise absorption due to other components (buffers, supporting electrolytes, solvents etc.) in the mixture. Most problems arise in the far UV and the absorption characteristics of many commonly used buffers and other reagents have been extensively studied [7]. Phosphate, borate and low molarity (20 mM) Tris have low absorbances above 190 nm in cells of pathlength 0.1 cm or less; these buffers can between them give suitable coverage of pH values from 6 to about 9.5. Buffers which are appropriate for pH values from 4 to 6 usually contain carboxylate groups and thus have relatively high absorbances below 200 nm. In such cases it is important to work with as dilute a buffer solution as is required to maintain the pH, and to run appropriate "blank" CD spectra to ensure that the buffer components do not lead to excessive noise or other artefacts in the spectra.

One point often overlooked is that chloride ions absorb strongly below 195 nm; thus high concentrations of chloride salts are not recommended as a component of a buffer solution for CD work. Fluoride and sulphate salts are much more appropriate to maintain the ionic strength, if necessary.

In order to be able to estimate the secondary structure content of a protein from CD it is essential to know the protein concentration accurately (to within at least $\pm5\%$, and preferably better). The most reliable methods are those which can be calibrated against absolute values such as A_{280} measurements calibrated by reference to amino acid analysis data [11].

2.3 Units for CD Measurement

For far UV CD, where the dichroism of the peptide bond is being measured, it is the molar concentration of these bonds which is required; this can be obtained by dividing the concentration in mass terms (i.e. mg/ml) by the mean residue weight (strictly speaking, mass) or MRW. The MRW is obtained by dividing the molecular mass by N - 1, where N is the number of amino acids in the polypeptide chain. For most proteins the MRW is about 110, but it can deviate significantly from this value if the amino acid composition is unusually skewed or if the protein undergoes extensive posttranslational modifications.

When CD data are expressed in terms of absorbance, the units are those of the *difference in molar absorbance* $\Delta\varepsilon = \varepsilon_L - \varepsilon_R$, i.e. $cm^{-1}.M^{-1}$.

In terms of ellipticity, the *mean residue ellipticity* $[\theta]_{mrw,\lambda}$ at a wavelength λ is quoted in units of $deg.cm^2.dmol^{-1}$ and is given by:-

$$[\theta]_{mrw,\lambda} = MRW\theta/10d.c$$

where θ is the observed ellipticity (degrees), d is the pathlength (cm) and c is the concentration (in units of g/ml).

Reservations have been expressed as to whether mean residue ellipticity is an appropriate unit for expressing the amplitude of the near UV and visible CD signals of a protein, since only a small number of aromatic amino acids or prosthetic groups contribute to the CD signals in these regions. In such cases, it may be appropriate to express the data in terms of the molar concentration of intact protein, rather than of the repeating unit. At any wavelength, the numerical relationship between values in the two sets of units is:-

$$[\theta]_{mrw} = 3298\Delta\varepsilon$$

2.4 Amount of Sample Required for CD

The amount of protein required for CD measurements can be gauged from the need to keep the absorbance less than about one unit. Typical cell pathlengths for far UV CD work are in the range 0.01 to 0.05 cm and protein concentrations are in the range 0.2 to 1 mg/ml. Depending on the design of the cell being used, the volume of sample required can range from about 1 ml to as little as 50 μl. Thus the amount of protein required to obtain an acceptable far UV CD spectrum can be as little as 10 μg, but generally 100 to 500 μg sample is required in order to explore the optimal conditions for recording spectra. The CD signals in the near UV and visible regions of the spectrum are much weaker than those in the far UV, reflecting the much lower molar concentrations of the relevant chromophores compared with those of the peptide bonds. For measurements in these spectral regions it would be fairly typical to use a protein concentration of 0.5 to 2 mg/ml and a pathlength of 0.5 to 2 cm. The amounts of protein required for such experiments are thus of the order of several mg.

It should be noted that very useful information can be obtained by recording CD spectra at low temperatures (typically down to that of liquid N_2, i.e. 77 K) by use of mixed solvent systems or water/glycerol "glasses". The resolution can be considerably enhanced at low temperatures [4,12] and it may also be possible to examine the structures of individual "frozen" conformational states which would rapidly interconvert at higher temperatures.

2.5 Stopped Flow CD

In order to examine early events during a reaction, it is necessary to employ rapid mixing and detection system, such as that provided by a stopped-flow apparatus. The first applications of

stopped flow CD were reported in the late 1970s and since that time a number of such devices have become available commercially, either as stand alone instruments or as accessories to conventional CD spectropolarimeters [13]. The dead times of such devices are generally around 5 msec, though this has been reduced to less than 1 msec by recent design improvements [10]. In order to detect changes which occur on the stopped flow time scale (tens to hundreds of msec), it is necessary to employ low time constants (typically of the order of 1 msec), and hence to accumulate a substantial number of individual measurements (often up to 100 or more) to reduce the noise to an acceptable level [14].

Stopped flow CD has been used extensively to study the early events in protein folding; typically a solution of denatured protein in 6 M GdmCl is diluted with 10 or more volumes of buffer using appropriate syringes. This will reduce the final concentration of the denaturant to a low level (0.55 M or less), and for most proteins promote refolding to a state very similar to the native state. The changes in ellipticity in the far UV (225 nm) and near UV (290 nm) give information about the regain of secondary and tertiary structure respectively, providing valuable information about the mechanism of protein folding. This is discussed further in Section 4.3.

2.6 CD Using Synchrotron Radiation

Synchrotron radiation is produced when charged particles such as electrons moving at velocities close to that of light are accelerated through magnetic fields. Intense radiation is produced over a wide range of wavelengths from X-rays to the IR and this has afforded structural biologists new opportunities. Thus the greatly enhanced intensity of X-rays produced by the synchrotron compared with conventional sources has revolutionised X-ray crystallography allowing much smaller crystals to be examined in only a fraction of the time previously required. As far as CD is concerned the key advantage of the synchrotron is the very large increase (over 1000-fold) in intensity in the far UV region (below 200 nm) compared with the xenon arc sources in conventional CD instruments [9]. The application of synchrotron CD to a study of clathrin coat proteins [10] is described in Section 4.1.

2.7 Strengths and Weaknesses of CD as a Structural Technique

The purpose of this section is to evaluate CD as a structural technique, principally by comparison with X-ray crystallography and high resolution NMR.

X-ray crystallography [15,16] requires crystals which diffract to sufficiently high resolution, and the availability of one or more isomorphous heavy atom derivatives; either of these requirements can take several weeks if not months. The X-ray structure of a protein, although providing detail at the atomic level, has been generally thought of as providing essentially a static picture with only limited insights into the dynamic aspects of protein structure, which can be crucial to function. However, recent developments in synchrotron sources and detector technology are making time-resolved crystallography a practical proposition [15].

The study of proteins by NMR requires high concentrations of sample (0.5 - 2 mM; i.e. 10 - 40 mg/ml for a 20 kDa protein) and a number of proteins are either not sufficiently soluble under these conditions or show a tendency to aggregate. For a complete determination of the structure, the individual resonances must be assigned and this requires multidimensional NMR techniques with (expensive!) stable isotope (^{13}C, ^{15}N) labelling [17]. At present, full structural determinations are limited to proteins of molecular mass up to about 30 kDa, but it should be emphasised that the NMR approach is capable of giving a wealth of information on both the structure and dynamics of proteins.

The principal advantages of CD derive from the speed and convenience of the technique. Compared with X-ray crystallography and NMR, CD measurements can be carried out rapidly; thus good quality spectra in the far and near UV can each be obtained within 30 min. Far UV CD studies, in particular, require only small amounts of material (Section 2.4) and since the technique is non-destructive, it is usually possible to recover most or all of the solution and hence conduct multiple experiments on the same sample. The ability to use cells of pathlengths differing by several orders of magnitude means that it is relatively easy to study a very wide range of protein concentrations. (This latter point

represents an important advantage of CD over Fourier Transform Infra Red (FTIR) spectroscopy which can give useful information about the secondary structure of proteins, but under a more restricted range of experimental conditions [18]). The changes in the far UV CD spectrum of a 58-residue DNA-binding peptide derived from transcription factor GCN4 over a 1000-fold range of concentrations have been interpreted in terms of a dimeric species with a high α-helical content (70-73%) dissociating upon dilution into two unfolded polypeptide chains [19]. More recent work involving theoretical and CD studies on related coiled-coil systems has, however, been interpreted in terms of a model involving partially structured monomeric forms of the polypeptide [20]. By contrast, CD data have been used to show that a model helix-turn-helix 38 residue peptide did not undergo any concentration-dependent changes over a wide (40-fold) range of concentrations [21]. In addition, CD studies can be performed over a wide range of experimental conditions, including pH, temperature and (with suitable attention to the requirement to avoid excess absorption or scattering of the incident radiation) in both solution and the solid state.

Developments in instrumental design have now also made it possible to use stopped flow CD to study structural changes in proteins which occur on the timescale of a few tens of milliseconds [10,13,14] (see Section 2.5). The application of stopped flow CD measurements in both the far UV and the near UV has been used in conjunction with other techniques to define the structural changes occurring during some of the early steps in protein folding (see Section 4.3).

The main limitation of CD is that it only provides relatively low resolution structural information. Thus, although far UV CD can give reasonably reliable estimates of the secondary structure content of a protein (in terms of the proportions of α-helix, β-sheet and β-turns), it must be noted that these are overall figures and do not indicate which regions of the protein are of which structural type. The value of the information from CD can be enhanced if used in conjunction with secondary structure prediction algorithms [22], since these are capable of pinpointing particular regions of secondary structure along the sequence. At present, the development of CD theory is insufficient to allow detailed interpretation of the near UV CD spectra in structural terms [1,12,23] and thus by itself CD is unable to give detailed insights into the tertiary structure of a protein. However, even with these limitations, CD can still play a very useful role in structural work, especially when it is employed in a complementary fashion with other structural techniques. Apart from the studies of protein unfolding and folding (see Section 4.3), CD has been widely applied in the protein engineering field. The secondary and tertiary structural characteristics of a range of mutant proteins can be assessed very rapidly by CD; this would allow the rapid selection of those mutants of altered structure which can then be characterised by the more detailed structural techniques. If the far UV and near UV CD spectra of the mutants are superimposable on those of the wild type protein, it can be concluded that the mutations have not resulted in any significant distortion of structure. This was found to be the case in two mutant forms of isocitrate lyase from *Escherichia coli* in which Cys 195 had been replaced by alanine or serine, leading to loss of activity [24] (see Fig. **2**). The conclusion that Cys 195 is indeed essential for the catalytic activity of the enzyme is strengthened by these results. In the case of phosphoglycerate mutase from *Schizosaccharomyces pombe*, mutation of His 163 to Gln leads to essentially complete loss of activity. However the far UV and near UV CD spectra show that the overall structure of this mutant is affected by the mutation, making it difficult to draw any definitive conclusions regarding the importance of His 163 in the catalytic mechanism [25].

A further limitation of CD is that in general it gives little detailed information on the quaternary structure of proteins. In some cases, however, the formation of oligomers is accompanied by changes either in the secondary structure of the individual subunits (as in the case of the GCN4-derived peptide [19]) or in the environment of aromatic amino acid residues (as in the case of insulin where the association of subunits to form hexamers is accompanied by changes in the environment of the Tyr side chains [1]). In these cases, the changes in quaternary structure can be monitored by far UV or near UV CD measurements respectively. The more usual approach would be to employ techniques such as gel permeation, ultracentrifugation, radiation scattering, or chemical cross-linking in conjunction with CD
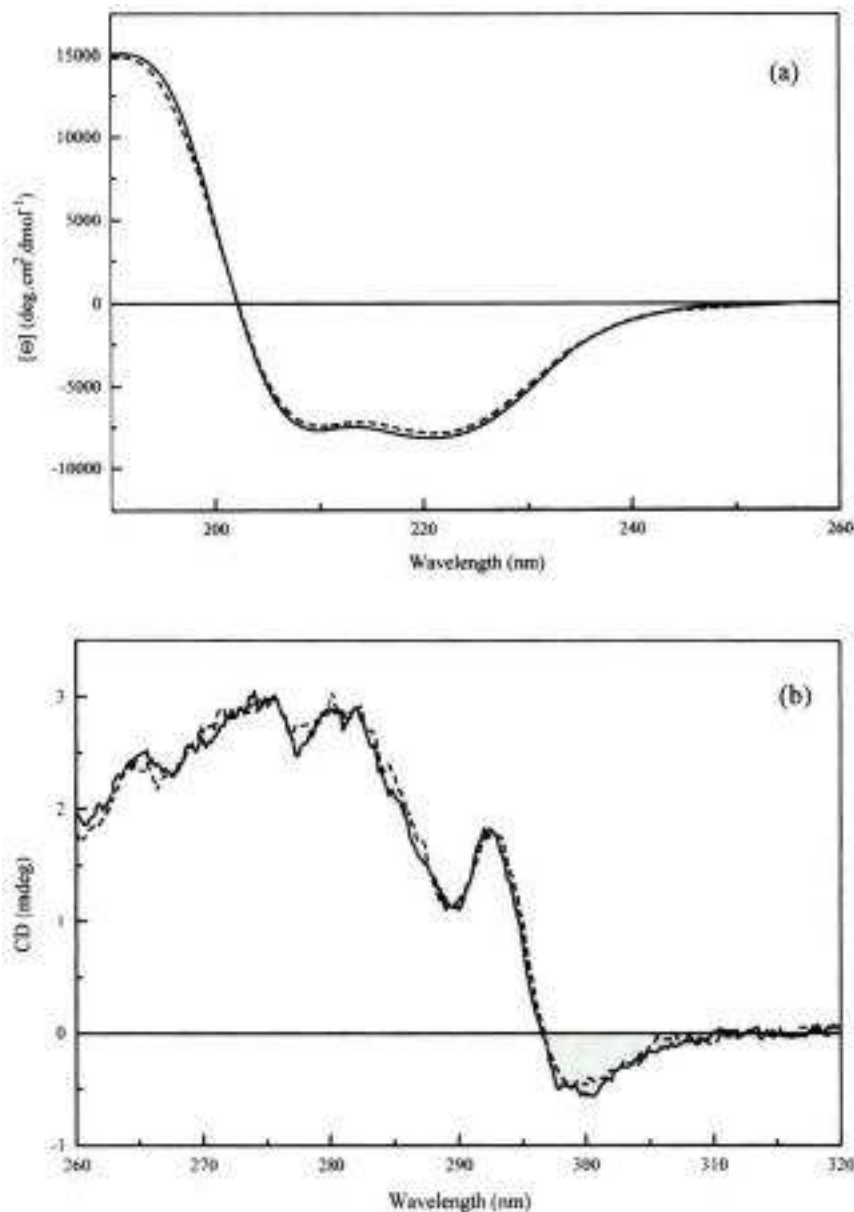
Fig. (2). CD spectra of isocitrate lyase from *E.coli*. The far UV (a) and near UV (b) spectra for wild type enzyme and a mutant in which Cys 195 has been replaced by Ala are shown as solid and dashed lines respectively. The spectra of the mutant in which Cys 195 has been replaced by Ser are superimposable on those shown [24].

measurements in order to explore the relationships between quaternary and secondary or tertiary structures [26].

3. CHROMOPHORES COMMONLY OBSERVED IN CD STUDIES OF PROTEINS

One of the strengths of CD is that various aspects of protein structure can be measured. In the far UV, the peptide bond is the principal

absorbing group; studies in this region can give information on the secondary structure. In the near UV, the aromatic amino acid side chains (phenylalanine, tyrosine and tryptophan) absorb in the range 250 to 290 nm. The tertiary folding of the polypeptide chain can place these side chains in chiral environments, thus giving rise to CD spectra which can serve as characteristic fingerprints of the native structure. In model compounds containing a disulphide bond, the dihedral angle of this bond can often be deduced

from the appearance of the CD spectrum in the 240 to 290 nm region. In the case of extensively disulphide bonded proteins, there can be difficulties in performing such analyses because of the overlap of the disulphide signals with others from the aromatic side chains [1,3,14]. In addition, non-protein components or cofactors, such as pyridoxal-5'-phosphate, haem, flavin, bacteriochlorophyll, or added ligands may absorb in regions of the spectrum well separated from those of amino acids and peptide bonds. The CD signals in these regions can be used to provide detailed information on the environment of, and possible interactions between, these cofactors or ligands. Each of these aspects will now be discussed in turn.
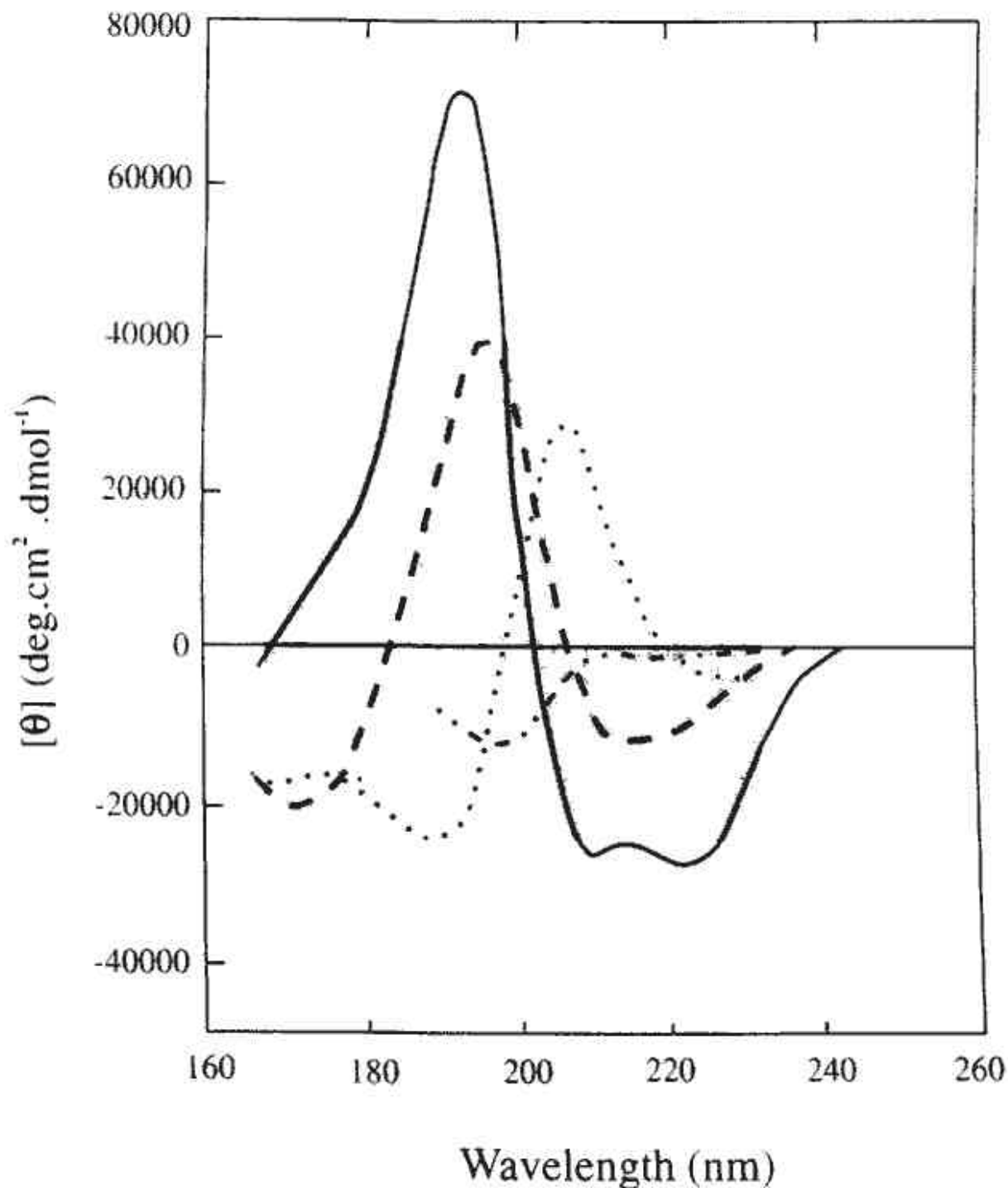
Fig. (3). Far UV CD spectra associated with various types of secondary structure: solid curve, α-helix; long dashes, anti-parallel β-sheet; dots, type I β-turn; dots and short dashes, irregular structure [6].

