Using circular dichroism spectra to estimate protein secondary structure

Norma J Greenfield

Department of Neuroscience and Cell Biology, Robert Wood Johnson Medical School, 675 Hoes Lane West, Piscataway, New Jersey 08854-8021, USA. Correspondence should be addressed to N.J.G. (greenfie@umdnj.edu).

Published online 25 January 2007; doi:10.1038/nprot.2006.202

Circular dichroism (CD) is an excellent tool for rapid determination of the secondary structure and folding properties of proteins that have been obtained using recombinant techniques or purified from tissues. The most widely used applications of protein CD are to determine whether an expressed, purified protein is folded, or if a mutation affects its conformation or stability. In addition, it can be used to study protein interactions. This protocol details the basic steps of obtaining and interpreting CD data, and methods for analyzing spectra to estimate the secondary structural composition of proteins. CD has the advantage that measurements may be made on multiple samples containing ≤20 μg of proteins in physiological buffers in a few hours. However, it does not give the residue-specific information that can be obtained by x-ray crystallography or NMR.

INTRODUCTION

The rapid characterization of new proteins is of great importance for the fields of proteomics and structural genomics. Circular dichroism (CD) is an excellent method for rapidly evaluating the secondary structure, folding and binding properties of proteins. Briefly, CD is defined as the unequal absorption of left-handed and right-handed circularly polarized light. A beam of light has time-dependent electric and magnetic fields associated with it. If the light is polarized by passing through suitable prisms or filters, its electric field, E, will oscillate sinusoidally in a single plane. When viewed from the front, the sinusoidal wave can be visualized as the resultant of two vectors of equal length, which trace out circles, one that rotates clockwise (E_R) and the other that rotates counterclockwise (E_L). The two circularly polarized waves have physical existence. The waves are 90 degrees out of phase with each other and can be separated using a variety of prisms or electronic devices that utilize Pockel's effect[1]. When asymmetric molecules interact with light, they may absorb right- and left-handed circularly polarized light to different extents (hence the term circular dichroism) and also have different indices of refraction for the two waves. The result is that the plane of the light wave is rotated and that the addition of the E_R and E_L vectors results in a vector that traces out an ellipse and the light is said to be elliptically polarized. CD is reported either in units of ΔE, the difference in absorbance of E_R and E_L by an asymmetric molecule, or in degrees ellipticity, which is defined as the angle whose tangent is the ratio of the minor to the major axis of the ellipse. [θ], the molar ellipticity in deg cm^2 dmol^{-1} = 3,298ΔE. For illustrations of the phenomena of CD see Beychok (Ref. 2). There is also a website that illustrates the production of circularly polarized light and the CD of optically active molecules with animated graphics (http://www.enzim.hu/~szia/cddemo/).

CD is an excellent method of determining the secondary structure of proteins. When the chromophores of the amides of the polypeptide backbone of proteins are aligned in arrays, their optical transitions are shifted or split into multiple transitions as a result of 'exciton' interactions (see Ref. 3 for a recent review). The result is that different structural elements have characteristic CD spectra (**Fig. 1a**). For example, α-helical proteins have negative bands at 222 nm and 208 nm and a positive band at 193 nm (Ref. 4). Proteins with well-defined antiparallel β-pleated sheets (β-helices) have negative bands at 218 nm and positive bands at 195 nm (Ref. 5), whereas disordered proteins have very low ellipticity above 210 nm and negative bands near 195 nm (Ref. 6). The collagens are a unique class of proteins, which have three chains that wrap together in a triple helix. Each strand has a conformation resembling that of poly-L-proline[7] in an extended helical conformation in which all of the bonds are *trans* to each other (poly-L-proline II). Charged polypeptides, such as poly-L-glutamate or poly-L-lysine at neutral pH (originally thought to be in random coil conformation), have a similar extended poly-L-proline II-like conformation[8–10]. The spectra of some representative proteins, with widely varying conformations, are shown in **Figure 1b**. Because the spectra of proteins are so dependent on their conformation, CD can be used

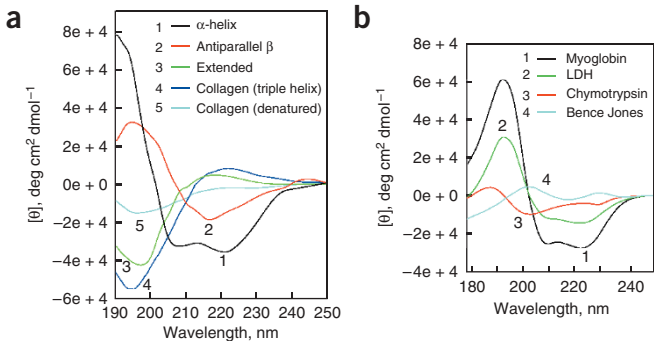

Figure 1 | CD spectra of polypeptides and proteins with representative secondary structures. (**a**) CD spectra of poly-L-lysine at pH 11.1 in the (1, black) α-helical and (2, red) antiparallel β-sheet conformations and at pH 5.7 in the (3, green) extended conformations[5] and placental collagen in its (4, blue) native triple-helical and (5, cyan) denatured forms[64]. (**b**) CD spectra of representative proteins with varying conformations: 1 (black), sperm whale myoglobin; 2 (green), chicken heart lactate dehydrogenase; 3 (red), bovine α-chymotrypsin and 4 (cyan), human Bence Jones protein REI light chain, which is a human immunoglobulin light chain of κ type. Spectra are from data sets supplied by Dr. W.C. Johnson (Oregon State University, Corvallis, Oregon, USA).

to estimate the structure of unknown proteins and monitor conformational changes due to temperature, mutations, heat, denaturants or binding interactions. Although CD does not give the secondary structure of specific residues, as do x-ray crystallographic and NMR structural determinations, the method has the advantage that data can be collected and analyzed in a few hours on solutions of samples containing $\leq$20 μg of protein in aqueous buffers under physiological conditions. Secondary structure can also be estimated from Fourier transform infrared spectroscopy of proteins[11–13] and Raman spectroscopy[14], but in some cases the measurement must be made on protein films or in deuterium oxide.

In addition to the intrinsic CD of the protein backbone, when ligands with chromophores bind to proteins they may develop strong extrinsic CD bands that can also be used to follow binding. The aromatic chromophores of proteins, which have bands in the near-UV region, are often in asymmetric environments and can be used to examine whether mutations change the tertiary structure of proteins.

Key considerations in designing and implementing CD experiments are discussed here.

Choosing CD cuvettes

CD spectra are collected in high-transparency quartz cuvettes (cells). Both rectangular and cylindrical cells are available, with pathlengths ranging from 0.01–1 cm. Water-jacketed cylindrical cells are available for CD machines that do not have temperature-regulated cell holders. Most cylindrical cells and 0.1-cm rectangular cells have relatively low birefringence and give reasonably straight baselines, but all cells are different and baseline spectra must always be collected. Rectangular cells with pathlengths $>$0.2 cm may have high birefringence due to strain. Cells designed for fluorescence, in which all four sides are made of the same material, usually have lower birefringence than cells where two sides are frosted. Rectangular cells with pathlengths $<$0.1 cm often have a very small total sample volume and a very small surface area facing the light beam of the CD machine. It is important that the light beam be very tightly focused if these cells are used, because large artifacts are produced if the light does not go directly through the sample. Dismountable short-pathlength cells are available, but the pathlength can be affected by the viscosity of the sample, and they tend to leak if temperatures are changed. Their use is not recommended. If they are necessary to collect data at low wavelengths, the scale should be checked by comparing data at higher wavelengths with that collected under the same conditions in cells with 0.1-cm or higher pathlengths.

Preparation of buffers

Buffers for CD spectroscopy must not contain any materials that are optically active and should be as transparent as possible. The total absorbance of the sample, including the buffer and cell, should be below one for high-quality data. Samples in which the protein is dissolved in water alone have the highest transparency, but some proteins denature in the absence of salt. The low wavelength cutoffs of several buffers commonly used for CD measurements (buffer + 0.1 mg ml^{-1} protein in 0.1-cm cuvettes) are given in **Table 1**. Oxygen absorbs light below 200 nm; for optimum transparency, buffers should therefore be prepared with glass-distilled water or the water should be degassed before use. Samples for CD analysis should be free of particulate matter. They can be filtered through filters (e.g., 0.1–0.2 μm) such as those sold by Millipore (http://www.millipore.com) or centrifuged at 100,000*g*.

Preparation of proteins and peptides

Samples for CD spectroscopy must be at least 95% pure by the criteria of high-performance liquid chromatography (HPLC), mass spectroscopy or gel electrophoresis. For secondary structure measurements, sample concentrations may range from 0.005–5 mg ml^{-1} depending on the pathlength of the cell. The most difficult part of obtaining high-quality CD data is the correct determination of protein concentration. The method of Lowry[15] and the Bradford dye binding methods[16] give different results for different proteins and must not be used for determining protein concentrations for CD.

The most accurate method of determining protein concentration is quantitative amino acid analysis, using the concentrations of the stable amino acids (e.g., alanine and lysine) to calculate the concentration of the intact protein. This method is very sensitive and can be performed on aliquots of the actual CD samples; however, many laboratories do not have the equipment to do their own measurements. Concentration can be determined using published molar extinction coefficients if they are available; the protein should be dialyzed or desalted into the CD buffer immediately before the spectrum is obtained and filtered though 0.1–0.2 μm filters to reduce light scattering. The spectrum of the CD buffer alone must be subtracted from the spectrum of the sample.

TABLE 1 | Properties of buffers.

