PROTEINS: Structure, Function, and Genetics 35:307–312 (1999)

Analyzing Protein Circular Dichroism Spectra for Accurate Secondary Structures

W. Curtis Johnson*
Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon

ABSTRACT **We have developed an algorithm to analyze the circular dichroism of proteins for secondary structure. Its hallmark is tremendous flexibility in creating the basis set, and it also combines the ideas of many previous workers. We also present a new basis set containing the CD spectra of 22 proteins with secondary structures from high quality X-ray diffraction data. High flexibility is obtained by doing the analysis with a variable selection basis set of only eight proteins. Many variable selection basis sets fail to give a good analysis, but good analyses can be selected without any a priori knowledge by using the following criteria: (1) the sum of secondary structures should be close to 1.0, (2) no fraction of secondary structure should be less than –0.03, (3) the reconstructed CD spectrum should fit the original CD spectrum with only a small error, and (4) the fraction of α-helix should be similar to that obtained using all the proteins in the basis set. This algorithm gives a root mean square error for the predicted secondary structure for the proteins in the basis set of 3.3% for α-helix, 2.6% for 3_{10}-helix, 4.2% for β-strand, 4.2% for β-turn, 2.7% for poly(L-proline) II type 3_1-helix, and 5.1% for other structures when compared with the X-ray structure. Proteins 1999;35:307–312.** © 1999 Wiley-Liss, Inc.

Key words: circular dichroism; secondary structure; proteins

INTRODUCTION

Over the years, many methods have been offered to analyze the circular dichroism (CD) of proteins in the amide region for their secondary structure.[1–21] As more ideas were presented, the accuracy of these analyses increased. A number of reviews have discussed this work,[22–27] and Greenfield[27] has recently reviewed the methods that workers are presently using to estimate the secondary structure of proteins from their CD data. The most successful methods use the CD spectra of proteins whose structure is known from X-ray diffraction as the basis for analyzing the CD of a protein with unknown structure. That is because there are more features than pure secondary structures that affect the CD of proteins in the amide region. For instance, the CD due to aromatic and sulfur-containing side chains, the length of α-helices, and twists in β-sheets all contribute to the CD in the amide region. Proteins contain all the features that affect their CD. If we use the CD of proteins with known structures as a basis, then all these features will be in the analysis, even if we do not recognize them directly.

In this paper we present an algorithm to estimate the secondary structure of proteins from CD data that reaches a new level of accuracy. Indeed, the accuracy is about the same as the variation in secondary structure found in X-ray diffraction data.[33] Workers cannot expect the accuracy in analyzing the CD spectra of proteins to be any better than the variation in the X-ray structures used for the proteins in the basis set. The method combines many of the ideas presented over the years in a new algorithm that gives a root mean square error of 4% or better for the secondary structures α-helix (H), 3_{10}-helix (G), β-strand (E), turn (T), and poly(L-proline) II type 3_1-helix (P).

A Fortran program called CDsstr that implements this algorithm is available over the internet, and is free of charge to anyone. Simply ftp to alpha.als.orst.edu. Login as anonymous, and please use your email address as the password. Change the directory by typing: cd /pub/wcjohnson/cdsstr. Notice that these are standard slashes since we are using a unix system. This directory contains the Fortran source code, test data and results, and the compiled binary version for a PC. To ensure that the binary version remains executable, type: bin. You can retrieve all files by typing: mget *.*

THE METHOD AND ITS RATIONALE

When workers use the same basis set of protein CD spectra together with the same known secondary structures, they are all stuck in the same vector space. Then the analysis of a protein with unknown structure should not be very dependent on the method of investigating this vector space. For the three-dimensional vector space in which we live, a vector will be the same whether it is described in a Cartesian coordinate system, a cylindrical coordinate system, or a spherical coordinate system. Different methods of analysis of CD spectra such as least squares fitting, singular value decomposition (SVD), convex constraint analysis, and neural networks simply apply different coordinate systems in the vector space of protein CD spectra. All of these methods should give about the same answer. We choose to use SVD in our algorithm.

