PROTEINS: Structure, Function, and Genetics 7:205–214 (1990)

Review Article

Protein Secondary Structure and Circular Dichroism: A Practical Guide

W. Curtis Johnson, Jr.
Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331-6503

Key words: protein secondary structure, circular dichroism

INTRODUCTION

This journal is testimony to the recent explosive interest in proteins. Now that we have learned how to manipulate genes, it is only natural to use these techniques to engineer gene products. Natural proteins are more easily isolated when they are overexpressed; site-directed mutagenesis allows selective changes to be made in the sequence of natural proteins. Furthermore, proteins can now be designed from scratch both by isolating products of synthetic genes and by synthesizing polypeptides with the desired sequence. These advances for producing proteins of all kinds have concomitantly produced a resurgence of interest in circular dichroism (CD), a technique which is sensitive to secondary structure, and which can be applied to proteins and polypeptides in solution (Fig. 1).

How reliable are secondary structures predicted from CD? Some modern methods have been shown to have high reliability. A method utilizing the statistical procedure called "variable selection" boasts correlation coefficients of 0.97 for α-helix, 0.76 for β-sheet, 0.49 for β-turn, and 0.86 for other structures when the predicted secondary structure for CD data measured to 178 nm is compared to X-ray structure for the sixteen proteins in the data base.[1]

To get reliable results it is important to understand the limitations of the technique. CD spectra of proteins look complicated, but the long wavelength bands are dominated by the CD due to any α-helical structure. Thus, the information content of a protein CD spectrum is lower than one might guess, the equivalent of two equations for spectra truncated at 200 nm or three to four equations for spectra truncated at 190 nm.[2,3] Nevertheless, some workers who use our methods carry out analyses on spectra truncated at 190 or even 200 nm. As we will show in the next section, CD data can give statistically significant estimates of multiple classes of secondary structure *only if the data extend to 184 nm.* If data extending only to 200 nm are used, α-helix is generally the only secondary structure that can be determined with confidence. The situation is not much better if data extending to 190 nm are used.

The obvious problem is in obtaining CD spectra of proteins measured to at least 184 nm. Fortunately, any commercial CD machine that utilizes a stressed plate modulator as a quarter wave retarder can be modified so that it will measure CD spectra to 180 nm. Our Jasco J-40 (Jasco Inc., Easton, MD) routinely measures CD spectra to 180 nm. A colleague who recently spent his sabbatical year in my laboratory now uses his Jobin Yvon Model V (Instruments SA Inc., Metuchen, NJ) to 180 nm. AVIV and associates (AVIV, Lakewood, NJ) have modified Cary spectrophotometers so that they measure CD spectra to 180 nm. A discussion of CD instrumentation and the procedures that make it possible to measure spectra into the vacuum UV comprise the second part of this mini-review.

With the CD spectrum measured to 180 nm in hand, there are a number of methods for its analysis. These are discussed in the third part of the review. Finally, some of the recent CD analyses that have come out of our laboratory will be discussed to illustrate the various applications of CD spectroscopy to determining protein secondary structure. More comprehensive reviews of CD spectroscopy are available in refs. 4–6.

INFORMATION CONTENT IN THE CIRCULAR DICHROISM OF PROTEINS

The CD of proteins is primarily the CD of the amide chromophore; secondary structure as measured by CD counts amide–amide interactions, a somewhat different number from counting residues in X-ray diffraction structures. This amide chromophore begins absorbing far into the UV region with the first band at about 220 nm. Figure 2 gives CD spectra of polypeptides that are presumed to assume a single type of secondary structure in solution.[7] We see that at most each of these structures has two CD bands between 240 and 200 nm.

Received August 24, 1989; revision accepted October 31, 1989.

Address reprint requests to W. Curtis Johnson, Jr., Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR 97331-6503.

© 1990 WILEY-LISS, INC.

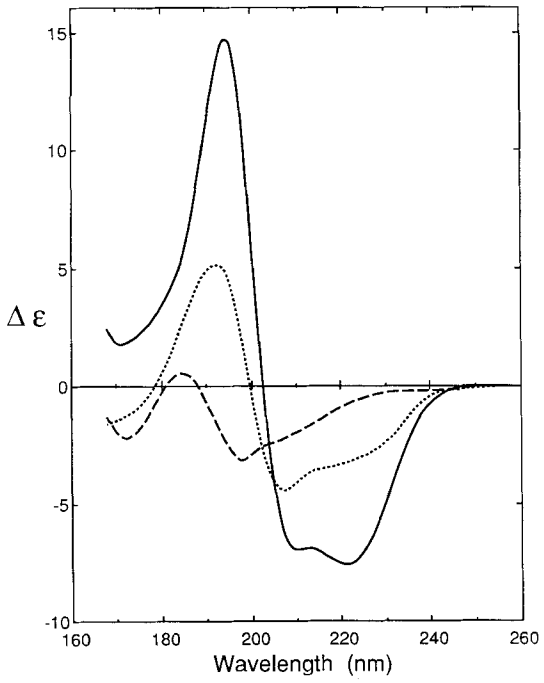

Fig. 1. The CD of three representative proteins: hemoglobin (——), elastase (– – –), and lysozyme (····), redrawn from Ref. 2.