Buffer	~Lower wavelength limit, nm*
10 mM potassium phosphate, 100 mM potassium fluoride	185
10 mM potassium phosphate, 100 mM $(NH4)_2SO_4$	185
10 mM potassium phosphate, 50 mM Na_2SO_4	185
10 mM potassium phosphate, 100 mM KCl	195
20 mM sodium phosphate, 100 mM NaCl	195
Dulbecco's phosphate buffered saline (PBS): 9.33 mM potassium phosphate, 136 mM NaCl, 2.7 mM KCl, 0.6 mM $MgCl_2$, 0.9 mM $CaCl_2$	200
2 mM Hepes, 50 mM NaCl, 2 mM EDTA, 1 mM dithiothreitol	200
50 mM Tris, 150 mM NaCl, 1 mM dithiothreitol, 0.1 mM EDTA	201

*The lower limit values are typical for solutions containing ~ 0.1 mg ml^{-1} protein in 0.1-cm cells. Below the lower wavelength cutoffs the dynode voltages rapidly increase, the signal to noise is poor and the ellipticity is not a linear function of the path length of the cell. DMSO and formamides have high absorbance and cannot be used for CD measurements. Many organic solvents, e.g. trifluoroethanol, hexafluoropropanol and hexane are transparent to 185 nm and below but will change the structure of proteins and polypeptides. Buffers can contain up to 20% glycerol, but measurements can only be made to 200 nm at this concentration.

Three different methods of rapid and accurate protein quantification, which are independent of protein composition, are described in this protocol (Step 1, Options A–C). Because CD is usually carried out on samples with relatively low concentrations, it is best to prepare a filtered stock solution of the protein of ~1 mg ml^{-1} in the CD buffer, determine its protein concentration and dilute the stock for the CD measurements.

None of these rapid methods will work if the protein is highly stable and the protein is not fully unfolded in 6 M guanidine HCl or in 3% NaOH (wt/vol). In these cases the protein can be hydrolyzed and the amino acid composition quantified using ninhydrin[17]; the protein may be subjected to Kjeldhal digestion followed by determination of ammonium sulfate using Nessler's reagent[18]; or the protein may be ashed with 72% $HClO_4$ in water at 210 °C and the nitrogen determined with Berthelot phenol–hypochorite reagent as described by Jaenicke[19]. The phenol–hypochorite method has high sensitivity, needing only 0.05–10 μg of protein in the sample, and has an error within of 0.01 μg, but it and the micro-Kjeldhal method cannot be used for samples dissolved in buffers containing nitrogen, such as Tris-HCl. A protocol for the Kjeldhal method is provided in **Supplementary Methods**.

Some proteins, such as collagen and collagen fragments, fold very slowly. If the folding properties are not known the proteins may be pre-folded for several days on ice in the refrigerator, or for several hours at 25 °C before performing CD experiments.

Data analysis

There are many different methods to analyze CD spectra to estimate secondary structure; the most commonly used are described here. Links to sites where the software can be obtained and descriptions of the sites are given in **Table 2**. In addition, some of the software is available in the **Supplementary Software**.

All methods of analyzing CD spectra assume that the spectrum of a protein can be represented by a linear combination of the spectra of its secondary structural elements, plus a noise term, which includes the contribution of aromatic chromophores and prosthetic groups.

$$\theta_\lambda = \Sigma\varepsilon_i S_{\lambda i} + \text{noise} \quad (1)$$

where θ_λ is the CD of the protein as a function of wavelength; ε_i is the fraction of each secondary structure, i; and $S_{\lambda i}$ is the ellipticity at each wavelength of each i^{th} secondary structural element. In constrained fits the sum of all the fractional weights, ε_i, must equal one, and all of the fractional contributions must be greater than or equal to zero.

There are two general classes of methods to evaluate protein conformation. The first uses standards of polypeptides, with defined compositions in known conformations, which have been determined by x-ray scattering of films or by IR in solution[5,20,21]. The second uses the spectra of proteins that have been characterized by x-ray crystallography as standards. These are then compared to the spectra of unknown proteins using least-squares analysis[22–26]; ridge regression[27]; singular value decomposition (SVD)[17]; SVD with variable selection[28–31]; the self-consistent method[32–36]; or neural network analysis[37–39].

All of the listed methods described here are useful for determining the α-helical content of globular proteins. The protein-based analyses are superior when analyzing the conformation of globular, well-folded proteins (see later). However, most programs using data based on protein spectra do not correctly analyze the conformation of proteins with a majority of pure β-helices, such as found in some synthetic polypeptides and in amyloids, or many fibrous proteins such as collagen or coiled-coil proteins. For analysis of the spectra of these samples, more accurate results will be obtained using constrained least-squares fitting analysis programs (e.g. LINCOMB (Ref. 40)) with polypeptide-based reference sets[5,20,21], or the programs K2D (Ref. 38) and CONTIN (Ref. 27). **Table 3** compares the results obtained using various methods for the conformational analyses of the proteins shown in **Figure 1b,** and for the analysis of the conformation of two polypeptides, one with a pure α-helical conformation and one with a pure β-pleated sheet conformation. The effects of wavelength range on the goodness of fits for the programs listed in **Table 3** have been compared elsewhere[26]. The effect of the nature of the protein database on the goodness of the estimates for SELCON3, CONTIN and CDSSTR have been described in detail[36].

Linear regression. Linear regression fits the spectrum of an unknown protein by comparison to the spectra of a set of fixed standards[5]. It is useful for evaluating the effects of mutations, ligands and solvents on protein conformation. Suitable standards include polypeptides with known conformations[5,20,21], which are essential for the analysis of nonglobular polypeptides and fibrous proteins; standard spectra extracted from a database of known proteins using the method of least squares[5,22,24,25,41,42], which are useful for analyzing globular proteins; and standard spectra extracted from a database of protein with known conformations using the convex constraint algorithm (CCA)[40]. Either constrained least-squares or nonconstrained multilinear regression (MLR) can be used.

In constrained least-squares fits, the sum of the contribution of each spectrum is constrained to equal one: $\theta_\lambda = \Sigma\varepsilon_i S_{\lambda i}$, $\Sigma\varepsilon_i = 1$ and $\varepsilon_i \geq 1$. Least-squares analysis can be performed using the program LINCOMB (Ref. 40), which is available in the **Supplementary Software**. It can also be done by loading a database of reference spectra into a commercial graphics program such as SigmaPlot (Systat Software Inc.; http://www.systat.com/products/sigmaplot/), with the wavelength in the first column, the ellipticity values of the references (e.g., α-helix, antiparallel β-sheet, parallel β-sheet, turn and disordered) in columns 2–6 and the unknown protein in column 7, and fitting the data in column 7 to the equation $\theta_{\lambda i} = \Sigma\varepsilon_i S_i$ in which ΣS_i is constrained to equal 1 and ε_i is constrained to be ≥ 0. An example of the needed equations is given in SigmaPlot format in **Supplementary Equations**. The use of commercial graphics programs is recommended because they quickly find the best fit to the data with all of the fractional components being positive and one can easily plot the best fit to the unknown data. The major advantages of this approach are the use of an invariant database, which is useful for direct comparisons, and that better fits are obtained than with nonconstrained least squares methods. However, there are no good single standards for β-turns.

In nonconstrained least-squares fit the sum of the conformations is not constrained to equal one[20].

$$\theta_\lambda = \Sigma\varepsilon_i S_{\lambda i} + C \quad (2)$$

A DOS-based program for calculating nonconstrained least-squares fits is MLR. It is available in the program in **Supplementary**

TABLE 2 | Sources of circular dichroism analysis software.

Website	Software	Operating systems
CDPro http://lamar.colostate.edu/~sreeram/CDPro/ListPro.htm **Advantages:** Data conversion program included. Superior fits of data on globular proteins. Source code available and can be compiled for use with LINUX or UNIX machines.	SELCON3 CONTIN CONTINLL CDSSTR	Windows 95, 98, XP; LINUX, UNIX
Circular Dichroism at UMDNJ http://www2.umdnj.edu/cdrwjweb/. These programs are also in the file **CD.Zip** in **SUPPLEMENTARY MATERIALS.** **Advantages:** Data conversion programs are supplied. Data collected at any range of wavelengths can be analyzed. Peptide references for least squares analyses are included. **Disadvantages:** Programs are not user friendly. Data conversion and text editing necessary.	LINCOMB MLR SELCON SELCON2 CONTIN VARSLC K2D CCA	MS-DOS, Windows 95, 98, XP
CONTIN http://s-provencher.com/index.shtml http://s-provencher.com/pages/contin.shtml **Disadvantages:** Programs must be compiled by user.	CONTIN	LINUX, Source Code
CCA + the CD Spectrum Analyser System http://www2.chem.elte.hu/protein/programs/cca/ **Advantage:** Windows operating system.	CCA	Windows 95, 98
DICROPROT http://dicroprot-pbil.ibcp.fr/ ftp://ftp.ibcp.fr/pub/C_DICROISM/ **Advantages:** Easy to use. Every method uses the same input format. Output files are shown on the screen immediately. **Note:** Programs to convert data to the DicroProt 'dic' format are in **SUPPLEMENTARY MATERIALS.** **Disadvantages:** The package does not have a constrained least squares analysis module where the fractional spectral components are constrained to be positive. Many of the programs do not run unless the CD data is in the range of 260–178 nm.	LINEAR REGRESSION SELCON2 SELCON3 CONTIN VARSLC	Windows 95, 98, XP
DICHROWEB http://www.cryst.bbk.ac.uk/cdweb/html/home.html **Advantages:** On-line analysis available. Accepts input directly from the output of many different CD machines.	CONTINLL SELCON3 CDSSTR VARSLC K2D	Online
K2D http://www.embl-heidelberg.de/%7Eandrade/k2d.html **Advantages:** Simple to use.	K2D	Online, DOS, Windows 95, 98, XP
SOMCD http://geneura.ugr.es/cgi-bin/somcd/index.cgi **Advantages:** Update of K2D algorithm. Analyzes turns as well as α-helix and β-sheets and uses data from 240–190 nm as well as from 240–200 nm.	SOMCD	Online

Software. It also can be evaluated using graphics programs with curve-fitting routines. An example of such a routine is also given in **Supplementary Equations**. The advantages of this approach are that there is no need to know protein concentration, invariant standards are used and it is useful for analyzing difference spectra. However, it is the least accurate method.

Ridge regression (CONTIN). CONTIN (Ref. 27) fits the CD of unknown proteins by comparison to a linear combination of the spectra of a large database of proteins with known conformations. In this method the contribution of each reference spectrum is kept small unless it contributes to a good agreement between the theoretical best fit curve and the raw data. This method results in relatively good estimates of α-helices and β-sheets. Different references are used for every fit; this is an advantage for obtaining the best fits of the data, but it complicates the quantitative analysis of the effect of a mutations or denaturant because a different set of standards is used for each analysis.

