International comparability in spectroscopic measurements of protein structure by circular dichroism: CCQM-P59.1

IOP PUBLISHING METROLOGIA
Metrologia **47** (2010) 631–641 doi:10.1088/0026-1394/47/6/001

International comparability in spectroscopic measurements of protein structure by circular dichroism: CCQM-P59.1

Jascindra Ravi[1], Paulina D Rakowska[1], Tommaso Garfagnini[1], Bruno Baron[2], Philippe Charlet[3], Christopher Jones[4], Stoyan Milev[5], Julie DeSa Lorenz[5], David Plusquellic[6], Frank Wien[7], Liqing Wu[8], Curtis W Meuse[9] and Alex E Knight[1,10]

[1] Quality of Life Division, National Physical Laboratory, Teddington, Middlesex, TW11 0LW, UK
[2] Institut Pasteur, Plate Forme de Biophysique des Macromolécules et de leurs Interactions, CNRS URA 2185, 25 rue du Dr Roux, 75015 Paris, France
[3] Laboratoire National d'Essais, Paris, France
[4] Laboratory for Molecular Structure, National Institute for Biological Standards and Control, South Mimms EN6 3QG, UK
[5] Olis, Inc. Bogart, GA, USA
[6] Optical Technology Division, National Institute of Standards and Technology, Gaithersburg, MD 20899-8443, USA
[7] Synchrotron SOLEIL, L'orme des Merisiers Saint-Aubin, BP48, 91192 Gif-sur-Yvette Cedex, France
[8] NIM, Beijing, People's Republic of China
[9] Biochemical Science Division, National Institute for Standards and Technology, Gaithersburg, MD 20899-8312, USA

E-mail: alex.knight@npl.co.uk

Received 14 May 2010, in final form 26 August 2010
Published 4 October 2010
Online at stacks.iop.org/Met/47/631

Abstract
Circular dichroism (CD) is a spectroscopic technique that is widely used to obtain information about protein structure, and hence is an important tool with many applications, including the characterization of biopharmaceuticals. A previous inter-laboratory study, CCQM-P59, showed that there was a poor level of comparability between laboratories in CD spectroscopy. In a follow-up study reported here, we achieved our goal of demonstrating improved comparability and data quality, primarily by addressing the problems identified in the previous study, which included cell path-length measurement, instrument calibration and good practice in general. Multivariate analysis techniques (principal component analysis and soft independent modelling of class analogies) were shown to be useful in comparing large spectral data sets and in classifying spectra. However, our results also show that there is more work to be done to improve confidence in the technique as the discrepancies observed were partially due to systematic effects, which the statistical approaches do not consider. We therefore conclude that there is a need for an improved understanding of the uncertainties in CD measurement.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

1.1. Circular dichroism (CD) spectroscopy

CD is defined as the difference in the absorption of left and right-handed circularly polarized light by a sample [1, 2]. This technique is often applied to the study of chiral molecules; there is a particularly wide range of applications in the spectroscopy of biomolecules. The most common use in biology is to obtain *a priori* information about the secondary structure of protein molecules [3]. CD spectra in the far UV region (typically 180 nm to 260 nm) can give an overall indication of the secondary structure content, because the principal

[10] Author to whom any correspondence should be addressed.

chromophore in this region is the peptide backbone of the molecule. A wide variety of algorithms exist for extracting structural information from these spectra [3, 4].

Useful information about protein structure can also be obtained from other spectral regions. The visible region is useful for those proteins that have prosthetic groups or ligands that absorb in this region, such as haems. However, much more widely used is the near UV region, typically 240 nm to 320 nm. The dominant chromophores in this region are the side chains of the amino acids, particularly aromatic residues such as tryptophan, and also cystines. Although *a priori* structural interpretation is difficult, the spectral signature in this region provides a useful 'fingerprint', which may be very sensitive to changes in protein structure. This region is therefore often useful for stability studies, or in the characterization, comparability assessment and process development of biopharmaceuticals.

CD in the infrared region is known as vibrational CD, and is a distinct technique, measured with different instrumentation, and is beyond the scope of this paper.

1.2. Accuracy of CD measurements

Unfortunately, CD is a challenging phenomenon to measure. The greatest difficulty arises from the relatively small size of the effect. Typically the difference in absorption between left- and right-handed circularly polarized light is of the order of 1 part in 10^3 or less [5]. Commercially available instrumentation takes a number of different approaches to measuring the effect, some of them rather indirect [5]. Therefore CD measurements are usually not absolute, but are calibrated using a reference material.

Regrettably, the reference materials that are used are not truly fit for purpose. The reference values used are based on the literature [6] and have unknown uncertainties. Ideally, such reference values should be traceable to the SI. Furthermore, the standards typically only provide calibration at one or two wavelengths, which are not necessarily the critical wavelengths for the measurements being made. Evidence from the literature shows that there may be considerable wavelength-dependent variation of calibration between instruments [7, 8]. Also there are issues relating to the correct formulation and storage of these materials [9].

Another issue that appears to bedevil the field is a lack of good practice in CD measurement. The technique requires a degree of knowledge to be carried out and interpreted correctly. Unfortunately, it is often the case that users are inadequately trained and poor measurements are common, even in the published literature [10]. There have been several previous attempts to formulate good practice in the measurement of CD [10, 11].

With all measurement techniques, the ideal is that they should be traceable to the SI. For example, UV/visible absorbance spectroscopy is traceable, allowing absolute measurements to be made, and there are a number of standards to support this. Whilst CD spectroscopy of proteins has a relatively long history (dating back to the 1960s [12]), it does not have such well-established measurement traceability. Indeed, the measurement problems of the technique have been recognized for many years, but few solutions have been found.

What are the consequences of such a lack of traceability? If measurements are not traceable to the same absolute reference, then measurements made in different laboratories, or with different instrumentation, or at different times, or under different conditions, cease to be comparable. If the data are not comparable, then this limits the applicability of the data. For example, scale (magnitude) differences to CD spectra can cause different results in secondary structure predictions [4]. Such scale differences cannot be unambiguously assigned to calibration, path length, concentration or structural differences, without additional data. Furthermore wavelength differences, or distortions in spectral shapes, will also influence results [4, 7, 8].

Similar problems will also occur in other applications of CD. For example, where CD is used as a tool to compare different production batches of biopharmaceuticals, artefactual differences between spectra that are collected at different times, or in different laboratories, will cause structurally identical samples to appear to be different. This could lead to the rejection of a batch of a drug that is actually sound—at huge expense to the manufacturer. Conversely, if spectral comparability is poor, tolerances on spectral differences could be set too high, resulting in problems being missed. The net result is to undermine industry confidence in a technique that should be extremely useful.