Grant sponsor: National Institutes of Health; Grant number: GM 21479
*Correpondence to: W. Curtis Johnson, Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR 97331–7305. E-mail: johnsowc@ucs.orst.edu
Received 22 September 1998; Accepted 5 January 1999

Mathematically, the number of features that can be determined from the CD spectrum of a protein is equal to the number of protein CD spectra in the basis set. In practice accuracy is a problem, and the number of features that can be determined is limited by the information content of the data. Thus a central problem in the analysis of the CD spectrum of a protein for its secondary structure is that there is not enough data in the CD spectrum to accurately solve for all the features that determine the CD spectrum. SVD has been used to show that because of the experimental error in the CD spectrum of a protein measured to 178 nm, it has an information content of five.[11,28] This means that the CD spectrum is the equivalent of only five independent equations and therefore can accurately solve for only five unknowns. The problem is *underdetermined*, in the sense that there are more than five features that determine the CD of proteins.

However, the problem is not as bad as we might imagine. SVD can also be used to evaluate the information content of the secondary structures. The information content of the secondary structures has been shown to be four,[11,28] so one equation remains to help determine the other features.

The problem is also *overdetermined*, in the sense that there are more proteins in the basis set than the information content of the CD data. Then many different combinations of the CD spectra for the basis proteins will fit the CD to be analyzed within its experimental error. Small changes in the data can cause large changes in the analysis, and the results may well be inaccurate. This problem can be overcome with SVD by using only the five most important singular values and setting the rest equal to zero. The matrix algebra of using SVD has been described in a number of publications, and will not be repeated here.[12–14,19]

Truncating the CD spectra of proteins at 190 nm reduces the information content to three or four, and truncating the CD spectra at 200 nm reduces the information content to two. The equation that the sum of structures must be 1.0 adds another equation to the information content. However, it has been shown that when this equation is used as a constraint, it makes the analysis less accurate.[29] Requiring that the fractions of structure be positive will further destroy an analysis, predicting an inaccurate amount of α-helix, which without the constraint is usually fairly good.[29]

A basis set consisting of the CD for 22 proteins digitized at 4 nm intervals from 234 to 178 nm was used in this work. Many of these CD spectra have been published previously,[11–13,30] and all were available as the basis set with the earlier program, VARSLC.[27] Of course they are now available over the internet with CDsstr. Corresponding X-ray diffraction data with a resolution of at least 2.0 Å that has been refined are required for a protein to be included in the basis set. This criterion eliminated some of the 33 proteins contained in the earlier basis set. The accompanying publication describes the method we used to analyze the X-ray diffraction data for secondary structure, and the results are given in Table I of that publication. Note that hydrogen-bond and non-hydrogen-bonded β-turns have been combined under the symbol T for analysis of the protein CD spectra. The CD spectrum for each of the 22 proteins in the basis set can be analyzed for secondary structure using the other 21 proteins in the basis set. The results of this analysis (HJ), which is essentially our original algorithm,[11] are compared with the X-ray secondary structures in Table I. We see that the analysis for α-helix is quite good, as has been noted previously.[11,20] However, the analysis for other structures is variable, and in particular the sum of fractions of secondary structure is often very different from 1.0.

The fact that the sum of structures is not 1.0 for every protein using the HJ analysis is independent of the X-ray secondary structures assigned to the proteins in the basis set. Indeed, if instead of analyzing for component secondary structures, we simply analyze for the sum of secondary structure by assigning 1.0 for the sum to each protein in the basis set, we would still end up with the HJ sum of secondary structures given in Table I.

The sum of structures problem is undoubtedly related to the fact that there are only five independent equations in a CD spectrum of a protein measured to 178 nm, causing the analysis for secondary structure to be underdetermined. Tukey developed "variable selection" to get around the underdetermined problem.[31,32] This powerful idea is a standard procedure in the statistical analysis of data. With variable selection proteins are removed from the basis set to achieve an accurate analysis. Changing the coordinate system won't change the analysis, but with variable selection we change the vector space, which in turn will change the analysis. This kind of flexibility is the only way, outside of a constraint, to get the sum of structures to be 1.0 and eliminate negative values for some structures. In previous work, flexible methods like variable selection,[13] local linearity,[14] ridge regression,[10] and cluster analysis[20] have been used to change the basis set. Of course variable selection is not without its own problems. How do we know which proteins to remove, and how do we know when the analysis is satisfactory?