3.1 The Peptide Bond

In the far UV region (typically 240 nm to 190 or 180 nm, or even lower in the case of synchrotron CD (see section 2.6), the absorbing group is principally the peptide bond. There is a weak but broad $n{\rightarrow}\pi^*$ transition centred around 210 nm and an intense $\pi{\rightarrow}\pi^*$ transition about 190 nm [3,27]. Studies of far UV CD can be used to assess quantitatively the overall secondary structure content of the protein, since it has been known for many years that the different forms of regular secondary structure found in peptides and proteins exhibit distinct spectra (Fig. **3**). The application of CD in determining the secondary structures of peptides and proteins is discussed in section 4.1.

3.2 Aromatic Amino Acid Side Chains

The near UV CD of proteins arises from the environments of each aromatic amino acid side chain as well as possible contributions from disulphide bonds, or non-protein cofactors which might absorb in this spectral region [1,12,23]. Small model compounds of the aromatic amino acids exhibit CD spectra because the chromophore is linked to the nearby chiral α-carbon atom. In the case of proteins in their native states, the side chains of these amino acids will be placed in a variety of asymmetric environments characteristic of the tertiary structure of the folded protein. The analysis of near UV CD spectra has been discussed in detail [1,12,23]. Each of the aromatic amino acids tends to have a characteristic wavelength profile:- tryptophan, a peak close to 290 nm with fine structure between 290 and 305 nm; tyrosine, a peak between 275 and 282 nm (the fine structure at longer wavelengths may be obscured by that from tryptophan); phenylalanine, sharp fine structure between 255 and 270 nm. The fine structure in these bands arises from vibronic transitions in which different vibrational levels of the excited state are involved. A number of factors can influence the CD spectra of aromatic amino acids. Among these are:- the rigidity of the protein, with the more highly mobile side chains having lower intensities, and the nature of the environment in terms of hydrogen bonding, polar groups and polarisability. In addition the CD spectrum can be altered by interactions between aromatic amino acids which are especially significant if the distance between them is less than 1 nm. In the "exciton coupling" model, two excited states of chromophores in close proximity result from exciton mixing of the two strong transitions; the states correspond to symmetric and anti-symmetric combinations of the excited state wave functions. The two resulting CD bands will overlap with some cancellation (the extent of which depends on the size of the interaction), giving rise to a sigmoidal CD curve [1].

A final factor is the number of aromatic amino acids in a protein. Proteins with large numbers of such amino acids can have smaller CD bands than might be expected because of cancelling effects of positive and negative contributions, as in the cases of aspartate carbamoyltransferase [28] and carbonic anhydrase [29].

The contributions of individual aromatic amino acids to the near UV CD spectrum of a protein can be assessed by examining the CD spectra of suitable mutant proteins. This approach was adoted by Craig *et al.* [30] to show that of the four tyrosines and one tryptophan in interleukin 1-β, the major contributions were made by Trp 120 and Tyr 68. Freskgård *et al.* [29] used a similar approach to evaluate the contributions of each of the seven tryptophan residues in human carbonic anhydrase II, using a range of mutants in which each tryptophan had been replaced in turn. The fact that the mutant enzymes retained substantial activity (at least 34% of the wild-type) indicated that the structures were likely to be very similar to wild-type enzyme. Each tryptophan was found to make distinctive contributions to the near UV CD spectrum (Fig. **4**); when these were summed algebraically the resulting spectrum was qualitatively similar to that of the wild-type enzyme. In this case each of these side chains was found to make significant contributions not only to the near UV but also to the far UV CD (Fig. **4**), varying between +2500 and -1500 deg.cm^2.dmol^{-1} at different wavelengths. The contributions are particularly significant in the case of carbonic anhydrase where the ellipticity in the 220 to 230 nm region does not exceed -4000 deg.cm^2.dmol^{-1}. This example serves to emphasise the need to exercise care in the interpretation of far UV CD spectra of proteins of low helical content, particularly when mutations involving aromatic amino acids are involved. In view of the various factors involved, it is hardly surprising that near

UV CD spectra of proteins are not readily amenable to detailed interpretation in terms of tertiary structural features. Nevertheless, near UV CD spectra can be very useful "fingerprints" for comparisons of tertiary structures between related proteins, e.g. wild-type and mutant proteins (section 2.7), and have been invaluable in studies of the "molten globule" state of proteins (section 4.2).

3.3. The Disulphide Bond

The near UV absorption of the disulphide bond occurs near 260 nm and is generally quite weak. However, the intensity depends on a number of factors including the dihedral angle of the disulphide bond, the C-S-S bond angle and the effects of neighbouring groups. In addition the position of the absorption (and CD) bands depends on the dihedral angle. From theoretical and experimental studies of the CD spectra of model cystinyl compounds including small peptide hormones, it has been possible to formulate a "quadrant rule" relating the sign of the long wavelength band to the dihedral angle [1,12,14,31,32]. For angles less than 90°, a positive long-wavelength band indicates a right-handed sense of the disulphide and a negative band a left-handed sense. For angles greater than 90°, the assignments are reversed. The extent to which this rule may apply to larger proteins is not clear,
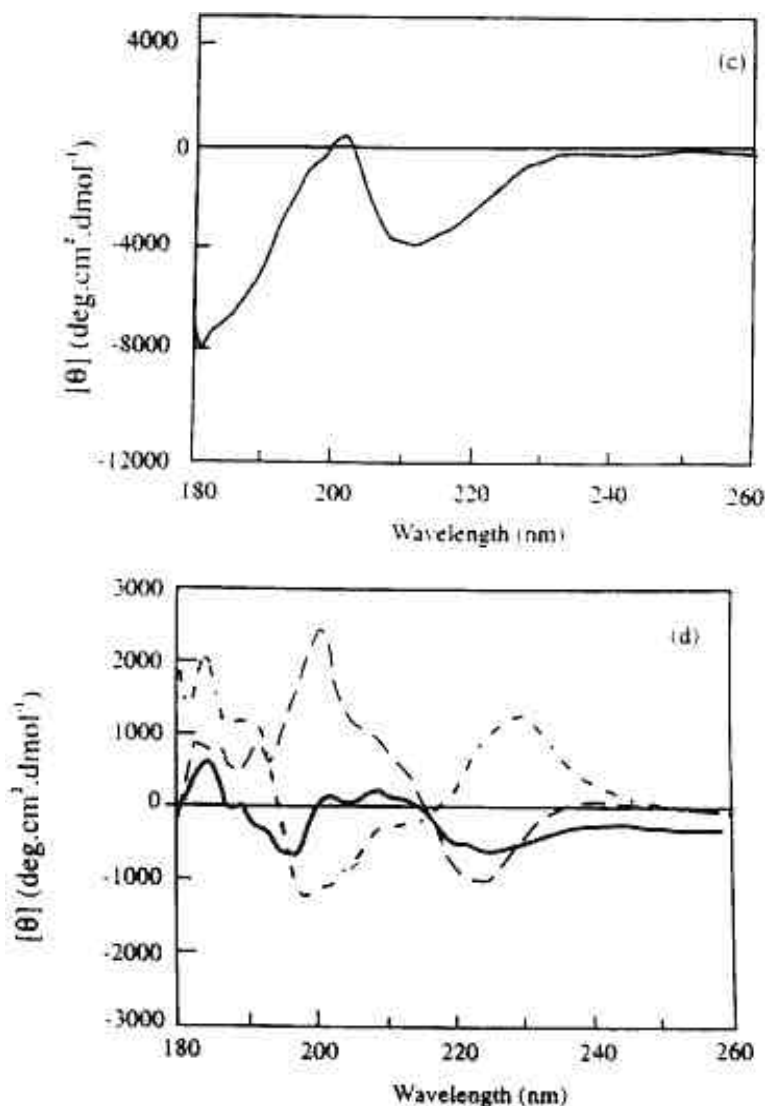

Wavelength (nm)

Wavelength (nm)
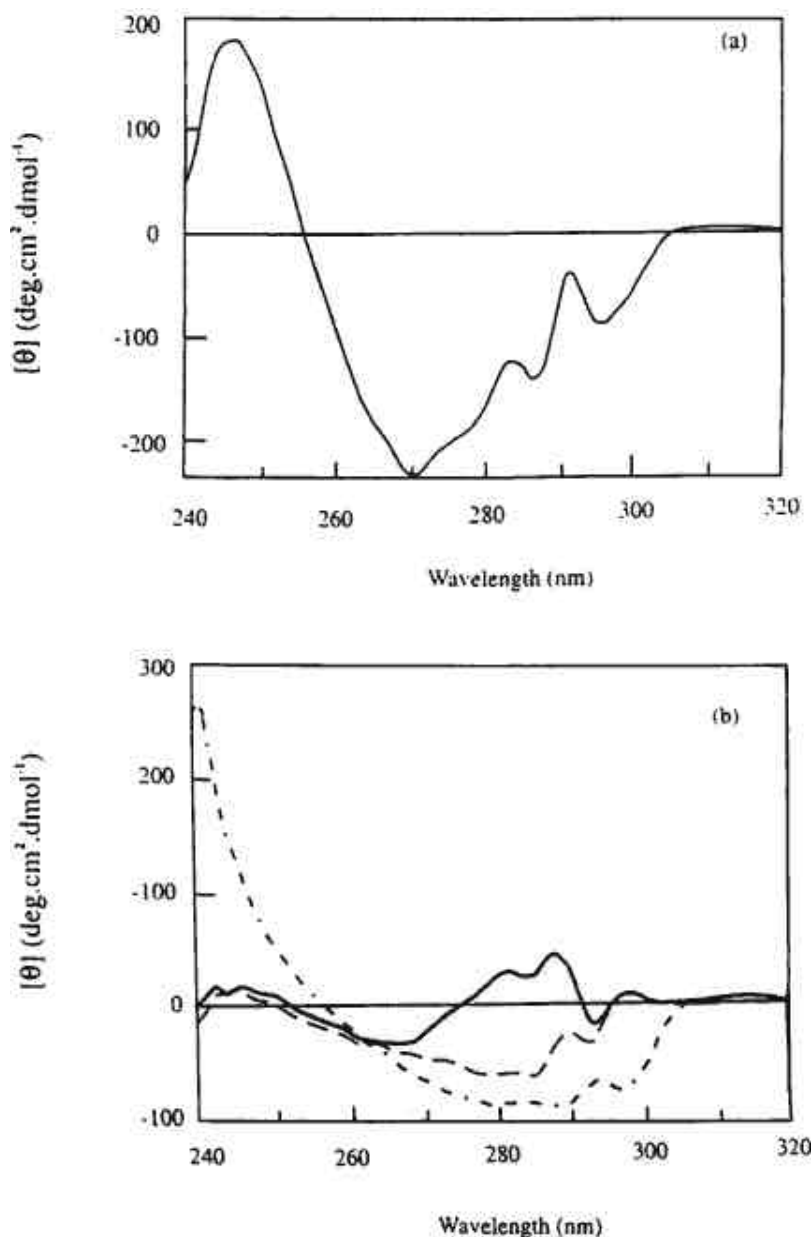
Fig. (4). CD spectra of carbonic anhydrase II. (a) and (c) Near UV and far UV CD spectra respectively of wild type enzyme. (b) and (d) Contributions of Trp 5 (solid line), Trp 16 (dashed line), and Trp 97 (dots and dashes) to the near UV and far UV CD spectra respectively, as deduced from the spectra of appropriate mutants in which the tryptophan had been replaced by another amino acid [29].

because the CD bands of disulphides are usually much broader than those of aromatic amino acids and it is difficult to evaluate the contribution of the former to the overall spectrum.

As described in section 4.3, the contribution made by disulphide bonds to the far UV CD of lysozyme may well account for the observed "overshoot" in the regain of ellipticity during the refolding of the enzyme [14].

3.4. Non-Protein Chromophores

CD in the near UV, visible and near IR can give a great deal of information on the environments of cofactors (which play an integral role in the biological activity of the protein) or of other non-covalently bound ligands. Typically, the free ligand or cofactor has little or no CD signal; the observed CD signals in the complex therefore indicate that the binding site of ligand or cofactor confers chirality. The loss of integrity of the binding site during unfolding of the protein can be conveniently monitored by changes in the CD signal. It should be noted that there is generally a much smaller change in the absorbance of the cofactor on dissociation from the protein. Some representative examples of cofactors and ligands are discussed below.
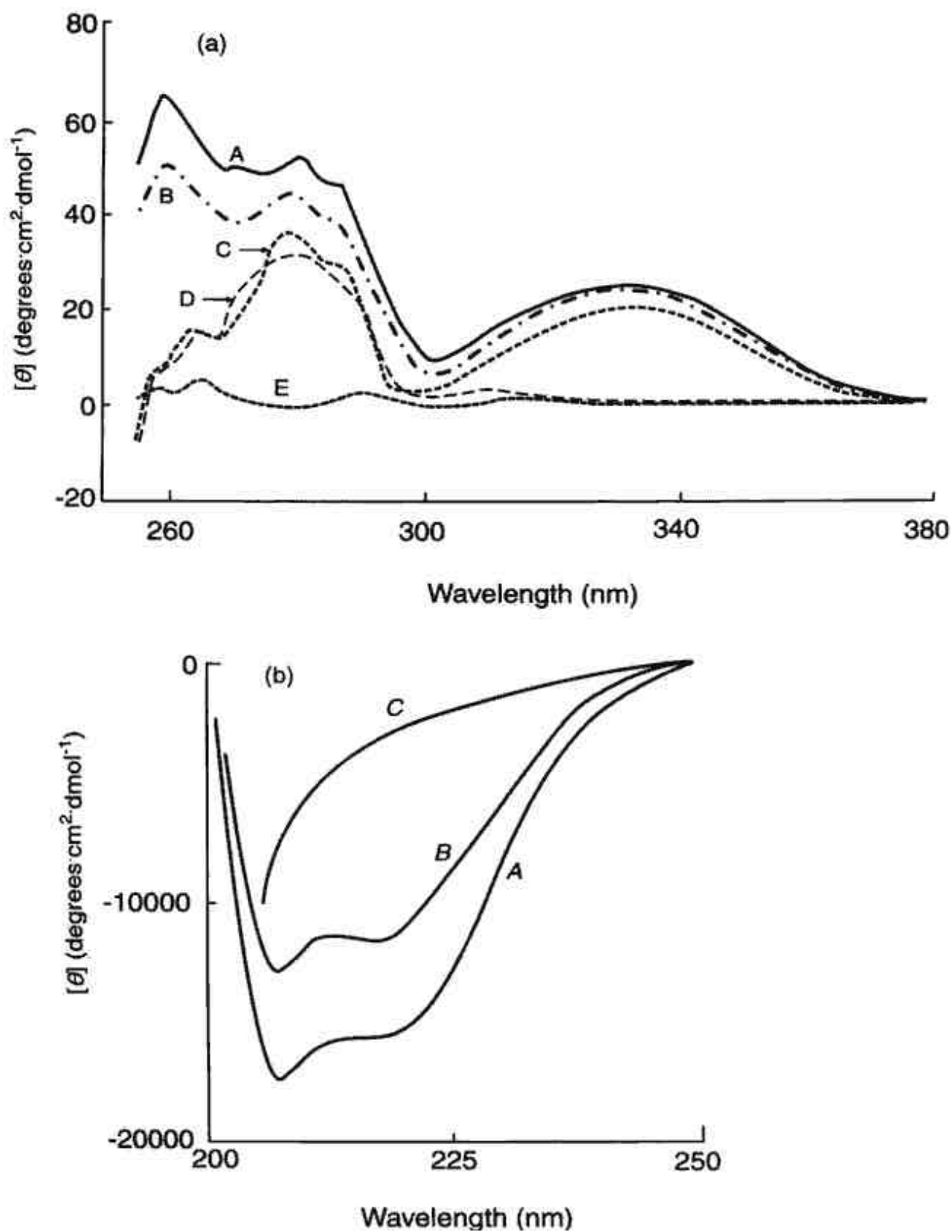
Fig. (5). The unfolding of phosphorylase *b* by GdmCl monitored by CD. (a) CD spectra over the range 255 to 380 nm were recorded in the absence of GdmCl and in the presence of 0.1 M (————), 0.1 M (— • • —), 0.53 M (• • • •), 0.8 M (– – – –) and 3 M GdmCl (— • —). (b) Far UV CD spectra recorded in the absence of GdmCl (A), and in the presence of 1 M (B) and 3 M (C) GdmCl.

