Synchrotron radiation circular dichroism spectroscopy of proteins: secondary structure, fold recognition and structural genomics

BA Wallace*† and Robert W Janes‡

Recent developments in instrumentation and bioinformatics show that the technique of synchrotron radiation circular dichroism spectroscopy can provide novel information on protein secondary structures and folding motifs, and has the potential to play an important role in structural genomics studies, both as a means of target selection and as a high-throughput, low-sample-requiring screening method. This is possible because of the additional information content in the low-vacuum ultraviolet wavelength data obtainable with intense synchrotron radiation light sources, compared with that present in spectra from conventional lab-based circular dichroism instruments.

Addresses
*School of Crystallography, Birkbeck College, University of London, London WC1E 7HX, UK
†Centre for Protein and Membrane Structure and Dynamics, Daresbury Laboratory, Warrington WA4 4AD, UK
‡School of Biological Sciences, Queen Mary and Westfield College, University of London, London E1 4NS, UK
Correspondence: BA Wallace; e-mail: ubcg25a@mail.cryst.bbk.ac.uk

Current Opinion in Chemical Biology 2001, **5**:567–571

1367-5931/01/$ – see front matter

Abbreviations

CD	circular dichroism
SRCD	synchrotron radiation circular dichroism
VUV	vacuum ultraviolet

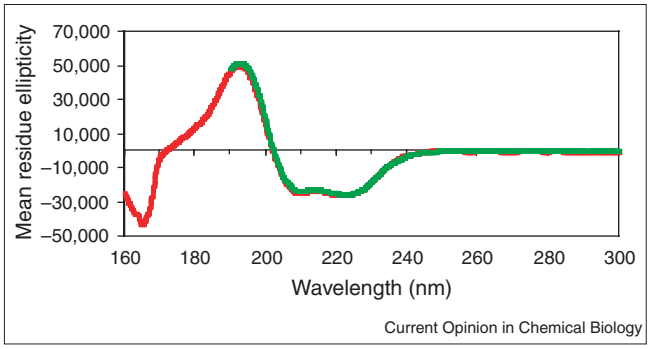
Introduction

Circular dichroism (CD) spectroscopy measures the difference in absorption of left- and right- circularly polarised light as it passes through an optically active or chiral sample. Spectra in the far ultraviolet wavelength range (typically from ~190 to 250 nm) provide information on the polypeptide backbone conformations of proteins. Further information exists in the vacuum ultraviolet (VUV) wavelength region (below 190 nm), but its measurement is generally limited in a conventional, lab-based CD instrument by the high absorption of the sample, buffer and solvent, and the low intensity of the light source. Empirical methods utilising reference databases consisting of the spectra of proteins with known structures allow decomposition of the CD spectrum of an unknown protein, providing information on its secondary structural features.

CD spectrometers using synchrotron radiation as their light sources were first developed in 1980 [1–3], but until recently saw limited use in studies of biological systems [4–8]. In the past two years or so, developments in instrumentation [9•,10] have resulted in synchrotron radiation circular dichroism (SRCD) spectrometers suitable for examining protein conformations and folding under a wide range of conditions, and have, for the first time, enabled protein spectra in aqueous solutions to be obtained to wavelengths as low as 160 nm (Figure 1) [11•]. Currently, SRCD stations suitable for protein studies are available at the Synchrotron Radiation Source (SRS; UK) (http://www.srs.dl.ac.uk/VUV/CD/cpmsd.html), the National Synchrotron Light Source (NSLS; USA) (http://bnlstb.bio.bnl.gov/biodocs/nsls/u9b/u9b.htmlx), and Aarhus Storage Ring in Denmark (ASTRID; Denmark) (http://www.isa.au.dk/SR/sr-facilities.html), and a station at the Beijing Synchrotron Radiation Facility (BSRF; China) (http://www.ihep.ac.cn/bsrf/bio.html) is planned to come on line in the next year. Parallel developments in bioinformatics have enabled better definition and classification of a wide range of protein structural types, and the rapid growth in crystal structure analyses has provided a large number of protein structures from which more comprehensive reference databases can be constructed. Furthermore, it has recently been demonstrated that VUV irradiation in an SRCD instrument is non-damaging to protein integrity [12•], an important result if this technique is to find general utility in studies of biological systems. Taken together, these developments have meant that SRCD has recently emerged as an exciting technique for protein structural studies [13••]. Recent SRCD studies have included examinations of the kinetics of protein folding and unfolding [14•,15], the nature of macromolecular interactions [15,16•], and environmental effects on secondary structures [17•,18].

