Electronic Circular Dichroism of Proteins from First-Principles Calculations

Jonathan D. Hirst,* Karl Colella, and Andrew T. B. Gilbert

School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom

Received: June 23, 2003; In Final Form: August 19, 2003

The circular dichroism (CD) spectra of 47 proteins in the far-ultraviolet have been calculated from first principles, using a parameter set derived from ab initio calculations on *N*-methylacetamide. These spectra agree well with experimental data, as shown by the Spearman rank correlation coefficients of 0.86, 0.80, and 0.94 between the computed and experimental intensities at 190, 208, and 220 nm, respectively. The computed spectra are most accurate for proteins that have a high α-helical content and are least accurate for a class of β-sheet-rich proteins, which have some irregular structure and are known as β-II proteins. To address the lack of resolution between the two negative peaks around 208 and 220 nm in the calculated spectra of α-helical proteins, narrower bandwidths have been explored. Other factors were investigated, including the dielectric constant of the protein, higher energy transitions of the amide chromophore, and the orientation of the $\pi\pi^*$ electric transition dipole moment vector. Combining some of these aspects made it possible to obtain accurate spectra with the desired resolution between the negative peaks. Although not fully quantitative, the first-principles calculations of protein CD presented in this study are the most accurate reported to date.

Introduction

Polypeptides are chiral. This asymmetry renders them amenable to study by electronic circular dichroism (CD) spectroscopy. CD is the differential absorption of left- and right-handed circularly polarized light. The basic principles underlying the CD of biopolymers are understood. Electronic transitions of (achiral) monomer groups mix in the context of the chiral environment of helices (and other structures) to give rise to transitions that are both electrically and magnetically allowed. Despite knowledge of the fundamentals, fully quantitative calculations of CD spectra of proteins from first principles have yet to be realized. Such calculations require quantum chemical models of the monomer chromophores and more approximate models of the interactions between monomers. Our work aims to improve the quality of first-principles calculations of protein CD and, in the longer term, to enhance the utility of CD spectroscopy as a biophysical technique for studying protein conformation.

Proteins give distinctive CD spectra in the far-ultraviolet (far-UV). From these spectra it is possible to determine the approximate proportion of the different elements of secondary structure (α-helices, β-sheets, etc.) in the protein.[1−7] For example, a protein with a high α-helical content will display an intense positive peak around 190 nm and negative peaks around 208 and 220 nm. CD has been widely used to probe the mechanism of protein folding, and recent developments have made it possible to obtain CD spectra on the nanosecond time scale.[8−10] A better understanding of the origin of CD spectra would facilitate a fuller interpretation of protein CD experiments.

Each secondary structure type has a distinctive contribution to the CD spectrum of a protein. Figure 1 shows the experimental CD spectra of myoglobin, concanavalin A, and elastase. Myoglobin has a high α-helical content. Its CD spectrum has an intense positive peak around 193 nm and negative peaks around 208 and 222 nm, which is typical for α-helical proteins. The intensity of the peak around 222 nm is correlated with the helical content.[11,12] Proteins that contain mostly β-sheets can be categorized as class I (β-I) and class II (β-II).[13] Proteins in the former class, such as concanavalin A, contain fairly regular sheets, and their CD spectra generally exhibit a negative band around 217 nm and a positive band at ∼195 nm. This positive band is much less intense than the one seen in α-helices. Proteins in the latter class, such as elastase, often contain disulfide bridges and irregularities such as β-bulges. The CD spectra of β-II proteins tend to look more like those of a random coil,[13] with a negative band around 198 nm.

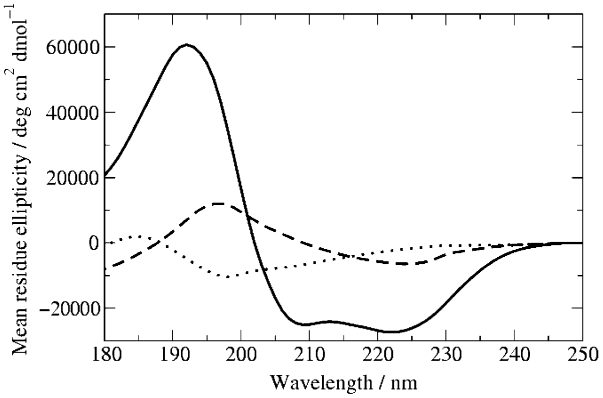

Figure 1. Typical CD spectra for myoglobin (solid line), concanavalin A (dashed line), and elastase (dotted line).

We can learn more about the origin of these spectra through first-principles calculations of the CD spectra of proteins from their three-dimensional structures. Several computational methods have been developed to do this. In the dipole interaction model[14] atoms and chromophores are considered to act as point dipole oscillators that interact through mutually induced dipole moments in the presence of an electric field. Another approach is the matrix method,[15−18] which is based on classical electrostatic interactions between the individual chromophoric groups

* Corresponding author. Tel: +44 115 951 3478. Fax: +44 115 951 3562. E-mail: jonathan.hirst@nottingham.ac.uk.

