Globular Protein Ultraviolet Circular Dichroic Spectra. Calculation from Crystal Structures via the Dipole Interaction Model

Kimberly A. Bode and Jon Applequist*

Contribution from the Department of Biochemistry and Biophysics, Iowa State University, Ames, Iowa 50011

Received July 16, 1998. Revised Manuscript Received August 31, 1998

Abstract: The dipole interaction model is used to calculate ultraviolet circular dichroic spectra of 16 globular proteins with various amounts of helix, sheet, and disordered regions. Structures for the calculations are based on crystal coordinates taken from the Brookhaven Protein Data Bank. Calculations for "reassembled fragment" models of the proteins reproduce the main features of observed solution CD spectra in the region of the amide $\pi-\pi^*$ transition near 200 nm. The predicted spectra for helix and sheet regions show considerable variation in magnitudes and wavelengths of the CD peaks, and a still wider variation is predicted for the disordered regions. A comparison of reassembled-molecule spectra with weighted averages of the spectra of the component fragments shows that deviations from additivity of the component spectra are generally small. The implications of these findings for attempts to resolve observed CD spectra in terms of a small number of components are discussed.

Introduction

The electronic circular dichroic (CD) spectra of globular proteins present one of the major challenges in the theory of optical activity. The systems are composed of a large number of chromophores in more-or-less fixed array; there are thousands of protein molecules for which the structures are known to atomic resolution. The amide NC′O group is a relatively simple chromophore, and the nature of the intramolecular interactions is understood at least in principle. It would seem a straightforward application of exciton theory to calculate the solution CD spectra of proteins and thereby establish a theoretical connection (as distinct from an empirical correlation) between structures and spectra. Yet the few applications of quantum mechanical exciton theory to whole protein molecules in the literature[1−4] have not proven generally capable of reproducing the main CD spectral features in the conformationally sensitive region of the amide chromophore around 170−230 nm.

There are practical reasons for seeking a theoretical basis for protein CD spectra. These spectra are widely used as empirical means of detecting and measuring the various conformational components of a protein in solution. Various approximate methods of analysis of observed CD spectra are in use for this purpose.[5] These depend on two important assumptions: (i) the protein contains a small number (e.g. 3−6) of conformational components with fixed, known CD spectra, and (ii) the spectrum of the protein is the average of the component spectra weighted by the fraction of residues in each component (i.e., the spectra of the components are additive). Actually, there is little direct evidence to support either of these assumptions. The crystal structures of proteins show considerable variation even in the conformations considered to be "ordered" (helices and sheets), and disordered chains occur in still wider variety. A direct experimental test of the assumptions would be to determine CD spectra of peptides in each possible conformation in pure form, but this would seem virtually impossible for typical globular proteins.

In principle, a physical theory of CD should be able to provide tests of the above assumptions by numerical calculations of the spectra for known protein structures and their conformational components. Despite over 40 years of work on the theory of protein and peptide optical activity by many investigators,[6] no theory has yet proven realistic enough to provide suitable tests. Our work on a dipole interaction model, employing empirical polarizabilities for the amide NC′O chromophore and the nonchromophoric atoms, has shown that the model reproduces the main features of the CD spectra of the low-energy amide $\pi-\pi^*$ transition for a variety of polypeptide and protein conformations.[7−11] As a next step in the application of this model, we report here its predictions of CD spectra for 16 globular proteins whose crystal structures and solution CD spectra are both known. We examine the predictions for the various conformational components found in each protein and test the assumption of additivity by comparing spectra of the reassembled protein with the weighted average spectrum of the components.

Methods

CD spectra for the low-energy amide $\pi-\pi^*$ transition were calculated from fixed atom coordinates of each model protein via the dipole interaction model[7] by using polarizability set H_y.[8] This model

(1) Madison, V.; Schellman, J. *Biopolymers* **1972**, *11*, 1041.
(2) Manning, M. C.; Woody, R. W. *Biochemistry* **1989**, *28*, 8609.
(3) Grishina, I. B.; Woody, R. W. *Faraday Discuss.* **1994**, *99*, 245.
(4) Kurapkat, G.; Krüger, P.; Wollmer, A.; Fleischhauer, J.; Kramer, B.; Zobel, E.; Koslowski, A.; Botterweck, H.; Woody, R. W. *Biopolymers* **1997**, *41*, 267.
(5) Venyaminov, S. Yu.; Yang, J. T. In *Circular Dichroism and the Conformational Analysis of Biomolecules;* Fasman, G. D., Ed.; Plenum Press: New York, 1996; Chapter 3, pp 69−107.
(6) Woody, R. W. In *Circular Dichroism and the Conformational Analysis of Biomolecules;* Fasman, G. D., Ed.; Plenum Press: New York, 1996; Chapter 2, pp 25−67.
(7) Applequist, J.; Sundberg, K. R.; Olson, M. L.; Weiss, L. C. *J. Chem. Phys.* **1979**, *70*, 1240. Erratum: *J. Chem. Phys.* **1979**, *71*, 2330.
(8) Bode, K. A.; Applequist, J. *J. Phys. Chem.* **1996**, *100*, 17825. Erratum: *J. Phys. Chem. A* **1997**, *101*, 9560.

10.1021/ja982509o CCC: $15.00
Published on Web 10/10/1998

considers atoms and NC′O chromophores to act as point dipole oscillators under the influence of mutually induced dipole moments when in the presence of an electric field. Analysis of the full system of dispersive and nondispersive coupled oscillators produces a set of normal modes, each of which contributes a Lorentzian band to the CD spectrum. The number of such bands is equal to the number of NC′O groups, which is over 200 in several proteins treated here. Many of the bands overlap, resulting in relatively simple, but non-Lorentzian, CD spectra.

The amide n−π* transition is omitted, as this is a magnetically allowed transition and not amenable to treatment by our dipole interaction model. Thus the observed CD band in the vicinity of 220 nm is not present in our calculations. The $\pi-\pi$* transitions of the aromatic side chains are likewise omitted, as we do not have suitable parameters for these. The aromatic residues are minor components of the proteins, and their contributions to the observed spectra are relatively weak and observable primarily above about 230 nm. We do not expect these omissions to have major effects on the spectra in the amide π-π* region near 200 nm.

Crystal structures were obtained from the Brookhaven Protein Data Bank (http://www.pdb.bnl.gov). The structures of all proteins came from X-ray diffraction studies except for that of bacteriorhodopsin, which came from electron diffraction. The structures cannot be used directly in our calculations because (i) the H atoms, required in our calculations, are usually missing from the PDB files, and (ii) small deviations of experimental bond geometries from those of the small model compounds which were used in the optimization of our polarizability parameters cause serious artifacts in the optical properties or outright failure of the computer program. The following describes the manner in which suitable protein models were adapted from the published crystal structures.

A protein chain was generated by sequential addition of the repeating unit $-C^{\alpha}HR-C'O-NH-$ where R is the amino acid side chain, using methods described previously.[12,13] The PDB data were used to calculate backbone torsion angles ϕ, ψ, ω, side chain torsion angles χ^1, χ^2, and the bond angle $C'C^{\alpha}N$ for all residues. The remaining bond lengths and bond angles were those used previously by this group.[12,13]

We are limited to aliphatic side chains by our present atom polarizabilities. Hence alanine (Ala) was substituted for serine, phenylalanine, cysteine, histidine, tryptophan, and tyrosine. α-Aminobutyric acid (Abu) replaced aspartate, asparagine, arginine, lysine, glutamate, glutamine, leucine, methionine, and theronine. Valine (Val) was used in place of isoleucine. These side chain residues were generated as previously described[12,13] with hydrogen atoms located in staggered positions. Proline residues were generated by using ring parameters corresponding to χ^2 equal to +20°, 0°, or −20° given by Thomasson and Applequist,[14] according to the sign of χ^2 obtained in the crystal structure.