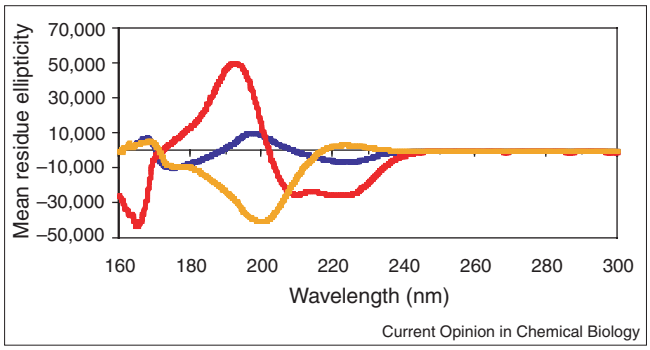
SRCD spectroscopy has a number of advantages over conventional lab-based CD spectroscopy [13••]. Besides the ability to measure lower-wavelength data, these include

Figure 1



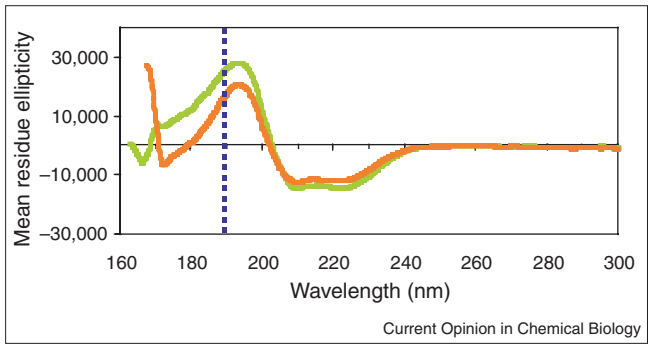
CD (green) and SRCD (red) spectra of horse myoglobin, showing additional transitions in the low-wavelength data present in the SRCD spectra (adapted from [11•]).

Figure 2



SRCD spectra of proteins representing primarily α-helical (myoglobin; red), β sheet (concanavalin A; blue), and polyproline II helical (type VI collagen; yellow) secondary structures, showing substantial differences are present in the low-wavelength region.

Figure 3



SRCD spectra of two proteins with substantially different secondary structures (74% helix, 0% sheet, 10% turn, 16% other [orange] and 48% helix, 5% sheet, 16% turn, 31% other [green]). Over the wavelength range obtained in conventional CD spectroscopy (to the right of the dotted vertical line), the spectra are nearly identical. It is only when the low-wavelength data (left of the vertical line) are considered that the differences in these structures are obvious.

higher signal-to-noise ratios (thus resulting in smaller sample requirements), more rapid measurements (because of the diminished requirement for signal averaging as a result of stronger signals), and the possibility of examining samples in the presence of high concentrations of buffers and other absorbing components. This review focuses on new developments and applications in SRCD spectroscopy over the past two years, which have enabled its use as a technique for examining protein structures, and have shown its potential for use in structural genomics.