10.1021/jp035775j CCC: $25.00 © 2003 American Chemical Society
Published on Web 09/30/2003

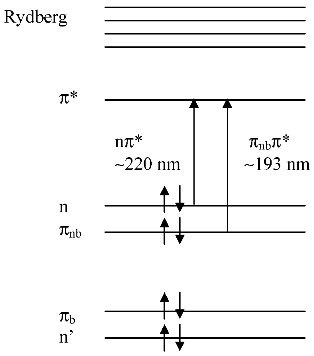

Figure 2. Electronic transitions of the amide group in the far-UV region. The molecular orbitals shown are the bonding, nonbonding, and antibonding π orbitals (π_b, π_{nb}, and π^*) and two lone pairs on the oxygen atom (n and n′).

in a protein. It has recently been applied with some success on sets of 23 and 29 proteins.[19,20]

The overall thrust of this study is to improve the accuracy of calculations of protein CD spectra. On one hand, we incorporate additional physics into the calculations, through consideration of the interior dielectric of the protein and by incorporating higher energy electronic transitions of the amide group. On the other, we pursue a more phenomenological approach to investigate two aspects. It is nontrivial to estimate from first principles the appropriate bandwidth of electronic transitions in CD spectra, but inspection of experimental spectra can provide some guidance. Another issue, where experiment can have some direct input, is the precise orientation of the electric transition dipole moments, especially of intense transitions. Previous work[21] has used experimental data[22] on the $\pi\pi^*$ transition moment direction in protein CD calculations, and we assess the desirability of this further.

The first aim of this study was to extend the calculations of Besley and Hirst[19] using the matrix method to a larger set of proteins, whose CD spectra have recently been collated and published.[1,2] Several aspects of the matrix method calculations were then investigated. The bandwidths applied to the calculated line spectra were reduced in an attempt to resolve peaks more clearly. Previous studies have used bandwidths (half-width at e^{-1} of the maximum) as broad as 15.5 nm, which do not permit peaks at 208 and 222 nm to be properly resolved. Narrower bandwidths lead to sharper peaks, with, in some cases, too large a maximum intensity. To mitigate against this, the dielectric constant of the protein was varied. Two higher energy transitions of the amide chromophore were included in the calculation, as these can, in principle, interact with transitions in the far-UV. The orientations of the electric transition moment vector of the $\pi\pi^*$ transitions were explored. These aspects were systematically studied to improve the accuracy of the computed CD spectra. Several combinations gave well-resolved spectra that replicated the experimental data satisfactorily, especially for proteins with a high α-helical content.

Method

The CD of proteins is usually measured in the far-UV region of the electromagnetic spectrum, typically between 180 and 250 nm. This corresponds to electronic transitions of the backbone chromophore in proteins (Figure 2). For example, the absorption peaks at ∼193 and ∼208 nm in the spectrum of an α-helical protein are due to electrically allowed $\pi_{nb}\pi^*$ transitions (transitions from the nonbonding π orbital to the antibonding π orbital). The peak around 220 nm is due to the magnetically allowed nπ^* transition (lone pair on oxygen to π^*) and the static-field mixing of this transition with the $\pi_{nb}\pi^*$ transitions.[23] Higher energy transitions may also occur, such as $\pi_b\pi^*$ (bonding π orbital to π^*) and n′π^* (second lone pair on oxygen to π^*). Although these transitions occur at energies outside the range of conventional instrumentation, their inclusion may be important to capture potential coupling between the electrically allowed transitions ($\pi_b\pi^*$ and $\pi_{nb}\pi^*$), which could alter the intensity and position of the bands in the far-UV.

An important aspect of calculating protein CD spectra is the accurate parametrization of the ground and excited electronic states of the amide chromophore. A small molecule such as *N*-methylacetamide (NMA) usually serves as a model for a single chromophore. Woody and Sreerama[21] employed a semiempirical approach, in which the electric transition dipole moment vector is taken from experimental results[22] and the rest of the parameters are calculated from the intermediate neglect of differential overlap/spectroscopic (INDO/S) wave functions for NMA.[21] Besley and Hirst[19] used the complete active space self-consistent-field method implemented within a self-consistent reaction field (CASSCF/SCRF) combined with multiconfigurational second-order perturbation theory (CASPT2-RF) to calculate the ground and electronic excited states of NMA.[24] This is a continuum dielectric model of solvent, where bulk properties of water are accounted for, but individual water molecules are not explicitly considered. Besley and Hirst[25] have also tried a semicontinuum approach, where water molecules that are hydrogen-bonded to NMA are explicitly considered and bulk water is treated as a continuum dielectric. However, parameters from the semicontinuum calculations do not lead to more accurate protein CD calculations, implying that hydrogen-bonding interactions are probably adequately modeled in the matrix method calculation and so not necessary to account for in the monomer.

From such calculations, the magnitudes and directions of the transition moments can be obtained. The electrostatic potential is calculated at a series of points on a grid centered on the NMA monomer for the permanent and transition electronic densities. Then charges (either five or eight charges around each atom, depending on the symmetry of the wave functions) are fitted around the carbon, nitrogen, and oxygen and the amide hydrogen atoms to reproduce the electrostatic potential as closely as possible. There is typically a 5% error on a given point. Thus, we have discrete charges that can be used in eq 5 in the next section.

The matrix method[15−18] utilizes classical electrostatic interactions between chromophores to determine transition energies and rotational strengths in large molecules such as proteins. Proteins are considered to consist of *M* chromophoric groups with independently defined electronic eigenstates. The total wave function ψ_k for state *k* of the protein is expressed as a sum of basis functions Φ_{ia} (where *i* is a particular chromophore and *a* is an excited state):

$$\psi_k = \sum_i^M \sum_a^{n_i} c_{ia}^k\, \Phi_{ia} \tag{1}$$

where n_i is the number of excitations associated with a group. Each basis function is a product of *M* monomer wave functions:

$$\Phi_{ia} = \phi_{10} \cdots \phi_{ia} \cdots \phi_{j0} \cdots \phi_{M0} \tag{2}$$

where the first subscript labels the chromophore and the second subscript denotes the electronic state: 0 is the ground state and *a* is an excited state. Thus ϕ_{ia} represents the wave function of monomer *i*, which has undergone an electronic transition from the ground state 0 to the excited state *a*. These monomer wave functions are obtained using methods described earlier.

