

Review

How to study proteins by circular dichroism

Sharon M. Kelly, Thomas J. Jess, Nicholas C. Price*

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

Received 21 April 2005; received in revised form 8 June 2005; accepted 8 June 2005

Available online 5 July 2005

Abstract

Circular dichroism (CD) is being increasingly recognised as a valuable technique for examining the structure of proteins in solution. However, the value of many studies using CD is compromised either by inappropriate experimental design or by lack of attention to key aspects of instrument calibration or sample characterisation. In this article, we summarise the basis of the CD approach and its application to the study of proteins, and then present clear guidelines on how reliable data can be obtained and analysed.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Circular dichroism; Protein structure; Secondary structure; Protein folding; Ligand binding

1. Introduction

Since the late 1980s, there has been an explosive growth in structural biology with the number of high resolution structures of proteins added to the protein data bank (PDB) currently growing at more than 2000 per year. This has allowed much more detailed insights into the function of systems of ever-increasing size, including complex cellular assemblies such as the proteasome [1] and the large ribosomal subunit [2]. To a large extent, the growth in structural biology has been driven by developments in recombinant DNA technology which allow proteins to be produced in the (often substantial) quantities required, as well as by advances in data analysis and bioinformatics. However, there is a growing realisation of the need to perform structural studies under the conditions in which proteins actually operate (i.e., generally in solution), as well as under other conditions and to provide measures of the rates of structural changes of proteins, which are often essential to their biological function. Circular dichroism (CD) has become increasingly recog-

nised as a valuable structural technique for addressing these issues. A significant improvement in the provision of CD instrumentation has occurred in recent years; unfortunately, it is not always clear that such instruments are being used to their best advantage. The aim of this article is to provide a brief summary of the CD technique and its applications with particular reference to the study of proteins. It will then go on to address the important practical aspects of performing CD experiments on proteins and provide clear guidance as to how reliable data can be obtained and interpreted. We hope that the article will help users to avoid most of the common errors which, regrettably, occur all too frequently in the published literature. This article is confined to the CD of electronic transitions in molecules (electronic CD, ECD); for details of more specialist aspects of CD such as vibrational CD (VCD) or fluorescence-detection CD (FD CD) other articles should be consulted [3,4].

2. Origin of the CD effect

Plane polarised light can be viewed as being made up of 2 circularly polarised components of equal magnitude, one rotating counter-clockwise (left handed, L) and the

* Corresponding author. Tel.: +44 141 330 2889; fax: +44 141 330 4620.
E-mail address: n.price@bio.gla.ac.uk (N.C. Price).

other clockwise (right handed, R). Circular dichroism (CD) refers to the differential absorption of these 2 components (see Fig. 1). If, after passage through the sample being examined, the L and R components are not absorbed or are absorbed to equal extents, the recombination of L and R would regenerate radiation polarised in the original plane (Fig. 1). However, if L and R are absorbed to different extents, the resulting radiation would be said to possess elliptical polarisation (Fig. 1). A CD signal will be observed when a chromophore is chiral (optically active) for one of the following reasons: (a) it is intrinsically chiral because of its structure, for example, a C atom with 4 different substituents, or the disulphide bond which is chiral because of the dihedral angles of the C–S–S–C chain of atoms, (b) it is covalently linked to a chiral centre in the molecule, or (c) it is placed in an asymmetric environment by virtue of the 3-dimensional structure adopted by the molecule. The theoretical basis of the CD technique and its application to the study of other types of molecules have been well covered in a number of books and review articles [5–8].

CD instruments (known as spectropolarimeters) measure the difference in absorbance between the L and R circularly polarised components ($\Delta A = A_L - A_R$), but will generally

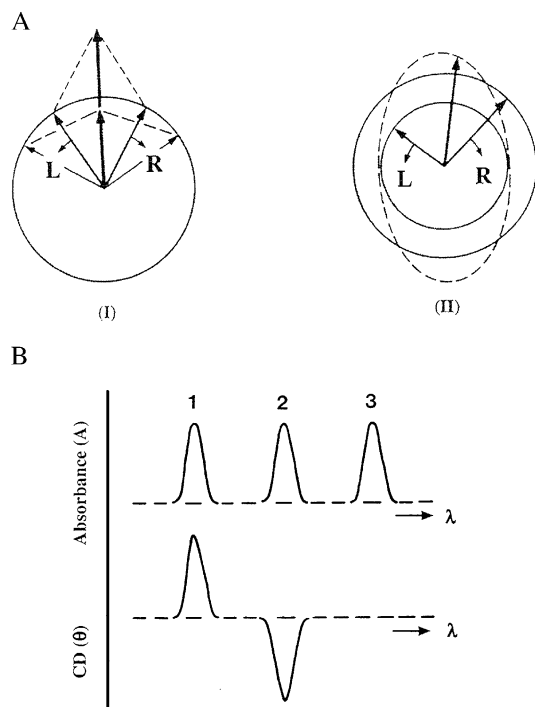


Fig. 1. Origin of the CD effect. (A) The left (L) and right (R) circularly polarised components of plane polarised radiation: (I) the two components have the same amplitude and when combined generate plane polarised radiation; (II) the components are of different magnitude and the resultant (dashed line) is elliptically polarised. (B) The relationship between absorption and CD spectra. Band 1 has a positive CD spectrum with L absorbed more than R; band 2 has a negative CD spectrum with R absorbed more than L; band 3 is due to an achiral chromophore.

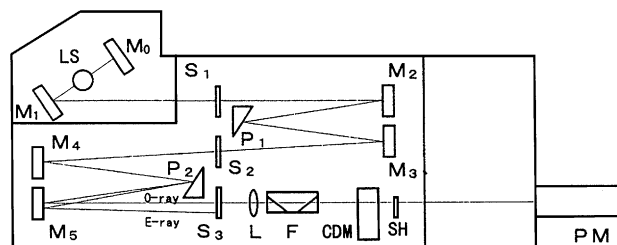


Fig. 2. Block diagram of a spectropolarimeter (Jasco J-810). Plane polarised radiation is produced by passage of light from the source (LS) through 2 prisms (P_1 and P_2) and a series of mirrors (M_0 to M_5) and slits (S_1 to S_3). The ordinary ray (O) is focussed by a lens (L), and passed through a filter (F) to the modulator (CDM). The circularly polarised components are then passed through the shutter (SH) to the sample compartment, before detection by the photomultiplier (PM). (E represents the extra-ordinary ray).

report this also in terms of the ellipticity (θ) in degrees. It should be noted that $\theta = \tan^{-1}(b/a)$ where b and a are the minor and major axes of the resulting ellipse (Fig. 1). There is a simple numerical relationship between ΔA and ellipticity (in degrees), namely $\theta = 32.98 \Delta A$. The CD spectrum is obtained when the dichroism is measured as a function of wavelength.

There are various methods by which the CD effect can be measured in a spectropolarimeter: (a) modulation, in which the incident radiation is continuously switched between the L and R components, (b) direct subtraction, in which the absorbances of the 2 components are measured separately and subtracted from each other, and (c) ellipsometric, in which the ellipticity of the transmitted radiation is measured [9]. Although it is possible that methods (b) and (c) have some potential advantages in terms of measuring time-resolved CD [9], the modulation method is by far the most commonly used. In such a CD instrument (Fig. 2), plane polarised light is split into the L and R components by passage through a modulator subjected to an alternating electric field (50 kHz is the frequency most commonly employed). The modulator normally used consists of a piezoelectric quartz crystal and a thin plate of isotropic material (e.g., fused silica) tightly coupled to the crystal. The alternating electric field induces structural changes in the quartz crystal; these make the plate transmit circularly polarised light at the extremes of the field. As the transmitted radiation is switched between L and R components, these are detected in turn by the photomultiplier.

It should always be remembered that in most, if not all, biological studies the observed CD signals are very small, i.e., ellipticities are typically in the range 10 mdeg, corresponding to a difference in absorbance ($\Delta A (=A_L - A_R)$) of the order of 3×10^{-4} . It is therefore especially important in CD work to pay attention to the experimental conditions in order to ensure that meaningful data are obtained. The purpose of this article is to explain how such conditions can be chosen with particular reference to the study of proteins.

3. Information available from CD studies of proteins

CD signals only arise where absorption of radiation occurs, and thus spectral bands are easily assigned to distinct structural features of a molecule. An advantage of the CD technique in studies of proteins is that complementary structural information can be obtained from a number of spectral regions.

In proteins, the chromophores of interest include the peptide bond (absorption below 240 nm), aromatic amino acid side chains (absorption in the range 260 to 320 nm) and disulphide bonds (weak broad absorption bands centred around 260 nm). In addition, non-protein cofactors can absorb over a wide spectral range [6], including pyridoxal-5'-phosphate around 330 nm, flavins in the range 300 nm to 500 nm (depending on oxidation state), haem groups strongly around 410 nm with other bands in the range from 350 nm to 650 nm (depending on spin state and coordination of the central Fe ion), and chlorophyll moieties in the visible and near IR regions. If a number of chromophores of the same type are in close proximity, they can behave as a single absorbing unit (exciton) which will give rise to characteristic spectral features. Finally, induced CD signals can arise from ligands which have no intrinsic chirality but acquire chirality when bound in an asymmetric environment such as provided by a protein.

The types of information which can be obtained from CD studies of proteins include:

3.1. Secondary structure composition (% helix, sheet, turns, etc.) from the peptide bond region

Absorption in this region (240 nm and below) is due principally to the peptide bond; there is a weak but broad $n \rightarrow \pi^*$ transition centred around 220 nm and a more intense $\pi \rightarrow \pi^*$ transition around 190 nm. (As noted in Section 3.2, in certain cases aromatic amino acid side chains can also contribute significantly in this spectral range). The different types of regular secondary structure found in proteins give rise to characteristic CD spectra in the far UV (Fig. 3).

A number of algorithms exist which use the data from far UV CD spectra to provide an estimation of the secondary structure composition of proteins. Most procedures employ basis datasets comprising the CD spectra of proteins of various fold types whose structures have been solved by X-ray crystallography. Detailed descriptions of the algorithms and datasets have been given in recent review articles [10,11]. Widely used algorithms include SELCON (self-consistent) [12], VARSLC (variable selection) [13], CDSSTR [14], K2d [15], and CONTIN [16]. An online server DICHROWEB [17,18] has been developed, hosted at Birkbeck College, University of London, U.K. which allows data to be entered in a number of formats including those from the major CD instrument manufacturers, and to be analysed by the various algorithms with a choice of databases. One point to be noted

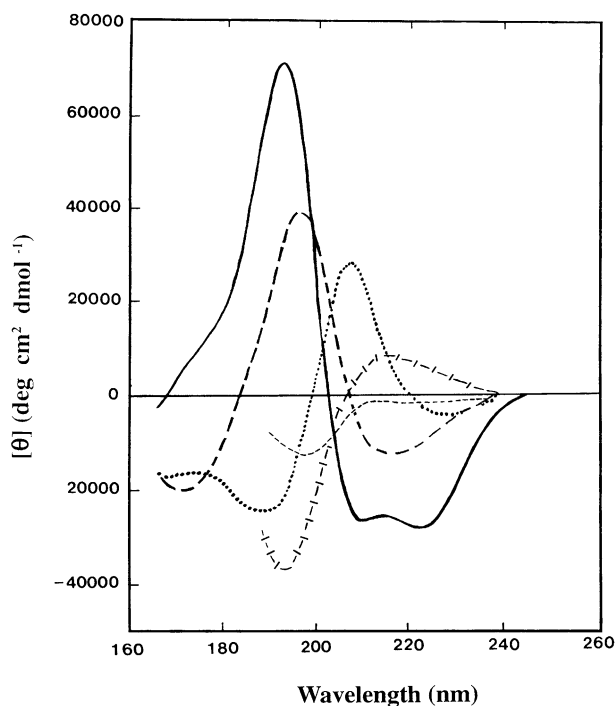


Fig. 3. Far UV CD spectra associated with various types of secondary structure. Solid line, α -helix; long dashed line, anti-parallel β -sheet; dotted line, type I β -turn; cross dashed line, extended 3_1 -helix or poly (Pro) II helix; short dashed line, irregular structure.

is that the databases do not include structures of oligopeptides and thus CD spectra of these compounds except under conditions where particular secondary structures are predominant (Section 4.2) cannot be analysed reliably at present.

In Section 10, we will describe how an assessment of the reliability of the various methods of structural analysis of proteins can be made.

For conventional CD instruments with a Xe arc light source, it is difficult to make measurements much below 180 nm, partly because the intensity of the radiation falls off in this region, but also because both the N_2 used for purging the sample compartment and optics and the H_2O solvent absorb significantly. However, estimates of secondary structure content are significantly more reliable if CD data further down into the far UV region (170 nm and below) are included; these can be obtained using synchrotron radiation CD (SRCD) [19–21]. However, in order to take full advantage of these developments, it will be necessary to build up larger datasets containing spectral data on proteins of different fold types to 170 nm and below.

It is possible to obtain estimates of the α -helical content of peptides and proteins from more limited data by using the values of the CD signals at 208 nm and 222 nm [22]. However, it must be emphasised that these estimates are to be treated with caution and, if at all possible, CD data should be gathered over a more extended range of wavelengths in the far UV.

It must be emphasised that for reliable analysis of secondary structure, it is necessary to ensure that the concentration of the protein solution is accurately known (see Section 6) and that the CD instrument is properly calibrated and operated (see Sections 7 and 8). Sophisticated analysis will not transform bad data into good data!