Singular value decomposition. SVD (Ref. 17) extracts basis curves with unique nodes from a set of spectra of proteins with known structures. The basis curves are each characterized by a mixture of secondary structures and are then used to analyze the conformation of unknown proteins. The sum of weights is not constrained to equal 1. This method provides the best estimates of α-helical content of proteins. However, it provides terrible estimates of β-sheets and turns if data are not collected to at least 184 nm. It is also unsuitable for the analysis of polypeptides and protein fragments.

Variable selection (VARSLC, CDSSTR). In the variable selection method[28] an initial large database of standard spectra from proteins with known spectra and secondary structures is created. The protein structure is then determined using SVD. Some of the protein spectra are then eliminated systematically to create new databases with a smaller number of standards. SVD is performed using all of the reduced data sets, and those fulfilling selection criteria for a good fit are averaged. Two versions of the program are available: VARSLC (also called Varselec) and CDSSTR. This method provides superior fits of the conformation of globular proteins. However, it does not always give fractional content of various conformations that add up to 100%, and it can be very slow if a large number of reference sets are used and more than two spectra are eliminated from the combinations. It is also unsuitable for the analysis of polypeptides and protein fragments.

Self-consistent method (SELCON). In SELCON, the spectrum of the protein to be analyzed is included in the basis set and an initial guess is made for the unknown structure as a first approximation. The resulting matrix equation is solved using the SVD algorithm, and the initial guess is replaced by the solution. The process is repeated until self-consistency is attained. The program works well for globular proteins. The original program, SELCON (Ref. 32), evaluates α-helix, antiparallel and parallel β-sheets, turns and remainder. SELCON2 (Ref. 43) is modified to use a data set where the poly-L-proline II conformation is evaluated in addition to α-helix, total β-sheets and turns. In the version on the CDpro website (see **Table 2**), SELCON3 (Ref. 36), the α-helix and β-sheet conformations in globular protein structures are divided into regular and distorted fractions by considering a certain number of terminal residues in a given helical or strand segment to be distorted. The number of α-helical and β-strand segments and their average length in a given protein are estimated from the fraction of distorted helical and strand conformations relative to the total helix and strand content. The main advantage of this method is that it provides very good estimates of the structure of globular proteins. However, SELCON3 gives poor estimates of turns compared to SELCON1 and SELCON2, VARSLC and CDNN. All the versions of SELCON are unsuitable for the analysis of polypeptides and protein fragments.

Neural networks (CDNN, K2D, SOMCD). A neural network is an artificial intelligence program used to find correlations in data. Two widely used programs are CDNN (Ref. 37) and K2D (Ref. 38). CDNN analyzes data to determine helix, antiparallel and parallel β-structure, turns and remainder and K2D determines helix, total β-structure and remainder. SOMCD is a variation of K2D that uses a larger reference set to train the network and also analyzes turns[39]. A neural network is first trained using a set of known proteins, so that the input of the CD at each wavelength results in the output of the correct secondary structure. The trained network is then used to analyze unknown proteins. K2D gives a good estimate of the helical and sheet contents of both proteins and polypeptides. However, the K2D program does not estimate turns. CDNN is not suitable for the analysis of polypeptides, and it currently is not being distributed.

Convex constraint algorithm. The CCA algorithm[44] deconvolutes a set of spectra into a desired number of basis spectra, which, when recombined, generate the entire data set with a minimum deviation between the original data set and the reconstructed curves. It is very useful to determine how many different states contribute to the changes in CD as a function of ligand or temperature[26,45–47]. The method was developed to estimate protein conformation but is poorer than least squares, SVD or neural net analysis.

The protocol described here will cover setting up CD machines to collect data; procedures for obtaining high-quality, reproducible data; and methods to analyze CD spectra to estimate the secondary structures of proteins. For more detailed information there are several reviews that discuss CD spectroscopy in detail, including the preparations of buffers, protein samples and cuvettes, and converting raw data to molar and mean residue ellipticity[48–53], and analysis of data to yield secondary structure information[3,26,50,54]. Three additional protocols will cover (i) the determination of the thermodynamics of protein folding from CD data collected as a function of temperature[47]; (ii) determination of the free energy of folding and binding constants from CD data collected as a function of denaturants, osmolytes or ligands[55]; and (iii) determination of the kinetics of folding from CD data collected as a function of time[56].

MATERIALS

REAGENTS

- Proteins to be analyzed, dissolved in an appropriate buffer
  All of the following reagents are available from Sigma–Aldrich (http://www.sigmaaldrich.com)
- Guanidine HCl (Step 1, options A and B (cat. no. G7153)
- Sodium hydroxide (NaOH) reagent grade pellets (Step 1, options A–C; cat. no. 22146)
- Sodium carbonate (Step 1, option C; cat. no. 222321)
- Sodium citrate (Step 1, option C; cat. no. S1804)
- BSA Cohn Fraction V (BSA) (Step 1, option C; cat. no. A3059)
- Copper sulfate (Step 1, option C; cat. no. C1297)
- Whatman #1 filter paper, diam. 11 inches; Step 1, option C; cat/ mp/ Z240087)

EQUIPMENT

- Circular dichroism instrument (see EQUIPMENT SETUP)
- Circular dichroism cuvettes (Hellma Cells Inc. (http://www.HellmaUSA.com or http://www.hellma-worldwide.de); NSG Precision Cells Inc. (http://www.nsgpci.com/) or Luzchem Research Inc. (http://www.luzchem.com/products/)).
- Spectrophotometer (Step 1, options A–C)
- Microtiter plate reader (Step 1, option C)
- Microtiter plates (Step 1, option C)
- 0.1–0.2 μm filters (Millipore; http://www.millipore.com)

REAGENT SETUP ● **TIMING 30 min to 2 h**

! CAUTION Solutions of guanidine HCl, NaOH, and Benedict's reagent are caustic. Wear gloves.

6 M Guanidine HCl, pH 7.1 Dissolve 57.3 mg guanidine HCl in ~90 ml glass-distilled or de-ionized water. Adjust the pH to 7.1 by adding either 1 M HCl or solid NaOH pellets. Adjust the volume to 100 ml. Keeps indefinitely in a glass bottle with a plastic cap.

6 M Guanidine HCl, pH 12.5 Dissolve 57.3 mg guanidine HCl in ~90 ml of water. Adjust the pH to 12.5 by adding NaOH pellets. Adjust the volume to 100 ml. This solution should be checked before use and the pH adjusted if necessary to >12 by adding solid NaOH pellets, because carbon dioxide absorbed from the air will lower the pH.

6 M Guanidine HCl, pH 6.5 Dissolve 57.3 mg guanidine HCl in ~90 ml water. Adjust the pH to 6.5 by adding either 1 M HCl or solid NaOH pellets. Adjust the volume to 100 ml. Keeps indefinitely in a bottle with a plastic cap.

Benedict's reagent Combine 50 g of sodium carbonate with 86.5 g of sodium citrate in 300 ml of water. Dissolve by stirring on a hot plate. Filter through Whatman no. 1 filter paper. Add 8.63 g of $CuSO_4$ dissolved in 50 ml of water. Dilute to 500 ml. This solution lasts about 1 year in a brown bottle at room temperature.

3% NaOH Dissolve 3 g of NaOH pellets in 100 ml of water.

BSA, 1 mg ml^{-1} Dissolve 2 mg of BSA in 1 ml of water and filter through a Millipore filter. The A_{280} of a 1 mg ml^{-1} solution of BSA is 0.68. Dilute to a final concentration of 1 mg ml^{-1}.

EQUIPMENT SETUP

CD machines For data collection from ~700–175 nm, machines can be obtained from Applied Photophysics Ltd. (http://www.photophysics.com), Aviv Biomedical (http://www.avivbiomedical.com/), Jasco Inc. (http://www.jascoinc.com/) or Olis (http://www.olisweb.com/products/cd/). For data collection at lower wavelengths there are CD machines than use beamline radiation from synchrotrons: the NSLS Brookhaven, USA (beamlines U9b and Ul1); ISA in Aarhus, Denmark (beamline UV_1); the SRS Daresbury, UK (beamline CD12); HSRC/HiSOR, Hiroshima, Japan; BESSY2 in Berlin, Germany; the BSRF in Beijing, China; and the NSRL in Hefei, China. ▲ CRITICAL CD machines must be calibrated on a regular basis to check that the ellipticity values and wavelengths are correct[52,57]. A commonly used calibration standard, crystallized CSA, (1*S*)-(+)-Camphor-10-sulfonic acid (Sigma-Aldrich cat. no. C2107; http://www.sigmaaldrich.com), 1 mg ml^{-1} in a 1-cm cell, has an A_{285} of 0.149 and an ellipticity band with a peak at 291 nm of 335 millidegrees. In addition, the ratio of ellipticity of CSA at 192.5–290.5 nm should be between 2.05 and 2.08. These numbers correspond to a $\Delta\varepsilon$ of 2.36 at 290.5 nm and 4.90 at 192.5 nm[17,30].

Cleaning cuvettes This procedure can be completed in 10 min or can be done overnight. Cuvettes for CD measurements must be clean and dry. Quartz cells can be cleaned by soaking in: mild detergent solutions available from several cell manufacturers, such as Hellma; a mixture of 30% concentrated HCl and 70% ethanol; or concentrated nitric acid. Protein residues can be dissolved using 6 M guanidine HCl. After soaking, cells should be rinsed with water and ethanol and either dried by suction using an aspirator or blown dry with nitrogen or compressed air that has passed through a filter to remove impurities. If residual proteins are not removed by the previous cleaning agents, filling the cells with Chromerge (a mixture of potassium chromate in concentrated sulfuric acid) (Fisher Scientific cat. no. C577-12; https://www1.fishersci.com) and immediately rinsing out with water and drying usually is effective. This method is the best method for cleaning cells with 0.01- or 0.02-cm pathlengths.