Of course, we do not know a priori which proteins to remove from the basis set. In previous work[13] we assumed that the final basis set should be as large as possible, and this assumption has been called into question.[14,21] We then removed proteins so that the sum of structures became close to 1.0 and negative fractions of structure were eliminated. In this work we find that it is best to use a small basis set, and eight proteins in the basis set gives the best results for the 22 proteins in the basis set where we already know the answer. Note that mathematically at least six proteins are required to be in the basis set to solve for the six features we are considering explicitly. There are 319,770 combinations of selecting eight proteins from a basis set of 22, and there will be more as the number of proteins in the basis set is increased. Rather than generating all these combinations, we follow Dalmas and Bannister,[21] and randomly choose the eight proteins for variable selection. We keep only those combinations where the sum of structures is between 0.952 and 1.05, where no fraction of secondary structure is less than –0.03, and where the

TABLE I. Comparison of Secondary Structures Predicted From CD to X-ray Results[†]

Protein	Method[a]	H	G	E	T	P	O	Sum
1. Azurin	X-ray	0.09	0.08	0.34	0.13	0.06	0.29	1.00
	HJ	0.10	0.04	0.32	0.11	0.04	0.37	0.99
	This work	0.07	0.04	0.36	0.11	0.04	0.39	1.00
2. Bence Jones protein	X-ray	0.00	0.00	0.34	0.10	0.12	0.44	1.00
	HJ	−0.07	0.01	0.28	0.10	0.04	0.30	0.74
	This work	−0.00	0.04	0.37	0.11	0.08	0.40	0.99
3. α-Chymotrypsin-a	X-ray	0.08	0.03	0.16	0.15	0.14	0.44	1.00
	HJ	0.08	0.04	0.15	0.12	0.08	0.35	0.82
	This work	0.05	0.04	0.24	0.14	0.11	0.42	1.00
4. Concanavalin-a	X-ray	0.01	0.01	0.36	0.12	0.07	0.42	1.00
	HJ	0.11	0.01	0.26	0.08	0.03	0.26	0.76
	This work	0.06	0.03	0.39	0.09	0.09	0.34	1.00
5. Cytochrome-c	X-ray	0.39	0.03	0.08	0.09	0.08	0.34	1.00
	HJ	0.31	0.04	0.10	0.11	0.05	0.27	0.88
	This work	0.36	0.05	0.09	0.17	0.04	0.29	1.00
6. Elastase	X-ray	0.06	0.04	0.21	0.14	0.11	0.44	1.00
	HJ	0.03	0.03	0.14	0.13	0.09	0.36	0.78
	This work	0.02	0.03	0.22	0.15	0.13	0.45	1.00
7. Flavodoxin	X-ray	0.25	0.06	0.17	0.20	0.02	0.30	1.00
	HJ	0.21	0.04	0.24	0.12	0.07	0.38	1.06
	This work	0.21	0.04	0.23	0.11	0.06	0.36	1.00
8. Hemerythrin	X-ray	0.59	0.12	0.04	0.06	0.03	0.17	1.00
	HJ	0.52	0.08	−0.19	−0.02	−0.06	−0.18	0.60