3.4.1 Pyridoxal-5′-Phosphate

This cofactor is utilised by many enzymes involved in amino acid metabolism such as aminotransferases and decarboxylases; the reaction cycles involve protonated imine intermediates. Pyridoxal-5′-phosphate is also a cofactor in glycogen phosphorylase where its phosphate group is thought to play a role as a general acid in promoting attack by inorganic phosphate on the glycogen substrate [33]. When attached to proteins, the absorption spectrum of the cofactor usually shows a maximum around 350 nm.

In studies of the two isoenzymes of aspartate aminotransferase, CD was used to examine the unfolding of the proteins. The specific CD peaks at 365 nm (cytoplasmic isoenzyme) and 355 nm (mitochondrial isoenzyme) were used to monitor loss of the cofactor from the enzyme, since free pyridoxal-5′-phosphate shows no significant CD signal in this region. The disruption of the cofactor binding site occurred at much lower guanidinium chloride (GdmCl) concentrations than the loss of secondary structure. From this work it was shown that the mitochondrial isoenzyme was considerably less stable than the cytoplasmic isoenzyme; this was correlated with a much greater difficulty in refolding the former even in the presence of added pyridoxal-5′-phosphate [34,35].

A similar pattern was observed in the GdmCl-induced unfolding of phosphorylase *b*; dissociation of the pyridoxal-5'-phosphate cofactor (monitored by the loss of the CD signal at 335 nm) was complete at a concentration of denaturant (0.8 M). Under these conditions only very small changes in the secondary and tertiary structure of the enzyme had occurred as indicated by the far and near UV CD respectively [36] (see Fig. **5**).

3.4.2 Flavins

The two flavins (flavin mononucleotide (FMN) and flavin adenine nucleotide FAD)) play important redox roles in the electron transport chain as well as in a number of enzyme catalysed oxidation reactions of substrates such as amino acids and hydroxyacids. The conversion of oxidised to reduced flavin involves a two electron transfer. However, it is the ability of flavins to form a semiquinone radical intermediate between the oxidised and reduced states that allows them to mediate electron transfer between organic substrates (two electron donors/acceptors) and one electron donors/ acceptors. In the case of flavocytochromes (see Section 3.4.4), electrons can be passed one at a time to the acceptor via the haem moiety.

The oxidised form of the free flavins shows two absorption peaks at 360 nm and 450 nm, whereas the reduced form shows a broad absorption over the range from 300 nm to 500 nm. The neutral semiquinone shows a characteristic weak absorption around 600 nm, whereas the anionic semiquinone shows strong absorbance at 380 nm [37].

Since the free flavins exhibit only very small visible CD, dissociation of the cofactor usually leads to an almost complete loss of this signal. At intermediate temperature or denaturant concentration, alterations in the flavin environment may be detected by changes in the visible CD signal.

The magnitude of the CD signal of a particular protein-bound flavin can be difficult to predict; indeed the CD signal may be either positive or negative, so that in a multiflavin protein or system, there could be a certain amount of "cancelling out" of signals [38]. For example, the FMN-containing *E. coli* flavodoxin shows negative ellipticity over the range of approx. 310-450 nm, with a maximal visible CD signal at 365 nm (-160 deg.cm^2.dmol^{-1}). By contrast, its redox partner flavodoxin reductase exhibits positive ellipticity over the range 300-420 nm, with a maximum at approximately 390 nm (+95 deg.cm^2.dmol^{-1}) (Fig. **6**).

An important factor to be borne in mind during the analysis of the visible CD of flavoproteins is the strength of binding of the flavin to the protein. If the flavin is only rather weakly bound in a non-covalent fashion (*e.g.* the FMN moiety of flavocytochrome P-450 BM3, see section 4.6) then, under the experimental conditions employed, a significant proportion of the flavin may not be bound to the protein and thus may not contribute to the visible CD spectrum. However, for proteins such as bacterial flavodoxin, for which the K_d for the FMN is very small (in the nanomolar range), dissociation will be minimal under the usual conditions for collection of CD data. CD studies of flavoproteins have been recently reviewed [38].
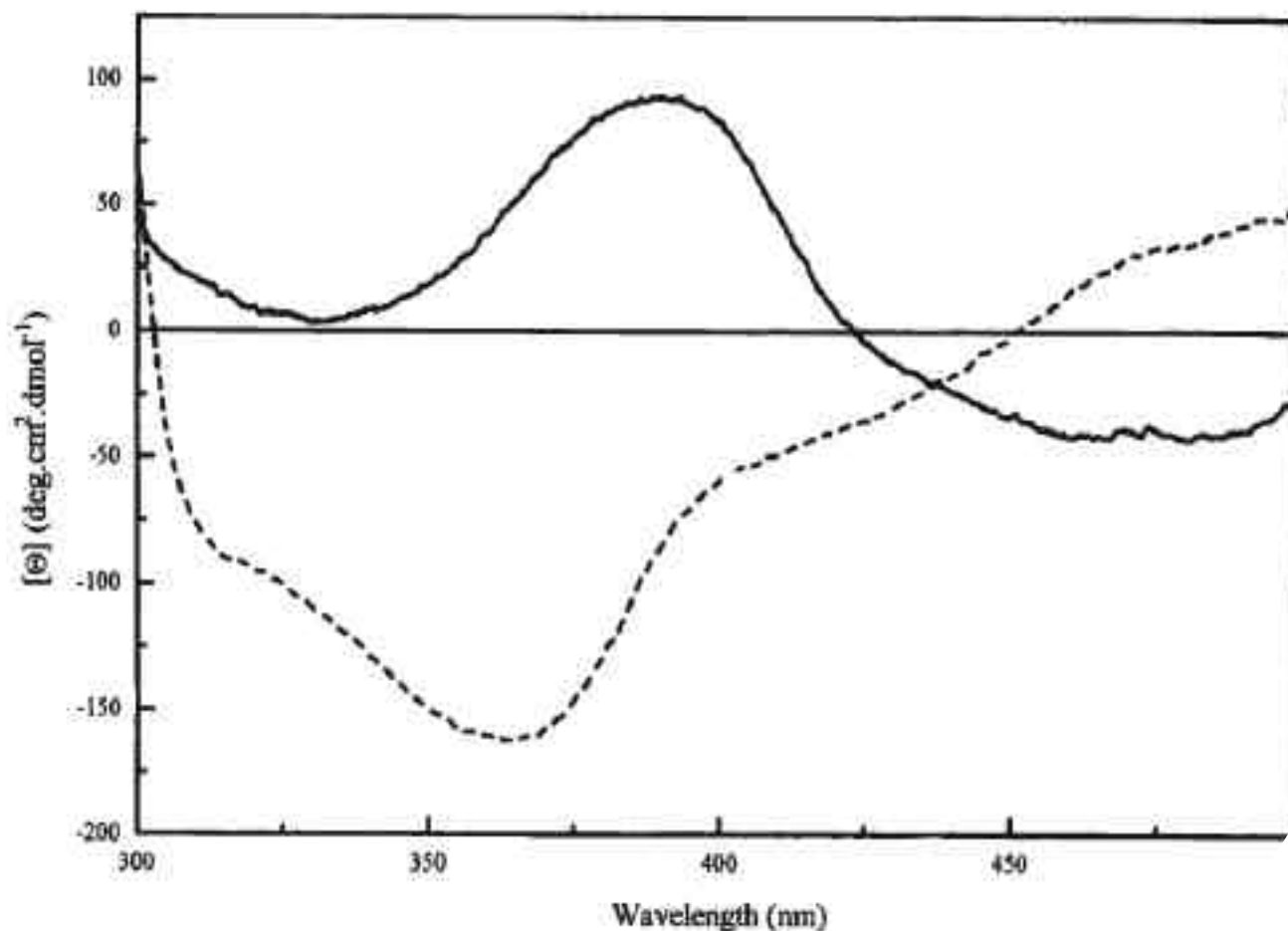
Fig. (6). CD spectra of flavoproteins in the near UV and visible regions. The spectra of *E. coli* flavodoxin and flavodoxin reductase are shown as dashed and solid lines respectively.

3.4.3 Haem

The haem cofactor can play a number of roles in proteins, such as binding of oxygen in haemoglobin, cytochrome *c* oxidase or cytochrome P-450, a cofactor for hydroperoxidases such as catalase, or as a redox cofactor as in cytochromes *b* and *c*. The haem moiety absorbs strongly around 410 nm (Soret band) and, depending on the nature of the coordinating groups and the spin state of the central Fe atom, at wavelengths in the region between 500 nm and 650 nm (alpha and beta bands) and around 350 nm (delta band).

CD has been used extensively to study the cytochrome P-450 enzymes; members of this large family of enzymes catalyse the O_2-dependent hydroxylation of a range of organic compounds. (The term P-450 derives from the fact that on addition of CO, the absorption maximum shifts to 450 nm).

Analysis of the CD spectra of a number of P-450 enzymes of known X-ray structures has allowed a distinct pattern to be discerned. The ratio of the CD bands at 410 nm (Soret) and 350 nm (delta) is correlated with whether a hydrogen bond is formed between the 6th ligand to the haem and the I-helix of the protein. The CD spectrum of a P-450 can thus be used to predict active site structural properties such as hydrogen bonding and polarity [39].

3.4.4 Flavocytochromes

Flavocytochromes are a group of enzymes which contain two redox cofactor systems, namely flavin (FMN or FAD) and a haem group. As mentioned above the advantage of such systems is that the flavin can act as a one electron carrier "bridge" between the substrate and haem group. Examples of flavocytochromes include several 2-hydroxyacid dehydrogenases such as L-lactate

dehydrogenase from *Saccharomyces cerevisiae* and the cytochrome P-450 BM3 system from *Bacillus megaterium*. Since the haem groups absorb in the same spectral region as flavins but with extinction coefficients generally around 10-fold higher than those of flavins, it can be difficult to perform spectrophotometric analysis of flavins in such proteins. CD in the visible region can therefore provide a particular valuable spectroscopic probe of the flavins in such enzymes. A good example is the nitrate reductase from *Chlorella*; where the flavin contribution is masked in the visible absorption spectrum of this molybdopterin- and haem *b*-containing enzyme. By contrast, strong visible CD signals from the FAD cofactor can be observed. Positive CD signals from the FAD are located at 311 nm and 387 nm, with negative CD signals at 460 nm and 487 nm. These features are lost on reduction of the flavin [38].

CD has proved to be very useful in studying the interactions between the reductase (flavin) and P450 (haem) domains of the BM3 flavocytochrome system from *Bacillus megaterium* and within the reductase domain itself; these studies are described in section 4.6.

3.4.5 Photosynthetic Pigments

The visible and near IR CD spectra of solubilised light harvesting complexes from photosynthetic bacteria (such as *Rhodopseudomonas acidophila*) show a number of peaks due to the carotenoid and bacteriochlorophyll components, which absorb in distinct spectral regions (Fig. **7**). The CD spectrum in the 750 to 900 nm region provides a sensitive probe for interactions between neighbouring bacteriochlorophyll molecules in the complex, as described in section 4.7 [40,41]. In addition, CD signals in the 450 to 550 nm region arise because in the complex the carotenoids are distorted by twisting from planarity. By contrast the extracted carotenoids dissolved in organic solvents are optically inactive [41] (see Section 4.7).

3.4.6 Induced CD Signals of Bound Ligands

The CD spectra of complexes formed between carbonic anhydrases and a range of (intrinsically optically inactive) azosulphonamide inhibitors has been used to explore detailed differences between the active sites of different isoenzymes and their mode of binding of the inhibitors [42]; this type of approach could be of potential value in assisting rational drug design (see Section 4.8).

3.5. Nucleic Acid Chromophores

The bases in DNA and RNA show quite broad absorption peaks with maxima in the near UV (around 260 nm) due to a multiplicity of $\pi \rightarrow \pi^*$ transitions. Additional peaks occur in the far UV (around 200 nm or lower). The bases themselves
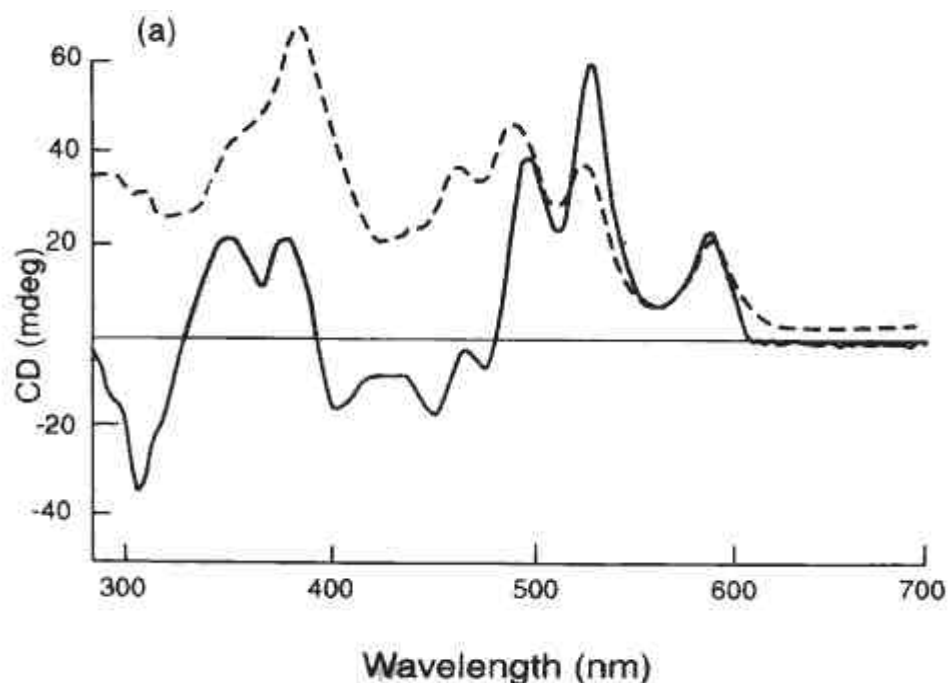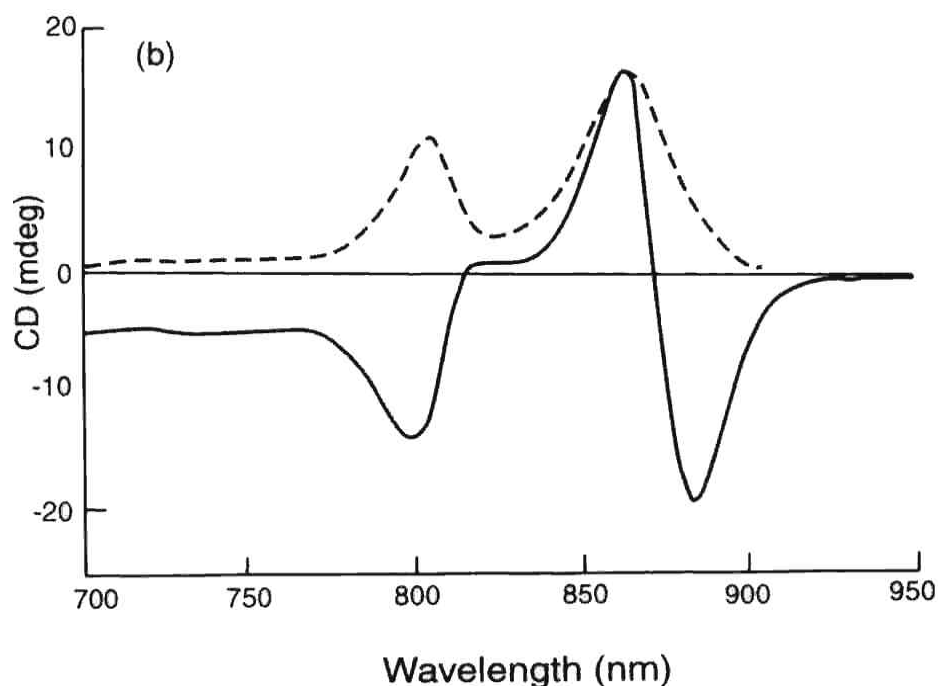
Fig. (7). Absorption and CD spectra of the LH2 light harvesting complex from *Rhodopseudomonas acidophila* strain 10050. For convenience the spectra in the near UV and visible (a) and near IR (b) regions are shown separately. In each panel, the dashed and solid lines represent the absorption and CD spectra respectively.

are planar and do not show any intrinsic CD; the chirality of the sugar residues induces a small CD in the nucleotide units. The stacking of bases which occurs in the helical structures adopted by oligonucleotides and nucleic acids gives rise to very large CD signals due to exciton coupling between the bases (see Section 1). The CD spectrum over the range from 230 nm to 300 nm is thus a very sensitive measure of secondary structure, with the various forms of DNA (A-, B- and Z-DNA) and RNA giving characteristic patterns in this region. The use of CD in analysing nucleic acid structure and nucleic acid- protein interactions has been reviewed in detail [43-45].

4. TYPES OF PROBLEM SUITABLE FOR INVESTIGATION BY CD

4.1 Secondary Structure Content of Peptides and Proteins

As outlined in section 3.1, the different types of regular secondary structures in proteins give rise to distinct types of far UV CD spectra (Fig. **3**). The task of taking any observed spectrum and deducing the contributions of the different structural forms would thus appear to be limited to the solution of a (large) number of simultaneous

equations using data over the far UV. In practice, this has turned out to be more complex for a number of reasons. Firstly, it is clear that aromatic amino acid side chains (especially Trp) can contribute to the far UV spectrum (see section 3.2) [1,3,30]. Secondly, disulphide bonds can contribute to the spectrum (see section 3.3) [3,14]. Thirdly, the characteristics of the spectra can depend on the length and regularity of structural elements in peptides and proteins [46]. More recent approaches have been based on sophisticated curve fitting procedures involving sets of spectra of well-characterised proteins of known secondary structure. The CONTIN procedure introduced by Provencher and Glöckner [47] involves the direct analysis of a CD spectrum over the range from 240 to 190 nm as a linear combination of the CD spectra of 16 proteins whose structures have been determined to high resolution by X-ray crystallography. The approach avoids the problem of defining reference spectra and provides a flexible analysis of actual protein spectra. In the CONTIN procedure, greater weight is automatically given to those proteins in the reference set whose spectra more closely resemble the protein being analysed. In the variable selection method of Johnson and coworkers [48], spectra can be deleted from the

reference set in a systematic fashion in order to assist the assessment of the best fit. In any case, it is clear from Fig **3** that the analysis of far UV CD spectra can be undertaken more confidently when data are available to low wavelengths [6]. For example, collection of CD data to 175 nm improves the accuracy of secondary structure determinations to within about 5% or better of that obtained from X-ray crystallography [49]. This finding explains the current interest in using synchrotron CD where the wavelength range can be extended to 170 nm [9,10].

