

Chem

Electronic circular dichroism spectroscopy of proteins

--Manuscript Draft--

Manuscript Number:	CHEMJOURNAL-D-19-00473R2
Full Title:	Electronic circular dichroism spectroscopy of proteins
Article Type:	Review
Keywords:	exciton; quantum; electrostatic potential; electronic excited states; aromatic, structure prediction
Corresponding Author:	Jonathan Hirst University of Nottingham Nottingham, UNITED KINGDOM
First Author:	David Rogers
Order of Authors:	David Rogers
	Sarah Jasim
	Naomi Dyer
	Francois Auvray
	Matthieu Refregiers
	Jonathan Hirst
Abstract:	Electronic circular dichroism (CD) spectroscopy is an important tool for the elucidation of biomolecular structure. This review describes the latest progress and developments in experimental and theoretical studies of proteins using CD spectroscopy, including time-resolved measurements, oriented CD, and state-of-the-art experiments using polarised UV light from high-energy synchrotron radiation. Statistical and machine learning methods for the analysis of experimental spectra are surveyed. Computational methods employed to predict CD spectra from structure include ab initio quantum chemistry techniques, time-dependent density functional theory and exciton theory. We describe recent computations using exciton theory, where we outline the importance of electronic-vibrational coupling and the influence of electrostatics of the protein environment on the electronic transitions in the chromophores responsible for CD signals in the near-UV. Improvements in the accuracy of the computational approaches should allow more quantitative studies applying a combination of experimental data and modelling to a variety of interesting questions.
Manuscript Classifications:	Other
Opposed Reviewers:	
Suggested Reviewers:	Alison Rodger alison.rodger@mq.edu.au Expert on protein circular dichroism
	Robert Woody Robert.Woody@ColoState.edu Expert on protein circular dichroism
	Tim Keiderling tak@uic.edu Expert on protein circular dichroism
	Bonnie Wallace b.wallace@mail.cryst.bbk.ac.uk Expert on protein circular dichroism
	Kathryn Thomasson kathryn.thomasson@und.edu Expert on protein circular dichroism



University of
Nottingham

UK | CHINA | MALAYSIA

School of Chemistry
University Park
Nottingham NG7 2RD
Tel: +44 115 9513478
FAX: +44 115 9513562
Email: jonathan.hirst@nottingham.ac.uk

Ilaria Cianchetta, Ph.D
Scientific Editor Chem
Consulting Editor Matter
Cell Press
50 Hampshire Street, 5th floor
Cambridge, MA 02139
United States of America

12th July, 2019

Dear Dr Cianchetta,

We are delighted that our manuscript “Electronic circular dichroism spectroscopy of proteins” is accepted in principle. We have uploaded a final revised version, as requested.

I trust that all is in order, but please do not hesitate to contact me with any questions. Many thanks for the invitation to submit a paper and for your efforts in this matter.

Best wishes,

A handwritten signature in black ink, appearing to read 'J. Hirst'.

Jonathan D. Hirst BA PhD CChem FRSC
Professor of Computational Chemistry

Revision of Rogers et al. 'Electronic circular dichroism spectroscopy of proteins' submitted to Chem

Reference: CHEMJOURNAL-D-19-00473

Revisions to address comments from Reviewer #1

Comment: "I would recommend a publication subject to slightly more substantial review of the correlation between protein dynamics and CD properties."

Section 5 "Protein dynamics and CD" has been added to page 25 of the revised manuscript. This section briefly discusses CD in the far- and near-UV and its correlation with dynamical changes of protein structure. Reference is made to the coupling mechanisms, described in section 2.1, that contribute to CD. Mention is made of two previously discussed examples that emphasize these points.

Revisions to address comments from Reviewer #2

Comment: "I suggest the authors to try to focus less on the numerical data and more on the physical concepts."

The following superfluous numerical details have been removed from the revised manuscript. Section, page, paragraph and line numbers refer to the initial submission draft. Eight deletions in total.

S2.3. P14, para 1, lines 10 and 11. "For the test set of 14 proteins, the overall normalised root-mean-square deviation (NRMSD) for PDB2CD is 0.10 and for DichroCalc is 0.17."

S3.1. P15, para 2, line 9. "from 62 to 91 % \pm 10 %".

S3.1. P16, para 1, line 3. ", with an angle of approximately 17° to 53°".