3.2. Tertiary structure fingerprint

The spectra in the region 260–320 nm arise from the aromatic amino acids. Each of the amino acids tends to have a characteristic wavelength profile. Trp shows a peak close to 290 nm with fine structure between 290 and 305 nm; Tyr a peak between 275 and 282 nm, with a shoulder at longer wavelengths often obscured by bands due to Trp; Phe shows weaker but sharper bands with 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 (Fig. 4). It should be noted that in certain cases, aromatic side chains can contribute significantly to the CD signals in the far UV region (below 250 nm) [23].

The actual shape and magnitude of the near UV CD spectrum of a protein will depend on the number of each type of aromatic amino acid present, their mobility, the nature of their environment (H-bonding, polar groups and polarisability) and their spatial disposition in the protein; near neighbours (generally less than 1 nm apart) may be able to couple as excitons, although the signals are generally too weak for this to be significant. Although the theoretical treatment of near UV CD spectra is not sufficiently advanced to yield significant structural insights, important progress has been made in assigning features of the spectrum to particular residues by sequential removal of aromatic side chains by site-directed mutagenesis, as in the case of bovine ribonuclease [24], human carbonic anhydrase II [25] and the molybdate-binding transcription factor ModE

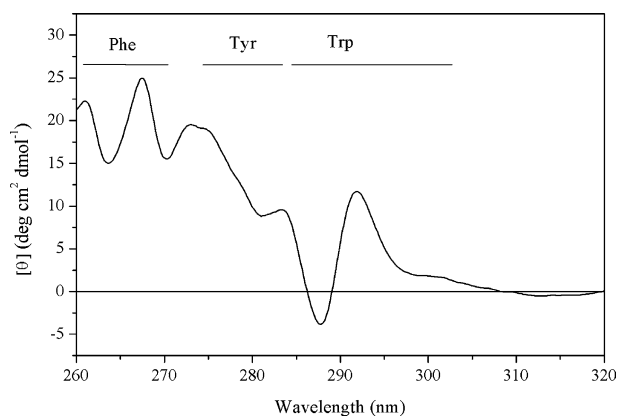


Fig. 4. The near UV CD spectrum for type II dehydroquinase from *Streptomyces coelicolor*. The wavelength ranges corresponding to signals from Phe, Tyr and Trp side chains are indicated, but it should be emphasised that there can be considerable overlap between the Tyr and Trp signals. The data are adapted from Krell et al. [84].

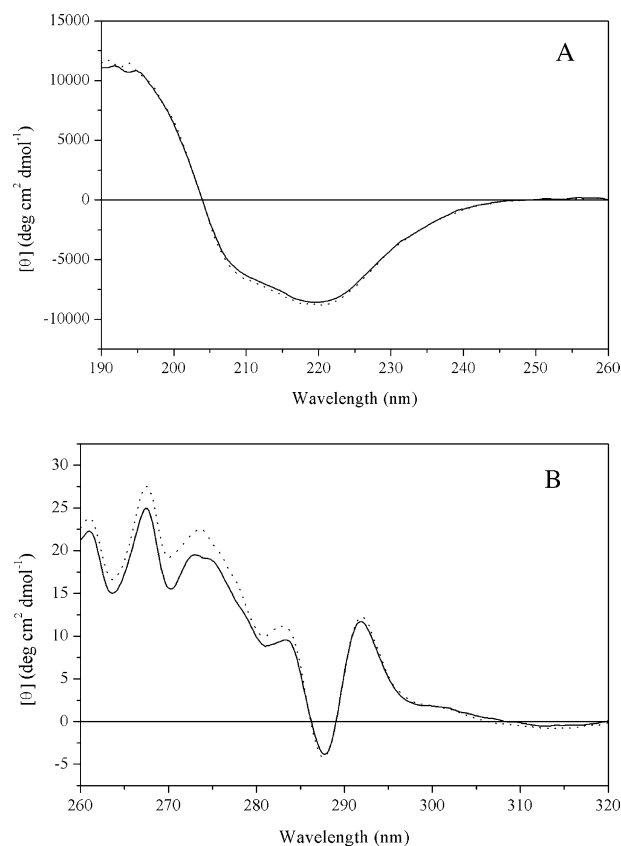


Fig. 5. CD spectra of wild type and mutant (R23Q) type II dehydroquinase from *Streptomyces coelicolor*. The far UV spectrum (panel A) and near UV spectrum (panel B) show that the wild-type (solid line) and mutant (dotted line) enzymes have very similar secondary and tertiary structures. The loss of catalytic activity in the mutant cannot therefore be due to an inability to acquire the correct folded structure. The data are adapted from Krell et al. [84].

from *Escherichia coli* [26]. Nevertheless, the near UV CD spectrum of a protein provides a valuable fingerprint of the tertiary structure of that protein, which can be used to compare, for example, wild type and mutant forms of proteins (Fig. 5). It can also provide important evidence for the existence of “molten globule” states in proteins, which are characterised *inter alia* by very weak near UV CD signals, reflecting the high mobility of aromatic side chains [27,28] (Section 3.6). It is clear that flexible or disordered regions are commonly found in proteins and may play a number of important functional roles [29].

3.3. Integrity of cofactor binding sites

Generally organic cofactors such as pyridoxal-5'-phosphate, haem, flavin, etc. show little if any CD signals when in free solution, but only when bound to their protein partner in sites which confer chirality. The CD signals in the appropriate spectral region are thus excellent indicators of the integrity of the cofactor-binding site. In some cases, the CD spectrum has been used to give detailed structural information about the cofactor sites; examples include the

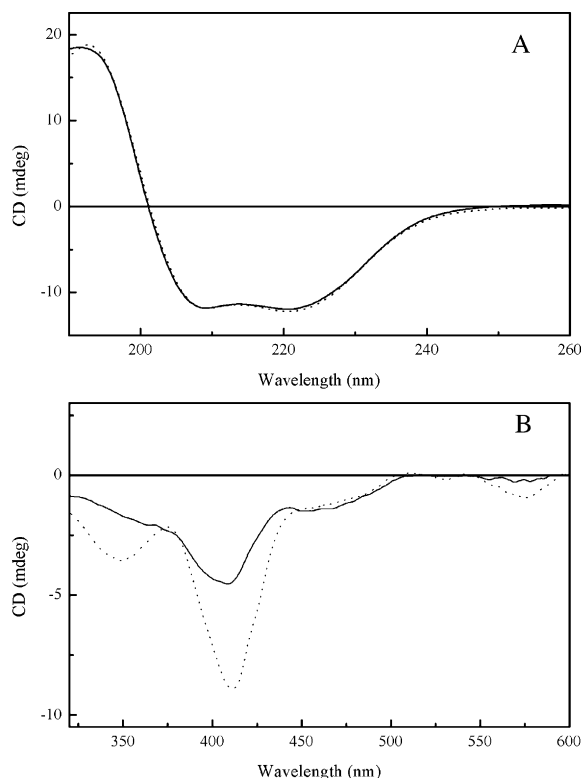


Fig. 6. CD spectra of intact cytochrome P450 BM3 from *Bacillus megaterium* and a mixture of its constituent domains expressed as recombinant proteins in *E. coli*. Panels A and B show the far UV and near UV/visible CD spectra respectively for the intact BM3 enzyme (solid line) and an equimolar mixture of the purified P450 and reductase domains (dotted line). The protein solutions are all present at 9.45 μ M and the cell pathlengths are 0.02 cm and 0.5 cm for panels A and B respectively. Note that although panel A shows that the secondary structures of the individual domains in the intact enzyme are retained when they are separately expressed, it is clear from panel B that the environment of the haem group (signal at 410 nm) is modified in the intact BM3 enzyme. The data are adapted from Munro et al. [33].

haem site in cytochrome P450s [30] and the bacteriochlorophyll sites in the LH2 light harvesting complex from *Rhodospseudomonas acidophila* [31].

3.4. Conclusion about the overall structure features of proteins

CD measurements in the far UV can give quantitative estimates of secondary structure which can be compared with those from X-ray crystallography or NMR. CD spectra in the different spectral regions are invaluable for assessing the structural relationships between native and recombinant protein, and between wild-type and mutant proteins (Fig. 5). CD data can be used to confirm the integrity of expressed domains of a multi-domain protein (Fig. 6) [32,33], an essential prerequisite before detailed structural studies, e.g., by X-ray crystallography are undertaken. In addition the loss of CD signals either on addition of denaturing agents (such as urea or guanidinium chloride) or by an increase in temperature can be used to

provide quantitative estimates of the stability of the folded state of the native protein [34–36].

3.5. Conformational changes in proteins

Structural changes in proteins caused by the binding of ligands are an essential part of the mechanism of action and regulation of biological activity. CD provides an experimentally very convenient means of detecting such changes which can be examined in different spectral regions (Fig. 7) [26]. In addition CD can be used to assess the range of ligand concentrations over which structural changes take place, the extent of the changes in the protein of interest and (using time-resolved CD studies) the speed at which such changes occur. CD is especially invaluable in the study of peptides where X-ray crystallography is not generally practicable. Examples of this are provided by the switch between α -helical and β -sheet structures in prion peptides [37] and designed peptide sequences which can switch conformation from a helical hairpin to a coiled-coil [38].

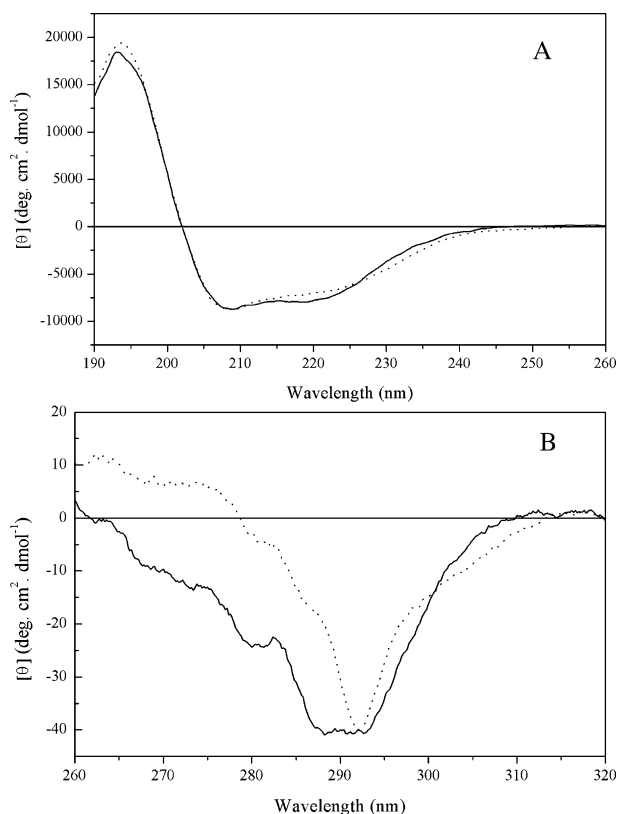


Fig. 7. CD spectra of the molybdate-sensing protein ModE from *E. coli* in the absence (solid line) and presence (dotted line) of 1 mM molybdate. Panels A and B show the far UV and near UV CD spectra respectively. The far UV CD spectra show that there is relatively little change in secondary structure on binding ligand. The marked changes in the near UV CD spectrum (signal at 292 nm) on addition of molybdate have been shown to be largely due to changes in the environment of one of the 3 Trp residues (Trp 186) in ModE. The data are adapted from Anderson et al. [85] and Boxer et al. [26].

3.6. Protein folding

A special class of a conformational change occurs during the acquisition of the native structure of a protein during biosynthesis. The experimental approach to this problem usually involves studies of the refolding of denatured proteins. CD has been used to measure the rate of acquisition of secondary and tertiary structure. Continuous- or stopped-flow CD methods can be used to detect events happening on the ms or sub-ms time scale; such information has been used to explore the mechanism of protein folding (Fig. 8). A number of excellent accounts of this topic are available [39–41]. As a result of these studies, it appears that small proteins (100 amino acids or less) can fold rapidly in a two-state mechanism with no kinetically detectable intermediates. However, the folding of larger proteins often involves a multi-stage pathway (or “folding funnel”), at the early stages of which native-like secondary structure has largely been acquired prior to the formation of correct tertiary structure contacts. These early intermediates are thought to be of the “molten globule” type (Section 3.2).

3.7. CD as a structural technique compared with X-ray and NMR approaches

By contrast with X-ray crystallography or NMR which are both capable of giving structural information on proteins at atomic level of resolution, CD is a low resolution structural technique in which overall structural features are described. It is, however, a much less demanding technique both in terms of sample and time requirements. Thus, good quality CD spectra can be obtained on ≤ 0.1 mg (far UV) or 1 mg (near UV and

visible) protein in 30 min or less and the technique is non-destructive. By contrast, X-ray crystallography requires that suitably diffracting crystals of protein are available; NMR requires high concentrations of the protein (typically 0.5 mM) and is limited to relatively small proteins or fragments of proteins (≤ 40 kDa or so). In addition, CD is versatile because it can explore protein structure under a very wide variety of experimental conditions, and can be used to measure the rates at which structural changes occur. It is also worth noting that CD could be used to assess the structure and stability of a protein under conditions used for NMR (generally high concentration of protein, slightly acid pH, elevated temperatures, long acquisition times). In addition CD could be used to explore the structure of a protein under conditions close to those used for crystallisation and compared with that under conditions which might be regarded as more relevant physiologically.