! CAUTION Nitric acid, HCl and Chromerge are very caustic and will burn holes in lab coats and damage clothing. Wear gloves.

Starting the CD machine This procedure requires 30 min. CD machines have very powerful lamps that promote the ionization of oxygen to ozone. Ozone is toxic and also will quickly destroy the mirrors in the optics of the machines. Most CD machines must be flushed with nitrogen to remove oxygen before the machine is turned on and during operation. Nitrogen sources include tanks of prepurified nitrogen, which last ~5 h, and high-pressure liquid nitrogen tanks, which produce gas and will last 1–2 weeks depending on usage. Commercial nitrogen should be free of oxygen and most other impurities, but some manufacturers suggest using a trap to remove impurities for added safety. If you are new to CD and there is no one to teach you the operation of your specific machine, call the manufacturer and ask for their start-up protocol.

PROCEDURE

Determination of protein concentration ● TIMING 0.5–1 h

1| Determine the concentration of the protein stock solutions. The following three simple spectroscopic methods give rapid, accurate measurements of protein concentration and are independent of protein composition, provided the protein is unfolded in 6 M guanidine HCl or 3% NaOH. Options A (Refs. 58, 59) and B (Ref. 58) require that the protein have tyrosines or tryptophans. Option C (Ref. 60) does not work with collagen-like proteins with high proline contents.

(A) Determination of protein concentration from the difference spectrum of the protein dissolved in 6 M guanidine at pH 12.5 versus pH 7.1

(i) Pipette exactly the same volume (0.4–1 ml depending on the sample volume of the cuvette) of each solution into two cuvettes with 1-cm pathlengths, and scan the baseline from 320–270 nm with the pH 7.1 solution in the reference compartment and the pH 12.5 solution in the sample compartment.
▲ CRITICAL STEP The dilutions of the samples must be exactly the same in the reference and sample compartments.

(ii) Add exactly the same volume (e.g., 10–100 µl) of protein solution to each cuvette and obtain the difference spectrum. Correct the spectrum for the baseline.

(iii) Calculate the concentration of protein. The molar concentration (in mol liter^{-1}) of protein in the cuvette = $A/(2{,}357Y + 830W)$ (Ref. 59), where A is the absorbance at 294 nm, Y is the number of tyrosines and W is the number of tryptophans. Correct the measured concentration for dilution. The mean residue concentration can be calculated by multiplying the molar concentration by the number of amino acids in the protein. The number of milligrams per milliliter of protein is calculated by multiplying the molar concentration by the molecular weight. If the proteins are denatured by the guanidine solution, the difference method should give one band at 294 nm.
? TROUBLESHOOTING

(B) Determination of protein concentration from the aromatic spectrum determined in 6 M guanidine HCl, pH 6.5

(i) Run a baseline spectrum of two cells with an equal volume of guanidine HCl in each side.

(ii) Add a small aliquot of protein solution to the sample side and an equal volume of the buffer to the reference side.
▲ CRITICAL STEP The protein must be free of scattering material and 2-mercaptoethanol or DTT. The oxidized form of these compounds absorb strongly at 280 nm, and the oxidation rates are faster in solutions containing protein than in plain buffers, so it is difficult to subtract their contributions.

(iii) Collect the spectrum of the protein between 350–250 nm, and calculate the protein concentration using the formulas:

$$\varepsilon_{288} = 4{,}815W + 385Y + 75C \text{ and } \varepsilon_{280} = 5{,}690W + 1{,}280Y + 120C$$

where *W*, *Y* and *C* are the numbers of tryptophans, tyrosines and cystines (oxidized) per mole of protein and ε_{288} and ε_{280} are the molar extinction coefficients of the protein at 288 nm and 280 nm, respectively. The protein concentrations (in mol liter^{-1}) are the absorbance value at 288nm/ε_{288} and 280nm/ε_{280}. The concentrations determined at 288 and 280 nm should agree.

? TROUBLESHOOTING

(C) Determination of protein concentration using a microbiuret procedure

(i) Aliquot protein samples in buffer (e.g., 0, 0.025, 0.05 and 0.1 ml) and add the buffer to a final volume of 0.1 ml in small, clean test tubes.

(ii) Prepare a standard curve containing 0.02, 0.04, 0.06, 0.08 and 0.10 ml of BSA diluted to a final volume of 0.1 ml for concentrations of 0.2, 0.4, 0.6, 0.8 and 1.0 mg ml^{-1}, respectively.

(iii) Add 0.5 ml of 3% NaOH and 0.02 ml of Benedict's reagent to the standards and samples. Mix with a vortex mixer.

(iv) Allow to stand at least 15 min for the color to develop.

(v) Read the absorbance at 330 nm (y-axis).

(vi) Plot the standard curve of absorbance as a function of protein concentration in mg ml^{-1} (x-axis). Correct the absorbance of each unknown sample for the contribution of the buffer, and read the protein concentration from the curve. Correct for dilution. This assay can be done in microtiter plates using half the volume of each reagent. The plates can be read at 340 nm, although the color intensity is lower at this wavelength than at 330 nm. The microbiuret method, using freshly prepared Benedict's reagent, should be linear for protein concentrations ranging between 0 and 1.5 mg ml^{-1}, with the absorbances ranging from ~0.2 for the blank (100 µl buffer) + 0.5 ml of 3% NaOH and 0.02 ml of Benedict's reagent to ~0.5–0.6 for the sample with 100 µl of BSA, 1.5 mg ml^{-1}.

? TROUBLESHOOTING

Sample preparation ● TIMING 10–30 min

2| Prepare the protein samples. For typical measurements in a 0.1-cm cell, depending on the buffer (see **Table 1**), make solutions of 0.05–0.2 mg ml^{-1} protein. For measurements in 0.01- to 0.02-cm cells use 0.2–1 mg ml^{-1} protein, and for 1-cm cells use 0.005–0.01 mg ml^{-1} protein.

Equipment preparation ● TIMING 30–40 min

3| Turn on the nitrogen, and flush the optics compartment for 15–20 min before starting the machine (see manufacturer's suggested time).

! CAUTION Nitrogen displaces oxygen. Operate a CD machine only in a well-ventilated room. Do not close the door. If a tank of liquid nitrogen begins to vent because pressure has built up, leave the room and allow the excess nitrogen to disperse before reentering.

4| Turn on the water supply or circulating water bath chiller to the lamp housing, if the lamp is water cooled. If a water supply is used, make sure the filter is clean. If a bath is used, make sure the water is clean.

▲ CRITICAL STEP Avoid using ethylene glycol in the water in the circulating bath, because it can damage the pumps.

5| Turn on the circulating water bath for temperature control. If the temperature is controlled using a temperature-regulated cell compartment that requires a heat sink, set the temperature of the bath to 20 or 25 °C. If the temperature is controlled using water-jacketed cells, set the temperature of the bath to the desired temperature (see Step 9).

▲ CRITICAL STEP Be sure that the water is circulating before turning on the power to the thermal regulators, or you can burn out the heating units.

? TROUBLESHOOTING

6| Turn on the lamp, if it has a separate switch, before you turn on the power to the rest of the CD machine or computers.

▲ CRITICAL STEP Firing the lamp may cause a voltage surge that can destroy electronic boards or computers in some machines if they are turned on before the lamp is lit.

? TROUBLESHOOTING

7| Turn on the CD machine and computer, and start the CD collection program.

? TROUBLESHOOTING

8| Set the data path of the operating program to store your data.

9| Set the desired temperature. For a previously uncharacterized protein, collect spectra at multiple temperatures and correlate the spectra with some measurement of the activity or the protein (e.g., enzyme activity, or ability to bind ligands or antibodies). Typically, preliminary spectra of proteins are collected at 4, 25, 37, 50, 60 and 70 °C. Once a stability range has

been established, data may be collected at more closely spaced intervals to determine whether there are spectral changes indicative of folding intermediates using the CCA algorithm[44] or SVD[47,61,62].

Collecting CD spectra for the determination of secondary structure ● TIMING 2–4 h for collection of five samples and five baselines at a single temperature, depending on wavelength range and interval and number of repeat spectra

10| Set the desired equilibration time. Usually globular proteins reach equilibrium within 2 min , but some proteins (e.g., collagen) can require literally days to fold (see preparation of proteins). If unsure of the folding time, incubate the CD sample on ice for several days before starting the CD measurements for measurements at 4 °C or at 25 °C for measurements near room temperature.

11| Set the half-bandwidth between 1 and 1.5 nm. These values give spectra with good signal-to-noise ratios and adequate spectral resolution, because UV bands are broad (**Fig. 2a,b**).

12| Set the wavelength range: from 260–185 nm for 0.1–0.2 mg ml^{-1} protein samples in transparent buffers (see **Table 1**) in 0.1-cm cells; from 260–178 nm for 0.2–0.8 mg ml^{-1} samples in 0.01-cm cells and from 260–200 nm for 0.01–0.02 mg ml^{-1} proteins in 1-cm cells.

13| Set the wavelength interval, 0.5 nm for samples with signal-to-noise ratios $>$20:1 or 0.1, 0.2 or 0.25 nm for samples with low ellipticity. Data collected at 0.10–0.5 nm intervals with half-bandwidths of 1–1.5 nm will give well-defined spectra (see **Fig. 2a,b**, where data were collected using a half-bandwidth of 1.5 nm and a wavelength spacing of 0.25 nm). ▲ CRITICAL STEP Only use intervals which have reciprocals that are integers, because most data analysis programs have databases with ellipticity values collected at 1-nm intervals.

14| Set the time for data collection at each point (i.e., signal averaging time). For samples with a concentration of $\sim$0.1 mg ml^{-1} in a transparent buffer, collecting for 1 s at each point should be sufficient. If collecting replicate spectra, at intervals from 0.1–0.5 nm, 0.5 s per point should be adequate. If the protein concentration is low or the buffer has high absorbance, increase the averaging time. The relative signal-to-noise ratio will increase as the square root of signal averaging time.

15| Set the instrument time constant. For routine collection of CD spectra this should be 100 ms. For rapid collection of data (e.g., in measurements of stopped-flow CD) the time constant should be decreased to $\leq$100 μs (no greater than one-tenth the data averaging time at each point).