S3.3. Figure 5 (P37), Figure 6 (P38) and Table 1 (P41) have all been removed.

S4.2. P25, para 2, lines 6-8. "At 175 nm, comparing computed with experimental spectra, the Spearman rank correlation coefficient between the experimental and computed intensities increases from 0.44 (backbone transitions only and no charge transfer in the calculations) to 0.80 (with charge transfer in the calculations)."

S4.2. P25, para 3, lines 8 and 9. "At 175 nm, the Spearman rank correlation coefficient is 0.12 with local peptide transitions only, 0.73 when charge-transfer transitions are included and 0.79 upon addition of side-chain transitions."

Comment: "I also think that a reorganization of the manuscript into smaller sections and a better use of the figures which, in the present selection, are rather useless, is needed."

The initial sections of the original submission have been divided into subsections as follows.

1 Fundamentals of the phenomenon of electronic circular dichroism (CD).

2 Far-ultraviolet CD.

2.1 Exciton theory and the matrix method: Theory; An example matrix method computation; Mechanisms that contribute to CD.

2.2 Recent experimental studies.

2.2.1 Time-resolved measurements.

2.2.2 Orientated CD.

2.3 Recent computational studies.

2.3.1 Exciton, theoretical and quantum chemical approaches.

2.3.2 Machine learning and statistical approaches.

3 Near-ultraviolet CD.

3.1 Recent experimental studies.

3.1.1 Site-directed mutagenesis and tertiary structure.

3.1.2 Time-resolved measurements.

3.2 Recent computational studies.

3.2.1 Aromatic side chains and vibrational structure.

3.2.2 Influence of electrostatics.

4 Vacuum-ultraviolet CD.

4.1 Recent experimental studies.

4.1.1 Synchrotron radiation CD and time-resolved measurements.

4.1.2 Applications of synchrotron radiation CD.

4.2 Recent computational studies.

4.2.1 Charge transfer transitions.

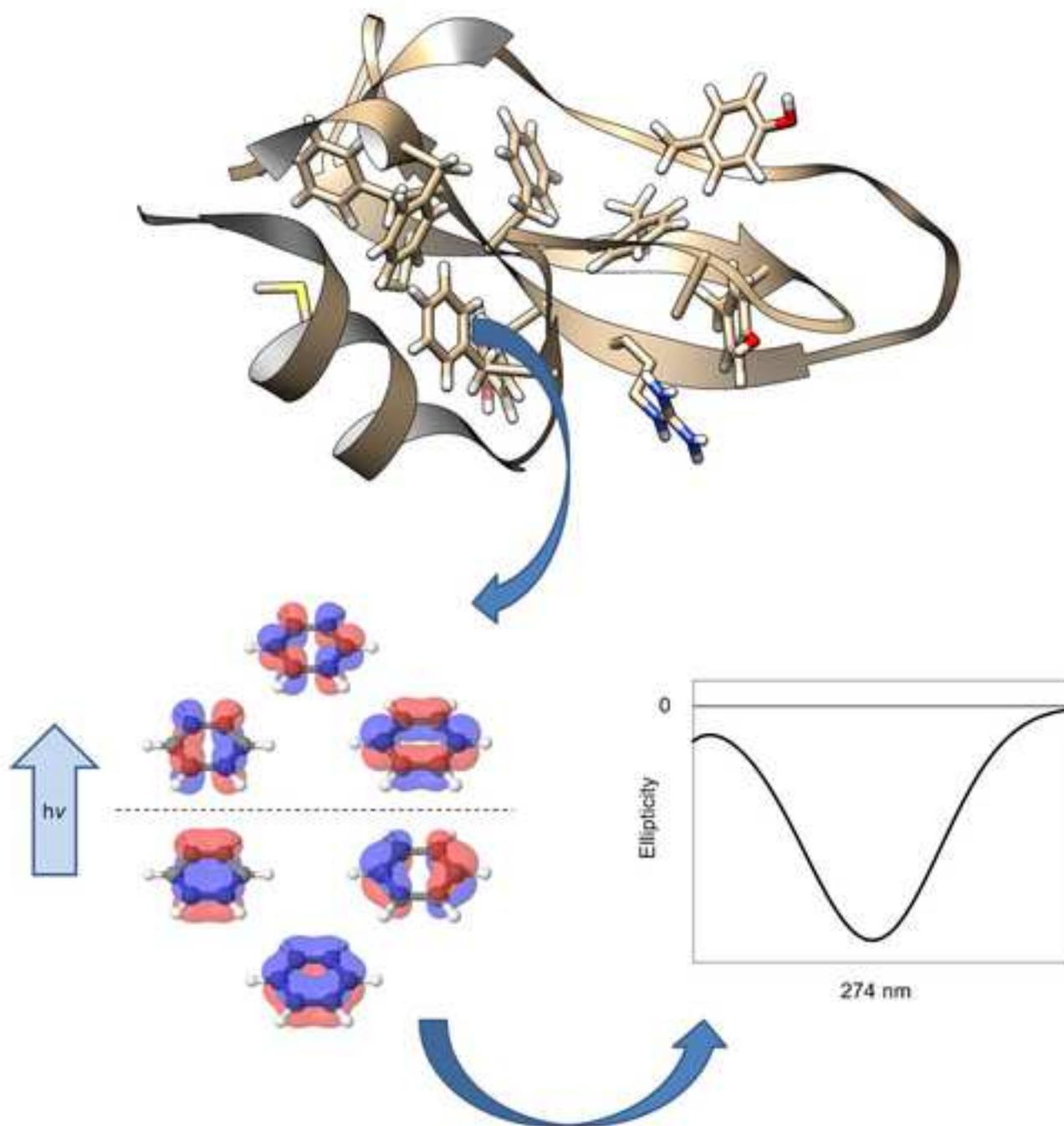
5 Protein dynamics and CD.