4. Preparation of protein samples for CD studies

4.1. Protein preparation, characterisation and storage

4.1.1. Preparation

Currently, the majority of protein samples are produced by over-expression of the gene encoding the protein in a suitable host system [42,43]. Such systems are usually bacteria such as *Escherichia coli*, lower eukaryotes such as yeast, or insect cells such as the fall army-worm *Spodoptera frugiperda*. The choice of host system is dictated by a number of factors including the size of the polypeptide chain (large chains often form insoluble inclusion bodies in bacteria) and the need to ensure that any post-translational modifications are correctly carried out; the methylotrophic yeast *Pichia pastoris* is often very suitable in this respect. In order to facilitate purification, it is common practice to add a small tag to the gene being expressed, such as a hexa-His tag which can be added to either the N- or C-terminal end of the protein, or to express the protein of interest as a fusion protein with glutathione-S-transferase (GST) or maltose-binding protein (MBP). The additional component can be used to provide a recognition site for purification by affinity chromatography. Generally the tag should be removed by a suitable protease unless it can be demonstrated that its presence does not affect the folding or stability of the protein. Detailed studies of α -lactalbumin, for example, have shown that a single additional N-terminal Met residue in the recombinant protein can have a marked effect on protein stability, destabilising the native form by some 15 kJ/mol [44]. In the case of GST- or MBP-fusion proteins it is clearly necessary to remove the GST or MBP portions before structural studies are carried out on the protein of interest, since these large moieties (26 kDa and 40 kDa, respectively) will contribute substantially to the spectroscopic signals from the fusion protein.

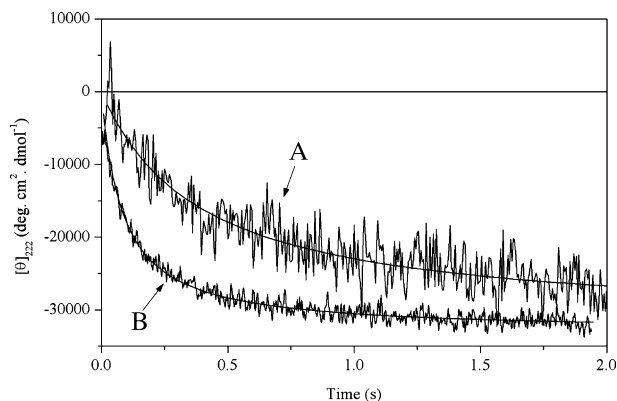


Fig. 8. The refolding of a leucine zipper peptide monitored by stopped flow CD at 222 nm. Data are shown from experiments in which the 33-residue peptide at concentrations of 6 μ M (A) and 26 μ M (B) was refolded after denaturation in GdmCl. The traces could be fitted (solid lines) to a kinetic model in which the 2 unfolded peptide chains refold to give a native dimer with a rate constant of $2.2 \pm 0.4 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$. The data are adapted from Zitzewitz et al. [76], and are reprinted in part with permission from the author and the American Chemical Society.

4.1.2. Characterisation

Proteins should be of at least 95% purity on SDS-PAGE, using Coomassie Blue and/or silver staining as appropriate. The identity of the protein and the authenticity of any post-translational modifications can be confirmed by mass spectrometry measurements on the whole protein and on peptide fragments.

The protein should also be free from nucleic acids or oligonucleotide fragments, which can often contaminate recombinant proteins. The presence of such contaminants can be most easily detected by running an absorption spectrum of the sample. For proteins the A_{280}/A_{260} ratio is typically 1.7 and for nucleic acids the A_{280}/A_{260} ratio is typically 0.6; it should be noted that the precise ratios will depend to some extent on the amino acid and base compositions, respectively. The best way of eliminating nucleic acids (or fragments) is to treat the extract obtained on cell lysis with an appropriate nuclease before protein purification is continued [43,45].

Dialysis or gel permeation should be used to remove protective agents or buffer ions which might cause problems with CD. This is especially the case with the high concentrations (typically 100–500 mM) of (a) imidazole used to elute His-tagged proteins from immobilised metal chromatography (Ni-NTA) columns, and (b) chloride ions (NaCl) used to elute proteins from ion-exchange columns, since these show high absorbance in the far UV region (Section 4.2).

The solution of the protein should be clear with no insoluble protein aggregates present, since these will cause artefacts due to (a) differential light scattering, which arises when light falls on chiral particles of dimensions comparable to or greater than its wavelength, and (b) absorption flattening which arises from the high concentrations of protein in such aggregates. These factors distort the shape and magnitude of the CD spectrum and will also decrease the signal/noise ratio [9,46,47]. Apart from visual inspection of the sample, spectrophotometry (Section 6.2.1), analytical ultracentrifugation and/or gel permeation can be used to confirm the presence of aggregates. Solutions of protein can be centrifuged (e.g., in a bench top microcentrifuge at 5000×g; 5 min) or filtered through a 0.2- μ m Millipore filter to remove aggregated material and dust particles; the concentration of the sample should be checked (Section 6) after centrifugation or filtration.

4.1.3. Storage

The integrity of the folded state of proteins in solution represents a complex balance between the protein/solvent and protein/protein forces. The majority of proteins are soluble and stable in aqueous systems over a reasonably narrow range of pH. Thus, it is usually necessary to employ suitable buffer systems [48] and it is likely that a certain minimum ionic strength is required in order to disperse the surface charges of the protein. It should also

be remembered that buffers should be used within a range of about 1 pH unit on either side of the appropriate pK_a [48] and should be at a concentration sufficient to resist changes in pH on addition of a highly charged ligand such as ATP^{4-} . Certain buffers, such as Tris, also have a high temperature coefficient [48] and the pH should be checked at or near the temperature at which they are to be used.

In addition, a number of protective agents can be added to protein solutions; the effect of these is to resemble cellular conditions in which the water is present at lower activity (concentration) than in dilute buffer solutions [49,50]. The most widely used agents include salts (e.g., $(NH_4)_2SO_4$) or osmolytes such as proline, β -alanine betaine or polyhydric alcohols such as glycerol or sucrose [51,52]. Addition of glycerol to 50% (v/v) is commonly used for storage because solutions can be kept at $-20^\circ C$ without freezing. The exact ways in which the various protective agents cause stabilisation have been extensively investigated. Polyhydric alcohols are thought to function by increasing preferential hydration of proteins. The effects of different salts on proteins are rather complex; the Hofmeister series represents an ordering of anions and cations based on their ability to precipitate or to stabilise proteins. The series has often been proposed to reflect the ability of ions to affect the H-bonding in water, although recent studies suggest that the differential effects of the ions arise from ion-specific perturbation of the macromolecular structure [53]. Other protective agents which are added to solutions of proteins during storage include inhibitors to prevent degradation by traces of proteases (these are often added as a cocktail), and dithiothreitol (a reducing agent which maintains Cys side chains in their reduced state). These protective agents are typically added at concentrations of about 1 mM in the protein solution. It should be noted that dithiothreitol absorbs quite strongly below 220 nm, so the absorbance of a buffer containing this compound should be carefully checked under the conditions used to record far UV CD spectra (Section 4.2). EDTA is often added as a stabilising agent since it will chelate not only the essential metal ions required for the action of some proteases, but also those heavy metal ions which could cause damage to cysteine side chains. Although the 4 carboxylate groups of EDTA absorb significantly below 200 nm (Section 4.2 and Table 1), the inclusion of 1 mM EDTA should not cause significant problems in a short pathlength cell (e.g., 0.02 cm).

If proteins are stored in solution, they will usually be dialysed or gel filtered to replace the protective agents by the buffer system in which the CD studies will be performed (Section 4.2). If the protein is stored or supplied as a lyophilised powder it should be carefully redissolved in buffer, with only gentle agitation to avoid denaturation. The resulting solution may require centrifugation or filtration to remove insoluble aggregates, and possibly

Table 1
Absorption properties of selected buffer components in the far UV

Component	Absorbance (50 mM solution in 0.02 cm pathlength cell)			
	180 nm	190 nm	200 nm	210 nm
NaCl	>0.5	>0.5	0.02	0
NaF	0	0	0	0
NaClO ₄	0	0	0	0
Boric acid	0	0	0	0
Na borate (pH 9.1)	0.3	0.09	0	0
Na ₂ HPO ₄	>0.5	0.3	0.05	0
NaH ₂ PO ₄	0.15	0.01	0	0
Na acetate	>0.5	>0.5	0.17	0.03
Tris/H ₂ SO ₄ (pH 8.0)	>0.5	0.24	0.13	0.02
HEPES/Na ⁺ (pH 7.5)	>0.5	>0.5	0.5	0.37
MES/Na ⁺ (pH 6.0)	>0.5	0.29	0.29	0.07

Data are adapted from Schmid [81]. Additional data can be found in Rosenheck and Doty [82] and Buck et al. [83].

dialysis or gel permeation to remove any materials added as protective agents prior to lyophilisation.

4.2. Choice of solvent/buffer system for CD studies

The solvent/buffer system chosen to perform the CD studies should be chosen with the requirements of the protein and of the experimental technique in mind. As far as the protein is concerned, some important factors concerning the preservation of conformational integrity have been discussed in Section 4.1.

As far as the requirements of CD are concerned, it should be noted that certain buffer systems may also cause problems because of their high absorbance at low wavelengths in the UV (see below and Section 8.2) or because of undesirable interactions such as chelation of essential metal ions; this can be a problem with the phosphate buffers for example. The absorption properties of a number of commonly used buffer components are given in Table 1.

If there is a conflict between the biological requirements (i.e., for protein stability), and the spectroscopic requirements (i.e., low absorbance) of a solvent, then it is generally best to give the former precedence, and to try to adapt to the latter, for example, by using salts other than chloride to maintain ionic strength (see below), or by using a cell of very short pathlength so as to minimise the absorbance due to the buffer. The latter approach is, of course, feasible only when the concentration of protein is reasonably high so that CD signals of appropriate magnitude can be obtained (Section 8.2).

Suitable buffer ions for far UV studies (in the range 260 nm to 190 or 180 nm) include phosphate, Tris and borate. Between them, these buffers are suitable for studies in the pH range from 6 to 10. In the pH range 1 to 3, phosphate is

suitable since the first ionisation of phosphoric acid has a pK_a of 2. The major problem arises in the pH range 4 to 6, since most of the buffers in this region are based on the ionisation of carboxylic acid groups (e.g., ethanoic (acetic) acid, citric acid) which absorb strongly below 200 nm (Fig. 9). In addition, chloride ions (often occurring in Tris buffers where HCl has been used to titrate to the required pH, or as a component of phosphate buffered saline (PBS) typically with a chloride concentration of about 0.14 M) absorb strongly below 200 nm (Fig. 9). If it is important that the ionic strength is to be maintained in order to preserve the structure of the protein, it is better to add anions such as sulphate or fluoride, which do not absorb significantly in this spectral range.

Buffers based on the zwitterionic sulphonic acids of heterocyclic nitrogen bases, such as HEPES, MOPS, MES, PIPES [54] generally absorb strongly at or below 200 nm and should only be used at low concentrations for detailed far UV CD studies.

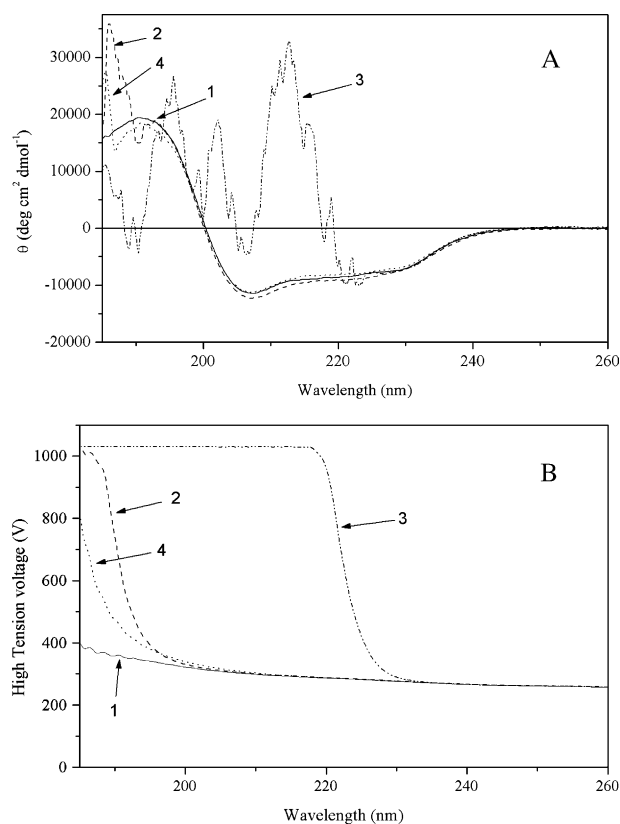


Fig. 9. The effects of buffer components on far UV CD spectra. Lysozyme (0.2 mg/ml) was dissolved in 50 mM sodium phosphate buffer, pH 7.5 (spectrum 1, solid line), or sodium phosphate buffer containing either 150 mM NaCl (spectrum 2, dashed line) or 150 mM imidazole (spectrum 3, dash-dot-dot line), or in 50 mM Tris/acetate, pH 7.5 (spectrum 4, dotted line). Spectra were recorded in a 0.02-cm pathlength cell on a Jasco J-810 spectropolarimeter using a scan speed of 50 nm/min, a time constant of 0.5 s and a bandwidth of 1 nm. 8 scans were accumulated. The upper panel shows the CD spectra and the lower panel the corresponding High Tension voltage traces.

It should be noted that imidazole which is used at concentrations of 100 mM or above to elute His-tagged proteins from immobilised metal chromatography (Ni-NTA) columns absorbs strongly in the far UV and should be removed by extensive dialysis or gel permeation prior to recording CD spectra in this range (Fig. 9). The extinction coefficient for imidazole at its absorption maximum (207.5 nm) is $3500 \text{ M}^{-1} \text{ cm}^{-1}$ [55]. Thus, a solution containing 100 mM imidazole for example in a cell of pathlength 0.02 cm has an absorbance at 207.5 nm of 7, which is far too large for meaningful CD studies (Section 8.2). Ideally the concentration of imidazole should be reduced to ≤ 1 mM.