Now that the total wave function can be calculated, we need to consider the total electronic Hamiltonian operator Ĥ for the protein. This is written as

$$\hat{H} = \sum_{i=1}^{M} \hat{H}_i + \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \hat{V}_{ij} \qquad (3)$$

where $\hat{H}_i$ is the Hamiltonian for chromophore *i* and $\hat{V}_{ij}$ is the interaction between monomers *i* and *j*. In the Hamiltonian matrix, the diagonal elements are the excitation energies of the electronically excited states and each off-diagonal element arises from either interactions between groups or the static-field mixing of excitations within a single group. The interactions between groups are assumed to be purely electrostatic. They can be considered as perturbations, with the effect of splitting the otherwise degenerate transitions over a wider range of energies. It is these interactions that give protein CD spectra their dependency on secondary (and tertiary) structure. They take the form

$$V_{i0a;j0b} = \int\int \frac{\rho_{i0a}(\mathbf{r}_i)\ \rho_{j0b}(\mathbf{r}_j)}{4\pi\epsilon_0 r_{ij}} \mathrm{d}\mathbf{r}_i \mathrm{d}\mathbf{r}_j \qquad (4)$$

where $\rho_{i0a}(\mathbf{r}_i)$ and $\rho_{j0a}(\mathbf{r}_j)$ represent the permanent and transition electron densities on chromophores *i* and *j*, respectively; ϵ_0 is the vacuum permittivity and r_{ij} is the distance between the chromophores. For mixing of excitations on a single group the same equation can be used, but $\rho_{i0a}(\mathbf{r}_i)$ would be replaced by the ground state electronic density of the whole molecule except for group *i*, and $\rho_{j0a}(\mathbf{r}_j)$ would be replaced by the nπ^* to $\pi_{\mathrm{nb}}\pi^*$ transition electron density for group *i*.

The dielectric constant in the interior of a protein is not that of a vacuum. Introducing a relative dielectric constant ϵ to the denominator of eq 4 allows for this. Higher values of ϵ lead to weaker electrostatic interactions. Organic liquids have values around 2; ionic compounds have much higher values. The dielectric constant of a protein can be calculated using the Poisson equation and other methods,[26,27] but this is beyond the scope of this paper. Values between 2 and 4 are generally accepted because proteins are similar to organic liquids, but contain some polar and charged residues. However, care should be taken when using a dielectric constant in the calculations, because the value depends on how explicitly interactions are taken into account in the model used.[27] A model that accounts for everything explicitly will not contain a dielectric constant. Also, the dielectric constant may vary between different regions of the protein.

The electron densities are normally represented by point charges, rather than a continuous electrostatic potential, to reduce the computational expense. Hence, the double integral in eq 4 is replaced by a sum over these charges:

$$V_{i0a;j0b} = \sum_{s=1}^{N_s} \sum_{t=1}^{N_t} \frac{q_s q_t}{r_{st}} \qquad (5)$$

where q_s and q_t are the point charges on chromophores *i* and *j*, and N_s and N_t are the number of these charges on the chromophore.

The Hamiltonian matrix is then diagonalized by a unitary transformation, so that the diagonal elements now represent the transition energies of the interacting system. The intensity, or rotational strength, of each transition is obtained from the following expression, derived[28] from the Rosenfeld equation,[29] which involves the electric transition dipole moment, μ, and the magnetic transition moment, *m*:

$$R_{0A} = \mathrm{Im}(\langle\psi_0|\vec{\mu}|\psi_A\rangle\cdot\langle\psi_A|\vec{m}|\psi_0\rangle) \qquad (6)$$

Thus, the rotational strength of a transition from the ground electronic state 0 to the excited state *A* is given by the imaginary part (Im) of the scalar product of μ and *m*. ψ_0 and ψ_A are the relevant ground and excited state wave functions. From this equation it is clear that the direction of the transition moments affects the rotational strengths of the transitions and also that if there were no interactions between transitions of different groups, there would be no CD spectrum, because either μ or *m* would be zero. The rotational strength derived using the Rosenfeld equation is origin-independent only for exact wave functions,[21] so it is common to calculate the origin-independent chiral strength and then convert this to rotational strength.[30] The transition dipole moment vectors are transformed by the same unitary transformation used to diagonalize the Hamiltonian matrix to give the rotational strengths of the interacting system, which are proportional to the areas underneath the peaks in the CD spectrum.[21] A CD spectrum is obtained by centering a band function such as a Gaussian curve around the rotational strengths calculated for the specific transition energies.

The matrix method uses several approximations. The protein wave function is constructed from single electronic excitations on single groups. Some calculations performed by Koslowski[31] have shown that including double excitations does not significantly affect the calculated spectrum. Side chains are often not incorporated into the calculations. They may influence the CD spectrum of a protein, especially if the protein contains a high proportion of aromatic amino acids[32−36] or a low α-helix content.[37] However, their inclusion seems to lead to a decrease in the correlation between the calculated and experimental CD spectra in the far-UV region.[38] Parametrizations of the side chain wave functions are in the process of being improved. The structures used in the calculations are taken from X-ray crystallography, whereas CD spectra are taken in solution. There is, therefore, the assumption that the static crystal structure of a protein is the same as the solution ensemble. In some cases, conformational dynamics may be important, and recent studies are starting to take account of this,[39,40] but this is beyond the scope of the current study.