6 Summary and outlook.

Figures 5 and 6 have been removed from the revised manuscript. Figure 1 is referenced in sections 2 and 5. Figure 2 is referenced in sections 2.1 and 5.

Comment: “Along the same line, I do not see the point of reporting the data of Table 1: as a matter of fact, the whole section 3.3 should be largely revised so to make it easier to follow.”

Table 1 has been removed. Section 3.3 has been incorporated into section 3.2 (as 3.2.2) and has been revised, where it has been reduced in length (from ~3 pages to ~1.5 pages) and Figures 5 and 6 have been removed. In addition, references [67] Davis et al. and [70] Jmol have been removed.



Electronic circular dichroism spectroscopy of proteins

David M. Rogers¹, Sarah B. Jasim^{1,2}, Naomi T. Dyer¹, François Auvray^{1,3}, Matthieu Réfrégiers³, and Jonathan D. Hirst^{1,*}

¹School of Chemistry, University of Nottingham, University Park, Nottingham, NG7 2RD,
UK

²Chemistry Department, College of Science, Wasit University, Kut, Wasit, Iraq

³Synchrotron SOLEIL, L'Orme des Merisiers, Gif-sur-Yvette 91192, France

*Correspondence jonathan.hirst@nottingham.ac.uk

July 12, 2019

Summary

Electronic circular dichroism (CD) spectroscopy is an important tool for the elucidation of biomolecular structure. This review describes the latest progress and developments in experimental and theoretical studies of proteins using CD spectroscopy, including time-resolved measurements, oriented CD, and state-of-the-art experiments using polarised UV light from high-energy synchrotron radiation. Statistical and machine learning methods for the analysis of experimental spectra are surveyed. Computational methods employed to predict CD spectra from structure include ab initio quantum chemistry techniques, time-dependent density functional theory and exciton theory. We describe recent computations using exciton theory, where we outline the importance of electronic-vibrational coupling and the influence of electrostatics of the protein environment on the electronic transitions in the chromophores responsible for CD signals in the near-UV. Improvements in the accuracy of the computational approaches should allow more quantitative studies applying a combination of experimental data and modelling to a variety of interesting questions.

Keywords exciton, quantum, electrostatic potential, electronic excited states, aromatic, structure prediction.

