
Access from the University of Nottingham repository:
http://eprints.nottingham.ac.uk/48783/8/1-s2.0-S0022283617305892-main.pdf

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the Creative Commons Attribution licence and may be reused according to the conditions of the licence. For more details see:
http://creativecommons.org/licenses/by/2.5/

A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk
DichroCalc: Improvements in Computing Protein Circular Dichroism Spectroscopy in the Near-Ultraviolet

Sarah B. Jasim, Zhuo Li, Ellen E. Guest and Jonathan D. Hirst

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

Correspondence to Jonathan D. Hirst: jonathan.hirst@nottingham.ac.uk

https://doi.org/10.1016/j.jmb.2017.12.009

Edited by Michael Sternberg

Abstract

A fully quantitative theory connecting protein conformation and optical spectroscopy would facilitate deeper insights into biophysical and simulation studies of protein dynamics and folding. The web server DichroCalc (http://comp.chem.nottingham.ac.uk/dichrocalc) allows one to compute from first principles the electronic circular dichroism spectrum of a (modeled or experimental) protein structure or ensemble of structures. The regular, repeating, chiral nature of secondary structure elements leads to intense bands in the far-ultraviolet (UV). The near-UV bands are much weaker and have been challenging to compute theoretically. We report some advances in the accuracy of calculations in the near-UV, realized through the consideration of the vibrational structure of the electronic transitions of aromatic side chains. The improvements have been assessed over a set of diverse proteins. We illustrate them using bovine pancreatic trypsin inhibitor and present a new, detailed analysis of the interactions which are most important in determining the near-UV circular dichroism spectrum.

© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

Circular dichroism (CD) in the near-ultraviolet (near-UV) is widely used to study structural changes in proteins, because of its sensitivity [1,2] to, for example, ligand binding, changes of the environment and catalysis [3]. The hydrophobic packing of aromatic side chains of amino acids is a key determinant of structure of proteins [4]. Aromatic side chains are often found in enzyme active sites, where they play an important role in molecular recognition and biological function. Thus, biospectroscopic investigations of aromatic–aromatic interactions can provide new information on protein structure and dynamics. The local conformation of the protein chain around an aromatic ring influences the ππ* electronic excitations of phenylalanine, tyrosine, and tryptophan, and the coupling of the various excitations is sensitive to the relative orientation and proximity of the aromatic groups.

CD is the differential absorption of left-handed and right-handed circularly polarized light. Bands in CD spectra in the near-UV occur between 240 and 300 nm. They provide useful information on the environment and interactions of aromatic side chains and the measurement of near-UV CD of proteins has been instrumental, for example, in characterizing intermediates in the folding process [5]. A combination of CD measurement and computational approaches (if sufficiently accurate) can offer a powerful means of investigating the conformations of large biopolymers.

Methods for computing the CD of proteins from first principles are well established. The background and methodology has been reviewed elsewhere [6]. These techniques are necessarily more approximate than fully quantum mechanical methods, such as time-dependent density functional theory, which are only tractable at the moment on systems of the size of small peptides [7]. A central component of approximate approaches for computing protein CD is the set of parameters used to model the electronic excitations of the various chromophores in proteins. Over the years, we have used ab initio quantum chemical methods to develop parameter sets that have improved the accuracy of protein CD
Calculations, in the vacuum-UV [8], where bands arise due to charge-transfer transitions, in the far-UV [9], where the backbone nπ* and ππ* transitions appear, and in the near-UV [10,11], where the main contribution is from the aromatic Lα and Lπ ππ* transitions. These parameters and the computational methodology are available via our DichroCalc web interface [12] and have been used by over 1400 different researchers to compute the CD spectra of more than 3.3 million structures, including all the proteins in the Protein Data Bank (PDB) [13], models and snapshots from molecular dynamics (MD) simulations. Software implementing analogous methodology to that in DichroCalc has recently been made available, for example, ProtPol [14], but the DichroCalc web interface makes the computational methodology, including some unique theoretical aspects, accessible to a wider community.

