Theoretical Studies toward Quantitative Protein Circular Dichroism Calculations

Nicholas A. Besley[†] and Jonathan D. Hirst*[,†]

Contribution from the Department of Molecular Biology, TPC-6, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received March 1, 1999. Revised Manuscript Received August 16, 1999

Abstract: Proteins have characteristic circular dichroism spectra in the far-ultraviolet, depending on their secondary structure content. Perhaps the most distinctive spectrum is that of α-helical proteins, with an intense positive band centered about 190 nm and a negative, double-peaked band with minima at 208 and 220 nm. Traditionally, calculations of such spectra from first principles have involved parametrizations of the charge distributions associated with the electronic states and transitions of the constituent chromophoric groups. The amide group is the most important of these chromophores. In this study, using solution phase ab initio parametrizations of the amide chromophore, we present first-principles calculations of protein circular dichroism. Over a set of 29 proteins, there is a significant correlation between the calculated and measured intensities at 190, 208, and 220 nm. The agreement is highest at 220 nm, with a Spearman rank correlation coefficient of 0.90. This near-quantitative accuracy has allowed us to investigate, with some confidence, the dependence of the intensity at 220 nm on helix length and backbone conformation for a number of real and model helices. In this study, a better understanding of the electronic structure of amides and improved calculations and parametrizations of the relevant charge distributions has led to significantly more accurate protein circular dichroism calculations.

Introduction

Circular dichroism (CD) spectroscopy is an important technique in the study of proteins.[1,2] In particular, far-UV CD spectra yield information on protein secondary structure. Increasingly, time-resolved methods are used to study peptide and protein folding[3] and time-resolved far-UV CD measurements have been developed on the microsecond[4] and nanosecond[5] time scales. Such experiments have the time-resolution to follow early events in protein folding. Furthermore, the nanosecond time scale is becoming increasingly accessible to molecular dynamics simulations.[6] This convergence of experiment and theory stimulates the need for accurate calculations of protein CD spectra to provide greater insight into the relationship between protein conformation and the associated spectra.

The major elements of protein secondary structure, α-helix, β-sheet, and coil, all have characteristic CD spectra. The most distinctive of these is perhaps that of the α-helix. A typical α-helical protein has a spectrum that consists of an intense positive band at 190 nm and negative bands at 208 and 220 nm. The bands at 190 and 208 nm arise from the exciton splitting of the amide $\pi_{nb}\pi^*$ (nonbonding π orbital to π^* orbital) transition.[7,8] All electronic transitions can, in principle, contribute to the CD spectrum and another important transition is the magnetically allowed amide $n\pi^*$ (lone pair on oxygen to π^* orbital) transition. This results in the band at 220 nm, which has been observed to correlate with helical content of the protein.[9] Electronic transitions of higher energy can influence the far-UV CD spectra through coupling with the $n\pi^*$ and $\pi_{nb}\pi^*$ transitions.

While measuring the CD spectra of proteins is routine, the calculation of electronic CD spectra has challenged theoreticians for a number of years. Several approaches have been devised. In the dipole interaction model of Applequist,[10] the CD spectrum is calculated by considering individual atoms and the amide chromophore as point dipole oscillators. These interact through mutually induced dipole moments in the presence of an electric field. A number of systems have been studied,[11−14] based on the $\pi_{nb}\pi^*$ transition and an empirical optimization of the polarizabilities of the amide chromophore.[11] Recently, this model was tested on a set of 16 proteins.[15] In another recent application of polarizability theory, a cyclic pentapeptide was studied considering both the $\pi_{nb}\pi^*$ and $n\pi^*$ transitions.[16]

[†] Current address: The School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom.

(1) Nakanishi, K.; Berova, N.; Woody, R. W. *Circular Dichroism Principles and Applications*; VCH: New York, 1994.
(2) Fasman, G. D. In *Circular Dichroism and the Conformational Analysis of Biomolecules*; Fasman, G. D., Ed.; Plenum Press: New York, 1996.
(3) Plaxco, K. W.; Dobson, C. M. *Curr. Opin. Struct. Biol.* **1996**, *6*, 630−636.
(4) Chen, E.; Wittung-Stafshede, P.; Kliger, D. S. *J. Am. Chem. Soc.* **1999**, *121*, 3811−3817.
(5) Zang, C.-F.; Lewis, J. W.; Cerpa, R.; Kuntz, I. D.; Kliger, D. S. *J. Phys. Chem.* **1993**, *97*, 5499−5505.
(6) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740−744.
(7) Moffitt, W. *J. Chem. Phys.* **1956**, *25*, 467−478.
(8) Moffitt, W.; Fitts, D. D.; Kirkwood, J. D. *Proc. Natl. Acad. Sci. U.S.A.* **1957**, *43*, 723−730.
(9) Chen, Y.-H.; Yang, J. T.; Martinez, H. M. *Biochemistry* **1972**, *11*, 4120−4131.
(10) Applequist, J.; Sundberg, K. R.; Olsen, M. L.; Weiss, L. C. *J. Chem. Phys.* **1979**, *70*, 1240−1246.
(11) Bode, K. A.; Applequist, J. *J. Phys. Chem.* **1996**, *100*, 17825−17834.
(12) Bode, K. A.; Applequist, J. *Macromolecules* **1997**, *30*, 2144−2150.
(13) Bode, K. A.; Applequist, J. *Biopolymers* **1997**, *42*, 855−860.
(14) Applequist, J.; Bode, K. A.; Apella, D. H.; Christianson, L. A.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 4891−4892.
(15) Bode, K. A.; Applequist, J. *J. Am. Chem. Soc.* **1998**, *120*, 10938−10946.

10.1021/ja990627l CCC: $18.00
Published on Web 10/01/1999

Several recent papers[17−20] have reported calculations of the molar optical rotation angle based on ab initio methods. Comparison of calculated optical rotation angles with those from experiment has enabled the determination of the absolute stereochemistry of several molecules. In one study, hennoxazole was divided into three weakly interacting fragments,[17] and another study[18] proposed a method that determines individual atomic contributions to the optical rotation. Larger systems, such as proteins, have not been studied.

The calculations presented in this study are based on the matrix method.[21,22] This approach has been used to study a number of proteins[23−25] and related systems.[26] While these calculations reproduce the general features of the CD spectra, quantitative agreement between experiment and theory has not been achieved. The matrix method requires parameters that describe the charge distributions associated with the different electronic states of the chromophoric groups of the protein. This parametrization is crucial to the success of this approach. In earlier work, parameters describing the *N*-methylacetamide (NMA) and acetamide chromophores derived from the semiempirical complete neglect of differential overlap (CNDO/S) method have been used. In a recent study,[27] new parameters were presented for NMA based on gas-phase multireference configuration interaction (MRCI) ab initio calculations.[28] When tested on a set of 23 different proteins, an improved correlation between the calculated and experimental intensity at 220 nm was found.

The process of improving the parametrization of the amide chromophoric group has not been exhausted. In this work we introduce two new advances. First, recognizing that protein CD spectra are measured in solution and the protein itself is, in some sense, a condensed phase, we base the new parameters on ab initio calculations of NMA in solution. Second, the representation of the charge distribution is improved. The interaction between chromophores is approximated in the matrix method by the electrostatic interaction between charge densities. These interactions have previously been evaluated using the monopole−monopole approximation,[29] where charge densities are represented by a set of point charges (or monopoles) that may be derived in a variety of ways. In our previous study,[27] the point charges describing the various charge distributions were derived by reproduction of the appropriate dipole and quadrupole moments. In this study, the point charges are fitted to reproduce the ab initio electrostatic potential arising from the various states. This eliminates a significant approximation in matrix method calculations. In what follows, we describe the matrix method and the origin of our new parameters in more detail. The new parametrization is tested against a set of experimental CD spectra and then used to investigate the CD of a number of real and model helices. We conclude with a broader discussion of the CD of β-strands.