1 Fundamentals of the phenomenon of electronic circular dichroism (CD)

CD is the differential absorption of left- and right-handed circularly polarised light. An elliptically polarised light wave results when a linearly polarised light wave passes through an optically active chiral compound. The magnitude of the effect is given by

$$\phi = \frac{\pi}{\lambda(\eta_l - \eta_r)} \quad (1)$$

where ϕ is the ellipticity (in radians) of the emerging light wave and η_l and η_r are, respectively, the absorption indices of the left- and right-handed circularly polarised light at wavelength λ . When the ellipticity is plotted as a function of wavelength, a curve results with a maximum corresponding to the wavelength of zero angle in the optical rotatory dispersion (ORD). Optical rotation is rotation of the plane of polarised light to lie along the major axis of the ellipse. Optical isomers give CD curves which are identical except that in one case the effect is positive (i.e. ϕ is positive throughout) whereas for the other isomer the effect is negative. The ellipticity $[\Theta]$ (with units $\text{deg cm}^2 \text{ dmol}^{-1}$) of the emerging light is the ratio of the minor over the major axis of the ellipse. CD is used to characterise optically active molecules that possess a chiral centre, for example the C_α carbon atom in the amino acid alanine. Alanine has two optically active stereoisomers or enantiomers: D-alanine and L-alanine. Prefix D refers to Latin *dexter* and prefix L to Latin *laevus* and they reflect the right- and left-handedness of the chemical structures. The optical activities for L- and D-alanine are, respectively, dextrorotary (rotates the plane of polarised light to the right), prefix (+), and levorotary (rotates the plane to the left), prefix (-). The (*R*)/(*S*) notation for stereoisomers is similar, where prefix (*R*) refers to Latin *rectus* (right, proper, or straight) and prefix (*S*) to Latin *sinister* (left). (*R*)/(*S*) naming follows the Cahn-Ingold-Prelog rules. Optical electronic transitions within a chiral compound give rise to the differential absorption and these electronic transitions are discussed for proteinaceous systems in the sections that follow.

CD in the far-UV is used to characterise protein secondary structure, where the chromophore is the peptidic bond. CD in the near-UV can give insight into a protein's tertiary structure. CD in these wavelength ranges can be used to study temporal changes in local and non-local structure. A protein's primary structure is the sequence of its amino acids and its secondary structure is the spatial arrangement of the amino acids. An α -helix has hydrogen bonds between residues i and $i+4$ in the same strand; a β -sheet has hydrogen bonds between residues in neighbouring strands; and a random coil has an irregular arrangement of amino acids. Secondary structural elements have distinct far-UV CD spectra. The tertiary structure of a protein

is its three-dimensional structure, which helps to give a protein its function.

The application of CD spectroscopy to investigate protein structure followed on from ORD studies of polypeptides and proteins in the early to mid 20th century. The ORD and CD studies were brought about by advances in instrumental techniques and the increasing availability of ORD and CD spectrometers. A modern and practical approach to designing an experimental study of proteins using CD spectroscopy has been outlined by Kelly et al.¹ The authors suggest that after the explosion in structural biology studies in the 1980s - and subsequent deposition of highly-resolved X-ray crystal and structures in the Protein Data Bank (PDB) - CD spectroscopy studies gained momentum due to the technique's ability to study proteins in solution, i.e. closer to their native environment.

2 Far-ultraviolet CD

Far-UV CD is concerned with the spectral region over the wavelength range 160 to 240 nm. This range is of particular interest because singlet electronic transitions in the backbone peptide bond of a protein occur at 190 and 220 nm (the amide chromophore N-C=O). The transition at 220 nm is from the lone pair on oxygen to a π^* anti-bonding orbital ($n\pi^*$ transition) and the transition at 190 nm is from a non-bonding π orbital to the π^* orbital ($\pi_{nb}\pi^*$ transition). These two transitions mix in the chiral environment of a protein and result in distinct spectra for each secondary structural element present within the protein. Exciton splitting of the $\pi_{nb}\pi^*$ transition gives rise to the positive peak at 190 nm and the negative peak at 208 nm in the CD spectrum for an α -helix alongside a negative peak at 220 nm due to the $n\pi^*$ transition. The electronic transitions at 208 nm and 190 nm are polarised parallel and perpendicular, respectively, to the helix axis in α -helical structures. Characteristic CD spectra for other secondary structural motifs are also observed: a β -sheet has a positive peak around 195 nm and a negative peak around 215 nm; a protein with no dominant secondary structure (a random coil) has a negative peak around 200 nm. Figure 1 displays illustrative CD spectra for three proteins which have dominant secondary structures of an α -helix, a β -sheet and a random coil. Characteristic CD spectra for additional secondary structural motifs can also be observed, for example, a type I β -turn.¹

2.1 Exciton theory and the matrix method

Theory

An exciton can be defined as an electron combined with a hole in a crystalline solid. A hole is an unoccupied vacancy in an electronic energy level and behaves as if it were a positive electronic charge with positive mass. In the exciton, the electron has sufficient energy to be in an excited state and is bound to the positive hole by electrostatic attraction. The exciton can migrate through the solid and eventually the electron and hole recombine with emission of a photon. The ‘matrix method’ is a theoretical approach based on exciton theory to calculate the ORD and CD of polypeptides, where the first matrix method computations were carried out by Woody and Tinoco² and by Bayley, Nielsen and Schellman.³ Here the exciton is an electron-hole pair describing an electronic excited state of the backbone peptide bond of the biomolecule.