SRCD and CD spectral characteristics

CD spectra arise from electronic transitions between ground and excited states of molecular orbitals. The most significant chromophore for proteins is the amide group of the polypeptide backbone. Secondary structures impose positional and intensity constraints on these transitions, and give rise to characteristic far-UV spectra [19] (Figure 2): for an α helix, the negative band at ~222 nm is the $n\rightarrow\pi^*$ transition, and the positive and negative bands at ~190 and ~208 nm, respectively, are the perpendicular and parallel exciton peaks of the $\pi_0\rightarrow\pi^*$ transition. For β sheets, the negative band at ~215 is the $n\rightarrow\pi^*$ transition, and the positive and negative ~198 and ~175 nm bands are the $\pi_0\rightarrow\pi^*$ transitions. For polyproline-II-type structures, the positive band at ~220 nm is the $\pi_0\rightarrow\pi^*$ parallel, and the negative band at ~200 nm is the $\pi_0\rightarrow\pi^*$ perpendicular transition, whereas the $n\rightarrow\pi^*$ transition is not seen. A number of additional transitions are visible in the VUV region detectable by SRCD (Figure 2). For helices, the shoulder seen in the VUV spectrum at ~175 nm had originally been ascribed to an $n\rightarrow\sigma^*$ transition [20], superimposed on a proposed helix transition. However, both the 160 nm and 175 nm peaks are now thought to arise from interamide charge transfer transitions (R Woody, personal communication). Spectral differences in this region may result also, in part, from tertiary structural interactions.

At wavelengths above 190 nm, the spectra of helices and sheets are roughly similar in shape, with the sheet signal dwarfed by the much more intense helix signal (Figure 2); one consequence of this is that analyses of sheets from CD data tend to be much less accurate than those of helices [21]. Below 190 nm, the SRCD spectra of sheets and helices have markedly different peak shapes and opposite signs, so analyses that include these low wavelength data are better able to quantitate sheet content and other types of secondary structures, such as polyproline II helices (Figure 2). A further consequence resulting from these low wavelength VUV SRCD data is the ability to distinguish the major differences between structures of proteins that have similar spectra in the far UV region, but differ considerably in the VUV region (Figure 3).

Information content of SRCD and CD data

Many methods have been developed for calculating protein secondary structure contents from CD spectra [22,23•,24••,25,26•] based on reference databases of spectra produced from proteins with known three-dimensional structures [27•]. One question that has arisen is how many different types of secondary structures can accurately be distinguished from a CD spectrum. The singular value decomposition approach [28,29] makes use of matrix algebra to determine the information content present within a spectrum. It is highly dependent on the wavelength range of the spectrum [30]: using data between 200 nm and 260 nm, the information content is two; if the data extends to 190 nm, the value is between three and four; to 178 nm, five; to 168 nm, six; and by taking spectra to ~160 nm, the information content could rise to seven or possibly even eight. What this means in terms of secondary structural component analyses can be illustrated thus: data measured to 178 nm have an information content of five, hence each spectrum may be represented by five independent equations or eigenvectors. However, because not all secondary structural types are independent variables [31], these can be used to solve for around eight different structural

features (e.g. helices, antiparallel and parallel sheets, four classes of turns, and 'other' — often incorrectly referred to as 'random coil' — structure). By extension, the 160 nm data may represent around eight eigenvectors and thus permit identification of a proportionately greater number of structural types, thus enabling the discrimination of more secondary or supersecondary structural components and possibly even motifs or tertiary structure fold features.

Fold recognition by SRCD

The accuracies of empirical analyses depend on the extent of coverage of all possible structures by the proteins used to create the reference databases [24••,25]. Hence, as our knowledge of protein structures expands, additions of spectra of novel structures to the spectral reference databases further improve the accuracies of the CD-based determinations of secondary structures. Recently, SRCD spectra obtained for a wide range of β-strand-containing motifs (antiparallel and parallel sheets, β helices, TIM barrels, β barrels of different strand and sheer numbers, Greek keys, β sandwiches and β meanders) have shown the spectra of these folds are very different, especially at low wavelengths (J Lees, RW Janes, BA Wallace, unpublished data) and support the idea that the low-wavelength data are relevant to supersecondary or tertiary structures. These spectra are forming the basis for fold recognition procedures using neural networks.