Methods

CD is the differential absorption of left and right circularly polarized light or the difference between the respective extinction coefficients:

$$\Delta\epsilon = \epsilon_L - \epsilon_R \tag{1}$$

CD can also be given in terms of ellipticity, as reported in this study. The differential absorption results in the elliptical polarization of the transmitted light to a degree directly proportional to the CD. Another useful measure is the rotational strength of a transition $A \leftarrow 0$, R_{0A}, which represents the integrated intensity beneath a single band in the CD spectrum:[30]

$$R_{0A} \propto \int \frac{\Delta\epsilon}{\lambda}\, \mathrm{d}\lambda \tag{2}$$

The rotational strength can be expressed as the imaginary part of the product of the electronic and magnetic transition dipole moments using the Rosenfeld equation:[31]

$$R_{0A} = \mathrm{Im}(\langle\psi_0|\vec{\mu}_e|\psi_A\rangle\cdot\langle\psi_A|\vec{\mu}_m|\psi_0\rangle) \tag{3}$$

For small systems, the electronic and magnetic dipole moments can be computed ab initio. However, for large systems, such as proteins, this is computationally prohibitive. The matrix method[21,22] provides a framework for computing the CD of large systems. In the matrix method, the protein is considered to consist of M chromophoric groups. The total wave function is expressed as a linear superposition of basis functions Φ_{ia}. Each basis function is a product of M monomer wave functions. Electronic excitations may occur within a group but not between groups. The basis is further restricted to allow only one group to be excited. Thus:

$$\Phi_{ia} = \varphi_{10}...\varphi_{ia}...\varphi_{j0}...\varphi_{M0} \tag{4}$$

where φ_{ia} represents the wave function of chromophore i, which has undergone an electronic excitation $a \leftarrow 0$. The excited-state wave functions of the whole molecule, Ψ_k, can each then be written as a linear superposition of these basis functions, involving the n_i excitations within each chromophoric group, i:

$$\psi_k = \sum_i^M \sum_a^{n_i} c_{\mathrm{ia}}\Phi_{ia} \tag{5}$$

In general, each transition from the ground state to one of the excited states may have a nonzero rotational strength at its particular transition energy, and the CD spectrum is the sum of all these rotational strengths. By considering the electronic Hamiltonian of the system:

$$\hat{\mathbf{H}} = \sum_{i=1}^{M}\hat{\mathbf{H}}_i + \sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\hat{\mathbf{V}}_{ij} \tag{6}$$

a Hamiltonian matrix is constructed in which the diagonal elements are given by the energies of the electronically excited states and the

(16) Ito, H. *J. Chem. Phys.* **1998**, *108*, 93−108.
(17) Kondru, R. K.; Wipf, P.; Beraton, D. N. *J. Am. Chem. Soc.* **1998**, *120*, 2204−2205.
(18) Kondru, R. K.; Wipf, P.; Beraton, D. N. *Science* **1998**, *282*, 2247−2250.
(19) Polavarapu, P. L.; Chakraborty, D. K. *J. Am. Chem. Soc.* **1998**, *120*, 6160−6164.
(20) Polavarapu, P. L.; Zhao, C. *J. Am. Chem. Soc.* **1999**, *121*, 246−247.
(21) Woody, R. W. *J. Chem. Phys.* **1968**, *49*, 4797−4806.
(22) Bayley, P. M.; Nielsen, E. B.; Schellman, J. A. *J. Phys. Chem.* **1969**, *73*, 228−243.
(23) Manning, M. C.; Woody, R. W. *Biochemistry* **1989**, *28*, 8609−8613.
(24) Kurapkat, G.; Krüger, P.; Wollmer, A.; Fleischhauer, J.; Kramer, B.; Zobel, E.; Koslowski, A.; Botterweck, H.; Woody, R. W. *Biopolymers* **1997**, *41*, 267−287.
(25) Grishina, I. B.; Woody, R. W. *Faraday Discuss.* **1994**, *99*, 245−262.
(26) Cooper, T. M.; Woody, R. W. *Biopolymers* **1990**, *30*, 657−676.
(27) Hirst, J. D. *J. Chem. Phys.* **1998**, *109*, 782−788.
(28) Hirst, J. D.; Hirst, D. M.; Brooks, C. L., III *J. Phys. Chem. A* **1997**, *101*, 4821−4827.
(29) Tinoco, I. *Adv. Chem. Phys.* **1962**, *4*, 113−160.
(30) Schellman, J. A. *Chem. Rev.* **1975**, *75*, 323−331.
(31) Rosenfeld, L. *Z. Phys.* **1928**, *52*, 161−174.

off-diagonal elements are typically of the form:

$$V_{i0a;j0b} = \int_i \int_j \varphi_{i0}\varphi_{ia}\hat{V}_{ij}\varphi_{j0}\varphi_{jb}\,d\tau_i\,d\tau_j \quad (7)$$

These off-diagonal elements describe the interaction between the different chromophoric groups, while other off-diagonal elements account for the mixing of electronic excitations on a given chromophoric group arising from the field of the rest of the molecule. If the interaction between individual chromophoric groups is assumed to be purely electrostatic in nature, then the off-diagonal elements in eq 7 have the form:

$$V_{i0a;j0b} = \int_{\mathbf{r}_i} \int_{\mathbf{r}_j} \frac{\rho_{i0a}(\mathbf{r}_i)\rho_{j0b}(\mathbf{r}_j)}{4\pi\epsilon_0 r_{i,j}}\,d\tau_i\,d\tau_j \quad (8)$$

where $\rho_{i0a}(\mathbf{r}_i)$ and $\rho_{j0b}(\mathbf{r}_j)$ represent permanent and transition electron densities on chromphores *i* and *j*, respectively. The Hamiltonian matrix is then diagonalized by a unitary transformation. The diagonal elements of the new matrix are the excited state energies of the interacting system. The electric and magnetic dipole moments describing the localized transitions can also be transformed to the interacting system by the same unitary transformation. From these, the rotational strengths in the interacting system are readily calculated.

In this procedure, a crucial step is the evaluation of the matrix elements of the original Hamiltonian matrix. In the current study, the CD spectra are calculated based on the amide group in the protein backbone, although side-chain chromophores do also contribute to the CD spectra.[23,32−34] Here we focus on improving the amide parameters. To generate analogous parameters describing all the side-chain chromophoric groups is beyond the scope of this study. Previous calculations and the observed correlation of CD spectra with protein secondary structure suggest that the backbone chromophores are the primary determinant of protein CD spectra in the far-UV. NMA provides a model of the amide chromophore in the protein backbone. To construct the Hamiltonian matrix, the permanent and transition densities associated with the ground and excited states of NMA must be described. This is a problem well suited to modern ab initio techniques.