	This work	0.58	0.09	0.03	0.07	0.06	0.17	1.00
9. Hemoglobin	X-ray	0.67	0.08	0.00	0.10	0.00	0.15	1.00
	HJ	0.82	0.13	0.18	0.19	0.10	0.53	1.95
	This work	0.68	0.13	0.01	0.06	−0.01	0.13	1.00
10. Lactate dehydrogenase	X-ray	0.35	0.07	0.14	0.12	0.03	0.29	1.00
	HJ	0.31	0.09	0.23	0.11	0.04	0.32	1.09
	This work	0.34	0.09	0.17	0.11	0.02	0.28	1.00
11. β-Lactoglobulin	X-ray	0.09	0.04	0.34	0.13	0.04	0.37	1.00
	HJ	0.13	0.06	0.26	0.14	0.08	0.43	1.09
	This work	0.10	0.05	0.25	0.13	0.07	0.40	1.00
12. Lysozyme	X-ray	0.30	0.11	0.04	0.19	0.02	0.33	1.00
	HJ	0.26	0.08	0.20	0.15	0.10	0.43	1.21
	This work	0.30	0.08	0.12	0.12	0.06	0.32	1.00
13. Myoglobin	X-ray	0.70	0.11	0.00	0.05	0.02	0.12	1.00
	HJ	0.80	0.15	−0.05	0.08	−0.02	0.17	1.20
	This work	0.72	0.11	−0.01	0.04	−0.01	0.13	0.99
14. Papain	X-ray	0.24	0.05	0.15	0.11	0.08	0.37	1.00
	HJ	0.21	0.03	0.11	0.13	0.09	0.36	0.91
	This work	0.21	0.04	0.13	0.14	0.10	0.39	1.00
15. Pepsinogen	X-ray	0.09	0.09	0.27	0.13	0.10	0.33	1.00
	HJ	0.06	0.05	0.26	0.14	0.07	0.42	1.00
	This work	0.04	0.04	0.29	0.14	0.07	0.42	1.00
16. Prealbumin	X-ray	0.05	0.04	0.35	0.07	0.04	0.46	1.00
	HJ	0.04	0.06	0.29	0.14	0.07	0.36	0.95
	This work	0.03	0.06	0.32	0.15	0.07	0.37	0.99
17. Ribonuclease-a	X-ray	0.17	0.05	0.19	0.11	0.08	0.40	1.00
	HJ	0.24	0.06	0.09	0.14	0.08	0.34	0.94
	This work	0.24	0.05	0.12	0.14	0.09	0.36	1.00
18. Superoxide dismutase	X-ray	0.00	0.04	0.24	0.16	0.09	0.47	1.00
	HJ	0.07	0.08	0.34	0.20	0.12	0.56	1.37
	This work	0.04	0.08	0.26	0.19	0.05	0.40	1.01
19. T4 lysozyme	X-ray	0.59	0.08	0.06	0.04	0.01	0.22	1.00
	HJ	0.53	0.07	0.05	0.06	0.00	0.12	0.83
	This work	0.58	0.08	0.04	0.10	0.01	0.18	0.99
20. Thermolysin	X-ray	0.34	0.06	0.14	0.12	0.04	0.30	1.00
	HJ	0.33	0.07	0.14	0.08	0.02	0.21	0.86
	This work	0.35	0.08	0.17	0.09	0.04	0.27	1.01
21. Triose phosphate isomerase	X-ray	0.33	0.07	0.15	0.10	0.04	0.31	1.00
	HJ	0.38	0.06	0.18	0.11	0.07	0.34	1.12
	This work	0.40	0.06	0.13	0.11	0.02	0.27	1.00
22. Trypsin	X-ray	0.09	0.03	0.19	0.14	0.14	0.41	1.00
	HJ	0.06	0.01	0.05	0.07	0.05	0.20	0.44
	This work	0.07	0.02	0.22	0.12	0.12	0.43	0.98

[†]H, α-helix; G, 3_{10}-helix; E, β-strand; T, turns; P, poly(L-proline) II type 3_1-helix; O, other amides not in the previous categories.
[a]HJ is the original Hennessey and Johnson SVD analysis.[11]

TABLE II. Typical Secondary Structures for Successful Combinations of Concanavalan-A[†]