The matrix method considers a protein to comprise M chromophores with each chromophore possessing a set of characteristic electronic transitions. For a protein the number of backbone peptide-bond chromophores is the number of amino acids minus one. The interaction of the chromophores and their electronic transitions gives a series of delocalised electronic 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 electronic transition. The charge densities can be approximated by sets of point charges that reproduce either the electrostatic potentials arising from *ab initio* densities or by the appropriate electric transition dipole moment associated with each electronic transition.

The matrix method considers the total wave function Ψ_k for a state k of a protein as a linear combination of singly-excited basis functions Φ_{ia}

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

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 singly-excited basis function is a product of M monomer wave functions

$$\Phi_{ia} = \phi_{10} \cdots \phi_{ia} \cdots \phi_{j0} \cdots \phi_{M0} \quad (3)$$

where the first and second subscripts denote, respectively, the chromophore and the electronic state. For example, wave function ϕ_{ia} is a transition from the ground state to excited state a on monomer i . The

Hamiltonian operator \hat{H} for the protein may then be defined as

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

where \hat{H}_i is the Hamiltonian for chromophore i and \hat{V}_{ij} is the operator for the interaction between monomers i and j . The diagonal elements of the Hamiltonian matrix are the excitation energies of the electronically excited states; the off-diagonal elements, V_{ij} , arise from interactions between groups or from interactions between excitations on the same group. An off-diagonal element typically has the form

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

where $\rho_{i0a}(\mathbf{r}_i)$ and $\rho_{j0b}(\mathbf{r}_j)$ are transition electron densities on, respectively, chromophores i and j , ϵ_0 is the permittivity of free space, and r_{ij} is the distance between the two chromophores. The densities in equation 5 may be represented by point charges, thereby replacing the double integral in equation 5 with a sum over the discrete charges as follows

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

where q_s and q_t are, respectively, point charges on monomers i and j with the number of charges in each set given by N_s and N_t . Diagonalisation of the Hamiltonian matrix, equation 4, yields the transition energies (the eigenvalues) of the composite system and the configuration (mixing) coefficients (the eigenvectors) that describe the contributions of the excited states of individual chromophores to the interacting system. For each delocalised transition, the rotational strength may be computed from the following expression,⁴ derived from the Rosenfeld equation,⁵ 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} .

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

Equation 7 is origin-independent only for electric moments in the dipole velocity representation; the current implementation of the matrix method employs the dipole length representation, and following Goux and Hooker,⁶ the origin-independent chiral strength is calculated as

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

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 transition moment vectors of the individual chromophores. The unitary transformation that diagonalised the Hamiltonian is used to transform the linear and angular momentum moment vectors before application of equation 8. The resulting chiral strength is then readily converted to rotational strength.⁶ Finally, a CD spectrum is obtained by centering a band function, such as a Gaussian or a Lorentzian curve, around the rotational strengths calculated for the specific transition energies.

An example matrix method computation

As an example of a matrix method computation, consider the $n\pi^*$ (220 nm) and $\pi_{nb}\pi^*$ (190 nm) electronic transitions in a diamide with dihedral angles $\phi = -48^\circ$ and $\psi = -57^\circ$ (i.e. an α -helical conformation, see Figure 2). The (symmetric) Hamiltonian matrix for the diamide (with just the lower triangle shown for clarity) is expressed as

$$\hat{H} = \begin{pmatrix} E_{n\pi^*}^1 & \cdot & \cdot & \cdot \\ V_{n\pi^*\pi_{nb}\pi^*}^{11} & E_{\pi_{nb}\pi^*}^1 & \cdot & \cdot \\ V_{n\pi^*n\pi^*}^{12} & V_{n\pi^*\pi_{nb}\pi^*}^{21} & E_{n\pi^*}^2 & \cdot \\ V_{n\pi^*\pi_{nb}\pi^*}^{12} & V_{\pi_{nb}\pi^*\pi_{nb}\pi^*}^{12} & V_{n\pi^*\pi_{nb}\pi^*}^{22} & E_{\pi_{nb}\pi^*}^2 \end{pmatrix} \quad (9)$$