A variation of the method of "reassembled fragments" [10] was used to position the backbone to minimize deviations from the crystal structure arising from use of standard bond lengths. In our previous calculations[10] a chain fragment was generated from the appropriate torsion angles and bond angles and the coordinates transformed such that the coordinate system defined by two C^{α} atoms at one end of the fragment and a third C^{α} atom at the other end coincided with the coordinate system of the crystal structure defined by the same three atoms. In the present study we used this method to give the fragment an initial alignment with the experimental structure and then optimized the translation and rotation matrices to minimize the root-mean-square (rms) deviation in C^{α} positions between the model and crystal structures.

A fragment was selected as a sequence of contiguous residues of the same conformational designation. These were usually given as helix (H) or β-sheet (S) in the PDB files. In a few cases where the designations were not given in the PDB, those found by the RasMol molecular graphics program (version 2.6-beta-2, cartoons view, from http://www.umass.edu/microbio/rasmol/getras.htm) were used. Residues not designated as helix or sheet were considered disordered (D). To avoid artifacts in the calculated spectra arising from close contacts between the end of one fragment and the beginning of the next fragment, the last residue of every fragment was omitted from the calculations. Fragments having 3 residues or less were omitted entirely. In the cases where a low rms deviation could not be found for a given fragment (e.g., in a long disordered region), the fragment was split into smaller fragments with a residue removed from the end of each new fragment and the optimization of C^{α} positions repeated.

After optimization of the backbone fragments and substitution of Ala, Abu, and Val when necessary, close contacts of side chain atoms from neighboring fragments or within tight turns may occur which cause artifacts (indicated by negative eigenvalues or the occurrence of normal modes at wavelengths longer than about 220 nm) or program failure (indicated by a nonpositive-definite interaction matrix of the nondisperive subsystem). Such contacts could often be relieved by replacing a methyl or larger side chain with an isotropic point X whose polarizability is the mean polarizability of a CH_3 group (2.164 Å^3) and whose distance from C^{α} was set at 1.78 Å to reproduce the theoretical CD spectra of a 12-residue α-helix of poly(Ala) from parameter set H_y. The residue is designated Alx. If substitution of Alx did not remove the close contacts, the residue was replaced by glycine (Gly).

Prosthetic groups and metal ions present in the native structures were omitted from our models.

Calculations were carried out for the following systems: (i) the "reassembled molecule", where all fragments together in their optimized locations interact as a single molecule; (ii) the set of all helix fragments in their optimized locations; (iii) each set of fragments comprising a single sheet region; and (iv) each disordered fragment. The weighted average spectrum was computed from those of (ii)−(iv) for comparison with (i). The treatment of all helix fragments together in (ii) is not expected to introduce interaction effects between helices, as a similar study of intimate helix bundles in the α-spectrin and tropomyosin[10] showed no such effects.

The CD is expressed as the molar absorption coefficient difference $\Delta\epsilon$ on a residue-molar basis, which is related to the commonly used molar ellipticity $[\theta] = 3298\Delta\epsilon$. Our recent publications incorporated a chromophore half-peak bandwidth Γ of 4000 cm^{-1}, which is the same as the half-peak bandwidth for all normal modes. In some cases, a Γ of 6000 cm^{-1} yields a better agreement with experimental CD peak amplitudes. We present calculations for both bandwidths to indicate the range of uncertainty associated with Γ.

Calculations were performed with double precision Fortran programs on either a DEC 3000/300L workstation with 64 MB memory or a cluster of six DEC 600 Alpha workstations, the largest of which has 1 GB of memory and 4 parallel processors. The programs were written for general applications of the theory, and no special measures were taken to economize on memory requirements for large proteins where many interatomic interactions could be neglected if one chose to do so. Calculations for lactate dehydrogenase involved inversion of an interaction matrix of order 8115, which required 70 min of CPU time on the parallel workstation. For the rubredoxin molecule the interaction matrix was of order 1173, requiring 1.5 min of CPU time on the DEC 3000/300L workstation. The programs are available on the World Wide Web at http://www.public.iastate.edu/~jba/CaPPS.

Experimental spectra shown here are for proteins from the same biological sources as those of the crystal structures for all proteins except plastocyanin. The spectra were digitized from published graphs with Arc software in the GIS computing facility at Iowa State University.

Results

Table 1 lists the fragments included for the 16 proteins. Table 2 shows the percent helix, sheet, and disordered region for each reassembled model protein along with the number of residues in the model, number of residues omitted from the reference structure, and the rms deviation between C^{α} positions of the reference structure and the model. The fraction of residues deleted from the reference protein is 16% on the average and

(9) Bode, K. A.; Applequist, J. *Macromolecules* **1997**, *30*, 2144.
(10) Bode, K. A.; Applequist, J. *Biopolymers* **1997**, *42*, 855.
(11) Applequist, J.; Bode, K. A.; Appella, D. H.; Christianson, L. A.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 4891.
(12) Applequist, J. *Biopolymers* **1981**, *20*, 387.
(13) Applequist, J. *Biopolymers* **1981**, *20*, 2311.
(14) Thomasson, K. A.; Applequist, J. *Biopolymers* **1991**, *31*, 529.