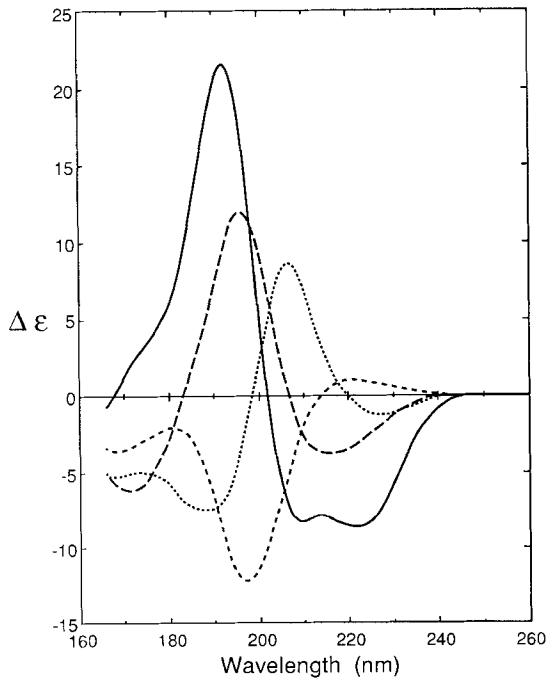

Fig. 2. The CD for various secondary structures: α-helix (——), antiparallel β-sheet (——), β-turn (····), and random coil (– –), redrawn from Ref. 7.

With only two CD bands, we might guess that the information content of this region is equivalent to two equations. Furthermore, we note that the CD for an α-helix dominates this region with a reasonably intense negative band at about 222 nm, and another reasonably intense negative band at 208 nm.

Figure 2 demonstrates that extending the CD spectra to 178 nm doubles the information content, since there are clearly two more CD bands between 200 and 178 nm. Thus, we might expect CD spectra measured from 240 to 178 nm to have an information content of about four, equivalent to four equations. Still the CD of the α-helix dominates with an intense positive band at 192 nm, but the β-sheet also has a fairly intense band at 198 nm, and the random coil has an intense negative band at 198 nm.

Singular value decomposition (SVD) is a mathematical method[8] that is well suited for analytically determining information content. A CD spectrum can be digitized by recording its intensity at each wavelength, say every 0.5 nm. This is the way a spectrum is stored in a computer, and converts the analog graph into a digitized vector. We took the CD spectra for 16 reference proteins recorded between 260 and 178 nm and digitized these spectra to form 16 vectors.[1] Taken together, the vectors can be written as 16 columns in a matrix that describes all of the CD data. SVD will decompose the data matrix into a product of three matrices with special properties. One matrix will contain 16 new CD basis vectors that are all orthogonal. Another matrix contains just a few values that give the importance of each orthogonal CD basis vector in reconstructing the original protein CD spectra. The third matrix contains the coefficients used to reconstruct the protein CD spectra from the basis vectors. Only the most important basis vectors will be needed to reconstruct the original protein CD spectra within experimental error. The number of new CD spectra necessary for this reconstruction is the information content of the CD, the number of equations that can be used for determining secondary structure.

Basis CD spectra are created by weighting basis vectors according to their importance, and the five most important are shown in Figure 3. We note that the α-helix shape dominates the most important basis spectrum, as we might expect from our qualitative discussion of the polypeptide CD spectra in Figure 2. All of these basis spectra represent a mixture of secondary structures, just as protein CD spectra represent a mixture of secondary structures. The fifthmost important basis spectrum is beginning to resemble noise with a low intensity and many crossovers.

In one report,[3] we carried out SVD on the CD spectra of the 16 reference proteins truncated at 200, 190, and 178 nm. Data were presented for three representative proteins, and the results for lactate dehydrogenase are shown in Figure 4. We see that the spectrum truncated at 200 nm is well fit by only two basis spectra, the spectrum truncated at 190 by three, and the spectrum truncated at 178 by four. The number of basis spectra necessary for a good fit

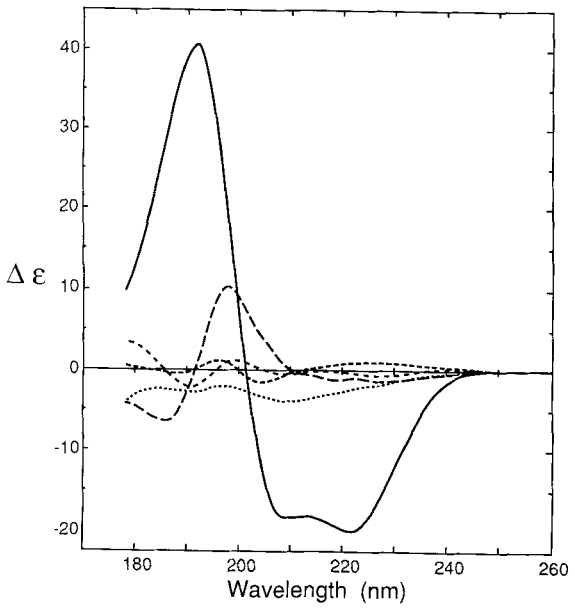

Fig. 3. The five most important basis CD spectra from singular value decomposition: most important (——), second (— —), third (····), fourth (– – –), and fifth (––), redrawn from Ref. 2.

to both shape and intensity within experimental error varies from protein to protein, but spectra truncated at 200 nm are always well fit by two, spectra truncated at 190 nm by three or four, and spectra truncated at 178 nm by four to six. CD spectra of proteins measured to 190 nm simply do not correspond to enough equations to reliably solve for the many parameters that determine their CD.

We also analyzed the CD spectra of the 16 reference proteins for secondary structure over various wavelength ranges using the five basis CD spectra. The results for lactate dehydrogenase are compared with X-ray diffraction data in Table I. We see that data truncated at both 178 and 184 nm analyzed to give secondary structures in agreement with X-ray diffraction. Truncating the data at 190 nm gives poor results including a fairly large and negative amount of antiparallel β-sheet. Truncating the CD data at 200 nm actually improves the analysis somewhat, although this is accidental. However, regardless of the wavelength range, the amount of α-helix is predicted correctly. The agreement in secondary structure measured by X-ray diffraction and predicted from CD spectra measured to 178 nm is not always as good as shown here for lactate dehydrogenase, but the trend with truncation is always the same. Predicted structures for data truncated at 178 and 184 nm agree closely. Data truncated at 190 and 200 nm give poor results. The analysis for α-helix is good regardless of the wavelength range, probably because the CD for an α-helix dominates a spectrum.