In practice, with conventional CD instruments, it can be difficult to acquire data of sufficiently high quality below 190 nm, or even 200 nm on many biological systems if other absorbing components such as buffers are unavoidably present. However, given the additional benefits of recording CD data below 190 nm, it is usually worthwhile attempting to access this region. This could be done by lowering the concentrations of these other absorbing components, using buffer components of low absorbance, or using short pathlength cells (or a combination of these methods). If data from wave-lengths only above 200 nm are available, it is still possible to make fairly reliable estimates of the α-helical content of peptides and proteins from the ellipticities at 208 and 222 nm [50], the wave-lengths corresponding to the characteristic double minima of the α-helix spectrum (Fig. **3**). The analysis of far UV CD spectra has been extensively reviewed [3,6,27,49,51,52].

The power of synchrotron CD in protein analysis has been well demonstrated by studies on the assembly of clathrin coat proteins which are involved in receptor-mediated endocytosis [10]. The clathrin coat protein is composed of triskelia which consist of 3 heavy chains (190 kDa), 3 light chains (27 kDa) and assembly polypeptides (APs). The assembly polypeptide associated with the plasma membrane is known as AP-2. Previous CD studies of such protein assembly systems had been hampered not only by the limited wavelength range but also by the distortion caused by scattering in turbid samples. In the synchrotron CD system used, the scattering problem was reduced by the close proximity of the sample and detector, and the increased intensity of the source allowed the wavelength range to be extended down to 175 nm even for the turbid samples. The whole

coat protein as well as a variety of constituent polypeptide chains and proteolytic fragments were examined at both pH 6 and pH 8, corresponding to assembly and disassembly conditions. The results of detailed secondary structure analyses showed that the assembly process involved no significant change in secondary structure of the proteins. In the absence of AP-2, the assembly process leads to the formation of cages, with a small (10%) loss in helical content. Stopped-flow CD measurements showed that this loss occurs within 1 msec of the start of the assembly process.

4.2 Conformational Changes in Proteins

CD has proved to be an ideal technique to monitor conformational changes in proteins which can occur as a result of changes in experimental parameters such as pH, temperature, binding of ligands etc. During the unfolding of the protein, the extent of the conformational changes can be analysed to provide quantitative estimates of the stability of the folded state of the protein; CD has been extensively used for this purpose [4,53].

In recent years there has been considerable interest in the ability of certain proteins to interconvert between different forms of secondary structure. The formation of β-sheet structures in certain proteins has been shown to act as a precursor to the formation of amyloid fibrils which accompany degenerative diseases such as scrapie, bovine spongiform encephalopathy (BSE), Alzheimer's disease etc. Far UV CD has been an invaluable tool for monitoring the transitions from α-helical to β-sheet structures in such proteins and peptides. Thus a 21 amino acid peptide based on residues 106-126 of the human prion protein was found to switch from an α-helical to a β-sheet structure in the presence of sodium dodecylsulphate (SDS) as the concentration of detergent was lowered from 28 mM to 3.5 mM [54] (see Fig. **8**). As described in section 4.4, an analogous switch between α-helical and β-sheet states in a designed 16-residue peptide can be brought about by changes in temperature. In a longer peptide consisting of residues 91 to 231 of the human prion protein, the transition from α-helical to β-sheet structure can be brought about by lowering the pH from 8 to 4 and reducing the disulphide bond [55].
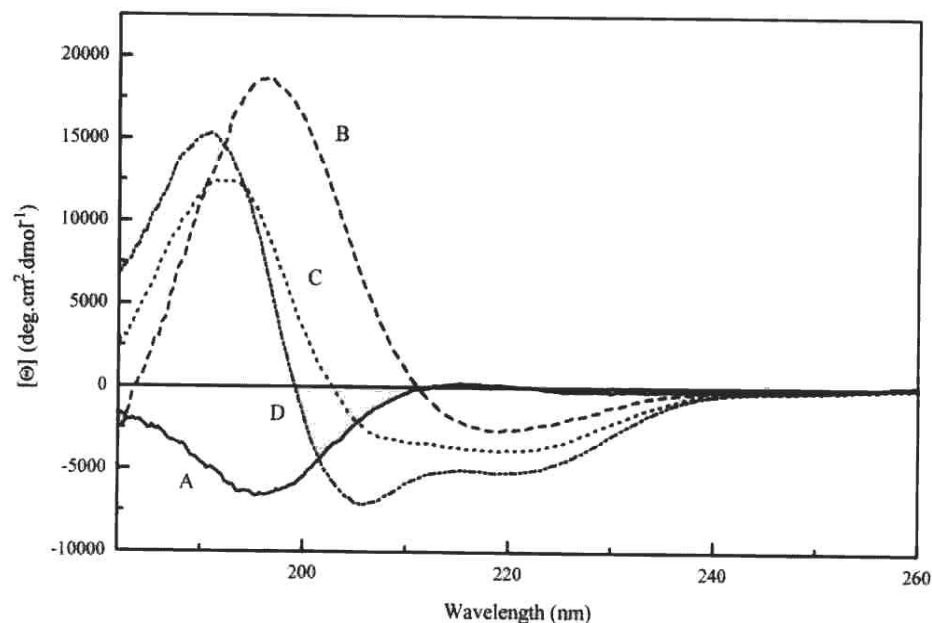
Fig. (8). Far UV CD spectra of a prion protein peptide. Spectra were recorded in the absence of SDS (A), and in the presence of 3.5 mM (B), 7 mM (C) and 28 mM SDS (D) respectively.

Addition of specific ligands such as substrates, cofactors or regulatory molecules to proteins can lead to structural changes which are essential for their function [56]. Such structural (or conformational) changes can be observed by CD in the far UV, the near UV or both regions. It should be noted that small conformational changes are more likely to be detected in the near UV since the CD contributions of the aromatic side chains is usually very sensitive to their environment. By contrast, changes in the far UV CD will usually reflect major backbone changes in the protein. One of the first applications of CD was to show that addition of the competitive inhibitor N-acetylglucosamine to lysozyme led to large (100%) increases in the near UV CD peaks at 285 and 295 nm characteristic of Trp side chains [57].

CD was used to monitor the binding of Ca^{2+} to troponin C, the protein which confers Ca^{2+}-sensitivity to muscle. Addition of Ca^{2+} led to an increase in the size of the negative CD signal over the range from 200 to 230 nm, reflecting the increased stability of helical regions of the protein. The overall helical content increased from about 35% to 60% on addition of saturating levels of Ca^{2+} to the Ca^{2+}-free protein. Changes in CD signal at 222 nm could be used to monitor the binding of Ca^{2+}; analysis of the titration curve yielded the dissociation constants for the strong and weak sites which have dissociation constants of the order of 1 nM and 1 μM respectively [58].

Changes in the near UV CD of the calcium binding protein calmodulin accompanying binding of Ca^{2+} have been analysed by Martin and Bayley [59]; the changes reflect alterations in the environments of Phe and Tyr side chains near the Ca^{2+}-binding sites.

CD has been used to monitor the binding of substrate (ATP) to a mutant form of the molecular chaperone GroEL. This protein consists of two stacked rings of seven subunits with a central cavity within which protein folding is assisted. The mutation K3E reduces the subunit interactions so that the mutant is monomeric, and shows no chaperone activity. The far UV CD confirmed that the mutant was folded, although there were some differences from the wild type protein. Binding of ATP to the K3E mutant was shown by the finding that addition of the nucleotide led to a 50% increase in the size of the CD signals over the wavelength range from 230 to 190 nm, consistent with a large increase in the helical content of the protein. By contrast, addition of ATP does not lead to any increase in change in the far UV CD signal of the wild type protein [60].

The near UV CD of the hexameric enzyme glutamate dehydrogenase has been used to monitor the binding of NAD^+ in the presence of the competitive inhibitor glutarate. Addition of NAD^+ led to a sharp change in the near UV CD signal at a point corresponding to half saturation of the sites.

The findings were consistent with the proposal that half saturation had led to a conformational change, resulting in a form of the enzyme with lower affinity for NAD+, thereby accounting for the observed negative cooperativity in coenzyme binding [61].

CD has been used extensively to monitor the R ↔ T allosteric transition in haemoglobin. The CD spectrum of the protein shows a number of bands in the 270-300 nm region due to the Tyr and Trp side chains, and a large band at 260 nm due to the haem group. Deoxyhaemoglobin shows a marked negative band at 287 nm; oxyhaemoglobin shows weak negative bands at 283 nm and 290 nm. From an analysis of the behaviour of mutant haemoglobin such as haemoglobin Kemsey and the effect of agents such as inositol hexakisphosphate which inhibit the structural transition, it was concluded that the marked changes in the near UV CD spectrum could be ascribed to changes in the environments of Trp 37 on the β-chain and Tyr 42 on the α-chain. The latter side chain was hydrogen bonded to Arg 40 on the β-chain at the $\alpha_1\beta_2$ interface in the T (deoxy) state; an interaction which was lost in the R (oxy state) [62].

One particularly important use of CD has been to characterise the overall structural states of proteins; in particular the technique provided some of the first convincing demonstrations of a state intermediate between the native and fully unfolded states. For example, when α-lactalbumin is incubated at pH 2, it retains most of its native secondary structure as shown by the far UV CD. However, under these conditions the near UV CD signal of the protein is very much reduced (Fig. **9**) indicating that native tertiary interactions are absent [63]. Studies of the Stokes radius (using gel permeation) show that at pH 2 the molecule is compact, leading to the term "molten globule" or "compact intermediate". In addition to the CD properties, a number of other characteristics of this type of structure have been described, including the lack of chemical shift dispersion in NMR and enhanced binding of the fluorescent dye 1-anilino-8-naphtha-lenesulphonate (ANS) [64]. Molten globule states may represent important intermediates in protein folding pathways [65,66] (see Section 4.3.3), and it has been suggested that they can be recognised by certain types of molecular chaperones [67]. For these reasons, the
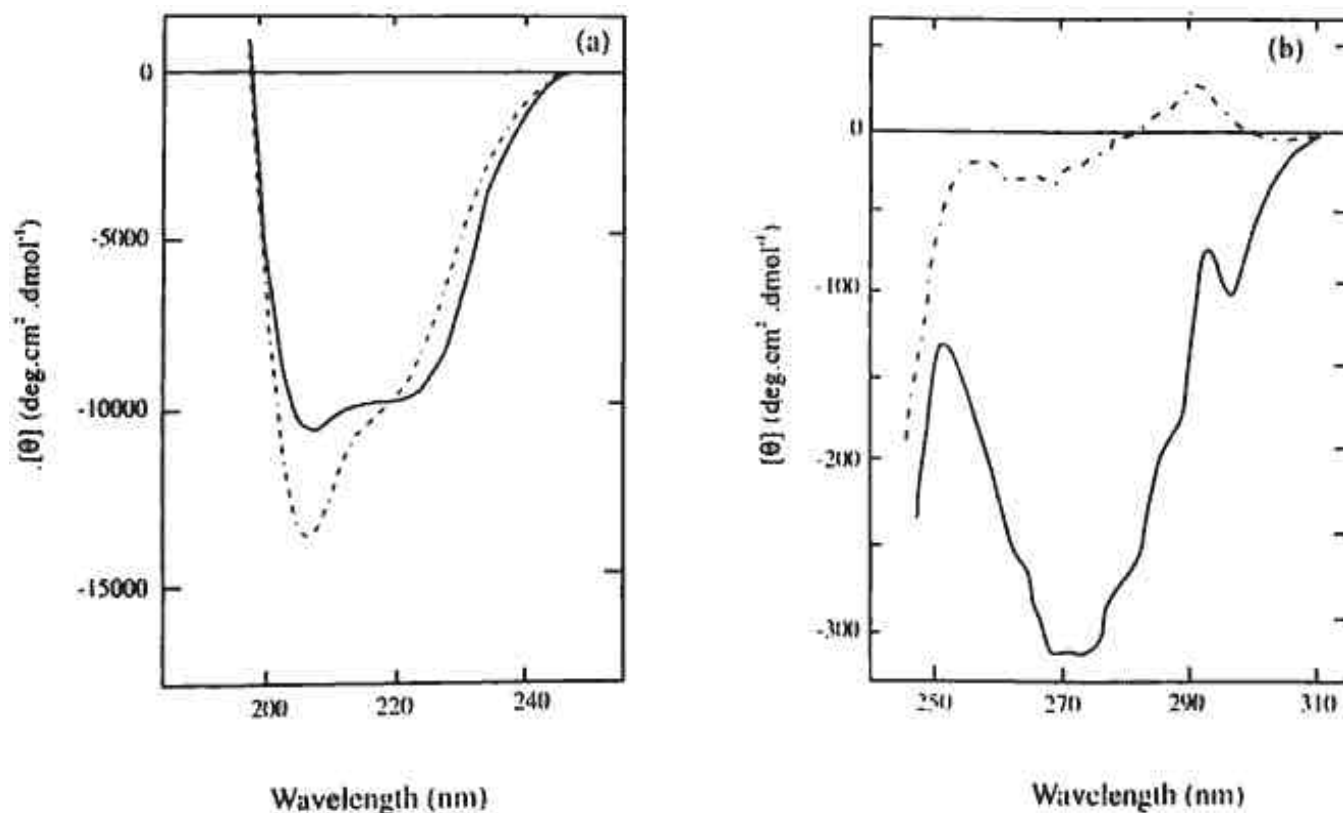

Fig. (9). CD spectra of α-lactalbumin. (a) and (b) represent far UV and near UV CD spectra respectively. The solid and dashed lines represent spectra obtained at pH 7 and pH 2 respectively [63].

molten globule state of a number of proteins has been widely studied. An interesting study on the heterodimeric protein bacterial luciferase suggested that the isolated individual subunits existed essentially as molten globules, as indicated by far UV CD and NMR measurements. The molecular chaperone GroEL was shown to bind to both subunits and to assist folding [68].

A number of other possible roles for molten globule states of proteins have been suggested [65]. For example, the flexibility of the molten globule state may be an important factor in the recognition of DNA by certain proteins [69] and in virus and phage capsid assembly processes [70,71]. CD studies played an important part in defining the molten globule characteristics of such proteins.

4.3 The Refolding of Proteins and Peptides

4.3.1 General Consideration

The mechanism of protein folding represents one of the major unsolved problems in molecular biology. Apart from its intrinsic scientific interest, an understanding of protein folding is important for the successful production of expressed proteins [72] and for understanding the molecular basis of an increasing number of disease states which appear to arise from protein misfolding [73]. For nearly 30 years, since Levinthal pointed out that if the folding of a protein involved a sampling of all possible conformations available to it, the process would take an impossibly long time [74], it has been clear that protein folding is likely to involve specific pathways. Amongst the stages which must occur during the folding of proteins are:- (a) the formation of elements of recognisable secondary structure (helices, sheets, turns etc.), (b) the collapse of the extended polypeptide chain to form a more compact state in which non-polar side chains are largely buried away from the solvent, (c) the formation of the distinct long-range interactions which characterise the native tertiary structure and are a prerequisite for the formation of specific binding and catalytic sites, and (d) the association between subunits in oligomeric proteins [26,75,76]. Whether a single all-embracing model for protein folding exists is still a matter of debate. The task for the experimental scientist is to employ a variety of techniques to gather structural and kinetic data which allow models for folding to be proposed and tested. CD has proved to be an extremely important technique for examining some of the early events in the folding process, as indicated below.

### 4.3.2 Experimental Approaches to Protein Folding

By far the most detailed insights into the kinetics and mechanism of protein folding have come from studies of refolding, in which a mature protein is unfolded (usually by high concentrations of urea or GdmCl), and then refolding is initiated by rapid dilution of the denaturing agent. Thus a typical experiment might involve incubating a protein in 6 M GdmCl, and performing a dilution into buffer to initiate refolding. It is important to show that (a) the protein is completely unfolded prior to dilution and (b) that the residual concentration of denaturant after dilution has a negligible effect on the protein. In order to establish the mechanism of refolding it is necessary to study the time scale of the recovery of structure from the denatured state, and to correlate these data with the rate of recovery of biological activity. The regain of secondary structure is often monitored by far UV CD and protection against amide hydrogen exchange (using NMR); that of tertiary structure by near UV CD and fluorescence. The regain of quaternary structure can be monitored by rapid cross-linking with glutar-aldehyde or by subunit hybridisation [26,76]. In view of the rates of the various processes, it is usually necessary to use rapid kinetic methods (stopped or quenched flow) as well as manual mixing methods (dead time typically 10-15 sec) in order to probe as much of the refolding process as possible. Stopped flow methods typically have dead times in the range 1 to 10 msec, and involve dilutions of 11-fold or greater; section 2.5).

