Published in final edited form as:

Anal Biochem. 2008 April 15; 375(2): 379–381.

Concentration Independent Estimation of Protein Secondary Structure by Circular Dichroism; A Comparison of Methods

Peter McPhie

Laboratory of Biochemistry and Genetics, NIDDK, NIH, Building 8, Room 215, Bethesda, MD 20892
–0830 Phone 301–496–5890 Fax 301–402–0240 E-mail pmcphie@helix.nih.gov.

Abstract

Estimation of a protein's secondary structure from its circular dichroism spectrum usually requires accurate knowledge of the concentration and pathlength of the sample. Two recently described methods avoid this problem by analysis of g-factor spectra (McPhie, *Anal. Bioch.* **293**, 109–119) or scaling of relative intensities (Raussens et al, *ibid.* **319**, 114–121). Application of the two methods to the same samples shows that they can have similar efficacies. Calculation with the latter method is more rapid, but the performance of the former is maintained over reduced wavelength ranges.

Circular dichroism spectroscopy (CD) is the most widespread technique for the estimation of the secondary structure of proteins, using sophisticated mathematical algorithms for the deconvolution of spectra (which must be converted to units of mean residue ellipticity) [1]. Recently, interest has returned to the role of experimental errors in these analyses [2,3]. If the spectropolarimeter is calibrated and operated in accordance with the manufacturer's specifications, then two major sources of error remain. The magnitude of the spectrum is affected by uncertainties in protein concentration and in cuvette path length. These errors are exacerbated if the concentration is measured in inaccurate ways [4] or if demountable, short path length cells or films are employed. The wavelength dependence of the spectrum may be shifted by misalignment of the instrument's monochromator, introducing systematic errors between the spectrum to be analyzed and the basis spectra in the reference database (which are usually measured in different laboratories).

Methods have been proposed which do not require knowledge of protein concentration or path length, potentially circumventing one of these problems. The first method follows the traditional approach, but analyzes the g-factor spectrum of the sample [5], (the ratio of its CD and absorbance spectra, which is an intensive property) with experimentally derived basis spectra characteristic of secondary structures. The second method is more pragmatic, using the relative intensities of the CD spectrum at four wavelengths (193, 196, 211 and 234nm) normalized against that at 207nm, in quadratic equations derived by numerical analysis of the 50 member RASP database of CD spectra of proteins of known structures [6a,b]. Since measurement of the CD spectrum is required for calculation of the g-factor spectrum, it is possible to compare these two methods with the same samples. Similar comparisons have shown that each of the algorithms normally used for analysis of CD spectra has its own strengths [1,7].

Carbonic anhydrase, chymotrypsinogen, ovalbumin and trypsinogen (purity >95%) were obtained from Sigma Chemical Co. (St. Louis, MO). Interferon 4A was a gift of Dr. Hana Schmeisser, NCI. Their CD and g-factor spectra were measured as before and combined with

those of 27 other globular proteins [5,8], to give an enlarged database (32gfact) used in the analyses reported below.

Each g-factor spectrum was analyzed for secondary structure by constrained least squares, with the four published basis spectra (g-LSQ) [5] and also with the CONTIN program (version of Dr. K.S. Vassilenko), using 28 other spectra as reference sets (g-CONTIN), in cross-validation tests. The corresponding CD spectra were scaled and analyzed using the revised equations of Raussens et al (CD-RRG) [6b]. As usual, the efficacies of the methods were compared by values of the correlation coefficient, r , and the root mean square deviation, δ , [7] between the estimated secondary structures and those derived from published X-ray structures in .PDB files, with the DSSP program [9]. The results are summarized in Table 1. Application of the two methods (CD-RRG and g-LSQ) to their parent databases, showed them to be very similar in their abilities to estimate secondary structures. However, CONTIN analysis of the full g-factor database showed a marked deterioration in performance. When the CD-RRG equations were applied to the CD spectra associated with the 32gfact database, they retained their ability to estimate α -helix, but their accuracy in estimation of other structures was reduced.

Use of the CD-RRG technique needs measurement of spectra down to 190nm. Physiological buffers or denaturants prevent measurements at short wavelengths. Truncation of g-factor spectra at wavelengths up to 210nm had little effect on estimates of helix and sheet content by least squares analysis. Surprisingly, improved results were obtained by analysis of such truncated spectra with g-CONTIN. The information in g-factor spectra is heavily weighted to long wavelengths [5,8], but further reductions in wavelength range degraded both the g-LSQ and g-CONTIN analyses (not shown). The g-factor value at one wavelength (230nm) is well correlated with helix content [5]. Pancoska et al. showed that there are strong correlations between the helix content of a protein and its fractions of other structural elements [10]. Thus, the possibility must be entertained that analyses of truncated spectra determine sheet content through these correlations.

Table 2 shows the results of artificial wavelength errors on the analyses. The measured CD and g-factor spectra of three representative proteins (HSA, all α , RNase, α/β , ConA, all β) were shifted by one nanometer to longer (+1) or shorter (-1) wavelengths before analysis by g-CONTIN or CD-RRG techniques. The consequences of these shifts are mostly similar to those reported earlier, with concentration dependent methods[2,3], but larger deviations occur when the CD-RRG equations are used with ConA spectra. CD intensities are normalized against the intensity at 207nm [6], a wavelength close to the 208nm negative maximum in the characteristic spectrum of the α -helix [1]. For proteins which contain this structure, small wavelength errors will have little effect on this normalization factor. However, for proteins which contain little α -helix, spectral slopes are very steep in this region [1] and wavelength errors may have more drastic consequences. Small wavelength deviations may explain the reduced performance of this technique on moving from the RASP to the 32gfact spectral data base (measured on different instruments).