It is possible to add one equation to the number of equations inherent in the data by noting that the sum of secondary structures should be 100%. Table II compares the use of constraints in analyzing the CD spectrum of lactate dehydrogenase truncated at 190 nm when four basis spectra are used for the analysis. The unconstrained analysis is poor, but forcing the sum of secondary structures to be 100% does not improve things; the amount of antiparallel β-sheet is largely negative. When the constraint is used in conjunction with forcing all fractions of secondary structure to be positive, even the predicted amount of α-helix is poor. The same trends were found for all 16 reference proteins. In general, using either type of constraint did not improve the analysis, and additionally destroyed the prediction of α-helix, which we would expect to be good. Constraining the secondary structures to sum to 100% and requiring them to be positive is particularly dangerous. If you do not know the structure of the protein a priori, then you will be fooled because the results will always look good.

MEASURING A PROTEIN CD SPECTRUM TO 180 NM

The CD spectra of polypeptides measured by pioneers Holzwarth, Gratzer, and Doty[9] extended to 190 nm. That was over 25 years ago, and today's CD instrumentation is significantly improved. It is not easy, but by using the same care as these early workers it is now possible to measure the CD spectra of proteins below 180 nm on commercial instrumentation.

It is usually not necessary to buy a new CD spectrometer. All Jasco and Jovin Yvon instruments that are equipped with a stress-plate modulator are capable of making measurements into the vacuum UV. Cary spectrometers refurbished by AVIV Associates must be equipped with a magnesium fluoride polarizer as well as the stress-plate modulator to achieve short wavelength measurements. The real key is that, like any other research equipment (HPLCs, computers, NMR machines, etc.), these instruments must be well maintained.

On a daily basis, this means flushing the instrument with pure dry nitrogen regardless of the wavelength range being scanned. If your instrument dies at 200 nm, chances are someone neglected to purge the instrument while they were recording a spectrum. The high-intensity light source converts oxygen into ozone, which damages the optics rapidly. As a matter of routine, workers should purge the instrument for 15 minutes before turning on the source and while making measurements. We use the boil-off from a high-pressure liquid nitrogen tank at a flow rate of about 16 liters/minute. However, impurities concentrate in the liquid as the nitrogen evaporates, so it is important not to use the last 10% of the liquid nitrogen for purging.

CD measurements require a bright light source, so it is important to use source tubes made of far-UV