Two measures of the accuracy of the calculated spectra with respect to the experimental spectra are used. One is the Spearman rank correlation coefficient, which is calculated at the key wavelengths of 190, 208, and 220 nm. This facilitates comparison with previous work. The other measure is the mean absolute error in the range 190−230 nm. This provides an assessment of the overall accuracy of the calculations. CD spectra were calculated for 47 proteins (Table 1). To improve the resolution between the peaks around 208 and 220 nm, a variety of bandwidths was explored, from 7.5 to 15.5 nm. The Spearman rank correlation coefficients are barely affected by changing the bandwidth or band shape, because they are based on rank not absolute value. Hence, the most accurate calculated CD spectra may be taken to be those with the smallest mean absolute error. The next factor that was investigated was the dielectric constant ϵ of the protein. Values between 1 and 2.2 were explored; above ϵ = 2.2 the calculation significantly underestimates the strength of the transitions. Varying the

TABLE 1: The 47 Proteins Used in This Study and Their Secondary Structure Class

class	protein (PDB code)
α-helix	colicin A (1col), bacteriorhodopsin (2brd), hemerythryin (2hmz), hemoglobin (2mhb), insulin (4ins), myoglobin (4mbn), parvalbumin (5cpv), cytochrome C (5cyt)
mixed α, *β*	*β*-lactoglobulin (1beb), azurin (1e5z), green fluorescent protein (1ema), ecoRI endonuclease (1eri), flavodoxin (1fx1), rat intestinal fatty acid binding protein (1ifc), rhodanese (1rhd), subtilisin (1sbt), tumor necrosis factor (1tnf), T4 lysozyme (2lzm), pepsinogen (2psg), subtilisin novo (2sbt), staphylococcal nuclease (2sns), adenylate kinase (3adk), glyceraldehyde-3-phosphate dehydrogenase (3gpd), glutathione reductase (3grs), phosphoglycerate kinase (3pgk), triosephosphate isomerase (3tim), *γ*-crystallin (4gcr), alcohol dehydrogenase (5adh), carboxypeptidase A (5cpa), lactate dehydrogenase (6ldh), thermolysin (8tln), papain (9pap)
β-I	carbonic anhydrase (1ca2), plastocyanin (1plc), *λ*-immunoglobulin (1rei), concanavalin A (2ctv), prealbumin (2pab), erabutoxin (3ebx), porin (3por), staphylococcal nuclease (3rn3)
β-II	α-bungarotoxin (2abx), α-chymotrypsinogen (2cga), superoxide dismutase (2sod), elastase (3est), trypsin (3ptn), α-chymotrypsin (5cha), bovine pancreatic trypsin inhibitor (5pti)

TABLE 2: Spearman Rank Correlation between Experimental and Calculated CD Intensities at 190, 208, and 220 nm

method	no. of proteins	correlation coefficient		
		$r_{[\lambda=190]}$	$r_{[\lambda=208]}$	$r_{[\lambda=220]}$
dipole interaction method[14]	15	0.89	0.75	0.74
matrix method and semiempirical parameters[21]	23	0.69[a] (0.66)[b]	0.72	0.86 (0.84)
	47	0.68[a]	0.67	0.93
matrix method and ab initio parameters[19]	15	0.87	0.71	0.96
	23	0.81	0.73	0.89
	29	0.84	0.73	0.90
	47	0.86	0.80	0.94

[a] Values computed using just the nπ* and $\pi_{nb}\pi$* transitions. [b] Values in parentheses were previously reported[21] and include the uncoupled $\pi_b\pi$* transition.

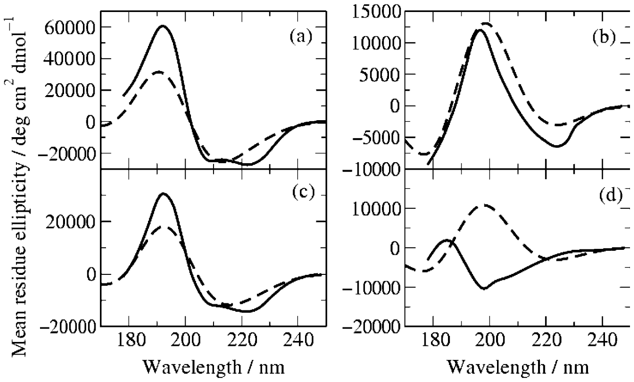

Figure 3. Experimental (solid lines) and calculated (dashed lines) CD spectra of (a) myoglobin (α-helical), (b) concanavalin A (*β*-I), (c) lactate dehydrogenase (mixed α, *β*), and (d) elastase (*β*-II). Calculations use a bandwidth of 15.5 nm.

dielectric constant was systematically combined with varying the bandwidth.

In the calculations above, two amide transitions, the $\pi_{nb}\pi$* and nπ* transitions, have been considered. The inclusion of higher energy amide transitions, the $\pi_b\pi$* and the n′π* transitions, has been investigated. This appears to be problematic. We have explored several possible orientations of the strongly allowed $\pi_{nb}\pi$* and $\pi_b\pi$* transitions to assess the sensitivity of the calculated interaction to orientation. The parameters for all four amide transitions have been previously made available as Supporting Information in an earlier publication.[19]

Results and Discussion

The Spearman rank correlation coefficients between the calculated and experimental intensities of the 47 proteins at 190, 208, and 220 nm were calculated and compared to previously reported results (Table 2). From these values it is apparent that the accuracy of the matrix method calculations with the ab initio parameters is maintained with a larger set of proteins and outperforms other approaches and parameter sets. The proteins can be divided into four categories (Table 1): α-helical, mixed α−*β*, class I *β*-sheet (*β*-I), and class II *β*-sheet (*β*-II). For ease of comparison, the figures in this paper will consist of one representative of each class: myoglobin (4mbn), lactate dehydrogenase (6ldh), concanavalin A (2ctv), and elastase (3est), respectively. Calculated spectra, with bandwidths of 15.5 nm, for these proteins are shown Figure 3.