16| Set the instrument to record the ellipticity and the photomultiplier tube (PMT) voltage. When light hits the photomultiplier of the CD machine a current is induced. Most CD machines maintain constant current by raising the voltage as the amount of light decreases. As it scans to lower wavelengths, the absorbance will increase and the PMT voltage will rise. The signal-to-noise ratio will greatly diminish once the PMT voltage exceeds 500 V, and the data often become very noisy and unreliable.

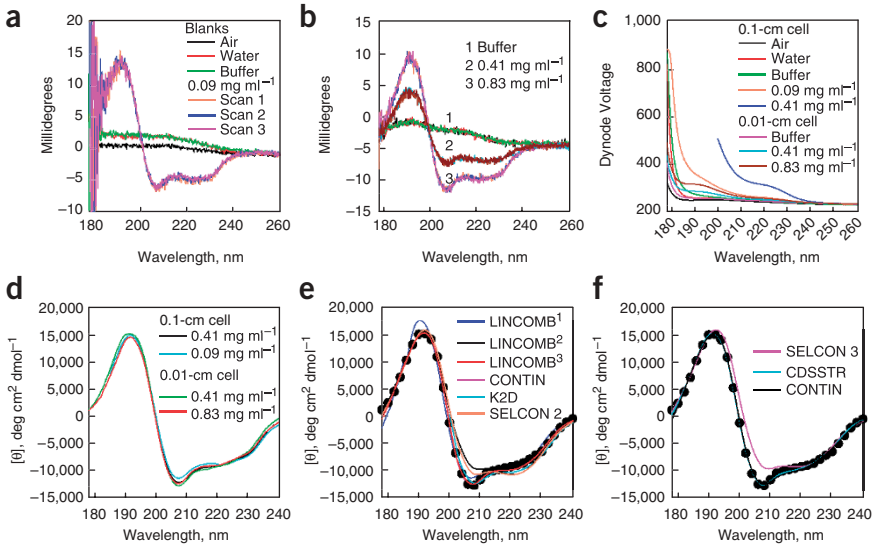

Figure 2 | Circular dichroism spectra of lysozyme in 10 mM sodium phosphate pH 7.0. (**a**) The spectra of air (black), water (red), buffer (green) and three replicate spectra of lysozyme, 0.09 mg ml^{-1} in a 0.1-cm cell (orange, blue and magenta). (**b**) The spectra of buffer (black, red, green) and three replicate spectra each of lysozyme at 0.41 mg ml^{-1} (cyan, purple, brown) and 0.83 mg ml^{-1} concentration (blue, magenta, orange) in a 0.01-cm cell. (**d**) The change in the photomultiplier tube dynode voltage as a function of wavelength for the conditions illustrated in **Figure 1a** and **b** and for 0.41 mg ml^{-1} lysozyme in a 0.1-cm cell. The mean residue ellipticity of lysozyme in a 0.1-cm cell (black), 0.41 mg ml^{-1} (raw data not shown); (cyan) 0.09 mg ml^{-1}; and in a 0.01-cm cell (green), 0.41 mg ml^{-1}; (red), 0.83 mg ml^{-1}. (**e**) The mean residue ellipticity of lysozyme (black circles) fitted using the method of least squares: using a peptide data base[20] (blue); four basis spectra extracted from 17 proteins (black)[26]; five basis spectra extracted from 33 proteins (orange)[26] and CONTIN (magenta)[27]; K2D (cyan)[38] and SELCON2 (orange)[43]. (**f**) The mean residue ellipticity of lysozyme (black circles) fitted using the CDPro Package[33]: SELCON3 (magenta), CDSSTR (cyan) and CONTIN (black). Note: Lysozyme was obtained from Sigma (L6876) and dissolved in sodium phosphate, 10 mM. The protein concentration was determined using the published extinction coefficient of $\varepsilon_{1\%}$ of 26.5 ($\varepsilon_M = 38.2 \times 10^3$) (Ref. 65) for comparison with previous data[5]. Data were obtained on an Aviv Model 215 spectrometer (Aviv Biomedical).

(On an Aviv instrument the PMT voltage is called the dynode voltage, whereas on a Jasco it is called HT voltage, on an Applied Photophysics machine it is called the detector gain and on an Olis machine it is called the PMT HV). For protein concentrations ranging from 0.05–0.1 mg ml^{-1} in 0.1-cm cells in 10 mM phosphate buffer, the signal-to-noise ratios are usually better than 10:1 over a wavelength range of 260–185 nm (**Fig. 2a**) with dynode voltages below 500 V (**Fig. 2c**). With 0.01-cm cells and 0.4–0.8 mg ml^{-1} concentrations, the signal-to-noise ratio of the data is usually $>$30:1 between 260–178 nm with dynode voltages $<$500 V (**Fig. 2b,c**). When the dynode voltage (**Fig. 2c**) exceeds 500 V (e.g., below 185 nm for data collected in a 0.1-cm cell at $\sim$0.1 mg ml^{-1} concentration), the signal-to-noise ratio usually becomes very poor (**Fig. 2a**), although the signal may still be linear as a function of concentration on some instruments up to 700 V.

17| Determine the spectrum of the blank. Fill the cell with water and determine its spectrum. Suitable buffers should have no ellipticity, but their increased absorbance as compared with water decreases the signal-to-noise ratio. The spectrum of the cell containing water should be relatively flat but may be displaced from that of pure air because of the birefringence of the cell (**Fig. 2a**).
? TROUBLESHOOTING

18| Collect a spectrum of the buffer to make sure that it does not have any ellipticity due to dichroic components or a high absorbance leading to a very poor signal-to-noise ratio and possibly false peaks. Note that phospholipids are asymmetric and have CD bands. If samples are suspended in phospholipids it is essential that the spectrum of the blank contains the same concentration of protein-free lipids. The spectrum of the buffer and water should overlay each other, within the experimental error, but the spectrum of the buffer usually has a lower signal-to-noise ratio than the spectrum of water at low wavelengths (see **Fig. 2a**).
? TROUBLESHOOTING

19| Clean the cell, fill with protein solution and collect the CD spectrum. It is best to collect data two or three times for new samples to make sure that the sample is at equilibrium and the signal is not changing as a function of time. Many CD machines can collect multiple spectra automatically. If the protein is at equilibrium, replicate scans of the protein solutions should overlay each other and not drift as a function of time (**Fig. 2a,b**).
? TROUBLESHOOTING

20| If the data sets overlay each other one, average the data sets. For the most accurate estimates of protein secondary structure, data should be collected to 178 nm or lower wavelengths, in 0.01–0.02 cm cells. because data collected at $<$200 nm may have low signal-to-noise ratios, one should collect three to five scans and average them.

21| Save the raw data on the hard drive, floppy disk or other media.
▲ CRITICAL STEP Always immediately save the data to prevent loss if there is a power failure.

22| Smooth the spectra of the sample and blank. Most CD machines have built-in smoothing algorithms, and some will automatically pick the best smoothing parameters; refer to the manual. The smoothing algorithms that are used depend on the manufacturer. If the machine uses Savitsky–Golay smoothing[63] and data are collected at 0.5-nm intervals, a polynomial order of 3 and a smoothing window of 15 points usually gives a good fit. If data is collected at shorter wavelength intervals, increase the number of points. Some smoothing protocols give estimates of the goodness of smoothing by calculating whether the difference between the raw and smoothed data has the statistical characteristics of noise.

23| Check that the data have not been oversmoothed by subtracting the smoothed curve from the raw data. The points should be evenly distributed around zero. Some CD machines will automatically calculate the residuals from the smoothing, and these can be viewed on the spectrometer in its data viewing mode.
? TROUBLESHOOTING

24| Subtract the smoothed baseline from the smoothed spectrum of the sample. The ellipticity for most proteins should be close to zero between 260 and 250 nm.
? TROUBLESHOOTING

25| Convert the data to mean residue ellipticity or $\Delta\varepsilon$ using the formulas:

Ellipticity, $[\theta]$, in deg cm^2 $dmol^{-1}$ = (millidegrees $\times$ mean residue weight)/(pathlength in millimeters $\times$ concentration in mg ml^{-1}) or
$[\theta]$ = millidegrees/(pathlength in millimeters $\times$ the molar concentration of protein $\times$ the number of residues)
$\Delta\varepsilon = [\theta]/3{,}298$

The mean residue weight of a peptide is the molecular weight divided by the number of backbone amides (number of amino acids – 1 if the protein is not acetylated). It is $\sim$115 for most proteins if the molecular weight of the sample is not known, but it should be calculated directly if the molecular weight of the protein is known for accurate results. If the protein or peptide is

TABLE 3 | Comparison of various methods to estimate the secondary structure of representative proteins and polypeptides.

METHOD		X-Ray	LINCOMB*		MLR*		CONTIN		SELCON		VARSLC		K2D	CDNN	
Lowest wavelength, nm			200	178	200	178	200	190	200	178	200	178	200	200	180
Myoglobin	α-helix	78	96	93	89	97	67	89	73	79	76	74	74	83	84
	β-sheet	0	0	0	0	0	0	0	−3	0	0	0	8	3	2
	Turn	10	4	5	8	3	16	0	13	20	4	18	ND	9	9
Lactate	α-helix	37	46	40	63	42	46	40	41	39	40	39	55	42	44
dehydrogenase	β-sheet	14	21	29	15	33	7	39	12	27	15	28	11	13	10
	Turn	25	15	11	13	9	26	1	22	27	17	13		15	14
Chymotrypsin	α-helix	10	15	21	33	28	11	9	7	15	24	16	12	19	20
	β-sheet	38	25	14	6	0	16	32	16	26	0	11	33	29	50
	Turn	26	10	16	5	9	44	32	13	16	42	39	ND	21	22
Bence Jones protein	α-helix	3	0	0	0	0	6	0	3	9	14	0	3	13	16
	β-sheet	50	43	40	68	40	42	81	47	34	7	69	50	39	22
	Turn	24	25	29	17	28	25	10	23	40	28	17	ND	22	22
Lowest wavelength, nm			200	190	200	190	200	190	200	178	200	178	200	200	180
Poly(lys-leu)n	α-helix	0	0	9	0	2	12	9	31	24	31	30	5	34	39
In 0.5 M NaF, pH 7	β-sheet	100	89	65	89	99	73	91	24	39	51	55	89	18	13
	Turn	0	11	26	7	0	0	0	26	26	1	3	ND	16	13
Lowest wavelength, nm			200	190	200	190	200	190	200	190	200	190	200	190	190
Poly-L-lysine 0.01%	α-helix	100	97	100	-	-	100	100	99	89	100	96	100	92	94
helical form, pH 11.1	β-sheet	0	3	0	-	-	0	0	0	1	6	18	0	1	1
In water	Turn	0	0	0	-	-	0	0	1	6	13	15	0	7	7

*The polypeptide reference set of Brahms and Brahms[20] for α-helix, β-structure, generic β-turn and disordered were used as standards for the evaluation the helical contents of the four proteins. The polypeptide reference set of Reed and Reed[21] for α-helix, β-structure, β-turns (type 1 and type 2) and disordered were used for the evaluation of the spectra of poly(lys-leu) in the β-structure conformation and poly-L-lysine in the α-helical conformation.

monomeric and does not aggregate under the experimental conditions used to collect the CD data, the spectra collected at different protein concentrations and pathlengths should give the same mean residue ellipticities (**Fig. 2d**).