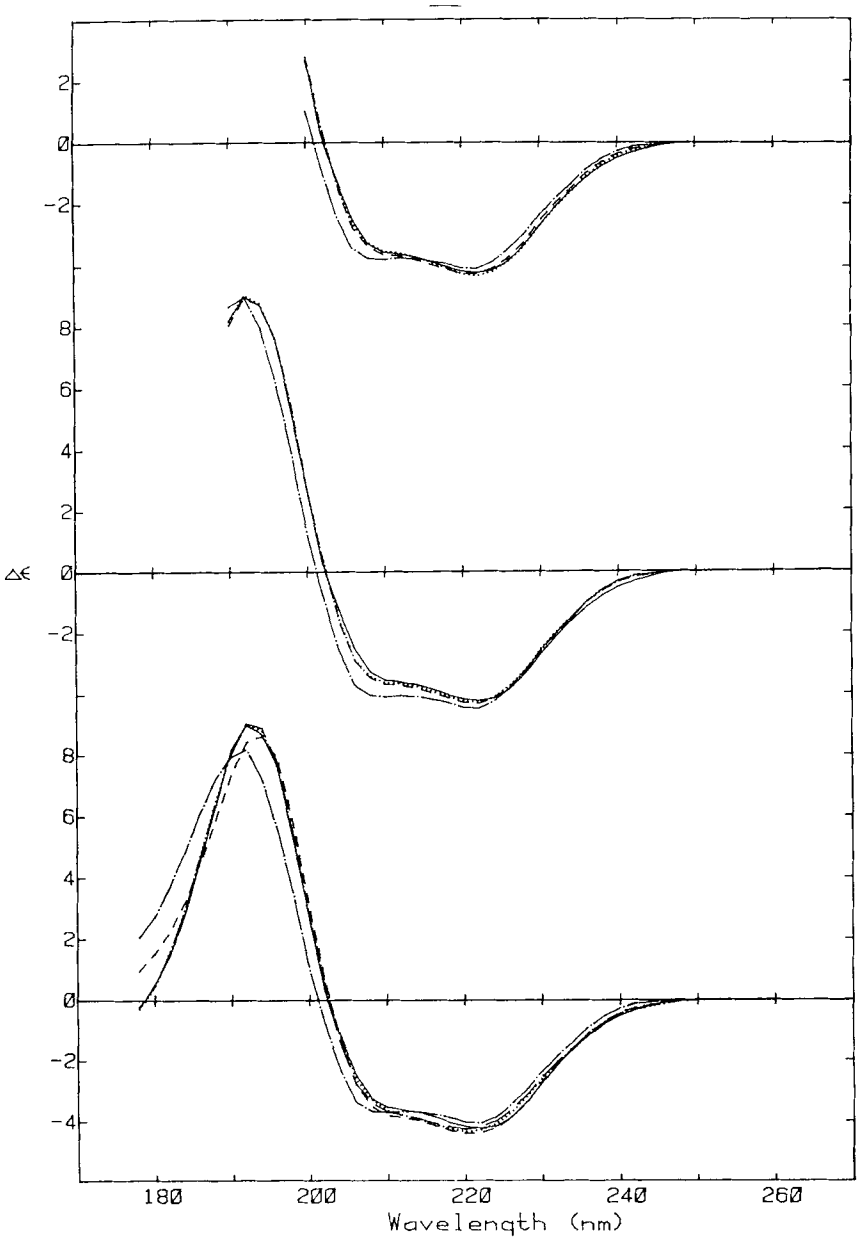

Fig. 4. Reconstruction of the CD spectrum for lactate dehydrogenase with various numbers of basis spectra. Top: measured (——), one basis (—·—·—), two bases (——), and three bases (·····); middle: measured (——), one basis (—·—·—), two bases (---), and three bases (·····); bottom: measured (——), one basis (—·—·—), two bases (– – –), four bases (——), and five bases (·····); redrawn from Ref. 3.

TABLE I. Secondary Structure of Lactate Dehydrogenase Predicted From CD Over Various Wavelength Ranges Using Five Basis CD Spectra

Wavelength range (nm)	H	A	P	T	O	Total
200–260	0.42	−0.03	0.07	0.11	0.24	0.81
190–260	0.41	−0.09	0.05	0.05	0.12	0.55
184–260	0.40	0.04	0.13	0.14	0.30	1.01
178–260	0.40	0.06	0.13	0.13	0.29	1.01
X-Ray	0.41	0.06	0.11	0.11	0.31	1.00

TABLE II. Secondary Structure of Lactate Dehydrogenase Predicted From CD Over the 260–190 nm Range Using Four Basis CD Spectra Both With and Without a Constraint for the Total

H	A	P	T	O	Total	
0.41	−0.06	0.06	0.04	0.13	0.58	Unconstrained
0.48	−0.14	0.14	0.06	0.45	1.00	Total constrained
0.66	0.00	0.06	0.09	0.19	1.00	Total constrained, no negatives
0.41	0.06	0.11	0.11	0.31	1.00	X-Ray

quartz for measurements into the vacuum UV. The optics should be cleaned and damaged mirrors replaced about every 2 years. If the instrument has been treated well, probably only the first mirror after the source that receives the high-intensity undispersed light need be replaced to restore the instrument to mint condition.

CD instruments can't make accurate measurements on samples with an absorbance (A) greater than 1.0. Thus, it is mandatory to measure the total absorbance of your cell, solvent, and sample before investing the time in measuring the CD. Solvent transparency is usually the central problem, and this can be improved substantially by going to shorter pathlength cells. We routinely use cylindrical quartz cells with a 0.05 mm pathlength, available from Helma (P.O. Box 544, Borough Hall Station, Jamaica, NY 14424). This short pathlength cell cuts solvent absorbance by a factor of 200 over the commonly used 1 cm cell, and extends the transparency of water to 176 nm. Table III lists the wavelength cut off for a variety of solvents. Of course, the transparency of your sample also depends on what you have added to the solvent. The sample itself will usually have a maximum optical density of about 0.3. For proteins, this means a concentration of about 1 mg ml^{-1} for the 0.05 mm cell, but only 100 μl of solution is needed since capillary action keeps the sample between the windows. Buffers and other added chemicals can also reduce transparency. We like to use 10 mM phosphate as a buffer, but perchlorate, Tris, and borate are also reasonably transparent. Dithiothreitol or 2-mercaptoethanol can be added to solutions at a concentration of 1 mM, and EDTA at 0.1 mM. Detergents 3-(*N*-morpholino)propanesulfonic acid (MOPS) at 15 mM, lubrol PC at 0.02%, and sodium dodecyl sulfate (SDS) at any reasonable concentration are fairly transparent.

Even with a well-tuned instrument, a short pathlength cell, and a transparent solvent, extensive data collection will be required to reach 180 nm. On an analog instrument, spectra of proteins can usually be scanned from 260 to 200 nm with a 16 second time constant and a scan speed of about 1 nm min^{-1}. The time constant is increased to 64 seconds for the 200 to 180 nm region, and the scan speed reduced to

TABLE III. Solvent Cut-Off ($A = 1.0$) for Two Cell Pathlengths

Compound	1.0 mm	0.05 mm
H_2O	182	176
F_6iPrOH	174.5	163
F_3EtOH	179.5	170
MeOH	195.5	184
EtOH	196	186
MeCN	185	175
Dioxane	231	202.5
Cyclohexane	180	175
n-Pentane	172	168

0.2 nm min^{-1}. In all, it would take about 2.5 hours to scan the spectrum and another 2.5 hours to scan the baseline. Computerized instruments should collect data for an equivalent time period. Because CD instruments are plagued by slow drift, it would actually be better to scan the spectrum four times at four times the speed and one-fourth the time constant, with baseline collection interspersed with sample collection.

In order to collect accurate CD spectra, your instrument needs to be calibrated. We calibrate ours daily by using an aqueous solution of (+)-10-camphorsulfonic acid (CSA) from Aldrich Chemical Company at a concentration of 1 mg ml^{-1} in a 1 mm cell. This compound has a $\Delta\epsilon$ of 2.36 at its CD maximum of 290.5 nm, or a ΔA of 1.02×10^{-3} (an ellipticity of 33.5 millideg) for our sample.[10] CSA is particularly hydroscopic, but this problem can be avoided by measuring the concentration through absorption spectroscopy. A solution of 1 mg ml^{-1} will give an A of 0.743 at 285 nm in a 5 cm cell, since the extinction coefficient for the anhydrous material is 34.5 at this absorption maximum. CSA also has a CD minimum at 192.5 nm with a $\Delta\epsilon$ of about −4.9. The ratio for the absolute intensities of these peaks tells you how well your instrument is working at short wavelengths. It should be 2.0 or greater. A ratio of less than 2.0 usually means that either the quarter wave retarder is out of adjustment at short wavelength, or that the amount of light emerging from the monochromator at short wavelength is insufficient to overwhelm the scattered light so that a new source tube or some new mirrors are needed.