Table 1. Protein Fragment Residues

Protein	Fragments[a]
Br	H(9-30),D(32-36),H(38-45),H(47-61),D(63-68),D(70-75), H(77-99),D(101-103),H(105-126),D(128-132),H(134-156), D(158-164),H(166-177),H(179-190),D(192-201),H(203-226)
CA	S(4A-8A),D(11A-16A),D(18A-22A),S(24A-28A),D(30A-34A), S(36A-38A),D(40A-45A),S(47A-54A),D(56A-58A),S(60A-65A), D(67A-71A),S(73A-77A),D(79A-86A),S(88A-95A), D(97A-101A),S(105A-116A),D(118A-121A),S(124A-129A), D(131A-138A),S(140A-143A),D(145A-151A),D(153A-159A), D(161A-168A),S(170A-174A),D(176A-184A),S(188A-199A), D(201A-207A),S(210A-215A),D(217A-225A),D(227A-236A)
Cy	H(3-12),D(14-18),D(20-24),D(26-30),D(32-36),D(38-48), H(50-53),D(55-59),H(61-67),H(71-73),D(75-86),H(88-103)
Er	S(1-4),D(6-11),S(13-16),D(18-21),S(23-31),S(33-40), D(42-48),S(50-54),D(56-61)
Fd	S(2-10),H(12-28),S(30-37),D(39-49),S(51-59),D(61-68), H(70-80),S(85-94),D(96-101),H(103-114),D(116-118), S(120-129),H(131-147)
LD	H(3-5),D(7-18),S(21-27),H(29-39),D(41-44),S(46-50), D(52-54),H(56-65),D(67-74),S(76-79),D(81-88),S(90-94), D(96-99),D(101-105),H(107-119),H(122-129),S(131-134), D(136-138),H(140-149),D(151-155),S(157-159),H(164-176), D(178-183),S(185-189),D(193-195),S(197-199),S(204-206), S(208-210),D(212-221),H(226-232),H(234-240),D(243-245), H(247-252),H(254-260),S(265-273),D(275-281),S(284-291), D(293-297),S(299-301),D(303-305),H(308-314),H(316-322), D(324-328)
Ly	H(5-13),D(15-18),S(20-22),H(27-35),D(37-41),S(43-45), D(47-49),S(51-53),D(55-58),D(60-64),D(66-77),H(80-83), D(85-87),H(89-98),D(100-102),H(104-107),H(109-113), D(115-118),H(120-123),D(125-128)
Mb	H(3-18),H(20-34),H(36-41),D(43-49),H(51-56),H(58-76), D(78-84),H(86-94),D(96-98),H(100-117),D(119-122), H(124-148),D(150-152)
Pa	S(4-6),D(8-14),D(16-22),H(24-42),D(44-48),H(50-56), D(58-65),H(67-77),D(79-86),D(88-95),D(97-105),S(108-112), H(117-128),S(130-133),H(138-142),D(144-149),D(151-156), S(158-162),S(164-167),S(169-174),D(176-183),S(185-190), D(192-204),S(206-210)
Pc	S(1-5),D(7-11),S(13-15),S(17-20),D(22-24),S(26-32), D(34-38),S(40-42),D(44-50),H(52-54),D(56-59),D(61-64), S(67-73),S(78-83),D(85-90),S(92-98)
Po	S(2-14),S(17-25),S(27-35),S(39-45),H(50-53),D(55-57), S(59-64),S(67-74),H(76-82),D(86-94),D(96-104),D(106-115), S(118-125),S(127-134),D(136-144),S(148-158),S(160-170), D(172-179),S(181-191),S(195-206),H(208-213),D(215-222), S(227-239),S(242-254),S(258-271),S(275-284),D(286-290), S(292-300)
Pr	S(11A-18A),D(20A-26A),S(28A-34A),D(36A-40A),S(45A-48A), S(53A-55A),D(57A-60A),D(62A-65A),S(67A-73A),H(75A-82A), D(84A-89A),S(91A-96A),D(98A-103A),S(105A-112A), S(114A-120A),S(11B-18B),D(20B-26B),S(28B-34B), D(36B-40B),S(45B-48B),S(53B-55B),D(57B-60B),D(62B-65B), S(67B-73B),H(75B-82B),D(84B-89B),S(91B-96B), D(99B-103B),S(105B-112B),S(114B-121B)
Ri	H(4-11),D(13-23),H(25-31),D(33-41),S(43-46),H(51-59), S(61-63),D(65-70),S(72-74),S(79-85),D(87-95),S(97-104), S(106-108),D(110-117),S(119-123)
Ru	S(4-6),D(8-12),D(14-18),H(20-22),D(24-28),H(30-32), D(34-37),D(39-44),H(46-51)
Sn	H(6-10),H(12-18),D(20-25),S(27-31),D(33-42),S(44-48), D(50-62),H(64-72),D(74-87),S(89-93),D(95-102),H(104-115), D(117-119),S(121-123),D(125-131),H(133-143),S(148-151), D(153-162),D(164-167),D(169-173),S(175-179),D(181-196), S(198-200),S(205-208),S(213-217),H(219-236),D(238-241), H(243-251),D(253-258),H(260-262),D(264-268),H(270-273)
SD	S(3A-7A),D(9A-13A),S(15A-23A),S(26A-33A),D(35A-37A), S(39A-41A),S(44A-46A),D(48A-54A),H(56A-58A),D(60A-66A), D(68A-74A),D(76A-82A),S(84A-86A),D(88A-91A),S(93A-98A), D(100A-105A),D(107A-112A),S(114A-117A),D(119A-129A), H(131A-134A),D(136A-139A),S(141A-145A),D(147A-151A), S(3B-7B),D(9B-13B),S(15B-23B),S(26B-33B),D(35B-37B), S(39B-41B),S(44B-46B),D(48B-52B),H(54B-57B),D(59B-66B), D(68B-74B),D(76B-82B),S(84B-86B),D(88B-91B),S(93B-98B), D(100B-112B),S(114B-117B),D(119B-128B),H(130B-134B), D(136B-139B),S(141B-145B),D(147B-150B)

[a] Fragments are H, helix; S, sheet; D, disordered.

Table 2. Conformation and Size of Protein Structures

protein	type	H[a]	S[a]	D[a]	no. of res[b]	diff[c]
Br	α	79	0	21	203 (18)	0.474
CA	β	0	43	57	197 (38)	0.393
Cy	α	45	0	55	88 (16)	0.492
Er	β	0	57	43	53 (9)	0.399
Fd	α/β	44	35	21	131 (16)	0.625
LD	α/β	41	23	36	266 (63)	0.337
Ly	α	45	9	46	101 (28)	0.358
Mb	α	83	0	17	138 (15)	0.468
Pa	α+β	31	21	48	177 (35)	0.372
Pc	β	4	53	43	79 (20)	0.399
Po	β	7	69	24	250 (51)	0.468
Pr	β	9	56	35	180 (45)	0.284
Ri	α+β	24	33	43	100 (24)	0.475
Ru	α+β	30	8	62	40 (12)	0.281
Sn	α/β	34	17	49	228 (47)	0.326
SD	β	6	37	57	250 (53)	0.311

[a] Percent helix, sheet, and disordered, respectively. [b] Number of residues in calculation given first. Number in parentheses is the number omitted from the PDB structure. [c] Root-mean-square difference between C^a positions in the calculated structure and the crystal structure (Å).

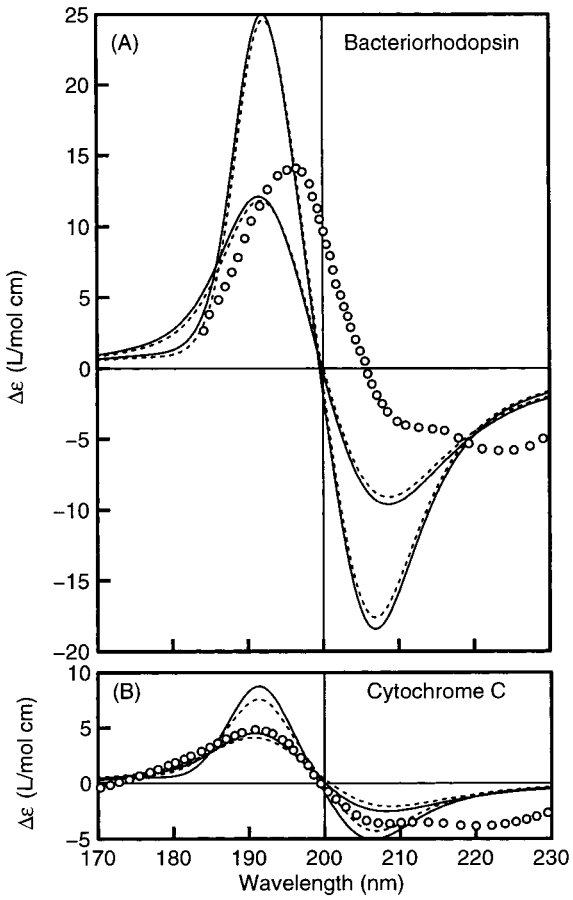

Figure 1. CD spectra of α-type proteins. The following information applies to Figures 1−6: (—) reassembled-molecule calculations; (- - -) weighted average of calculations for isolated fragments; (○) experimental data cited in the text. Half-peak bandwidths for calculated spectra are 4000 (larger peak amplitudes) and 6000 cm^{-1} (smaller peak amplitudes).

varies from 8% for bacteriorhodopsin to 23% for rubredoxin. Figures 1−6 show the calculated and experimental CD spectra for the 16 globular proteins. Theoretical curves are for both reassembled-molecule and weighted average spectra, both shown with half-peak bandwidths of 4000 and 6000 cm^{-1}. In each case the smaller bandwidth corresponds to the CD peaks with

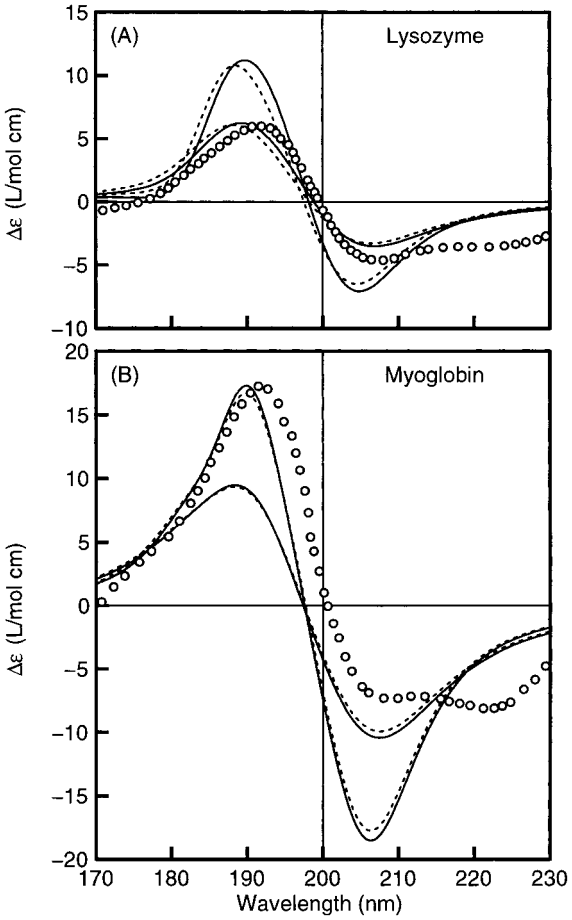

Figure 2. CD spectra of α-type proteins. See caption for Figure 1.