The spectra for myoglobin and lactate dehydrogenase (Figures 3a and 3c) are typical of α-helical proteins. Myoglobin is highly helical, and lactate dehydrogenase is mostly helical with some *β*-sheet character. Methods of calculating CD spectra of proteins, including the matrix method, tend to perform well with these classes, because they have the most regular repetitive secondary structure. The peak around 193 nm is fairly well reproduced by the calculation, although less intense than the experimental data; the peaks at ∼208 and ∼220 nm are poorly resolved, due to the bandwidth of the fitted Gaussian curves.

Concanavalin A is a *β*-I type protein. A typical *β*-I protein spectrum has a negative peak around 217 nm and a positive peak around 195 nm, but this is quite variable. The matrix method is fairly successful at predicting the spectra for this class, but often overestimates the intensity of the positive peak (Figure 3b). Elastase is a *β*-II type protein. The matrix method has some difficulty predicting spectra for this class of protein, perhaps partly because of their structural diversity, such as the presence of *β*-bulges and perhaps partly due to the presence of conformations similar to the poly(Pro)II-type structure[40] or to conformational flexibility.[41] Some *β*-II proteins also contain disulfide bridges, which have electronic transitions that are not accounted for in the calculations. Spectra for this class are often similar to that of random coils, with a negative peak around 198 nm, but the matrix method predicts *β*-I type spectra (Figure 3d).

The resolution of the peaks around 208 and 220 nm was improved using narrower bandwidths. Values below 12.5 nm generally give the two minima for proteins with a high α-helical content. However, this typically makes the peaks around 193 and 208 nm too intense (Figure 4 shows this for a bandwidth of 9.5 nm), as the area under a band must be conserved for a given rotational strength. Bode and Applequist[14] have noted that no single bandwidth fits all experimental CD spectra well. Use of Lorentzian curves rather than Gaussian band shapes gave better resolution, but the overall agreement with the experimental spectra is quite poor (Table 3), because the bands become too intense. Gaussian curves are used exclusively from hereon.

Empirically, it is clear that a bandwidth of 15.5 nm is too broad and that a narrower bandwidth is physically more

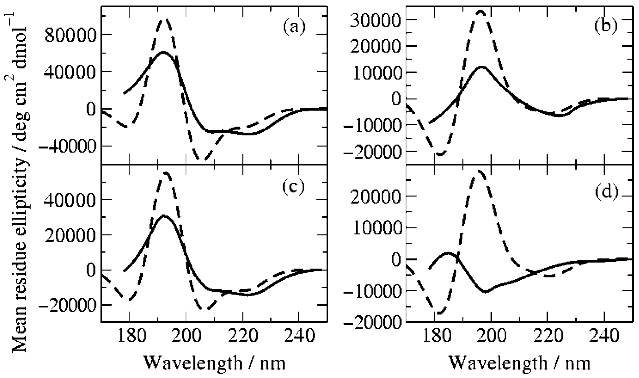

Figure 4. Experimental (solid lines) and calculated (dashed lines) CD spectra of (a) myoglobin, (b) concanavalin A, (c) lactate dehydrogenase, and (d) elastase. Calculations use a bandwidth of 9.5 nm.

TABLE 3: Quality of Calculated Spectra Using Gaussian (G) and Lorentzian (L) Bands, as Measured by the Mean Absolute Error Computed over 190−230 nm and the Correlation between the Experimental and Calculated Intensities

bandwidth/nm	curve	correlation coefficient			mean absolute error deg cm^2 dmol^{-1}
		$r_{[\lambda=190]}$	$r_{[\lambda=208]}$	$r_{[\lambda=220]}$	
9.5	G	0.85	0.82	0.90	10312
	L	0.84	0.81	0.90	14232
11.5	G	0.85	0.81	0.92	6954
	L	0.84	0.81	0.91	10896
13.5	G	0.86	0.80	0.92	5561
	L	0.84	0.81	0.92	8658
15.5	G	0.86	0.80	0.94	5439
	L	0.85	0.81	0.92	7194

TABLE 4: Values of the Dielectric Constant and the Bandwidth that Gave the Most Accurate Calculated CD Spectra