Accurate analysis of CD for secondary structure demands accurate concentrations. The concentration of protein samples to be used for CD measurements is most easily determined from the absorption spectroscopy used to monitor the absorbance of the sample. To do this, it is necessary to know the extinction coefficient for the amide absorption which occurs at about 190 nm, and which for most proteins ranges between 8,500 and 11,400 liter mol^{-1} cm^{-1} on a per residue basis (2). Fortunately, Elwell and Schellman[11] have worked out a fairly simple method for determining protein concentration, although it does require considerably more material than is needed once the amide extinction coefficient at 190 nm has been determined. For a fully unfolded protein in 6 M guanidinium chloride, the total extinction coefficient at 280 nm will be due to the tyrosines and tryptophans, and will be the sum of the contributions for each tryptophan (5690 liter mol^{-1} cm^{-1}), and each tyrosine (1280 liter mol^{-1} cm^{-1}). Quantitative amino acid analysis is an alternative method of determining concentration, and will work on proteins that contain no tyrosines or tryptophans. Lowry analyses are often in error by as much as a factor of 2, and are not sufficiently accurate to be used with CD measurements.

ANALYZING CD SPECTRA OF PROTEINS FOR SECONDARY STRUCTURE

The direct way of analyzing the CD spectrum of a protein is to fit it with the CD spectra for the various secondary structures. This was done in the groundbreaking work of Greenfield and Fasman[12] and a number of laboratories have subsequently adopted this method. CD spectra are measured for polypeptides that are presumed to assume a single secondary structure in solution, giving results like those found in Figure 2. These spectra are then fit to the CD spectrum of the protein being analyzed, and the fraction of each secondary structure (in terms of amide chromophores) is taken to be the fraction of the corresponding polypeptide CD spectrum that is contained in the protein spectrum.

However, there are a number of problems with this idea. First, the CD spectrum of an α-helical polypeptide is for the infinitely long structure, and only short α-helices are found in proteins. Second, β-sheets come in both antiparallel and parallel varieties, and it is usually not known which variety the polypeptide assumes in solution, nor is there overwhelming proof that the polypeptide assumes the β-sheet. Further, there are undoubtedly β-turns mixed in with any β-sheet structure. While the CD spectra for β-sheets found in the literature all have features in common, they all have substantial differences in intensity, peak positions, and crossovers. Third, there are various types of β-turns, and their CD spectra are also not well documented. Fourth, there is disagreement over what conditions produce a true random coil in solution, and the corresponding CD spectra, while they have some features in common, do differ in shape and intensity. Moreover, it is not clear that a random coil CD should be used for the residues in a protein that do not fall into the other secondary structure classifications, since they are not really "randomly" oriented. Finally, proteins have many features such as chromophoric side chains and distortions of secondary structure that can contribute to protein CD, but are not represented in polypeptide spectra.

An alternative is to use the CD spectra of proteins that have a known secondary structure to analyze the CD spectrum of a protein with unknown structure. Obviously, one cannot be sure that the known secondary structures, at this time obtained in terms of amides from X-ray diffraction data, are really the structures in solution. Nevertheless, proteins contain by definition all of the features that contribute to the CD spectrum of a protein, even if we do not recognize them directly. Thus, short lengths of α-helix, both parallel and antiparallel β-sheet, all types of β-turns, other residues that do not fit into these categories, chromophoric side chains, distortions of secondary structure, etc., all go into the analysis. Saxena and Wetlaufer[13] and Chen et al.[14] used this system early on for their analyses, and many other workers have adopted the CD spectra of proteins with known secondary structure as their basis for analysis. The problem with this alternative is that there are generally more proteins in the reference set than there is information content in the CD of the protein being analyzed. This means that the fit of the reference spectra to the CD can be unstable. In the most recent work, Provencher and Glöckner[15] have used a mathematical device called a regularizor, which penalizes fits that favor a particular protein in the set, thus stabilizing the solution. Our laboratory has made use of SVD to create basis CD spectra that are a mixture of the spectra of all the proteins in the reference set, but which are limited in number to be consistent with the information content of the data. For protein CD spectra measured to 178 nm, this means limiting the number of basis spectra used in the analysis to five.

How can a CD spectrum with an information content of five be reliably analyzed when so many protein features contribute to the spectrum? Actually, it cannot, but there are mitigating circumstances. First, the protein features that contribute to the CD may be related, so that the number of independent variables may be fewer than we think. SVD can be used to investigate this relationship, because it is a general mathematical method that is applicable to any data matrix. Hennessey and Johnson[2] applied SVD to the matrix of secondary structures corresponding to the 16 proteins in their reference set. When the structures are divided as α-helix, antiparallel β-sheet, parallel β-sheet, type I β-turn, type II

β-turn, type III β-turn, other turns, and other structures, the information content is only about four. Since β-turns of types I, II, and III predominate, this means that it is reasonable to analyze for β-turns even though there are many different types. Moreover, there is one independent variable left over in spectra measured to 178 nm that can be used to help account for other features that contribute to the CD spectrum.