1.3. Inter-laboratory comparison of CD

Given the importance of this technique in such a commercially and medically significant sector, and our concerns about the reliability of the measurement, we set out to investigate the seriousness of the problem on an international scale. The international system of units—the SI—is administered by the BIPM[11]. Development and support of the SI in various scientific areas is the responsibility of the consultative committees; the committee responsible for chemical measurement is the CCQM, which is responsible for the mole, the SI unit for amount of substance. The Bioanalysis Working Group (BAWG) of the CCQM is responsible for international metrology activities in bioanalysis.

The Bioanalysis Working Group and the CCQM approved an inter-laboratory pilot study of CD, designated CCQM-P59. In this previous study [13], we investigated the comparability between laboratories of CD measurements of protein solutions, and we showed that there were many problems which were preventing comparable measurements from being made. We also demonstrated the utility of multivariate data analysis techniques, such as principal component analysis (PCA) and soft independent modelling of class analogies (SIMCA),

[11] Abbreviations: CD: circular dichroism; PCA: principal component analysis; SIMCA: soft independent modelling of class analogies; NMI: national measurement institute; NPL: National Physical Laboratory; ACS: ammonium *d*-10-camphorsulfonate; SI: Système International d'Unités (International System of Units); CCQM: Comité consultatif pour la quantité de matière—métrologie en chimie (Consultative Committee for Amount of Substance—Metrology in Chemistry); BIPM: Bureau International des Poids et Mesures (International Bureau for Weights and Measures).

to this type of data. In response to the observed poor level of comparability, the Bioanalysis Working Group agreed to support a follow-up study with the aim of demonstrating improved performance. In this paper, we describe the design and results of this study, which was designated CCQM-P59.1.

1.4. Aims of study

The aim of a pilot study is typically to compare the ability of NMIs to make a particular measurement. This, indeed, was the principal goal of this study, although some expert non-NMI laboratories also participated. Furthermore, we also attempted to identify the source of any errors in participant measurements. We were also keen to investigate the utility of pattern recognition techniques in analysing large spectral data sets of this type. In this study, we chose to achieve these aims by distributing samples to the participating laboratories for them to measure.

Because we were primarily interested in the biological applications of the technique, we chose to distribute biological samples to the study participants. Measuring biologicals (in this case, proteins) introduces additional concerns, such as sample stability and light scattering. These issues are exacerbated by the necessity of shipping the samples internationally. A key element of the study design, therefore, was to control for changes in the samples.

In general, the philosophy of the P59.1 study was to improve upon the design of the previous study, with the aim of progressing towards comparable CD measurement among the national measurement institutes and in the wider user community. Since many of the difficulties seen previously were a result of poor practice, we set out to provide good practice guidance and more detailed protocols to the participants. For example, we provided protocols for the measurement of cell path length and for instrument calibration. Similarly, we made some changes to the study protocol and the questionnaire to improve their clarity. Some issues with the distribution of the samples were experienced in the first study, and although these did not markedly affect the results obtained, we used improved packaging methods that kept the samples chilled for longer and reduced leakage.

1.5. Study measurand

The most significant change we made from P59, however, was to define the measurand of the study as the molar CD spectrum of the samples, rather than the ellipticity spectra that were reported previously. CD is defined as the differential absorbance between left and right circularly polarized light:

$$\Delta A = A_{\mathrm{L}} - A_{\mathrm{R}}, \tag{1}$$

where A_{L}, A_{R} and ΔA are dimensionless quantities. However, for historical reasons, CD measurements are typically reported as ellipticities, in units of millidegrees[12] (mdeg). These are related very simply to the dimensionless CD as follows:

$$\theta = 32\,980 \cdot \Delta A, \tag{2}$$

[12] The millidegree (abbreviated to mdeg) is a non-standard unit that is widely used in the field of CD spectroscopy for historical reasons [5]. 1 mdeg is equal to 10^{-3} degrees or approximately 17.5 µrad.

where θ is in mdeg. As with absorbance, CD follows a form of the Beer–Lambert law,

$$\Delta A = \Delta\varepsilon c \ell. \tag{3}$$

If the concentration c is expressed in mol dm^{-3} and the path length ℓ in cm, then $\Delta\varepsilon$ is the molar CD with units of mol^{-1} dm^3 cm^{-1}. Using this quantity, we are taking into account any variations in the cell path length used by the participants, *provided that this has been measured, and measured correctly!* There should be no variation in the concentration, as all participants will be measuring aliquots of the same solutions. We feel that this measurand is the most appropriate for inter-laboratory comparisons, as derived measurands, such as secondary structure composition, are dependent on further processing steps and sequence databases with unknown, and probably complex, uncertainties.

The concentration used can either be the concentration of the protein molecules or the concentration of peptide bonds. The latter is usually the most appropriate in the far UV region, where the principal chromophores are the peptide bonds themselves.

2. Methods and materials[13]

2.1. Study design

A significant concern influencing the design of the study was that changes in the samples during shipping or storage over the course of the study would be responsible for differences between the data from participants. To eliminate this concern, a number of steps were taken. In the previous study, all the samples were subjected to a battery of stability tests [13]. Stability trials were not necessary prior to this study, as the same proteins were used; however, the stability of the samples was monitored by repeated measurements throughout the study. Furthermore, participants were asked to return unused samples to NPL for re-measurement. This last step allowed us to place an upper limit on the changes in the sample at the time of measurement in the participants' laboratory.

2.2. Study protocol and data collection

Participants were provided with a detailed protocol for the experimental work in the study. This was revised from the previous study to improve clarity. The detailed instrument parameters used are given in table 1. These apply to instruments from manufacturer A (see also table 4); where instruments from other manufacturers were used, the nearest equivalent parameter sets were chosen.

Participants were encouraged, but not required, to follow these parameters. However, variations from these parameters were captured in an electronic questionnaire completed by

[13] Certain commercial materials, instruments and equipment are identified in this paper in order to specify the experimental procedure as completely as possible. In no case does such identification imply recommendation or endorsement by the National Physical Laboratory or the National Institute of Standards and Technology, nor does it imply that the material, instrument or equipment identified is necessarily the best available for the purpose.

Table 1. Instrument parameters in study protocol.

	Sample		
	Proteins		
Parameter	Far UV	Near UV	ACS
Cell path length/mm	0.10	10	10
Wavelength range/nm	180 to 260	240 to 360	200 to 400
Accumulations	6	4	1
Response time/s	1	1	1
Bandwidth/nm	1	1	1
Sensitivity	Standard	Standard	Standard
Data pitch/nm	0.1	0.1	0.1
Scan speed/nm min^{-1}	50	50	50

the participants. The questionnaires additionally recorded information necessary for the calculation of uncertainties, where participants were able to provide it. As previously, all of the data and the electronic questionnaires were returned to the organizers by electronic mail.