ϵ	bandwidth/nm	mean absolute error/ deg cm^2 dmol^{-1}	correlation coefficient		
			$r_{[\lambda=190]}$	$r_{[\lambda=208]}$	$r_{[\lambda=220]}$
1.0	14.5	5387	0.86	0.80	0.93
1.0	15.5	5439	0.86	0.80	0.94
1.0	13.5	5561	0.86	0.80	0.92
1.2	13.5	5645	0.86	0.80	0.93
1.2	12.5	5664	0.86	0.80	0.92
1.2	14.5	5788	0.86	0.80	0.93
1.2	11.5	5983	0.86	0.81	0.90
1.2	15.5	6030	0.86	0.80	0.93
1.0	12.5	6060	0.85	0.80	0.92
1.5	11.5	6220	0.86	0.80	0.90
1.5	12.5	6271	0.86	0.80	0.91
1.5	10.5	6377	0.86	0.81	0.88

appropriate. The overly intense bands observed with the narrower bandwidths may indicate that a dielectric constant larger than unity should be employed in the matrix method calculations. To investigate this, calculations were performed in which both the dielectric constant and bandwidth were varied. The combinations that give the smallest errors are shown in Table 4, along with the corresponding correlation coefficients. These results show that it is possible to use a narrower bandwidth with only a small deterioration in the absolute error between the calculated and experimental spectra. This is exemplified by the spectra in Figure 5, in which a dielectric constant value of 1.5 and a bandwidth of 10.5 nm have been used. The bands in these spectra are not too intense, and the two negative peaks can be distinguished. However, as can be seen in Table 4, the spectra with the smallest absolute errors are generally those in which the dielectric constant is unity and where the bandwidth is too broad to give the required resolution.

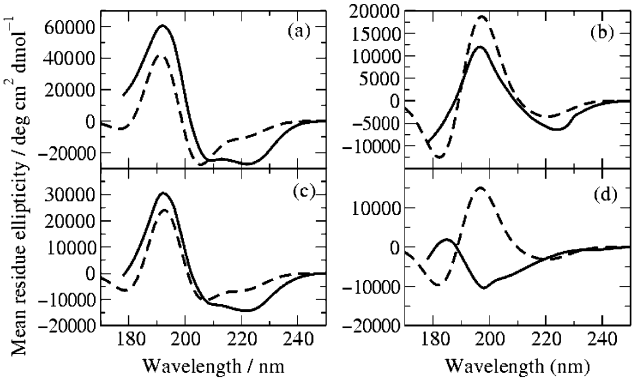

Figure 5. Experimental (solid lines) and calculated (dashed lines) CD spectra of (a) myoglobin, (b) concanavalin A, (c) lactate dehydrogenase, and (d) elastase. Calculations use a bandwidth of 10.5 nm and a dielectric constant of 1.5.

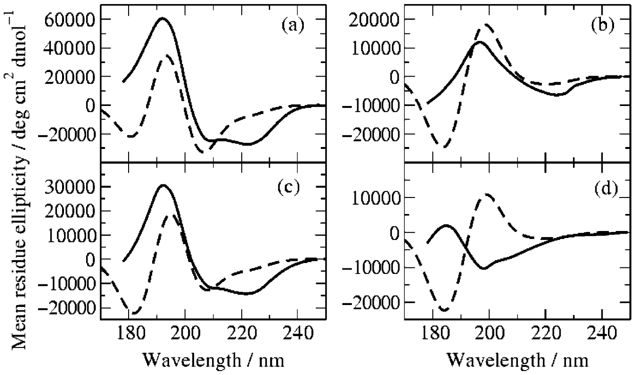

Figure 6. Experimental (solid lines) and calculated (dashed lines) CD spectra of (a) myoglobin, (b) concanavalin A, (c) lactate dehydrogenase, and (d) elastase. Calculations use a bandwidth of 10.5 nm and a dielectric constant of 1.2; four amide transitions are included.

TABLE 5: Values of the Dielectric Constant and the Bandwidth that Gave the Most Accurate CD Spectra from Calculation Including the Amide $\pi_b\pi^*$ and n′π^* Transitions

ϵ	bandwidth/nm	mean absolute error/ deg cm^2 dmol^{-1}	correlation coefficient		
			$r_{[\lambda=190]}$	$r_{[\lambda=208]}$	$r_{[\lambda=220]}$
1.0	11.5	6088	0.76	0.81	0.91
1.0	12.5	6093	0.75	0.80	0.92
1.0	13.5	6368	0.74	0.80	0.93
1.0	10.5	6628	0.77	0.81	0.89
1.2	10.5	6673	0.78	0.81	0.85
1.2	11.5	6705	0.77	0.81	0.89
1.0	14.5	6729	0.72	0.79	0.93
1.2	10.0	6820	0.78	0.81	0.83

Figures 4 and 5 suggest that the calculated rotational strengths of the $\pi_{nb}\pi^*$ transitions comprising the bands at ~190 and 208 nm are too high in relation to the rotational strengths of the nπ^* transitions at 220 nm. The electrically allowed $\pi_{nb}\pi^*$ transitions may couple with higher energy transitions more significantly than the nπ^* transitions. We have examined this by including four electronic transitions of the amide chromophore instead of two. Although the two higher energy transitions (n′π^* and $\pi_b\pi^*$) occur at ~140 nm, outside the region 180−250 nm, they may affect the intensities and location of the lower energy bands via coupling. It is possible that such coupling might reduce the intensity of the bands at 193 and 208 nm relative to the band at 220 nm. However, inclusion of the extra transitions in the calculations does not improve the quality of the calculated CD spectra (Table 5 and Figure 6). The intensity of the band around 193 nm decreases and the peak shifts to a lower energy, and concomitantly there is an increase