As mentioned in Section 3.4, agents such as urea (8 M) or GdmCl (6 M) are routinely used to denature proteins. At such concentrations, these agents absorb too strongly to allow reliable CD data to be collected much below 210 nm even using cells of short pathlength (0.02 cm). This, of course, is not a problem if changes in the CD signals at 222 or 225 nm are used to assess the unfolding of a protein. It has been suggested that LiClO_4 is a useful denaturing agent for proteins, since it absorbs much less in the far UV (Table 1). However, it should be noted that the CD signals obtained in the presence of LiClO_4 may not always be readily interpreted in terms of the secondary structure of the protein [56,57].

Detailed structural studies of membrane proteins can only be carried out once they are extracted from their membrane environments by detergents [46]. The choice of detergent must be such as to maintain the structural integrity of the protein, but should not lead to excessive absorbance in the far UV. CD is an ideal technique for assessing the structural integrity of extracted membrane proteins and their stability under conditions used for crystallisation. It should be noted that the study of membrane proteins by CD can be complicated by artefacts such as light scattering and absorption flattening (Section 4.1.2) especially when the proteins are present in large structures such as sonicated membrane fragments [47]. However, even in small vesicles where the above artefacts are much less significant, the structural analysis of membrane proteins can be complicated by shifts of a few nm in the wavelengths of absorption bands due to changes in solvent polarity. It has been suggested that these spectral shifts can make the use of reference spectra based solely on soluble proteins in the analysis of membrane proteins less reliable [58]. However, other studies [11] have indicated that the use of soluble protein reference spectra together with careful attention to the method of data analysis can yield reasonably reliable results especially when membrane proteins are included in the dataset. It is clear that further progress in this area will be greatly assisted by the expansion of reference data sets for membrane proteins.

Detergent systems based on alkyl glycosides (e.g., lauryl maltoside and octyl glucoside) are very suitable for far UV studies; they are, however, expensive and it may be that other detergents are used in the early stages of preparations

with a detergent exchange step incorporated at a later stage in the process. However, it should be noted that some commonly used detergents such as Triton X-100 can be difficult to remove from proteins and lead to high absorbance around 280 nm. Although not generally useful for protein studies, SDS is a valuable reagent for assessing the structure-forming potential of peptides and protein fragments [59,60] since it does not absorb in this spectral range.

Non-aqueous solvents are rarely used except in studies of certain very non-polar integral membrane proteins. However, such solvents are commonly used in studies of small peptides and organic molecules such as natural products and drug molecules. Solvents containing carbon-chlorine bonds tend to absorb strongly below 230 nm; dimethylsulphoxide and dimethylformamide absorb strongly below 240 nm and 250 nm respectively. Organic solvents which can be used down to 190 nm include acetonitrile, ethanol and methanol (these should be of high purity, e.g., HPLC grade). 2,2,2-trifluoroethanol (TFE) is a useful solvent for the study of peptides and protein fragments, since it is usually found to promote helix formation and can thus be used to assess structure-forming potential [61]. Although the origin of the effect of TFE has been a matter of considerable debate, recent studies on synthetic peptides suggest that it enhances the dehydration of specific residues within a helix, promoting helix stability [62].

The suitability of any buffer or solvent system for CD should be checked by running appropriate blank spectra. These will ensure that the absorbance is not too high (as indicated by the High Tension voltage, Section 8.2) and will indicate whether background signals are arising from any chiral components present.

Finally, it is important to ensure that the protein being studied is stable over the period of the experiment under the conditions employed. This could involve checking that the protein is not damaged by exposure to the light source (typically a Xe arc in conventional CD instruments or a synchrotron source in SRCD). Stability would be assessed by looking for evidence of loss of biological activity or of a time-dependent loss of signal.

5. Presentation of CD data

5.1. Units of CD data

CD data are presented in terms of either ellipticity [θ] (degrees) or differential absorbance (ΔA). The data are normalised by scaling to molar concentrations of either the whole molecule or the repeating unit of a polymer.

For far UV CD of proteins, the repeating unit is the peptide bond. The Mean Residue Weight (MRW) for the peptide bond is calculated from $\text{MRW} = M/(N - 1)$, where M is the molecular mass of the polypeptide chain (in Da), and N is the number of amino acids in the chain; the number of

peptide bonds is $N-1$. For most proteins the MRW is 110 ± 5 Da.

The mean residue ellipticity at wavelength λ ($[\theta]_{\text{mrw},\lambda}$) is given by:

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

where θ_{λ} is the observed ellipticity (degrees) at wavelength λ , d is the pathlength (cm), and c is the concentration (g/ml).

If we know the molar concentration (m) of a solute, the molar ellipticity at wavelength λ ($[\theta]_{\text{molar},\lambda}$) is given by:

$$[\theta]_{\text{molar},\lambda} = 100 \times \theta_{\lambda} / m \times d$$

where θ_{λ} and d have the same meaning as above.

The units of mean residue ellipticity and molar ellipticity are $\text{deg cm}^2 \text{ dmol}^{-1}$.

For data in absorption units, it is usual to calculate the molar differential extinction coefficient, $\Delta\epsilon$. If the observed difference in absorbance at a certain wavelength of a solution of concentration m in a cell of pathlength d (cm) is ΔA , then $\Delta\epsilon$ is given by:

$$\Delta\epsilon = \Delta A / m \times d$$

The units of molar differential extinction coefficient are $\text{M}^{-1} \text{ cm}^{-1}$.

There is a simple numerical relationship between $[\theta]_{\text{mrw}}$ and $\Delta\epsilon$ namely $[\theta]_{\text{mrw}} = 3298 \times \Delta\epsilon$.

The size of signals in the far UV is a good check on the various experimental parameters and calibration of the instrument. Thus, a hypothetical protein which is 100% α -helix has a mean residue ellipticity at 222 nm of about $-30000 \text{ deg cm}^2 \text{ dmol}^{-1}$ ($\Delta\epsilon = -9 \text{ M}^{-1} \text{ cm}^{-1}$) (Fig. 3). Clearly the observed value for a real protein cannot be greater than this. For near UV CD, the mean residue ellipticities for the aromatic amino acid side chains of proteins should be generally less than $200 \text{ deg cm}^2 \text{ dmol}^{-1}$ ($\Delta\epsilon$ less than $0.06 \text{ M}^{-1} \text{ cm}^{-1}$).

There is some debate about the correct choice of molar units for near UV/visible CD; should it refer to the MRW (repeating unit) or to the whole protein? In reporting any work, it is essential to specify the actual molar unit used.

Reference to the equations above shows that in order to calculate the molar or mean residue ellipticity or molar differential extinction coefficient, we need to know the quantities listed below.

5.1.1. The concentration and mean residue weight of the protein under study

The determination of protein concentration is discussed in Section 6. The mean residue weight (MRW) can be precisely calculated if we know the sequence of the protein, since we will know the exact molecular mass and the number of amino acids. If the sequence it is not known, it may be possible to estimate the MRW from the measured molecular mass, e.g., by analytical ultracentrifugation or gel

permeation, coupled with the number of amino acids estimated by amino acid analysis. In any case the value of MRW for most proteins is likely to be close to 110, and this value can be used for preliminary purposes.

5.1.2. The measured ellipticity or differential absorbance at a selected wavelength

The calibration of the CD instrument in terms of wavelength and spectral magnitude are discussed in Sections 7.1 and 7.2, respectively.

5.1.3. The pathlength of the cell

The determination of cell pathlength is discussed in Section 7.3

In addition, it is important that CD data are acquired using appropriate experimental conditions; these are discussed in detail in Section 8.

6. Determination of protein concentration

As mentioned in Section 3.1, accurate determination of protein concentration is particularly important for reliable determination of secondary structure by far UV CD.

6.1. Methods for determination of protein concentration

A number of methods are routinely used for determination of protein as indicated in Table 2. Further details of, and references to, these methods are given in Price [63].

The biuret method depends on the formation of a purple complex between Cu(II) ions and adjacent peptide bonds in a protein under alkaline conditions. The method gives a fairly uniform response for different proteins but requires large amounts of sample (0.5 to 5 mg).

The Lowry method is based on the reduction of the phosphomolybdic tungstic mixed acid chromogen (in the Folin–Ciocalteu reagent) by a protein to give a blue product. Although the Lowry method is sensitive (5 to 100 μg sample), different proteins give different responses and a large number of substances can interfere.

The bicinchoninic acid (BCA) method is based on the ability of BCA to combine with Cu(I) ions (produced by reduction of Cu(II) by a protein under alkaline conditions) to give a yellow-green product. The method is convenient and sensitive (1 to 100 μg sample), but gives different responses for different proteins.

The Coomassie Blue binding method (for example the Bradford method) relies on the change in the absorption spectrum of the dye when it binds to proteins under acid conditions (from orange-red to blue). The method is convenient and sensitive (1 to 100 μg sample), but gives different responses for different proteins.

For a purified protein, measurement of A_{280} is a very commonly used method and is discussed in detail below (Section 6.2).

Table 2
Methods for determination of protein concentration

Method	Amount of protein required (μg)	Complexity of method ^a	Response of identical masses of different proteins	Reference protein used	Major sources of interference
Biuret	500–5000	2	Very similar	Yes	Tris, NH_4^+ , glycerol
UV (280 nm)	100–1000	1	Variable	No	Nucleic acids and other chromophores
UV (205 nm)	5–50	1	Similar	No	Many buffer components and other solutes
Lowry	5–100	3	Variable	Yes	Amino acids, NH_4^+ , thiol compounds, certain buffers and detergents
Bicinchoninic acid (BCA)	5–100 (1–10 in microprotocol)	3	Variable	Yes	Glucose, NH_4^+ , EDTA
Coomassie blue binding	5–50 (1–10 in microprotocol)	2	Variable	Yes	Triton, SDS
Amino acid analysis	10–200	4	Variable	No	Other contaminating proteins

Data adapted from Price [63], which should be consulted for further details and references for each method.

^a Graded on a 1–4 scale; 1 involves pipetting of the sample only; 2 involves mixing the sample with one reagent solution; 3 involves mixing the sample with >1 reagent solution; 4 involves lengthy manipulation of the sample and specialised equipment.

A second less commonly used absorbance method is based on measurements of the absorbance of a protein in the far UV (205 nm). At this wavelength, absorption is primarily due to peptide bonds which occur at a very similar frequency in virtually all proteins. There is a small contribution to the A_{205} from other chromophores, principally the side chains of Tyr, Phe, Trp, and His, and these are taken into account in an empirical formula devised by Scopes [64] which requires measurements of A_{205} and A_{280} .

$$A_{205}(1 \text{ mg/ml}) = 27 + 120(A_{280}/A_{205})$$

The far UV absorbance method is sensitive (5 to 50 μg), and gives reliable results for different proteins. However, there can be problems due to the fact that (a) the deuterium arc sources in most spectrophotometers have a relatively poor output at wavelengths as low as 205 nm, and (b) many buffer components can absorb strongly at 205 nm, thereby compromising the validity of the measurements of absorbance of the protein solution. It should be pointed out that the A_{280} of a protein solution is generally of the order of only 5% of the A_{205} , so it can be difficult to measure both absorbance values accurately for a solution of protein unless either cuvettes (cells) of different pathlength are employed, or alternatively the solution is accurately diluted to measure the A_{205} .

A relatively small number of proteins contain cofactors which absorb in the near UV or visible; this absorption can be used as the basis for estimating the protein concentration, e.g., cytochrome P450 is estimated on the basis of absorbance of the haem group at 450 nm in the reduced form of the enzyme complexed with CO [33]. It should be noted that this method assumes that there is stoichiometric occupation of the cofactor binding sites in the protein, which may not always be the case with proteins over-expressed in heterologous systems.

6.2. Protein concentration based on A_{280}

This convenient, non-destructive, method depends on the absorption of a protein at 280 nm (assuming that the protein contains at least some Tyr and Trp). The expected A_{280}

value for a 1 mg/ml solution of protein in 6 M GdmCl (i.e., under denaturing conditions) can be calculated from the following formula, which is based on the absorbance properties of these amino acids as well as a small contribution from any disulphide bonds [65]. It should be noted that although these bonds are rarely found in intracellular proteins, this term makes only a very small contribution to the calculated absorbance.

$$A_{280}(1 \text{ mg/ml}; 1 \text{ cm}) = (5690n_W + 1280n_Y + 60n_C)/M$$

where n_W , n_Y , and n_C are the number of Trp, Tyr, and Cys per polypeptide chain and M is the molecular mass (in Da). In the above equation, the coefficient of n_C in the equation given by Gill and von Hippel [65] has been halved since their value (120) refers to the disulphide-bonded unit, i.e., cystine. If there are no disulphide bonds in the protein the term in n_C should be neglected.

For example the formula can be applied to the proteins hen egg white lysozyme, bovine chymotrypsinogen and bovine insulin as shown in Table 3.

The calculated value of A_{280} (1 mg/ml; 1 cm) is available from analytical tools such as ProtParam within the ExPasy system (<http://www.us.expasy.org/tools/protparam.html>) for any given protein stored in Swiss-Prot or TrEMBL databases or for a user-entered sequence. Two sets of calculated A_{280} values are given, one assuming that all the cysteine side chains form disulphide bonds, the other assuming that none do.

The calculated value of A_{280} is valid for a homogeneous sample of protein, provided that ALL the following conditions are met:

- there is no contribution from light scattering (Section 6.2.1);
- there is no other chromophore (e.g., cofactor) in the protein (Section 6.2.2);
- there is no other absorbing contaminant, e.g., nucleic acids (Section 6.2.2);
- a correction is applied for the difference between native and denatured protein (Section 6.2.3).