Ab Initio Calculations

It has been shown that improving the quality of the parametrization results in more accurate calculations of CD spectra.[27] To date, parameters have been derived from gas-phase calculations. Experimental studies[35,36] have shown that solvation has a large influence on the electronic structure of amides. It follows that failure to account for solvation in the parametrization of the amide chromophore may lead to significant errors in the calculated CD spectra. We have used the complete-active-space self-consistent-field method implemented within a self-consistent reaction field (CASSCF/SCRF)[37−40] combined with multi-configurational second-order perturbation theory (CASPT2-RF)[39,40] to calculate the electronic spectrum of NMA in solution. The MOLCAS4[41] program package was used. These calculations are described in detail elsewhere,[42] but we review the findings. In the gas phase, the electronic spectrum of NMA features transitions to Rydberg states. However, in solution the diffuse Rydberg states were found to be destabilized by the Pauli repulsion of solvent, resulting in an electronic spectrum consisting of transitions to valence states. The spectrum is dominated by a broad $\pi_{nb}\pi^*$ transition that undergoes a red-shift in solution. The intensity of the $\pi_{nb}\pi^*$ transition in solution is lower than in the gas phase. The orientation of the $\pi_{nb}\pi^*$ transition dipole moment also changes. Using the carbonyl C−O bond as a reference axis with negative angles representing rotations away from the nitrogen atom,[43] an 8° rotation from −36° to −44° is observed. This compares well with the experimentally determined values of −41° (myristamide),[44] −35° (propanamide),[43] and −55° (*N*-acetylglycine).[43]

The reaction field calculations reproduce the experimental transition energies well, and may provide the diagonal elements of the Hamiltonian matrix directly. To give better agreement with experimental CD spectra, the $\pi_{nb}\pi^*$ transition energy was adjusted slightly. However, this change was within the error of the ab initio calculation. Traditionally, the off-diagonal elements have been evaluated using the monopole−monopole approximation,[29] where the permanent and transition densities are represented by a number of point charges. Thus the integral in eq 8 becomes a sum over discrete charges:

$$V_{i0a;j0b} = \sum_{s=1}^{N_s}\sum_{t=1}^{N_t}\frac{q_s q_t}{r_{st}} \quad (9)$$

The density $\rho_{i0a}(\mathbf{r}_i)$ is represented by N_s charges q_s and similarly $\rho_{j0b}(\mathbf{r}_j)$ is represented by N_t charges q_t.

In the first model presented in this paper, charges that reproduced the dipole and quadrupole moments of the distributions are used. Six charges represent each charge density. Two charges reproduce the dipole and four reproduce the quadrupole. This is similar to a previous calculation using the matrix method,[27] except here solution-phase ab initio calculations have been used. In the second model presented, we use the same solution phase calculations, but improve the derivation of the charges. The role of the fitted charges is to model the ab initio electrostatic potential, which is not necessarily well reproduced by a multipole expansion to the second-order moment. The most direct approach is to evaluate the electrostatic potential and then fit charges to reproduce this potential. In this study we have implemented this approach, the technical details of which are now discussed.

Calculation and Fitting of the Electrostatic Potential

The electronic component of the electrostatic potential between two states described by wave functions:

$$\psi_1 = \sum_\mu C_\mu^a\,\Phi_\mu^a$$
$$\psi_2 = \sum_\mu C_\mu^b\,\Phi_\mu^b \quad (10)$$

in which Φ_μ^a represents a configuration state function constructed from orbitals χ_p^a (and similarly for index *b*), is given by a product of the density matrix, Γ, with a matrix, **f**, whose

(32) Woody, R. W. *J. Polym. Sci. Macromol. Rev.* **1977**, *12*, 181−321.
(33) Woody, R. W. *Biopolymers* **1978**, *17*, 1451−1467.
(34) Chakrabartty, A.; Kortemme, T.; Padmanabhan, S.; Baldwin, R. L. *Biochemistry* **1993**, *32*, 5560−5565.
(35) Nielsen, E. B.; Schellman, J. A. *J. Phys. Chem.* **1967**, *71*, 2297−2304.
(36) Basch, H.; Robin, M. B.; Kuebler, N. A. *J. Chem. Phys.* **1967**, *47*, 1201−1210.
(37) Karlström, G. *J. Phys. Chem.* **1988**, *92*, 1315−1318.
(38) Karlström, G. *J. Phys. Chem.* **1989**, *93*, 4952−4955.
(39) Bernhardsson, A.; Lindh, R.; Karlström, G.; Roos, B. O. *Chem. Phys. Lett.* **1996**, *151*, 141−149.
(40) Serrano-Andrés, L.; Fülscher, M. P.; Karlström, G. *Int. J. Quantum Chem.* **1997**, *65*, 167−181.
(41) Andersson, K.; Blomberg, M. R. A.; Fülscher, M. P.; Karlström, G.; Lindh, R.; Malmqvist, P.-Å.; Neogrády, P.; Olsen, J.; Roos, B. O.; Sadlej, A. J.; Schütz, M.; Seijo, L.; Serrano-Andrés; Siegbahn, P. E. M.; Widmark, P.-O. *MOLCAS Version 4*; Lund University, Sweden, 1997.
(42) Besley, N. A.; Hirst, J. D. *J. Phys. Chem. A* **1998**, *102*, 10791−10797.
(43) Clark, L. B. *J. Am. Chem. Soc.* **1995**, *117*, 7974−7986.
(44) Peterson, D. L.; Simpson, W. T. *J. Am. Chem. Soc.* **1957**, *79*, 2375−2382.

Table 1. Proteins Used To Evaluate Parameters

class	protein (PDB code)
all-α	cytochrome *c* (3cyt), hemoglobin (1hco), myoglobin (1mbn), bacteriorhodopsin (2brd)
mixed α, β (mainly α)	alcohol dehydrogenase (5adh), glutathione reductase (3grs), lactate dehydrogenase (6ldh), lysozyme (7lyz), papain (9pap), rhodanese (1rhd), subtilisin (1sbt), thermolysin (4tln), triose phosphate isomerase (1tim), flavodoxin (2fx2)
β (class I)	carbonic anhydrase (1ca2), concanavalin A (3cna), λ-immunoglobulin (1rei), ribonuclease A (3rn3), ribonuclease S (2rns), erabutoxin (3ebx), plastocyanin (1plc), porin (3por), prealbumin (2pab)
β (class II)	α-chymotrypsinogen A (2cga), α-chymotrypsin II (5cha), elastase (3est), superoxide dismutase (2sod), trypsin inhibitor (4pti), trypsin (3ptn)

elements involve the orbitals and the operator for the property of interest:[45]

$$\left\langle \psi_1 \left| \frac{1}{r} \right| \psi_2 \right\rangle = \sum_{p,q} \Gamma_{pq}^{ab} f_{pq}^{ab} \tag{11}$$

where

$$\Gamma_{pq}^{ab} = \sum_{\mu,v} A_{pq}^{\mu v} C_{\mu}^{a} C_{v}^{b} \tag{12}$$

$$f_{pq}^{ab} = \left\langle \chi_p^a \left| \frac{1}{r} \right| \chi_p^b \right\rangle \tag{13}$$

and *A* values are the coupling coefficients. In our implementation, the permanent and transition density matrices produced in the RASSI program[41,45] are imported into our modified version of the PROPERTY program in the MOLPRO96[46] suite of ab initio programs. This enables us to calculate the electronic contribution to the electrostatic potential at a given point for permanent and transition densities (for wave functions of all symmetries) obtained from our solution-phase calculations. The total electrostatic potential is then evaluated by addition of the appropriate nuclear component.