2.3. Study samples

The composition of the samples used in this study is given in table 2. The same proteins as previously [13] were used, but a different ratio of proteins was used in the mixed sample (003). In addition to the three protein samples, one sample of a commonly used CD calibrant (ACS[10]) and corresponding solutions for use as blanks were distributed. All proteins were obtained from Sigma (UK). ACS was obtained from JASCO (Great Dunmow, UK). All weights were measured using a Sartorius (Göttingen, Germany) 'Genius' or CP2P balance, as appropriate, and solution volumes were measured using volumetric flasks and calibrated pipettes. All solutions were filter-sterilized to prevent bacterial or fungal growth during the study. Proteins were made up as concentrated stock solutions at 5 mg ml^{-1} in the phosphate buffer, before dilution into the final solutions[14]. Samples were stored at 4 °C and protected from light.

To avoid the issues experienced with the shipping of the samples experienced previously, we used polyurethane boxes with rigid cells filled with coolants (Emball'infor, Glisy, France). Sealing the lids of the sample tubes with Parafilm and placing them in a rack prevented leakage of the samples.

2.4. Instrument calibration

Where appropriate, CD instruments were calibrated using a solution of ammonium D-10-camphorsulfonate (ACS) as described previously [9–11, 13]. Briefly, 60.0 mg of ACS (JASCO, Great Dunmow, Essex) were dissolved in 100 ml of ultra-pure water in a volumetric flask. Solutions were stored at 4 °C until used and allowed to warm to room temperature before measurement. CD spectra of water (for baselines) and the ACS solution were acquired in a 1 cm cell. If necessary, the instrument was adjusted to give a reading of 190.4 mdeg at 291 nm. This protocol was recommended to study participants and also used in the organizing laboratory. Some of the participants' instruments do not require calibration as they operate on a direct subtraction approach [5].

[14] Since all participants received aliquots of the same solutions, uncertainties in protein concentration were not a factor in the comparability of participant results, and were therefore not evaluated. The purity of the proteins was, however, checked qualitatively by SDS-PAGE.

2.5. Path-length measurement

Path-length measurements were made by the methods recommended elsewhere [10, 11]. For 1 cm cells, the potassium chromate method was used. For 0.01 cm cells, where possible, an interference method was used; otherwise the chromate method was applied. Again, these protocols were distributed to the participants and used in the organizing laboratory.

2.5.1. Potassium chromate method. Briefly, 194.19 mg of analytical grade potassium chromate were dissolved in 50 ml of ultra-pure water in a volumetric flask, with one pellet of potassium hydroxide, and stored overnight at 4 °C. This 0.02 mol dm^{-3} solution was used to measure path lengths of 0.01 cm cells, or diluted to 0.2 mmol dm^{-3} to measure 1 cm cells.

Correcting for the absorbance of the cell (measured using a water blank), three replicate absorbance measurements were made at 372 nm. The path length was then calculated using the Beer–Lambert law, given an extinction coefficient of 4830 mol^{-1} dm^3 cm^{-1}.

2.5.2. Interference method. Briefly, the empty cell was placed in visible or near infrared spectrometer and a spectrum obtained at 800 nm to 900 nm, or at longer wavelengths if necessary. If a sinusoidal interference pattern was seen, two widely separated fringes were chosen, with peak wavelengths w_1, w_2 and n intervening fringes, and the path length was then calculated from

$$\ell = \frac{n w_1 w_2}{2(w_2 - w_1)}. \tag{4}$$

2.6. Analysis of study data

Data processing, including collation, baseline subtraction, interpolation and curve fitting, were performed in MATLAB (The MathWorks, Natick, MA, USA), using specially written scripts. Data were stored in Excel spreadsheet format (Microsoft, Redmond, WA, USA). Typically an average was taken of the two baselines provided by a participant, and this was subtracted from each spectrum. Curve fitting was used to derive calibration values for the participants from the ACS spectra. Spectra were truncated to 280 nm to 300 nm and fitted with a Gaussian model using the MATLAB Curve Fitting Toolbox.

After the raw (ellipticity) data had been collated and baseline subtracted, the protein concentrations and path lengths were used to convert the protein spectral data to molar CD, and it is this data that was used for subsequent analyses. For the far UV data, concentrations were expressed in terms of peptide bonds; for the near UV, they were expressed in terms of protein molecules (see table 2). Calibrant data were processed in the form of ellipticities, as previously.

Table 2. Samples distributed to the study participants. The ACS sample was made up in ultra-pure water and the protein solutions were made in the phosphate buffer.

Sample ID	Description	Concentration
Baselines and calibrants		
Water	Ultra-pure water	—
ACS	Ammonium *d*-10-camphorsulfonate	0.6 mg ml^{-1}
Buffer	Sodium phosphate buffer, pH 7.2	30 mmol dm^{-3}

			Molar concentration	
Sample ID	Description	Mass concentration/ mg ml^{-1}	Molecules/ (μmol dm^{-3})	Peptide bonds/ (mmol dm^{-3})
Proteins				
001	Lysozyme	0.47	31.8	4.07
002	Cytochrome C	0.47	38.1	3.92
003	10% Cytochrome, 90% Lysozyme	0.47	32.4	4.06

2.7. Multivariate analysis

All multivariate analysis was performed in SIMCA P+ 11.5 (Umetrics, Umeå, Sweden) [14]. Three sets of protein spectra were used in the analysis:

(1) The reference data set collected in the organizing lab, which was used to build the models.
(2) The re-measurement data, collected in the organizing lab from participants' returned samples, which were used to detect any changes in the samples.
(3) The participant data.

Far and near UV data were analysed separately. For each data set, the reference data were imported into SIMCA and transposed such that each spectrum was treated as an observation, and each wavelength point was treated as a variable. No mean centring or scaling was applied to the data set. Observations (spectra) were grouped into three classes, corresponding to the sample which was measured. For each spectral data set, two types of analysis were performed:

(1) A PCA of all classes of spectra, to identify clusters of similar spectra.
(2) A SIMCA analysis, where a separate PCA model is built for each class. This allowed new spectra to be classified according to the model.

Therefore, four principal component models (one global and three class-specific) were built for each spectral dataset. Once the models had been built, scores plots were used to visualize the clustering of the data. 'Scree plots' were used to evaluate how many of the principal components were useful for further analysis. Loadings plots were also used to assess how physically meaningful each component of the model was. For further details, please consult the Umetrics documentation and references therein [14].