The focus of this paper is on new developments. First, however, we give a few illustrative examples of some of the applications of DichroCalc. In contrast to the analysis of protein CD spectra to extract information (perhaps on secondary structure content) through the application of statistical or machine learning methods, which is an active field in its own right [15,16], the focus of DichroCalc and its applications usually start with the generation of atomistic models (or hypotheses) of structure and/or conformational change, followed by computation, from first principles, of the CD spectra that would be predicted to arise. DichroCalc has been used [17] to compute the CD spectra of amyloid structures, based on MD simulations. Calculated CD spectra have also been used in the comparison of simulated conformations of Mets7, bound and unbound to cis-platin [18]; Mets7 is an eight-residue peptide, which mimics an important function, that is, cellular uptake of platinum-based drugs, of the high-affinity copper influx transporter protein, Ctr1. Simulations of a phosphate binding protein labeled with two rhodamine fluorophores have been conducted [19], because the protein was not amenable to study by X-ray crystallography or NMR spectroscopy; comparison with experimental biophysical data was facilitated by computing the CD from the MD simulations. A similar approach has been adopted to help design mutants of the coiled-coil domain of the oncoprotein Bcl–Abl [20]; vacuum-UV CD spectrum of human myelin P2 protein has been computed [21] using DichroCalc and qualitatively agreed with experimental measurements made using synchrotron radiation. Simulations and DichroCalc have been used to predict the loss of helicity for peptides binding to the surface of silica [22]. As is evident from the preceding selected examples, DichroCalc is finding utility across a range of challenging problems.

There have been many biochemical and biophysical studies utilizing near-UV CD spectroscopy. The contribution of aromatic residues to the near-UV spectra has been investigated using site-directed mutagenesis [23,24]. Recently, there has been additional interest in the near-UV spectroscopy of proteins, with the advent of two-dimensional electronic spectroscopy, realizable through ultrafast laser spectroscopy, which has been used, for example, to study tryptophan-to-heme electron transfer and excitation energy transfer in myoglobin [25]. Several theoretical studies [26–28] have investigated the distinctive signatures that amyloid fibrils show in the near-UV as well as those of other proteins [29]. MD simulation and calculations of two-dimensional near-UV spectroscopy have been used to monitor the unfolding of a tetrapeptide after photo-induced reduction of an internal disulfide bond [30].

In contrast to intense bands in the far-UV, near-UV bands are much weaker and have been challenging to compute theoretically. We have recently [11] made some advances in the accuracy of calculations in the near-UV, realized through the consideration of the vibrational structure of the electronic transitions of aromatic side chains and significant improvements have been established over a set of diverse proteins. The conformational influences of protein structure on near-UV CD spectra are rooted in the same fundamental physical origins as those on the far-UV CD spectra, but the structural relationship is most likely rather more subtle than in the far-UV, where the regular, repeating nature of the backbone chain in well-defined secondary structures dominates. Herein, we investigate these origins, using our new parameters for the aromatic chromophores.

The intensity or rotational strength of a transition in a CD spectrum may arise from three mechanisms coupling electronic transitions. The one-electron mechanism [31] involves two transitions within the same chromophore and is also known as intra-chromophore mixing. The υ–υ mechanism is the coupling mediated by electric transition dipole moments of two separate chromophores. The υ–m mechanism is the coupling between the electric and magnetic transition moments of two separate chromophores [32]. Several studies have investigated the role of these mechanisms in determining the nature of the near-UV CD spectra of the proteins TEM-1 β-lactamase [33–37] and human carbonic anhydrase [38].

We have recently updated and enhanced the DichroCalc web server with new models of the aromatic side-chain chromophores, which incorporate the vibrational structure within the electronic bands in the near-UV. The near-UV CD spectra of 40 proteins computed with the new parameter set agree well with experiment and much better than spectra calculated with previous parameters [10,39]. Accounting for conformational flexibility by using families of NMR structures, instead of the X-ray structure, improved the calculated spectrum for several proteins. This...
suggested an informative level of sensitivity to side-chain conformation and provides a means for us to examine the detailed influences of structure on the spectrum. In what follows, we investigate the contributions to the near-UV CD spectra computed for a family of NMR structures for bovine pancreatic trypsin inhibitor (BPTI), a protein comprising 58 residues, including four phenylalanines (Phe4, Phe22, Phe33, Phe45), four tyrosines (Tyr10, Tyr21, Tyr23, Tyr35) and three disulfide bonds.

Results and Discussion

The quality of the calculated near-UV CD spectra has been systematically assessed [11] over a set of 40 proteins, for which there are published experimental spectra. The correlation between the computed and the experimental intensity from 270 to 290 nm is much improved; the root mean squared error (RMSE) in the computed intensity in this wavelength range is reduced by typically a factor of two compared to the previous best calculations. The calculations do not yet give a fully quantitative agreement with experiment. However, in addition to the significant improvement in statistical measures such as RMSE, mean relative error and correlation, there is a striking qualitative improvement which arises from reproduction of the vibrational features clearly evident in the near-UV spectra.