This comparison indicates that under optimal conditions, these two techniques are equally useful for the analysis of CD spectra of proteins of unknown concentration and/or path length. The main advantage of the CD-RRG technique lies in its ease of calculation (approx. 5min). However spectra must be recorded over a wide wavelength range and for the best result increased attention must be given to synchronisation of wavelengths between instruments. Usually, little regard is given to this aspect of CD measurements [2,3]. Analysis of g-factor spectra is a computationally more intensive solution (approx. 15min), but can yield reasonable estimates of extended secondary structures using spectra measured over the reduced wavelength ranges enforced by many solvents.

Acknowledgements

I would like to thank Dr. Erik Goormaghtigh for correspondence about use of the CD-RRG analysis and a copy of the RASP database and Dr Jan Wolff and Dr. Allen Minton for critical reviews of the manuscript. This research was supported by the intramural research program, NIDDK, NIH..

Abbreviations

CD, circular dichroism; HSA, human serum albumin; RNase A, bovine pancreatic ribonuclease A; ConA, concanavalin A.

References

1. Greenfield NJ. Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. *Anal. Biochem* 1996;235:1–10. [PubMed: 8850540]
2. Hennessey JP Jr, Johnson WC Jr. Experimental errors and their effect on analyzing circular dichroism spectra of proteins. *Anal. Biochem* 1982;125:177–188. [PubMed: 7149253]
3. Miles AJ, Wien F, Lees JG, Wallace BA. Calibration and standardisation of synchrotron radiation and conventional circular dichroism spectrometers. Part 2: factors affecting magnitude and wavelength. *Spectroscopy* 2005;19:43–51.
4. Szollosi E, Hazy E, Szasz C, Tompa P. Large systematic errors compromise quantitation of intrinsically unstructured proteins. *Anal. Biochem* 2006;360:321–323. [PubMed: 17129571]
5. McPhie P. Circular dichroism studies on proteins in films and in solution. Estimation of secondary structure by g-factor analysis. *Anal. Biochem* 2001;293:109–119. [PubMed: 11373086]
6. a Raussens V, Ruyschaert J-M, Goormaghtigh E. Protein concentration is not an absolute prerequisite for the determination of secondary structure from circular dichroism spectra: a new scaling method. *Anal. Biochem* 2003;319:114–121. [PubMed: 12842114] b *Anal. Biochem* 2006;359:150.erratum in
7. Sreerama N, Woody RW. Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON and CDSSTR Methods with an expanded reference set. *Anal. Biochem* 2000;287:252–260. [PubMed: 11112271]
8. McPhie P. CD studies on films of amyloid proteins and polypeptides: quantitative g-factor analysis indicates a common folding motif. *Biopolymers* 2004;75:140–147. [PubMed: 15356868]
9. Kabsch W, Sander C. Dictionary of protein secondary structures: Pattern recognition of hydrogen bonded and geometrical features. *Biopolymers* 1983;22:2577–2637. [PubMed: 6667333]
10. Pancoska P, Bitto E, Janota V, Urbanova M, Gupta VP, Keiderling TA. Comparison of and limits of accuracy for statistical analyses of vibrational and electronic circular dichroism spectra in terms of correlations to and predictions of protein secondary structure. *Protein Science* 1995;4:1384–1401. [PubMed: 7670380]

Table 1
Comparison of Estimates of Secondary Structure from CD and g-factor Spectra

Method	Database	Wavelength Range	Helix		Sheet		Turn		r	Rem.	δ
			r	δ	r	δ	r	δ			
CD-RRG	RASP	240–190nm	0.84	0.10	0.77	0.09	0.29	0.01	0.45	0.05	
g-LSQ	32gfac	240–190nm	0.85	0.10	0.76	0.11	0.39	0.12	0.43	0.15	
g-CONTIN	32gfac	240–190nm	0.69	0.12	0.44	0.10	0.23	0.07	0.31	0.06	
CD-RRG	32gfac	240–190nm	0.85	0.09	0.47	0.12	0.12	0.01	0.43	0.04	
g-LSQ	32gfac	240–210nm	0.85	0.11	0.68	0.12	0.38	0.17	0.36	0.17	
g-CONTIN	32gfac	240–210nm	0.84	0.10	0.76	0.07	0.25	0.07	0.03	0.06	

The correlation coefficient (r) and rms deviation (δ) between CD and crystallographic estimates of structural elements are defined in [7].

Table 2
Effect of artificial wavelength shifts on estimated extended secondary structures for three representative proteins.

Protein (DSPP values)	Shift (nm)	helix	g-CONTIN sheet	helix	CD-RRG	sheet
HSA (61% helix, 0% sheet)	+1	57%	0%	40%		5%
	0	69%	0%	40%		3%
	-1	76%	0%	38%		3%
RNase A (21% helix, 33% sheet)	+1	28%	30%	17%		28%
	0	30%	29%	18%		28%
	-1	32%	28%	17%		26%
Con A (4% helix, 46% sheet)	+1	4%	46%	399%		49%
	0	10%	36%	-5%		58%
	-1	12%	30%	0%		47%