The re-measurement and participant data were then compared against the models. Predicted scores plots against the global model were used to visualize the clustering of the new data in relation to the reference data. Scores plots for the individual class models were used to assess whether the NPL data formed an outlier when compared with the participant data. The relationship between the new data and the models is summarized by the combined distance-to-model statistic (DModXPS+). These distances were normalized in units of standard deviation, to permit comparison between different models, and were unweighted. Threshold or 'critical' distances (D_{crit}) were calculated for each model using a significance level of 0.05 (or 95% confidence). Contribution plots were used to investigate why some spectra were further from the model than others [14].

3. Results

3.1. Study practicalities

The study participants are listed in table 3. The CD data, experimental questionnaires and temperature loggers were returned by all participants. The samples were also returned successfully, with the exception of participant 2, where not all of the samples were recovered. The data from the temperature loggers showed that the temperature was maintained between 4 °C and 5 °C for about 4 days. This clearly shows an improvement compared with the previous study [13], and ensured that the samples were correctly stored until they reached participants' laboratories. Table 4 shows anonymized information on the participants, including the reported path lengths of their cells and the manufacturer of the instrument used.

3.2. Calibration status of participants

It is apparent from figure 1 that there is significant variation in the participants' calibration state. This variation is apparent in both the peak intensity and wavelength. The plot also shows the peak wavelength and intensity of the same samples when returned to NPL; it is clear that the samples have not changed significantly and that the observed variability is an accurate reflection of the calibration of the participants' instruments. However, the spread of participants' calibration values is much

Table 3. Participants in the study.

Country	Organization	Contact	Status	Role
UK	National Physical Laboratory (NPL)	Alex Knight	NMI	Organizer
	National Institute for Biological Standards and Control (NIBSC)	Chris Jones	Expert lab.	Participant
USA	National Institute for Standards and Technology (NIST)	Curtis Meuse, David Plusquellic	NMI	Organizer and participant
	Olis, Inc.	Stoyan Milev, Julie DeSa Lorenz	Expert lab.	Participant
China	National Institute of Metrology	Liqing Wu	NMI	Participant
France	Institut Pasteur	Bruno Baron	Expert	Participant
	Synchrotron SOLEIL	Frank Wien	Expert	Participant

Table 4. Participant information. Participants' instrument manufacturers and reported cell path lengths. Participant numbers were assigned randomly, and do not correspond to those used in the previous study; participant 0 refers to the organizing lab at NPL. Manufacturer codes do, however, correspond to the previous study. Participants used a variety of methods to measure cell path lengths as indicated by the footnotes.

			Reported cell path length/cm	
Participant number	Instrument number	Manufacturer code	0.01 cm nominal	1 cm nominal
0	1	A	0.0122[c]	0.958[c]
1	1	E	0.01[a]	1[a]
2	1	A	0.01[b]	1.004[c]
3	1	A	0.0093[b]	1[a]
4	1	D	0.01[b]	0.987[c]
6	1	C	0.01[a]	1[a]
7	1	C	0.009 683[b]	0.966[d]
7	2	C	0.009 683[b]	1[a]

[a] Path length not measured.
[b] Measured by interference method.
[c] Measured by chromate method.
[d] Measured by ratio to short path-length cell.

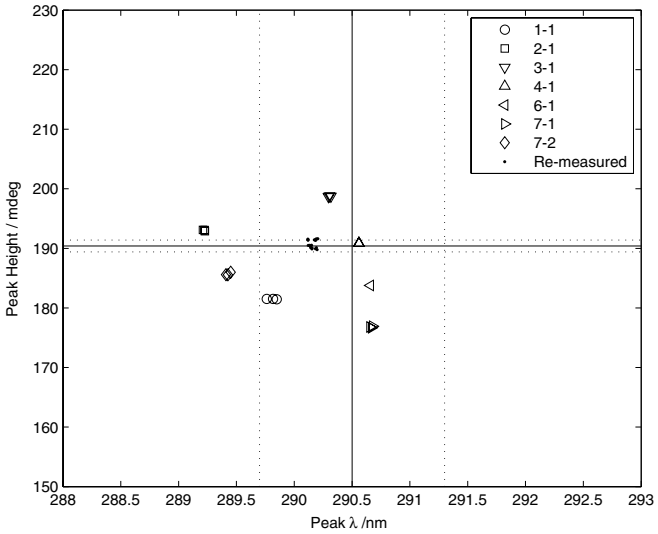

Figure 1. Calibration status of participants' instruments. Summary of ACS peak wavelengths and intensities obtained by curve fitting to the spectra reported by the participants. The codes in the legend indicate the participant and machine numbers. The black dots indicate the same samples re-measured in the organizing lab after their round-trip to the participants' laboratories. The spread of calibration values is between ∼175 mdeg and ∼200 mdeg, a significant improvement on the previous study, although only one participant's calibration is within the arbitrary ±1 mdeg limit.

smaller than seen in the previous study [13]. The reference spectra collected at NPL through the course of the study show no significant changes (data not shown).

3.3. Reference data set

A reference data set for the three protein samples was collected throughout the study, with 15 (far UV) or 16 (near UV) separate sets of measurements being made in triplicate for each sample. The purpose of this reference data set was to check for any instability on the samples throughout the duration of the study, and to act as a 'training set' for the multivariate data analysis. These data were collected under identical conditions to those specified in the study protocol. The spectra are shown in figure 2. As expected, samples 001 and 003 gave similar spectra in the far and near UV regions, although in both cases the spectra can be distinguished by eye in the overlay plots. In each case there is good consistency within each sample, although a few outlier spectra are apparent.

3.4. PCA analysis

A PCA of the reference data set was carried out to examine the structure of the reference data set; the corresponding scores plots are shown in figure 3, and the details of the principal component models are given in table 5. We found that two components were sufficient to describe both the far and near UV datasets (with a cumulative R2X better than 99%). We found that in both the near and far UV the spectra for each sample formed tight, distinct clusters, suggesting that they could readily be distinguished by this approach (figure 3).

