

# Comment on "Improving protein circular dichroism calculations in the far-ultraviolet through reparametrizing the amide chromophore" [J. Chem. Phys. 109, 782 (1998)]

Robert W. Woody<sup>a)</sup> and Narasimha Sreerama

Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, Colorado 80523-1870

(Received 3 March 1999; accepted 13 May 1999)

[S0021-9606(99)00430-4]

Hirst<sup>1</sup> has recently reported calculations of the circular dichroism (CD) spectra of a set of 23 proteins using parameters derived from both semiempirical and ab initio molecular orbital (MO) calculations on model amides. The semiempirical parameters were derived from complete neglect of differential overlap/spectroscopic (CNDO/S) wave functions<sup>2</sup> for acetamide<sup>3</sup> and for *N*-methylacetamide. The ab initio parameters were obtained from multireference configuration interaction (MRCI) calculations<sup>4</sup> on *N*-methylacetamide. The results of these calculations were compared to experimental data<sup>5</sup> for the 23 proteins at 220 and 190 nm, the approximate wavelengths of the  $n\pi^*$  and  $\pi\pi^*$  amide transitions, respectively. The results for both semiempirical parameter sets were disappointing, showing only weak correlation at 220 nm. The Spearman rank correlation coefficients,<sup>6</sup>  $r_{sp}$ , were 0.41 and 0.32 for the acetamide- and *N*-methylacetamide-based calculations, respectively, neglecting side-chain contributions. Results including side chains were poorer. The MRCI parameters led to some improvement, giving  $r_{sp}=0.62$  neglecting side chains (0.25 with the inclusion of side chains). None of the three parameter sets gave a significant correlation at 190 nm.

We wish to report semiempirical calculations for the same protein set that give much better results at 220 and 190 nm than those reported by Hirst,<sup>1</sup> including the ab initio parameters. Three peptide transitions were considered,  $n\pi^*$  at 220,  $\pi\pi^*$  ( $NV_1$ ) at 190,  $\pi\pi^*$  ( $NV_2$ ) at 140 nm. For the  $\pi\pi^*$  transitions, Clark's transition moment directions<sup>7</sup> for the secondary amide of *N*-acetylglycine ( $NV_1$ ,  $-55^\circ$ ;  $NV_2$ ,  $+61^\circ$ ) were used. The  $NV_1$  and  $NV_2$  transitions were not allowed to mix, i.e., the transition connecting these excited states was neglected. Transition monopole charges for the  $n\pi^*$  transition and for transitions connecting the  $\pi\pi^*$  and  $n\pi^*$  excited states were calculated from intermediate neglect of differential overlap/spectroscopic (INDO/S) wave functions<sup>8</sup> for NMA. The ground-state monopole charges were those used by Woody.<sup>9</sup> Side-chain transitions of Phe, Tyr, and Trp were included in the calculations, as described previously.<sup>10</sup> The calculations used the matrix method<sup>11</sup> in its origin-independent form,<sup>12</sup> as did those of Hirst.<sup>1</sup> The monopole charges were located at atomic centers for aromatic side chains and at the positions given by the semiempirical  $Z$  values described by Woody<sup>9</sup> for amide transitions.

Gaussian band shapes were assigned for each transition

in the protein. The bandwidths were assigned as described in the supplemental material. The two most important bandwidths were identical (10.5 nm for the amide  $n\pi^*$  transition) or nearly identical (11.3 nm) to the values used by Hirst<sup>1</sup> for all transitions.

The Spearman rank correlation coefficient for the comparison of experimental vs calculated  $\Delta\epsilon$  at 220 nm was 0.84, well above the best value reported by Hirst<sup>1</sup> (0.62, for ab initio parameters). The correlation coefficient from a linear regression analysis,  $r_{lin}$ , was 0.93, and the slope of the regression line was 0.75, higher than the best cause of ab initio parameters in Hirst's results, estimated to be  $\sim 0.4$ .