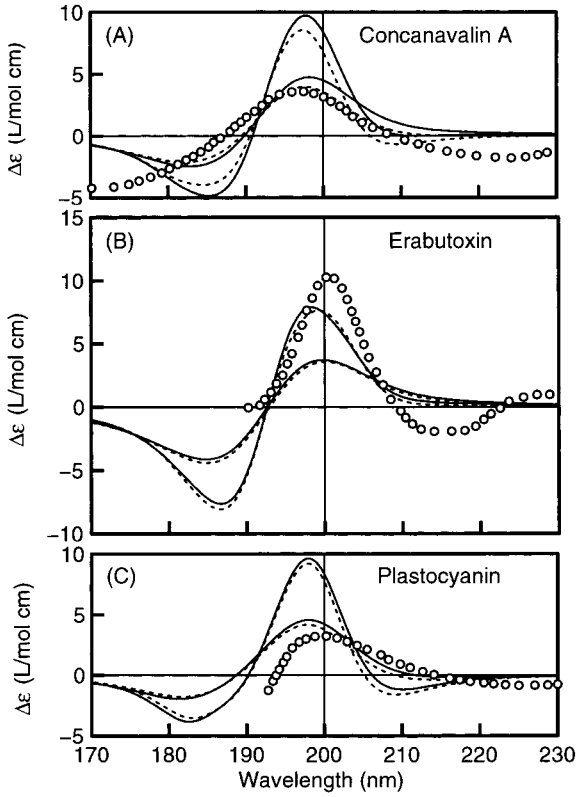

Figure 3. CD spectra of β-type proteins. See caption for Figure 1.

the greater amplitudes. Information on individual proteins follows, grouped by conformation type.

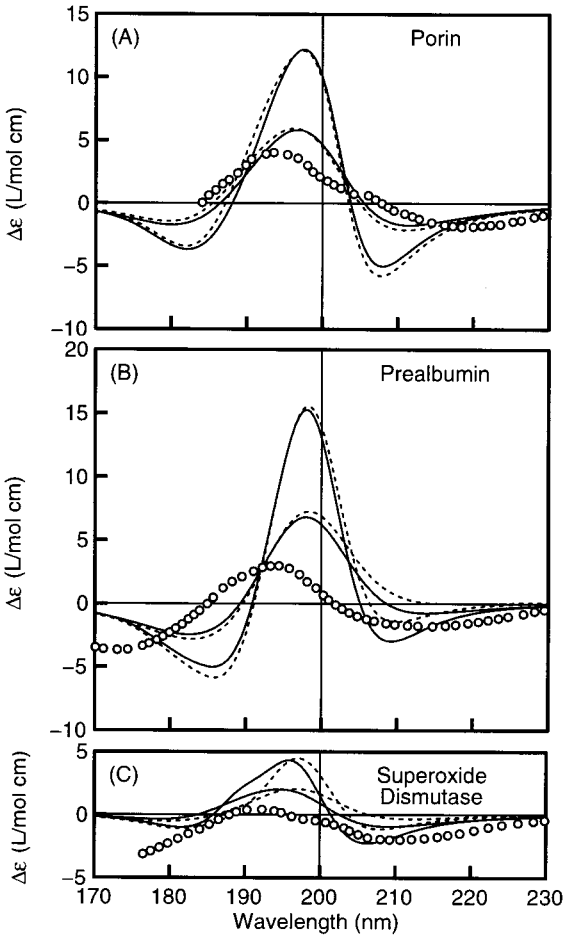

Figure 4. CD spectra of β-type proteins. See caption for Figure 1.

Conformation Type α (Mostly α-Helices). (a) Bacteriorhodopsin (Br) from *Halobacterium halobium*. Crystal coordinates from Grigorieff et al.[15] (PDB file 2BRD, resolution 3.5 Å, residues 7−227) were used to generate a model protein with the 16 fragments listed in Table 1. Residues 128, 134, and 145 were modeled with Alx instead of Abu and residue 148 was modeled with Alx instead of Val to avoid close contacts between side chain atoms. Figure 1A shows the calculated CD spectra along with the experimental data of Fasman[16] for the protein in water. It is likely that the experimental CD spectrum shows distortions due to particulate scattering,[16] and it is noteworthy that the CD spectrum of detergent-solubilized bacteriorhodopsin[17] shows about 80% greater intensity in the peak at 210 nm, which is more consistent with our calculations. Unfortunately that spectrum did not extend below 207 nm, so we have not reproduced it here.

(b) Cytochrome C (Cy) from Horse Heart. Crystal coordinates from Bushnell et al.[18] (PDB file 1HRC, resolution 1.9 Å, residues 1−104) were used to generate a model protein with the 12 fragments shown in Table 1. No substitutions of Alx or Gly were necessary. The acetyl group at the N-terminus of the crystal structure was not included in the model compound.

(15) Grigorieff, H.; Ceska, T. A.; Downing, K. H.; Baldwin, J. M.; Henderson, R. *J. Mol. Biol.* **1996**, *259*, 393.

(16) Fasman, G. D. In *Circular Dichroism and the Conformational Analysis of Biomolecules;* Fasman, G. D., Ed.; Plenum Press: New York, Chapter 10, 1996; pp 381−412.

(17) Reynolds, J. A.; Stoeckenius, W. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 2803.

(18) Bushnell, G. W.; Louie, G. V.; Brayer, G. D. *J. Mol. Biol.* **1990**, *214*, 585.

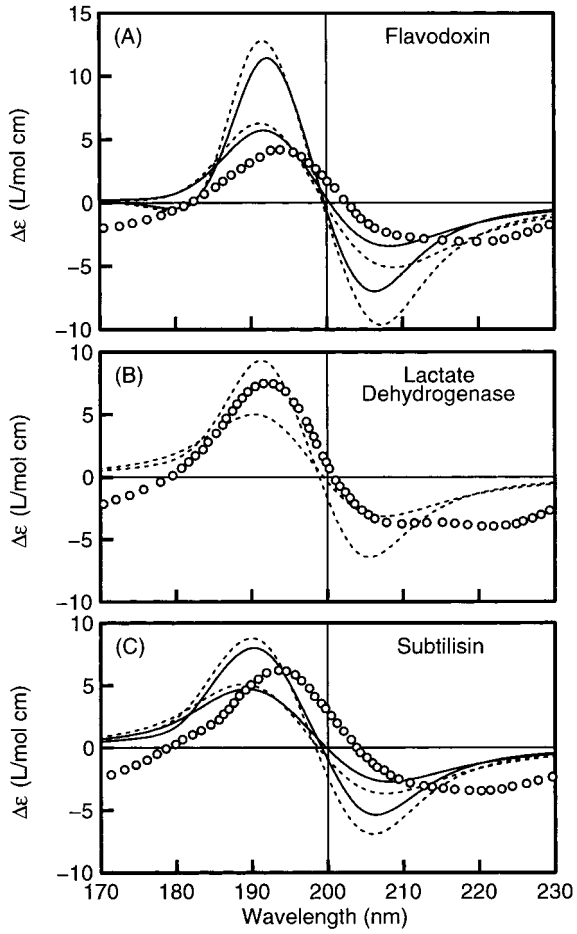

Figure 5. CD spectra of α/β-type proteins. See caption for Figure 1.