Even so, there are undoubtedly more than five independent features in a protein that contribute to its CD spectrum. That is where the statistical procedure called a variable selection comes in. According to variable selection, proteins in the basis set that have features contributing to the CD that are not found in the protein being analyzed should be eliminated so that these variables do not have to be considered. Of course, we do not know a priori which proteins should be eliminated, but we can systematically look at the results of eliminating proteins. Because the number of combinations rise quickly, we look at all combinations of eliminating three proteins at a time, and choose the best truncated basis sets by using the following criteria: (1) the sum of secondary structures should be between 0.9 and 1.1, or if possible between 0.96 and 1.05; (2) secondary structures are expected to be positive, but should never be less than −0.05; (3) the fit of the reconstructed CD should be improved by truncating the basis set, and in the final analysis should be within the noise level of experimental data, about 0.22 $\Delta\epsilon$ units for the root mean square residual; (4) truncated basis sets should have as many reference proteins as possible. Proteins are eliminated until these criteria are met, and usually similar results are obtained for many subsets. The predicted secondary structure for the protein is an average of these subsets. When variable selection was carried out on the 16 reference proteins from Hennessey and Johnson using the other 15 proteins as the basis set, correlation coefficients were 0.97 for H, 0.78 for A, 0.67 for P, 0.49 for T, and 0.86 for O. The low correlation coefficient for T is not as bad as it seems. The absolute error in predicting T is small, but the variation among T for the various proteins is also small, so the correlation coefficient is low.

To be on the cutting edge in the analysis of their CD data, researchers will want to make use of the most modern methods of analysis. From our laboratory, that means using variable selection, which will yield superior results to the original Hennessey and Johnson method alone. An interim paper[16] that permitted analysis of protein CD spectra by using a simple matrix multiplication merely put the Hennessey and Johnson method in an easy-to-use format, and when correctly used yields results that are identical to the earlier method. Although workers are encouraged to write their own code for these methods, the program that we use in our laboratory for variable selection, and the digitized CD for the 16 reference proteins are available upon request on a disk suitable for an IBM-type computer.

Provencher and Glöckner[15] have also developed a viable method of analyzing the CD spectra of proteins for secondary structure. The regularizor in their procedure provides the same type of flexibility that we gained through variable selection. Although their method was developed by using reference protein spectra measured to 190 nm, it can certainly be used with reference spectra, such as ours, that have been measured further into the vacuum UV. CONTIN is the computer program written by Provencher for employing this method. The primary source is the CPC Program Library, Department of Applied Mathematics and Theoretical Physics, Queens University of Belfast, BT71NN, Northern Ireland. Nonprofit institutions only may request the program from the author: Stephen Provencher, European Molecular Biology Laboratory, Postfach 10.2209, D-6900 Heidelberg, Federal Republic of Germany. CONTIN calculates a number of solutions, which are a function of the strength of the regularizor. The program then uses several criteria to choose one of these solutions, but we have found that the "chosen solution" works well only for a few all-α proteins. We recommend choosing a solution with about five degrees of freedom, so that it is consistent with the information content of the data measured to 178 nm. When applied to the CD spectra for the 16 reference proteins, using the 15 other proteins as a basis, we obtain correlation coefficients of 0.98 for H, 0.63 for A, 0.56 for P, 0.65 for T, and 0.83 for O.

ILLUSTRATIVE PROTEIN STUDIES USING CD

This section of the mini-review contains the results of a few recent CD studies from our laboratory that serve to illustrate how CD spectroscopy can be used to solve structural problems for proteins. Figure 5 shows the CD spectrum of *Eco*RI endonuclease,[17] which was measured and published before the X-ray structure.[18] This was also before variable selection, and the Hennessey and Johnson method gave a good analysis, predicting 33% H, 20% A, 5% P, 17% T, and 25% O in terms of backbone amides. Prediction of secondary structure from the primary sequence gave α-helical segments alternating with β-strand segments, an indication of an α/β tertiary structure with the helices going in one direction and β-sheet in the returning direction, which would be expected to give a great deal of parallel β-sheet. Nevertheless, the CD analysis agrees well in all respects with the X-ray data, which shows β-sheet that is mostly antiparallel (26% H, 20% A, 8% P, 25% T, 21% O in terms of amide–amide interactions).

The CD spectrum of calmodulin with 4 mol of Ca^{2+} per mol of protein, shown in Figure 5, was

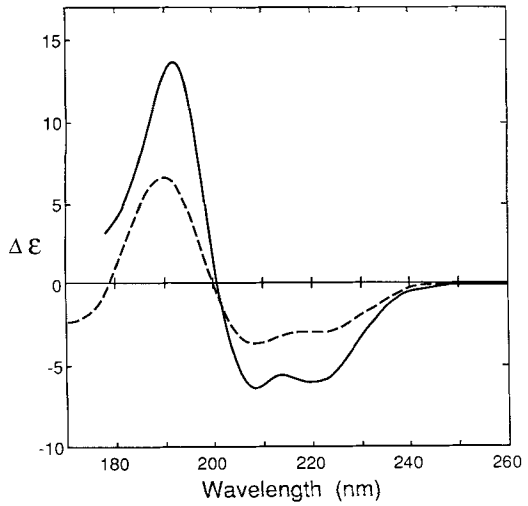

Fig. 5. The CD of calmodulin (——), redrawn from Ref. 19, and *Eco*RI endonuclease (– – –), redrawn from Ref. 17.