where E^1 and E^2 are the transition energies of the monomer excited states; matrix elements V^{12} and V^{21} represent interactions between transitions in different amides; and V^{11} and V^{22} mix excited states within a single amide group. Diagonalisation of the Hamiltonian matrix, equation 9, yields eigenvalues that are the energies of the transitions of the diamide, and eigenvectors that give the contribution of the excited states of individual amides to the delocalised excited states of the diamide. The unitary transformation that diagonalised the Hamiltonian is used to transform the electric and magnetic transition moments of the individual amides. Then equation 8 is employed to compute the chiral strength. Finally, the rotational strength is calculated. The computed rotational strengths for the four transitions in the diamide have Gaussian functions of full-width half-maximum (FWHM) 12.5 nm fitted to generate the predicted CD spectrum shown in Figure 3. The parameters in the matrix method that describe the peptide backbone $n\pi^*$ and $\pi_{nb}\pi^*$ singlet transitions at 220 and 193 nm, respectively, are derived from *ab initio* calculations on *N*-methylacetamide (NMA).⁷

Mechanisms that contribute to CD

The contribution of a chromophore to a CD spectrum results from the rotational strength generated by three mechanisms. i) *The one-electron mechanism* involves two transitions within the same chromophore and is also known as intra-chromophore mixing. ii) *The μ - μ mechanism* is the coupling between electric transition dipole moments of two transitions on different chromophores. iii) *The μ - m mechanism* is the coupling between electric and magnetic transition moments of two different chromophores.⁸ The last two mechanisms are called inter-chromophore mixing or a coupled oscillator type of mechanism.

2.2 Recent experimental studies

2.2.1 Time-resolved measurements

The amplitude of the CD signal is several orders of magnitude lower than the absorption of the achiral background. This difference in intensity makes CD measurements challenging and sensitive. CD measurements require a stable light source and a low noise detector. The latest CD commercial setups usually use deuterium lamps or Xenon flash lamps that provide radiation of sufficient intensity down to ~ 160 nm. Below this value, poor stability and the low intensity of the source preclude reliable CD measurements. The main use of CD spectroscopy in the UV range is related to protein structure determination. However, the low stability and the low intensity of the available light sources make time-resolved measurements complicated in the far-UV range. The following part details the development of the time-resolved CD setups, from the birth of the method to the latest breakthroughs.

The first time-resolved CD techniques appeared in 1974.^{9,10} Researchers recorded, with millisecond temporal resolution, transient CD of photolysis processes in proteins and biomolecular interactions. The UV probe beam was generated by a mercury arc or a Xenon flash lamp, filtered with an interference filter and polarised alternately circularly left and right (50 kHz) by a quartz modulator. The CD signal was determined from the variation of intensity of the probe beam after passing through the sample. These first monochromatic techniques paved the way for time-resolved CD spectroscopy and allowed the first studies of biomolecular dynamics. The technique was improved to reach 300 μ s temporal resolution and coupled to a T-jump to study the relaxation of protein conformation after an increase in temperature.¹¹ The first stopped-flow CD study of protein secondary structure was carried out by Luchins and Beychook.¹² Akiyama and co-workers coupled a commercial CD spectrometer based on the same principle with a mixing microfluidic cell to obtain CD spectra sequentially with 400 μ s time resolution.¹³ A significant enhancement of the time

resolution was made in 1989 using a picosecond pulsed laser.¹⁴ The quartz modulator was replaced by a Pockel’s cell, which circularly polarises the 80 ps pulse of the laser probe. The CD signal was still determined through a phase sensitive detection. This novel approach was used to study the photo-dissociation of the CO ligand in carbonmonoxymyoglobin. With this setup, Xie and Simon¹⁴ made a breakthrough in CD spectroscopy, but its spectral range was limited to 280 nm in the UV range and its use was limited for biological applications.

