First-Principles Calculations of Protein Circular Dichroism in the Near Ultraviolet†

David M. Rogers and Jonathan D. Hirst*

School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, U.K.

Received May 13, 2004; Revised Manuscript Received June 24, 2004

ABSTRACT: Electronic transitions in aromatic side chains are responsible for the characteristics of proteins in the near UV. We present the first systematic study of a large number of proteins focused on the accurate calculation of near-UV circular dichroism (CD) spectra. We report new parameter sets derived from ab initio calculations for benzene, phenol, and indole that describe the valence electronic transitions to the ^1L_b, ^1L_a, ^1B_b, and ^1B_a states in the side chains of amino acids phenylalanine, tyrosine, and tryptophan. CD spectra were calculated, using the matrix method with the new side-chain parameters, for 30 proteins whose CD spectra and crystal structures have been made publicly available. The new parameter sets are fully self-consistent and yield near-UV spectra better than those obtained using previous parameter sets. The mean absolute errors for computed wild-type spectra in the near UV are reduced by a factor of ∼2. A similiar reduction is found for the near-UV spectra (and difference spectra) of mutants involving aromatic amino acids. Empirical modifications to model vibronic coupling in the side-chain chromophore of phenylalanine offer no overall improvement. Protein CD calculations from first principles coupled with atomic-level modeling enhance the utility and interpretability of CD measurements in the near UV.

Circular dichroism (CD)[1] spectroscopy is a useful experimental technique for protein secondary structure determination. The method is often used to study the impact that various physiochemical factors, for example, pH, have on the conformation and on the secondary structural content of a globular protein. The technique can furnish insight into protein folding events. Time-resolved CD allows one to study the time evolution of a protein's secondary structure in various native and non-native environments, and recent advances have enabled protein dynamics to be monitored on sub-microsecond time scales (*1*). Recent progress has yielded theoretical CD calculations that are sufficiently accurate that, when allied with molecular dynamics simulations, they may be of use in predicting and interpreting such experiments. Inclusion of chromophores other than the peptide or amide group in theoretical descriptions of CD should enhance the agreement between theory and experiment (*2*).

There are many examples of the application of CD in biochemical and biophysical investigations. One recent study (*3*) on a homodimeric DNA binding protein, the factor for inversion stimulation (FIS) which possesses four tyrosines, employed far- and near-UV CD (among other techniques) in an effort to understand how the side-chain residues contribute to its conformational and spectroscopic properties. CD was measured for the wild type and for four Tyr → Phe mutants. It was concluded that the Tyr51Phe mutation perturbs the hydrogen bonding network within the C-terminus, and that a tyrosine residue (Tyr51) is central for modulating the local environment in FIS. This structural difference accounts for the near-UV CD spectra of the wild type and the Tyr51Phe mutant. A more precise understanding of the origins of near-UV CD would enhance its utility in such studies. This motivates efforts toward accurate first-principles calculations in the near UV.

The far-UV CD of a protein reflects its secondary structure; for example, a predominantly α-helical protein, such as myoglobin, will have a CD spectrum with negative peaks at 222 and 208 nm and a positive peak around 190 nm. Excitations of valence electrons in the amide group are responsible for the dominating features in this region, where the transitions are from a lone pair on the oxygen to the π^* orbital ($n\pi^*$) and from the nonbonding π_{nb} orbital to the π^* orbital ($\pi_{nb}\pi^*$). Exciton splitting of the latter gives rise to the peaks at 208 and 190 nm. In addition to the polypeptide backbone, other chomophoric groups contribute to the CD in both the far and near UV. Particularly important are the aromatic side chains of amino acids phenylalanine, tyrosine, and tryptophan, due to four associated electronically excited valence states: ^1L_b, ^1L_a, ^1B_b, and ^1B_a as labeled in Platt's notation (*4*). Therefore, the influence of the aromatic residues should be taken into account, both in the near UV, where transitions due to the amide groups do not feature, and in the far UV, where strongly allowed transitions to the aromatic ^1B_b and ^1B_a states occur.

Bands in the near UV of a protein's CD spectrum reflect the side chains that are present, since the aromatic chromophores have characteristic electronic transitions in that region. Exciton coupling of these transitions in closely neighboring side chains may result in more intense bands and couplets in the near UV that allow the proximity of the side chains and their local environment to be determined (*5*). Site-directed mutagenesis may be used to generate a mutant

† This work was supported by grants from the BBSRC and the EPSRC/JREI (Grants 42/B15240 and GR/62052/01, respectively).

[1] Abbreviations: CD, circular dichroism; UV, ultraviolet; FIS, factor for inversion stimulation; PDB, Protein Data Bank; NMA, *N*-methylacetamide; CASSI, complete active space state interaction.

10.1021/bi049031n CCC: $27.50
Published on Web 08/06/2004

in which an aromatic side chain in the wild type is replaced with a nonaromatic residue, for example. The difference spectrum generated from the respective CD spectra gives the contribution of the particular side-chain residue to the total CD (*5*).

Theoretical CD calculations employing the dipole interaction model (*6*) or the matrix method (*7*) have been successful at predicting spectra in the far UV (*8*−*11*), in particular, for proteins with a high α-helical content. A recent study (*12*) employed the matrix method in its origin-independent form (*13*) to calculate from structures determined by X-ray crystallography the far-UV CD spectra of 47 proteins, whose CD spectra have been made publicly available (*14*, *15*). The agreement between the calculated and experimental spectra was good; considering two amide transitions (nπ^* and $\pi_{nb}\pi^*$), a mean absolute error of 5400 deg cm^2 dmol^{-1} over the wavelength range of 190−230 nm was obtained. The parameters describing the amide electronic transitions were derived from ab initio calculations (*10*, *11*) on *N*-methylacetamide (NMA), where a dielectric continuum model was employed to represent bulk solvent.

Using parameters derived from semiempirical calculations and experiment, Woody and co-workers have incorporated the side-chain chromophores of phenylalanine, tyrosine, and tryptophan into matrix method calculations. The total and difference spectra of several proteins have been calculated (*16*−*20*), and the side chains influenced the computed CD in both the far- and near-UV regions. For instance, the CD computed for bovine pancreatic trypsin inhibitor was sensitive to the number and character of the side-chain transitions included in the calculations and the local environment of the side chains (*16*, *18*). The overall spectra computed for mutants Tyr21Leu and Phe22Leu (the residues with the greatest number of neighboring aromatic side chains) agreed poorly with experiment, implying that local conformation and aromatic coupling of side chains in the proximity affected the computed CD.

The aim of our study is to improve the description of the aromatic side chains for use in the matrix method and to investigate their inclusion in the calculation of protein CD from first principles. New parameters describing electronic transitions to the ^1L_b, ^1L_a, ^1B_b, and ^1B_a valence states of the aromatic side chains of phenylalanine, tyrosine, and tryptophan are presented. They are enhanced variants of the parameters presented in a recent study (*21*). Rotational strengths and CD were calculated for 30 proteins whose near- and far-UV CD spectra are available in the literature, along with their X-ray crystal structures. The calculated CD for each protein is compared with experiment and with spectra obtained using an alternative semiempirical side-chain parameter set.

METHODS

Computational Details. The matrix method (*7*) enables the calculation of CD spectra for large biomolecules, such as proteins. The method is well-established; however, for completeness, it is summarized in the Appendix. The parameters in the matrix method that describe the peptide backbone nπ^* and $\pi_{nb}\pi^*$ transitions at 220 and 193 nm, respectively, are derived from calculations on NMA and have been described previously (*11*). To calculate the side-chain contributions, new parameter sets describing electronic transitions to the four electronically excited valence states (^1L_b, ^1L_a, ^1B_b, and ^1B_a) of the side chains of phenylalanine, tyrosine, and tryptophan were employed. These sets are derived from ab initio calculations on the chromophores benzene, phenol, and indole, respectively (*22*), which were performed at the complete active space self-consistent field level of theory (*23*) with the polarizable continuum model (*24*), a self-consistent reaction field technique, to mimic bulk solvent. Transition energies were calculated using multiconfigurational second-order perturbation theory (*25*). Complete active space state interaction (CASSI) calculations (*26*) furnished permanent and transition electric and magnetic moments and the associated charge densities. The derivation of the new side-chain parameters follows analogous work (*11*) on the peptide backbone based on calculations on NMA; the new side-chain parameters are derived from the ab initio calculations with water as the bulk solvent.

Charges for use in eq 5 (in the Appendix) are constructed to reproduce the total electrostatic potentials such that the least-squares difference is minimized. Five charges are located around each non-hydrogen atom, giving a total of 30, 35, and 45 charges for benzene, phenol, and indole, respectively. The electronic transitions involve a translation of charge in the symmetry plane, and the charges are arranged in the symmetry plane (one centrally located on each atom with two orthogonal pairs of dipoles that intersect at the atomic center). The charges were adjusted to conserve the total charge of each system. Using the procedure described above, charges describing the five permanent electronic states (S_0, ^1L_b, ^1L_a, ^1B_b, and ^1B_a) for each system were generated, in addition to charges that represent transitions between the states. The electrostatic potentials evaluated for each system were rotated in the symmetry plane to match the ab initio dipole orientations (*22*, *27*). In the parameters presented in this work, the electric dipole moments take values from the charge fitting calculations. The mean absolute errors for the electric dipole moments described by the fitted charges are 0.01, 0.07, and 0.10 D for benzene, phenol, and indole, respectively. The new side-chain parameters and further details of their calculation are provided in the Supporting Information.