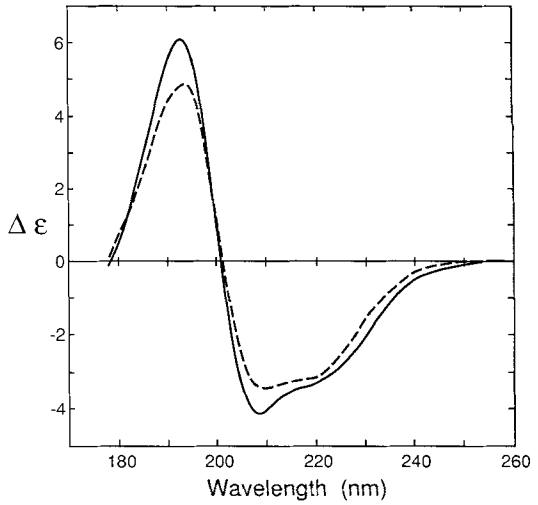

Fig. 6. The CD of thymidylate synthetase (——) and thymidylate synthetase complex with FdUMP and 5,10-methylenetetrahydrofolate (– – –), redrawn from Ref. 21.

measured in 1978. However, work on this protein was not published; the initial analysis was not entirely successful using the Hennessey and Johnson method, which gave values of 50% H, 20% A, −11% P, 21% T, and −2% O. However, variable selection was successful in analyzing calmodulin CD spectra,[19] with values of 61% H, 2% A, 2% P, 21% T, and 14% O, which is in close agreement with the X-ray structure[20] (59% H, 3% A, 0% P, 41% T and O combined, in terms of amide–amide interactions).

CD can be used to follow changes in secondary structure with the binding of ligands,[21] as shown in Figure 6. Here the CD of thymidylate synthase loses intensity and changes shape as the protein binds FdUMP and the coenzyme 5,10-methylenetetrahydrofolate. Binding of either FdUMP or the coenzyme alone does not change the CD spectrum, an interesting finding. Using variable selection, the CD of thymidylate synthase analyzed with 33% H, 24% A, 2% P, 21% T, and 20% O is in good agreement with subsequently published X-ray diffraction work at 3 Å resolution[22] (33% H, 11% A, 3% P, 53% T and O combined, in terms of amide–amide interactions). Analysis of the CD for thymidylate synthase binding the ligand and coenzyme indicates a loss of 5% A, and 6% T, with a concomitant gain of 8% O.

Figure 7 shows that the formation of a protein quaternary structure can result in a change in secondary structure.[23] Here the intact $12S_H$ subunit of transcarboxylase loses intensity and changes shape as it dissociates into the six component $2.5S_H$ monomers. The monomers analyze to give 19% H, 23% A, 0% P, 19% T, and 32% O. The CD measurements indicate that formation of the quaternary structure increases α-helix by 19%, and decreases A by 8% and O by 7%.

Random coil polypeptides have a characteristic

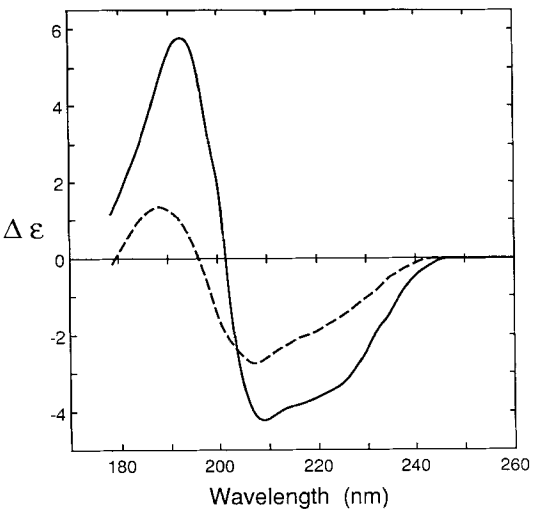

Fig. 7. The CD of the hexameric $12S_H$ subunit of transcarboxylase at pH 5.8 (——) and dissociated into its component monomers at pH 9.0 (– – –), redrawn from Ref. 23.

CD spectrum that is illustrated for one case in Figure 2. Analysis of this type of spectrum by our methods gives little H and P, but a fair amount of A and T, as well as the expected O. Clearly, random coil polypeptides do not have any stable antiparallel β-sheet or β-turns as such, but the φ–ψ angles associated with these two structures are favored by the dynamic structure as is clear in NMR COSY spectra. CD cannot tell the difference between dynamic and static structures, so that other methods would have to be used for that distinction. For instance, when studying oligopeptides in solution, melting studies should be carried out to monitor cooperativity and decide between a dynamic or static structure. The

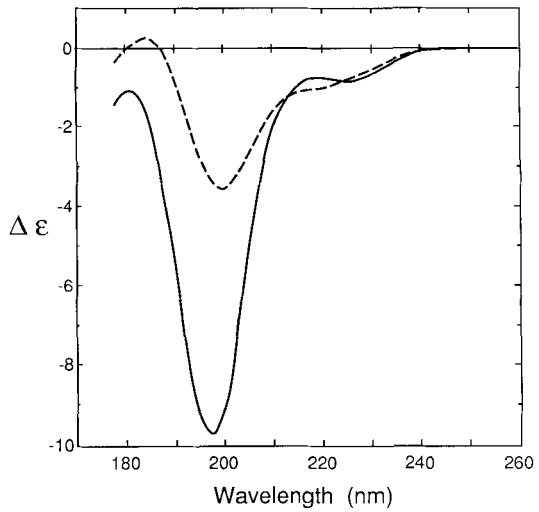

Fig. 8. The CD of the protamine-like protein from *Spisula solidissima* (——) and its trypsin-resistant core (– – –), redrawn from Ref. 24.

protamine-like protein from *Spisula solidissima* and its trypsin-resistant core both have CD spectra reminiscent of a random polypeptide,[24] as shown in Figure 8. Nevertheless, hydrodynamic measurements demonstrate that both have a viable tertiary structure. Variable selection predicts 10% H, 33% A, 0% P, 18% T, and 37% O for the compact core. The more intense CD for the whole protein analyzes to predict that the tail is primarily β-turn and other structure.