The near-UV CD spectra of BPTI calculated using the X-ray crystal structure form II (PDB code: 5PTI) are compared to the experimental spectrum [40] in Fig. 1. The X-ray crystal structure leads to a computed spectrum that reproduces the overall shape and location of the peaks reasonably well, but the intensity across the wavelength range is noticeably underestimated. The spectrum calculated by averaging over the NMR structures is about 20% more intense and the RMSE agreement with experiment is reasonably well, but the intensity across the wavelength range is noticeably underestimated. The spectrum calculated by averaging over the NMR structures is about 20% more intense and the RMSE with respect to the experimental spectrum is correspondingly reduced (from 128 to 91 deg cm² dmol⁻¹). The conformational diversity of BPTI in solution is reflected, at least in part, by the family of NMR structures. Figure 1 shows the CD spectrum calculated using specific NMR structures: frame 3 and frame 5, which give the more intense spectra (in closest agreement with experiment), and by way of contrast frame 17, which gives the calculated spectrum of weakest intensity. The structural difference across these three frames is small, varying from 0.8 to 1.1 Å all-atom RMSD.

The spectrum computed from averaging over all the NMR frames is less intense than the experimental spectrum. However, two particular NMR frames, 3 and 5, lead to computed spectra in excellent agreement (both with an RMSE of 49 deg cm² dmol⁻¹) with experiment. The contributions of the phenylalanine and tyrosine chromophores are clearly well reproduced. The good agreement and the sensitivity to modest structural differences motivate us to investigate the detailed interactions that govern the spectrum.

The interactions of different transitions in the protein environment are computed in the exciton matrix method calculation as the purely electrostatic, Coulombic interaction energy between two transition charge densities, each represented by a set of point charges (as part of the parameter set, described in detail in our earlier work [11]). They depend on the charge densities and on their relative orientation and separation. In the case of the one-electron mechanism, there is a dependence on the surrounding ground state electronic density. The significant interaction energies (i.e., those greater than 1 cm⁻¹) for frame 3 are shown in Table 1. Coupling between L states and B states (the latter occurring in the far-UV) is weak. The largest interaction energy is 84.9 cm⁻¹ and corresponds to the one-electron mixing of the ¹Lb, v2 and ¹La transitions of Tyr23 (v2 denotes that the transition is to the second vibrational level of the excited state). The analogous transitions on Tyr10 interact via the same mechanism and with a very similar magnitude. Inter-chromophore interactions are much weaker, and with the exception of the interaction between Phe22 and Phe33 (27.4 cm⁻¹), they are all less than 10 cm⁻¹; for example, the interaction between Phe22 ¹Lb,v1 and Phe45 ¹La is −1.4 cm⁻¹. The interaction energies for frame 5 have a very similar pattern (data not shown).
From the interaction energies in Table 1, it is clear that the one-electron mixing of the tyrosine $1^L_b$ states is an important factor in the appearance of the near-UV spectrum. This is confirmed by a detailed examination of the intensity (i.e., rotational strength) of each transition in the calculated spectrum. The most intense transitions in this region arise from the tyrosine residues, and the phenylalanine transitions make a more modest contribution. Table 2 shows for each tyrosine $1^L_b$ transition the contribution of the angle between the electric and magnetic transition dipole moments to the intensity. An angle less than 90° would lead to a positive intensity; an angle greater than 90° gives a negative rotational strength. The angles in Table 2 exhibit some variation and a delicate balance (of positive and negative contributions) across the transitions is responsible for the overall intensity. The predominance of the one-electron mechanism in the near-UV CD of BPTI leads to significant sensitivity of the spectrum to the precise nature of the local environment surrounding the tyrosine residues.

The above exemplar, taken from a wide range of proteins used to benchmark our latest calculations, demonstrates that a significant advance in quantitative modeling of the near-UV CD spectra of proteins has been achieved through more sophisticated modeling of the electronic and vibronic structure of the relevant chromophores. The conformational dynamics, as reflected in the calculations using a family of NMR structures compared to an X-ray crystal structure, and the influence of the local electrostatic environment are other factors that can also be important. The latter aspect, in particular, will be a focus for future development, along with the development of parameters for the disulfide bond.