4.3.3 The Nature of Early Folding Intermediates

Stopped flow CD has been extensively used to examine the properties of the early intermediates in protein folding. In some of the first studies of this type [77], it was shown that during the refolding of denatured ferricytochrome *c* and β-lactoglobulin the native ellipticity in the far UV was regained within the dead time of the instrument (estimated as 18 msec). By contrast the

ellipticity in the visible region (ferricytochrome *c*) or the near UV (β-lactoglobulin) was only regained over a period of minutes. These results were taken to indicate that within 20 msec an intermediate had formed in which a substantial proportion of native secondary structure was present; however this intermediate had little stable tertiary structure. Similar results have been obtained for a number of other proteins including α-lactalbumin [78] and apomyoglobin [79], giving rise to the proposal that the folding process proceeds via an intermediate with the principal characteristics of a "molten globule" (see Section 4.2). In a study of the refolding of native and mutant forms of staphylococcal nuclease, it was found that a transient kinetic intermediate formed within 10 msec [80]. This intermediate possessed only about 30% of the native ellipticity at 225 nm; in conjunction with hydrogen exchange experiments it was shown that this relatively low amplitude reflected formation of the 5-stranded β-sheet structural core of the N-terminal domain of the enzyme, with the helices (which contribute the bulk of the CD signal at 225 nm) formed at a later stage [81].

Some of the most detailed studies on protein folding to date have been those on the refolding of hen egg white lysozyme [14,82,83]; it should be noted that in these studies the four disulphide bonds of the enzyme are kept intact. In addition to stopped flow CD and fluorescence, NMR (pulsed amide hydrogen exchange labelling) has been used to define those regions of the protein where secondary structure and tertiary interactions had formed at various stages of the refolding process. The NMR results showed that the folding process appears to involve multiple pathways of folding with a variety of structured intermediates in which secondary structure had formed in the two domains (α-helical and β-sheet) of the enzyme. Additional insights into the folding of lysozyme have been obtained from stopped flow CD measurements in the far and near UV (Fig. **10**). The near UV results (289 nm) show that the native ellipticity at this wavelength develops in a single kinetic process, with a rate very similar to the protection of amides in the β-sheet domain. Since the majority of the tryptophan side chains are in the α-domain, where the amides are protected faster, this indicates that the partially folded intermediate form of this domain has the
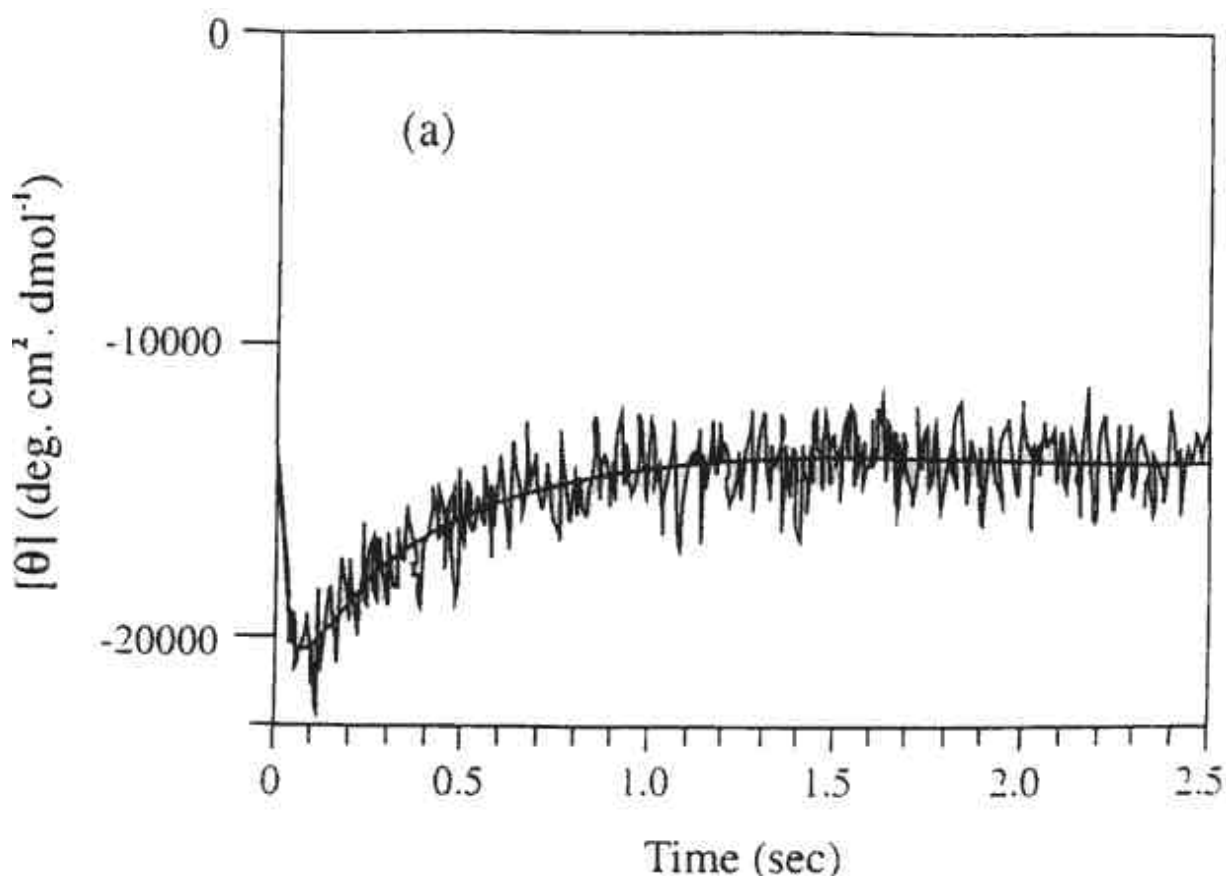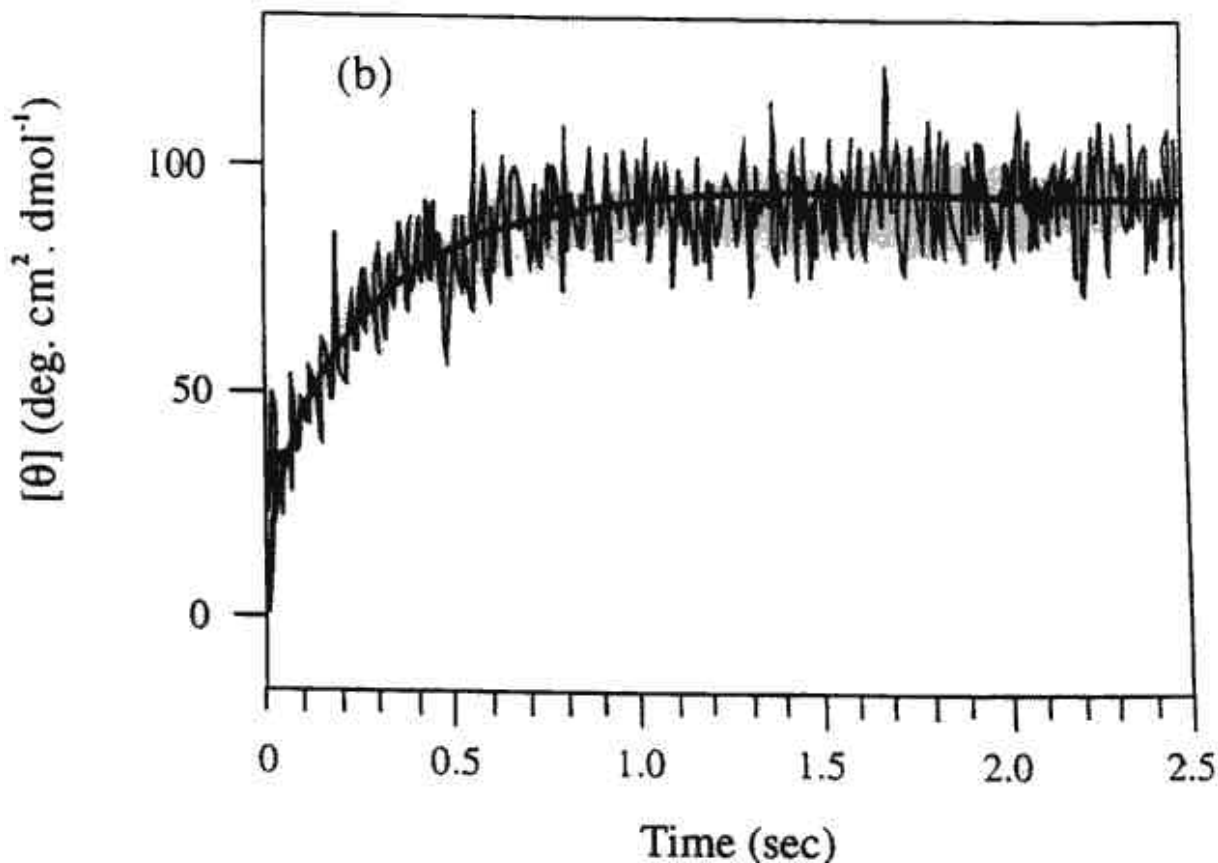
Fig. (10). The refolding of lysozyme after denaturation in GdmCl [82,83]. (a) Refolding monitored by changes in the ellipticity at 225 nm. The mean residue ellipticities of denatured and folded protein samples are -1500 and -14000 deg.cm^2.dmol^{-1} respectively. (b) Refolding monitored by changes in the ellipticity at 289 nm. The mean residue ellipticities of denatured and folded protein are 0 and 100 deg.cm^2.dmol^{-1} respectively.

characteristics of a molten globule, with the tryptophan side chains not yet having adopted their native orientations. The kinetics of the changes in CD signal at 289 nm are also very similar to those of the development of a binding site for a fluorescent derivative of chitobiose which is a competitive inhibitor of the enzyme.

By contrast the results from stopped flow CD in the far UV show some of the complexities involved in the interpretation of these experiments (Fig. **10**). Within the dead time of the instrument (about 5 msec) about 80% of the native ellipticity at 225 nm has been regained. In a second phase, essentially complete after 80 msec, the signal at 225 nm increases to about 150% of the native value. The CD signal returns to the native value in a third phase (of half time 300 msec). From these results it is clear that a large proportion of secondary structure is formed very quickly (within

2 msec) by which time there is only a very small degree of amide protection in the NMR experiments. This suggests that such protection arises from the stabilisation of already formed helices (see below). The origin of the apparent "overshoot" of the far UV CD signal in the second phase of the process has been examined in detail [14]. The most likely explanation for the overshoot is contributions from side chains, of which the largest component comes from disulphide bonds, with only a small contribution from aromatic side chains. The sign of the CD signal of disulphide bonds in the far UV is critically dependent on the -C-S-S-C- dihedral angle (section 3.3). The large negative ellipticity at 225 nm observed during the early stages of folding of lysozyme could therefore arise from four disulphide bonds with dihedral angles close to 90° in the P (right-handed helix) chiral conformation [31] and has been observed in a number of

disulphide cyclic peptides of various lengths containing a proline residue [32].

The relationship between the hydrogen exchange method and far UV CD in assessing the degree of secondary structure formation in early folding intermediates of several proteins has been examined in detail by Guijarro *et al.* [84].

The above studies have shown that compact molten globule-like intermediates possessing the majority of the native secondary structure are formed on the millisecond time scale during the folding of several proteins. However, the relationship between the rate of hydrophobic collapse and the rate of secondary structure formation has been less well explored. The studies by Agashe *et al.* [85] on the refolding of barstar (an 89 amino acid protein which is a powerful inhibitor of barnase) after GdmCl-induced unfolding have allowed the situation to be resolved at least for this protein. The degree of compactness of the molecule was measured by the efficiency of energy transfer between the tryptophan side chains (there are three distributed at both ends of its bundle-like structure) and dansyl groups attached to the cysteine side chains. Using this approach, it was possible to show that the collapse of the polypeptide chain to a compact form had occurred within 4 msec after the start of refolding of barstar. This burst intermediate was able to bind ANS, but showed essentially none of the secondary structure (far UV CD) or tertiary structure (near UV CD or intrinsic tryptophan fluorescence) of the native protein. The elements of secondary and tertiary structures were both regained in slower biphasic processes. The generality of the conclusions reached for barstar remains to be established.

4.4 Protein and Peptide Design

CD is an extremely useful tool for the structural characterisation of peptides and protein fragments which have been designed to adopt specific structures and hence display particular biological functions. This type of approach is not only important in understanding structure-function relationships, but is of considerable technological importance since such fragments could well be much more stable than the corresponding full size proteins. There is increasing interest in the idea that stable protein fragments or designed peptides could form the basis of "molecular switching" devices [86,87]. The study of protein fragments could also provide insights into some of the early stages in protein folding, since such fragments could act as nuclei for directing further folding [88] (see Section 4.3).

Using information on the secondary structure preferences of individual amino acids or sequences of amino acids, it has been possible to design oligopeptides which show a strong tendency to form α-helices in solution [88]. For example, peptides based on a repeating unit AEAAKA ranging in length from 14 to 50 residues form helices in solution as shown by far UV CD. The urea-induced unfolding of these peptides has been monitored by CD and the energetics of the unfolding process have been shown to correspond reasonably well with those of the unfolding of helices in proteins such as myoglobin [89].

Model peptides for β-sheet structures have been rather less well studied but such peptides could have interesting structural and mechanical properties. The 24-residue peptide KLEALYVLGFF GFFTLGIMLSYIR is based on the sequence of the transmembrane domain of the IsK (K$^+$ channel) protein. The peptide forms β-sheet structures in lipid bilayers suggesting that in amphiphilic solvents such as methanol or 2-choroethanol the peptide chains would self assemble into "molecular tape" structures in which the β-sheets are aligned [86]. The CD and FTIR properties of the peptides indicated the presence of the anti-parallel β-sheet structural motif and the electron microscopy and wide angle X-ray diffraction were used to confirm the existence of the tapes. The tapes were shown to be a single molecule thick by rheological and atomic force microscopy. From the results it was possible to design novel β-sheet tape forming peptides of differing polarity, which could be of use in mechanical or physical trigger devices, as well as providing insights into fibril formation in various degenerative diseases [86].

Peptides of the type Ac-KAKAKAKAEAEA EAEA-NH$_2$ are termed self-complementary because of the pattern of distribution of opposite charges. Such peptides show interesting structural changes which can be induced by temperature or pH changes. At 20°C, the far UV CD spectrum of

the peptide shows that it forms a stable macroscopic β-sheet structure in water, which stains with the dye Congo red, a characteristic property of fibrils. However, on raising the temperature to 70°C there is an abrupt change in the CD spectrum which shows that a helical structure is formed. On cooling back to 20°C, the reformation of the β-sheet structure is very slow, taking several weeks [90]. It has been proposed this latter process in such peptides may be a suitable model for the slow formation of amyloid fibrils in a number of neurological disorders. This type of approach has been extended to include peptides in which a β-sheet-forming core (VT)$_n$ is flanked by sequences of polar amino acids to promote water solubility. Far UV CD studies show that when n>6, the peptides form water-soluble β-sheet complexes. These fibrillar structures are of molecular mass >2000 kDa and show some similarities in terms of appearance under the electron microscope and Congo Red staining to natural amyloid fibrils [91].

When individual secondary structural elements are linked together a number of new possibilities emerge. Fezoui et al. [21] designed a 38-residue peptide which consisted of two amphipathic α-helices each 17 amino acids in length linked by a turn sequence (GTDS). CD and NMR studies showed that the peptide formed the expected helices which associated with each other in an anti-parallel fashion. The extent of interaction between the helices could be judged by observing the melting of the peptide. This type of model peptide is important because the helix-turn-helix motif is an important feature in protein folding mechanisms, and is also found in many DNA-binding proteins [92].

The synthetic 33-residue peptide GCN4-p1, derived from the yeast transcriptional activator GCN4, forms a stable bimolecular coiled coil structure typical of the leucine zipper structural motif in a number of DNA-binding proteins [93]. In an ingenious experiment [94], two different helix-turn-helix peptides (27 amino acids in length) were "dimerised" by formation of disulphide bonds between appropriately positioned cysteine side chains. This led to the formation of an anti-parallel four helix bundle, a stable structural "core" found in many proteins from hyperthermophilic organisms. As shown by far UV CD, the peptide was remarkably stable towards thermally-induced

unfolding. Loss of helical structure occurred in a co-operative fashion with mid points in the range 78 to 95°C depending on pH. Such protein "cores" might form the basis of thermally stable artificial enzymes.

The four helix bundle motif has been exploited in the design of a model system for flavocytochromes (see Section 4.6). The system consisted on two 62 amino acid helix-loop-helix peptides, each of which contains a single cysteine to which flavin was linked covalently and two histidines which could coordinate the haem cofactor. The flavin and haem groups were both located within the hydrophobic core of the protein. Laser-pulse activated electron transfer occurs from flavin to haem on a rapid time scale (100 nsec), showing that the peptide matrix can mimic the environment provided by the protein in flavocytochromes [95].

4.5 Membrane Proteins

There have been a number of impressive advances in the study of integral membrane proteins in recent years, as exemplified by the solution of the structures of cytochrome *c* oxidases from bacterial and mammalian sources [96,97]. However, it is still the case that membrane proteins make up only a small proportion of known protein structures [98]. The principal reason for this has been the difficulty in obtaining suitable crystalline preparations of integral membrane proteins under conditions where it can be confidently assumed that the structure resembles that of the native protein in the membrane. Indeed the first structure determination of a membrane protein, bacteriorhodopsin, was performed by electron diffraction on intact purple membranes from the archaebacterium *Halobacterium salinarium* where the protein (which acts as a light-driven proton pump) forms regular two-dimensional arrays in the purple patches [99]. The majority of structural studies on integral membrane proteins, however, are performed on preparations which have been extracted from the membranes by suitable detergents. Addition of crystallising agents and amphiphiles to the extracted proteins can give crystals of good quality for diffraction studies [100]. A second problem in working with membrane proteins is that it can be difficult to set up assays for biological activity of

the extracted proteins, especially for those which play a transport, binding or structural role. CD is an important technique in the study of membrane proteins, particularly to address the question of the retention of native structure on extraction, purification and crystallisation. For example, in studies of the extraction of myelin basic protein (MBP) from equine myelin (for which there is no convenient assay), CD was used to assess the effect of several detergents on the protein. An optimum procedure was developed using extraction with lauryldimethylamine-N-oxide followed by cation-exchange chromatography which led to high levels of extraction with the retention of secondary structure. From these preparations, crystals suitable for diffraction were obtained and structural work is currently in progress [101].