Benzene possesses D_{6h} symmetry; therefore, the symmetry-allowed transitions are from the ground state S_0 to the degenerate ^1B_a and ^1B_b ($^1E_{1u}$) states, where the D_{6h} label is in parentheses. The two lowest transitions to ^1L_b ($^1B_{2u}$) and ^1L_a ($^1B_{1u}$) are dipole-forbidden; however, they become accessible due to vibronic coupling. Therefore, for benzene, we only consider electronic excitations from S_0 to the ^1L_b, ^1L_a, ^1B_b, and ^1B_a states, i.e., four transitions. Transitions between the excited states are symmetry-forbidden (in D_{6h} symmetry). The ab initio calculations on benzene neglected the effect of vibronic coupling on the excitations to the ^1L_b ($^1B_{2u}$) and ^1L_a ($^1B_{1u}$) states. Therefore, following Woody and co-workers (*18*), we augmented the ab initio parameters describing phenylalanine with four sets of atom-centered charges (six charges for ^1L_b and four for ^1L_a) that reproduce the two orthogonal components of the transition moments for the ^1L_b and ^1L_a states such that, when combined with the calculated transition energies, they yield the observed oscillator strengths of 0.015 and 0.063 for ^1L_b and ^1L_a, respectively (*28*). In atomic units, the absolute magnitude, $|\vec{\mu}|$, of

Table 1: Thirty Proteins Studied and Their Secondary Structure Class

class and proteins (PDB entry)	no. of side-chain residues in monomer chain			no. of chains
	Phe	Tyr	Trp	
α-helix				
human growth hormone (1axi) (*30*)	11	9	1	1
cytokine (1iar) (*31*)	6	2	1	1
relaxin (6rlx) (*32*)	1	1	1	2
calmodulin (1cfd) (*33*)	8	2	0	1
interleukin-6 (1alu) (*34*)	7	3	1	1
serum albumin (1n5u) (*35*)	31	18	1	1
phospholipase C (1qm6) (*36*)	15	26	10	2
staphylococcal nuclease (1snc) (*37*)	3	7	1	1
hen egg white lysozyme (1hf4) (*38*)	3	3	6	2
apolipophorin III (1aep) (*39*)	2	0	2	1
horse cytochrome *c* (1hrc) (*40*)	4	4	1	1
β-lactamase (1btl) (*19*, *41*)	5	4	4	1
mixed α,*β*				
sticholysin II (1gwy) (*42*)	6	12	5	2
papain (9pap) (*43*)	4	19	5	1
cyclodextrin glucanotransferase (1pam) (*44*)	33	34	13	1
monellin (1mol) (*45*)	5	7	1	2
subtilisin (1sbt) (*46*)	2	7	3	1
ribonuclease T-1 (1i0v) (*47*)	4	9	1	1
thioredoxin (2trx) (*48*)	4	2	2	2
phosphatidylethanolamine-binding protein (1a44) (*49*)	3	10	5	1
glucose oxidase (1cf3) (*50*)	18	27	10	1
barstar (1ay7) (*51*)	2	3	3	1
dihydrofolate reductase (5dfr) (*52*)	6	4	5	1
ribonuclease A (3rn3) (*20*)	3	6	0	1
barnase (1ban) (*53*)	4	7	3	3
β-I				
pertactin (1dab) (*54*)	10	5	7	1
human carbonic anhydrase II (2cba) (*55*)	12	8	7	1
β-II				
chymotrypsinogen A (2cga) (*56*)	6	4	8	2
human tissue factor (2hft) (*57*)	8	11	4	1
bovine pancreatic trypsin inhibitor (5pti) (*18*)	4	4	0	1

one component of the transition moment is

$$|\vec{\mu}| = \left(\frac{3f}{4\Delta E}\right)^{1/2} \quad (1)$$

where f is the observed oscillator strength and ΔE is the transition energy. However, given that vibronic coupling should be accounted for from first principles (as are the unperturbed transitions), this approach may prove to be unsuitable.

To assess our calculations in the far and near UV, we use a set of 30 proteins whose far- and near-UV CD spectra we collated from the literature. Table 1 displays the secondary structure class for each of the 30 proteins. Proteins that contain mostly *β*-sheets may be categorized as class I (*β*-I) or class II (*β*-II), on the basis of their CD spectra (*29*). Class *β*-I proteins, for example, prealbumin, exhibit minima in the wavelength ranges of 170−180 and 210−220 nm. Proteins belonging to the latter class, such as α-chymotrypsin, have spectra that resemble a random coil's spectrum, with minima around 200 nm and positive bands in the range of 185−190 nm. This categorization may reflect the underlying secondary structure; class *β*-I proteins contain fairly regular sheets, whereas class *β*-II proteins contain more irregular strands and *β*-bulges. The proteins in this set were chosen on the basis of the availability of their X-ray crystal structures and on the basis of the resolution of their published CD spectra. Proteins were not included if their experimental ellipticity at 222 nm indicated a helicity significantly different from that suggested by a DSSP (*58*) analysis of the crystal structure, as this might reflect a large conformational change. For example, we did not include human apolipoprotein A-I (PDB entry 1av1), since the experimental ellipticity at 222 nm (*59*) indicated an α-helical content of approximately 60%, whereas the crystal structure shows 85% α-helical secondary structure.

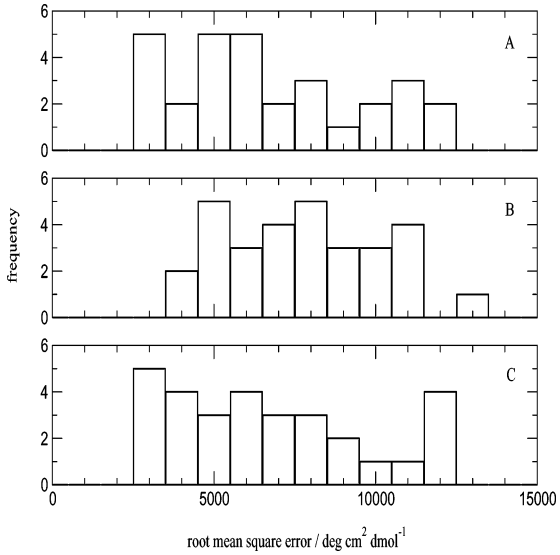

FIGURE 1: Root-mean-square errors in the far UV for the 30 proteins: (A) ab initio backbone, (B) ab initio backbone and semiempirical side-chain parameters, and (C) ab initio backbone and ab initio side-chain parameters.

Rotational strengths were calculated using all distinct chains for multimeric proteins, from the appropriate PDB atomic coordinates. Spectra were generated with Gaussian bands with a width of 12.5 nm. All calculations included the peptide $n\pi^*$ and $\pi_{nb}\pi^*$ transitions described by parameters derived from calculations on NMA (*11*). The alternative semiempirical side-chain set employs four transitions for the side chains of phenylalanine and tyrosine and six transitions for the tryptophan side chain; energies and monopole charges describing the transitions were taken from the URL cited in ref *9*, with the transition dipole moments assigned values consistent with the monopole charges.

RESULTS

Absolute Errors in the Far and Near UV. Figures 1 and 2 show, as histograms, the root-mean-square errors of the CD calculated for the 30 proteins in the far- and near-UV regions, respectively. In the far UV, transitions to the aromatic ^1B_b and ^1B_a states occur in addition to transitions to the ^1L_a states of phenylalanine and tyrosine. Absolute errors for the α-helical proteins are similar for each of the parameter sets in the far UV; inclusion of the semiempirical side-chain parameters improves the agreement with experiment for the proteins interleukin-6 and serum albumin. The ab initio side-chain parameters improve the quality of the calculated CD for relaxin, interleukin-6, serum albumin, and hen egg white lysozyme. The situation is analogous for the mixed α,*β*-proteins; inclusion of the semiempirical side-chain parameters reduces the errors for subtilisin and thioredoxin, with the ab initio set improving the computed spectra for cyclodextrin

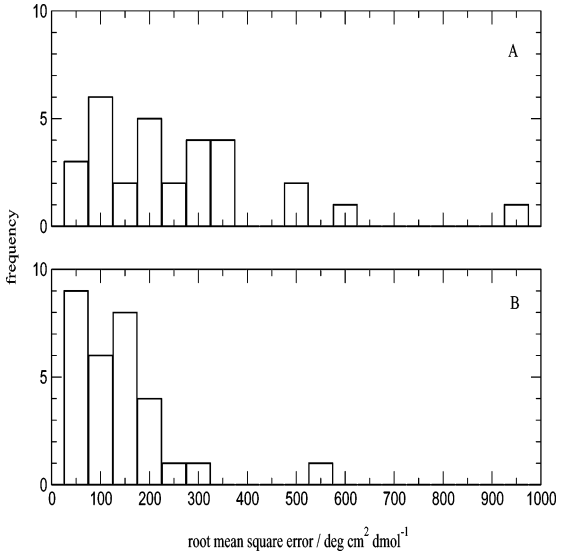

FIGURE 2: Root-mean-square errors in the near UV for the 30 proteins: (A) ab initio backbone and semiempirical side-chain parameters and (B) ab initio backbone and ab initio side-chain parameters.