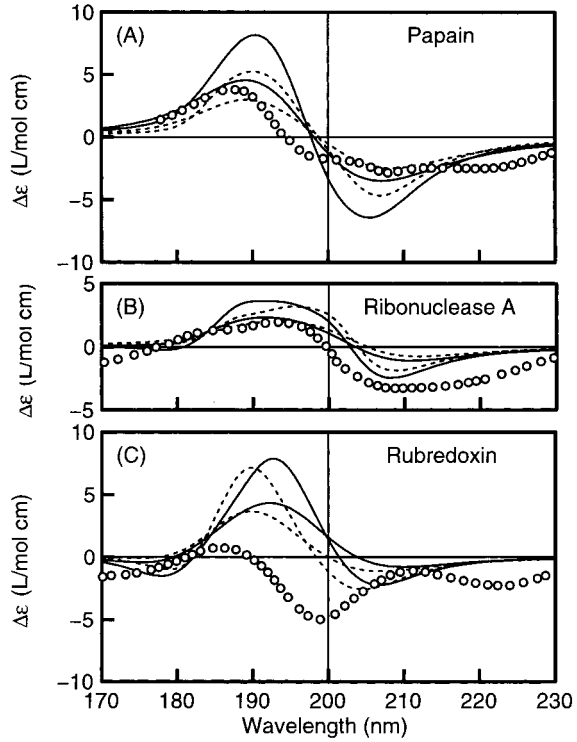

Figure 6. CD spectra of ($\alpha + \beta$)-type proteins. See caption for Figure 1.

Figure 1B shows the calculated spectra and the experimental spectra at neutral pD in 0.1 M NaF in D_2O from Brahms and Brahms.[19]

(c) Lysozyme (Ly) from Hen Egg White. Crystal coordinates from Kurinov and Harrison[20] (PDB file 1LSE, resolution 1.7 Å, residues 1−129) were used to generate a model protein with the 20 fragments listed in Table 1. Fragment assignments were made based on the cartoons view in RasMol. No substitutions of Alx or Gly were made. Figure 2A shows the calculated spectra and the experimental CD spectrum in 0.1 M NaF at pH 7 from Brahms and Brahms.[19]

(d) Myoglobin (Mb) from Sperm Whale Muscle. Crystal coordinates from Yang and Phillips[21] (PDB file 1VXA, resolution 2.0 Å, residues 1−153) were used to generate the 13-fragment model listed in Table 1. Model residues Abu 18 and Val 21 were modeled with Alx. Figure 2B displays the calculated spectra and the experimental CD spectrum in 0.01 M sodium phosphate buffer at pH 6.8 from Hennessey and Johnson.[22]

Conformation Type β (mostly β-Sheets). (a) Concanavalin A (CA) from Jack Bean. Crystal coordinates from Weisgerber and Helliwell[23] (PDB file 2CTV, resolution 1.95 Å, residues 1A−237A) were used to create the model protein with 30 fragments of antiparallel β-strands and disordered regions shown in Table 1. Model residues Ala 6A, 95A, 213A and Val 7A, 27A, 91A, 210A, 214A were replaced with Alx. Figure 3A compares the calculated spectra with the experimental spectrum in 0.1 M NaF at pH 5.1 and 4 °C from Brahms and Brahms.[19]

(b) Erabutoxin (Er) from the Venom of Sea Snake *Laticauda semifasciata.* Crystal coordinates from Smith et al.[24] (PDB file 3EBX, resolution 1.4 Å, residues 1−62) were used to construct two model proteins from the A and B coordinates arranged in 9 antiparallel β-strands and disordered fragments as shown in Table 1. No substitutions of Alx or Gly were necessary. Figure 3B shows the calculated spectra from the A coordinates and the experimental spectrum in 1 mM phosphate buffer at pH 6.5 from Dahms and Szabo.[25] Nearly identical spectra were produced from the B coordinates (not shown).

(c) Plastocyanin (Pc) from Poplar Leaves. Crystal coordinates from Guss et al.[26] (PDB file 1PLC, resolution 1.33 Å, residues 1−99) were used to generate two 16-fragment models of the mixture of parallel and antiparallel β-sheet structure using the A and B coordinates as listed in Table 1. Val 28 and Glu 71 were modeled with Alx. Figure 3C shows the calculated spectra for the A coordinates and the experimental spectrum for spinach plastocyanin in 10 mM HEPES buffer at pH 7.0 from Draheim et al.[27] According to Durell et al.[28] the CD spectra of spinach and poplar plastocyanin are indistinguishable in the range 190−270 nm. Spectra predicted from the B coordinates are nearly identical to those for the A coordinates.

(d) Porin (Po) from *Rhodobacter capsulatus.* Crystal coordinates from Weiss and Schultz[29] (PDB file 3POR, resolution 2.5 Å, residues 1−301) were used to construct the 28-fragment

(19) Brahms, S.; Brahms, J. *J. Mol. Biol.* **1980**, *138*, 149.
(20) Kurinov, I. V.; Harrison, R. W. *Acta Crystallogr. Sect. D* **1995**, *51*, 98.
(21) Yang, F.; Phillips, G. N., Jr. *J. Mol. Biol.* **1996**, *256*, 762.
(22) Hennessey, J. P., Jr.; Johnson, W. C., Jr. *Biochemistry* **1981**, *20*, 1085.
(23) Weisgerber, S.; Helliwell, J. R. *J. Chem. Soc., Faraday Trans.* **1993**, *89*, 2667.
(24) Smith, J. L.; Corfield, P. W. R.; Hendrickson, W. A.; Low, B. W. *Acta Crystallogr., Sect. A* **1988**, *44*, 357.
(25) Dahms, T. E.; Szabo, A. G. *Biophys. J.* **1995**, *69*, 569.
(26) Guss, J. M.; Bartunik, H. D.; Freeman, H. C. *Acta Crystallogr., Sect. B* **1992**, *48*, 790.
(27) Draheim, J. E.; Anderson, G. P.; Duane, J. W.; Gross, E. L. *Biophys. J.* **1986**, *49*, 891.
(28) Durell, S. R.; Gross, E. L.; Draheim, J. E. *Arch. Biochem. Biophys.* **1988**, *267*, 217.
(29) Weiss, M. S.; Schulz, G. E. *J. Mol. Biol.* **1993**, *231*, 817.

antiparallel β-barrel listed in Table 1, a single subunit of the homotrimer. Ala replaced Pro 287; Gly replaced Ser 44, Ala 45, and Phe 62; Alx replaced model Abu 60, 108, 144, 185, 195, 198, 253, 259, model Val 61, 139, 189, 277, 297, and model Ala 43, 187, 200, 202, 238, 235, 248. Figure 4A displays the calculated spectra and the experimental spectrum in 0.3 M LiCl, 20 mM Tris buffer, 0.6% octyltetraoxyethylene, 3 mM NaN_3 at pH 7.2 from Fasman.[16]

(e) Prealbumin (Pr) from Human Plasma. Crystal coordinates from Blake et al.[30] (PDB file 2PAB, resolution 1.8 Å, residues 10A−123A and 10B−123B) were used to generate the 30-fragment model of the mixture of parallel and antiparallel β-strands as listed in Table 1, which is two subunits representing half of the heterotetramer. Arg 103A and 103B were replaced with Ala since the C^β coordinates were not included in the PDB file. Model Val 14B, 107B, 73B and Ala 91B were replaced with Alx. Figure 4B shows the calculated spectra and the experimental CD spectrum in D_2O at pD 7 from Brahms and Brahms.[19] The experiments in this case were performed on the same preparation as used in the crystal study.[19] Predicted spectra for the individual subunits (not shown) were similar to the dimer spectra though not exact reproductions.