Hache and co-workers improved this approach, enhancing the alignment of the Pockel’s cell.¹⁵ They also developed a novel technique based on the use of a Babinet–Soleil compensator.¹⁶ This approach significantly enhances the CD measurement, because it removes all the errors arising from polarisation artefacts, and can achieve sub-picosecond temporal resolution. They demonstrated the strength of their method in the deep-UV range (>225 nm) with the measurement of binaphthol spectra and following the photo-dissociation of carbonmonoxymyoglobin.¹⁷

Kliger’s group developed¹⁸ a time-resolved CD setup able to access a spectral range down to 195 nm. It is still today the only setup built with commercial devices that reaches that wavelength. They first proposed a novel method, based on the measurement of the ellipticity change of a highly eccentric elliptically polarised light after passing through a dichroic sample. They also demonstrated the reliability of this new approach measuring the CD change of the photo-dissociation of the CO ligand in carbonmonoxymyoglobin with a 300 ns time resolution. The method has been extended later to higher energy¹⁹ and higher time resolutions (~ 10 ns).²⁰ Oppermann et al.²¹ have developed the first time-resolved femtosecond CD setup that incorporates shot-to-shot broadband detection in the deep-UV region with artifact-free measurement of static and transient CD spectra with a sensitivity below 2×10^{-5} optical density (OD). The setup represents a significant advancement in the field of time-resolved chiral spectroscopy and has the potential to give important new insights into the structural dynamics of biologically relevant chromophores in the UV region.

2.2.2 Orientated CD

Oriented circular dichroism (OCD) is a method that can determine the conformation and the orientation of membrane-active peptides in oriented membrane systems. OCD relies on the phenomenology predicted by Moffitt’s theory,²² that the electric transition dipole moments of the $\pi_{nb}\pi^*$ electronic transition in backbone amide bonds in helical polypeptides are polarised either parallel or perpendicular to the helix axis.²³ This $\pi_{nb}\pi^*$ transition splits into three energy levels by exciton splitting, where each resultant energy level possesses a corresponding dipole moment. One of these transitions is polarised parallel while two are po-

larised perpendicular with respect to the helical axis. A membrane-bound α -helical peptide can be arranged in a surface-bound S-state, a tilted T-state, and an inserted I-state. The transition dipole moment for the parallel-polarised “fingerprint” band at 208 nm interacts with the electric field vector, which oscillates perpendicular to the propagation direction of light. This results in distinct spectra for each of the three states of the membrane-bound α -helical peptide when the incident circularly polarised light is normal to the membrane surface. An S-state, lying flat on a membrane surface, has an intense negative band around 208 nm due to the electric field strongly interacting with the parallel-polarised transition (where the transition dipole is aligned with the electric field). A T-state has a less intense band around 208 nm and for an I-state the band around 208 nm vanishes due to the helix (and transition dipole) being aligned perpendicular to the electric field.²⁴ Bürck et al.²⁵ have given an overview of amphiphilic peptides, hydrophobic transmembrane helical systems and nonhelical peptides investigated using OCD.

2.3 Recent computational studies

The Protein Circular Dichroism Data Bank (PCDDDB),²⁶ established in 2009, operates as an online public repository for archiving CD spectroscopic data and associated bioinformatics and experimental metadata. The resource has been put to various uses, including to help develop new computational methods, some of which are outlined in the following sections.

2.3.1 Exciton, theoretical and quantum chemical approaches

A key early theoretical study of the far- and near-UV CD spectrum of bovine pancreatic trypsin inhibitor (BPTI), by Manning and Woody,²⁷ considered the contribution of aromatic side chains to its CD spectrum. BPTI has an unusual far-UV CD spectrum which has made determination of its secondary structure from its CD spectrum difficult. The protein comprises 58 residues including four Phe and four Tyr residues. Manning and Woody used the matrix method with parameters describing electronic transitions in the peptide backbone (four transitions, $n\pi^*$, $\pi_{nb}\pi^*$, $\pi_b\pi^*$ and $n'\pi^*$) and $\pi\pi^*$ transitions in the aromatic side-chains Phe (four transitions, 1L_b at 260 nm, 1L_a at 208 nm, 1B_b at 181 nm and 1B_a at 180 nm) and Tyr (four transitions, 1L_a at 276 nm, 1L_b at 227 nm, 1B_b at 192 nm and 1B_a at 191 nm). The L and B notation for the aromatic excited states is due to Platt.²⁸ Manning and Woody predicted far-UV CD spectra that were in best agreement with the experimental spectrum when the four peptide backbone transitions and the four Phe and four Tyr transitions were included in the matrix method calculations. They concluded that contributions from aromatic side chains may significantly perturb the far-UV CD spectrum of a protein, and note that in

BPTI this is due to the low amounts of α -helical structure and the clusters of the aromatic amino acids, in particular the Tyr21, Phe22, Tyr23, Phe4, and Phe45 cluster.