3.5. Participant data

Example participant spectra are shown in figure 4. While in the far UV region the spectra are typically quite noisy, they show a broad agreement in terms of absolute intensity. The spectra shown here are for sample 001; the other samples are not shown

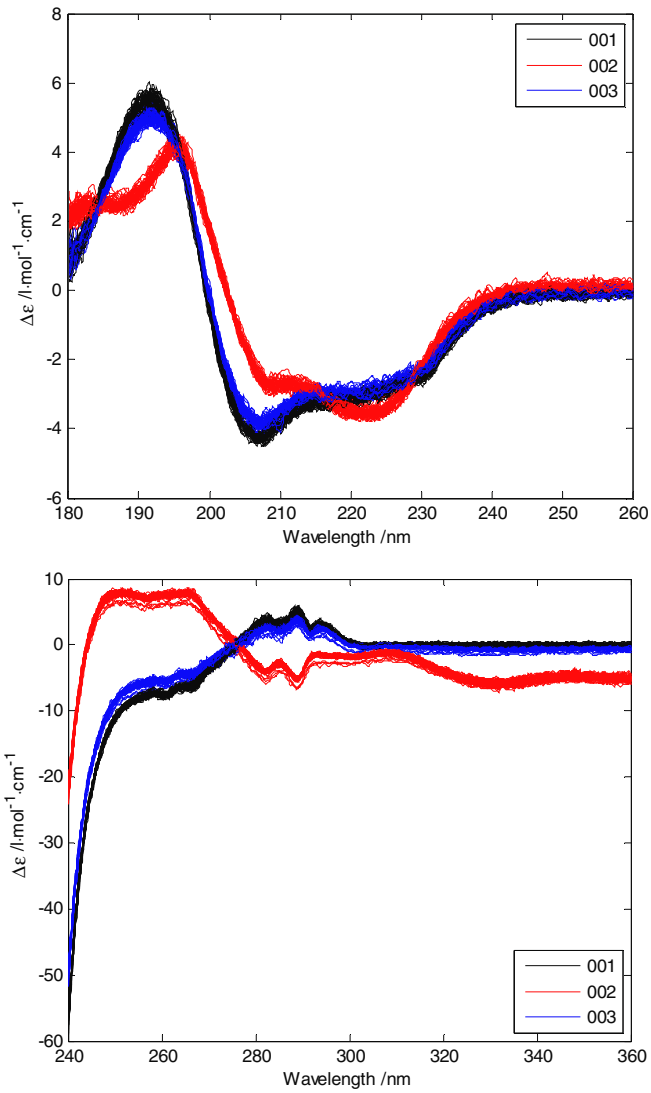

Figure 2. Reference data set: spectra. (*a*) Far UV spectra ($n = 45$). (*b*) Near UV spectra ($n = 48$). All spectra have been converted to molar CD; note that peptide bond concentration was used for the far UV and protein concentration for the near UV spectra (see table 2). Note that samples 001 and 003 can be distinguished visually.

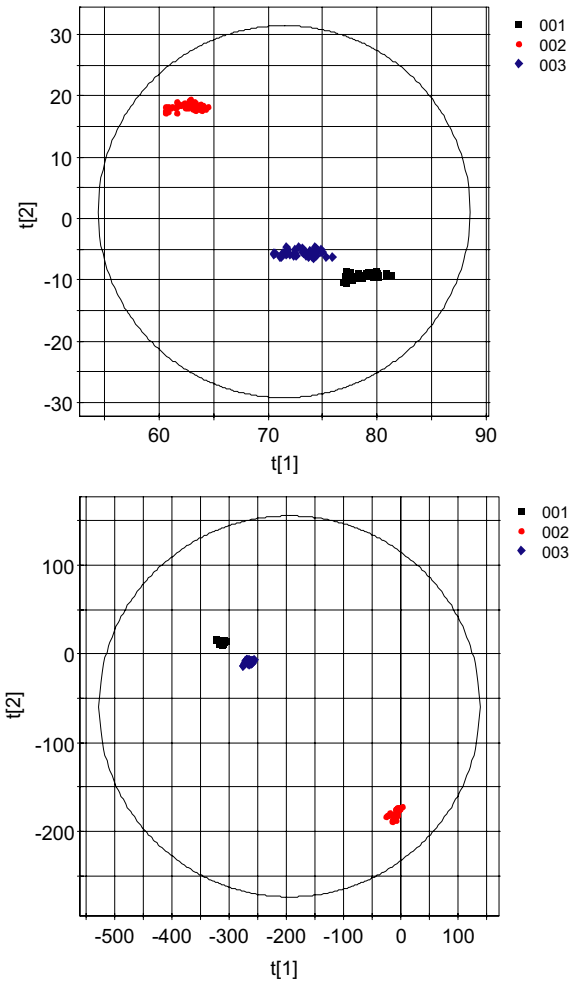

Figure 3. PCA scores plots for reference data. (*a*) Far UV data and (*b*) near UV data both form tight clusters in the reference data set. The values $t[1]$ and $t[2]$ are the PCA scores for the first and second principal components, respectively, and are in arbitrary units.

for reasons of space, but show similar results. The noise level varies notably between instruments, and some scale differences are apparent. Participant 7 reported problems with the nitrogen purging on their instrument 2, which may explain some of the discrepancy between their data and other participants' spectra.

The near UV spectra are less noisy and show much closer agreement, with participant 1 being an outlier. Here sample 002 is shown, which shows clear outlier spectra from participant 1; the agreement seen with the other samples was somewhat better (not shown).

3.6. SIMCA analysis

The models used for the SIMCA analysis are detailed in table 5; one component was sufficient in each case. The 'distance-to-model' statistics for the reference data set are shown in figure 5 (here the distance is plotted between each observation and the model corresponding to that particular sample). As expected, the vast majority of the data points fall within the D_{crit} 95% confidence limit, indicating that the models are a good description of the data. The same models were then used to analyse the participants' spectra and the spectra of the returned samples. In figure 6 the distance-to-model data for the participant spectra are shown. In the far UV region, none of the spectra fell within the 95% confidence limit, indicating that all of the spectra showed differences to the reference data set. However, the range of distances was markedly lower than reported previously. In the near UV, as might be expected, the agreement was much better, with most of the data from participants 3 and 4 falling within the confidence limit and overall the distances being lower than in the previous study.

The results from the analysis of the samples on their return to the organizers are presented in figure 7. Here, the majority of the samples were not significantly different from the reference data set; that is they had not changed throughout shipping or storage in the participating laboratories. There are, however, a few outliers where the samples have clearly undergone some changes; for example sample 002 from participants 2 and 7-1

Table 5. Models used in PCA and SIMCA analysis.