(f) Superoxide Dismutase (SD) from Bovine Erythroctyes. Crystal coordinates from Rypniewski et al.[31] (PDB file 1SXA, resolution 1.9 Å, residues 1A−151A and 1B−151B) were used to generate the antiparallel β-sheet structure modeled with the 45 fragments given in Table 1, a heterodimer of the A and B chains. Arg 77A and 77B and Asn 81A and 81B were modeled with Alx in place of Abu. Figure 4C compares the calculated spectra and the experimental CD spectrum in 0.2 M NaF at pH 7 from Brahms and Brahms.[19] Theoretical spectra for the monomers (not shown) were somewhat different from the predicted dimer spectra.

Conformation Type α/β (Alternating α-Helices and β-Sheets). (a) Flavodoxin (Fd) from *Desulfovibrio vulgaris.* Crystal coordinates from Watt et al.[32] (PDB file 2FX2, resolution 1.9 Å, residues 2−148) were used to construct the 13-fragment model listed in Table 1, a mixture of helix and parallel β-sheets. Model Ala 71 and 140, Val 6 and 144, and Abu 54 and 68 were changed to Alx. Figure 5A compares the calculated spectra and the experimental CD spectrum in 0.05 M KF in D_2O at pD 7 from Brahms and Brahms.[19]

(b) Lactate Dehydrogenase (LD) from Dogfish. Crystal coordinates from Abad-Zapatero et al.[33] (PDB file 6LDH, resolution 2.0 Å, residues 1−329) were used to generate the 43-fragment mixture of helix and parallel and antiparallel β-sheet structure listed in Table 1, a single subunit of the homotetramer. Model residues Ala 34, Val 38, and Abu 218 were replaced with Alx. The acetyl group at the N-terminus of the crystal structure was not included in the calculations. Figure 5B displays the calculated spectra and the experimental spectrum in D_2O at neutral pD from Brahms and Brahms.[19]

(c) Subtilisin (Sn) from *Bacillus amyloliquifaciens.* Crystal coordinates from Gallagher et al.[34] (PDB file 1SUP, resolution 1.6 Å, residues 1−275) were used to construct two 32-fragment models from the A and B coordinates as shown in Table 1. The β-sheets occur in both parallel and antiparallel forms. Ala 74, Val 81, Ile 205, and Met 222 were modeled with Alx. Figure 5C displays the spectra calculated from the A coordinates and the experimental CD spectrum in 0.2 M NaF in D_2O at neutral pD from Brahms and Brahms.[19] Spectra predicted with the B coordinates (not shown) were nearly identical to those for the A coordinates.

Conformation Type α + β (Mixtures of α-helix and β-sheets). (a) Papain (Pa) from Papaya Fruit Latex. Crystal coordinates from Kamphuis et al.[35] (PDB file 9PAP, resolution 1.65 Å, residues 1−212) were used to construct a model with the 24 fragments given in Table 1. Ile 171 and 189 were modeled with Alx in place of Val. The β-strands in papain are all antiparallel. Figure 6A displays the calculated spectra and the experimental spectra in 0.01 M sodium phosphate buffer at pH 6.8 from Hennessey and Johnson.[22]

(b) Ribonuclease A (Ri) from Bovine Pancreas. Crystal coordinates from Dunbar et al.[36] (PDB file 1RBX, resolution 1.69 Å, residues 1−124) were used to generate a model protein with the 15 fragments shown in Table 1. No substitutions of Alx or Gly were made. Ribonuclease A β-strands are all antiparallel. Figure 6B displays the calculated spectra and the experimental spectrum in 0.01 M sodium phosphate buffer at pH 6.8 from Hennessey and Johnson.[22]

(c) Rubredoxin (Ru) from *Desulfovibrio vulgaris.* Crystal coordinates were taken from Dauter et al.[37] (PDB file 8RXN, resolution 1.0 Å, residues 1−52). No secondary structure information was included for this protein. Fragment assignments were made based on the cartoons view in RasMol. Our classification with type α + β is somewhat arbitrary, as the sheet content is low by these assignments, and the helices are short and disordered. Brahms and Brahms[19] classified rubredoxin as type β from a different crystal structure. Two models were constructed from the A and B coordinates by using the 9 fragments listed in Table 1. No substitutions of Alx or Gly were necessary. Figure 6C shows the calculated spectra for the A coordinates and the experimental spectrum in 0.1 M NaF at pH 4 from Brahms and Brahms.[19] The CD spectrum for the B coordinates (not shown) was similar to that for the A coordinates. The experimental spectrum in this case resembles that of random-coil polypeptides, and deviates widely from the spectra predicted from the crystal structure. Brahms and Brahms[19] noted a similar discrepancy in their analysis, and suggested that the crystal structure is unstable in solution due to the small size of the sheet regions. Interestingly, the solution structure of the exceptionally thermostable rubredoxin from *Pyrococcus furiosus* was shown by NMR to be very similar to the crystal structure.[38,39] While the crystal secondary structures of the *D. vulgaris* and *P. furiosus* forms are similar, the sequence identity is only 60%, and there is no reason to expect them to have the same stability in solution. Our CD spectrum for *P. furiosus* rubredoxin calculated from the NMR structure (PDB file 1ZRP) is qualitatively similar to that for the *D. vulgaris* form, but with peaks shifted to shorter wavelengths by 4−7 nm (not shown). We were not able to locate a solution CD spectrum of *P. furiosus* rubredoxin for comparison.

(30) Blake, C. C. F.; Geisow, M. J.; Oatley, S. J.; Rérat, B.; Rérat, C. *J. Mol. Biol.* **1978**, *121*, 339.
(31) Rypniewski, W. R.; Mangani, S.; Bruni, B.; Orioli, P.; Casati, M.; Wilson, K. S. *J. Mol. Biol.* **1995**, *251*, 282.
(32) Watt, W.; Tulinsky, A.; Swenson, R. P.; Watenpaugh, K. D. *J. Mol. Biol.* **1991**, *218*, 195.
(33) Abad-Zapatero, C.; Griffith, J. P.; Sussman, J. L.; Rossmann, M. G. *J. Mol. Biol.* **1987**, *198*, 445.
(34) Gallagher, D. T.; Bott, J. D. O. R.; Betzel, C.; Gilliland, G. C. *Acta Crystallogr. Sect. D* **1996**, *52*, 1125.
(35) Kamphuis, I. G.; Kalk, K. H.; Swarte, M. B. A.; Drenth, J. *J. Mol. Biol.* **1984**, *179*, 233.
(36) Dunbar, J.; Yennawar, H. P.; Banerjee, S.; Luo, J.; Farber, G. K. *Protein Science* **1997**, *6*, 1727.
(37) Dauter, Z.; Sieker, L. C.; Wilson, K. S. *Acta Crystallogr., Sect B* **1992**, *48*, 42.
(38) Blake, P. R.; Park, J.-B.; Zhou, Z. H.; Hare, D. H.; Adams, M. W. W.; Summers, M. F. *Protein Sci.* **1992**, *1*, 1508.
(39) Day, M. W.; Hou, B. T.; Joshua-Tor, L.; Park, J.-B.; Zhou, Z. H.; Adams, M. W. W.; Rees, D. C. *Protein Sci.* **1992**, *1*, 1494.