Table 3
Calculated and experimental A_{280} values for lysozyme, chymotrypsinogen and insulin

Protein	Mass (Da)	n_W	n_Y	n_C	A_{280} (1 mg/ml: 1 cm)		
					Calculated ^a	Calculated ^b	Experimental ^c
Lysozyme (Hen egg white)	14,314	6	3	8	2.69	2.65	2.65
Chymotrypsinogen (bovine)	25,666	8	4	10	2.00	1.97	1.98
Insulin (bovine)	5734	0	4	6	0.96	1.10	0.97

Note that in each protein all the cysteine side chains form disulphide bonds.

^a Calculated according to the modified equation of Gill and von Hippel [65] given in Section 6.2.

^b Calculated according to the equation of Pace et al. [66] given in Section 6.2.4.

^c The experimental values represent averages of data given in Gill and von Hippel [65] and Pace et al. [66].

6.2.1. Light scattering

It is important to check the absorption spectrum of the protein sample, ideally over the range from 400 nm to 240 nm (Fig. 10). If there is a gradually increasing baseline optical density (i.e., apparent absorbance) as the wavelength decreases from 400 nm to 310 nm, this indicates that light scattering is a problem; it is assumed that no

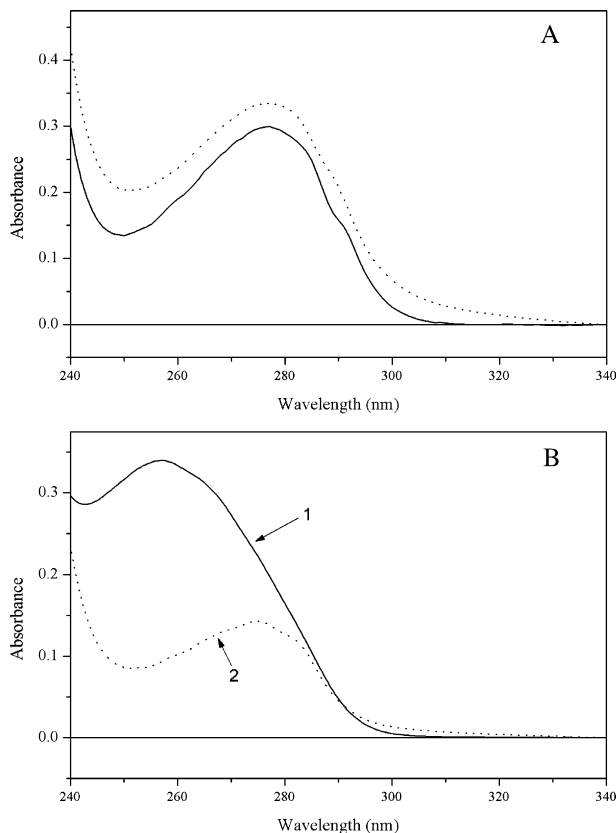


Fig. 10. Problems encountered in measuring the absorbance of protein solutions. Panel A shows the absorption spectrum of a solution of the actin-binding protein scp1p from *Saccharomyces cerevisiae* before and after centrifugation to remove aggregated protein. Panel B shows the absorption spectrum of the SNARE protein syntaxin 4 from *S. cerevisiae* expressed as a recombinant His-tagged protein and purified by immobilised metal-ion chromatography. Spectrum 1 (solid line) shows that there is considerable contamination by nucleic acids, since the peak of absorption is close to 260 nm. Spectrum 2 (dotted line) shows the absorption spectrum of syntaxin 4 where the cell extract has been incubated with DNase before chromatography.

cofactor such as flavin or pyridoxal-5'-phosphate which absorbs in this range is present. If light scattering is a problem, the solution can be centrifuged or filtered to try to eliminate or minimise the problem (see Section 4.1.2). If these approaches are not successful, it is possible to correct for the actual contribution of light scattering to the observed A_{280} by use of a log/log plot, but unless this is incorporated into the spectrophotometer software, it can be a time-consuming procedure. A plot is made of log absorbance against log wavelength, and the line obtained in the non-absorbing region, e.g., from 400 nm to 310 nm is extrapolated to lower wavelengths to determine the contribution from scattering to the measured absorbance at these wavelengths [47].

6.2.2. Interference from other chromophores

The presence of nucleic acids can be inferred from the A_{280}/A_{260} ratio of the sample (Fig. 10) or by the use of dyes which bind specifically to nucleic acids. Nucleic acid contaminants can be removed by treatment of the cell extract with the appropriate nuclease (e.g., DNase for DNA or its fragments) (Section 4.1.2).

Some non-protein chromophores such as haem groups absorb at 280 nm and their contribution has to be taken into account if A_{280} values are being used to determine the protein concentration.

6.2.3. Correction to native conditions

The calculated value of the A_{280} for a protein refers to the unfolded protein, i.e., where the chromophore side chains are fully exposed to solvent as they are in small model compounds; it should be noted that the spectral properties of the chromophores will depend to a small extent on the polarity of the environment. The correction can be made, i.e., the ratio of the A_{280} values for native and denatured protein can be determined, by performing parallel dilutions, e.g., (a) 0.25 ml of native protein in buffer is diluted into 0.75 ml buffer (native) or (b) 0.25 ml of native protein in buffer is diluted into 0.75 ml 8 M GdmCl in buffer (denatured). The experimentally determined ratio of the A_{280} values of native and denatured proteins generally lies between 0.9 and 1.1 [65] and can be used to correct the calculated A_{280} to give the appropriate value for native protein.

6.2.4. Calculation of A_{280} values of native proteins

Other equations have been developed for calculating the A_{280} values of proteins under native conditions from the amino acid composition; these use average extinction coefficients for Tyr and Trp in a number of proteins [66,67]. The equation proposed by Pace et al. [66] is:

$$A_{280}(1 \text{ mg/ml}; 1 \text{ cm}) = (5500n_W + 1490n_Y + 62.5n_C)/M$$

This equation assumes that all the cysteines form disulphide bonds; if there are no disulphide bonds the term in n_C is neglected. It is reliable for those proteins which contain Trp, but less so for those which do not. Although the equation avoids the need to correct to native conditions (Section 6.2.3), it is still important to check that the measured A_{280} value is not affected by light scattering (Section 6.2.1) or interference from other chromophores (Section 6.2.2).

6.3. Recommended method for accurate determination of protein

For accurate determination of protein concentration, the method of choice is quantitative amino acid analysis [63]. Although a number of amino acids are not released in quantitative fashion by hydrolysis in 6 M HCl, for example, Thr and Ser are partially destroyed, Trp is largely destroyed and bulky amino acids such as Ile and Val may not be fully released, the method can be made reliable by measuring the amounts of a number of stable, abundant amino acids (e.g., Ala, Arg, Lys, Tyr, etc.) in a sample relative to a known amount of an added internal amino acid standard such as norleucine. The relative amounts of the stable, abundant, amino acids should correspond to those predicted from the amino acid composition of the protein. The composition can be inferred from the sequence, which will be known if an over-expressed protein is being purified, or may have been determined directly for the wild-type protein. The correct relative amounts of the amino acids provide an additional check on the authenticity of the protein sample. From the measured amounts of the abundant amino acids, the number of moles of protein can be determined and hence the mass of protein calculated. Although this method is cumbersome, it can be used to calibrate a more readily applied everyday method for the protein in question, e.g., A_{280} measurements, dye binding, BCA assay, etc.

7. Calibration of CD instrument and cells

7.1. Wavelength calibration of spectropolarimeter

A number of methods are available for calibrating the wavelength of the CD instrument, including the use of rare earth element filters (e.g., holmium oxide which has peaks at 279.4 nm, 361.0 nm and 453.7 nm), benzene vapour (241.7 nm, 253.0 nm and 260.1 nm) and neodymium glass (586.0 nm). In addition solutions used for calibrating the

magnitude of the CD signal (Section 7.2) such as CSA or pantolactone can be used to check wavelength calibration. Miles et al. [68] recommend that an instrument is calibrated at a number of different wavelengths including at least one in the spectral range of interest, e.g., the far UV below 240 nm [69].

7.2. Magnitude calibration of spectropolarimeter

A number of compounds have been recommended as standards for magnitude calibration of CD instruments [70,71]. CD studies of proteins usually involve the spectral region from 320 nm to 180 nm, and it is not surprising that CSA (1S-(+)-10-camphorsulphonic acid) has been widely used as a standard, since it has convenient CD peaks in the near and far UV (290.5 nm and 192.5 nm). The formula mass of free acid form of CSA is 232.3 and solutions for calibration are typically made up at 0.06% (w/v) (0.6 mg/ml) in distilled water. As determined by Miles et al. [72] the extinction coefficient of CSA at 285 nm is $34.6 \pm 0.2 \text{ M}^{-1} \text{ cm}^{-1}$; so the absorbance of this solution of the free acid should be 0.089₄ in a 1 cm pathlength cell. The ellipticity of this solution in a 1-cm pathlength cell should be +202 mdeg at 290.5 nm and -420 mdeg at 192.5 nm. The ratio of the absolute signals at 192.5 nm and 290.5 nm provides an important check on the performance of the instrument; this ratio should be 2.05 or greater.

Because the free acid form of CSA is hygroscopic, the ammonium salt of CSA (formula mass 249.3) is preferred by some workers. A 0.06% (w/v) solution of the ammonium salt should have an absorbance at 285 nm of 0.083₃ in a 1-cm pathlength cell and ellipticities in a 1-cm cell of +188 mdeg at 290.5 nm and -391 mdeg at 192.5 nm.

It should be noted that there is a significant (2.5%) decline in the ellipticity of CSA solutions with temperature over the range 5° to 30 °C [64, Rodger, A., unpublished work; Jess, T.J., unpublished work]. The cause of this decline is not established, but it does not represent an irreversible change in the CSA, since reproducible changes in ellipticity can be observed by cooling and re-heating the solution. CSA solutions should therefore be allowed to equilibrate at the required temperature after being stored in the fridge or cold room. It is recommended that the calibration of the CD instrument should be carried out at 20 °C.

Other substances recommended for calibrating CD instruments include pantolactone (formula mass 130.15) which has a strong band at 219 nm. For (*R*)-(-)-pantolactone $[\Theta]_{219} = -16160 \text{ deg cm}^2 \text{ dmol}^{-1}$. A 0.015% (w/v) solution will have an ellipticity of -186 mdeg at 219 nm. Calibration in the visible region of the spectrum (490 nm) can be performed using the tris(ethylenediamine)-Co complex, salts of which can be prepared and used as described [70].

Although the evidence dealing with the stability of CSA solutions in water is limited, it is best to operate on the side

of caution and make up a new standard solution at least every 2 weeks. The solution should be kept in the dark at 4 °C apart from when it is being used for calibration. It is recommended that the calibration of the CD instrument is checked weekly, or more frequently if a particularly crucial set of experiments is planned.

7.3. Calibration of CD cells

The cells or cuvettes used for CD are expensive; for example cylindrical quartz Suprasil cuvettes suitable for work down to 170 nm are typically 200–300 euros (250–400 dollars) each and are usually manufactured to order. To avoid optical artefacts, each cell should be ordered as “strain free for polarimetry”. It is important to run a blank spectrum with the buffer or solvent in the same cell as used for the protein sample of interest.

Refillable cells of pathlength 0.01 cm to 1 cm (or even longer) can be used. These cells can be of a rectangular or circular geometry, depending on the design of the sample compartment; it is important to ensure that the light beam impinges on the transparent portion of the cell. Shorter pathlength cells (e.g., 0.001 cm, 10 μm or even shorter, e.g., 4 μm) are of a demountable type; these consist of 2 quartz plates, one of which is machined out to the required depth.

Because the pathlength of the cell is a term in calculating the molar ellipticity of a sample it is essential to check that this is accurately known. A study by Miles et al. [71] showed that whereas the pathlengths of cells in the range 0.1 to 1.0 cm are generally accurately reported by the manufacturers, the pathlengths of shorter cells can differ by as much as 50% from the stated value. In addition, depending on the volume of liquid loaded in a demountable cell, and the pressure applied by the operator, the pathlength can be increased by 0.5–1.0 μm due to the film of liquid between the flanges of the plates [69,71].

Two principal methods can be used to determine the actual pathlengths of cells. The first is the interference fringe method [73], which is based on a portion of the incident beam being retarded by double reflection from the inner surfaces of the empty cell. An interference pattern is produced in the visible and UV transmission spectrum, when the retarded beam interferes with the main beam. From the distance between the peaks, it is possible to calculate the cell pathlength. The second method is based on measurements of the absorbance of a solution of potassium chromate (molecular weight 194.2), which is widely used as a primary standard for spectrophotometry; solutions are made up gravimetrically. The extinction coefficient of potassium chromate in alkaline solution (0.05 M KOH) at 372 nm is 4830 $\text{M}^{-1} \text{cm}^{-1}$ [74]. Thus, the absorbance of a 0.01-M solution (1.942 g/l) in a cell of pathlength 0.01 cm would be 0.483. For shorter cells (e.g., 10 μm), a correspondingly more concentrated solution of potassium chromate should be used.

8. Experimental conditions

8.1. Instrument parameters

Various machine settings can be adjusted to improve the results

- bandwidth;
- time constant;
- scan rate;
- number of scans.

The following considerations are important in selecting these parameters:

8.1.1. Bandwidth

The bandwidth is a measure of the precision with which a monochromator selects light of a chosen wavelength. Increasing the bandwidth will allow more light to fall on the sample and hence on the photomultiplier, but will decrease the ability to resolve spectral bands. The bandwidth should be less than or equal to 1 nm for routine CD studies, but values down to 0.1 nm are useful, particularly to resolve fine structure in the near UV spectrum of proteins [69].