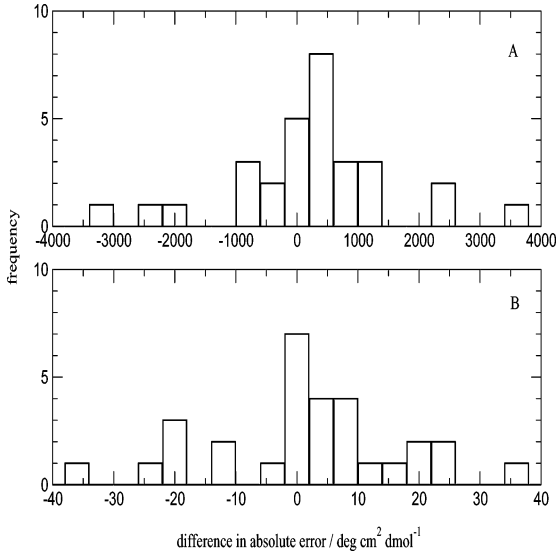

FIGURE 3: Difference in the absolute errors calculated with and without the vibronic monopoles in the (A) far UV and (B) near UV. The plot displays with minus without.

glucanotransferase and thioredoxin. For the β-I and β-II proteins, the smallest errors are obtained using the backbone-only parameters, with the exceptions being bovine pancreatic trypsin inhibitor and chymotrypsinogen A where the semiempirical and ab initio sets yield the smallest errors, respectively.

The errors in the near-UV region (Figure 2) are smaller than in the far-UV region, mainly because the ellipticity of a given protein is weaker in the near UV. Transitions to the ^1L_b states of phenylalanine and tyrosine and to the ^1L_b and ^1L_a states of tryptophan dominate this wavelength region. The absolute errors of the computed CD are generally smaller for the ab initio side-chain parameters, and only the calculated spectrum of monellin is in error by more than 400 deg cm^2 dmol^{-1}, where the ab initio side-chain parameters underestimate the negative band at 277 nm. The semiempirical side-chain parameters give calculated spectra with errors greater than 400 deg cm^2 dmol^{-1} for four proteins: chymotrypsinogen A, phosphatidylethanolamine-binding protein, horse cytochrome *c*, and glucose oxidase. The majority of the spectra obtained with the ab initio side-chain parameters are in error by not more then 200 deg cm^2 dmol^{-1}.

Inclusion of Vibronic Coupling. The contribution that vibronic coupling makes to the computed CD was investigated by augmenting the phenylalanine ab initio parameter set with four groups of monopoles that reproduce the observed $^1L \leftarrow S_0$ transition dipoles for benzene. The effect of these additional charges on the CD in the far and near UV is negligible for most of the 30 proteins (Figure 3). Slight improvements in the calculated spectra were found for some proteins in the far and near UV, but inclusion of the vibronic monopoles worsens the calculated spectra for other proteins. As the vibronic monopoles do not consistently improve the calculated spectra, we do not consider them further.

Mutants. We applied the matrix method with the different side-chain parameter sets to calculate CD spectra for mutants of barnase, human carbonic anhydrase II, and bovine pancreatic trypsin inhibitor. Table 2 displays the selected mutants. For each mutant, the appropriate side-chain residue was removed from the calculation, with the exception of the Tyr → Phe mutants of barnase where the phenolic oxygen was deleted. PDB atomic coordinates of the wild type were used throughout. Figures 4 and 5 display the root-mean-square errors for the computed total and difference (wild type minus mutant) spectra of the 20 mutants in the near UV. The errors of the computed total CD in the far UV for the mutants are comparable to those for the 30 wild-type proteins shown in Figure 1, with little difference between the side-chain parameter sets. However, the semiempirical side-chain parameters led to slightly better difference spectra than the ab initio set in the far UV. In the near UV, the calculated total spectra for the mutants using the ab initio side-chain parameters are in better agreement with experiment; all the errors are less than 300 deg cm^2 dmol^{-1} (Figure 4B). For the semiempirical side-chain parameters, the three proteins with errors greater than 300 deg cm^2 dmol^{-1} in the near UV (Figure 4A) are the Tyr97Phe mutant of barnase and human carbonic anhydrase II mutants Trp192Phe and Trp97Cys. The ab initio side-chain parameters yield difference spectra with errors of less than 200 deg cm^2 dmol^{-1} in the near UV (Figure 5). For the semiempirical side-chain parameters, the two proteins with errors greater than 200 deg cm^2 dmol^{-1} in the near UV (Figure 5A) are the Tyr97Phe mutant of barnase and the Trp16Phe mutant of human carbonic anhydrase II.

Table 2: Mutants of Barnase, Human Carbonic Anhydrase II, and Bovine Pancreatic Trypsin Inhibitor

protein	mutants
barnase (*53*)	Phe7Leu, Phe56Met, Phe82Leu, Tyr78Phe, Tyr90Phe, Tyr97Phe, Trp35Phe, Trp71Phe, Trp94Phe
human carbonic anhydrase II (*55*)	Trp16Phe, Trp97Cys, Trp192Phe
bovine pancreatic trypsin inhibitor (*18*)	Phe4Leu, Phe22Leu, Phe33Leu, Phe45Leu, Tyr10Leu, Tyr21Leu, Tyr23Leu, Tyr35Leu

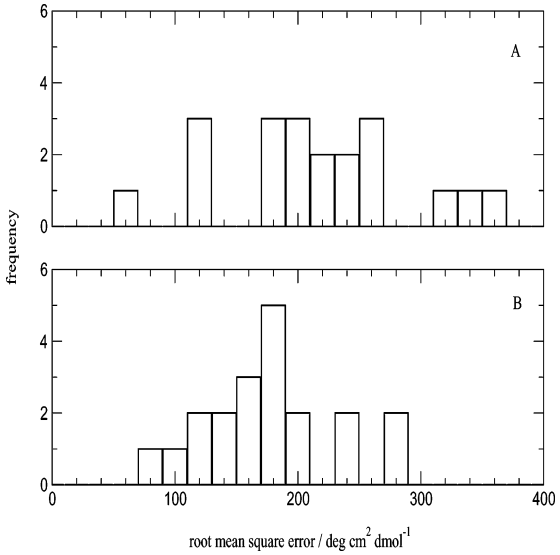

FIGURE 4: Root-mean-square errors in the near UV for the total spectra of the 20 mutants: (A) ab initio backbone and semiempirical side-chain parameters and (B) ab initio backbone and ab initio side-chain parameters.

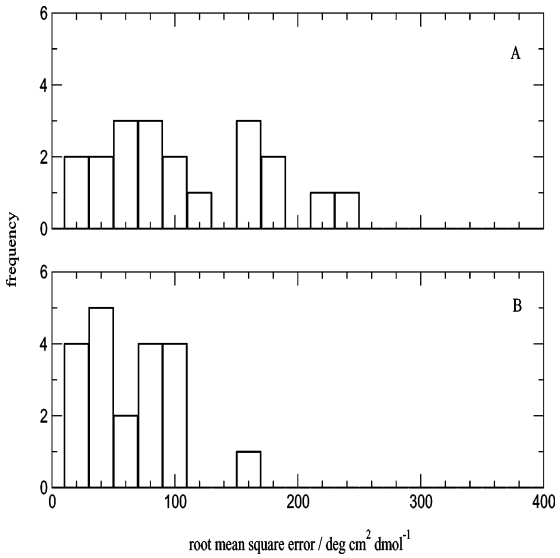

FIGURE 5: Root-mean-square errors in the near UV for the difference spectra of the 20 mutants: (A) ab initio backbone and semiempirical side-chain parameters and (B) ab initio backbone and ab initio side-chain parameters.

Selected Spectra. To illustrate the effectiveness of the new side-chain parameters, we present a representative selection of spectra from the set of 30 proteins. First, we consider β-lactamase, barnase, and human carbonic anhydrase II, since these proteins were included in a recent study (*21*). Figure 6 displays the far- and near-UV experimental (*19*, *41*) and calculated CD spectra for wild-type β-lactamase. This enzyme consists of 263 residues, including five phenylalanines, four tyrosines, and four tryptophans. The predominantly α-helical structure is reflected in the far UV (Figure 6A), where a positive band at 190 nm is observed, in addition to two negative bands around 210 and 220 nm. None of the computed spectra clearly resolves the two negative bands characteristic of an α-helix. However, application of the ab initio side-chain parameters reduces the extent of overpredicted intensity around 210 nm. In the near UV (Figure 6B), the negative band with a minimum around 280 nm is

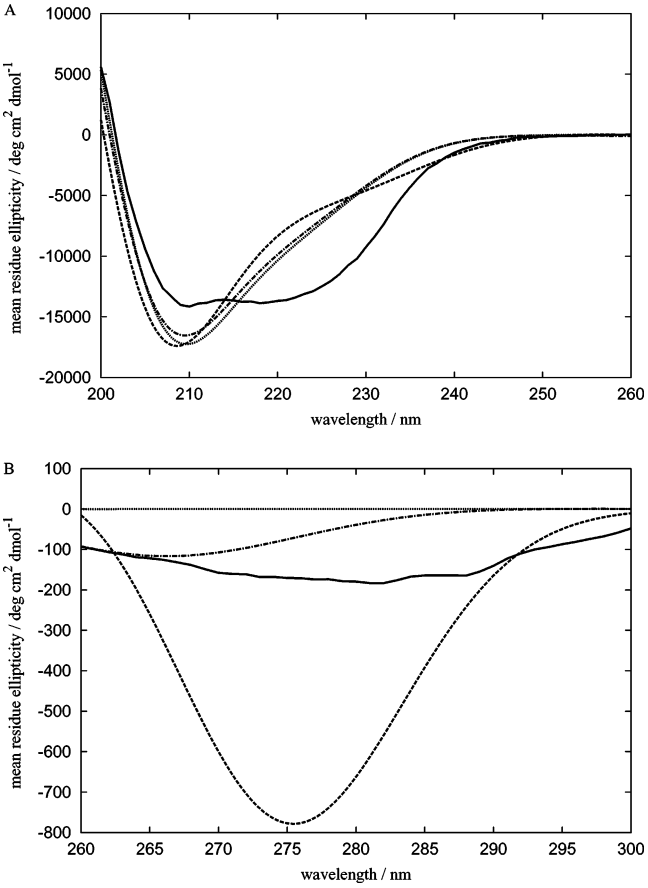

FIGURE 6: Experimental (*19*, *41*) and calculated CD spectra of wild-type β-lactamase in the (A) far UV and (B) near UV: experiment (—), ab initio backbone (•••), ab initio backbone and semiempirical side-chain parameters (- - -), and ab initio backbone and ab initio side-chain parameters (-•-).

overestimated with the semiempirical side-chain parameters, and better captured with the new side-chain set, albeit with the band center shifted to 266 nm.