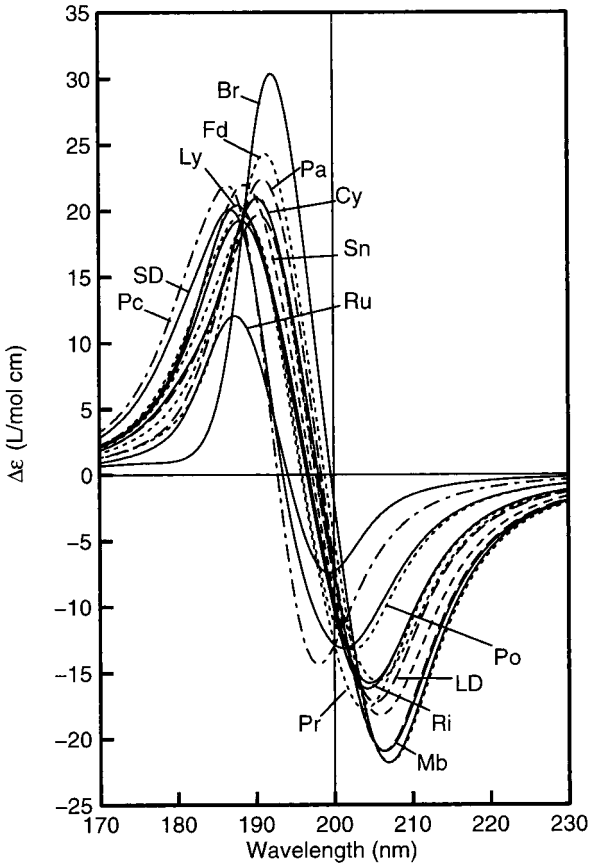

Figure 7. CD spectra of helix regions calculated with a half-peak bandwidth of 4000 cm^{-1}.

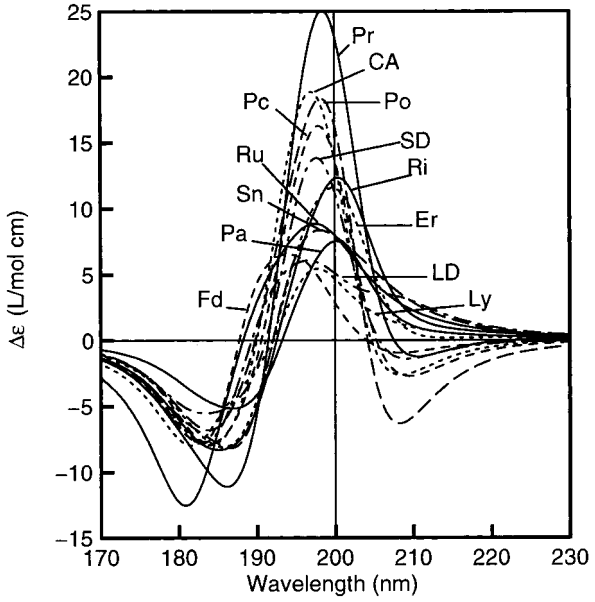

Figure 8. CD spectra of sheet regions calculated with a half-peak bandwidth of 4000 cm^{-1}.

Spectra of Conformational Components. To test the variability of the CD spectra for the various conformational components, calculations of CD spectra for only the helix, sheet, or disordered regions are presented in Figures 7-9. A consistent CD pattern is seen for the helices in Figure 7, yet the amplitude varies from 12 to 30 L/(mol cm) at the shorter wavelength band at 186−192 nm, and the negative peak has amplitude from −7 to −22 L/(mol cm) at 198−206 nm. The variation in ϕ and ψ for these helical regions is shown in Table 3. The sheet spectra in Figure 8 show a more variable band pattern, the most consistent feature of which is a peak in the range 194−200 nm with amplitude between 5.5 and 25 L/(mol cm). Table 4 shows the variation in ϕ and ψ for the sheet regions of these proteins. The CD spectra averaged over the disordered regions of the proteins are shown in Figure 9. A still wider range of band patterns is found, none of which is close to the experimental spectrum shown for poly(Pro-Lys-Leu-Lys-Leu),[19] which is typical of random-coil polypeptides. Figure 10 shows Ramachandran plots for the disordered residues in subtilisin, concanavalin A, and superoxide dimutase, which are representative of the distributions of disordered residues in most of the

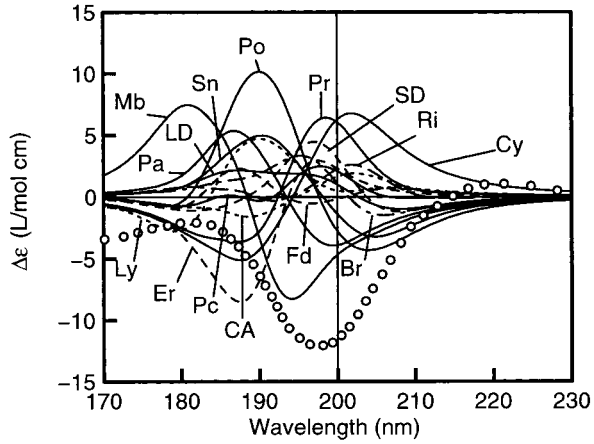

Figure 9. CD spectra for disordered regions calculated with a half-peak bandwidth of 4000 cm^{-1}. Experimental data (○) are for poly(Pro-Lys-Leu-Lys-Leu).[19]

Table 3. Average Torsion Angles (deg) in Helix Regions[a]

protein	ϕ	ψ	ω
bacteriorhodopsin	−58(10)	−50(10)	180(1)
cytochrome C	−64(29)	−40(11)	180(1)
flavodoxin	−65(26)	−35(19)	179(3)
lactate dehydrogenase	−66(14)	−36(32)	180(5)
lysozyme	−70(18)	−39(20)	176(4)
myoglobin	−68(12)	−40(37)	179(6)
papain	−68(30)	−25(47)	180(5)
plastocyanin	−86(22)	−28(3)	182(1)
porin	−71(14)	−58(52)	181(5)
prealbumin	−81(58)	−44(10)	179(2)
ribonuclease A	−70(18)	−37(15)	180(2)
rubredoxin	−78(52)	+28(72)	179(4)
subtilisin	−68(14)	−35(24)	178(4)
superoxide dismutase	−72(44)	−34(34)	180(2)

[a] Standard deviation given in parentheses.

Table 4. Average Torsion Angles (deg) in Sheet Regions[a]

protein	ϕ	ψ	ω
concanavalin A	−131(29)	140(24)	179(2)
erabutoxin	−123(29)	154(39)	178(3)
flavodoxin	−108(42)	126(47)	179(3)
lactate dehydrogenase	−119(41)	99(65)	179(4)
lysozyme	−47(86)	109(48)	179(8)
papain	−109(33)	128(51)	178(6)
plastocyanin	−118(35)	123(47)	184(28)[b]
porin	−137(34)	142(36)	178(6)
prealbumin	−117(24)	129(36)	180(3)
ribonuclease A	−114(28)	145(35)	179(2)
rubredoxin	−90(20)	134(10)	180(2)
subtilisin	−118(30)	134(35)	178(5)
superoxide dismutase	−124(40)	137(34)	179(3)

[a] Standard deviation given in parentheses. [b] One ω value is 0° in this case.