The electrostatic potential was evaluated at 0.5 au intervals on a 7 × 7 × 7 au grid around the NMA molecule. This is a total of 24389 points. A smaller grid in which data points were spaced by 1.0 au led to similar results. Grid points that lay within the van der Waals radius of an atom were removed from the data set. van der Waals radii of 1.7, 1.55, and 1.2 Å were assumed for carbon, nitrogen, and hydrogen, respectively, based on the work of Bondi.[47] A value of 1.4 Å was used for oxygen following the suggestion of Tuñún et al.[48] The removal of these points prevents biasing toward the nuclear potential, which was a concern particularly for the permanent densities. Furthermore, these regions are not important, because atoms within the protein do not become this close. The charges are then fitted to reproduce this electrostatic potential such that the least-squares difference is minimized.[49]

The carbon, oxygen, nitrogen, and amide hydrogen atoms of the peptide group were used as centers for the location of charges. For densities involving states of the same symmetry (i.e. two A′ or two A″ states, as the calculations assumed C_s symmetry) five charges were located on each atom, giving a total of 20 charges. These charges were arranged in the symmetry plane of the molecule, with one central charge and two orthogonal pairs of dipoles. For densities involving wave functions of two different symmetries, a different arrangement of charges was used because the symmetry plane is a nodal plane in the electrostatic potential. Eight charges were located around each atom center, arranged such that they occupied the corners of a cube centered around the atom site, giving a total of 32 charges. The charges were adjusted such that the overall charge of the system is conserved. The overall quality of the fit was good, with typically a 5% error on a given data point. Using this procedure, charges were obtained for the permanent densities of the five lowest valence states (ground state, nπ*, $\pi_{nb}\pi^*$, n′π*, and $\pi_b\pi^*$), in addition to the respective transition densities between the states. In this study only the ground, nπ*, and $\pi_{nb}\pi^*$ states are used in the CD calculations. Although the computational cost of the CD calculation is proportional to the square of the number of charges, it is sufficiently fast that the increase in the number of charges does not present a problem. On a present day workstation the CD calculations take a few minutes.

CD spectra were then calculated for a set of 29 proteins and compared with experiment. The experimental CD spectra of 23 proteins were determined by Pancoska et al.,[50] and to facilitate comparison with other work, these data were augmented by bacteriorhodopsin,[2] porin,[2] prealbumin,[51] flavodoxin,[51] erabutoxin,[52] and plastocyanin.[53] The spectra were calculated from the crystal structures from the Protein Data Bank (PDB).[54] This test set of proteins, shown in Table 1, encompasses a broad range, from highly helical proteins to those that are largely β-sheet.

Results and Discussion

Calculations using the matrix method yield rotational strengths. To produce a CD spectrum, a functional form has to be assumed for each band, with the area beneath a given band proportional to the rotational strength. Usually Lorentzian or Gaussian functions are used,[30] introducing a parameter, the bandwidth, that probably lies between 9 and 18 nm.[35] CD spectra are sensitive to the bandwidth, with a small bandwidth making all peaks narrower and higher, while a larger bandwidth makes peaks lower and wider. However, this does not alter the underlying correlation of the calculated rotational strengths with experiment. In this study, we have used Gaussian functions with a single bandwidth for all peaks. A bandwidth of 15.5 nm was chosen based on a comparison of the calculated and experi-

(45) Malmqvist, P.-Å.; Roos, B. O. *Chem. Phys. Lett.* **1989**, *155*, 189−194.
(46) MOLPRO is a package of ab initio programs written by H.-J. Werner and P. J. Knowles, with contributions from R. D. Amos, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, T. Leininger, R. Lindh, A. W. Lloyd, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, P. Palmieri, K. Peterson, R. Pitzer, P. Pulay, G. Rauhut, M. Schütz, H. Stoll, A. J. Stone, and T. Thorsteinsson.
(47) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441−451.
(48) Tuñún, I.; Silla, E.; Pascual-Ahuir, J.-L. *J. Am. Chem. Soc.* **1993**, *115*, 2226−2230.
(49) Press: W. H.; Teukolsky, S. A.; Vettering, W. T.; Flannery, B. P. *Numerical Recipes*; Cambridge University Press: Cambridge, 1992.
(50) Pancoska, P.; Bitto, E.; Janota, V.; Urbanova, M.; Gupta, V. P.; Keiderling, T. A. *Protein Sci.* **1995**, *4*, 1384−1401.
(51) Brahms, S.; Brahms, J. *J. Mol. Biol.* **1980**, *138*, 149−178.
(52) Dahms, T. E.; Szabo, A. G. *Biophys. J.* **1995**, *69*, 569−576.
(53) Draheim, J. E.; Anderson, G. P.; Duane, J. W.; Gross, E. L. *Biophys. J.* **1986**, *49*, 891−900.
(54) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535−542.

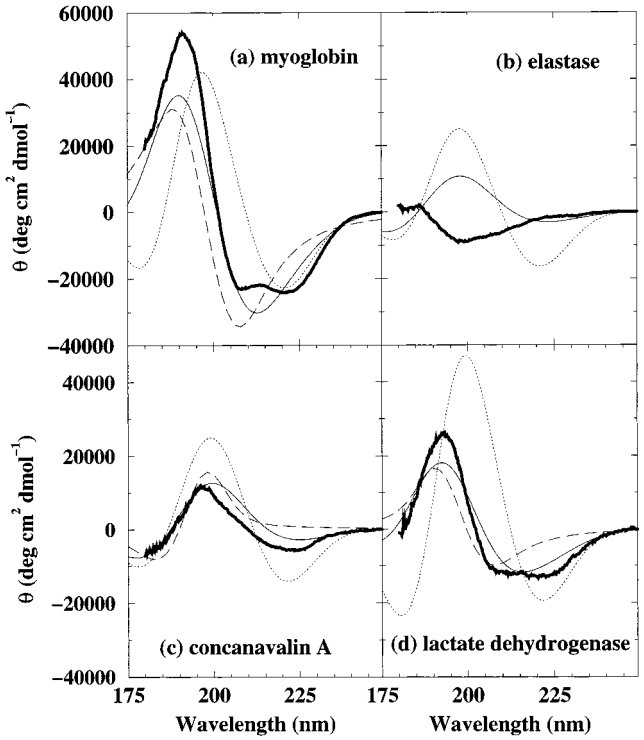

Figure 1. Calculated and experimental CD spectra: bold solid line, experiment; solid line, CASSCF/SCRF/ESP; dashed line, dipole interaction model;[15] and dotted line, CNDO/S paramters.[24]

mental spectra for the 29 proteins. Plots comparing experimental and calculated spectra for all 29 proteins are available in the Supporting Information. Figure 1 shows illustrative examples of the four different types of CD spectra.

The first type of protein is α-helical, for example, myoglobin. The experimental spectrum is typical of an α-helical protein, with an intense positive band at 190 nm and a minimum at 208 nm associated with the $\pi_{nb}\pi^*$ amide transition. The other minimum at 220 nm is associated with the amide $n\pi^*$ transition. The spectrum calculated using the CNDO/S parameters reproduces the intensity at 220 nm. The $\pi_{nb}\pi^*$ band is less well described and the intensities at 190 and 208 nm do not agree with experiment. The dipole interaction model describes the $\pi_{nb}\pi^*$ band more accurately and has a significant intensity at 208 nm. However, above 208 nm theory and experiment begin to diverge, because of the absence of the $n\pi^*$ transition in the model. For our latest parameters, the calculated spectrum agrees fairly well across the range of wavelengths. The intensity at 190 nm is underestimated and the small maximum between 208 and 220 nm is not reproduced. This is in part due to the choice of bandwidth. The second class of proteins, those with significant amounts of α-helices and β-sheets, has a similar spectrum to those that are α-helical, as illustrated by lactate dehydrogenase.

The CD spectra of predominantly β-sheet proteins are much less intense than those of α-helical proteins. They can be divided into two classes.[55] The first type, class I, have a maximum at about 195 nm and a minimum in the range 210−220 nm. An example, concanavalin A, is included in Figure 1. The calculated spectrum derived from the new parameters agrees with experiment qualitatively, reproducing the intensity at 195 nm and the minimum. Class II β-sheet proteins have a significant proportion of β-bulge in their secondary structure, reflecting disorder in the β-sheets. A β-bulge involves the mis-pairing of two strands, with two or more residues in one strand opposite a single residue on an adjacent strand, leading to a bulge.[56,57] The CD spectra of these proteins are very different and resemble the spectrum of a random coil. This is illustrated by elastase, which has a minimum at about 200 nm. The calculated spectra for this type are poor. Above 220 nm there is good agreement with experiment; however, below 220 nm theory predicts a band with the wrong sign. We discuss the source of this discrepancy in a later section.