Combination number	Protein numbers[a]	H	G	E	T	P	O
868	1, 3, 7, 10, 12, 15, 16, 19	0.06	0.02	0.37	0.16	0.09	0.35
22	2, 7, 8, 12, 13, 16, 17, 21	0.09	0.01	0.35	0.09	0.07	0.40
31	1, 5, 9, 10, 11, 12, 14, 20	0.07	0.02	0.40	0.07	0.09	0.33
513	1, 6, 7, 9, 11, 12, 14, 22	0.07	0.02	0.38	0.13	0.07	0.31
831	1, 3, 7, 9, 12, 14, 16, 22	0.09	0.03	0.39	0.12	0.07	0.35
397	3, 5, 6, 10, 14, 17, 20, 21	0.36	0.07	0.03	0.14	0.15	0.26
595	5, 7, 9, 11, 15, 16, 17, 22	0.27	0.05	0.15	0.17	0.03	0.33
10	2, 3, 7, 10, 14, 15, 21, 22	0.22	−0.02	0.18	0.20	0.01	0.36
614	10, 11, 12, 16, 17, 20, 21, 22	−0.02	−0.03	0.50	0.01	0.06	0.45
119	3, 8, 14, 15, 16, 17, 18, 22	0.16	0.08	0.48	−0.01	0.00	0.32
860	6, 8, 13, 14, 16, 17, 18, 22	0.07	0.02	0.43	−0.01	0.01	0.44
100	1, 2, 5, 11, 13, 15, 17, 21	0.15	0.03	0.26	0.10	0.09	0.35
259	2, 5, 8, 10, 13, 14, 17, 21	0.17	0.00	0.23	0.09	0.09	0.38
602	2, 7, 9, 11, 14, 15, 17, 19	0.18	0.00	0.24	0.15	0.07	0.36

[†]H, α-helix; G, 3_{10}-helix; E, β-strand; T, turns; P, poly(L-proline) II type 3_1-helix; O, other amides not in the previous categories.
[a]Refer to Table I for matching protein numbers to proteins.

reconstructed CD spectrum fits the original CD spectrum with an average root mean square error of less than 0.25 Δε units. Typical successful combinations for concanavalan-A analyzed with the 21 other proteins are given in Table II. We see that some analyses are quite good (the first five in the table), while others are not very good at all (the remaining nine in the table). How can we choose the good analyses without already knowing the answer?

We do know that if we analyze using the complete basis set (HJ), we get about the right amount of α-helix (Table I). Graphing the HJ prediction for α-helix versus the X-ray α-helix shows high correlation and accuracy. It is better than estimating from Δε at 222 nm, or using only the first SVD basis vector. We can use the amount of α-helix predicted by the complete basis set to select from the variable selection analyses using eight proteins in the basis set without any a priori knowledge. In the end we use slightly more complicated criteria in our algorithm. The HJ method tends to overestimate α-helix for proteins with a low content, so if the predicted amount of α-helix is less than 0.15, we average this fraction with the minimum α-helix in the successful combinations, and then select combinations that are within 3% of this value. If the predicted α-helix is between 0.15 and 0.25, we select successful combinations that are within 3% of this value. If the predicted α-helix is between 0.25 and 0.65, we average this with the maximum α-helix in the successful combinations, and select successful combinations within 3% of this value. Finally, if the predicted α-helix is greater than 0.65, we select successful combinations with the largest amount of α-helix, since the successful combinations tend to underestimate α-helix for all-α proteins. For concanavalan-A these criteria select the first five and the eleventh successful combinations in Table II.

Sreerama and Woody[19] improved SVD and variable selection by putting the protein with unknown structure that was being analyzed into the basis set and iterating until the analysis was self-consistent. We use this self-consistency in our algorithm.

RESULTS AND DISCUSSION

Table I shows the results of analyzing each of the 22 proteins in the basis set with the other 21 proteins by using our new algorithm. The predictions of secondary structure compare well with the X-ray diffraction numbers. The root mean square error in the secondary structures for the 22 proteins in the basis set are: 3.3% for H, 2.6% for G, 4.2% for E, 4.2% for T, 2.7% for P, and 5.1% for O. Greenfield[27] has recently compared various algorithms of analyzing CD for secondary structure. The best method, program SELCON from Sreerama and Woody,[19,20] gave a root mean square error of 8% for H+G, 7% for E, and 5% for T. When our new basis set is run on SELCON, the root mean square error is 6.2% for H, 2.7 for G, 5.2% for E, 3.6% for T, 2.5% for P, and 5.1% for O. Clearly the accuracy we have achieved in this work is due both to the algorithm and to using a basis set with secondary structures from high quality X-ray data. The new criterion in the algorithm of basing α-helix estimates on the HJ predictions allows great flexibility in choosing the basis set, improving accuracy. Our correlation between predicted and X-ray structures are 0.99 for H, 0.62 for G, 0.94 for E, 0.38 for T, 0.76 for P, and 0.87 for O. Our error based on the center of the dynamic range for each structure is 9.4% for H, 43.3% for G, 23.3% for E, 40.0% for T, 35.7% for P, and 17.3% for O.