SRCD in structural genomics and proteomics

A principal goal of structural genomics is to identify all unique protein folds that exist, and subsequently to use this information to improve the prediction of protein structures from sequences. Of the ~1000 unique folds that are postulated to exist, ~550 have already been identified. Currently only ~3–5% of new structures solved are novel folds, with the proportion of new versus previously identified folds diminishing with time. It has been estimated that the human genome has around 31,000 open reading frames, many of which are potential candidates for new unique folds. This has led to questions of how proteins with new folds can be efficiently identified. SRCD has a role to play in this: because the reference databases are composed of spectra from proteins of known structures, if an 'unknown' protein spectrum is not well-fit by the reference database, the protein is a potential candidate for a new fold or motif. The more proteins with different motifs used in creation of the reference database, the more likely is its ability to identify a protein as having a unique fold, and with the higher information content in SRCD spectra, the ability to distinguish different folds should be greatly enhanced. Hence, a role for SRCD could be in target selection for structural genomics, as a valuable adjunct to bioinformatics predictions of folds. In particular, it may be an especially useful screen for what is *not* a new fold, greatly improving the efficiency of candidate choices for crystal structure determinations.

One current approach to proteomics studies is to identify and characterise all expressed (or all soluble) proteins that have no known structural or functional homologues [32]. In an elegant study [33••] on the small organism *Mycoplasma genitalium*, bioinformatics techniques were used to select proteins whose functions were unknown. These were subsequently expressed and then CD spectroscopy was used to screen them for secondary structural characteristics and thermodynamic stabilities, thereby identifying stable globular proteins as well as proteins that might require partners for folding and function.

Consequently, structural genomics and proteomics studies could benefit from SRCD for the identification of both well-folded structures, and conditions for stabilisation of conformers by ligands, binding partners etc., in order to gain understanding of the nature of interactions with partner molecules, and as a means of increasing the success rate of crystal formation [34]. Because of the high intensity of the synchrotron radiation light source, highly absorbing samples can be examined (i.e. proteins at high concentrations in the presence of buffers, high salt concentrations and additives, conditions such as those typically used for crystallisations). Because SRCD uses only 10^{-2} to 10^{-3} the amount of material, and can take 10^{-3} to 10^{-4} or less of the time required for crystallographic studies, it can provide an efficient high-throughput screening method for proteomics and functional or structural genomics studies.

Conclusions

SRCD was first developed as a method more than 20 years ago, but only recently has its utility in biology begun to be effectively exploited. This follows recent developments that have included demonstrations of the additional low-wavelength data obtainable [11•], the additional information content present in that data [35], and that the technique is non-damaging so protein integrity is maintained [12•]. In this article, we have discussed potential new uses in secondary structural analyses, fold recognition, and structural genomics, as well as recent uses in protein folding studies, in monitoring conformational transitions, and as an adjunct to crystallography.

Two obvious areas for applications of SRCD in the future will be investigations of membrane proteins and in the pharmaceutical industry. Integral membrane proteins constitute an important class of proteins that is greatly under-represented in the databases of crystal structures. It has been estimated that about one-third of the proteins in the human genome may be membrane-associated, and at present ~60% of all drug targets are against membrane proteins. To date, CD studies of membrane proteins have been limited in the accuracy of their secondary structure determinations because none of the reference databases currently available include any membrane proteins (A Orry, A Lobley, RW Janes, J Lees, BA Wallace, unpublished data), even though it is clear that such proteins have different spectral characteristics from their soluble counterparts [36,37]. A new SRCD membrane protein database is currently under