(55) Manavalan, P.; Johnson, W. C., Jr. *Nature* **1983**, *305*, 831−832.

Table 2. Correlation between Experimental and Calculated CD Spectra

	r		
model	$\lambda = 190$ nm[a]	$\lambda = 208$ nm	$\lambda = 220$ nm
CNDO/S NMA (g)	0.44	−0.14	**0.48**
CNDO/S acetamide (g)	0.08	−0.16	**0.65**
MRCI NMA (g) μ, Q	0.12	−0.24	**0.74**
CASSCF NMA (aq) μ, Q	**0.50**	**0.51**	**0.67**
CASSCF NMA (aq) ESP	**0.84**	**0.73**	**0.90**

[a] Spearman rank correlation coefficient calculated for 29 proteins. Bold: probability of random correlation <0.01.Differences from values reported earlier (ref 27) are due to the larger data set, and to the correction of previous mis-processing of a couple of protein structures.

More generally, both our calculations and the dipole interaction model predict less negative bands at 175 nm than the CNDO/S parameters. It had been thought that there might be a monomer amide transition in this region, partly based on the prediction of the CNDO/S parameters. Both our CD and ab initio calculations indicate that this is not the case. The ab initio calculations place the $n'\pi^*$ transition at 123 nm and the $\pi_b\pi^*$ transition at 129 nm. Furthermore, they show no evidence of $n\sigma^*$ or $\pi\sigma^*$ transitions.

To compare different parameter sets objectively, it is necessary to have a measure of the agreement with the known experimental spectra. In this study, we measure the agreement at the three key wavelengths, 190, 208, and 220 nm. The agreement with the experiment is represented by a Spearman rank correlation coefficient,[49] *r*, which permits one to evaluate the significance of correlation through the Fisher *z* test. Table 2 shows the Spearman rank correlation between experiment and theory for different parameter sets. Where the probability of this correlation occurring by chance is less than 1%, *r* is shown in bold type. For our latest parameters, the probability of a chance correlation is less than 0.001%. This represents a significant agreement between experiment and theory.

Results are shown for five parameter sets. In addition to the two new parameter sets, previous parameters based on CNDO/S calculations on NMA or acetamide[24] and the recent parameters derived from MRCI calculations on NMA[27,28] are included. The labels "(g)" and "(aq)" indicate parameters derived from gas-phase and condensed-phase calculations, respectively. The label "μ, Q" indicates the charges reproduce dipole and quadrupole moments, while "ESP" indicates that the charges were fitted to reproduce the ab initio electrostatic potential. Results are reported for calculations based the two lowest electronic transitions, i.e., $n\pi^*$ and $\pi_{nb}\pi^*$.

Previous parameters only show a significant correlation at 220 nm, which is greatest for the MRCI NMA parameters. Introducing solvent effects produces an improvement in the agreement between experiment and theory. Moderate correlation

(56) Richardson, J. S.; Getzoff, E. D.; Richardson, D. C. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 2574−2578.

(57) Chan, A. W. E.; Hutchinson, E. G.; Harris, D.; Thornton, J. M. *Protein Sci.* **1993**, *2*, 1574−1590.

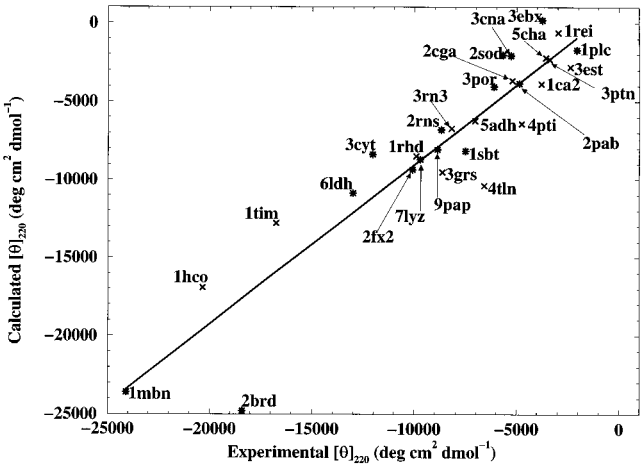

Figure 2. Calculated and experimental mean residue ellipticities at 220 nm. The subset of proteins studied by Bode and Applequist[15] are marked with asterisks.

is observed at 190 and 208 nm, while the correlation at 220 nm is retained. The results for the final parameter set, for which the charges were fitted directly to the ab initio potential, show a highly significant correlation at all three wavelengths. The increase in correlation observed for the "ESP" scheme suggests that there are some deficiencies in the "μ, Q" scheme, as noted in a preliminary comment elsewhere.[58] The calculated CD spectra are sensitive to changes in the parameters describing the charge distributions, and it is important to describe them as accurately as possible.

The agreement between experiment and theory is particularly strong at 220 nm. This is illustrated in Figure 2, which shows the calculated and experimental intensities at 220 nm. The data lie close to the least-squares fit line, which has a gradient of 0.92. The largest error occurs in bacteriorhodopsin. This may be due to distortions in the experimental spectrum arising from particulate scattering.[2] The magnetically allowed nπ* transition will couple mainly with the $\pi_{nb}\pi$* transition. Thus the two-state model used in these calculations should provide a good description of the band at 220 nm. In contrast, the electronically allowed $\pi_{nb}\pi$* transition may couple significantly with a large number of states not included in the two-state model. The inclusion of the higher energy n$'\pi$* and $\pi_b\pi$* transitions had little effect on the calculated correlations. However, they did reduce the intensity of the band at 190 nm.

The accuracy at all wavelengths is shown in Figure 3. The correlation is generally good except in the region 195−210 nm. The error in this region is partly due to the class II β proteins. In addition, the correlation is particularly sensitive in this region, because at these wavelengths the intensity changes from positive to negative, so small shifts in the calculated spectra can result in large errors. The upper panel shows the mean absolute error and mean absolute intensity. The mean absolute error decreases across the spectra. However, the mean absolute intensity shows a marked decrease in the region around 200 nm. The ratio of these two measures gives a normalized mean absolute error (center panel) that shows a large peak around 202 nm.

It is appropriate to compare this work with the recent study of Bode and Applequist,[15] who used the dipole interaction model to calculate the CD spectra of 16 proteins. The Spearman rank correlation coefficients between experiment and theory for 15 proteins (rubredoxin was omitted, as its solvated structure is probably quite different from X-ray structure, as discussed by

(58) Hirst, J. D.; Besley, N. A. *J. Chem. Phys.* **1999**, *111*, 2846−2847.