As shown by electron diffraction [99] and more recently by X-ray diffraction [102], bacteriorhodopsin possesses seven transmembrane α-helices (designated A-G); the retinal chromophore is linked by a Schiff base (imine) to the side chain of Lys 216 on helix G. Because of the considerable amount of structural detail available, bacteriorhodopsin has been used extensively as a model membrane protein for CD studies. Wallace and Teeters examined the spectra of bacteriorhodopsin in different preparations, namely intact purple membranes, sonicated purple membranes and small unilamellar vesicles [103]. It should be noted that the study of bacteriorhodopsin in purple membranes poses two major difficulties:- firstly, the optical artefacts caused by scattering of light by membrane fragments which are large compared with the wavelength of the radiation, and secondly the effect of "absorption flattening" which occurs in purple membranes and sonicated membrane preparations. The latter effect arises from the non-random distribution of chromophores in these samples, because the proteins are sequestered at very high local concentrations in discrete regions of the membrane. By contrast absorption flattening does not occur in the small unilamellar vesicle preparations; as a result the CD signals in the far UV are significantly higher (by 27% at 224 nm and 42% at 193 nm) [103]. The errors caused by absorption flattening can lead to errors in the determination of secondary structure content [103,104]. By careful determination of protein concentration and taking absorption flattening into

account, the secondary structure content of bacteriorhodopsin in the unilamellar vesicles was found to be 80% α-helix and 0% β-sheet, in accordance with the results of electron diffraction data. In further work, Swords and Wallace [105] showed that the solubilisation of bacteriorhodopsin in the non-ionic detergent octyl-glucoside led to significant changes in secondary structure, as indicated by the far UV CD spectrum and a small shift in the absorption maximum of the retinal chromophore compared with that of native membrane fragments. This result indicates that the lipid to protein ratio may be an important factor contributing to membrane protein structure, and that it cannot be assumed that the detergent environment is always the same as that provided by the membrane.

Bacteriorhodopsin can be denatured and refolded *in vitro*; the details of this process have been examined by the use of stopped-flow kinetic studies on the intact protein and fragments derived from it [106]. The refolding of bacteriorhodopsin from an SDS-denatured state in DMPC/DHPC (L-α-1,2-dimyristoylphosphatidylcholine/L-α-1,2-dihexano-ylphosphatidylcholine) lipid micelles was studied by stopped flow CD [107]. This showed that there was a surprisingly slow formation of a significant fraction of the α−helical structure with a half time of more than 10 sec. One possible interpretation proposed was that the formation of the native seven transmembrane helix structure of bacteriorhodopsin involves the rapid formation of a bundle consisting of the central regions of the helices with the ends of the helices forming much more slowly. However, other studies (including CD) on the folding of fragments of bacteriorhodopsin shows that only five of the helices (A-E) could form individual stable trans-bilayer helices in phospholipid vesicles. Thus in the folding of the intact protein, it is likely that folding and insertion of the two C-terminal helices (F and G) follows a more complex pathway and could well represent the rate limiting step [108,109]

In contrast to the transmembrane helix motifs found in many membrane proteins such as bacteriorhodopsin and the G-protein linked receptors, it appears that outer membrane proteins are rich in β-barrel structures. The archetypal protein of this type is porin which consists of 16 strands of anti-parallel β-sheet folded to form a β-

barrel structure [110]. The folding and membrane insertion of the outer membrane protein OmpA from *E. coli*, which in its native form is of the β-barrel type structure, have been examined [111]. The OmpA protein was unfolded in a solubilised form in 8 M urea; refolding was initiated by dilution of this solution, and simultaneous addition of sonicated DMPC/DMPG (L-α-1,2-dimyristoyl-phosphatidyl-choline/L-α-1,2-dihexanoylphospha-tidylglycerol) vesicles. The structural changes in the protein were monitored by far UV CD and protein fluorescence, and the degree of insertion into the vesicle by susceptibility to protease digestion. The changes in CD signal at 206 nm showed that there were three phases to the folding and insertion process. The first of these with a half time of less than one second corresponds to the formation of a misfolded state in water, the second process (half time of about 5 minutes) to the formation of a partially folded state in the membrane and the third step (half time about 60 minutes) to the formation of the fully folded state in the membrane.

The nicotinic acetylcholine receptor consists of 4 types of subunit with the stoichiometry $\alpha_2\beta\gamma\delta$. From hydropathy plots of the amino acid sequence of each subunit, it has been proposed that there are 4 transmembrane segments (designated M1-M4). Peptides containing these proposed segments were purified from proteolytic digests of different receptor subunits and incorporated into lipid vesicles. Far UV CD and FTIR measurements showed that the peptides were largely (> 80%) α-helical, though in the case of the M1 segment from the α subunit there was evidence from the CD spectrum that other structures could be present, or that the helix was distorted [112].

Membrane proteins which are not deeply embedded in a membrane are easier to characterise since they can be more easily prepared in a soluble form. An example is the kinetoplastid membrane protein-11 (KMP-11) which is a major component of the surface membrane of *Leishmania* parasites, and is thought to play a key role in the internalisation of the parasites within the host cell. CD studies of recombinant KMP-11 showed that it possessed a high helical content (75%) consistent with predictions from the amino acid sequence of two amphipathic helices separated by a short random-coil segment. In such a structure, the presence of the helices would facilitate the interaction of the protein with lipids and account for the ability of the protein to perturb membrane stability [113].

4.6 Interactions Between Domains in Proteins

Most polypeptide chains larger than about 30 kDa tend to exist as multiple domains, i.e. independent folding units [56]. Although in some multidomain proteins, it is possible to isolate the individual domains by limited proteolysis of the
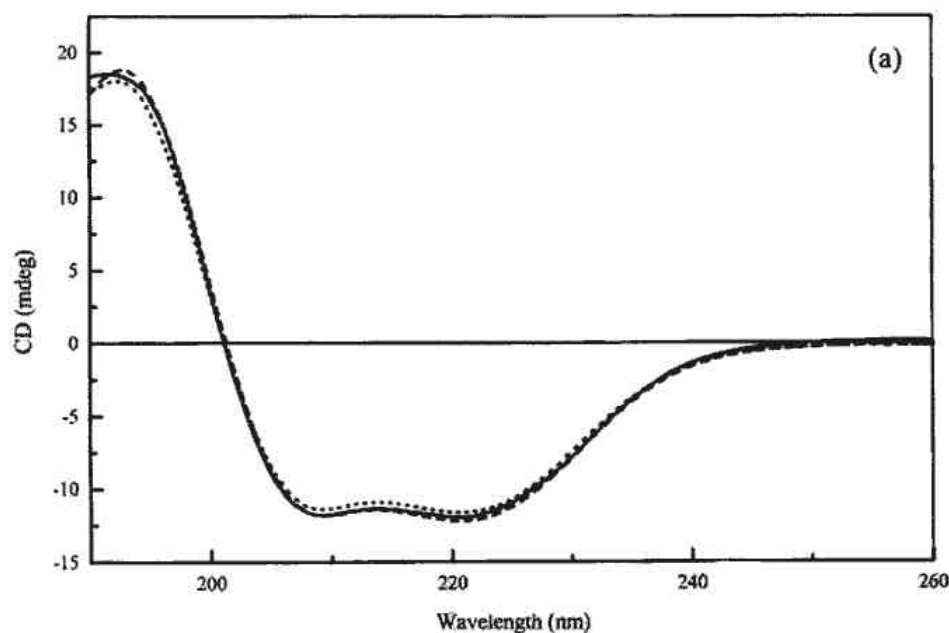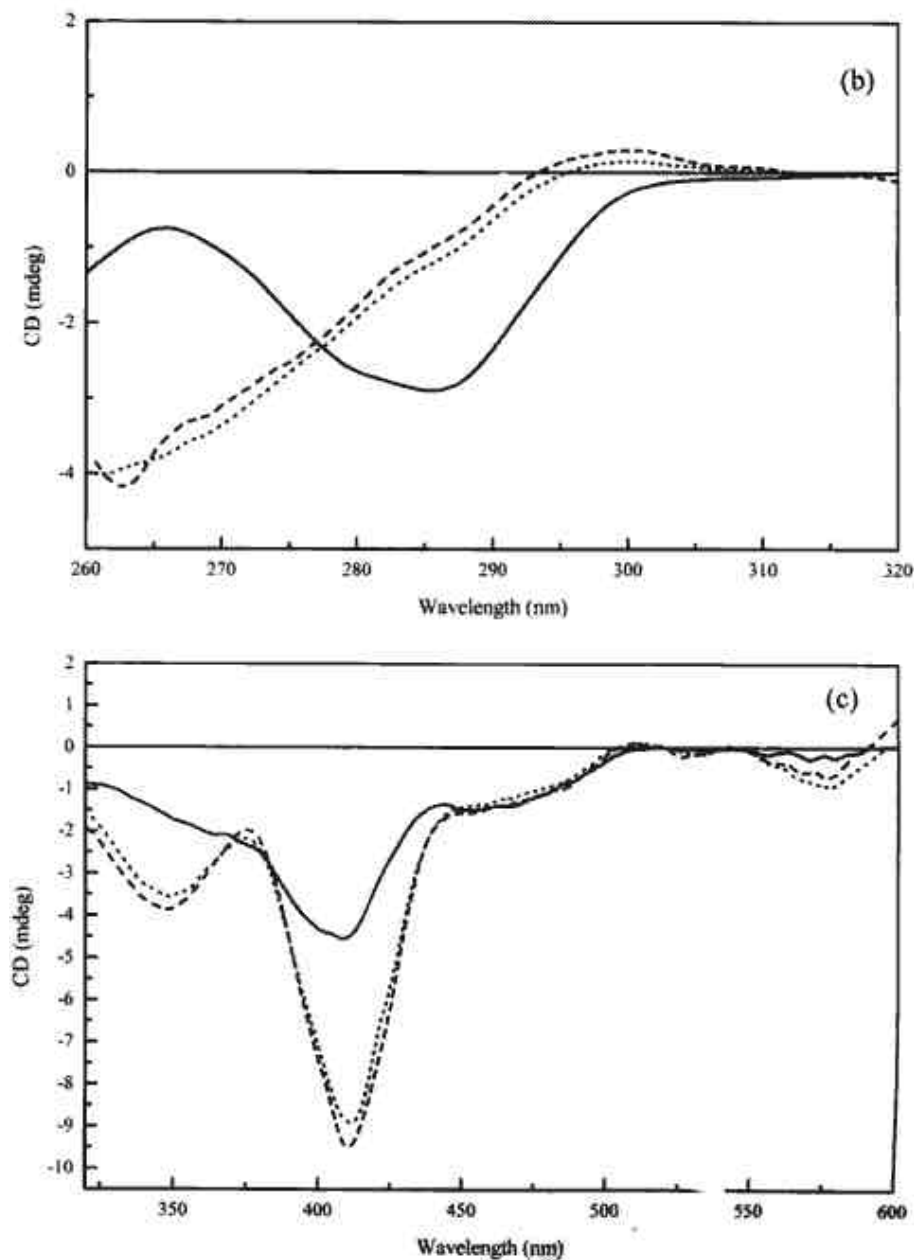
Fig. (11). CD spectra of BM3 flavocytochrome and its constituent domains [116]. In each panel the solid line represents the CD spectrum of the intact BM3 protein, the dotted line that of the stoichiometric mixture of the separately expressed P450 and reductase domains, and the dashed line the algebraic sum of the CD spectra of the separate domains. (a), (b) and (c) represent spectra recorded in the far UV, near UV and visible regions respectively.

native protein, the major developments in this field have been made by application of recombinant DNA techniques, which have allowed the preparation of large quantities of protein corresponding to defined domains. The individual domains can then be characterised and the interactions between them in the intact protein studied. CD has been widely used as a structural technique in such work, as illustrated by the examples below.

Phosphoglycerate kinase has a single polypeptide chain of 415 amino acids, divided into two domains each of approximately 200 amino acids which mostly correspond to the *N-* and *C-*terminal halves of the sequence, although the *C-*terminal 15 amino acids (401-415) form an integral part of the *N-*domain. In order to characterise the two separate domains of the enzyme from yeast, advantage has been taken of the fact that using recombinant DNA techniques each can be

produced in a folded form (as shown by CD) [114]. Using the CD signal at 218 nm to monitor the loss of structure caused by GdmCl, the unfolding of the intact enzyme and the isolated *N*- and *C*-domains was found in each case to be symmetric, monophasic and completely reversible. When the results were analysed according to a simple two-state model, the stabilities *(ΔG* (H$_2$O) of the intact protein, the *N*-domain and the *C*-domain were found to be 32.6, 18.8 and 16.7 kJ/mol respectively. Thus in terms of thermodynamic stability, the intact protein behaves essentially as the sum of its component domains. However, there is evidence from the reactivity of the single cysteine (Cys 97) in the enzyme that the *N*-domain is somewhat more flexible when isolated compared with the intact protein, pointing to the presence of at least limited interactions between the two domains.

The cytochrome P-450 BM3 system from *Bacillus megaterium* is an excellent model for mammalian P-450 systems, such as those involved in drug and xenobiotic metabolism. The two redox systems of the BM3 flavocytochrome, i.e. the flavin domain (consisting of FMN and FAD subdomains) and the P-450 haem-containing domain, are linked covalently in the intact polypeptide (119 kDa). The BM3 is a soluble enzyme and its constituent domains and sub-domains can be expressed in high yield [115].

Experiments in which the separate flavin and P-450 domains of the BM3 flavocytochromes were mixed in varying proportions showed that the rate of the overall reaction (as measured by the NADPH-dependent hydroxylation of dodecanoic acid) was at least 500-fold lower than the rate of the intact BM3 system. CD studies helped to shed some light on the reasons for this low rate of electron transfer between the redox centres. When the separate domains were physically mixed the CD spectra in the far UV, near UV and visible regions of the spectra were identical to those obtained by the algebraic addition of the spectra of the domains measured separately.

The far UV CD spectrum obtained for the intact BM3 (i.e. where the domains are covalently linked to each other) was identical to that for the physical mixture or the algebraic sum of the domains, indicating that the covalent linking has little effect on the overall secondary structure of the domains (Fig. **11**(a)). There were, however, significant differences in the CD spectra in the near UV (Fig. **11**(b)) and visible (Fig. **11**(c)) regions indicating that in intact BM3 there are characteristic environments of the aromatic amino acid side chains and flavin and haem cofactors. Although the CD spectra cannot define the nature of the changes in environments of these chromophores, it would be reasonable to presume that they have evolved to allow the efficient
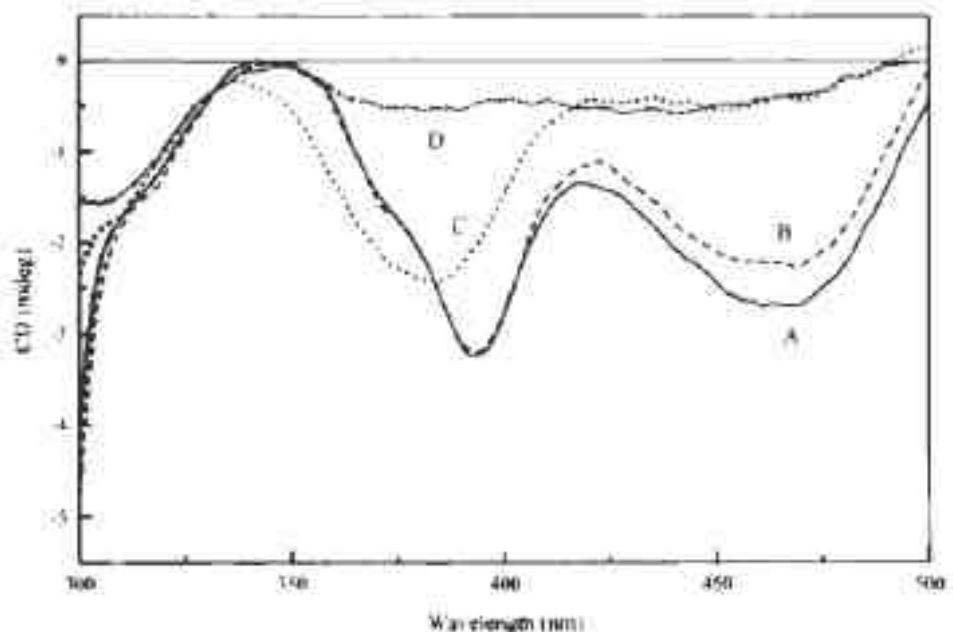

Fig. (12). The effect of GdmCl on the reductase domain of flavocytochrome BM3 as monitored by CD [117]. Spectra were recorded in the absence of GdmCl (A), and in the presence of 0.5 M (B), 1 M (C) and 3 M (D).

transfer of electrons in BM3 [116]. In further experiments with P-450 BM3, CD was used to compare the structural stability of the protein (as measured by far UV and near UV CD) with the stability of the binding sites for flavins (as measured by CD in the near UV and visible regions). The disappearance of one of the two visible CD bands was seen to occur over the range of denaturant (GdmCl) from 0.5–1M at which point all of the more weakly bound flavin (FMN) was lost from the enzyme. Between 1M and 2M GdmCl, all FAD was also dissociated, corresponding to the loss of the second visible CD band (Fig. **12**). All FMN was lost from the protein at a point where only very minor changes in secondary structure were observed; the results suggested that loss of this flavin represented a distinct step in the unfolding process of the enzyme [117].

CD has been used in conjunction with other biophysical techniques to examine the relationship between modules in a range of multi-module proteins. Many proteins of the immune system, the complement and clotting cascades and the extra-cellular matrix are of this type. Although in such proteins each module usually possesses a consensus sequence and thus a common structural frame-work, considerable diversity can be generated by variations in the non-conserved regions in the modules themselves and in the way in which modules are assembled. Studies have been made of fragments of the 220 kDa complement receptor protein type 1 (CR1), which consists of 30 modules. Recombinant soluble versions of this protein have been shown to have therapeutic value as inhibitors of complement. Attention has been focussed on fragments of CR1 which contain modules 15 to 17 as these encompass one of the two C3b-binding sites in the protein. The structure and stability of the 3-module fragment (15-17) was studied by a combination of NMR, differential scanning calorimetry, far UV CD and fluorescence measurements. These showed that on increasing temperature or concentration of denaturant (GdmCl) the loss of structure occurred in a number of discrete steps, with the junction 16-17 melting first, followed in turn by the 15-16 junction, module 17, and finally modules 15 and 16. The latter two modules adopted an intermediate state prior to complete unfolding [118]. These studies have been extended to examine the fragment consisting of modules 16 and 17, which have

confirmed the important role of module interactions in the CR1 system and indicated that the 15-16 interface is more extensive and stable than the 16-17 interface [119]. One interesting feature of the far UV CD spectra of these modular systems is that they show an unusual positive ellipticity at 240 nm. Although the exact cause of this signal is not known, it could well reflect the contributions of invariant core Trp side chains which are located on the conserved β-strands and which lie close to one of the two totally conserved disulphide bridges in each module [30,120]. CD has also been used in analogous studies on a related system, namely the 28 kDa Vaccinia virus complement control protein which consists of 4 domains of the type found in CR1. Analysis of the stability of a fragment containing the central modules (2 and 3) showed that these melted essentially independently, showing that the interface between these modules was not extensive [120].