Figure 7 displays far- and near-UV CD spectra for wild-type barnase (*53*), a protein of the $\alpha+\beta$ class. Each barnase monomer consists of 110 residues, including four phenylalanines, seven tyrosines, and three tryptophans. The CD was calculated for the trimer. A typical far-UV CD spectrum for a protein of this structural class has a positive band at around 195 nm and a negative band around 217 nm, both with low intensities. All the calculated spectra give positive bands between 195 and 200 nm, with the backbone only and with the backbone and the ab initio side-chain parameters predicting negative bands between 210 and 220 nm (Figure 7A). The backbone and semiempirical side-chain parameters reproduce the experimentally observed negative band at 230 nm, which is associated with the Trp94 residue (*53*). The semiempirical tryptophan parameters feature a transition at 225 nm ($^1B_b \leftarrow S_0$), whereas the ab initio calculations on indole place this transition at 204 nm. We discuss this difference later. In the near-UV region (Figure 7B), the observed positive band has a maximum around 280 nm. The semiempirical parameters yield a positive band at 268 nm and a weak negative band at 289 nm. The spectrum obtained using the ab initio side-chain set is qualitatively similiar to, but much weaker than, the semiempirical side-chain spectrum, with positive and negative bands at 260 and 283 nm, respectively.

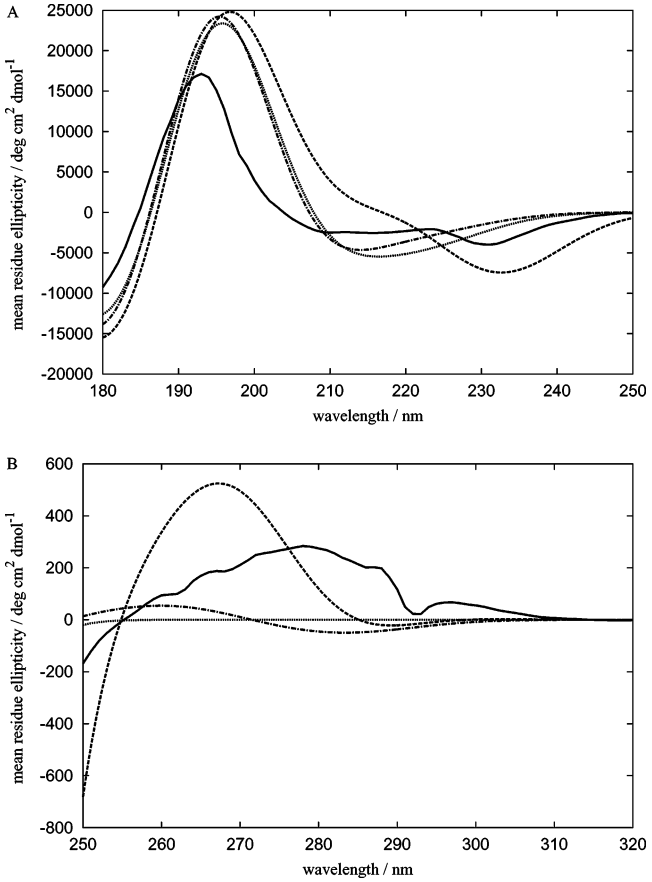

FIGURE 7: Experimental (*53*) and calculated CD spectra of wild-type barnase in the (A) far UV and (B) near UV: experiment (—), ab initio backbone (•••), ab initio backbone and semiempirical side-chain parameters (- - -), and ab initio backbone and ab initio side-chain parameters (-•-).

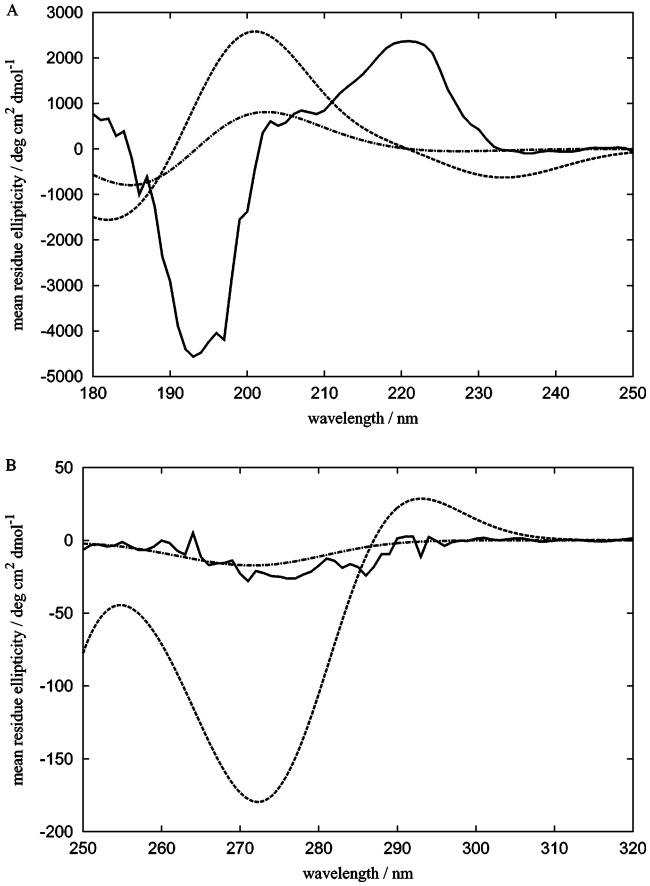

FIGURE 8: Experimental (*53*) and calculated CD difference spectra of the Tyr90Phe mutant of barnase in the (A) far UV and (B) near UV: experiment (—), ab initio backbone and semiempirical side-chain parameters (- - -), and ab initio backbone and ab initio side-chain parameters (-•-).

The far-UV spectrum of the Tyr90Phe mutant of barnase is similar to that of the wild type, with the exception that the negative band at 230 nm (due to Trp94 in the wild type) is less pronounced. The computed spectra are similar to those for the wild type, with the semiempirical side-chain parameters predicting the band around 230 nm and with the backbone-only and the ab initio side-chain parameters agreeing better with experiment in the 210−220 nm range. The experimental (*53*) and calculated near-UV spectra of the Tyr90Phe mutant are very similiar to their counterparts for the wild type. Panels A and B of Figure 8 display the difference spectra for the Tyr90Phe mutant of barnase in the far and near UV, respectively. In the far UV, the observed difference spectrum has two major regions: one negative between 185 and 200 nm and one positive with a maximum around 220 nm. Both the computed difference spectra follow a similiar trend, where the minima and maxima occur at wavelengths around 185 and 200 nm, respectively. The semiempirical side-chain parameters also predict a negative change centered at 233 nm. The observed difference spectrum in the near UV indicates no substantial changes, and is best reproduced with the ab initio side-chain parameters.

Figure 9 displays far- and near-UV CD spectra for wild-type human carbonic anhydrase II. This class I β-sheet protein comprises 260 residues, including 12 phenylalanines, eight tyrosines, and seven tryptophans. The experimental far-UV spectrum (*55*) consists of a very weak positive band at 200 nm and a negative band at 210 nm (Figure 9A). The calculated spectra all show positive bands at 195 nm with each of the parameter sets substantially overestimating the ellipticity at this wavelength. The negative band at 210 nm is not particularly well reproduced by any of the calculations. Figure 9B shows a negative band at 270 nm in the experimental spectrum in addition to two small bands above 280 nm with negative extrema at 287 and 296 nm (*55*). The backbone and semiempirical side-chain parameters give a spectrum with an overly intense negative band at 266 nm and a positive band at 287 nm. The ab initio side-chain parameters give less intense maxima in much better accord with experiment.

The above remarks for the calculated wild-type spectra also apply to the far- and near-UV CD spectra for the Trp192Phe mutant of human carbonic anhydrase II. Difference spectra for the Trp192Phe mutant of human carbonic anhydrase II are shown in panels A and B of Figure 10. In the far UV, the observed difference spectrum has one negative region and one positive region. Both side-chain sets correctly predict a positive change at 200 nm, and the semiempirical side-chain parameters also indicate a positive change at 220−230 nm. The observed difference spectrum (*55*) in the near UV shows two weak negative changes around 270 and 295 nm (Figure 10B). The ab initio set is in closer agreement than the semiempirical set.