construction that will contain spectra of membrane proteins whose three-dimensional structures have been determined (J Lees, A Orry, RW Janes, BA Wallace, unpublished data). This should improve the utility of CD spectroscopic studies in the secondary structure determination of membrane proteins, a class of macromolecules generally being by-passed at present by structural genomics [38]. Furthermore, CD has demonstrable utility in the pharmaceutical industry for drug-binding studies [39]. With its increased sensitivity and the potential for automation of data collection, SRCD could provide a regular assay for monitoring drug–target interactions. Additionally, SRCD is currently being used as a method for examining integrity of protein folding, and may assume an even greater role as more proteins are expressed in high yields as inclusion bodies, requiring refolding to regain their native structures. Indeed, SRCD could find a use as a standard reference method for assays of fidelity of folded products.

In summary, SRCD extends the measurements possible with conventional CD instruments, providing a much higher information content and the ability to examine biological samples under a richer variety of conditions.

Acknowledgements

We thank Jonathan Lees and Andrew Orry for help with data collection, the staff of station 3.1 at Daresbury (especially David Clarke) for their advice and aid, and Gareth Jones and Robert Woody for helpful discussions. We thank Anthony Willis of Oxford University for the quantitative amino acid analyses to determine protein concentrations. This work was supported by grants from the BBSRC, including a Structural Biology Centre grant to the CPMSD (Gareth Jones, PI), and a project grant and beamtime awards (to BAW).

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Sutherland JC, Desmond EJ, Takacs PZ: **Versatile spectrometer for experiments using synchrotron radiation at wavelengths greater than 100 nm.** *Nucl Instrum Methods* 1980, **172**:195-199.

2. Snyder PA, Rowe EM: **The first use of synchrotron radiation for vacuum ultraviolet circular dichroism measurements.** *Nucl Instrum Methods* 1980, **172**:345-349.

3. Sutherland JC: **Circular dichroism using synchrotron radiation.** In *Circular Dichroism and the Conformational Analysis of Biomolecules*. Edited by Fasman GD. New York: Plenum Press; 1996:599-633.

4. Sutherland JC, Griffin KP: **Vacuum ultraviolet circular dichroism of Poly (dI-dC)·Poly(dI-dC): no evidence for a left-handed double helix.** *Biopolymers* 1983, **22**:1445-1448.

5. Sutherland JC, Lin B, Mugavero J, Trunk J, Tomasz M, Santella R, Marky L, Breslauer KJ: **Vacuum ultraviolet circular dichroism of double stranded nucleic acids.** *Photochem Photobiol* 1986, **44**:295-301.

6. Garone L, Albaugh S, Steiner RF: **The secondary structure of turkey gizzard myosin light chain kinase and the nature of its interaction with calmodulin.** *Biopolymers* 1990, **30**:1139-1149.

7. France LL, Kieleczawa J, Dunn JJ, Hind G, Sutherland JC: **Structural analysis of an outer surface protein from the lyme disease spirochete, *Borrelia burgdorferi*, using circular dichroism and fluorescence spectroscopy.** *Biochim Biophys Acta* 1992, **1120**:59-68.

8. Sutherland JC, Emrick A, France LL, Monteleone DC, Trunk J: **Circular dichroism user facility at the National Synchrotron Light Source — estimation of protein secondary structure.** *Biotechniques* 1992, **13**:588-590.

9. • Clarke DT, Bowler MA, Fell BD, Flaherty JV, Grant AF, Jones GR, Martin-Fernandez ML, Shaw DA, Todd B, Wallace BA, Towns-Andrews E: **A high aperature beamline for vacuum ultraviolet circular dichroism on the SRS.** *Synchrotron Radiat News* 2000, **13**:21-27.
This paper describes new technical advances that enable the collection of low wavelength data on a synchrotron-based instrument.

10. Ojima N, Sakai K, Fukazawa T, Gekko K: **Vacuum-ultraviolet circular dichroism spectrophotometer using synchrotron radiation: optical system and off-line performance.** *Chem Lett* 2000, **7**:832-833.