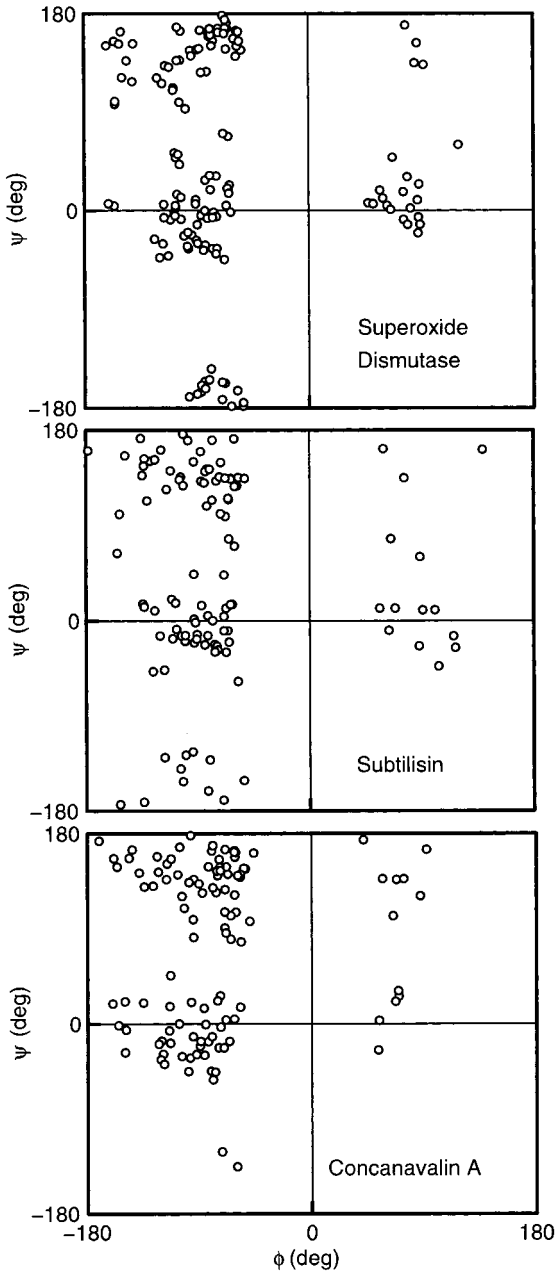

Figure 10. Ramachandran plots for disordered regions in concanavalin A, subtilisin, and superoxide dismutase.

proteins in that they show a wide scatter throughout the sterically allowed regions of the map.

The variability of the helix and sheet spectra found here is reminiscent of that predicted by Manning et al.[40] for hypothetical helices and sheets in which the torsion angles for all residues are uniformly distorted.

Discussion

The following are the major conclusions to be drawn from this study.

1. The dipole interaction model and its parameters as applied here to the crystal structures reproduce fairly well the main features of the observed CD spectra of most of the globular proteins in the region of the amide $\pi-\pi^*$ transition near 200 nm. The exception in the case of rubredoxin has been noted above. To the best of our knowledge this is the first study in which a physical theory of CD spectra in the amide region has proven successful for a broad range of globular proteins.

2. The predicted spectra for ordered regions of the 16 proteins (helices and sheets) show qualitative consistency within the two structure types, but there is considerable variation in the predicted magnitudes and wavelengths of the CD peaks. For the disordered regions a wider range of variation is predicted, and none of the mean spectra of disordered regions comes close to the spectra observed for random-coil polypeptides. These findings would impose severe limits on attempts to analyze protein CD spectra in terms of a small number of components with fixed spectra.

3. The comparison of reassembled-molecule spectra with weighted average spectra shows that any deviations from additivity are generally small relative to the overall theoretical uncertainty of the spectra. In a few cases (e.g., flavodoxin, papain, rubredoxin) potentially significant deviations are found, and these must arise from interactions among the reassembled fragments of our protein models. While we have attempted to remove artifactual interactions by deleting residues or side chain atoms where fragments are unrealistically close, we cannot rule out the possibility that some artifacts in the reassembled-molecule spectra remain in the cases where the deviations appear significant. On the whole, these findings indicate that, for our model, the CD spectra are insensitive to tertiary structure, as we found previously in the case of helix-bundle proteins.[10]

The following matters also require comment.

4. The comparison of disordered regions of superoxide dismutase, subtilisin, and concanavalin A shows widely differing CD spectra, yet the Ramachandran plots of conformations in these regions all look similar in that they appear to be random distributions of points throughout the allowed regions of the map. The inference is that the CD spectrum is sensitive to sequence information that is not given by the Ramachandran plot. The fact that the calculated spectra do not resemble the observed spectra for random-coil polypeptides supports the idea that such polypeptides are actually confined to a small portion of the Ramachandran map, probably near the extended chain region.[12,41] The disordered regions of the globular proteins are mostly involved in turns, parts of which would be classified as specific structures distinct from random coils. (See the review of turn structures by Perczel and Hollósi.[42]) For the present study we have not attempted to make this distinction, as the term "disordered" seems appropriate for the wide distribution of conformations included. It might be noted that calculations of CD spectra of model β-turn compounds by the dipole interaction model show widely varying spectra for various low-energy conformations,[43] most of which resemble the spectra calculated in this study for disordered fragments of the globular proteins.

5. No single half-peak bandwidth Γ of the NC′O chromophore fits all of the protein CD spectra equally well, and the range 4000−6000 cm^{-1} represents roughly the limits of the optimum Γ for various proteins. While our theory yields CD spectra with correct band signs and wavelengths from a consistent set of input parameters, Γ is a parameter that does not appear to be consistent for all molecules. Thus the ability of the theory to predict band intensities appears to be in question.

(40) Manning, M. C.; Illangasekare, M.; Woody, R. W. *Biophys. Chem.* **1988**, *31*, 77.

(41) Brant, D. A.; Flory, P. J. *J. Am. Chem. Soc.* **1965**, *87*, 2791.

(42) Perczel, A.; Hollósi, M. In *Circular Dichroism and the Conformational Analysis of Biomolecules*, Fasman, G. D., Ed.; Plenum Press: New York, 1996; Chapter 9, pp 285−380.

(43) Sathyanarayana, B. K.; Applequist, J. *Int. J. Peptide Protein Res.* **1986**, *27*, 86.

On this matter the following should be noted: (i) The intensity of a band should be measured by its rotational strength, which is related to the area under the band rather than the peak amplitude. The rotational strengths, as well as the wavelengths, of the normal modes are independent of the choice of bandwidth. However, the experimental determination of rotational strengths by resolution of overlapping bands is subject to large uncertainties due to the lack of knowledge of the actual band shape and the extent of mutual cancellation in overlap regions. (ii) Our theory assumes for convenience that the chromophore bands are Lorentzian functions, and the band shapes in the calculated spectra are sensitive to this choice.[7,44] It is possible that, with a different choice of band shape for the NC′O chromophore, a more consistent agreement with experimental band shapes might be found. The best we can say at present about the calculated band intensities is that they are roughly consistent with experiment within the theoretical uncertainty suggested by the range of Γ shown in the figures.

6. While the correspondence between theory and experiment is encouraging, the discrepancies between the two raise the question of what further modifications of the theory might be needed to obtain better agreement. Our approximations in generating structures and the matter of band shape are some obvious areas for seeking improvement. However, it is worth remembering that the molecular structure in the crystal is probably not exactly the same as that in solution in any protein. Given the sensitivity of the CD spectrum to secondary structure found in this study, one must expect some errors in a theory based on crystal structures to arise. The importance of this cannot be known until more precise information about solution structures is obtained, perhaps from NMR studies.

7. Finally, a brief summary of previous theortical studies of the CD spectra of globular proteins based on quantum mechanical exciton theory is in order. Madison and Schellman[1] found reasonable agreement with experiment for myoglobin but large discrepancies for lysozyme, α-chymotrypsin, and ribonuclease S in the amide spectral region. Manning and Woody[2] found good agreement with experiment for basic bovine pancreatic trypsin inhibitor in the amide region when tyrosine and phenylalanine chromophores were included along with the amide chromophores. On the other hand, Grishina and Woody[3] report poor agreement with experiment in the amide region for dihydrofolate reductase, chymotrypsin, chymotrypsinogen, and barnase, though reasonable agreement was found in the longer wavelength region of the aromatic side chain chromophores. Kurapkat et al.[4] likewise report discouraging results in the amide region for ribonuclease, while finding reasonable agreement for the tyrosine and disulfide chromophore regions. While all of these studies included more known electronic transitions of the amide and side chain chromophores than was done here, none has included the nonchromophoric backbone and side chain atoms which are vital parts of our model. As we have pointed out many times, this distinction is probably the most important one in accounting for the relative success of our dipole interaction model in the region of the amide $\pi-\pi^*$ transition.

Acknowledgment. We are grateful to Dr. James J. Coyle of the Iowa State University Computation Center for programming assistance on the Alpha Farm workstation cluster.

JA982509O

(44) Rabenold, D. A. *J. Phys. Chem.* **1988**, *92*, 4863.