The final class of protein, class β-II, is represented in Figure 11, which shows the far- and near-UV experimental spectra (*18*) alongside the calculated spectra for bovine

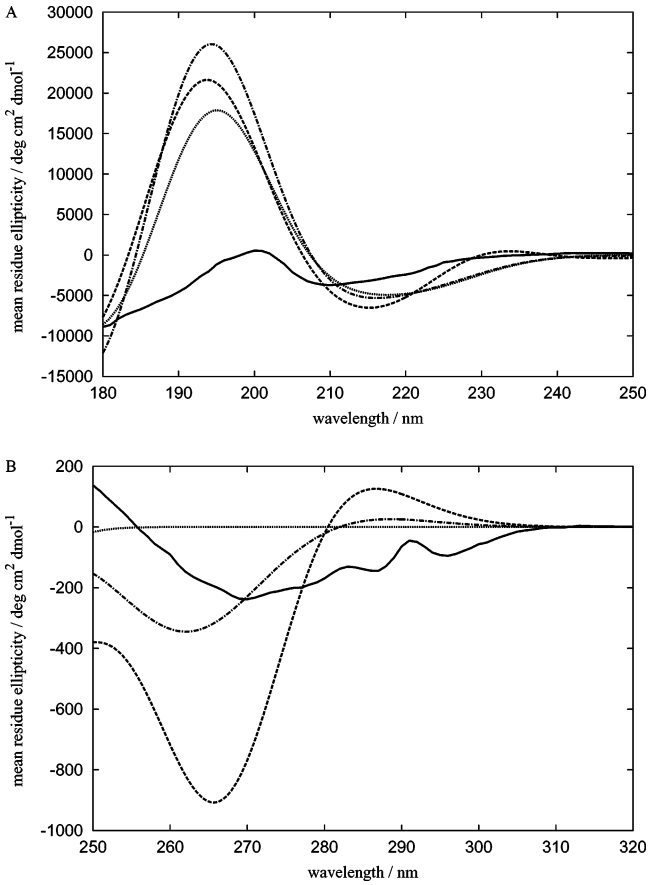

FIGURE 9: Experimental (*55*) and calculated CD spectra of wild-type human carbonic anhydrase II in the (A) far UV and (B) near UV: experiment (—), ab initio backbone (•••), ab initio backbone and semiempirical side-chain parameters (- - -), and ab initio backbone and ab initio side-chain parameters (-•-).

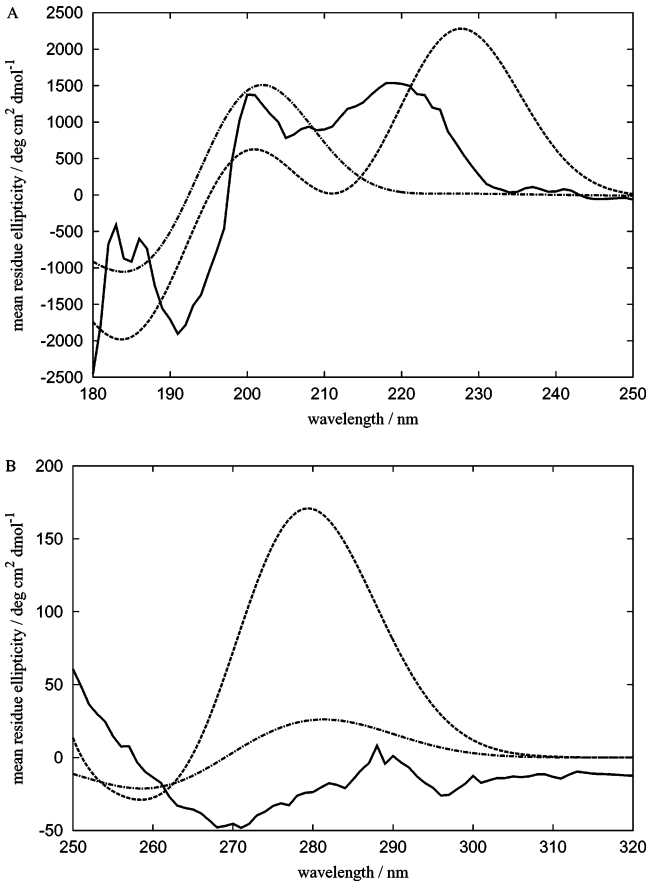

FIGURE 10: Experimental (*55*) and calculated CD difference spectra of the Trp192Phe mutant of human carbonic anhydrase II in the (A) far UV and (B) near UV: experiment (—), ab initio backbone and semiempirical side-chain parameters (- - -), and ab initio backbone and ab initio side-chain parameters (-•-).

pancreatic trypsin inhibitor, a small 58-residue protein with four phenylalanines and four tyrosines. In the far UV (Figure 11A), the experimental spectrum has a negative band around 202 nm, which is absent in the spectrum from the backbone-only parameters and is best reproduced by the semiempirical side-chain parameters. In the near UV (Figure 11B), a negative band centered around 275 nm is observed, which is overestimated with the semiempirical side-chain set and underestimated with the ab initio set.

The experimental (*18*) and calculated far- and near-UV CD spectra for the Tyr21Leu mutant of bovine pancreatic trypsin inhibitor have character similiar to that of the wild-type spectra. Figure 12 displays the computed difference spectra (along with the experimental spectra) for the Tyr21Leu mutant of bovine pancreatic trypsin inhibitor in the far and near UV. In the far UV, the negative difference around 200 nm is reproduced by both side-chain parameter sets, albeit with the minima shifted to longer wavelengths (Figure 12A). The agreement with experiment is good in the near UV, where both side-chain parameter sets give negative differences around 275 nm (Figure 12B). The semiempirical side-chain parameters give a difference spectrum that is closer in magnitude to the experimental difference.

DISCUSSION

We have derived a new set of parameters that describe the electronic transitions in the aromatic side chains of phenylalanine, tyrosine, and tryptophan. When these parameters are combined with the amide parameters of Besley and Hirst (*11*) and incorporated in the matrix method, fully first-principles CD calculations may be performed in the far- and near-UV regions. To determine the influence of the side-chain chromophores on the computed CD in the far and near UV, a set of 30 proteins was considered.

For the 30 wild-type proteins, the mean absolute errors in the far UV are 6100, 7300, and 6300 deg cm^2 dmol^{-1} for the backbone, semiempirical side-chain, and ab initio side-chain parameters, respectively. Mean absolute errors for the near UV are 240 and 110 deg cm^2 dmol^{-1} for the semiempirical side-chain parameters and for the ab initio side-chain parameters, respectively. The far-UV errors for this set may be compared with the absolute errors of Hirst et al. (*12*), where a value of 6100 deg cm^2 dmol^{-1} was obtained for 47 proteins. In the near UV, the ab initio side-chain set yields a mean absolute error that is a 2-fold improvement over the semiempirical side-chain set. The mean absolute errors for the computed total spectra of the 20 mutants are 6900 and 7500 deg cm^2 dmol^{-1} in the far UV and 200 and 170 deg cm^2 dmol^{-1} in the near UV for the semiempirical and ab initio side-chain sets, respectively. For the difference spectra, the mean absolute errors are 1600 and 1800 deg cm^2 dmol^{-1} in the far UV and 98 and 54 deg cm^2 dmol^{-1} in the near UV for the semiempirical and ab initio side-chain parameters, respectively. The difference spectra presented in Figures 8, 10, and 12 indicate the contribution of a single side-chain

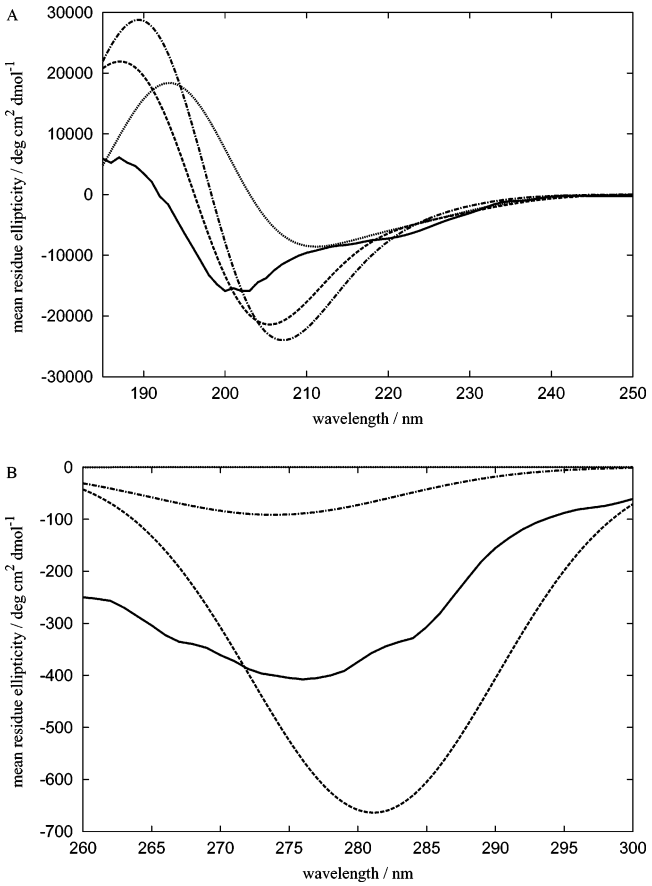

FIGURE 11: Experimental (*18*) and calculated CD spectra of wild-type bovine pancreatic trypsin inhibitor in the (A) far UV and (B) near UV: experiment (—), ab initio backbone (•••), ab initio backbone and semiempirical side-chain parameters (- - -), and ab initio backbone and ab initio side-chain parameters (-•-).