					R2X (component)	
Dataset	Model	Samples	Observations	Variables	1st	2nd
Far UV	M1	001, 002, 003	135	801	0.970	0.0283
	M2	001	45	801	0.998	—
	M3	002	45	801	0.997	—
	M4	003	45	801	0.998	—
Near UV	M1	001, 002, 003	145	1201	0.834	0.164
	M2	001	48	1201	0.999	—
	M3	002	49	1201	0.994	—
	M4	003	48	1201	0.998	—

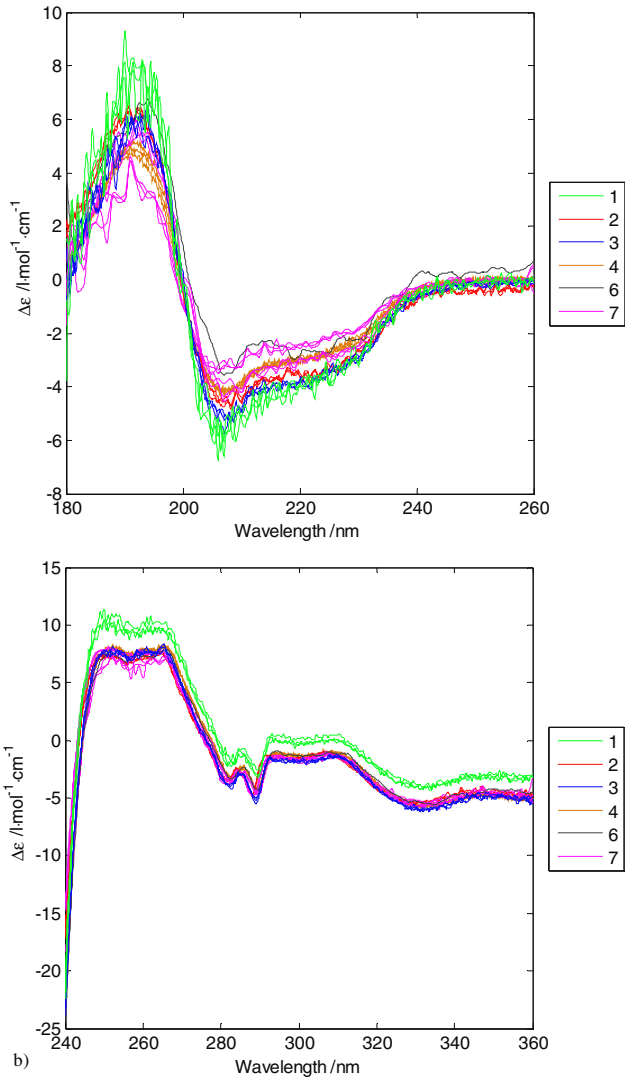

Figure 4. Example participant spectra. (*a*) Participant far UV data for sample 001. (*b*) Participant near UV data for sample 002. All spectra have been converted to molar CD; note that peptide bond concentration was used for the far UV and protein concentration for the near UV spectra (see table 2). Participant 7 used two instruments; the data from both of them are shown.

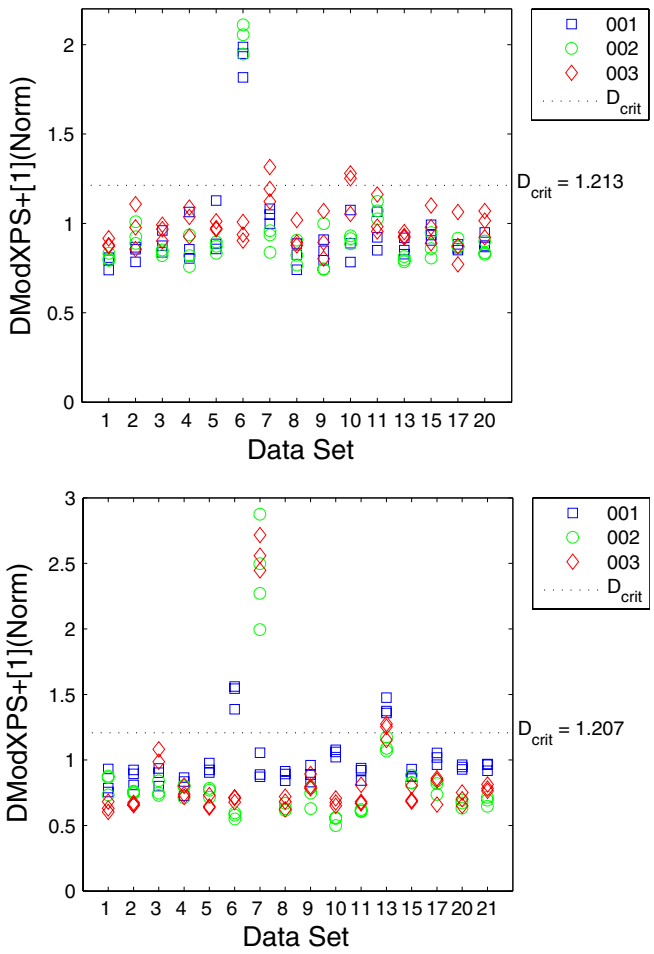

Figure 5. Distance-to-model: plots for reference data. The DModXPS+ statistic is plotted for one principal component. As expected, for each of the three models the majority of the data used to build the model fall within the 95% confidence limit. Data sets 6 in the far UV and 7 in the near UV appear to be outliers. Distances are normalized and shown in units of standard deviation to enable comparisons between samples; data sets are identified by code numbers, and were measured on a single day.

in the far UV and additionally sample 003 for participants 4 and 7. Since these changes are larger than the differences observed between the participant and reference spectra, it may be that these changes occurred on the return leg of the shipping.

The distance-to-model statistics can also be used to classify spectra. The Cooman's plot shown in figure 8 shows how the similar samples 001 and 003 can be distinguished in both near and far UV. The separation is greater in the near UV, but in neither spectral region are there any spectra that could be in either category; a few spectra are outliers that fall

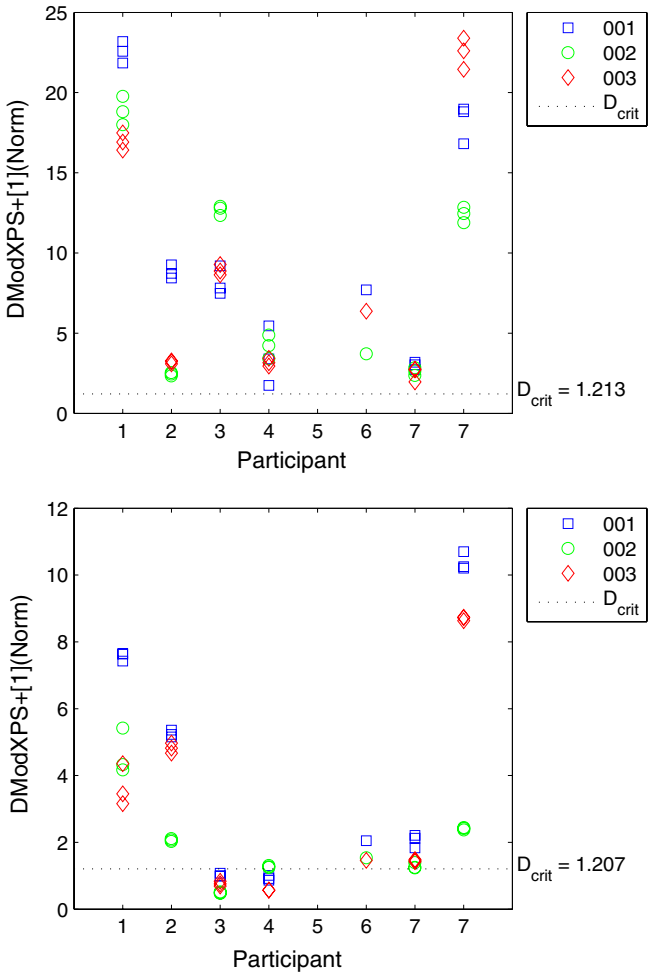

Figure 6. Distance-to-model plots for participant data. (*a*) In the far UV spectra, none of the participants' data were statistically indistinguishable from the reference data. However, the distances were much smaller than seen previously (up to 150 standard deviations in the uncorrected data [13]). (*b*) In the near UV, participants 3 and 4 had data largely indistinguishable from the reference data set. Again, however, the distance-to-model values were smaller than seen in the previous study (up to 50 standard deviations).