4.7 The Interactions Between Chromophores in the Light Harvesting System

The primary events in photosynthesis by purple bacteria involve absorption of light energy by antenna light harvesting complexes and the rapid and efficient transfer of this energy to the reaction centre where it is trapped as chemical potential. Generally there are two types of light harvesting (LH) complexes. The first type (LH1) is very closely associated with the reaction centres in the core complex. Arranged peripherally around this core complex is the second type of light harvesting complex, designated LH2 [41].

In both LH1 and LH2 complexes the chromophores are bacteriochlorophyll and carotenoid molecules which are noncovalently linked to a protein scaffold. CD has been extensively used as a tool to characterise LH2 complexes and preparations spectroscopically; however, these studies have been given greater impetus following the determination of the crystal structure of the complex [121].

As shown in Fig. **13**, the structure of the LH2 complex from *Rhodopseudomonas acidophila* 10050, contains a number of novel features. The complex shows 9-fold symmetry, with each of the 9 units consisting of two small polypeptide chains (α and β, consisting of 53 and 41 amino acids
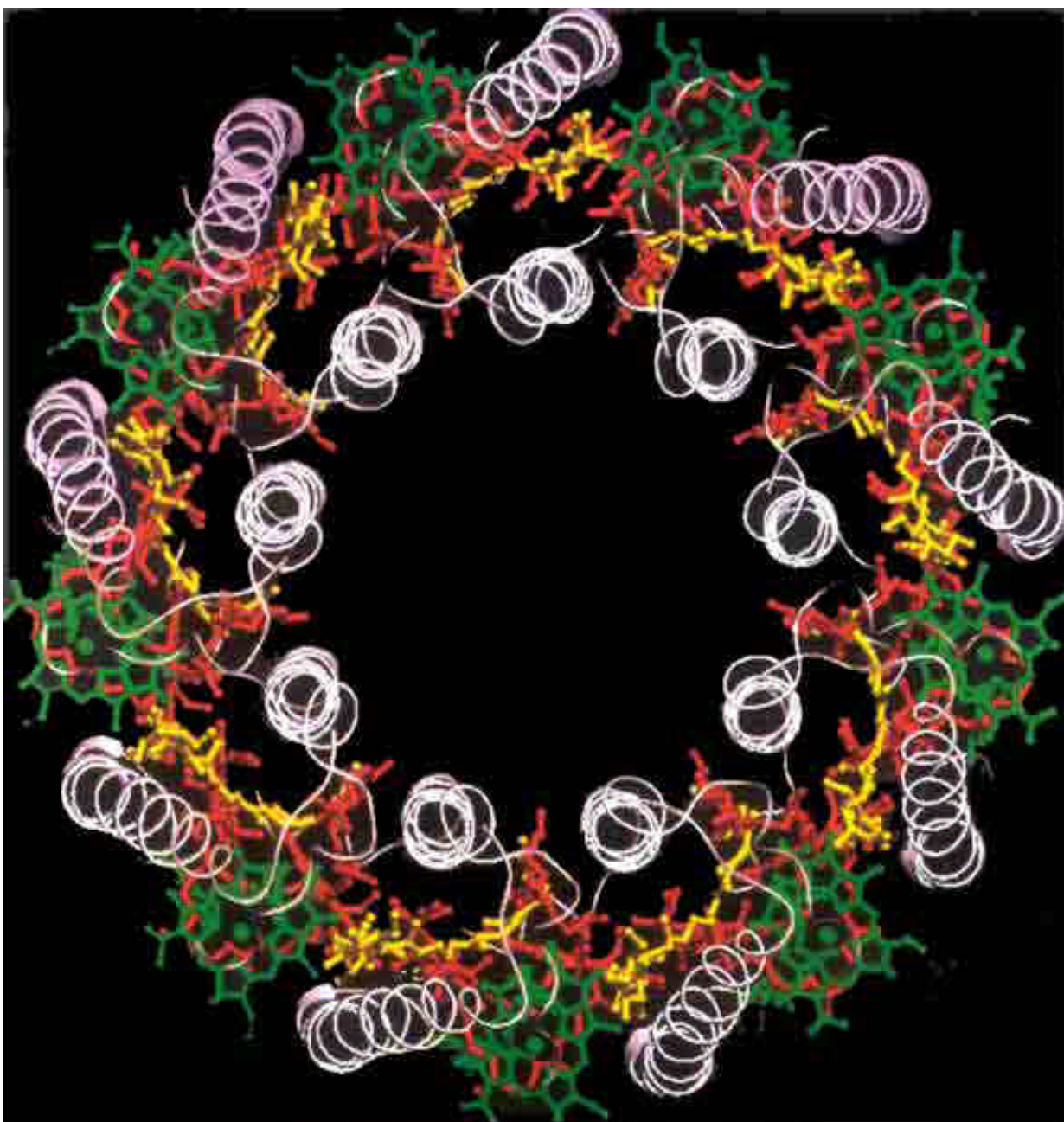
Fig. (13). The structure of the light harvesting antenna complex LH2 from *Rhodopseudomonas acidophila* [121] viewed perpendicular to the membrane. The α and β polypeptide chains are depicted as white coils. The bacteriochlorophyll *a* molecules are shown in red (B850, inner ring) and green (B800, outer ring). The carotenoid pigments (rhodopin glucoside) are shown in yellow.

respectively), 3 molecules of bacteriochlorophyll *a* (Bchl*a*) and one molecule of carotenoid (rhodopin glucoside). In the complex, the Bchl*a* molecules are of two distinct types. 18 of the total of 27 Bchl*a* molecules are in an inner ring sandwiched between the α and β polypeptide chains; the remaining 9 lie towards the cytoplasmic side of the membrane and are located between the transmembrane helices of the β polypeptide chains.

CD studies of the LH2 complex in the far UV region had shown that the α and β polypeptides had high (50-60%) α-helical content [41], as confirmed by the later X-ray structure [121]. However, the main area of interest for this review concerns the spectroscopic properties of the complex in the visible and near IR regions, since these underpin the biological function of the complex.

The absorption and CD spectra of LH2 are shown in Fig. **7**. The 3 main absorption peaks between 450 and 550 nm are due to the carotenoid molecules. The absorption peaks at 380 nm, 595 nm and in the 750 to 900 nm region are due to Bchl*a*. The peak at 801 nm is due to the outer set of 9 Bchl*a* (termed B800), and that at 863 nm is due to the inner set of 18 Bchl*a* (termed, somewhat misleadingly, B850).

The CD spectrum of the LH2 complex in the visible and near IR regions shows a number of interesting features. When the carotenoid is extracted and dissolved in organic solvents, there is no CD signal reflecting the symmetry of the molecule. The fact that a CD spectrum is observed in the intact LH2 complex is explained by the finding from the X-ray structure that the carotenoid is twisted; viewed along its long axis this degree of twist amounts to about half a turn of a helix, thereby generating chirality.

When Bchl*a* is extracted and dissolved in organic solvents (i.e. in a monomeric form), it shows an absorption maximum in the near IR at 772 nm; this represents the Q_y absorption transition. In the LH2 complex, not only is this absorption band red shifted, but it is also split into two. These effects arise from interactions between the polypeptides and the Bchl*a* molecules and between the Bchl*a* molecules themselves. The absorption at 800 nm is due to the outer ring of 9 Bchl*a* (B800), and that at 850 nm is due to the inner ring of 18 Bchl*a*, where there is extensive overlap of the orbitals to give exciton coupling (see sections 1 and 3.2). This coupling is the basis of the characteristic CD spectrum in this region where a positive CD peak is accompanied by a negative peak at longer wavelengths, with a cross over point to the red of the absorption maximum.

The availability of the high resolution structure of the LH2 complex [121] has allowed meaningful attempts to relate theory and experiment to be undertaken. Semiempirical quantum mechanical calculations using the QCFF/PI (quantum mechanical consistent force field/π-electron) approach incorporating exciton and charge-transfer interactions have been able to reproduce the main features of the absorption and CD spectra [122]. The results of these calculations supported the proposal that there was extensive overlap between the excited states of the B850 Bchl*a*s, whereas the

excited states of the B800 Bchl*a*s were much more localised.

CD can also be used to assess the structural integrity of LH2 complexes. The outer ring of Bchl*a* (B800) can be selectively removed by treatment with the non-ionic Triton detergent TBG10, followed by ion-exchange chromatography. Release of B800 was confirmed by the disappearance of the absorption peak at 800 nm. Reconstitution of the LH2 complex could be achieved by incubating the B850-only complexes with Bchl*a* in a variety of detergent systems including octyl-glucoside, dodecylmaltoside (laurylmaltoside) or Triton, followed by removal of excess Bchl*a* by ion-exchange chromatography [123]. Comparison of the CD spectra of native LH2, B850-only complexes and reconstituted complexes in both visible and near IR spectral regions showed that the structures of the inner core of carotenoids and Bchl*a*-B850 were very similar (indicating that the overlap of orbitals in the excited state of the latter had been preserved). In addition the CD of the reconstituted and native complexes in the region around 800 nm showed that the Bchl*a*-B800 molecules had bound in the correct orientation in the reconstituted complexes.

4.8 Ligand Binding and Drug Design

As mentioned in section 3.4.6, the binding of ligands to proteins can give rise to a CD signal from the bound ligand; this signal is induced because the ligand is bound in an asymmetric fashion. The finding that derivatives of azosulphonamide can discriminate between the active sites of the different isoenzymes of carbonic anhydrase [42] shows that CD has the potential to report on detailed differences in the mode of binding of ligands to proteins. However, except in a few cases such as that of the haem cofactor in a series of cytochrome P450s [39] and the binding of polyaromatic-containing compounds to DNA [124], it has not generally been possible to draw detailed structural conclusions at an atomic level about the mode of ligand binding.

As described in section 4.2, the binding of ligands can lead to conformational changes in proteins which are essential for their biological function. Such changes can be detected by far UV

and/ or near UV CD and changes in the spectroscopic signals can be used to determine the binding parameters (stoichiometry, dissociation constant, degree of cooperativity etc.).

There has been increasing interest in recent years in the possibility of using the wealth of structural information available about proteins in order to design suitable inhibitors which could act as specific targetted drugs or at least as potential lead compounds in the search for specific drugs; this field has been termed structure assisted (or rational) drug design [125]. Some notable successes have been achieved in this area, for example:- (a) the design of inhibitors of angiotensin-converting enzyme; these inhibitors are used extensively as anti-hypertensive agents, and (b) inhibitors of the HIV type 1 protease which has been shown to be essential for virion assembly and maturation and is therefore considered an excellent target for inhibitory drug design [126,127]. The design process involves the combination of structural information from X-ray or NMR and detailed computational modelling of the ligand-protein interaction [125-128] in order to assess the important structural features associated with strong, specific interactions. Up to now CD has not played a particularly significant role in this process, primarily because the structural information provided by CD is generally of lower resolution.

However, some indication of the role which CD could play in the drug design area is illustrated by recent work on other potential target proteins crucial to the process of HIV-1 viral infection. These targets include the glycosylated coat proteins (gp120 and gp41) which remain on the surface of the viral particle and are involved in the binding and fusion with host receptor cells. Experiments have shown that the gp41-gp120 complex undergoes a conformational change following binding of the gp120 protein to the target cell [129]. As this conformational change appears to be a crucial step in the conversion of gp41 into a fusion-competent form, it was concluded that prevention of this change would be an excellent strategy in the development of successful therapeutic agents against viral pathogenesis. X-ray crystallography studies on gp41 showed that the core of its ectodomain was a six-helix bundle comprising a trimeric α-helical core [130]. Each helical hairpin was found to consist of a coiled coil N helix with a hydrophobic groove into which was packed an anti-parallel C helix. This conserved hydrophobic groove was targeted as a potential interaction site for antiviral therapeutic compounds which would interfere with the pairing of the C-helices and so prevent the conformational change necessary for membrane fusion. Evidence for the importance of this groove was provided by experiments in which the behaviour of synthetic peptides corresponding to the amino acids of the C-helix portion of the helical hairpins was studied [131]. The synthetic wild type C-terminal peptide, C34 (comprising residues 628-661 of the C-helix) effectively inhibited viral infection. Studies of the helical content and thermal stability using CD were employed to examine the effects of mutating amino acids involved in contact with the N-helix in the hydrophobic groove. These mutations were shown to decrease the helical content and thermal stability of the N36 (residues 546-581 of the N-helix region) /C34 complex generated in each case. Conclusions drawn from these studies highlighted the correlation between the effective binding of the C34 peptide to the N-helix and the inhibition of viral fusion. Further studies using the C peptide as a lead compound and a combinatorial library of 61,275 potential ligands have led to the identification of non-natural ligands which when attached to a peptide (consisting of residues from the C peptide) bind to the hydrophobic groove thus inhibiting cell fusion by the prevention of activation of gp41 [132]. The similarity of the fusion-active HIV-1 gp41 core with fusion mediating proteins from other viruses provides a basis for future strategies in the design of therapies for many viral-induced diseases.

5. SUMMARY AND CONCLUSIONS

CD is a structural technique which plays a very important role in complementing the higher resolution structural approaches of X-ray crystallography and NMR. The strengths of CD are its economy of sample required, speed of acquisition of data and ability to perform time-resolved measurements. This review has highlighted a number of the important theoretical and practical aspects of designing suitable experiments to be addressed by CD and illustrated the wide range of potential applications of the technique in the study of protein structure and function. Such applications are set to increase

considerably in future years as the range of proteins amenable to study expands. This growth will reflect the identification of greater numbers of genes (and pseudogenes) from genome sequencing projects and of the development of improved expression systems for such genes. CD will play an important part in the characterisation of these gene products.

ACKNOWLEDGEMENTS

ABBREVIATIONS

ANS   =   1-Anilino-8-napthalenesulphonate

GdmCl =   Guanidinium chloride

SDS   =   Sodium dodecylsulphate

FTIR   =   Fourier transform infra red spectroscopy

DMPC=   L-α-1,2-Dimyristoylphos-phatidylcholine

DHPC =   L-α-1,2-Dihexanoylphos-phatidylcholine

DMPG=   L-α-1,2-Dihexanoylphos-phatidylglycerol

REFERENCES

[1]   Woody, R.W. and Dunker, A.K. (**1996**) in Circular Dichroism and the Conformational Analysis of Biomolecules (Fasman, G.D., ed.). pp. 109-157, Plenum Press, New York.

[2]   Woody, R.W. (**1995**) *Methods Enzymol.*, *246*, 34-71.

[3]   Greenfield, N.J. (**1999**) *Trends Analyt. Chem.*, *18*, 236-244.

[4]   Kelly, S.M. and Price, N.C. (**1997**) *Biochim. Biophys. Acta*,*1338*, 161-185.

[5]   Fasman, G.D. (ed.) (**1996**) Circular Dichroism and the Conformational Analysis of Biomolecules 738 pp., Plenum Press, New York.

[6]   Johnson, W.C., Jr. (**1990**) *Proteins: Struct., Funct. Genet. 7,*205-214.

[7]   Martin, S.R. (**1996**) in Proteins Labfax (Price, N.C., ed.). pp. 195-204, Bios Scientific Publishers, Oxford.

[8]   Johnson, W.C. jr. (**1996**) in Circular Dichroism and the Conformational Analysis of Biomolecules (Fasman, G.D., ed.). pp. 635-652, Plenum Press, New York.

[9]   Sutherland, J.C. (**1996**) in *Circular Dichroism and the Conformational Analysis of Biomolecules* (Fasman, G.D., ed.). pp. 599-633, Plenum Press, New York.

[10]   Clarke, D.T. and Jones, G.R. (**1999**) *Biochemistry 38*, 10457-10462.

[11]   Price, N.C. (**1996**) in Enzymology Labfax (Engel, P.C., ed.) pp. 34-41, Bios Scientific Publishers, Oxford.

[12]   Strickland, E.H. (**1974**) *Crit. Rev. Biochem.*, *2*, 113-175.

[13]   Kuwajima, K. (**1996**) in Circular Dichroism and the Conformational Analysis of Biomolecules (Fasman, G.D., ed.). pp. 159-182, Plenum Press, New York.

[14]   Chaffotte, A.F., Guillou, Y. and Goldberg, M.E. (**1992**) *Biochemistry ,31*, 9694-9702.

[15]   Wess, T.J. (**1997**) *Biotechnol. Appl. Biochem.*, *26*, 127-142.

[16]   Drenth, J. (**1994**) Principles of Protein Crystallography, Springer-Verlag, Berlin.

[17]   James, T.L. and Oppenheimer, N.J. (eds.) (**1994**) Nuclear Magnetic Resonance Part C, Methods Enzymol. vol 239, 813 pp., Academic Press, San Diego.

[18]   Wharton, C.W. (**1996**) in Proteins Labfax (Price, N.C., ed.). pp. 187-194, Bios Scientific Publishers, Oxford.

[19]   Weiss, M.A., Ellenberger, T., Wobbe, C.R., Lee, J.P., Harrsion, S.C. and Struhl, K. (**1990**) *Nature*, *347*, 575-578.

[20]   Myers, J.K. and Oas, T.G. (**1999**) *J. Molec. Biol.*, *289*, 205-209.

[21]   Fezoui, Y., Weaver, D.L. and Osterhout, J.J. (**1994**) *Proc. Natl. Acad. Sci.*, U.S.A. **91**, 3675-3679.