### Table 1. Calculated interaction energies (shown only if greater than 1 cm$^{-1}$) of the electronic transitions of aromatic side chains in frame 3 of the NMR structure of BPTI (PDB code: 1PIT)

<table>
<thead>
<tr>
<th>Chromophore/Transition</th>
<th>Chromophore/Transition</th>
<th>Separation (Å)</th>
<th>Interaction energy (cm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phe4 $1^L_b$ v1</td>
<td>Phe45 $1^L_b$ v1</td>
<td>9.6</td>
<td>1.3</td>
</tr>
<tr>
<td>Phe22 $1^L_b$ v1</td>
<td>Phe33 $1^L_a$ v1</td>
<td>5.2</td>
<td>27.4</td>
</tr>
<tr>
<td>Phe22 $1^L_b$ v1</td>
<td>Phe45 $1^L_a$ v1</td>
<td>7.8</td>
<td>−1.4</td>
</tr>
<tr>
<td>Phe2 $1^L_a$ v1</td>
<td>Tyr23 $1^L_b$ v2</td>
<td>9.8</td>
<td>−1.2</td>
</tr>
<tr>
<td>Phe33 $1^L_b$ v1</td>
<td>Tyr21 $1^L_b$ v1</td>
<td>8.8</td>
<td>1.9</td>
</tr>
<tr>
<td>Phe33 $1^L_a$ v1</td>
<td>Tyr35 $1^L_b$ v2</td>
<td>8.9</td>
<td>−5.4</td>
</tr>
<tr>
<td>Phe45 $1^L_b$ v1</td>
<td>Tyr21 $1^L_a$ v1</td>
<td>8.6</td>
<td>3.5</td>
</tr>
<tr>
<td>Tyr10 $1^L_b$ v2</td>
<td>Tyr10 $1^L_a$ v2</td>
<td>0.0$^b$</td>
<td>80.9</td>
</tr>
<tr>
<td>Tyr10 $1^L_a$ v1</td>
<td>Tyr35 $1^L_b$ v2</td>
<td>8.9</td>
<td>2.8</td>
</tr>
<tr>
<td>Tyr21 $1^L_b$ v2</td>
<td>Tyr35 $1^L_a$ v1</td>
<td>13.5</td>
<td>1.6</td>
</tr>
<tr>
<td>Tyr23 $1^L_a$ v2</td>
<td>Tyr23 $1^L_b$ v2</td>
<td>0.0</td>
<td>84.9</td>
</tr>
<tr>
<td>Tyr23 $1^L_a$ v1</td>
<td>Tyr35 $1^L_b$ v2</td>
<td>19.8</td>
<td>1.6</td>
</tr>
<tr>
<td>Tyr35 $1^L_b$ v2</td>
<td>Tyr35 $1^L_a$ v1</td>
<td>0.0</td>
<td>6.7</td>
</tr>
</tbody>
</table>

$^a$ Vibrational levels of $1^L_b$ states are denoted v1 or v2; only these levels showed significant interactions.

$^b$ A separation of 0 Å corresponds to the one-electron mixing of two transitions on the same chromophore.

### Table 2. The eight ($\nu = 1$ to 8) tyrosine $1^L_b$ vibrational states, their wavelengths ($\lambda$) and their calculated contributions to the rotational strength (in Debye Bohr magneton) at that wavelength ($R_0$) in the near-UV CD spectrum of BPTI (computed for frame 3 of the NMR structure)

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>$\lambda$ (nm)</th>
<th>Tyr35$^a$ $\mu_m$$^b$</th>
<th>$\cos\theta$$^b$</th>
<th>Tyr21 $\mu_m$</th>
<th>$\cos\theta$</th>
<th>Tyr10 $\mu_m$</th>
<th>$\cos\theta$</th>
<th>Tyr23 $\mu_m$</th>
<th>$\cos\theta$</th>
<th>$R_0$/DBM $\times 1000$$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>286.1</td>
<td>0.52 0.03</td>
<td>0.31 −0.06</td>
<td>0.45 −0.05</td>
<td>0.39 −0.01</td>
<td>−0.01</td>
<td>−29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>279.7</td>
<td>0.64 0.02</td>
<td>0.39 −0.05</td>
<td>0.59 −0.05</td>
<td>0.51 −0.01</td>
<td>−0.01</td>
<td>−34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>276.1</td>
<td>0.36 0.02</td>
<td>0.01 −0.07</td>
<td>0.02 −0.06</td>
<td>0.01 −0.09</td>
<td>−0.09</td>
<td>−8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>273.5</td>
<td>0.36 0.02</td>
<td>0.21 −0.07</td>
<td>0.33 −0.05</td>
<td>0.29 −0.01</td>
<td>−0.01</td>
<td>−25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>270.2</td>
<td>0.05 0.02</td>
<td>0.01 −0.42</td>
<td>0.03 −0.07</td>
<td>0.02 −0.07</td>
<td>−0.07</td>
<td>−9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>267.6</td>
<td>0.13 0.01</td>
<td>0.06 −0.16</td>
<td>0.11 −0.05</td>
<td>0.09 −0.02</td>
<td>−0.02</td>
<td>−16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>264.5</td>
<td>0.04 0.00</td>
<td>0.01 −0.37</td>
<td>0.01 −0.09</td>
<td>0.01 −0.19</td>
<td>−0.19</td>
<td>−7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>262.0</td>
<td>0.04 −0.01</td>
<td>0.01 −0.36</td>
<td>0.01 −0.09</td>
<td>0.01 −0.19</td>
<td>−0.19</td>
<td>−7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ BPTI has four tyrosines; thus, there are four contributions at each wavelength.