This algorithm demonstrates that our intuition is not always correct. For instance, we believed that the variable selection basis set should contain the maximum number of proteins, and stated this as one criterion in earlier work.[13] However, in this research we found that decreasing the number of proteins in the variable selection basis set improved the analysis, in the sense that there were some combinations that gave results close to the X-ray structure. All α-helix proteins analyzed best with six, seven, or

eight proteins in the basis set. Other proteins analyzed best with eight, nine, or ten proteins in the basis set, and some had no successful combinations with a basis set of only six proteins. We compromised on a basis set of eight proteins.

Another intuitive idea is the locally linear criterion,[14] that CD spectra in the basis set should resemble the CD spectrum being analyzed. Table II, which contains some successful combinations from randomly chosen basis sets of eight proteins applied to concanavalan-A, demonstrates that proteins with very different structure always appear in the variable selection basis sets that give the correct analysis. The eight proteins in each variable selection basis set are listed by number. We see that even though concanavalan-A contains a large amount of β-strand and very little α-helix, each analysis that agrees with the known secondary structure uses a basis set that contains at least two proteins with a large amount of α-helix and a very intense CD spectrum that is quite different from the weak CD spectrum of concanavalan-A.

There are two obvious criteria for a successful analysis: that the sum of fractions of secondary structure be about 1.0, and that there are no negative fractions of secondary structure. Without the flexibility of varying the basis set, these criteria can never be met, except by using them as constraints. However, it has been demonstrated that these criteria used as constraints destroy the analysis.[29] The solution has to be flexible in the choice of the basis set, and this research demonstrates that the tremendous flexibility available using variable selection when there are only eight proteins in the basis set leads to a number of good analyses. Indeed, variable selection with a minimum basis set is so flexible that even CD data truncated at 200 nm give analyses with a root mean square error of about 5%. Apparently, a well-chosen basis set with the important proteins can compensate for a woeful lack of information content. It must extract the component spectra in such a way that the problem is no longer underdetermined. However, the more data you have the better your answers will be. We strongly suggest you collect data to 178 nm.

Although we demonstrate here that flexibility in the basis set leads to successful analyses, in general flexibility has not been good for the prediction of α-helix. Sreerama and Woody investigated many methods for prediction of secondary structure,[20] and their work (see also ref. 27) shows that the best prediction of α-helix comes from the SVD method used with no flexibility. This is the method that we use to predict α-helix, which in turn is used to select from the many analyses generated by using the flexible variable selection basis set. In this work we have combined the ideas of many workers together with flexibility in creating the basis set to achieve a highly accurate analysis of secondary structure for a given protein.

ACKNOWLEDGMENT

It is a pleasure to thank Dr. Narasimha Sreerama for helpful conversations and for running our basis set on the SELCON algorithm.