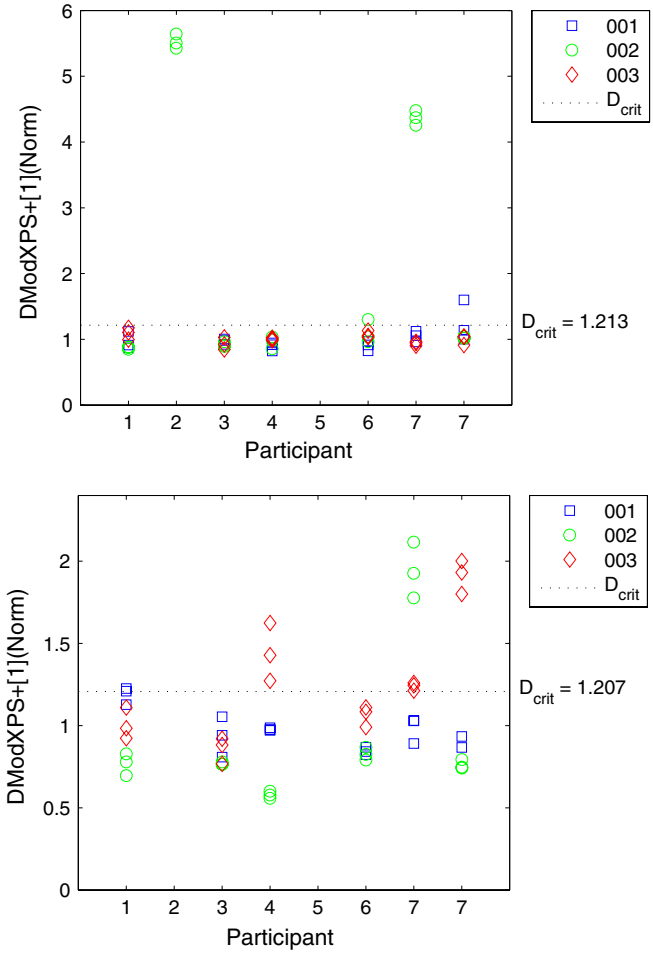

Figure 7. Distance-to-model plots for returned samples. (*a*) No significant changes were observed in the returned samples with the exception of sample 2 in the case of participants 2 and 7 (instrument 1). These changes most likely occurred on the return journey, as the distances are greater than for the corresponding participant spectra. (*b*) In the near UV, changes were also apparent in sample 3 for participants 4 and 7. There was insufficient sample from participant 2 to make the measurement.

outside the confidence limit for both models. When the same analysis is performed for the returned sample data, the number of outliers increases (not shown) as might be expected from the corresponding distance-to-model plots. The much greater variability of the participant data and the fact that many of the points fall outside the confidence limits means that this method is not useful for classifying the participant spectra.

4. Discussion

The primary aim of this study was to show that improved results in CD measurements could be obtained if best practice was followed, and this has clearly been achieved. Examination of the overlaid spectra (figure 4) and the SIMCA distance-to-model plots (figure 6), in comparison with the previous results [13], shows a marked improvement in the agreement between laboratories. Most strikingly, while the far UV spectra show some differences in scale, these are much smaller than were seen previously, when some participants used incorrect path-length cells, which in some cases rendered the data meaningless. Furthermore, on this occasion many of the participants measured the path length of their cells, which allowed this to be taken into account in the conversion to molar CD; the results are therefore in better agreement than was seen even with the path-length corrected data previously. It is noticeable that some of the far UV spectra have much higher noise levels than others. In all probability, this reflects the total amount of light throughput in these instruments. Ageing of the lamp and mirrors in CD instruments can cause this type of degradation in performance.

Interestingly, the variation in amplitude in the far UV is greater than one might expect from the small differences in instrument calibration. This may be due to a variety of factors. Firstly, the instruments are calibrated (unless they use the direct subtraction method) at 290.5 nm, whereas the far UV spectra are measured from 180 nm to 260 nm; therefore any variation in calibration with wavelength can affect the