Comparing experimental vs calculated  $\Delta\epsilon$  at 190 nm, the correlation coefficients are lower ( $r_{sp}=0.66$ ,  $r_{lin}=0.87$ ) than those for the 220 nm data, as is the slope of the regression line (0.47), but the results show a good correlation at 190 nm where none was found by Hirst.<sup>1</sup> Hirst<sup>1</sup> obtained poorer correlations when the aromatic side chains were included in the calculation. We also find that inclusion of side-chain transitions leads to somewhat poorer results, but the correlations were still reasonably good ( $r_{sp}=0.71$ ,  $r_{lin}=0.87$  at 220 nm;  $r_{sp}=0.46$ ,  $r_{lin}=0.72$  at 190 nm).

Figure 1 shows a comparison of the calculated and experimental CD spectra for the three proteins for which Hirst<sup>1</sup>

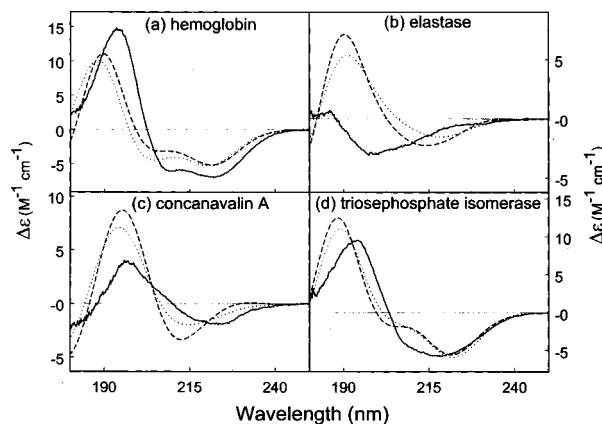


FIG. 1. Comparison of calculated and experimental far-UV CD spectra of four proteins: (a) hemoglobin; (b) elastase; (c) concanavalin A; (d) triosephosphate isomerase. The solid curve is the experimental spectrum, the dashed curve is the calculated spectrum, including aromatic side-chain contributions, and the dotted curve is the calculated spectrum neglecting aromatic side-chain contributions.

showed spectra: hemoglobin, elastase, and concanavalin A. Two calculated curves are shown for each protein, with and without aromatic side-chain contributions. For hemoglobin, the main difference from Hirst's results is that we predict a distinct negative shoulder or maximum near 205 nm, in qualitative agreement with experiment. For elastase, neither set of calculations reproduces the negative band observed near 200 nm, but our results give a somewhat weaker 190 nm positive band than those predicted in Hirst's calculations, which leads to a smaller discrepancy with experiment near 190 nm. With concanavalin A, our results do not agree in the 220 nm region as well as Hirst's *ab initio* parameters, but do not overestimate the 190 nm maximum as strongly. Figure 1 also includes the spectra of triose phosphate isomerase (TIM), which was described as an "outlier" by Hirst.<sup>1</sup> In our calculations, TIM is actually one of the better-behaved proteins, probably because of its high helix content.

The improved results reported here are primarily ascribable to two factors: the use of an experimental<sup>7</sup> transition moment direction for the  $NV_1$  transition and the choice of monopole positions. The transition moment directions used in the two semiempirical calculations of Hirst<sup>1</sup> were  $-17^\circ$  for acetamide<sup>3</sup> and  $-23^\circ$  for NMA.<sup>1</sup> The direction for NMA from *ab initio* calculations<sup>4,13</sup> is  $-36$  to  $-45^\circ$ . Calculations of the exciton splitting for the  $NV_1$  band in an  $\alpha$ -helix (unpublished results) show that the splitting increases as the  $NV_1$  transition moment direction becomes more negative. This is the main reason why our calculations for proteins with high helix content, using the angle of  $-55^\circ$ , show at least a shoulder near 205 nm, whereas those of Hirst<sup>1</sup> and Kurapkat *et al.*<sup>3</sup> do not.