? TROUBLESHOOTING

26| Save the corrected data of each sample in separate text files (not binary) as mean residue ellipticity, [θ] (*y*-axis) as a function of wavelength (*x*-axis), or in files with the ellipticity values, which also have information about the starting wavelength, ending wavelength and data interval so that it can be converted into formats that can be used to estimate protein conformation.

▲ **CRITICAL STEP** Save the data in text (ASCII) format. Binary files are in proprietary formats that can only be accessed by the correct CD machine and cannot be edited by a text editor or imported into a spreadsheet or converted to formats that can be used for analysis programs.

27| If data have been collected at different wavelength ranges, e.g. 260–190 nm in 0.1-cm cells and 200–178 nm in 0.01-cm cells, and the data agree between 200 and 190 nm, combine the data containing the mean residue ellipticity into a single file using a text editor. If the data do not agree with each other, make sure the proper baseline has been subtracted for each sample and that the pathlengths of the cells are correct. If they still do not agree, repeat the measurements.

▲ **CRITICAL STEP** For data analysis, ensure there is a single entry for each wavelength in the final file, or it will confuse the data analysis programs.

Data analysis ● TIMING 2–16 h

28| Analyze the data using appropriate methods, four of which are provided here as (A) through (D). To obtain the most accurate results, use as many methods as possible.

(A) Evaluating the secondary structure of globular proteins using data collected between 260 and 178 nm

● TIMING ~10–15 min per spectrum for data conversion, ~10–15 min per spectrum for analysis

The CDPro package or the online analysis programs at DicroWeb are recommended for the most accurate estimates of secondary structure. The DicroProt suite of programs uses older versions of the programs available at DicroWeb and in CDPro, but the results are similar to those of the more modern versions. It is easy to convert data to both the CDPro and the DicroProt 'dic' format, and the user interface is simple for both programs. A simple program to convert CD data in columns of ellipticity as a function of wavelength to the dic format, CONVERT.EXE, is in **Supplementary Software**. Converting data to the formats used by the CDPro package and analyzing it using SELCON, CONTINLL and CDSSTR will require ~10–15 min per spectrum. Converting data to the dic format and then analyzing it by all the programs in the DicroProt package including VARSLC, SELCON2 and SELCON3, CONTIN and K2D requires ~10–15 min per spectrum.

(B) Evaluating the secondary structure of globular proteins using data truncated below 260 nm or above 178 nm ● TIMING ~10 min per spectrum per program

For data sets collected over truncated wavelength ranges (e.g. 240–200 nm), use the programs SELCON, VARSLC, K2D, CONTIN or LINCOMB (with a database extracted from proteins as references). These programs are all included in **Supplementary Software**, which also includes conversion programs that will edit the reference sets so that they match the range of data in the sample files. Reference sets for LINCOMB are also included. Converting raw CD data and running each program will require ~10 min per spectrum per program, because each of these older versions of the program has a different format for its input files, and the CD spectrum must be converted separately for each program.

(C) Analyzing the effect of mutations on the helical or β-sheet content of a protein ● TIMING ~20 min per spectrum

Use the constrained method of least squares (e.g., LINCOMB) with a fixed reference set so that the same standards will be used to analyze the wild-type and mutant proteins. This method should also be used for quantifying the conformational changes in response to addition of a ligand. Analysis of data using the LINCOMB or MLR programs is slow and takes about 20 min per spectrum, because the programs are DOS based and use the old DOS graphic screens, which take time to display the results. In addition, one must first convert the data to the correct format, then run the program and finally examine the results with a text editor.

(D) Analysis of the conformation of polypeptides or short protein fragments ● TIMING 10–15 min per spectrum

Use LINCOMB (with polypeptide references), CONTIN or K2D. These programs are in **Supplementary Software**.

? TROUBLESHOOTING

● TIMING

Summary of procedure for collecting and analyzing one CD spectrum

Protein sample preparation and concentration determination: 0.5–1 h (can be done in advance)
Machine setup: 0.5 h
Data collection (assuming that three replicate data sets are collected for both the protein sample and the reference between 260 and 185 nm at 0.5 nm intervals with a 0.5 s signal averaging at each point): 0.5 h
Averaging the data, correcting the data for the baseline, smoothing the data and converting the data to mean residue ellipticity: 10 min
Converting the data to CDPro format and analyzing the data with SELCON3, CONTINLL and CDSSTR: 10 min
Cleaning the cuvette between and after spectra collection: 10 min

When collecting data on multiple samples, the data correction and analysis can be done on the first sample while data on subsequent samples are being collected. Overall, this procedure can be completed in approximately 2–2.5 h.

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 4**.

TABLE 4 | Troubleshooting table.

Step	Problem	Solution
1(A)	Difference spectra have multiple peaks between 280 and 300 nm	The protein is not fully denatured by guanidine HCl. Try a different method.
1(B)	Analysis at 280 and 288 nm gives different results	The protein is not fully denatured, or the sample has nonprotein contaminants. Try a different method.
1(C)	Standard curves are nonlinear	Make fresh Benedict's reagent or try a lower concentration range for the standard.
5	The water is not circulating through the sample compartment	The inlet or outlet lines may be clogged. Replace the fluid in the tank to make sure it is clean. Try reversing the flow to the temperature controller to dislodge any blockage. You may have to have the unit repaired.
6	The lamp does not turn on	Some machines require the lamp base to be warmed up before firing. Wait a minute and try again. Make sure that the ignition switch is not stuck. If this does not work, turn off the instrument. Disconnect the power switch, and check that the lamp is not burned out and that the leads to the lamp are connected properly. Each machine is different, so consult the manual. You may need a service call.
7	The machine does not turn on	First make sure everything is plugged in. If the machine is controlled by a computer, make sure that it is turned on and that the monitor is also turned on. Consult the manual to see how to start the collection program if it does not start automatically when the machine is turned on.

TABLE 4 | Troubleshooting table (continued).

Step	Problem	Solution
17	The spectrum of the blank has a very large ellipticity	(i) Make sure that the cell is not under pressure, caused by, e.g., cell adapters being too tight or the water pressure being too high if a water-jacketed cell is being used. (ii) Try recleaning the cell, or try a different cell. (iii) Check that the light is going directly through the cell by setting the wavelength to 500 nm and putting a piece of paper before or after the cell holder to visualize the light path. **! CAUTION** Do not look at the light beam in the UV region.
18	The spectrum of the buffer does not agree with that of water	Make sure that the buffer does not have any optically active components, such as glutamate or ATP, and that the buffer is transparent in the wavelength range of interest.
19	A. Successive scans of the same solution do not overlay each other	A. (i) Allow the solution to equilibrate for a longer time. (ii) Check the spectrum of air to make sure the lamp is stable and the baseline is not drifting. If there is a problem, contact the manufacturer.
	B. There is very low ellipticity	B. Check the concentration of the sample. If it was determined by absorbance at 280 nm or a dye binding technique, it could be too low. Oxidized DTT and 2-mercaptoethanol absorb at 280 nm, and dye binding is very dependent on the composition of a protein.
	C. The signal-to-noise ratio is very low	C. (i) Make sure that the half-bandwidth is set to at least 1 nm and that the slits are open. (ii) Check the absorbance of the sample + cell on a spectrometer. Signal-to-noise ratio decreases when the absorbance is >1. (iii) Measure the ellipticity of a standard solution of CSA, and make sure that the instrument is calibrated properly. (iv) Check the age of the lamp; most lamps begin to deteriorate above 1,000 h of use. (v) Mirrors in CD machines deteriorate with time. If the machines are old or used heavily, the mirrors in the optics may require replacement. Call the manufacturer.
	D. The dynode voltage does not increase as the wavelengths decrease, and the CD bands below 200 nm have much lower magnitude that those at higher wavelength	D. This result may be caused by 'stray' light. (i) Make sure that the half-bandwidth is not >2 nm. (ii) Try lowering the half-bandwidth. (iii) If the problem persists, contact the manufacturer to make sure that the slits are functioning properly, the optics are properly aligned and the light source is focused on the sample.
23	When the smoothed spectra are subtracted from the raw data, the noise is not distributed evenly around zero	Try either lowering or raising the number of points or the degree of the polynomial used for smoothing.
24	The ellipticity is flat between 260 and 250 nm but is greater or less than zero	Sometimes the baselines of CD machines will drift slightly when the machine is first turned on. The data can be corrected for the drift by adding or subtracting a constant so that the ellipticity between 260 and 250 nm is at zero. For highest accuracy, the spectrum of the baseline should be taken again to make sure that the displacement is due to instrumental drift and that the displacement is constant over the entire wavelength range.
25	The mean residue ellipticity seems much too high or too low. (Typically proteins will have ellipticity maxima or minima between 222 and 200 nm ranging from $\sim$5,000 to –45,000 deg cm^2 $dmol^{-1}$ (**Fig. 1**).	Check that the protein concentration is correct and that the data were divided by the correct pathlength to calculate the mean residue ellipticity.
23	The mean residue ellipticity changes as a function of protein concentration	The protein may not be monomeric or may be aggregating. Try to determine the oligomeric state of the protein using methods such as ultracentrifugation, light scattering or native gradient gel electrophoresis.
28	When analyzed by the curve fitting programs, the sum of the fractions of each conformation are much larger or much less than 1.	The protein concentration probably is not correct. Try a different method of determining the concentration.