8.1.2. Time constant and scan rate

The time constant is a measure of the time over which the CD data are averaged and will depend on the precise mode of operation of the instrument (response time, or dwell time have analogous meanings for different types of instruments).

A set of useful guidelines has been suggested:

$$\text{scan rate} \times \text{time constant} < \text{bandwidth} < W/10$$

where W = the width at half height of the spectral feature under investigation [64].

For example, if the bandwidth were assumed to be 0.5 nm, we could use combinations of time constant and scan rate of 2 s and 10 nm/min or 0.25 s and 100 nm/min. These values would need to be adjusted for high resolution work where a smaller bandwidth might be required. It is important that any departure from these guidelines (for example to accommodate a faster scan speed) should only be made if it has been shown that this does not distort the spectrum obtained on the particular CD instrument used (compare, for example, Panels A and B of Fig. 11).

8.1.3. Number of scans

Increasing the number of scans will improve the signal/noise (S/N) ratio; the S/N ratio is proportional to the square root of the number of scans.

In practice, the number of scans and the scan rate (and hence time constant) depend on the time available for the experiment and should also take into account any limitations imposed by (a) the stability of the sample

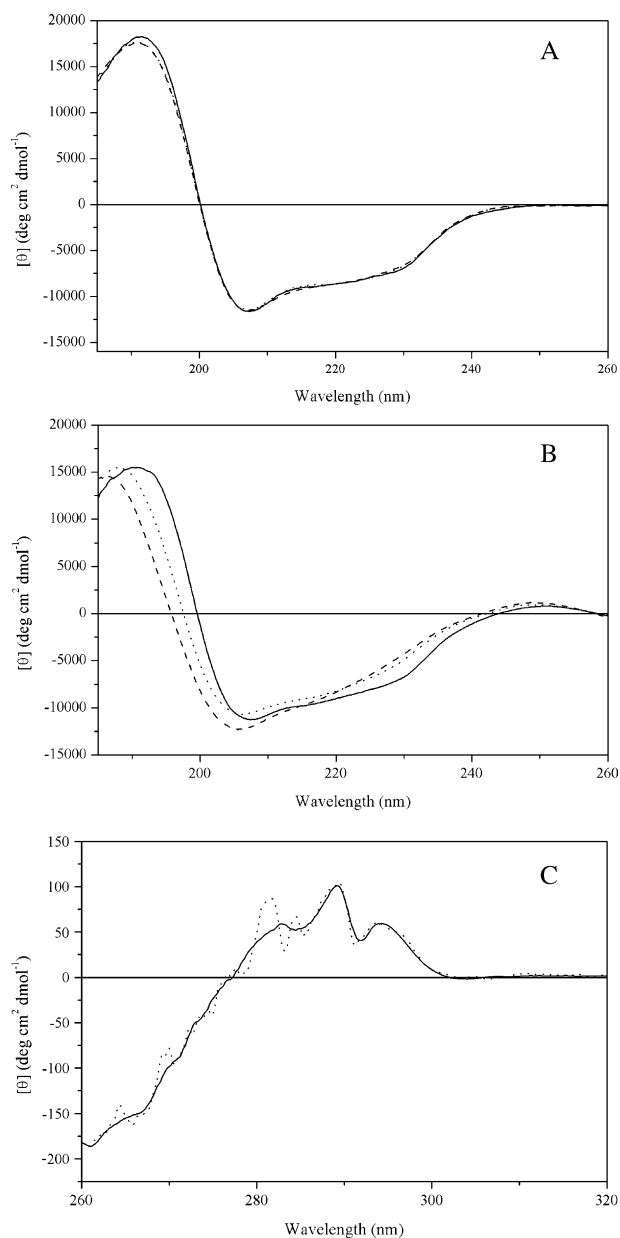


Fig. 11. The effects of instrument parameters on CD spectra. The spectra shown are those of lysozyme (0.5 mg/ml) in 50 mM sodium phosphate buffer, pH 7.5. Panels A and B show far UV CD spectra recorded in a cell of 0.02 cm pathlength on Jasco J-810 and J-600 spectropolarimeters, respectively. The solid line, dotted line and dashed line refer to time constants and scan rates of 0.5 s and 50 nm/min, 2 s and 50 nm/min, and 2 s and 100 nm/min, respectively. In each case, the bandwidth was 1 nm and 8 scans were accumulated. Panel C shows the near UV CD spectrum recorded in a cell of 0.5 cm pathlength on a Jasco J-810 spectropolarimeter. The solid and dotted lines refer to accumulations of 20 scans and 1 scan, respectively; the time constant, scan rate and bandwidth were 0.5 s, 50 nm/min, and 1 nm, respectively.

under the conditions employed, and (b) the stability of the instrument.

The effects of various machine settings on the quality of the CD spectra are illustrated in Fig. 11.

It is not possible to issue definitive recommendations about the combination of machine settings to be used

under all circumstances. The balance between scan speed, time constant and number of scans (even within the guidelines above) will also depend on factors such as the time available for the experiment and the stability of the sample under the conditions used, and must be investigated in each case.

In addition, the protein concentration and cell pathlength can be adjusted to improve the quality of the data (Section 8.2).

8.2. Choice of correct protein concentration and cell pathlength

One of the great strengths of CD is that it can be used to study proteins over a wide range of concentrations, which can give valuable insights into concentration-dependent processes such as the association between proteins or peptides containing the leucine-zipper motif to generate coiled-coil structures which can bind to DNA [75,76].

The aim in CD is to keep the total absorbance of the sample (due to the protein and to the buffer or solvent) within reasonable bounds (typically below about 1.0) in order to avoid excessive noise. In fact, the signal/noise ratio for CD is theoretically a maximum when the absorbance is 0.869 [9]. The absorbance of the sample is conveniently monitored by the trace of the High Tension voltage (the voltage applied to the photomultiplier). For reliable data, this should remain within specified bounds (generally the voltage should be less than 700 V, but this value will depend on the particular instrument being used).

The absorbance values of a 1 mg/ml solution of a typical protein in a cell of 1 cm pathlength at various wavelengths are shown in Table 4.

Table 4
Absorbance values for proteins in the UV^a

Wavelength (nm)	Absorbance (1 mg/ml solution in a 1-cm pathlength cell)
190	65
193	68
200	45
205	32
210	21
215	15
220	11
230	5
260	≈ 0.6 ^b
280	≈ 1 ^b
310	< 0.05 ^c
320	0 ^c

^a The absorbance values shown are to be regarded as typical for proteins, rather than exact values. The data are adapted from Stevens [48].

^b This value depends on the Trp and Tyr content of the protein; values for A_{280} are found to range from 0 to 4.

^c Assuming that no cofactor absorbing in this range, e.g., FAD, pyridoxal-5'-phosphate is present.

In order to meet the absorbance criterion (i.e., total absorbance of sample ≤ 1.0), we can adjust either the concentration of protein or the cell pathlength or both. For example, in order to keep the A_{190} at or below 1.0 we could use a protein concentration of 10 mg/ml and a pathlength of 0.001 cm or a concentration of 1 mg/ml and a pathlength of 0.01 cm or a concentration of 0.1 mg/ml and a pathlength of 0.1 cm. Thus, potentially a very wide range of concentrations can be studied in the far UV, although it should be noted that in very dilute solutions ($<10 \mu\text{g/ml}$) a significant proportion of the protein present may be adsorbed on the surfaces of the cell. Using refillable cells with pathlengths in the range 0.01 to 0.05 cm the volumes required are usually in the range 0.3 to 0.4 ml; most of this ($>90\%$ with care) can be recovered. With demountable cells, a drop of solution (typically 0.05 ml or less) is required; this cannot readily be recovered when the cell is taken apart after the spectral measurements have been made. The effects of varying protein concentration and cell pathlengths on the quality of far UV CD spectra are shown in Fig. 12. It is seen that high absorbances can lead to artefacts; the pattern of the High Tension voltage trace clearly indicates the range of wavelengths over which reliable data can be collected.

In the near UV and visible regions, the range of concentrations which can be studied is rather more limited. With a protein concentration of 0.5 mg/ml, the pathlength is typically 1 cm; at a concentration of 5 mg/ml, the pathlength would be 0.1 cm. Using a rectangular cell, it is possible to use 1 ml of a 0.5 mg/ml solution to obtain a spectrum; though essentially, all of this can be readily recovered. If the protein has a chromophore which absorbs in the visible/UV region (e.g., pyridoxal-5'-phosphate, flavin or haem) the concentrations and pathlengths used are in the same range as for the near UV region.

8.3. CD studies at different temperatures

As mentioned in Section 3.4, CD studies at different temperatures can be used to provide information on the stability of proteins. These types of experiments have become much easier to perform with the introduction of Peltier devices which allow the temperature inside a cell to be varied in a more systematic fashion than is possible with an external circulating bath. However, there are several points to be borne in mind when thermal studies are

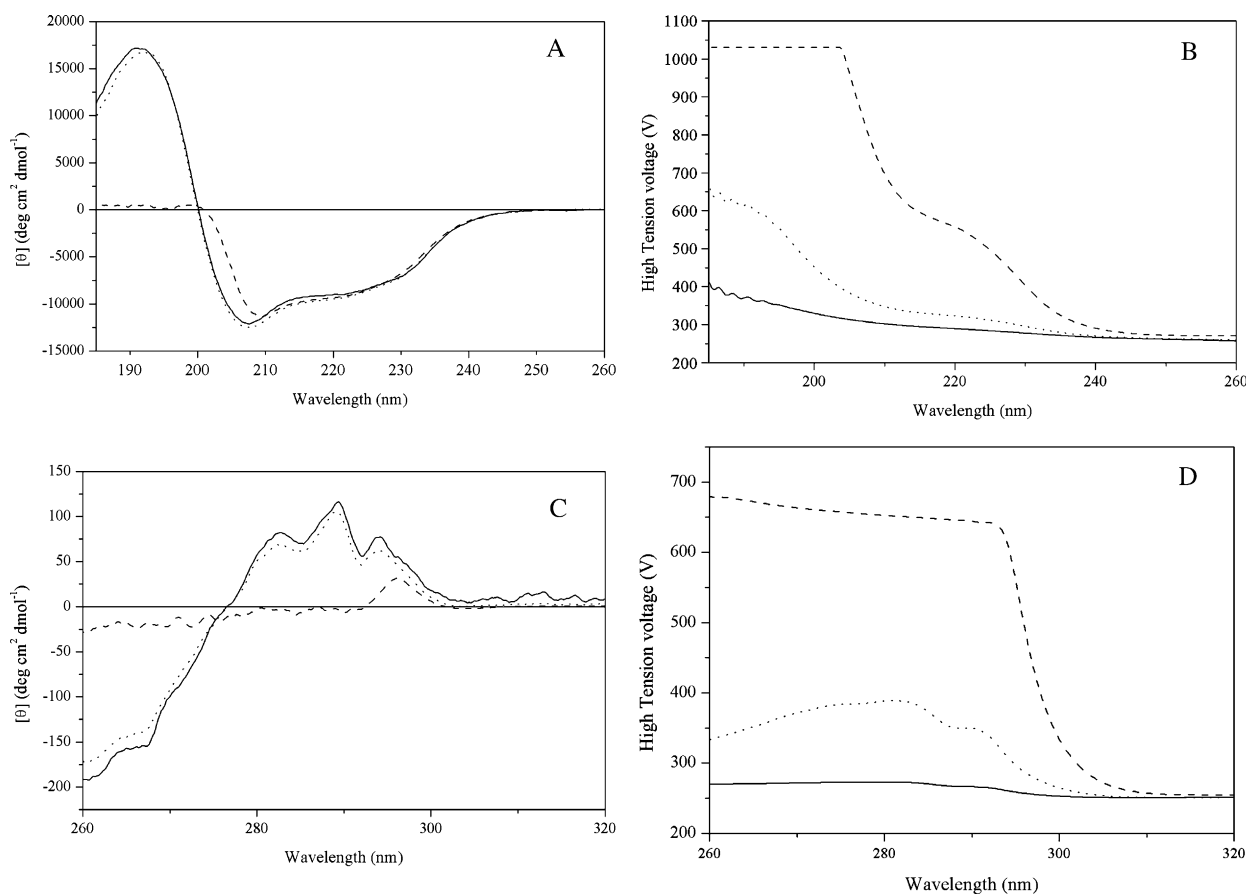


Fig. 12. The effects of protein concentration on CD spectra. The spectra shown are those of lysozyme in 50 mM sodium phosphate buffer, pH 7.5 and are recorded on a Jasco J-810 spectropolarimeter using 50 nm/min scan speed, 0.5 s time constant, 1 nm bandwidth and accumulation of 8 scans in each case. The solid, dotted and dashed lines refer to protein concentrations of 0.2 mg/ml, 1.0 mg/ml, and 5.0 mg/ml, respectively. Panel A shows the far UV CD spectra and panel B the corresponding High Tension voltage traces. Panel C shows the near UV CD spectra and panel D the corresponding High Tension voltage traces.

performed; these might require careful control experiments to be performed.

- There could be a small but significant change in the cell pathlength with temperature.
- The solvent could evaporate at high temperature, leading to a change in concentration.
- There could be inefficient heat transfer, so that temperature of the protein solution inside the cell might not reflect that of the external heat source.
- There could be an effect of temperature on the magnitude of the CD signals, as is found for CSA (Section 7.2).

9. General maintenance

9.1. Spectropolarimeter

The spectropolarimeter should be sited in a room where the temperature and humidity are kept reasonably constant (this may involve air-conditioning) and which is not subject to excessive mechanical vibration or the accumulation of atmospheric dust.