[22]   Ribas de Pouplana, L, Atrian, S., Gonzalez-Duarte, R., Fothergill-Gilmore, L.A., Kelly, S.M. and Price, N.C. (**1991**) *Biochem. J.*, *276*, 433-438.

[23]   Kahn, P.C. (**1979**) *Methods Enzymol, 61*, 339-378.

[24]   Robertson, A.G.S. and Nimmo, H.G. (**1995**) *Biochem. J.*,*305*, 239-244.

[25] Nairn, J., Price, N.C., Kelly, S.M., Rigden, D., Fothergill-Gilmore, L.A. and Krell, T. (**1996**) *Biochim. Biophys. Acta*,*1296*, 69-75.

[26] Jaenicke, R. (**1987**) *Prog. Biophys. Mol. Biol.*, *49*, 117-237.

[27] Venyaminov, S.Yu. and Yang, J.T. (**1996**) in Circular Dichroism and the Conformational Analysis of Biomolecules (Fasman, G.D., ed.). pp. 69-107, Plenum Press, New York.

[28] Griffin, J.H., Rosenbusch, J.P., Weber, K.K. and Blout, E.R. (**1972**) *J. Biol. Chem., 247*, 6482-6490.

[29] Freskgård, P.-O., Mårtensson, L.-G., Jonasson, P., Jonsson, B.-H. and Carlsson, U. (**1994**) *Biochemistry* ,*33*, 14281-14288.

[30] Craig, S., Pain, R.H., Schmeissner, U., Virden, R. and Wingfield, T.P. (**1989**) *Int. J. Peptide Protein Res., 33*, 256-262.

[31] Neubert, L.A. and Carmack, M. (**1974**) *J. Amer. Chem. Soc*.,*96*, 943-945.

[32] Kishore, R., Raghothama, S. and Balaram, P. (**1988**) *Biochemistry* , *27*, 2462-2471.

[33] Johnson, L.N. and Barford, D. (**1990**) *J. Biol. Chem., 265*, 2409-2412.

[34] West, S.M. and Price, N.C. (**1989**) *Biochem. J.*, *261*, 189-196.

[35] West, S.M. and Price, N.C. (**1990**) *Biochem. J., 265*, 45-50.

[36] Price, N.C. and Stevens, E. (**1983**) *Biochem. J.*, *213*, 595-602.

[37] Macheroux, P. (**1999**) in Methods in Molecular Biology vol. 131: Flavoprotein Protocols (eds. Chapman, S.K. and Reid, G.A.) Humana Press, Totowa, New Jersey, pp. 1-7.

[38] Munro, A.W., Kelly, S.M. and Price N.C. (**1999**) in Methods in Molecular Biology vol. 131: Flavoprotein Protocols (eds. Chapman, S.K. and Reid, G.A.) Humana Press, Totowa, New Jersey, pp. 111-123.

[39] Andersson, L.A. and Peterson, J.A. (**1995**) *Biochem. Biophys. Res. Commun., 211*, 389-395.

[40] Cogdell, R.J. and Scheer, H. (**1985**) *Photochem. Photobiol.*, *42*, 669-689.

[41] Cogdell, R.J., Isaacs, N.W., Freer, A.A., Arrelano, J., Howard, T.D., Papiz, M.Z., Hawthornthwaite-Lawless, A.M. and Prince, S. (**1997**) *Prog. Biophys. Molec. Biol.*, *68*, 1-27.

[42] Coleman, J.E. (**1968**) *J. Biol. Chem., 243*, 4574-4587.

[43] Johnson, W.C., Jr. (**1996**) in Circular Dichroism and the Conformational Analysis of Biomolecules (Fasman, G.D., ed.). pp. 433-468, Plenum Press, New York.

[44] Gray, D.M. (**1996**) in Circular Dichroism and the Conformational Analysis of Biomolecules (Fasman, G.D., ed.). pp. 469-500, Plenum Press, New York.

[45] Rodger, A. and Nordén, B. (**1997**) Circular dichroism and linear dichroism. Oxford University Press, Oxford. pp. 15-32.

[46] Hirst, J.D. and Brooks, C.L.,III (**1994**) *J. Mol. Biol., 243*, 173-178.

[47] Provencher, S.W. and Glöckner, J. (**1981**) *Biochemistry* ,*20*, 33-37.

[48] Manavalan, P. and Johnson, W.C.,Jr. (**1987**) *Analyt. Biochem., 167*, 76-85.

[49] Johnson, W.C., Jr. (**1999**) *Proteins: Struct., Funct. Genet., 35*, 307-312.

[50] Barrow, C.J., Yasuda, A., Kenny, P.T.M. and Zagorski, M.G. (**1992**) *J. Mol. Biol., 225*, 1075-1093.

[51] Yang, J.T., Wu, C.-S.C. and Martinez, H.M. (**1986**) *Methods Enzymol., 130*, 208-269.

[52] Greenfield, N.J. (**1996**) *Analyt. Biochem., 235*, 1-10.

[53] Pace, C.N. (**1986**) *Methods Enzymol., 131*, 266-280.

[54] Hope, J., Shearman, M.S., Baxter, H.C., Chong, A., Kelly, S.M. and Price, N.C. (**1996**) *Neurodegeneration* ,*5*, 1-11.

[55] Jackson, G.S., Hosszu, L.L.P., Power, A., Hill, A.F., Kenney, J., Saibil, H., Craven, C.J., Waltho, J.P., Clarke, A.R. and Collinge, J. (**1999**) *Science*, *283*, 1935-1937.

[56] Price, N.C.and Stevens, L. (**1999**) Fundamentals of Enzymology: the Cell Biology of Catalytic Proteins, 3rd edition, Oxford University Press, pp. 47-117.

[57] Glazer, A.N. and Simmons, N.S. (**1966**) *J. Amer. Chem. Soc*., *88*, 2335-2336).

[58] Smith, L., Greenfield, N.J. and Hitchcock-DeGregori, S.E. (**1994**) *J. Biol. Chem. 269*, 9857-9863.

[59] Martin, S.R. and Bayley, P.M. (**1986**) *Biochem. J.*, *238*, 485-490.

[60] White, Z.W., Fisher, K.E. and Eisenstein, E. (**1995**) *J. Biol. Chem., 270*, 20404-20409.

[61] Bell, J.E. and Dalziel, K. (**1973**) *Biochim. Biophys. Acta, 309*, 237-242.

[62] Perutz, M.F., Ladner, J.E., Simon, S.R. and Ho, C. (**1974**) *Biochemistry* , *13*, 2163-2173.

[63] Dolgikh, D.A., Gilmanshin, R.I., Brazhnikov, E.V., Bychkova, V.E., Semisotnov, G.V., Venyanimov, S. Yu. and Ptitsyn, O.B. (**1981**) *FEBS Lett., 136*, 311-315.

[64] Christensen, H. and Pain, R.H. (**1991**) *Eur. J. Biophys.*, *19*, 221-229.

[65]   Ptitsyn, O.B. (**1995**) *Adv. Protein Chem., 47*, 83-229.

[66]   Christensen, H. and Pain, R.H. (**1994**) in Mechanisms of Protein Folding (Pain, R.H., ed.), pp. 55-79, Oxford University Press, Oxford.

[67]   Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A.L. and Hartl, F.-U. (**1991**) *Nature ,352*, 36-42.

[68]   Flynn, G.C., Beckers, C.J.M., Baase, W.A. and Dahlquist, F.W. (**1993**) *Proc. Natl. Acad. Sci., 90*, 10826-10830.

[69]   Hornby, D.P., Whitmarsh, A., Pinabarsi, H. Kelly, S.M., Price, N.C., Shore, P.D., Baldwin, G.S. and Waltho, J. (**1994**) *FEBS Lett.*, *355*, 57-60.

[70]   Kirkitadze, M.D., Barlow, P.N., Price, N.C., Kelly, S.M., Boutell, C.J., Rixon, F.J. and McClelland, D.A. (**1998**) *J. Virol.*, *72*, 10066-10072.

[71]   Tuma, R. and Thomas, G.J. (**1997**) *Biophys. Chem., 68*, 17-31.

[72]   Thatcher, D.R. and Hitchcock, A. (**1994**) in Mechanisms of Protein Folding (Pain, R.H., ed.), pp. 229-261, Oxford University Press, Oxford.

[73]   Radford, S.E. and Dobson, C.M. (**1999**) *Cell*, *97*, 291-298.

[74]   Levinthal, C. (**1968**) *J. Chim. Phys.*, *65*, 44-45.

[75]   Jaenicke, R. (**1991**) *Biochemistry, 30*, 3147-3161.

[76]   Price, N.C. (**1994**) in Mechanisms of Protein Folding (Pain, R.H., ed.), pp. 160-193, Oxford University Press, Oxford.

[77]   Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S. and Nagamura, T. (**1987**) *FEBS Lett., 221*, 115-118.

[78]   Gilmanshin, R.I. and Ptitsyn, O.B. (**1987**) *FEBS Lett.*, *223*, 327-329.

[79]   Jennings, P.A. and Wright, P.E. (**1993**) *Science ,262*, 892-896.

[80]   Kalnin, N.N. and Kuwajima, K. (**1995**) *Proteins: Struct. Funct. Genet.*, *23*, 163-176.

[81]   Jacob, M.D. and Fox, R.O. (**1994**) *Proc. Natl. Acad. Sci.,*U.S.A. *91*, 449-453.

[82]   Radford, S.E., Dobson, C.M. and Evans, P.A. (**1992**) *Nature ,358*, 302-307.

[83]   Dobson, C.M., Evans, P.A. and Radford, S.M. (**1994**) *Trends Biochem. Sci., 19*, 31-37.

[84]   Guijarro, J.I., Jackson, M., Chaffotte, A.F., Delepierre, M., Mantsch, H.H. and Goldberg, M.E. (**1995**) *Biochemistry, 34*, 2998-3008.

[85]   Agashe, V.R., Shastry, M.C.R. and Udgaonkar, J.B. (**1995**) *Nature*, *377*, 754-757.

[86]   Aggeli, A., Bell, M., Boden, N, Keen, J.N., Knowles, P.F., McLeish, T.C.B., Pitkeathly, M. and Radford, S.E. (**1997**) *Nature*, *386*, 259-262.

[87]   Hungerford, G., Marinez-Insua, M., Birch, D.J.S. and Moore, B.D. (**1996**) *Angew. Chem. Int. Ed. Engl.*, *35*, 326-329.

[88]   Dyson, H.J. and Wright, P.E. (**1991**) *Annu. Rev. Biophys. Biophys. Chem., 20*, 519-538.

[89]   Scholtz, J.M., Barrick, D., York, E.J., Stewart, J.M. and Baldwin, R.L. (**1995**) *Proc. Natl. Acad. Sci.,* U.S.A. *92*, 185-189.

[90]   Zhang, S. and Rich, A. (**1997**) *Proc. Natl. Acad. Sci.*, USA. *94*, 23-28.

[91]   Janek, K., Behlke, J., Zipper, J., Fabian, H., Georgalis, Y., Beyermann, M., Bienert, M. and Krause, E. (**1999**) *Biochemistry, 38*, 8246-8252.

[92]   Harrison, S.C. (**1991**) *Nature*, *353*, 715-719.

[93]   Zitzewitz, J.A., Bilsel, O., Luo, J., Jones, B.E. and Matthews, C.R. (**1995**) *Biochemistry, 34*, 12812-12819.

[94]   Jiang, X., Bishop, E.J. and Farid, R.S. (**1997**) *J. Amer. Chem. Soc., 117*, 838-

[95]   Sharp, R.E., Moser, C.C., Rabanal, F. and Dutton, P.L. (**1998**) *Proc. Natl. Acad. Sci.,* USA, *95*, 10465-10470.

[96]   Iwata, S., Ostermeier, C., Ludwig, B and Michel, H. (**1995**) *Nature, 376*, 660-669.

[97]   Tsukihara, T., Aoyama, H., Yamashita, E., Tomazaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R. and Yoshikawa, S. (**1996**) *Science*, *272*, 1136-1144.

[98]   Garavito, R.M. (**1998**) *Curr. Opin. Biotechnol.*, *9*, 344-349.

[99]   Henderson, R. and Unwin, P.N.T. (**1975**) *Nature, 257*, 28-32.

[100]   Cogdell, R.J. and Lindsay, J.G. (**1996**) Proteins Labfax (Price, N.C., ed.). pp. 101-107, Bios Scientific Publishers, Oxford.

[101]   Ahmed, T., Isaacs, N.W., Freer, A.A., Griffiths, I.R., Kikham, D., Lawrence, A.J., Price, N.C. and Kelly, S.M. (**1997**) *FASEB J, 11*, no. 9 SS, p. 1383.

[102]   Luecke, H., Schobert, B., Richter, H.T., Cartailler, J.P. and Lanyi, J.K. (**1999**) *J. Mol. Biol.*, *291*, 899-911

[103]   Wallace, B.A. and Teeters, C.L. (**1987**) *Biochemistry, 26*, 65-70.

[104]   Fasman, G.D. (**1996**) in Circular Dichroism and the Conformational Analysis of Biomolecules (Fasman, G.D., ed.). pp. 381-412, Plenum Press, New York.

[105]   Swords, N.A. and Wallace, B.A. (**1993**) *Biochem. J.*, *289*, 215-219.

[106] Booth, P.J. and Curran, A.R. (**1999**) *Curr. Opin. Struct. Biol.*, *9*, 115-121.

[107] Riley, M.L., Wallace, B.A., Flitsch, S.L. and Booth, P.J. (**1997**) *Biochemistry*, *36*, 192-196.

[108] Hunt, J.F., Earnest, T.N., Bousche, O., Kalghati, K., Reilly, K., Horváth, C., Rothschild, K.J. and Engelman, D.M.(**1997**) *Biochemistry*, *36*, 15156-15176.

[109] Engelman, D.M. and Steitz, T.A. (**1981**) *Cell*, *23*, 411-422.

[110] Weiss, M.S., Kreusch, A., Schiltz, E., Nestel, U., Welte, W., Weckesser, J and Schultz, G.E. (**1991**) *FEBS Letters*, *280*, 379-382.

[111] Surrey, T and Jähnig, F. (**1995**) *J. Biol. Chem.*, *270*, 28199-28203.

[112] Corbin, J., Méthot, N., Wang, H.H., Baenziger, J.E. and Blanton, M.P. (**1998**) *J. Biol. Chem.*, *273*, 771-777.

[113] Fuertes, M.A., Berberich, C., Lozano, R.M., Gimenez-Gallego, G. and Alonso, C. (**1999**) *Eur. J. Biochem.*, *260*, 559-567.

[114] Missiakis, D., Betton, J.-M., Minard, P. and Yon, J.M. (**1990**) *Biochemistry*, *29*, 8683-8689.

[115] Miles, J.S., Munro, A.W., Rospendowski, B.N., Smith, W.E., McKnight, J. and Thomson, A.J. (**1992**) *Biochem. J.*, *288*, 503-509.

[116] Munro, A.W, Lindsay, J.G., Coggins, J.R., Kelly, S.M. and Price, N.C. (**1994**) *FEBS Lett.*, 343, 70-74.

[117] Munro, A.W., Lindsay, J.G., Coggins, J.R., Kelly, S.M. and Price, N.C. (**1996**) *Biochim. Biophys. Acta.*, *1296*, 127-137.

[118] Kirkitadze, M.D., Krych, M., Uhrin, D., Dryden, D.T.F., Smith, B.O., Wang, X., Hauhart, R., Atkinson, J.P. and Barlow, P.N.(**1999**) *Biochemistry*, *38*, 7019-7031.

[119] Kirkitadze, M.D., Dryden, D.T.F., Kelly, S.M., Price, N.C., Wang, X., Krych, M., Atkinson, J.P. and Barlow, P.N. (**1999**) *FEBS Lett.*, *459*, 133-138.

[120] Kirkitadze, M.D., Henderson, C. Price, N.C., Kelly, S.M., Mullin, N.P., Parkinson, J., Dryden, D.T.F. and Barlow, P.N. (**1999**) *Biochem. J.*, *344*, 167-175.

[121] McDermott, G., Prince, S.M., Freer, A.A., Hawthornethwaite-Lawless, A.M., Papiz, M.Z., Cogdell, R.J. and Isaacs, N.W. (**1995**) *Nature*, *374*, 517-521.

[122] Alden, R.G., Johnson, E., Nagarajan, V., Parson, W.W., Law, C.J. and Cogdell, R.J. (**1997**) *J. Phys. Chem. B.*, *101* 4667-4680.

[123] Fraser, N.J., Dominy, P.J., Ücker, B, Simonin, I., Scheer, H. and Cogdell, R.J. (**1999**) *Biochemistry*, *38*, 9684-9692.

[124] Rodger, A., Taylor, S., Adlam, G, Blagborough, I.S., and Haworth, I.S. (**1995**) *Bioorg. Med. Chem.*, *3*, 861-872.

[125] Appelt, K., Bacquet, R.J., Bartlett, C.A., Booth, C.L.J., Freer, S.T., Fuhry, M.A.M. et al. (**1991**) *J. Med. Chem.*, *34*, 1925-1934.

[126] Perutz, M.F. (**1992**) in Protein Structure: New Approaches to Disease and Therapy, Freeman, New York, pp. 139-163.

[127] Wlodawer, A. and Vondrasek, J. (**1998**) *Ann. Rev. Biophys. Biomol. Struct.*, *27*, 249-284.

[128] Blundell, T.L. (**1994**) *Trends Biotech.*, *12*, 145-148.

[129] Jones, P.L., Korte, T.and Blumenthal, R. (**1998**) *J. Biol. Chem.*, *273*, 404-409.

[130] Chan, D.C., Fass, D., Berger, J.M. and Kim, P.S. (**1997**) *Cell*, *89*, 263-273.

[131] Chan, D.C., Chutowski, C.T. and Kim, P.S. (**1998**) *Proc. Natl. Acad. Sci.*, *95*, 15613-15617.

[132] Ferrer, M., Kapoor, T.M., Strassmaier, T., Weissenhorn, W., Skehel, J.J., Oprian, D., Schreiber, S.L., Wiley, D.C. and Harrison, S.C. (**1999**) *Nature Struct. Biol.*, *6*, 953-960.