TABLE 5 | Analyses of the structure of lysozyme from the CD data in **Figure 2**.

Method	Basis sets	Path Length, cm	Conc. mg ml^{-1}	Wavelength range, nm	Fraction of each secondary structure: Helix regular	Helix ends	Helix total	β regular	β ends	β total	Turns	Other
CDPro Package[a]												
X-Ray[b]					0.20	0.22	0.42	0.02	0.05	0.07	0.30	0.22
CDSSTR	30	0.01	0.41	260–178	0.19	0.14	0.33	0.11	0.08	0.19	0.20	0.28
CDSSTR	30	0.01	0.83	260–178	0.19	0.14	0.32	0.11	0.08	0.18	0.21	0.28
CDSSTR	30	0.1	0.085	260–185	0.20	0.13	0.34	0.14	0.08	0.22	0.19	0.25
CDSSTR	30	0.1	0.41	260–200	0.20	0.14	0.34	0.12	0.07	0.20	0.20	0.25
CONTIN	30	0.01	0.41	260–178	0.18	0.17	0.35	0.18	0.17	0.35	0.18	0.17
CONTIN	30	0.01	0.83	260–178	0.18	0.17	0.35	0.07	0.07	0.14	0.27	0.23
CONTIN	30	0.1	0.085	260–185	0.16	0.16	0.32	0.10	0.08	0.18	0.25	0.24
CONTIN	30	0.1	0.41	260–200	0.21	0.21	0.42	0.00	0.04	0.04	0.27	0.26
CONTINLL	30	0.01	0.41	260–178	0.19	0.16	0.34	0.08	0.07	0.15	0.23	0.27
CONTINLL	30	0.01	0.83	260–178	0.19	0.17	0.36	0.07	0.06	0.13	0.24	0.27
CONTINLL	30	0.1	0.085	260–185	0.17	0.15	0.32	0.11	0.08	0.18	0.24	0.25
CONTINLL	30	0.1	0.41	260–200	0.18	0.15	0.33	0.12	0.07	0.19	0.22	0.25
SELCON3	30	0.01	0.41	260–178	0.17	0.14	0.30	0.11	0.08	0.19	0.23	0.28
SELCON3	30	0.01	0.83	260–178	0.16	0.14	0.30	0.12	0.08	0.20	0.23	0.27
SELCON3	30	0.1	0.085	260–185	0.16	0.13	0.29	0.11	0.08	0.18	0.23	0.28
SELCON3	30	0.1	0.41	260–200	0.17	0.14	0.31	0.09	0.07	0.16	0.24	0.27
AVERAGE					0.18	0.16	0.34	0.10	0.08	0.17	0.23	0.26
STDEV					0.02	0.02	0.04	0.04	0.03	0.07	0.03	0.03
Other methods							Total helix			Total β	Turns	Other
X-RAY[c]							0.39			0.11	0.34	0.16
Method of constrained least squares fits with peptide references[d]												
Peptide set 1	4	0.01	0.41	240–178			0.26			0.29	0.02	0.42
Peptide set 1	4	0.01	0.41	240–200			0.31			0.26	0.02	0.40
Peptide set 2	5	0.01	0.41	240–190			0.24			0.31	0.06	0.38
Peptide set 3	3	0.01	0.41	240–208			0.28			0.15	ND	ND
Average							0.27			0.25	0.03	0.40
Stdev							0.03			0.07	0.02	0.02
Method of constrained least squares with references extracted from sets of protein spectra[e]												
Protein set 1	4	0.01	0.41	240–178			0.30			0.22	0.22	0.26
Protein set 2	5	0.01	0.41	240–178			0.27			0.22	0.32	0.20
Methods that use singular value decomposition to deconvolute the reference data sets[f]												
SVD	33	0.01	0.41	260–178			0.27			0.15	0.17	0.25
VARSLC	33	0.01	0.41	260–178			0.28			0.18	0.18	0.27
SELCON	17	0.01	0.41	260–178			0.31			0.18	0.28	0.24
SELCON2	17	0.01	0.41	260–178			0.32			0.16	0.24	0.29
Ridge regression[g]												
CONTIN	16	0.01	0.41	240–190			0.30			0.30	0.23	0.18
Methods using neural network analyses[h]												
K2D	18	0.01	0.41	240–200			0.33			0.14	ND	ND
CDNN	17	0.01	0.41	260–180			0.32			0.20	0.18	0.27
CDNN	17	0.01	0.41	260–200			0.32			0.18	0.17	0.31
Average off all the methods using protein references excluding those in the CDPro[a] package												
Average							0.30			0.19	0.22	0.25
Stdev							0.02			0.05	0.05	0.04

[a]The CDPro Package[36] contains the programs CONTIN (Ref. 27), CONTINLL (Ref. 36), SELCON3 (Ref. 36) and CDSSTR (Ref. 31). [b]The secondary structure of lysozyme was determined by the method of Kabsch and Sander (Ref. 64) and separated into classes of regular α and β helices and helices found at the ends of helical segments, or short fragments[36]. [c]The secondary structure of lysozyme was analyzed to total helices and turns by the method of Kabsch and Sander (Ref. 64). [d]The peptide references sets were 1, α-helix, β-structure, generic turn and other)[20]; 2, α-helix, β-structure, type 1 turn, type 2 turn, and others[21]; and 3 poly-L-lysine in the α-helical, β-structure and extended conformations[5]. [e]Protein data sets containing 33 proteins[29] or 17 proteins[32] were deconvoluted into 5 (α-helix, antiparallel β, parallel β, turn and other) or 4 reference spectra (α-helix, total β-structure, turn and other) by the method of least squares[22]. [f]The programs for extracting data from variations of the method of singular value decomposition analysis were: SVD (Ref. 17), VARSLC (Ref. 28), SELCON (Ref. 33) and SELCON2 (Ref. 43). [g]The data were analyzed by the original CONTIN method using 16 proteins[27]. [h]The neural net programs employed were K2D (Refs. 37,38) and CDNN.

ANTICIPATED RESULTS

Typical results for the study of a globular protein (chicken hen's egg white lysozyme) are given in **Figure 2,** and the results of the analysis of the spectra to give the secondary structure are given in **Table 5**. The data collection of the 10 individual spectra used to create **Figure 2**, including washing the cells between measurements, smoothing the data and correcting the data for the baseline and protein concentration, took ~4 h, and the 30 analyses summarized in **Table 5** took ~12 h, including converting the data to the appropriate formats, running the calculations and tabulating the results.

All of the methods used to analyze protein spectra should give reasonable estimates of α-helical content (**Table 5**). The four programs included in the CDPro package (SELCON3, CONTIN, CONTINLL and CDSSTR) should give results that are almost identical to each other and relatively independent of the lower limit wavelength range from 200–178 nm (**Table 5**). Similar secondary contents should be also obtained with SELCON and SELCON2, CDNN and K2D (**Table 5**). When proteins are analyzed using the method of least squares, basis sets extracted from proteins should give good fits, but some peptide data sets tend to overestimate β-structure and underestimate turns if data lower than 208 nm are fitted[5].

Some of the analysis programs output the fitted curve to the raw data. It should be noted that the programs giving the best fits to the data do not necessarily give the best estimates of protein conformation, because better fits will be obtained when more variables are used. Therefore, the fits obtained using CONTIN, which fits protein CD data by a large number of reference spectra, almost always gives a perfect fit to the raw data as compared with methods that use fewer reference sets. Representative fits using programs with graphical output are shown in **Figure 2e** and **f**.

SUPPLEMENTARY MATERIALS

Supplementary materials include equations to estimate secondary structure of proteins fitted them to basis sets using contrained and unconstrained least squares fits. The equations are in SigmaPlot format in **Supplementary Equations**. **Supplementary Methods** has the procedure for estimating protein concentration using a micro-Kjeldahl procedure. **Supplementary Software** has DOS-based versions of SELCON1, SELCON2, MLR, LINCOMB, VARSLC, K2D and the CCA algorithm. They run under MS-DOS and in a command box under Microsoft Windows 95, 98 and XP. Also provided are programs from converting data from Aviv, Jasco and text files to the formats needed for each program.

Note: Supplementary information is available via the HTML version of this article.

ACKNOWLEDGMENTS I thank M.A. Andrade, W.C. Johnson, Jr., N. Sreerama, R.W. Woody, S. Venyaminov, and especially the late Gerald D. Fasman for supplying versions of their software for analyzing CD data for use, analysis and distribution. The research was supported by a National Institutes of Health grant, GM-36326 to N.J.G. and to Sarah E. Hitchcock-DeGregori and by the Circular Dichroism Facility at Robert Wood Johnson Medical School (UMDNJ).

COMPETING INTERESTS STATEMENT The author declares that she has no competing financial interests.