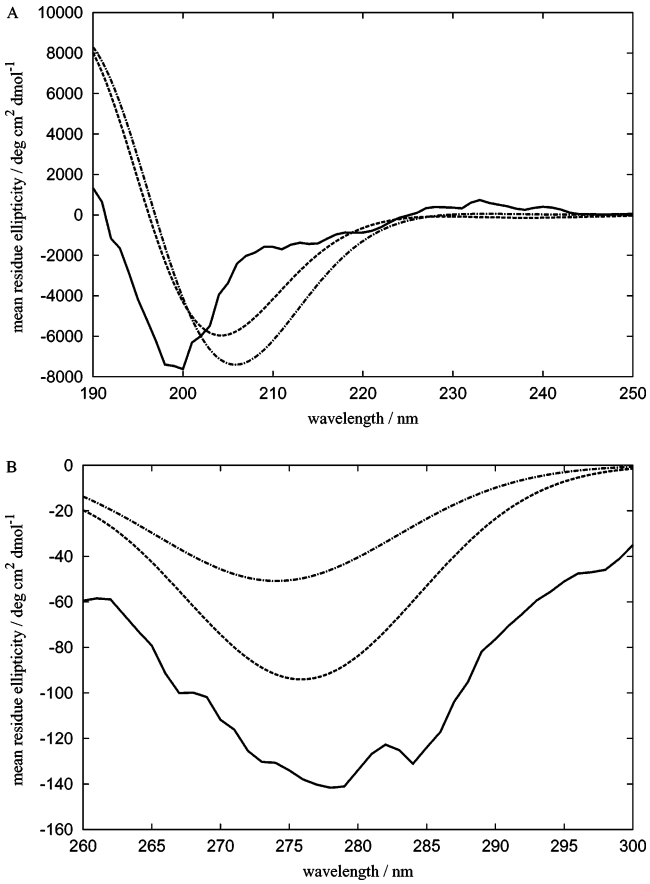

FIGURE 12: Experimental (*18*) and calculated CD difference spectra of the Tyr21Leu mutant of bovine pancreatic trypsin inhibitor in the (A) far UV and (B) near UV: experiment (—), ab initio backbone and semiempirical side-chain parameters (- - -), and ab initio backbone and ab initio side-chain parameters (-•-).

residue to the CD of barnase, human carbonic anhydrase II, and bovine pancreatic trypsin inhibitor, respectively. The contribution of side chains to the ellipticity of proteins is important when one is concerned, for example, with estimating the degree of helicity by measuring CD at 222 nm (*60*).

Spearman rank correlation coefficients at key wavelengths of 190, 208, and 222 nm were calculated. These wavelengths are points of extrema (positive and negative) in a helical CD spectrum and are of particular interest. At 190 nm, the side-chain parameter sets yield correlation coefficients of approximately 0.7. The backbone-only parameters give a correlation of 0.87, and a similiar level of correlation is evident at 208 and 222 nm. At these latter two wavelengths, both side-chain sets show correlations above 0.8. At 222 nm, for example, the side-chain chromophores are predicted to give a positive contribution to the CD for the majority of the 30 proteins. Further assessment based on a larger set of 47 proteins, as reported by Hirst et al. (*12*), suggests that any adverse effect in the far UV is relatively minor. For the larger set, the correlation at 222 nm between observed and computed CD (using each of the parameter sets) is greater than 0.9 and slightly enhanced by the new side-chain parameters, although the correlation at 190 nm falls by 0.06 for the backbone-only parameters. However, the emphasis in the current study is on the quality of the calculations in the near UV, where backbone amide transitions do not obscure the weaker side-chain transitions.

The ab initio calculations on benzene (the phenylalanine side-chain chromophore) neglected vibronic coupling and, therefore, gave small transition dipoles to the ^{1}L states. To investigate the effect of this, sets of charges that mimic the vibronic coupling in benzene were added to the phenylalanine parameters to account for the missing intensity for transitions to the ^{1}L states. The inclusion of the vibronic monopoles had a mixed effect on the CD computed for the 30-protein set, with a negligible change in the absolute error for the majority of the proteins in the far and near UV (Figure 3). The benefit of this sort of ad hoc approach is at best marginal. A more rigorous inclusion of vibronic coupling should be derived from first principles, for example, using a methodology that considers the vibrational modes of benzene that induce the coupling (*61*). However, such an approach is beyond the scope of the current study.

There are some notable differences between the semiempirical and the ab inito tryptophan side-chain parameters; the number of transitions is greater in the semiempirical set, and the wavelength of the $^1B_b \leftarrow S_0$ transition differs by 21 nm. In the gas phase, the vertical transition occurs at 206 nm (*62*). When indole is in a stretched polyethylene environment, the transition has a band maximum at 215 nm; for 3-methylindole, the transition has a maximum at 222 nm (*63*). To study the influence of the latter discrepancy on the CD computed with the ab initio side-chain parameters, we shifted the $^1B_b \leftarrow S_0$ transition to 225 nm in the ab initio parameters that describe tryptophan's side chain. For the 47-

protein set, the resulting correlation between observed and computed CD at 222 nm is 0.85, which is lower than the correlation found for the original set ($r_{\lambda=222} = 0.92$). With respect to the pseudosymmetry long axis of indole (with the nitrogen in the fourth quadrant), the orientation of the ab initio electric transition dipole moment is −34°, which differs from the experimental value of 0 ± 15° (*63*). The semi-empirical set gives the transition dipole an angle of 15°, which is close to previously reported orientations of 16° (*62*) and 23° (*22*) calculated in the gas phase and just falls within the error range for the condensed phase experimental value (*63*). CD was therefore recomputed for the 47 proteins using the ab initio side-chain parameters with the parameters describing indole's $^1B_b \leftarrow S_0$ transition (at 225 nm) rotated to give a transition electric moment of 0°. The resulting correlation at 222 nm is 0.90, which is comparable to the correlation of the initial parameter sets. For barnase (wild type and Tyr90Phe mutant), the total and difference far-UV spectra obtained using this hybrid set are similiar to the spectra obtained using the original set. The negative band at 230 nm, observed for the wild type, is not reproduced; however, the ellipticity recomputed for the wild type is more negative in the region of 215−240 nm. Spectra recomputed for human carbonic anhydrase II differ from the spectra shown in Figures 9 and 10. In the far UV, the recomputed total wild-type spectrum is less negative in the range of 211−224 nm, and for the difference spectrum, a second positive change is found with a maximum at 226 nm. Clearly, the computed CD is sensitive to both the orientation and energy of the $^1B_b \leftarrow S_0$ transition in indole. However, there is not a compelling case for empirical modifications of the ab initio side-chain parameter set, and as it stands, it represents a fully self-consistent model. The sensitivity of the ^1B transition moments to different environments arises from the extent of configurational mixing between the ^1B_b and ^1B_a states. The ratio of the oscillator strengths of ^1B_b to ^1B_a for indole (from the ab initio calculations in water) is 1.1, implying that a significant amount of configurational mixing will occur between these two states. The new side-chain parameters were derived from ab initio calculations that considered water as the bulk solvent; this choice of solvent is consistent with the backbone parameters derived from calculations on NMA in water. More generally, one could consider constructing alternative parameter sets to describe the chromophores in less polar environments, such as protein interiors, and using different parameter sets depending on the extent of burial of the chromophore. However, the additional complexity of the protein CD calculations would be less than desirable.

Computed CD may be sensitive to the conformational dynamics of the protein. We plan to explore this aspect in a future study, by computing spectra for structural snapshots from molecular dynamics simulations and for families of structures from NMR experiments. In addition, we have considered only two transitions for the peptide backbone. Higher-energy amide transitions and charge-transfer transitions between neighboring amide groups may couple with the aromatic side-chain transitions in the far UV and alter the character of the computed CD spectra. We are actively exploring this.

The incorporation of aromatic side chains into the matrix method is necessary for reproduction of qualitative features of a CD spectrum, and is essential in the near UV. The combination of the peptide backbone parameters with the improved ab initio side-chain parameters yields spectra that broadly agree with experiment for the majority of the proteins that have been studied. Although more remains to be done, this work takes us a step closer to fully quantitative first-principles protein CD calculations.

ACKNOWLEDGMENT

We thank Mr. T. M. Watson (University of Nottingham) for technical assistance and Drs. A. T. B. Gilbert and N. A. Besley (University of Nottingham) for helpful discussions.

APPENDIX

The methodology considers a protein to comprise M chromophores, with each having a set of characteristic electronic transitions. The interaction of these chromophores and their transitions gives a series of delocalized transitions (and corresponding intensities, i.e., rotational strengths) for the protein as a whole. The interaction of the isolated chromophores is described by the Coulombic interaction of charge densities associated with each transition. The charge densities are approximated by sets of point charges that reproduce either the electrostatic potentials arising from ab initio densities or the appropriate electric moment associated with each electronic transition.