The high intensity light source will convert residual O₂ to O₃ which damages the optics. It is recommended that the instrument should be purged with N₂ at about 10–15 l/min for 15 min before switching on the light source.

Cylinders of O₂-free N₂ are sometimes used, but purging at 15–20 l/min will mean that these have to be replaced frequently. A more convenient source of N₂ gas of sufficient purity (>99.9%) is to boil it off from a liquid N₂ kept in a vessel of high (200 l) capacity. However, it is important to note that on safety grounds, the tank should not be in the room where the CD instrument is located, and there must be adequate ventilation to ensure that asphyxiation from the build-up of N₂ (and consequent decrease in O₂) does not occur. In addition, it is important that the last 10% of the liquid N₂ is not used, since that will be where impurities are concentrated as the N₂ gas is evaporated off [77]. In recent years, N₂ generators which take a supply of compressed air and use a carbon molecular sieve to remove O₂, CO₂ and H₂O vapour have come on the market. These are very convenient in terms of continuity of supply of N₂ gas of suitable purity, and carry no risk of build-up of N₂ in the laboratory, since air is merely separated into its components which are then re-combined.

After switching on, it is necessary to allow the machine to warm up and achieve stability (at least 30 min is recommended). The stability of the instrument over the course of a day should be checked by recording the drift in the baseline. Any such drift should be within the limits stated by the manufacturer. The magnitude of this effect may influence the design of experiments, for instance, if a change in a CD signal is being monitored over an extended period of time, or a large number of scans are being accumulated.

It is important to check the light source and optical system and take corrective action when performance begins to deteriorate. The lifetime of the light source will generally be around 1000 h of use, and will then give progressively poorer output at low wavelengths, which can be detected by an increase in the High Tension voltage. As far as the optical system is concerned, the first mirror (which gathers light from the lamp) is most rapidly degraded, and may require cleaning or replacement; this will be indicated by a decrease in the signal/noise ratio. It should be remembered that the light source is made of quartz and is filled with xenon gas at moderate pressure (5 to 10 atm) which is increased approximately 4-fold when the source is switched on. There is always a risk, albeit small, of an explosion when the source is handled and the appropriate precautions recommended by the manufacturers should be taken.

The spectropolarimeter should be regularly calibrated (Sections 7.1 and 7.2); sudden deterioration in performance will indicate the need for corrective action.

9.2. Cells

The cells (cuvettes) should be handled, used and stored carefully. It is important to minimise mechanical and thermal damage; in addition care should be taken to avoid scratching the surface of the cell if microsyringes are used during filling and emptying procedures.

Quartz cells should be washed and dried between runs. The typical procedure we have used involves a short (1 min) wash with concentrated HNO₃ (CARE, corrosive and powerful oxidising agent, check local safety rules), followed by thorough washing with distilled/deionised water, then ethanol, followed by drying with a vacuum pump. It is advisable not to use a laboratory supply of compressed air for drying, since this can contain traces of oil from the compressor; the oil deposits are difficult to remove from cells of short pathlength. The inclusion of the short HNO₃ wash is particularly important when using refillable cells of short (<0.05 cm) pathlength to avoid the formation of small air bubbles when transferring solutions.

It should also be possible to use proprietary cleaning formulations such as those provided by the suppliers of cells in place of the HNO₃. If there is evidence for the accumulation of “sticky” protein deposits during a series of experiments (e.g., from changes in the baseline), it is probably necessary to treat the cells with concentrated HNO₃ for extended periods (1–2 h or longer), followed by the thorough washing and drying procedures described above.

9.3. Record keeping

This is very important for efficient storage and retrieval of data; for example, CD data are often recorded before the exact concentration of protein is known, so that conversion into molar units and subsequent processing is carried out

later. In addition, by monitoring the performance of the instrument, a decline in the quality of the data can point to the need for further investigation or maintenance procedures. Above all, the value of the CD spectropolarimeter will be enhanced if there is a dedicated specialist in charge of the instrument who will be able to spot the warning signs of poor performance and undertake appropriate checks and maintenance procedures.

A new integrated software package (*CDtool*) has been developed which allows for more streamlined processing, analysis and archiving of CD data [78]. The package also allows the CD data to be associated with the PDB structural files for proteins. Apart from its use in structural studies of proteins, *CDtool* can provide a data audit trail of the type required to comply with good laboratory practice procedures. This will be especially important as CD becomes increasingly recognised as a valuable means of characterising therapeutic proteins and other products of commercial interest [78]. Recently, an international consortium has established a new resource, the Protein Circular Dichroism Data Bank (PCDDDB) which will act as a deposition and searchable data bank for CD spectra of biomacromolecules, located at <http://pcddb.cryst.bbk.ac.uk> [Wallace, Whitmore and Janes, in preparation]. It will include a set of online and stand-alone validation tools that provide reports on data quality. The aim of the PCDDDB is to provide open access and archiving facilities for CD spectra in a parallel fashion to those of the long established Protein Data Bank for X-ray crystallography and NMR data.

10. Assessment of the reliability of secondary structure analysis by CD

When assessing the secondary structure estimates for a given protein, it is important to bear in mind that there are several different algorithms available (Section 3.1) which use different computational approaches and may employ distinct data sets of reference proteins obtained over various spectral ranges. This makes it difficult to give hard and fast rules which apply under all circumstances. However, the reliability of the analysis can be judged by keeping the following points in mind [10,18].

- (a) The “goodness of fit” parameter NRMSD (normalised root mean square deviation) is the most useful single measure of how well the theoretical CD spectrum calculated from the derived secondary structure composition matches the experimental data over the entire wavelength range of interest. NRMSD is defined by $[\sum(\theta_{\text{exp}} - \theta_{\text{cal}})^2 / \sum(\theta_{\text{exp}})^2]^{1/2}$ and ranges from 0 (perfect fit) to 1 (no fit whatsoever). It is usual to regard a value of NRMSD above 0.25 as constituting an error in the analysis procedure; in practice values of less than 0.1, or ideally 0.05 or lower should be aimed for.
- (b) The *R* value (if provided) represents a measure of the appropriateness of the secondary structure composition derived from CD. It is defined as the sum of the differences (irrespective of sign) between the fractions of each of the different secondary structure elements (helix, sheet, and turns) derived from CD and X-ray analysis of a given protein. Low *R* values indicate that the analysis is appropriate; in practice values of 0.1 or lower should be aimed for. The *R* value can help to indicate which is the most appropriate algorithm and reference dataset to be used in any particular case. It may be, for example, that proteins with an unusual type of fold are not well represented in certain reference databases.
- (c) The comparison between the plots of calculated and experimental spectra should always be inspected to look for signs of systematic differences. These might arise, for example, if a membrane protein was being compared with a reference dataset of soluble proteins, since there could be small wavelength shifts in the absorption maxima reflecting the changes in polarity between an aqueous solvent and the environment of a lipid bilayer or detergent [58].
- (d) The results of using different algorithms should be compared. An analysis can be judged as reliable if very similar estimates for the secondary structural features are obtained from different algorithms. It would appear that a “consensus estimate” of secondary structures based on using the average fractional components obtained from different algorithms gives better NRMSD and *R* values than using one algorithm alone [79]. The use of larger reference protein datasets (containing both soluble and membrane proteins) should lead to improvements in the accuracy of protein secondary structure estimation from CD. The creation of a database which also includes data from infra-red spectroscopy would combine the complementary strengths of each technique to improve reliability and accuracy [80].

11. Summary of key points for reliable data collection and analysis

- Make sure that the protein is characterised; check its identity and purity (Section 4.1.2). Determine the concentration of the protein solution accurately (Section 6).
- Make sure the protein is dissolved in a suitable solvent system in which it is stable and which, ideally, makes only small contributions to the overall absorbance of the sample over the wavelength range of interest (Sections 4.2 and 8.2). In general, it is best to try to adapt the composition of the solvent system to the requirements of the protein, e.g., for activity or stability (Section 4.2).

- Make sure the CD instrument is properly maintained and has been calibrated in terms of CD amplitude and wavelength using suitable reference compounds (Sections 7.1 and 7.2). Make sure that the cells used are properly cleaned and that the pathlength is accurately known (Sections 7.3 and 9.2). Make sure that the instrument has had time to warm up and achieve operational stability.
- Make sure that the instrumental settings (time constant, bandwidth, scan speed, number of scans) are appropriate to give the best signal-to-noise ratio possible under the experimental conditions (Sections 8.1 and 8.2). Always run a solvent blank under the same conditions in order to correct the spectrum of the sample. Always note the pattern of the High Tension voltage trace so that any artefacts arising from excessive absorbance can be identified (Section 4.2). If necessary, adjust experimental conditions to obtain more reliable data.
- Make sure that the data recorded are suitably annotated with key experimental conditions noted and given suitable file names for efficient storage and retrieval at a later date (Section 9.3).
- Make sure that the output from structural analysis is scrutinised in order to ascertain its reliability (Section 10).

Acknowledgements

The work of the CD facility at Glasgow has been supported by the Biotechnology and Biological Sciences Research Council of the U.K. We would like to thank Bonnie Wallace, Alison Rodger and Liz Rideout for helpful discussions and providing unpublished data, and Robert Matthews and Jill Zitzewitz for providing data files.

References

- [1] M. Groll, L. Ditzel, J. Löwe, D. Stock, M. Bochtler, H.D. Bartunik, R. Huber, Structure of 20S proteasome from yeast at 2.4 Å resolution, *Nature* 386 (1997) 463–471.
- [2] P. Nissen, N. Ban, J. Hansen, P.B. Moore, T.A. Steitz, The structural basis of ribosome activity in peptide bond synthesis, *Science* 289 (2000) 920–930.
- [3] A. Polyanchko, H. Wieser, The FTIR/VCD spectroscopy as an informative tool for the investigation of large supramolecular complexes of biological macromolecules. *Biopolymers* (in press).
- [4] T. Arvinte, T.T. Bui, A.A. Dahab, B. Demeule, A.F. Drake, D. Elhag, P. King, The multi-mode polarization modulation spectrometer: part 1. A simultaneous detection of absorption, turbidity, and optical activity, *Anal. Biochem.* 332 (2004) 46–57.
- [5] G.D. Fasman (Ed.), *Circular Dichroism and The Conformational Analysis of Biomolecules*, Plenum Press, New York, 1996, 738 pp.
- [6] N. Berova, K. Nakanishi, R.W. Woody (Eds.), *Circular Dichroism: Principles and Applications*, 2nd ed., Wiley-VCH, New York, 2000, 877 pp.
- [7] A. Rodger, B. Nordén, *Circular Dichroism and Linear Dichroism*, Oxford Univ. Press, Oxford, 1997.
- [8] S.M. Kelly, N.C. Price, The use of circular dichroism in the investigation of protein structure and function, *Curr. Prot. Pept. Sci.* 1 (2000) 349–384.
- [9] W.C. Johnson Jr., in: G.D. Fasman (Ed.), *Circular Dichroism Instrumentation in Circular Dichroism and The Conformational Analysis of Biomolecules*, Plenum Press, New York, 1996, pp. 636–652.
- [10] N. Sreerama, R.W. Woody, Computation and analysis of protein circular dichroism spectra, *Methods Enzymol.* 383 (2004) 318–351.
- [11] N. Sreerama, R.W. Woody, On the analysis of membrane protein circular dichroism spectra, *Protein Sci.* 13 (2004) 100–112.
- [12] N. Sreerama, R.W. Woody, A self-consistent method for the analysis of protein secondary structure from circular dichroism, *Anal. Biochem.* 209 (1993) 32–44.
- [13] P. Manavalan, W.C. Johnson Jr., Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra, *Anal. Biochem.* 167 (1987) 76–85.
- [14] W.C. Johnson, Analyzing protein circular dichroism spectra for accurate secondary structures, *Proteins: Struct. Funct. Genet.* 35 (1999) 307–312.
- [15] M.A. Andrade, P. Chacón, J.J. Merelo, F. Morán, Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised neural network, *Protein Eng.* 6 (1993) 383–390.
- [16] S.W. Provencher, J. Glöckner, Estimation of globular protein secondary structure from circular dichroism, *Biochemistry* 20 (1981) 33–37.
- [17] A. Lobley, L. Whitmore, B.A. Wallace, DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra, *Bioinformatics* 18 (2002) 211–212.
- [18] L. Whitmore, B.A. Wallace, DICHROWEB, An online server for protein secondary structure analyses from circular dichroism spectroscopic data, *Nucleic Acids Res.* 32 (2004) W668–W673.
- [19] W.C. Johnson Jr., Analyzing protein circular dichroism spectra for accurate secondary structures, *Proteins: Struct. Funct. Genet.* 35 (1999) 307–312.
- [20] B.A. Wallace, Synchrotron radiation circular dichroism spectroscopy as a tool for investigating protein structures, *J. Synchrotron. Radiat.* 7 (2000) 289–295.
- [21] B.A. Wallace, R.W. Janes, Synchrotron radiation circular dichroism spectroscopy of proteins: secondary structure, fold recognition and structural genomics, *Curr. Opin. Chem. Biol.* 5 (2001) 567–571.
- [22] C.J. Barrow, A. Yasuda, P.T. Kenny, M.G. Zagorski, Solution conformations and aggregational properties of synthetic amyloid beta-peptides of Alzheimer's disease. Analysis of circular dichroism spectra, *J. Mol. Biol.* 225 (1992) 1075–1093.
- [23] C. Krittanai, W.C. Johnson Jr., Correcting the circular dichroism spectra of peptides for contributions of absorbing side chains, *Anal. Biochem.* 253 (1997) 57–64.
- [24] A.Y. Woody, R.W. Woody, Individual tyrosine side-chain contributions to circular dichroism of ribonuclease, *Biopolymers* 72 (2003) 500–513.
- [25] P.-O. Freskgård, L.-G. Mårtensson, P. Jonasson, B.-H. Jonsson, U. Carlsson, Assignment of the contribution of the tryptophan residues to the circular dichroism spectrum of human carbonic anhydrase II, *Biochemistry* 33 (1994) 14281–14288.
- [26] D.H. Boxer, H. Zhang, D.G. Gourley, W.N. Hunter, S.M. Kelly, N.C. Price, Sensing of remote oxyanion binding at the DNA binding domain of the molybdate-dependent transcriptional regulator, *ModE*, *Org. Biomol. Chem.* 2 (2004) 2829–2837.
- [27] O.B. Ptitsyn, Molten globule and protein folding, *Adv. Prot. Chem.* 47 (1995) 83–229.
- [28] N.E. Price, N.C. Price, S.M. Kelly, J.M. McDonnell, The key role of protein flexibility in modulating IgE interactions, *J. Biol. Chem.* 280 (2005) 2324–2330.
- [29] A.K. Dunker, C.J. Brown, J.D. Lawson, L.M. Iakoucheva, Z. Obradovic, Intrinsic disorder and protein function, *Biochemistry* 41 (2002) 6573–6582.