Published online at http://www.natureprotocols.com

1. Velluz, L., Legrand, M. & Grosjean, M. *Optical Circular Dichroism: Principles, Measurements, and Applications* (Verlag Chemie Academic Press Inc., New York and London, 1965).
2. Beychok, S. Circular dichroism of biological macromolecules. *Science* **154**, 1288–1299 (1966).
3. Sreerama, N. & Woody, R.W. Computation and analysis of protein circular dichroism spectra. *Methods Enzymol.* **383**, 318–351 (2004).
4. Holzwarth, G. & Doty, P. The ultraviolet circular dichroism of polypeptides. *J. Am. Chem. Soc.* **87**, 218–228 (1965).
5. Greenfield, N. & Fasman, G.D. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8**, 4108–4116 (1969).
6. Venyaminov, S., Baikalov, I.A., Shen, Z.M., Wu, C.S. & Yang, J.T. Circular dichroic analysis of denatured proteins: inclusion of denatured proteins in the reference set. *Anal. Biochem.* **214**, 17–24 (1993).
7. Bovey, F.A. & Hood, F.P. Circular dichroism spectrum of poly-L-proline. *Biopolymers* **5**, 325–326 (1967).
8. Tiffany, M.L. & Krimm, S. Effect of temperature on the circular dichroism spectra of polypeptides in the extended state. *Biopolymers* **11**, 2309–2316 (1972).
9. Woody, R.W. Circular dichroism and conformation of unordered polypeptides. *Adv. Biophys. Chem.* **2**, 31–79 (1992).
10. Rucker, A.L. & Creamer, T.P. Polyproline II helical structure in protein unfolded states: lysine peptides revisited. *Protein Sci.* **11**, 980–985 (2002).
11. Nafie, L.A. Infrared and Raman vibrational optical activity: theoretical and experimental aspects. *Annu. Rev. Phys. Chem.* **48**, 357–386 (1997).
12. Jackson, M. & Mantsch, H.H. The use and misuse of FTIR spectroscopy in the determination of protein structure. *Crit. Rev. Biochem. Mol. Biol.* **30**, 95–120 (1995).
13. Cooper, E.A. & Knutson, K. Fourier transform infrared spectroscopy investigations of protein structure. *Pharm. Biotechnol.* **7**, 101–143 (1995).
14. Pelton, J.T. & McLean, L.R. Spectroscopic methods for analysis of protein secondary structure. *Anal. Biochem.* **277**, 167–176 (2000).
15. Lowry, O.H., Rosebrough, N.J., Farr, A.L. & Randall, R.J. Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* **193**, 265–275 (1951).
16. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
17. Hennessey, J.P. Jr. & Johnson, W.C. Jr. Information content in the circular dichroism of proteins. *Biochemistry* **20**, 1085–1094 (1981).
18. Lang, C.A. Simple microdetermination of Kjeldahl nitrogen in biological material. *Anal. Chem.* **30**, 1692–1694 (1958).
19. Jaenicke, L. A rapid micromethod for the determination of nitrogen and phosphate in biological material. *Anal. Biochem.* **61**, 623–627 (1974).
20. Brahms, S. & Brahms, J. Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. *J. Mol. Biol.* **138**, 149–178 (1980).
21. Reed, J. & Reed, T.A. A set of constructed type spectra for the practical estimation of peptide secondary structure from circular dichroism. *Anal. Biochem.* **254**, 36–40 (1997).
22. Saxena, V.P. & Wetlaufer, D.B. A new basis for interpreting the circular dichroic spectra of proteins. *Proc. Natl Acad. Sci. USA* **68**, 969–972 (1971).
23. Chen, Y.H., Yang, J.T. & Chau, K.H. Determination of the helix and β form of proteins in aqueous solution by circular dichroism. *Biochemistry* **13**, 3350–3359 (1974).
24. Chang, C.T., Wu, C.S. & Yang, J.T. Circular dichroic analysis of protein conformation: inclusion of the β-turns. *Anal. Biochem.* **91**, 13–31 (1978).
25. Yang, J.T., Wu, C.S. & Martinez, H.M. Calculation of protein conformation from circular dichroism. *Methods Enzymol.* **130**, 208–269 (1986).
26. Greenfield, N.J. Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. *Anal. Biochem.* **235**, 1–10 (1996).
27. Provencher, S.W. & Glöckner, J. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20**, 33–37 (1981).

28. Manavalan, P. & Johnson, W.C. Jr. Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal. Biochem.* **167**, 76–85 (1987).
29. Toumadje, A., Alcorn, S.W. & Johnson, W.C. Jr. Extending CD spectra of proteins to 168 nm improves the analysis for secondary structures. *Anal. Biochem.* **200**, 321–331 (1992).
30. Toumadje, A. & Johnson, W.C. Jr. Effects of relative band intensity on prediction of protein secondary structure from CD. *Anal. Biochem.* **211**, 258–260 (1993).
31. Johnson, W.C. Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins* **35**, 307–312 (1999).
32. Sreerama, N. & Woody, R.W. A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal. Biochem.* **209**, 32–44 (1993).
33. Sreerama, N. & Woody, R.W. Protein secondary structure from circular dichroism spectroscopy. Combining variable selection principle and cluster analysis with neural network, ridge regression and self-consistent methods. *J. Mol. Biol.* **242**, 497–507 (1994).
34. Sreerama, N., Venyaminov, S.Y. & Woody, R.W. Estimation of the number of α-helical and β-strand segments in proteins using circular dichroism spectroscopy. *Protein Sci.* **8**, 370–380 (1999).
35. Sreerama, N., Venyaminov, S.Y. & Woody, R.W. Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis. *Anal. Biochem.* **287**, 243–251 (2000).
36. Sreerama, N. & Woody, R.W. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287**, 252–260 (2000).
37. Böhm, G., Muhr, R. & Jaenicke, R. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng.* **5**, 191–195 (1992).
38. Andrade, M.A., Chacón, P., Merelo, J.J. & Morán, F. Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng.* **6**, 383–390 (1993).
39. Unneberg, P., Merelo, J.J., Chacón, P. & Morán, F. SOMCD: method for evaluating protein secondary structure from UV circular dichroism spectra. *Proteins* **42**, 460–470 (2001).
40. Perczel, A., Park, K. & Fasman, G.D. Analysis of the circular dichroism spectrum of proteins using the convex constraint algorithm: a practical guide. *Anal. Biochem.* **203**, 83–93 (1992).
41. Chen, Y.H. & Yang, J.T. A new approach to the calculation of secondary structures of globular proteins by optical rotatory dispersion and circular dichroism. *Biochem. Biophys. Res. Commun.* **44**, 1285–1291 (1971).
42. Chen, Y.H., Yang, J.T. & Martinez, H.M. Determination of the secondary structures of proteins by circular dichroism and optical rotatory dispersion. *Biochemistry* **11**, 4120–4131 (1972).
43. Sreerama, N. & Woody, R.W. Poly(pro)II helices in globular proteins: identification and circular dichroic analysis. *Biochemistry* **33**, 10022–10025 (1994).
44. Perczel, A., Hollósi, M., Tusnády, G. & Fasman, G.D. Convex constraint analysis: a natural deconvolution of circular dichroism curves of proteins. *Protein Eng.* **4**, 669–679 (1991).
45. Greenfield, N.J. & Hitchcock-DeGregori, S.E. Conformational intermediates in the folding of a coiled-coil model peptide of the N-terminus of tropomyosin and α-tropomyosin. *Protein Sci.* **2**, 1263–1273 (1993).
46. Safar, J., Roller, P.P., Gajdusek, D.C. & Gibbs, C.J. Jr. Thermal stability and conformational transitions of scrapie amyloid (prion) protein correlate with infectivity. *Protein Sci.* **2**, 2206–2216 (1993).
47. Greenfield, N.J. Using circular dichroism, collected as a function of temperature, to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* doi: 10.1038/nprot.2006.204 (2006).
48. Adler, A.J., Greenfield, N.J. & Fasman, G.D. Circular dichroism and optical rotatory dispersion of proteins and polypeptides. *Methods Enzymol.* **27**, 675–735 (1973).
49. Rosenkranz, H. Circular dichroism of helical and nonhelical proteins. A review of the limits of the methods for calculating secondary structure. *Z. Klin. Chem. Klin. Biochem.* **12**, 222 (1974).
50. Johnson, W.C. Jr. Protein secondary structure and circular dichroism: a practical guide. *Proteins* **7**, 205–214 (1990).
51. Woody, R.W. Circular dichroism. *Methods Enzymol.* **246**, 34–71 (1995).
52. Kelly, S.M., Jess, T.J. & Price, N.C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **1751**, 119–139 (2005).
53. Miles, A.J. & Wallace, B.A. Synchrotron radiation circular dichroism spectroscopy of proteins and applications in structural and functional genomics. *Chem. Soc. Rev.* **35**, 39–51 (2006).
54. Venyaminov, S.Y. & Yang,, J.T. in *Circular Dichroism and the Conformational Analysis of Biomolecules* (ed. Fasman, G.D.) (Plenum Press, New York and London, 1996).
55. Greenfield, N.J. Determination of the folding of proteins as a function of denaturants, osmolytes or ligands using circular dichroism. *Nat. Protoc.* doi: 10.1038/nprot.2006.229 (2006).
56. Greenfield, N.J. The use of circular dichroism to study the kinetics of protein folding and unfolding. *Nat. Protoc.* doi: 10.1038/nprot.2006.244 (2006).
57. Miles, A.J., Whitmore, L. & Wallace, B.A. Spectral magnitude effects on the analyses of secondary structure from circular dichroism spectroscopic data. *Protein Sci.* **14**, 368–374 (2005).
58. Edelhoch, H. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6**, 1948–1954 (1967).
59. Mihalyi, E. Numerical values of the absorbances of the aromatic amino acids in acid, neutral, and alkaline solutions. *J. Chem. Eng. Data* **13**, 179–182 (1968).
60. Goa, J. A micro biuret method for protein determination. *Scand. J. Clin. Lab. Invest.* **5**, 218–222 (1953).
61. Konno, T. Conformational diversity of acid-denatured cytochrome *c* studied by a matrix analysis of far-UV CD spectra. *Protein Sci.* **7**, 975–982 (1998).
62. Melberg, S.G. & Johnson, W.C. Jr. Changes in secondary structure follow the dissociation of human insulin hexamers: a circular dichroism study. *Proteins* **8**, 280–286 (1990).
63. Savitsky, A. & Golay, M.J.E. Smoothing and dIfferentiation of data by simplified least squares procedures. *Anal. Chem.* **36**, 1627–1639 (1964).
64. Bentz, H., Bächinger, H.P., Glanville, R. & Kühn, K. Physical evidence for the assembly of A and B chains of human placental collagen in a single triple helix. *Eur. J. Biochem.* **92**, 563–567 (1978).
65. Steiner, R.F. Structural transitions of lysozyme. *Biochim. Biophys. Acta* **79**, 51–63 (1964).