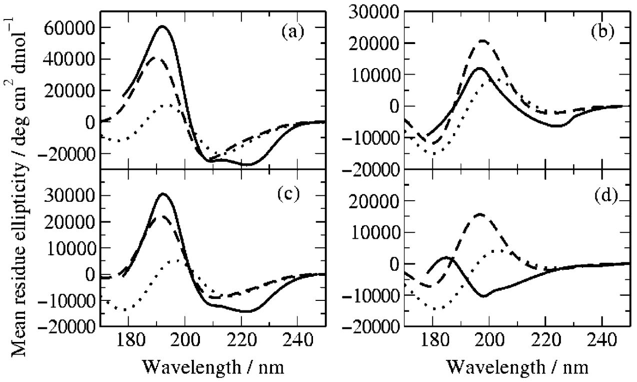

Figure 7. Experimental (solid lines) and calculated CD spectra of (a) myoglobin, (b) concanavalin A, (c) lactate dehydrogenase, and (d) elastase. Calculations use a bandwidth of 10.5 nm and a dielectric constant of 1.2; four amide transitions are included. The spectra calculated with rotations of −10° and −30° to the $\pi_{nb}\pi^*$ and $\pi_b\pi^*$ transitions (dashed line), respectively, are compared to those with no rotation (dotted line).

in the intensity of a negative peak around 182 nm. Such a feature is not apparent in experimental CD spectra,[22] although some protein CD spectra do show a negative peak close to this wavelength; the CD spectrum of myoglobin shows a positive shoulder at 175 nm and a negative peak at 165 nm.[42] The bands around 208 and 220 nm are virtually unaffected by the inclusion of the extra transitions, so the relative rotational strengths of these two bands do not improve. The lack of improvement with four transitions could be due to a number of factors. High-energy transitions are more difficult to describe accurately using ab initio techniques, and perhaps these two higher energy transitions need to be studied more closely. Their interaction with the lower energy transitions may need to be modeled more carefully. Charge transfer transitions in the region 150−180 nm may also need to be accounted for, and we are currently working on this.[43]

The intensity and location of the bands in the calculated far-UV CD spectra of proteins are quite sensitive to the orientation of the electric transition dipole moment, μ, of the amide $\pi_{nb}\pi^*$ transition. Small changes in its orientation significantly alter the interactions between monomers, which are dominated by the dipole−dipole interaction of the amide $\pi_{nb}\pi^*$ transitions, and this in turn has a noticeable effect on the calculated CD spectra. The direction of μ calculated using ab initio methods[24] for NMA differs by 15° from the experimentally determined value[22] advocated by Woody and co-workers in their recent matrix method calculations.[21,37,44] We rotated the ab initio parameters for the amide $\pi_{nb}\pi^*$ transition by −15° to agree with the experimental value and computed CD spectra for the 47 proteins using the two amide transitions, $\pi_{nb}\pi^*$ and nπ^*. The result was to reduce the agreement between the calculated and experimental spectra.

The factors investigated above were combined in a systematic evaluation on the 47 proteins, in which the bandwidth, the dielectric constant, and the orientations of the amide $\pi_{nb}\pi^*$ and $\pi_b\pi^*$ transitions were varied. One of the better combinations, as assessed by the correlation coefficients and the mean absolute error, used a bandwidth of 12.5 nm, a dielectric constant of 1.3, and rotations of −10° and −30° applied to the $\pi_{nb}\pi^*$ and $\pi_b\pi^*$ transitions, respectively. The resultant spectra are shown in Figure 7. The correlation coefficients improve slightly over those achieved without rotation of the parameters for the $\pi_{nb}\pi^*$ and $\pi_b\pi^*$ transitions, and the mean absolute error over the 47 proteins was 6619 deg cm^2 dmol^{-1}, which is comparable to the errors without rotation (Table 5).

Conclusions

The matrix method was used to calculate the rotational strengths for 47 proteins, and Gaussian curves were fitted to these rotational strengths to obtain far-UV CD spectra. The calculated spectra generally reproduce the experimental spectra well, especially for proteins that have a high α-helical content. The CD spectra of β-II proteins are not reproduced well. Reducing the bandwidth used in the calculations leads to better resolution of the two minima in α-helix-rich proteins, but reduces the overall accuracy of the spectra. Introducing a dielectric constant into the calculations attenuates the calculated rotational strengths. Combining the reduced bandwidths with the inclusion of the dielectric constant leads to accurate spectra with a partially resolved double minimum. However, in most spectra the rotational strength of the nπ^* transition is underestimated relative to the $\pi_{nb}\pi^*$ transition. Inclusion of higher energy transitions does not improve the quality of the calculated spectra. Neither does rotation of the orientation of the $\pi_{nb}\pi^*$ transition to the experimentally determined orientation. There is some suggestion that modest reorientation of both the $\pi_{nb}\pi^*$ and $\pi_b\pi^*$ transitions may lead to reasonable calculations with four amide transitions. However, the influence of charge transfer transitions needs to be assessed before this strategy is pursued further.

Overall, the calculated spectra obtained using the matrix method are very promising. The approximations underpinning the matrix method render the calculations quick to perform and yield calculated spectra in reasonable agreement with the experimental data. To achieve even higher accuracy, the parameters used for the higher energy transitions, charge transfer transitions, and transitions located in side chains require further study. The calculated rotational strengths of the nπ^* transitions compared to the $\pi_{nb}\pi^*$ transitions also warrant further investigation. Finally, increased study of β-sheet proteins will hopefully lead to more accurate calculations of their CD spectra in the near future.

Acknowledgment. We thank the Engineering and Physical Science Research Council of the United Kingdom for financial support (grant number GR/N 66971).

References and Notes

(1) Sreerama, N.; Venyaminov, S. Y.; Woody R. W. *Anal. Biochem.* **2000**, *287*, 243.

(2) Sreerama, N.; Woody, R. W. *Anal. Biochem.* **2000**, *287*, 252.

(3) Brahms, S.; Brahms, J. *J. Mol. Biol.* **1980**, *138*, 149.