$^b$ The $\mu_m$ value given in the table is not the scalar product, but simply the product of the magnitudes of the electric and magnetic transition dipole moments.

$^c$ In an achiral environment, the electric and magnetic transition dipole moments would be perpendicular ($\cos\theta = 0$).

$^d$ The rotational strength, $R = \mu_m\cos\theta$, where $\mu$ is the magnitude of the electric transition dipole moment, $m$ is the magnitude of the magnetic transition dipole moment and $\theta$ is the angle between them.
(which has not been considered in the current work). The more quantitative nature of the near-UV CD calculations should help molecular simulation studies of protein structure and dynamics make more direct connections with both equilibrium and time-resolved experiments, and potentially could provide additional useful information for NMR structure determination. There are, of course, many other molecules that form interesting chiral macromolecular structures and that contain aromatic chromophores; a couple of examples which we have studied previously include oligoureas [41] and naphthalenediimide nanotubes [42]. Thus, calculations of near-UV CD spectra, using our web server DichroCalc, will find applicability across many types of macromolecular system.

Materials and Methods

We took the coordinates of BPTI from the X-ray crystal structure (PDB code: 5PTI) and from the NMR models (PDB code: 1PIT). The theoretical framework, an exciton approach [43] or the so-called matrix method [44,45], embodied in DichroCalc has been described fully elsewhere [6,9]. Recently, we have described the extension of the approach to include vibrational structure in the far-UV [46]. Vibrational structure is an important, but not the sole, contributor to the bandwidth of transitions. Extension of the approach is exactly analogous for the near-UV [11], where we have developed new parameters for the phenylalanine (Phe), tyrosine (Tyr) and tryptophan (Trp) side chains, each of which has four electronic states (Lb, Lα, Bb, Bα) as labeled by Platt [47]. For near-UV CD spectra, the main focus of interest is the Lb and Lα transitions; the higher-energy Bb and Bα transitions occur in the far-UV [48]. For phenylalanine and tyrosine, the vibrational structure of only the 1Lb excited states was considered, since their 1Lα states are located at higher energies. Toluene is used to represent the aromatic side chain of phenylalanine and p-cresol is used to represent tyrosine. The vibrational transitions are included by extending the exciton Hamiltonian [11] and modifying by scaling (by a normalized Franck–Condon overlap integral) the electric transition dipole moments and the monopole charges representing the transition charge densities. Interactions between the vibrational levels of the same chromophore are not considered. We include the first six to nine transitions in a vibrational progression. The new parameters can be accessed via the DichroCalc web interface. Initially in the form of line spectra, which, to facilitate comparison with experimental spectra, are convolved with Gaussian bands. In this study, for the near-UV region, we have used a Gaussian function with a bandwidth of 4 nm.

Acknowledgments

This work was supported by the Engineering and Physical Sciences Research Council [grant numbers EP/P020232/1, EP/R512059/1]; the Iraqi Ministry of Higher Education and Scientific Research; and a University of Nottingham Vice-Chancellor’s scholarship. We are grateful for access to the University of Nottingham High Performance Computing facility, and to the HPC Midlands Plus Tier-2 Centre. DichroCalc is freely available for use via http://comp.chem.nottingham.ac.uk.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2017.12.009.

Received 25 October 2017; Received in revised form 24 November 2017; Accepted 10 December 2017

Available online xxxx

Keywords:

- aromatic;
- vibronic;
- ab initio;
- electronic excited states;
- quantum

Abbreviations used:

BPTI, bovine pancreatic trypsin inhibitor; CD, circular dichroism; MD, molecular dynamics; PDB, Protein Data Bank.

References