The total wave function Ψ_k for a state k of a biomolecule may be expressed as a linear combination of basis functions Φ_{ia}:

$$\Psi_k = \sum_i^M \sum_a^{n_i} c_{ia}^k \Phi_{ia} \tag{2}$$

where a is an excited state on chromophore i, n_i is the number of transitions on a group, and c_{ia}^k is an expansion coefficient. Each basis function is a product of M monomer wave functions

$$\Phi_{ia} = \phi_{10}\cdots\phi_{ia}\cdots\phi_{j0}\cdots\phi_{M0} \tag{3}$$

where the first and second subscripts denote the chromophore and the electronic state, respectively. For example, ϕ_{ia} labels a transition from the ground state to excited state a on monomer i. The Hamiltonian operator $\hat{H}$ for the biomolecule may then be defined as

$$\hat{H} = \sum_{i=1}^{M} \hat{H}_i + \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \hat{V}_{ij} \tag{4}$$

where $\hat{H}_i$ is the Hamiltonian for chromophore i and $\hat{V}_{ij}$ is the interaction between monomers i and j. The diagonal elements of the Hamiltonian matrix are the excitation energies of the electronically excited states, with the off-diagonal elements, V_{ij}, arising from interactions between groups or from interactions between excitations on the same group. A typical off-diagonal element takes the form

$$V_{i0a;j0b} = \int\int \frac{\rho_{i0a}(\mathbf{r}_i)\rho_{j0b}(\mathbf{r}_j)}{4\pi\epsilon_0 r_{ij}} \, d\mathbf{r}_i \, d\mathbf{r}_j \tag{5}$$

where $\rho_{i0a}(\mathbf{r}_i)$ and $\rho_{j0b}(\mathbf{r}_j)$ represent permanent and transition electron densities on chromophores i and j, respectively, ϵ_0

is the permittivity of free space, and r_{ij} is the distance between the chromophores. The densities in eq 5 may be represented by point charges, thereby replacing the double integral in eq 5 with a sum over the discrete charges:

$$V_{i0a;j0b} = \sum_{s=1}^{N_s}\sum_{t=1}^{N_t}\frac{q_s q_t}{r_{st}} \quad (6)$$

where q_s and q_t are point charges on monomers i and j, respectively, with the number of charges in each set defined by N_s and N_t, respectively. The Hamiltonian matrix is diagonalized to yield the transition energies (eigenvalues) of the composite system and the mixing coefficients (eigenvectors) that describe the contributions of the excited states of individual chromophores to the interacting system. The rotational strength of each delocalized transition may be calculated from an expression (eq 7), derived (*64*) from the Rosenfeld equation (*65*), which involves the imaginary part (denoted by Im) of the scalar product of the electric and magnetic transition dipole moments $\vec{\mu}$ and $\vec{m}$.

$$R_{0k} = \mathrm{Im}(\langle\Psi_0|\hat{\mu}|\Psi_k\rangle\langle\Psi_k|\hat{m}|\Psi_0\rangle) \quad (7)$$

Application of eq 7 is origin-independent only for electric moments in the dipole velocity representation; we employ the dipole length representation, following Goux and Hooker (*13*), and calculate the origin-independent chiral strength:

$$c_{0k} = \frac{\alpha}{3}\mathrm{Re}(\langle\Psi_0|\hat{p}|\Psi_k\rangle\langle\Psi_k|\hat{L}|\Psi_0\rangle) \quad (8)$$

where Re denotes the real part of the scalar product, α is the fine structure constant, $\vec{p}_{0k}$ is the linear momentum, and $\vec{L}_{k0}$ is the angular momentum moment. The latter quantities are obtained from the electric and magnetic moment vectors of the individual chromophores. The unitary transformation that diagonalized the Hamiltonian is used to transform the linear and angular momentum moment vectors before application of eq 8. The resulting chiral strength is then readily converted to rotational strength (*13*). Finally, a CD spectrum is obtained by centering a band function, such as a Gaussian curve, around the rotational strengths calculated for the specific transition energies.

SUPPORTING INFORMATION AVAILABLE

Ab initio side chain parameters for phenylalanine, tyrosine, and tryptophan. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Chen, E., Goldbeck, R. A., and Kliger, D. S. (2003) Earliest events in protein folding: Submicrosecond structure formation in reduced cytochrome C, *J. Phys. Chem. A 107*, 8149−8155.
2. Woody, R. W. (1994) Circular Dichroism of Peptides and Proteins, in *Circular Dichroism: Principles and Applications* (Nakanishi, K., Berova, N., and Woody, R. W., Eds.) pp 473−496, VCH Publishers, New York.
3. Boswell, S., Mathew, J., Beach, M., Osuna, R., and Colón, W. (2004) Variable contributions of tyrosine residues to the structural and spectroscopic properties of the factor for inversion stimulation, *Biochemistry 43*, 2964−2977.
4. Platt, J. R. (1949) Classification of spectra of cata-condensed hydrocarbons, *J. Chem. Phys. 17*, 484−495.
5. Woody, R. W., and Dunker, A. K. (1996) Aromatic and Cystine Side-Chain Circular Dichroism in Proteins, in *Circular Dichroism and the Conformational Analysis of Biomolecules* (Fasman, G. D., Ed.) pp 109−157, Plenum Press, New York.
6. Applequist, J., Sunberg, K. R., Olson, M. L., and Weiss, L. C. (1979) A normal mode treatment of optical properties of a classical coupled dipole oscillator system with Lorentzian band shapes, *J. Chem. Phys. 70*, 1240; *71*, 2330−2330 (Erratum).
7. Bayley, P. M., Nielsen, E. B., and Schellman, J. A. (1969) The rotatory properties of molecules containing two peptide groups, *J. Phys. Chem. 73*, 228−243.
8. Bode, K. A., and Applequist, J. (1998) Globular protein ultraviolet circular dichroic spectra. Calculation from crystal structures via the dipole interaction model, *J. Am. Chem. Soc. 120*, 10938−10946.
9. Woody, R. W., and Sreerama, N. (1999) Comment on "Improving protein circular dichroism in the far-ultraviolet through reparametrizing the amide chromophore", *J. Chem. Phys. 111*, 2844−2845.
10. Hirst, J. D., and Besley, N. A. (1999) Response to "Comment on 'Improving protein circular dichroism in the far-ultraviolet through reparametrizing the amide chromophore'", *J. Chem. Phys. 111*, 2846−2847.
11. Besley, N. A., and Hirst, J. D. (1999) Theoretical studies toward quantitative protein circular dichroism calculations, *J. Am. Chem. Soc. 121*, 9636−9644.
12. Hirst, J. D., Colella, K., and Gilbert, A. T. B. (2003) Electronic circular dichroism of proteins from first-principles calculations, *J. Phys. Chem. B 107*, 11813−11819.
13. Goux, W. J., and Hooker, T. M., Jr. (1980) Chiroptical properties of proteins. I. Near-ultraviolet circular dichroism of ribonuclease S, *J. Am. Chem. Soc. 102*, 7080−7087.
14. Pancoska, P., Bitto, E., Janota, V., Urbanova, M., Gupta, V. P., and Keiderling, T. A. (1995) Comparison of and limits of accuracy for statistical analyses of vibrational and electronic circular dichroism spectra in terms of correlations to and predictions of protein secondary structure, *Protein Sci. 4*, 1384−1401.
15. Sreerama, N., Venyaminov, S. Y., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: Inclusion of denatured proteins with native proteins in the analysis, *Anal. Biochem. 287*, 243−251.
16. Manning, M. C., and Woody, R. W. (1989) Theoretical study of the contribution of aromatic side chains to the circular dichroism of basic bovine pancreatic trypsin inhibitor, *Biochemistry 28*, 8609−8613.
17. Grishina, I. B., and Woody, R. W. (1994) Contributions of tryptophan side chains to the circular dichroism of globular proteins: Exciton couplets and coupled oscillators, *Faraday Discuss. 99*, 245−262.
18. Sreerama, N., Manning, M. C., Powers, M. E., Zhang, J.-X., Goldenberg, D. P., and Woody, R. W. (1999) Tyrosine, phenylalanine, and disulfide contributions to the circular dichroism of proteins: Circular dichroism spectra of wild-type and mutant bovine pancreatic trypsin inhibitor, *Biochemistry 38*, 10814−10822.
19. Christov, C., Gabriel, S., Atanasov, B., and Fleischhauer, J. (2001) Calculation of the CD spectrum of class A β-lactamase from *Escherichia coli* (TEM-1), *Z. Naturforsch. 56a*, 757−760.
20. Woody, A.-Y. M., and Woody, R. W. (2003) Individual tyrosine side-chain contributions to circular dichroism of ribonuclease, *Biopolymers 72*, 500−513.
21. Rogers, D. M., and Hirst, J. D. (2004) Calculations of protein circular dichroism from first principles, *Chirality 16*, 234−243.
22. Rogers, D. M., and Hirst, J. D. (2003) Ab initio study of aromatic side chains of amino acids in gas phase and solution, *J. Phys. Chem. A 107*, 11191−11200.
23. Roos, B. O. (1987) The Complete Active Space Self-Consistent Field Method and its Applications in Electronic Structure Calculations, in *Ab Initio Methods in Quantum Chemistry Part II* (Prigogine, I., and Rice, S. A., Eds.) Vol. 69, pp 399−446, Wiley, New York.
24. Miertus, S., Scrocco, E., and Tomasi, J. (1981) Electrostatic interaction of a solute with a continuum: a direct utilization of ab initio molecular potentials for the prevision of solvent effects, *Chem. Phys. 55*, 117−129.
25. Roos, B. O., Fülscher, M. P., Malmqvist, P.-Å., Merchán, M., and Serrano-Andrés, L. (1995) Theoretical studies of electronic spectra of organic molecules, in *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy* (Langhoff, S. R., Ed.) pp 357−438, Kluwer Academic Publishers, Dordrecht, The Netherlands.