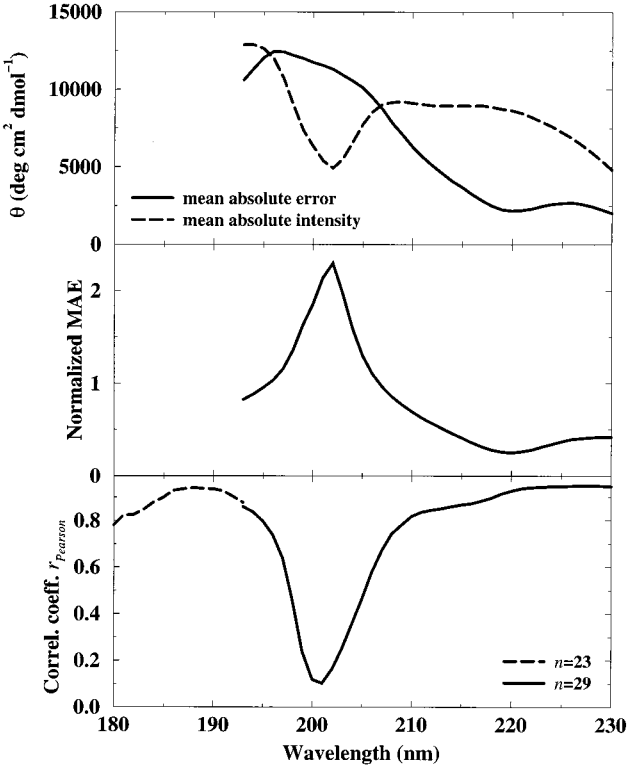

Figure 3. Error in computed intensities. Upper panel: Mean absolute error (solid line) and mean absolute intensity (dashed line). Center panel: Normalized mean absolute error. Lower panel: Pearson correlation coefficient. Data below 195 nm are only available for 23 of the proteins.

Table 3. Comparison with Dipole Interaction Model

	r		
model	$\lambda = 190$ nm[a]	$\lambda = 208$ nm	$\lambda = 220$ nm
dipole interaction model[15]	0.89	0.75	0.74[a]
CASSCF NMA (aq) ESP	0.87	0.71	0.96

[a] The dipole interaction model does not include the nπ* transition, so a high correlation is not anticipated.

Bode and Applequist[15]) examined by both methods are shown in Table 3. For this comparison the spectra from the dipole interaction model calculated at a half-peak bandwidth of 6000 cm^{-1} were used, as overall these appeared to correlate better with experiment. The two models perform comparably at 190 and 208 nm, while our calculations achieve better correlation at 220 nm. The dipole interaction model does not include the nπ* transition, so our calculations, which do, would be expected to perform better in this region.

The data set of 29 proteins contains three groups of proteins that are homologous. These include hemoglobin and myoglobin; ribonuclease A and ribonuclease S; and elastase, trypsin, α-chymotrypsin II, and chymotrypsinogen A. The data set was constructed to allow a direct comparison with many of the previously studied proteins. The original set and corresponding PDB structures of 23 proteins was chosen to match the work of Pancoska et al.,[50] with an additional six proteins added following the work of Bode and Applequist.[15] The homologues could result in a bias in the calculated correlation coefficients. The 16 proteins studied by Bode and Applequist[15] contain no homologues. The results show that a greater correlation is achieved for this second data set. The inclusion of the homologues actually provides an unfavorable bias, because one of

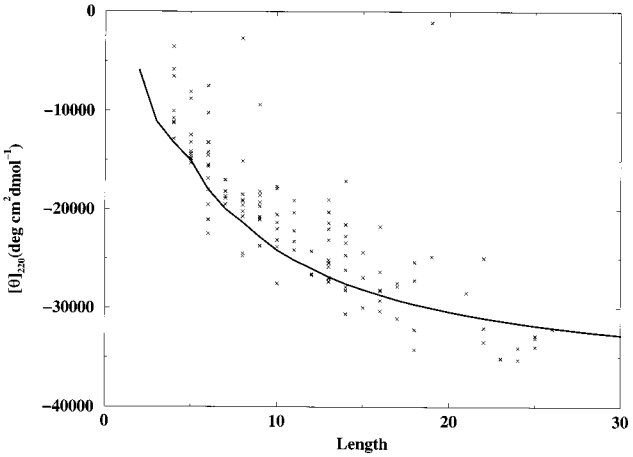

Figure 4. Variation of the mean residue intensity at 220 nm with helix length.

the homologous groups comprises class II β-sheet proteins for which our calculations perform poorly.

Helices

Theory coupled with knowledge of the precise conformation of X-ray structures provides an opportunity to investigate helical fragments of proteins, in a manner not possible by experiment alone. Helical peptides have been studied using CD to assess helical propensities of different amino acids.[59,60] The fullest interpretation of such experiments requires an understanding of the conformational effects on CD. The accuracy of our calculations at 220 nm is nearly quantitative (as shown in Figure 2) and this enables us to study, with some confidence, helical fragments and to examine the relationship between their conformation and the mean residue ellipticity at 220 nm ($[\theta]_{220}$). Helical fragments were extracted from the 29 proteins. Those helices that were severely distorted or had mean ϕ or ψ angles outside of the range $-90° < \phi,\psi < -20°$ were removed from the data set. This gave a set of 124 real helical fragments, varying in length from 4 to 26 peptide groups. The CD spectra of these fragments were then calculated and scaled by a factor of 1.08 (the inverse of the gradient of the line in Figure 2). The variation of $[\theta]_{220}$ with helix length of these fragments and the Pauling−Corey helix[61] is shown in Figure 4. The magnitude of $[\theta]_{220}$ increases with the helix length and slowly converges to an asymptotic value. This has been previously observed in a similar study on model helices.[62] The value for a Pauling−Corey helix of 200 peptide groups can give an estimate of $[\theta]_{220}$ for an infinite helix ($[\theta_{H\infty}]_{220}$). This is found to be −37000 deg cm^2 dmol^{-1}, which is at the lower end of the range of values of experimental estimates.[60,63,64] Our calculations probably underestimate the intensity at 220 nm slightly more than the scaling factor suggests. The real helices follow the same trend with a small spread of values around those of the Pauling−Corey helix.

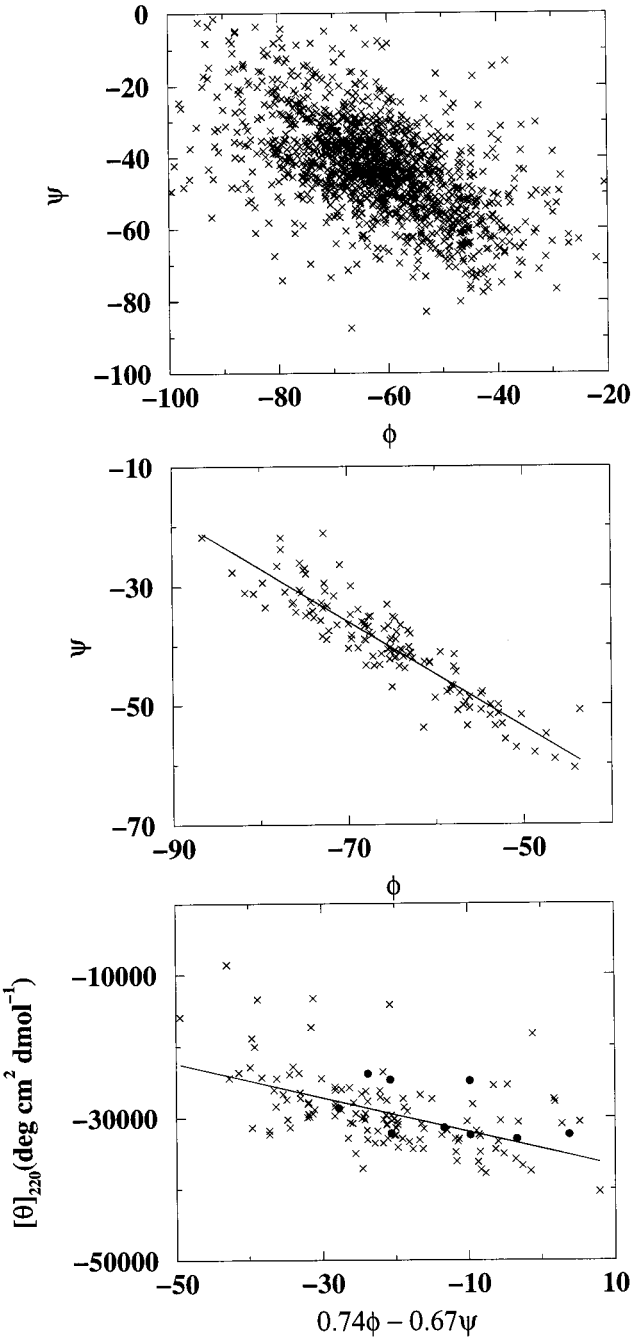

Figure 5. Variation of the mean residue intensity at 220 nm with helix conformation. Upper panel: Distribution of residue dihedral angles in helical fragments. Center panel: Distribution of mean dihedral angles. Lower panel: Variation of $[\theta]_{220}$ with dihedral angle distribution.