- [30] L.A. Andersson, J.A. Peterson, Active-site analysis of ferric P450 enzymes: hydrogen-bonding effects on the circular dichroism spectra, *Biochem. Biophys. Res. Commun.* 211 (1995) 389–395.
- [31] R.G. Alden, E. Johnson, V. Nagarajan, W.W. Parson, C.J. Law, R.J. Cogdell, Calculations of spectroscopic properties of the LH2 bacteriochlorophyll–protein antenna complex from *Rhodospseudomonas acidophila*, *J. Phys. Chem.* B101 (1997) 4667–4680.
- [32] D. Missiakas, J.-M. Betton, P. Minard, J.M. Yon, Unfolding–refolding of the domains in yeast phosphoglycerate kinase: comparison with the isolated engineered domains, *Biochemistry* 29 (1990) 8683–8689.
- [33] A.W. Munro, J.G. Lindsay, J.R. Coggins, S.M. Kelly, N.C. Price, Structural and enzymological analysis of the interaction of isolated domains of cytochrome P-450 BM3, *FEBS Lett.* 343 (1994) 70–74.
- [34] C.N. Pace, Determination and analysis of urea and guanidine hydrochloride denaturation curves, *Methods Enzymol.* 131 (1986) 266–280.
- [35] C.M. Johnson, A.R. Fersht, Protein stability as a function of denaturant concentration: the thermal stability of barnase in the presence of urea, *Biochemistry* 34 (1995) 6795–6804.
- [36] A. Giletto, C.N. Pace, in: N.C. Price (Ed.), *Protein Stability in Proteins Labfax*, Bios Scientific Publishers, Oxford, 1996, pp. 233–239.
- [37] J. Hope, M.S. Shearman, H.C. Baxter, A. Chong, S.M. Kelly, N.C. Price, Cytotoxicity of prion protein peptide (PrP106-126) differs in mechanism from the cytotoxic activity of the Alzheimer's disease amyloid peptide, A β 25–35, *Neurodegeneration* 5 (1996) 1–11.
- [38] M.J. Pandya, E. Cerasoli, A. Joseph, R.G. Stoneman, E. Waite, D.N. Woolfson, Sequence and structural duality: designing peptides to adopt two stable conformations, *J. Am. Chem. Soc.* 126 (2004) 17016–17024.
- [39] S.E. Radford, Protein folding: progress made and promises ahead, *Trends Biochem. Sci.* 25 (2000) 611–618.
- [40] P. Jemth, S. Gianni, R. Day, B. Li, C.M. Johnson, V. Daggett, A.R. Fersht, Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation, *Proc. Natl. Acad. Sci. U. S. A.* 101 (2004) 6450–6455.
- [41] C.M. Dobson, Experimental investigation of protein folding and misfolding, *Methods* 34 (2004) 4–14.
- [42] R.K. Scopes, *Protein Purification: Principles and Practice*, 3rd ed., Springer, New York, 1994, 380 pp. (See Chap. 9 (pages 270–282)).
- [43] N.C. Price, L. Stevens, *Fundamentals of Enzymology*, 3rd ed., Oxford Univ. Press, Oxford, 1999, 478 pp. (See Chap. 2 (pages 15–46)).
- [44] T.K. Chaudhuri, K. Horii, T. Yoda, M. Arai, S. Nagata, T.P. Terada, H. Uchiyama, T. Ikura, K. Tsumoto, H. Kataoka, M. Matsushima, K. Kuwajima, I. Kumagai, Effect of the extra N-terminal methionine residue on the stability and folding of recombinant alpha-lactalbumin expressed in *Escherichia coli*, *J. Mol. Biol.* 285 (1999) 1179–1194.
- [45] R.K. Scopes, *Protein Purification: Principles and Practice*, 3rd ed., Springer, New York, 1994, 380 pp. (See Chap. 2 (pages 22–43)).
- [46] B.A. Wallace, D. Mao, Circular dichroism analyses of membrane proteins: an examination of differential light scattering and absorption flattening effects in large membrane vesicles and membrane sheets, *Anal. Biochem.* 142 (1984) 317–328.
- [47] B.A. Wallace, C.L. Teeters, Differential absorption flattening optical effects are significant in the circular dichroism spectra of large membrane fragments, *Biochemistry* 26 (1987) 65–70.
- [48] L. Stevens, in: R. Eisenthal, M.J. Danson (Eds.), *Buffers and The Determination of Protein Concentrations in Enzyme Assays: A Practical Approach*, Oxford Univ. Press, Oxford, 1992, pp. 317–335.
- [49] R.K. Scopes, *Protein Purification: Principles and Practice*, 3rd ed., Springer, New York, 1994, 380 pp. (See Chap. 12 (pages 310–345)).
- [50] R.K. Scopes, Protein purification in the nineties, *Biotechnol. Appl. Biochem.* 23 (1996) 197–204.
- [51] T. Arakawa, S.N. Timasheff, Stabilization of protein structure by sugars, *Biochemistry* 21 (1982) 6536–6544.
- [52] D.W. Bolen, Effects of naturally occurring osmolytes on protein stability and solubility: issues important in protein crystallization, *Methods* 34 (2004) 312–322.
- [53] M.C. Gurau, S.-M. Lim, E.T. Castellana, F. Albertorio, S. Kataoka, P.S. Cremer, On the mechanism of the Hofmeister effect, *J. Am. Chem. Soc.* 126 (2004) 10522–10523.
- [54] N.E. Good, G.D. Winget, W. Winter, T.N. Connolly, S. Izawa, R.M. Singh, Hydrogen ion buffers for biological research, *Biochemistry* 5 (1966) 467–477.
- [55] M.F. Ansell (Ed.), *Rodd's Chemistry of Carbon Compounds*, Vol. IV part C, 2nd ed., Elsevier, Amsterdam, 1986 (see page 133).
- [56] F. Ahmad, Free energy changes in ribonuclease A denaturation. Effect of urea, guanidine hydrochloride, and lithium salts, *J. Biol. Chem.* 258 (1983) 11143–11146.
- [57] J.M. Scholtz, R.L. Baldwin, Perchlorate-induced denaturation of ribonuclease A: investigation of possible folding intermediates, *Biochemistry* 32 (1993) 4604–4608.
- [58] B.A. Wallace, J. Lees, A.J.W. Orry, A. Lobley, R.W. Jones, Analyses of circular dichroism spectra of membrane proteins, *Prot. Sci.* 12 (2003) 875–884.
- [59] W.L. Mattice, J.M. Riser, D.S. Clark, Conformational properties of the complexes formed by proteins and sodium dodecyl sulfate, *Biochemistry* 15 (1976) 4264–4272.
- [60] B. Jirgensons, Conformational transitions of non-helical proteins effected by dodecyl sulfate. Circular dichroism of alpha1-acid glycoprotein, Bence Jones protein, carbonic anhydrase B, deoxyribonuclease, pepsinogen, and plasminogen, *Biochim. Biophys. Acta* 434 (1976) 58–68.
- [61] M. Buck, Trifluoroethanol and colleagues: cosolvents come of age, *Q. Rev. Biophys.* 31 (1998) 297–355.
- [62] A. Starzyk, W. Barber-Armstrong, M. Sridharan, S.M. Decatur, Spectroscopic evidence for backbone desolvation of helical peptides by 2,2,2-trifluoroethanol: an isotope-edited FTIR study, *Biochemistry* 44 (2005) 369–376.
- [63] N.C. Price, in: P.C. Engel (Ed.), *The Determination of Protein Concentration in Enzymology Labfax*, Bios Scientific Publishers, Oxford, 1996, pp. 34–41.
- [64] R.K. Scopes, Measurement of protein by spectrophotometry at 205 nm, *Anal. Biochem.* 59 (1974) 277–282.
- [65] S.C. Gill, P.H. von Hippel, Calculation of protein extinction coefficients from amino acid sequence data, *Anal. Biochem.* 182 (1989) 319–326.
- [66] C.N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, How to measure and predict the molar absorption coefficient of a protein, *Prot. Sci.* 4 (1995) 2411–2423.
- [67] H. Mach, C.R. Middaugh, R.V. Lewis, Statistical determination of the average values of the extinction coefficients of tryptophan and tyrosine in native proteins, *Anal. Biochem.* 200 (1992) 74–80.
- [68] A.J. Miles, F. Wien, J.G. Lees, A. Rodger, R.W. Janes, B.A. Wallace, Calibration and standardisation of synchrotron radiation circular dichroism and conventional circular dichroism spectrophotometers, *Spectroscopy* 17 (2003) 653–661.
- [69] C. Jones, D. Schiffmann, A. Knight, S. Windsor, Val-CiD best practice guide: CD spectroscopy for the quality control of biopharmaceuticals (2004) National Physical Laboratory, Teddington, Middlesex, U.K., Crown Copyright 2004 (ISSN 1744–0602).
- [70] P.H. Schippers, H.P.J.M. Dekkers, Direct determination of absolute circular dichroism data and calibration of commercial instruments, *Anal. Chem.* 53 (1981) 778–782.
- [71] A.J. Miles, F. Wien, J.G. Lees, B.A. Wallace, Calibration and standardisation of synchrotron radiation circular dichroism and conventional circular dichroism spectrophotometers. Part 2: factors affecting magnitude and wavelength, *Spectroscopy* 19 (2005) 43–51.
- [72] A.J. Miles, F. Wien, B.A. Wallace, Redetermination of the extinction coefficient of camphor-10-sulfonic acid, a calibration standard for circular dichroism spectroscopy, *Anal. Biochem.* 335 (2004) 338–339.
- [73] J.P. Hennessey Jr., W.C. Johnson Jr., Experimental errors and their effect on analysing circular dichroism spectra of proteins, *Anal. Biochem.* 125 (1982) 177–188.

- [74] G.W. Haupt, An alkaline solution of potassium chromate as a transmittancy standard in the ultraviolet, *J. Res. Natl. Bur. Stand.* 48 (1952) 414–423.
- [75] M.A. Weiss, T. Ellenberger, C.R. Wobbe, J.P. Lee, S.C. Harrison, K. Struhl, Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA, *Nature* 347 (1990) 575–578.
- [76] J.A. Zitzewitz, O. Bilsel, J. Luo, B.E. Jones, C.R. Matthews, Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy, *Biochemistry* 34 (1995) 12812–12819.
- [77] W.C. Johnson Jr., Protein secondary structure and circular dichroism: a practical guide, *Proteins: Struct. Funct. Genet.* 7 (1990) 205–214.
- [78] J.G. Lees, B.R. Smith, F. Wien, A.J. Miles, B.A. Wallace, CDtool—an integrated software package for circular dichroism spectroscopic data processing, analysis and archiving, *Anal. Biochem.* 332 (2004) 285–289.
- [79] J.G. Lees, PhD Thesis, University of London (2005).
- [80] K.A. Oberg, J.-M. Ruyschaert, E. Goormaghtigh, The optimization of protein secondary structure determination with infrared and circular dichroism spectra, *Eur. J. Biochem.* 271 (2004) 2937–2948.
- [81] F.X. Schmid, in: T.E. Creighton (Ed.), *Spectral Methods of Characterizing Protein Conformation and Conformational Changes in Protein Structure: A Practical Approach*, Oxford Univ. Press, Oxford, 1989, pp. 251–285.
- [82] K. Rosenheck, P. Doty, The far ultraviolet absorption spectra of polypeptide and protein solutions and their dependence on conformation, *Proc. Natl. Acad. Sci. U. S. A.* 47 (1961) 1775–1785.
- [83] R.P. Buck, S. Singhadeja, L.B. Rogers, Ultraviolet absorption spectra of some inorganic ions in aqueous solutions, *Anal. Chem.* 26 (1954) 1240–1242.
- [84] T. Krell, M.J. Horsburgh, A. Cooper, S.M. Kelly, J.R. Coggins, Localization of the active site of type II dehydroquinases. Identification of a common arginine-containing motif in the two classes of dehydroquinases, *J. Biol. Chem.* 271 (1996) 24492–24497.
- [85] L.A. Anderson, T. Palmer, N.C. Price, S. Bornemann, D.H. Boxer, R.N. Pau, Characterisation of the molybdenum-responsive ModE regulatory protein and its binding to the promoter region of the modABCD (molybdenum transport) operon of *Escherichia coli*, *Eur. J. Biochem.* 246 (1997) 119–126.