REFERENCES

1. Greenfield N, Fasman GD. Computed circular dichroism spectra for the evaluation of protein conformation. Biochemistry 1969;8: 4108–4116.
2. Saxena VP, Wetlaufer DB. A new basis for interpreting the circular dichroism spectra of proteins. Proc Natl Acad Sci USA 1971;68:969–972.
3. Chen Y-H, Yang JT. A new approach to the calculation of secondary structures of globular proteins by optical rotatory dispersion and circular dichroism. Biochem Biophys Res Commun 1971;44:1285–1291.
4. Rosenkranz H, Scholten W. An improved method for the evaluation of helical protein conformation by means of circular dichroism. Hoppe-Seyler's Z Physiol Chem 1971;352:896–904.
5. Chen Y-H, Yang JT, Chan KH. Determination of the helix and β-form of proteins in aqueous solution by circular dichroism. Biochemistry 1974;13:3350–3359.
6. Bannister WH, Bannister JV. A study of three-component fitting of protein circular dichroism spectra. Int J Biochem 1974;5:679–686.
7. Chang CT, Wu C-SC, Yang JT. Circular dichroism analysis of protein conformation: inclusion of β-turns. Anal Biochem 1978;91: 13–31.
8. Brahms S, Brahms J. Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. J Mol Biol 1980;138:149–178.
9. Bolotina IA, Chekhov VO, Lugauskas VYu, Finkel'shtein AV, Ptitsyn OB. Determination of the secondary structure of proteins from the circular dichroism spectra. 1. Protein reference spectra for α-, β- and irregular structures. Mol Biol (Eng. Transl.) 1980;14:701–709.
10. Provencher SW, Glöckner J. Estimation of protein secondary structure from circular dichroism. Biochemistry 1981;20:33–37.
11. Hennessey JP, Jr, Johnson WC, Jr. Information content in the circular dichroism of proteins. Biochemistry 1981;20:1085–1094.
12. Compton LA, Johnson WC, Jr. Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. Anal Biochem 1986;155:155–167.
13. Manavalan P, Johnson WC, Jr. Variable selection method improves the prediction of protein secondary structure from circular dichroism. Anal Biochem 1987;167:76–85.
14. van Stokkum IHM, Spoelder HJW, Bloemendal M, van Grondelle R, Groen FCA. Estimation of protein secondary structure and error analysis from circular dichroism spectra. Anal Biochem 1990;191:110–118.
15. Pancoska P, Keiderling TA. Systematic comparison of statistical analysis of electronic and vibrational circular dichroism for secondary structure prediction of selected proteins. Biochemistry 1991;30: 6885–6895.
16. Perczel A, Hollosi M, Tusnady G, Fasman GD. Convex constraint analysis: a natural deconvolution of circular dichroism curves of proteins. Protein Eng 1991;4:669–679.
17. Böhm G, Muhr R, Jaenicke R. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. Protein Eng 1992;5:191–195.
18. Perczel A, Park K, Fasman GD. Analysis of the circular dichroism spectrum of proteins using the convex constraint algorithm: A practical guide. Anal Biochem 1992;203:83–93.
19. Sreerama N, Woody RW. A self-consistent method for the analysis of protein secondary structure from circular dichroism. Anal Biochem 1993;209:32–44.
20. Sreerama N, Woody RW. 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 1994;242:497–507.
21. Dalmas B, Bannister WH. Prediction of protein secondary structure from circular dichroism spectra: an attempt to solve the problem of the best-fitting reference protein subsets. Anal Biochem 1995;225:39–48.
22. Rosenkranz H. Circular dichroism of globular proteins. A review of the limits of the CD methods for the calculation of secondary structure. Klin Chem Klin Biochem 1974;9:415–422.
23. Bannister WH, Bannister JV. Minireview: circular dichroism and protein structure. Int J Biochem 1974;5:673–677.

24. Woody RW. Circular dichroism of peptides. In: Hruby VJ, editors, The Peptides, Vol. 7. New York:Academic Press; 1985. p 15–114.
25. Yang JT, Wu C-SC, Martinez HM. Calculation of protein conformation from circular dichroism. Meth Enzymol 1986;130:208–269.
26. Johnson WC, Jr. Secondary structure of proteins through circular dichroism spectroscopy. Annu Rev Biophys Chem 1988;17:145–166.
27. Greenfield NJ. Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. Anal Biochem 1996;235:1–10.
28. Johnson WC, Jr. Analysis of circular dichroism spectra. Meth Enzymol 1992;210:426–447.
29. Manavalan P, Johnson WC, Jr. Protein secondary structure from circular dichroism spectra. Proc Int Symp Biomol Struct Interactions, Suppl J Biosci 1985;8:141–149.
30. Toumadje A, Alcorn SW, Johnson WC, Jr. Extending CD spectra of proteins to 168 nm improves the analysis of secondary structures. Anal Biochem 1992;200:321–331.
31. Mosteller F, Tukey JW. Data analysis and regression. Reading, MA: Addison-Wesley; 1977. 588 p.
32. Weisberg S. Applied linear regression. New York: John Wiley & Sons; 1980. 323 p.
33. King SM, Johnson WC. Assigning secondary structure from protein coordinate data. Proteins 1999;35:313–320.