26. Malmqvist, P.-Å., and Roos, B. O. (1989) The CASSCF state interaction method, *Chem. Phys. Lett. 155*, 189−194.
27. The dipole moment components and orientations reported for indole in bulk water in ref *22* are incorrect. The parameters presented here are consistent with the correct orientation of the indole chromophore. Correct orientations of the dipole moments are as follows: −48° for S_0, +44° for $^1L_b \leftarrow S_0$, −48° for $^1L_a \leftarrow S_0$, −34° for $^1B_b \leftarrow S_0$, and +30° for $^1B_a \leftarrow S_0$ (angles defined by the direction of the dipole moment with respect to the long axis of the molecule with the nitrogen in the fourth quadrant).
28. Bayliss, N. S., and Hulme, L. (1953) Solvent effects in the spectra of benzene, toluene, and chlorobenzene at 2600 and 2000 Å, *Aust. J. Chem. 6*, 257−277.
29. Manavalan, P., and Johnson, W. C., Jr. (1983) Sensitivity of circular dichroism to protein tertiary structure class, *Nature 305*, 831−832.
30. DeFelippis, M. R., Kilcomons, M. A., Lents, M. P., Youngman, K. M., and Havel, H. A. (1995) Acid stabilization of human growth hormone equilibrium folding intermediates, *Biochim. Biophys. Acta 1247*, 35−45.
31. Windsor, W. T., Syto, R., Le, H. V., and Trotta, P. P. (1991) Analysis of the conformation and stability of *Escherichia coli* derived recombinant human interleukin 4 by circular dichroism, *Biochemistry 30*, 1259−1264.
32. Shire, S. J., Holladay, L. A., and Rinderknecht, E. (1991) Self-association of human and porcine relaxin as assessed by analytical ultracentrifugation and circular dichroism, *Biochemistry 30*, 7703−7711.
33. Findlay, W. A., Martin, S. A., Beckingham, K., and Bayley, P. M. (1995) Recovery of native structure by calcium binding site mutants of calmodulin upon binding of sk-MLCK target peptides, *Biochemistry 34*, 2087−2094.
34. Matthews, J. M., Hammacher, A., Howlett, G. J., and Simpson, R. J. (1998) Physicochemical characterization of an antagonistic human interleukin-6 dimer, *Biochemistry 37*, 10671−10680.
35. Dockal, M., Carter, D. C., and Rüker, F. (2000) Conformational transitions of the three recombinant domains of human serum albumin depending on pH, *J. Biol. Chem. 275*, 3042−3050.
36. Naylor, C. E., Jepson, M., Crane, D. T., Titball, R. W., Miller, J., Basak, A. K., and Bolgiano, B. (1999) Characterisation of the calcium-binding C-terminal domain of *Clostridium perfringens* α-toxin, *J. Mol. Biol. 294*, 757−770.
37. Filfil, R., and Chalikian, T. V. (2000) Volumetric and spectroscopic characterizations of the native and acid-induced denatured states of staphylococcal nuclease, *J. Mol. Biol. 299*, 827−842.
38. Buck, M., Radford, S. E., and Dobson, C. M. (1993) A partially folded state of hen egg white lysozyme in trifluoroethanol: Structural characterization and implications for protein folding, *Biochemistry 32*, 669−678.
39. Weers, P. M. M., Kay, C. M., Oikawa, K., Wientzek, M., Van der Horst, D. J., and Ryan, R. O. (1994) Factors affecting the stability and conformation of *Locusta migratoria* apolipophorin III, *Biochemistry 33*, 3617−3624.
40. Sasahara, K., McPhie, P., and Minton, A. P. (2003) Effect of dextran on protein stability and conformation attributed to macromolecular crowding, *J. Mol. Biol. 326*, 1227−1237.
41. Vanhove, M., Raquet, X., and Frere, J.-M. (1995) Investigation of the folding pathway of the TEM-1 β-lactamase, *Proteins: Struct., Funct., Genet. 22*, 110−118.
42. Mancheño, J. M., De Los Rios, V., Martinez Del Pozo, Á., Lanio, M. E., Oñaderra, M., and Gavilanes, J. G. (2001) Partially folded states of the cytolytic protein sticholysin II, *Biochim. Biophys. Acta 1545*, 122−131.
43. Edwin, F., and Jagannadham, M. V. (1998) Sequential unfolding of papain in molten globule state, *Biochem. Biophys. Res. Commun. 252*, 654−660.
44. Yamamoto, T., Shiraki, K., Fujiwara, S., Takagi, M., Fukui, K., and Imanaka, T. (1999) In vitro heat effect on functional and conformational changes of cyclodextrin glucanotransferase from hyperthermophilic archaea, *Biochem. Biophys. Res. Commun. 265*, 57−61.
45. Fan, P., Bracken, C., and Baum, J. (1993) Structural characterization of monellin in the alcohol-denatured state by NMR: Evidence for β-sheet to α-helix conversion, *Biochemistry 32*, 1573−1582.
46. Eder, J., Rheinnecker, M., and Fersht, A. R. (1993) Folding of subtilisin BPN′: Characterization of a folding intermediate, *Biochemistry 32*, 18−26.
47. Shastry, M. C. R., and Eftink, M. R. (1996) Reversible thermal unfolding of ribonuclease T_1 in reverse micelles, *Biochemistry 35*, 4094−4101.
48. Georgescu, R. E., Braswell, E. H., Zhu, D., and Tasayco, M. L. (1999) Energetics of assembling an artificial heterodimer with an α/β motif: Cleaved versus uncleaved *Escherichia coli* thioredoxin, *Biochemistry 38*, 13355−13366.
49. Vallée, B., Teyssier, C., Maget-Dana, R., Ramstein, J., Bureaud, N., and Schoentgen, F. (1999) Stability and physicochemical properties of the bovine brain phosphatidylethanolamine-binding protein, *Eur. J. Biochem. 266*, 40−52.
50. Haq, S. K., Ahmad, M. F., and Khan, R. H. (2003) The acid-induced state of glucose oxidase exists as a compact folded intermediate, *Biochem. Biophys. Res. Commun. 303*, 685−692.
51. Ramachandran, S., and Udgaonkar, J. B. (1996) Stabilization of barstar by chemical modification of the buried cysteines, *Biochemistry 35*, 8776−8785.
52. Svensson, A.-K. E., O'Neill, J. C., Jr., and Matthews, C. R. (2003) The coordination of the isomerization of a conserved non-prolyl cis peptide bond with the rate-limiting steps in the folding of dihydrofolate reductase, *J. Mol. Biol. 326*, 569−583.
53. Vuilleumier, S., Sancho, J., Loewenthal, R., and Fersht, A. R. (1993) Circular dichroism studies of barnase and its mutants: Characterization of the contribution of aromatic side chains, *Biochemistry 32*, 10303−10313.
54. Bolgiano, B., Crane, D. T., Xing, D., Williams, L., Jones, C., and Corbel, M. J. (1999) Physico-chemical analysis of *Bordetella pertussis* antigens, *Biologicals 27*, 155−162.
55. Freskgård, P.-O., Mårtensson, L.-G., Jonasson, P., Jonsson, B.-H., and Carlsson, U. (1994) Assignment of the contribution of the tryptophan residues to the circular dichroism spectrum of human carbonic anhydrase II, *Biochemistry 33*, 14281−14288.
56. Khan, F., Khan, R. H., and Muzammil, S. (2000) Alcohol-induced versus anion-induced states of α-chymotrypsinogen A at low pH, *Biochim. Biophys. Acta 1481*, 229−236.
57. Andersson, D., Carlsson, U., and Freskgård, P.-O. (2001) Contribution of tryptophan residues to the CD spectrum of the extracellular domain of human tissue factor, *Eur. J. Biochem. 268*, 1118−1128.
58. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers 22*, 2577−2637.
59. Gursky, O., and Atkinson, D. (1996) Thermal unfolding of human high-density apolipoprotein A-1: Implications for a lipid-free molten globular state, *Proc. Natl. Acad. Sci. U.S.A. 93*, 2991−2995.
60. Andrew, C. D., Bhattacharjee, S., Kokkoni, N., Hirst, J. D., Jones, G. R., and Doig, A. J. (2002) Stabilising interactions between aromatic and basic side chains in α-helical peptides, *J. Am. Chem. Soc. 124*, 12706−12714.
61. Borges, I., Jr., Varandas, A. J. C., Rocha, A. B., and Bielschowsky, C. E. (2003) Forbidden transitions in benzene, *THEOCHEM 621*, 99−105.
62. Serrano-Andrés, L., and Roos, B. O. (1996) Theoretical study of the absorption and emission spectra of indole in the gas phase and in a solvent, *J. Am. Chem. Soc. 118*, 185−195.
63. Albinsson, B., and Nordén, B. (1992) Excited-state properties of the indole chromophore. Electronic transistion moment directions from linear dichroism measurements: Effect of methyl and methoxy substituents, *J. Phys. Chem. 96*, 6204−6212.
64. Condon, E. U., Altar, W., and Eyring, H. (1937) One-electron rotatory power, *J. Chem. Phys. 5*, 753−775.
65. Rosenfeld, L. (1928) Quantenmechanische theorie der naturlichen optischen aktivitat von flussigkeiten und gasen, *Z. Phys. 52*, 161−174.

BI049031N