11. • Wallace BA: **Conformational changes by synchrotron radiation circular dichroism spectroscopy.** *Nat Struct Biol* 2000, **7**:708-709.
This paper contains the first example of an SRCD spectrum that includes data down to 160 nm for a protein in aqueous solution, showing the additional transitions present at the low wavelengths.

12. • Orry AJW, Janes RW, Sarra R, Hanlon M, Wallace BA: **Synchrotron radiation circular dichroism spectroscopy: vacuum ultraviolet irradiation does not damage protein integrity.** *J Synchrotron Radiat* 2001, **8**:1027-1029.
This is an enabling paper for the technique of SRCD because it demonstrates that the method does not cause damage to proteins.

13. •• Wallace BA: **Synchrotron radiation circular-dichroism spectroscopy as a tool for investigating protein structures.** *J Synchrotron Radiat* 2000, **7**:289-295.
This is an interesting paper that reviews the current state of the method of SRCD and discusses the advantages that arise because of the ability to collect low-wavelength data. It compares CD and SRCD instrumentation and facilities and discusses the utility of the technique as an adjunct to crystallography and in binding and dynamics studies, as well as its potential for use in studies of membrane proteins and in structural genomics.

14. • Clarke DT, Doig AJ, Stapeley BJ, Jones, GR: **The alpha-helix folds on a millisecond time scale.** *Proc Natl Acad Sci USA* 1999, **96**:7232-7237.
This paper describes a study examining the kinetics of helical nucleation and folding for a model polypeptide using stopped-flow SRCD spectroscopy.

15. Qi XL, Holt C, McNulty D, Clarke DT, Brownlow S, Jones GR: **Effect of temperature on the secondary structure of beta-lactoglobulin at pH 6.7, as determined by CD and IR spectroscopy: a test of the molten globule hypothesis.** *Biochemical J* 1997, **324**:341-346.

16. • Chang JF, Phillips K, Lundback T, Gstaiger M, Ladbury JE, Luisi B: **Oct-1 POU and octamer DNA co-operate to recognise the Bob-1 transcription co-activator via induced folding.** *J Mol Biol* 1999, **288**:941-952.
This paper includes an SRCD study examining DNA–protein interactions, and shows the utility of the low-wavelength data in examining conformational changes in proteins that are largely disordered or contain polyproline II helices.

17. • Clarke DT, Jones GR: **Extended circular dichroism measurements using synchrotron radiation show that the assembly of clatherin coats requires no change in secondary structure.** *Biochemistry* 1999, **38**:10457-10462.
This paper demonstrates the use of additional data in the low-wavelength range (in this case down to 175 nm) for defining protein secondary structures, especially in highly absorbing or scattering samples.

18. Symmons MF, Buchanan SGS, Clarke DT, Jones G, Gay NJ: **X-ray diffraction and far-UV CD studies of filaments formed by a leucine-rich repeat peptide: structural similarity to the amyloid fibrils of prions and Alzheimer's disease beta-protein.** *FEBS Lett* 1997, **412**:397-403.

19. Woody RW: **Theory of circular dichroism of proteins.** In *Circular Dichroism and the Conformational Analysis of Biomolecules*. Edited by Fasman GD. New York: Plenum Press; 1996:25-67.

20. Johnson WC Jr, Tinoco I Jr: **Circular dichroism of polypeptide solutions in the vacuum ultraviolet.** *J Am Chem Soc* 1972, **94**:4389-4390.

21. Keiderling TA: **Vibrational circular dichroism: applications to conformational analysis of biomolecules.** In *Circular Dichroism and the Conformational Analysis of Biomolecules*. Edited by Fasman GD. New York: Plenum Press; 1996:555-598.

22. Sreerama N, Venyaminov SY, Woody RW: **Estimation of the number of α-helical and β-strand segments in proteins using circular dichroism spectroscopy.** *Protein Sci* 1999, **8**:370-380.