This length dependence has been expressed as:[9]

$$[\theta_H]_{220} = [\theta_{H\infty}]_{220}\frac{r-k}{r} \qquad (14)$$

where $[\theta_{H\infty}]_{220}$ represents the mean residue ellipticity of an infinite helix, r is the number of residues in the helix, and k is a constant that accounts for an end-effect. An estimate of k can be made by optimizing both $[\theta_{H\infty}]_{220}$ and k in eq 14 to give a least-squares fit to the calculated $[\theta]_{220}$ of the Pauling−Corey helix. This procedure suggests values of 2.6−3.0 for k.

In addition to length, the CD of a helix may also be sensitive to conformation.[62] Figure 5 (upper panel) shows the distribution of the main chain dihedral angles in the 124 helical fragments.

(59) Platzer, K. E. B.; Ananthanarayanan, V. S.; Andreatta, R. H.; Scheraga, H. A. *Macromolecules* **1972**, *5*, 177−187.
(60) Padmanabhan, S.; Marqusee, S.; Ridgeway, T.; Laue, T. M.; Baldwin, R. L. *Nature* **1990**, *344*, 268−270.
(61) Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, 235−240.
(62) Manning, M. C.; Woody, R. W. *Biopolymers* **1991**, *31*, 569−586.
(63) Chen, Y.-H.; Yang, J. T.; Chau, K. H. *Biochemistry* **1974**, *13*, 3350−3359.
(64) Luo, P.; Baldwin, R. L. *Biochemistry* **1997**, *36*, 8413−8421.

	i		i+1		i+2	
i	$E_{n\pi^*}$					
	CAE	$E_{\pi\pi^*}$				
i+1	0	μ-m	$E_{n\pi^*}$			
	μ-m'	K_{i+1}	CAE	$E_{\pi\pi^*}$		
i+2	0	0	0	μ-m	$E_{n\pi^*}$	
	0	K_{i+2}	μ-m'	K_{i+1}	CAE	$E_{\pi\pi^*}$
i+3	0	0	0	0	0	μ-m
	0	K_{i+3}	0	K_{i+2}	μ-m'	K_{i+1}

Figure 6. Portion of the matrix computed in the matrix method for a helix of 20 peptide groups ($\phi = -57°$, $\psi = -47°$).

The center panel shows the mean dihedral angles of each helical fragment. This distribution, in contrast to the previous, is nearly linear indicating a single coordinate may be used to describe the different helical conformations. The lower panel shows $[\theta]_{220}$ plotted against a new ϕ/ψ coordinate that describes this line. The length dependence has been removed from $[\theta]_{220}$ by normalizing to a Pauling−Corey helix of 26 peptide groups. There is a weak, but nevertheless significant trend, that helices with a more negative ϕ have lower intensity at 220 nm than helices with a more negative ψ. Superimposed are values for polyalanine helices of 20 peptide groups with different conformations, which were studied by Manning and Woody.[62] The geometry of these helices was minimized with the dihedral angles restrained, allowing other degrees of freedom to relax. In particular, it is important to allow the hydrogen bond distances to relax. These points conform to the relationship observed for the helical fragments. The three data points that lie significantly above the line correspond to 3_{10} helices. In short peptides 3_{10}-helical structure may be common.[65−67] Our calculations predict that $[\theta_{H\infty}]_{220}$ is roughly 0.75 times smaller for 3_{10}-helices. Thus simple relationships of the type shown in eq 14 may not be appropriate for the CD of peptides with a significant fraction of 3_{10}-helical structure. Over an entire protein the conformational dependence will tend to average out. However, this conformational dependence is likely to be more important in the CD of smaller systems.

To illustrate the relative importance of different terms in the matrix method, a portion of the (symmetric) matrix computed by the matrix method is shown in Figure 6. The potentially significant interaction terms for the α-helix are labeled, and have been discussed by Bayley et al.[22] These interactions can result in mixing of states within a given group, in addition to coupling of states in different groups. The static coupling corresponding to the mixing of the nπ^* and $\pi\pi^*$ states on a single peptide group is denoted *CAE*, for the mechanism described by Condon, Altar, and Eyring.[68] The analogous dynamic coupling[69] (between neighboring peptide groups) is denoted by μ-m. Coupling between $\pi\pi^*$ states on different peptide groups[8,70] is denoted by K. For an idealized helix comprising 20 peptide groups, with ($\phi = -57°$, $\psi = -47°$), our parameters give the following: $CAE \sim 0$, μ-$m' \sim 0$, μ-$m \sim -280$, $K_{i+1} \sim -150$, $K_{i+2} \sim -640$, $K_{i+3} \sim -370$, and $K_{i+j} \sim 0$ (for $j > 3$), where the units are consistent with energies in inverse centimeters. The matrix is formulated in terms of peptide groups, not residues, so the K_{i+3} term corresponds to the hydrogen bond in the α-helix. Local effects are most important, with the principal terms being the dynamic coupling between the $\pi\pi^*$ transition of peptide group i with the nπ^* transition of peptide group i+1 and the coupling of $\pi\pi^*$ transitions between peptide groups i, i+2 and i, i+3. For the idealized 3_{10}-helix, the dynamic coupling is over a factor of 2 smaller, which leads to a less intense peak at 220 nm compared with the α-helix. The value of the K_{i+2} term is similar for the 3_{10}-helix and the α-helix, but the K_{i+3} term is essentially zero for the 3_{10}-helix. For the idealized α-helix, the earlier CNDO/S parameters[24] lead to matrix elements that are generally of significantly larger magnitude, with the exception of the dynamic coupling terms, which are much smaller. Thus, in contrast to previous work, our parameters suggest that dynamic coupling is much more important that static coupling in the CD of α-helices.

β-Strands

The CD spectra of β-sheet proteins fall into two classes[55,71] Class I β-proteins (β-I) exhibit a negative band at 216−218 nm and a positive band at 195 nm. Class II β-sheet proteins (β-II) have CD spectra that resemble those of random coil models, namely an intense negative band near 198 nm. The β-II proteins tend to contain some disulfide bridges (which may contribute to the CD in the far-UV[72]) and have β-sheets that are more irregular than the β-I proteins, reflected by a larger content of β-bulge structures. However, the structural origins of the CD spectra of β-II proteins are not fully understood. Our calculations of CD are unable to distinguish between β-I and β-II proteins, predicting β-I spectra for both classes. To investigate the CD calculations on β-I and β-II proteins further, we have analyzed the structures of concanavalin A (a β-I protein) and elastase (a β-II protein) in greater detail and we have also calculated the CD of idealized regular β-strands.