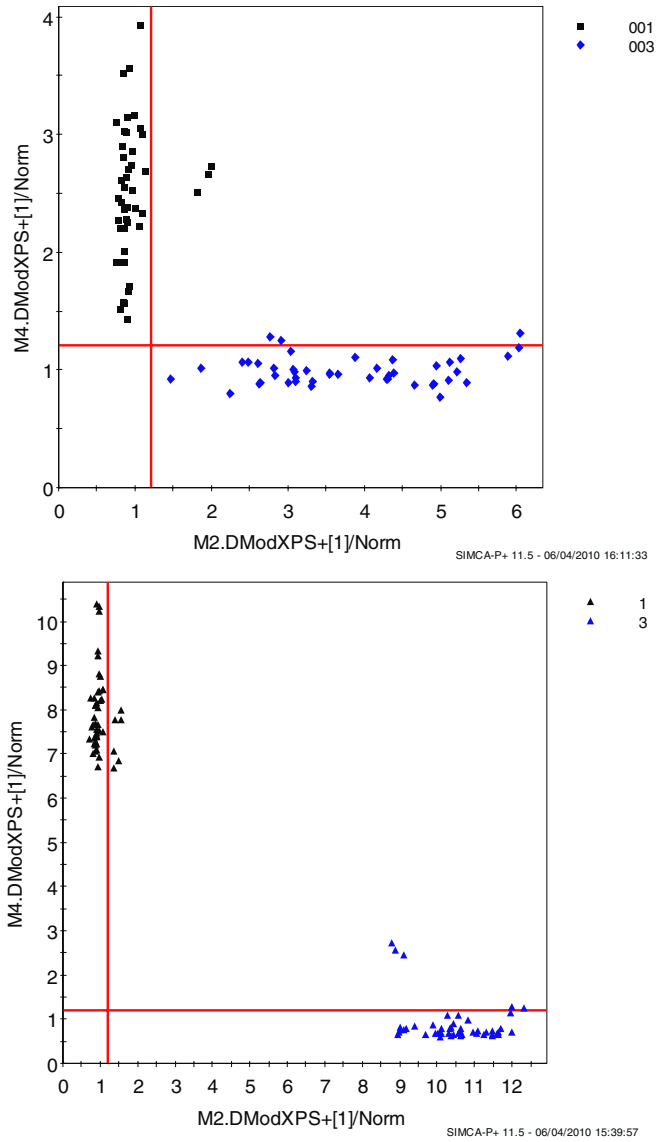

Figure 8. Cooman's plot for samples 001 and 003 (reference data set). Panel (*a*) shows analysis of far UV and (*b*) of near UV data. Spectra from the similar samples 001 and 003 can reliably be discriminated using the 'distance-to-model' statistics. The red lines (vertical and horizontal) show the D_{crit} values for the two models (at a 95% confidence limit). Points in the top left quadrant of the plot are classified as sample 001 and points in the bottom right as sample 003. Points in the bottom left quadrant could belong to either model, whereas points in the upper right quadrant do not fit either model. Data for sample 002 are omitted for clarity; these fall far into the upper right quadrant (beyond the axis limits shown here).

results. Also, some participants (e.g. participant 1) did not measure their cell path length, and this may explain some of the discrepancy.

In the previous study, the results in the near UV region showed better scale agreement than in the far UV, but there were problems with wavelength calibration, smoothing and apparent baseline issues. These problems were largely avoided in this study, with figure 4 showing the worst example, where one participant's results show an offset from the others. It is not unexpected that agreement is better in the near UV, as this region includes the ACS calibration peak wavelength, uses cells with proportionately better manufacturing tolerances, and there is typically more light available in this region.

The reported calibration state of the instruments is in better agreement than previously, with all participants reporting values within ∼±8% of the nominal value. This improvement is probably due in part to the participants being advised to calibrate their instruments before the study, although it should be noted that the direct subtraction instruments do not require calibration in the same way as the modulation instruments. It is also interesting that the participants with the most divergent calibration values did not measure the path length of their 1 cm cells, and that therefore the true spread of calibration states may be smaller than it appears.

There are three principal reasons why the study results show improved performance. Firstly, we provided best practice guidance and an improved protocol to the participants. Secondly, some of the participants will have benefited from their experience in the first study. Finally, there are fewer (and different) participants from the first study. Although the performance has improved markedly, there are still worrying inter-laboratory discrepancies. These could originate from a number of sources, but we feel that the biggest contributor is likely to be that some of the participants have still not measured their cell path lengths. This not only affects their measurements, but in the case of the 1 cm cells, will also affect their calibration. General instrument maintenance and performance also appears to be a factor in some cases. Another concern is variation of calibration with wavelength; instruments from the different manufacturers are likely to behave differently in this respect. A proposed new reference material could address this problem [15].

This study has once again demonstrated the utility of a multivariate statistical analysis approach for the comparison of sets of CD spectra. As a result of a change to the formulation of the 'mixed' or 'spiked' sample 003, it was possible to unambiguously discriminate between spectra from samples 001 and 003 in the reference set (although a few outlier spectra could not be assigned, see figure 8). Fewer and fewer spectra could be classified as one moves from the reference, to the re-measured samples, to the participant data, although spectra were never misclassified. This demonstrates that while this approach can account for random variability between spectra, it cannot account for uncontrolled systematic effects that occur when samples are measured in different laboratories. Therefore, we conclude that a more complete picture of the uncertainties in CD measurements is required to support inter-laboratory comparability of CD data, and we are working to achieve this [16].

Acknowledgments

The authors would like to acknowledge the contribution of the study participants, without whose efforts this study would not have been possible. They would also like to thank the members of the CCQM Bioanalysis Working Group for supporting the study and for their input and suggestions. This work was funded by the Chemical and Biological Metrology Programme of the National Measurement Office.

References

[1] Fasman G D (ed) 1996 *Circular Dichroism and the Conformational Analysis of Biomolecules* (New York: Plenum)

[2] Rodger A and Norden B 1997 *Circular Dichroism and Linear Dichroism* (Oxford: Oxford University Press)

[3] Venyaminov S Y and Yang J T 1996 Determination of protein secondary structure *Circular Dichroism and the Conformational Analysis of Biomolecules* ed G D Fasman (New York: Plenum) pp 69–107

[4] Miles A J, Whitmore L and Wallace B A 2005 Spectral magnitude effects on the analyses of secondary structure from circular dichroism spectroscopic data *Protein Sci.* **14** 368–74

[5] Johnson W C Jr 1996 Circular dichroism instrumentation *Circular Dichroism and the Conformational Analysis of Biomolecules* ed G D Fasman (New York: Plenum) pp 635–52

[6] Takakuwa, T, Konno T and Meguro H 1985 A new standard substance for calibration of circular dichroism: ammonium d-10-camphorsulfonate *Anal. Sci.* **1** 215–8

[7] Miles A J *et al* 2003 Calibration and standardisation of synchrotron radiation circular dichroism and conventional circular dichroism spectrophotometers *Spectroscopy* **17** 653–61

[8] Miles A J, Wien F, Lees J G and Wallace B A 2005 Calibration and standardisation of synchrotron radiation and conventional circular dichroism spectrometers: II. Factors affecting magnitude and wavelength *Spectroscopy* **19** 43–51

[9] Vives O C *et al* 2004 *Val-CiD Appendix B: The Use of Chemical Calibrants in Circular Dichroism Spectrometers* National Physical Laboratory

[10] Kelly S M, Jess T J and Price N C 2005 How to study proteins by circular dichroism *Biochim. Biophys. Acta* **1751** 119–39

[11] Jones C *et al* 2004 *Val-CiD Best Practice Guide: CD Spectroscopy for the Quality Control of Biopharmaceuticals* National Physical Laboratory

[12] Yang J T 1996 Remembrance of things past: a career in chiroptical research *Circular Dichroism and the Conformational Analysis of Biomolecules* ed G D Fasman (New York: Plenum) pp 1–24

[13] Ravi J *et al* 2010 International comparability in spectroscopic measurements of protein structure by circular dichroism: CCQM-P59 *Metrologia* **47** (Tech. Suppl.) 08022

[14] 2005 *User's Guide to SIMCA-P, SIMCA-P+*. Version 11, (Umeå: Umetrics AB)

[15] Damianoglou A *et al* 2008 A new reference material for UV–visible circular dichroism spectroscopy *Chirality* **20** 1029–38

[16] Knight A E *et al* Uncertainty in measurements of protein structure by circular dichroism, in preparation