23. • Johnson WC Jr: **Analyzing protein circular dichroism spectra for accurate secondary structures.** *Proteins* 1999, **35**:307-312.
This paper introduces a new algorithm and basis set derived from high-resolution X-ray structures for the analysis of secondary structural content from CD spectra.

24. Sreerama N, Woody RW: **Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set.** *Anal Biochem* 2000, **287**:252-260.
•• In this paper, the authors compare several algorithms for secondary structure determination from CD data and show the significant improvements that arise both when a more comprehensive database of reference proteins is used, and when lower wavelength data are included in the analyses.

25. Sreerama N, Venyaminov SY, Woody RW: **Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis.** *Anal Biochem* 2000, **287**:243-251.

26. Lobley A, Wallace BA: **DICHROWEB: a website for the analysis of protein secondary structure from circular dichroism spectra.** *Biophys J* 2001, **80**:373a.
• This abstract describes and points to a new website (http://www.cryst.bbk.ac.uk/cdweb) available for the analysis of protein secondary structures from CD and SRCD data.

27. King SM, Johnson WC Jr: **Assigning secondary structure from protein coordinate data.** *Proteins* 1999, **35**:313-320.
• This paper discusses the use of the program Xtlsstr for extracting secondary structural contents from crystal structures, using definitions of secondary structural types that are most appropriate for CD spectral characteristics.

28. Hennessey JP Jr, Johnson WC Jr: **Information content in the circular dichroism of proteins.** *Biochemistry* 1981, **20**:1085-1094.

29. Compton LA, Johnson WC Jr: **Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication.** *Anal Biochem* 1986, **155**:155-167.

30. Toumadje A, Alcorn SW, Johnson WC Jr: **Extending CD spectra of proteins to 168 nm improves the analysis for secondary structure.** *Anal Biochem* 1992, **200**:321-331.

31. Pancoska P, Blazek M, Keiderling TA: **Relationships between secondary structure fractions for globular proteins. Neural network analyses of crystallographic data sets.** *Biochemistry* 1992, **31**:10250-10257.

32. Zarembinski TI, Hung LW, Mueller-Dieckmann HJ, Kim KK, Yokota H, Kim R, Kim SH: **Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics.** *Proc Natl Acad Sci USA* 1998, **95**:15189-15193.

33. Balasubramanian S, Schneider T, Gerstein M, Regan L: **Proteomics of *Mycoplasma genitalium*: identification and characterization of unannotated and atypical proteins in a small model genome.** *Nucleic Acids Res* 2000, **28**:3075-3082.
•• This is a groundbreaking paper that demonstrates the use of CD in a proteomics/structural genomics programme both to characterise new proteins for which bioinformatics studies cannot readily identify fold types and to identify which proteins may require partners for folding.

34. Wallace BA: **Circular dichroism spectroscopy and X-ray crystallography: a dyanamic duo.** *CCP4 Newletter* 1999, **37**:29-30.

35. Wallace BA, Janes RW, Orry A: **Secondary structure information content in the VUV region of synchrotron radiation circular dichroism (SRCD) spectra.** *Biophys J* 2001, **80**:28a.

36. Cascio M, Wallace BA: **Red- and blue-shifting in the circular dichroism spectra of polypeptides due to dipole effects.** *Protein Peptide Lett* 1994, **1**:136-140.

37. Cascio M, Wallace BA: **Effects of local environment on the circular dichroism spectra of polypeptides.** *Anal Biochem* 1995, **227**:90-100.

38. Blundell TL, Mizuguchi K: **Structural genomics: an overview.** *Prog Biophys Mol Biology* 2000, **73**:289-295.

39. Rodi DJ, Janes RW, Sangasnee HJ, Soares A, Holton RA, Wallace BA, Makowski L: **Screening of a library of phage-displayed peptides identifies human Bcl-2 as a Taxol-binding protein.** *J Mol Biol* 1999, **285**:197-203.