Our calculations suggest that protein CD in the far-UV is determined primarily by local backbone structure. This seems reasonable, as short-range interactions between chromophores dominate over long-range interactions and local structure is much more regular and repetitive than long-range structure. Although β-bulges can have atypical dihedral angles, the population of ∼10% of residues in β-bulge conformations in elastase compared to ∼0% in concanavalin A is not sufficient for the distributions of the dihedral angles, ϕ and ψ, of the two proteins to be significantly different. There is little difference between the calculated CD of β-strands excised from elastase and concanavalin A, probably reflecting the similarity of the backbone dihedral angles. The calculations on individual strands also suggest that the calculated CD of these proteins may be crudely approximated by the sum of the CD of the secondary structure fragments. It appears that our CD calculations reflect the local structural similarity of the two proteins, as defined by their static X-ray crystal structures.

Calculations of the CD of idealized β-strands, whose structures were generated in a fashion similar to the idealized α-helices described in the previous section, show that just two

(65) Miick, S. M.; Martinez, G. V.; Fiori, W. R.; Todd, A. P.; Millhauser, G. L. *Nature* **1992**, *359*, 653−655.
(66) Millhauser, G. L. *Biochemistry* **1995**, *34*, 3873−3877.
(67) Sheinerman, F. B.; Brooks, C. L., III *J. Am. Chem. Soc.* **1995**, *117*, 10098−10103.
(68) Condon, E. U.; Altar, W.; Eyring, H. *J. Chem. Phys.* **1937**, *5*, 753−777.
(69) Kuhn, W. *Trans. Faraday Soc.* **1930**, *26*, 293−308.
(70) Kirkwood, J. G. *J. Chem. Phys.* **1937**, *5*, 479−491.
(71) Wu, J.; Yang, J. T.; Wu, C.-S. C. *Anal. Biochem.* **1992**, *200*, 359−364.
(72) Woody, R. W.; Dunker, K. A. In *Circular Dichroism and the Conformational Analysis of Biomolecules*; Fasman, G. D., Ed.; Plenum Press: New York, 1996.

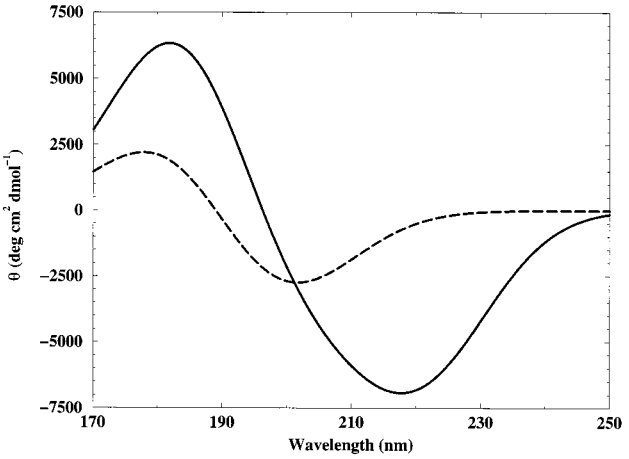

Figure 7. Calculated CD of an ideal β-strand: solid line, dynamic coupling = 300 cm^{-1}; dashed line, dynamic coupling = 0 cm^{-1}. In both cases K_{i+1} = 900 cm^{-1} and all other interactions are zero.

local interactions are dominant: the interaction between nearest neighbor $\pi\pi^*$ transitions, $K_{i+1} \sim -900$ cm^{-1}, and the dynamic mixing term, μ-$m \sim 300$ cm^{-1}. The former is large and negative, as might be anticipated from β-strand conformations, which place neighboring $\pi\pi^*$ transitions almost in the same plane and pointing in almost opposite directions. The dynamic mixing term is almost independent of ϕ in the β-strand region of conformational space, but changes quite steeply with ψ. As ψ changes from ~140° to ~170°, the dynamic mixing term falls from 300 to 0. A similar change is seen as ψ goes from ~30° to ~0°. Strikingly, the location of the calculated negative band moves from 218 to 200 nm, as the dynamic mixing term falls from 300 to 0 (Figure 7), leading to a calculated CD spectrum that is much closer to the β-II spectrum. It may be that our calculations do not reproduce the dynamic mixing term sufficiently accurately, but it is also conceivable that in solution there are relatively minor fluctuations of 30° in backbone dihedral angles which lead to structures whose calculated CD spectra would be much closer to those observed experimentally.

Conclusions

In this work we have reported improvements in the calculation of protein CD spectra using the matrix method, by recalculating the parameters which describe the electronic transitions on the amide chromophore. Including effects of solvation and providing a more accurate representation of the charge distributions improves agreement with a set of experimental CD spectra. In particular, significant correlation between experiment and theory is found at the three key wavelengths of 190, 208, and 220 nm. The agreement with experiment at 220 nm, which is almost quantitative, is encouraging and important because CD measurements at this wavelength are used to assess the helical content of peptides and proteins. However, there remains one class of proteins whose CD our calculations do not describe adequately. The calculated CD of these class II β-type proteins are largely dependent on the dynamic coupling term of the matrix method. This dynamic term is sensitive to conformation and relatively small differences between the crystal structure and the solvated structure may be partly responsible for the discrepancy between calculated and experimental CD.

The complex relationship between the parameters and the computed CD makes it difficult to isolate the advantageous properties of a given parameter set. However, it is clear that it is important to describe the $\pi_{nb}\pi^*$ intensity and transition dipole orientation accurately. Very recently, CNDO/S parameters combined with experimental data of the $\pi_{nb}\pi^*$ transition have been used to compute CD spectra with some success.[73]

There are several ways in which our calculations of protein CD may be further improved. First, our calculations only consider the protein backbone. Side-chain chromophoric groups can contribute to the CD in the far-UV. In principle, the inclusion of side chains into the matrix method is straightforward through parameters describing the side-chain groups. Using existing CNDO/S parameters for side chains did not improve the calculated spectra.[27] The calculation of parameters at the CASSCF/SCRF level is a considerable undertaking, and beyond the present study. The effect of nonchromophoric groups is also likely to have a significant role, through coupling with the $\pi_{nb}\pi^*$ transition.[15] In principle, all possible electronic transitions should be included. The calculations reported here have been limited to electronic transitions to the nπ^* and $\pi_{nb}\pi^*$ states of the backbone amide groups, but the inclusion of additional $\pi_b\pi^*$ and n′π^* amide transitions did not significantly affect the correlation with experiment. Additional effects that may be of importance include protein dynamics.[74]

While work remains to be done before fully quantitative calculations of protein CD spectra are readily performed, the attraction of the matrix method is clear. First, the calculations are relatively straightforward to perform. The entire protein structure is used in the calculations without modification. Second, the full potential of the method has yet to be realized, and we are currently addressing some of the limitations outlined above. Nevertheless, the current calculations are sufficiently reliable and accurate to allow some questions concerning the relationship between protein structure and CD to be addressed. This has been demonstrated by our study of real and model helices, which has reexamined the dependence of CD on helical length and conformation. Investigations of this type illustrate the practical value of calculations from first principles.

Acknowledgment. Financial support from the National Science Foundation (Grant No. MCB-9632124) is gratefully acknowledged. Acknowledgment is made to the donors of the Petroleum Research Fund, administered by the American Chemical Society, for partial support of this work. The CD calculations were performed using a modified version of code generously provided by Professor Jörg Fleischhauer at RWTH Aachen, Germany. We also thank Professor Jon Applequist (Iowa State University) for sharing his data.

Supporting Information Available: Calculated and experimental CD spectra for the 29 proteins (PDF). Parameters used to describe the various charge distributions of the amide chromophore (ASCII). This material is available free of charge via the Internet at http://pubs.acs.org.

JA990627L

(73) Woody, R. W.; Sreerama, N. *J. Chem. Phys.* **1999**, *111*, 2844−2845.
(74) Hirst, J. D.; Brooks, C. L., III *J. Mol. Biol.* **1994**, *243*, 173−178.