(4) Sreerama, N.; Woody, R. W. *J. Mol. Biol.* **1994**, *242*, 497.

(5) Provencher, S. W.; Glöckner, J. *Biochemistry* **1981**, *20*, 33.

(6) Hennessey, J. P., Jr.; Johnson, W. C., Jr. *Biochemistry* **1981**, *20*, 1085.

(7) Johnson, W. C., Jr. *Proteins: Struct. Funct. Genet.* **1999**, *35*, 307.

(8) Zhang, C. F.; Lewis, J. W.; Cerpa, R.; Kuntz, I. D.; Kliger, D. S. *J. Phys. Chem.* **1993**, *97*, 5499.

(9) Chen, E. F.; Wood, M. J.; Fink, A. L.; Kliger, D. S. *Biochemistry* **1998**, *37*, 5589.

(10) Goldbeck R. A.; Kim-Shapiro, D. B.; Kliger D. S. *Annu. Rev. Phys. Chem.* **1997**, *48*, 453.

(11) Chen, Y. H.; Yang, J. T.; Martinez, H. M. *Biochemistry* **1972**, *11*, 4120.

(12) Chen, Y. H.; Yang Y. T. *Biochem. Biophys. Res. Commun.* **1971**, *44*, 1285.

(13) Manavalan, P.; Johnson, W. C., Jr. *Nature* **1983**, *305*, 831.

(14) Bode, K. A.; Applequist, J. *J. Am. Chem. Soc.* **1998**, *120*, 10938.

(15) Bayley, P. M.; Nielsen, E. B.; Schellman, J. A. *J. Phys. Chem.* **1969**, *73*, 228.

(16) Woody; R. W.; Tinoco, I., Jr. *J. Chem. Phys.* **1967**, *46*, 4927.

(17) Tinoco, I., Jr. *Adv. Chem. Phys.* **1962**, *4*, 113.

(18) Woody, R. W. *J. Chem. Phys.* **1968**, *49*, 4797.

(19) Besley, N. A.; Hirst, J. D. *J. Am. Chem. Soc.* **1999**, *121*, 9636.

(20) Hirst, J. D.; Besley, N. A. *J. Chem. Phys.* **1999**, *111*, 2846.

(21) Woody, R. W.; Sreerama, N. *J. Chem. Phys*. **1999**, *111*, 2844.

(22) Clark, L. B. *J. Am. Chem. Soc*. **1995**, *117*, 7974.

(23) Koslowski, A.; Sreerama, N.; Woody, R. W. In *Circular Dichroism: Principles and Applications*, 2nd ed.; Berova, N., Nakanashi, K., Woody, R. W., Eds.; Wiley-VCH: New York, 2000.

(24) Besley, N. A.; Hirst, J. D. *J. Phys. Chem. A* **1998**, *102*, 10791.

(25) Besley, N. A.; Hirst, J. D. *J. Mol. Struct*. **2000**, *506*, 161.

(26) Schutz C. N.; Warshel, A. *Proteins: Struct. Funct. Genet*. **2001**, *44*, 400.

(27) King, G.; Frederick, S. L.; Warshel, A. *J. Chem. Phys.* **1991**, *95*, 4366.

(28) Condon, E. U.; Altar, W.; Eyring, H. *J. Chem. Phys.* **1937**, *5*, 753.

(29) Rosenfeld, L. *Z. Phys.* **1928**, *52*, 161.

(30) Goux, W. J.; Hooker, T. M., Jr. *J. Am. Chem. Soc.* **1980**, *102*, 7080.

(31) Koslowski, A. Masters Thesis, RWTH, Aachen, Germany, 1994.

(32) Kurapkat, G.; Krüger, P.; Wollmer, A.; Fleischhauer, J.; Kramer, B.; Zobel, E.; Koslowski, A.; Botterweck, H.; Woody, R. W. *Biopolymers* **1997**, *41*, 267.

(33) Grishina, I. B.; Woody, R. W. *Faraday Discuss.* **1994**, 99, 245.

(34) Manning, M. C.; Woody, R. W. *Biochemistry* **1989**, *28*, 8609.

(35) Sreerama, N.; Manning, M. C.; Powers, M. E.; Zhang, J. X.; Goldenberg, D. P.; Woody, R. W. *Biochemistry* **1999**, *38*, 10814.

(36) Chakrabartty, A.; Kortemme, T.; Padmanabhan, S.; Baldwin, R. L. *Biochemistry* **1993**, *32*, 5560.

(37) Sreerama, N.; Woody, R. W. In *Circular Dichroism: Principles and Applications*, 2nd ed.; Berova, N., Nakanashi, K., Woody, R. W., Eds.; Wiley-VCH: New York, 2000.

(38) Hirst, J. D. *J. Chem. Phys*. **1998**, *109*, 782.

(39) Glätti, A.; Daura, X.; Seebach, D.; van Gunsteren, W. F. *J. Am. Chem. Soc*. **2002**, *124*, 12972.

(40) Sreerama, N.; Woody, R. W. *Protein Sci*. **2003**, *12*, 384.

(41) Hirst, J. D.; Bhattacharjee, S.; Onufriev, A. V. *Faraday Discuss.* **2003**, *122*, 253.

(42) Wallace, B. A.; Janes, R. W. *Curr. Opin. Chem. Biol.* **2001**, *5*, 567.

(43) Gilbert, A. T. B.; Hirst, J. D. *J. Mol. Struct. (THEOCHEM)*, in press.

(44) Chin, D.-H.; Woody, R. W.; Rohl, C. A.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15416.