CONCLUSION

Circular dichroism spectroscopy has greatly improved in recent years, and it is now possible to analyze the spectrum of a protein for secondary structure with high reliability. It is incumbent upon protein researchers using CD to make use of this modern technology, rather than to use older techniques that have been shown to be unsatisfactory. When CD is used as part of a project, it deserves the same effort that was put into the rest of the research. This means measuring CD spectra to at least 180 nm so that it contains enough information to solve for the amounts of the various secondary structures. When the mathematics of the problem are taken into account, it is clear that spectra truncated at 190 nm or longer wavelengths cannot be analyzed reliably. If data truncated at 190 nm are analyzed, then one or another method of analysis may appear to give accurate results, but this is accidental. The favored method can change with each new protein.

All things considered, it would seem best to use reference spectra derived from the CD of proteins with known secondary structure for the analysis, rather than the CD of particular secondary structures, since proteins will contain all of the factors that contribute to the CD. Two methods have been developed that make use of protein CD spectra as references, and also take into account the instability generated by such an overdetermined reference set, one method in our laboratory and another by Provencher and Glöckner. Using all refinements, both methods can be expected to give reliable, and of course, similar results.

ACKNOWLEDGMENTS

This work was supported by NIH Grant GM-21479.

REFERENCES

1. 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.
2. Hennessey, J.P., Jr., Johnson, W.C., Jr. Information content in the circular dichroism of proteins. Biochemistry 20:1085–1094, 1981.
3. Manavalan, P., Johnson, W.C., Jr. Protein secondary structure from circular dichroism spectra. Suppl. J. Biosci. 8:141–149, 1985.
4. Woody, R.W. Circular dichroism of peptides. In: "The Peptides" (Blout, E.R., Bovey, F.A., Goodman, M., Lotan, N., eds.), Vol. 7. New York: Academic Press, 1985: 15–114.
5. Johnson, W.C., Jr. Circular dichroism and its empirical application to biopolymers. In: "Methods of Biochemical Analysis" (Glick, D., ed.), Vol. 31. New York: Wiley, 1985: 61–163.
6. Johnson, W.C., Jr. Secondary structure of proteins through circular dichroism spectroscopy. Annu. Rev. Biophys. Biophys. Chem. 17:145–166, 1988.
7. Brahms, S., Brahms, J. Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. J. Mol. Biol. 138:149–178, 1980.
8. Noble, B., Daniel, J.W. "Applied Linear Algebra," 2nd ed. Englewood Cliffs, NJ: Prentice-Hall. 323–342, 1977.
9. Holzwarth, G., Gratzer, W.B., Doty, P. The optical activity of polypeptides in the far ultraviolet. J. Am. Chem. Soc. 84:3194–3196, 1962.
10. Chen, G.C., Yang, J.T. Two point calibration of circular dichrometer with d-10-camphorsulfonic acid. Anal. Lett. 10:1195–1207, 1977.
11. Elwell, M.L., Schellman, J.A. Native properties and thermal stability of wild type and two mutant lysozmes. Biochim. Biophys. Acta 494:367–383, 1977.
12. Greenfield, N., Fasman, G.D. Computed circular dichroism spectra for the evaluation of protein conformation. Biochemistry 8:4108–4116, 1969.
13. Saxena, V.P., Wetlaufer, D.B. A new basis for interpreting the circular dichroic spectra of proteins. Proc. Natl. Acad. Sci. U.S.A. 68:969–972, 1971.
14. 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.
15. Provencher, S.W., Glöckner, J. Estimation of globular protein secondary structure from circular dichroism. Biochemistry 20:33–37, 1981.
16. Compton, L.A., Johnson, W.C., Jr. Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. Anal. Biochem. 155: 155–167, 1986.
17. Manavalan, P., Johnson, W.C., Jr., Modrich, P. Prediction of secondary structure for EcoRI endonuclease. J. Biol. Chem. 259:11666–11667, 1984.
18. McClarin, J.A., Frederick, C.A., Wang, B-C., Greene, P., Boyer, H.W., Grable, J., Rosenberg, J.M. Structure of the DNA-Eco RI endonuclease recognition complex at 3 Å resolution. Science 234:1526–1541, 1986.
19. Hennessey, J.P., Jr., Manavalan, P., Johnson, W.C., Jr., Malencik, D.A., Anderson, S.R., Schimerlik, M.I., Shalitin, Y. Conformational transitions of calmodulin as studied by vacuum-UV CD. Biopolymers 26:561–571, 1987.

20. Babu, Y.S., Bugg, C.E., Cook, W.J. Structure of calmodulin refined at 2.2Å resolution. J. Mol. Biol. 204:191–201, 1988.
21. Manavalan, P., Mittelstaedt, D.M., Schimerlik, M.I., Johnson, W.C., Jr. Conformational analysis of thymidylate synthase from amino acid sequence and circular dichroism. Biochemistry 25:6650–6655, 1986.
22. Hardy, L.W., Finer-Moore, J.S., Montfort, W.R., Jones, M.O., Santi, D.V., Stroud, R.M. Atomic structure of thymidylate synthase: Target for rational drug design. Science 235:448–455, 1987.
23. Hennessey, J.P., Jr., Johnson, W.C., Jr., Bahler, C., Wood, H.G. Subunit interactions of transcarboxylase as studied by circular dichroism. Biochemistry 21:642–646, 1982.
24. Ausio, J., Toumadje, A., McParland, R., Becker, R.R., Johnson, W.C., Jr., van Holde, K.E. Structural characterization of the trypsin-resistant core in the nuclear sperm-specific protein from *Spisula solidissima*. Biochemistry 26: 975–982, 1987.