The monopole positions are very important. We have used the positions derived by Woody<sup>9</sup> for semiempirical  $Z$  values. For the  $n\pi^*$  transition, this places quadrupolar arrays of charges around the O, C, and N atoms, the largest and most important of which are those at the O atom. Hirst<sup>1</sup> represented the  $n\pi^*$  quadrupole moment from the *ab initio* calculation by four charges of  $\pm 100$  au clustered very tightly near the center of mass of NMA, i.e., effectively a point quadrupole. Hirst's calculations with semiempirical parameters used monopoles positioned by the method of Kurapkat *et al.*<sup>3</sup> These monopoles are distributed over the molecule, but they are closer to the nuclei than those used by Woody.<sup>9</sup> Other differences between the present calculations and those of Hirst<sup>1</sup> are likely to be of minor importance.

Both sets of calculations give rather poor results for

$\beta$ -sheet-rich proteins in the 190 nm region, predicting much larger amplitudes in this region compared with those experimentally observed. We believe that this is largely due to the neglect of the effect of the local electrostatic environment on the transition energies and the neglect of through-bond coupling of the peptide transition. We are extending our calculations to take these effects into account.

In summary, semiempirical wave functions in conjunction with suitable representations of transition and static charge distributions can provide the basis for calculations of protein CD spectra that reproduce the major features of these spectra. We do not question the desirability of utilizing *ab initio*-derived parameters, but to provide real improvements over the semiempirical methods, more realistic representations of transition and static charge densities must be used. The complete set of theoretical and experimental CD curves, and the parameters used in the calculation, are available as supplemental material (EPAPS)<sup>14</sup> or at our website: <http://lamar.colostate.edu/~sreeram/JCP1999>.

This work was supported by USPHS Grant No. GM22994.

<sup>a</sup>Electronic mail: [rww@lamar.colostate.edu](mailto:rww@lamar.colostate.edu)

<sup>1</sup>J. D. Hirst, *J. Chem. Phys.* **109**, 782 (1998).

<sup>2</sup>J. DelBene and H. H. Jaffé, *J. Chem. Phys.* **48**, 1807 (1968).

<sup>3</sup>G. Kurapkat, P. Krüger, A. Wollmer, J. Fleischhauer, B. Kramer, E. Zobel, A. Koslowski, H. Botterweck, and R. W. Woody, *Biopolymers* **41**, 267 (1997).

<sup>4</sup>J. D. Hirst, D. M. Hirst, and C. L. Brooks, III, *J. Phys. Chem. A* **101**, 4821 (1997).

<sup>5</sup>P. Pancoska, E. Bitto, V. Janota, M. Urbanova, V. P. Gupta, and T. A. Keiderling, *Protein Sci.* **4**, 1384 (1995).

<sup>6</sup>W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes* (Cambridge University Press, Cambridge, 1989), pp. 489–494.

<sup>7</sup>L. B. Clark, *J. Am. Chem. Soc.* **117**, 7974 (1995).

<sup>8</sup>J. Ridley and M. C. Zerner, *Theor. Chim. Acta* **32**, 111 (1973).

<sup>9</sup>R. W. Woody, *J. Chem. Phys.* **49**, 4797 (1968).

<sup>10</sup>I. B. Grishina and R. W. Woody, *Faraday Discuss.* **99**, 245 (1994).

<sup>11</sup>P. M. Bayley, E. B. Nielsen, and J. A. Schellman, *J. Phys. Chem.* **73**, 228 (1969).

<sup>12</sup>W. J. Goux and T. M. Hooker, Jr., *J. Am. Chem. Soc.* **102**, 7080 (1980).

<sup>13</sup>L. Serrano-Andrés and M. P. Fülcher, *J. Am. Chem. Soc.* **118**, 12190 (1996).

<sup>14</sup>See AIP Document No. EPAPS: EJCPSA6-111-004930 for the complete set of theoretical and experimental CD curves, and the parameters used in the calculation. EPAPS document files may be retrieved free of charge from AIP's FTP server (<http://www.aip.org/pubservs/paps.html>) or from [ftp.aip.org](ftp://ftp.aip.org) in the directory /epaps/. For further information, e-mail: [paps@aip.org](mailto:paps@aip.org) or fax: 516-576-2223.