A detailed theoretical analysis of empirical protein CD data was performed by Micsonai and co-workers²⁹ to improve the estimation of the secondary structure content for β -sheet and mixed α/β containing proteins from their far-UV CD spectra. This led to the derivation of an algorithm named β -structure selection (BeStSel), which employs the assumption that the CD spectrum of a protein is a linear combination of the basis spectra characteristic of the secondary structural elements present in the protein. The BeStSel algorithm has four characteristic features: (1), the selection of eight secondary structure components for the residues of the known structures of the reference set by a DSSP³⁰ analysis; (2), a reference set of protein spectra with known structures was used to optimise the basis spectra sets; (3), to enable secondary structure determination, basis-spectra sets were calculated on optimised subsets of the reference database by the linear least-square approximation (subsets of reference proteins and wavelength ranges were optimised for each secondary structure separately); (4), each optimised basis set is used to determine the secondary structure of an unknown spectrum. To optimise the basis CD spectra sets (characteristic 2), the authors used a defined reference set containing 73 spectra based on the SP175 reference dataset³¹ augmented with additional examples with spectra of proteins with structural compositions absent or scarce in the SP175 set (native β 2-microglobulin, amyloid fibrils of the K3 fragment of β 2-microglobulin, and Alzheimer’s amyloid- β (1–42) peptide). The BeStSel algorithm reliably distinguished parallel from anti-parallel β -sheets by CD spectroscopy. The authors showed that the twisting of the β -sheets has a strong influence on the CD spectrum. By taking into account the twisting angles between β -strands, the BeStSel algorithm improves secondary structure prediction when compared to previously published algorithms, and specifically for β -structure-rich proteins and amyloid fibrils. Moreover, the increased information content obtained from the CD spectra makes protein fold prediction possible down to the topology level, in terms of the CATH protein structure classification: Class, Architecture, Topology/fold, and Homologous superfamily (CATH).³² BeStSel is available to access online via a user-friendly web interface.³³

Quantum chemical techniques can be applied to the study of CD in the far-UV. Kaminsky and co-workers³⁴ applied time-dependent density functional theory (TDDFT) (and two other methods) to assess the dominant factors that contribute to the CD spectrum of an α -helix. They employed a model oligopeptide Ac-(Ala)_N-NH-Me, where $N = 1, \dots, 18$, to study the dependence on the peptide chain length and the role of the flexibility of the peptide and its solvent environment. The TDDFT calculations showed that a characteristic CD spectrum for an α -helix is discernible for peptides with four or five residues, from a

qualitative agreement with experimental spectra of short α -helices stabilised by metal-histidine binding. A combined molecular mechanics / quantum mechanics (MM/QM) approach indicated that explicit hydration and temperature fluctuations of the peptide geometry had a significant influence on the UV absorption and the CD spectrum. Consequently, a large number of geometries from molecular dynamics (MD) simulations would be needed for adequate structural averaging in a combined MM/QM approach. The authors found that the TDDFT-computed CD results were extremely sensitive to the one-electron basis set and the solvent model employed. This sensitivity was explained in part by the close to perpendicular orientation of the electric and magnetic transition dipole moments in the geometry of the helix for the dipeptide Ac-Ala-NH-Me. If the electric and magnetic transition dipoles are perpendicular (i.e. at 90° to one another) the rotational strength is zero.

The adsorption of polypeptides upon surfaces in solution and their resulting conformational change is an interesting event which CD spectroscopy can monitor. Adsorption is important as it underpins processes in materials science, biotechnology and medicine. Meißner and co-workers³⁵ employed MD simulations to study the conformational change of an α -helical oligopeptide (4DAR5) adsorbing on an anionic amorphous silica (SiO_2) surface. The MD-based simulations employed replica exchange with solute tempering (REST) in combination with metadynamics (metaD), which enabled the prediction of the helicity loss of the oligopeptide upon interaction with silica colloids in water and estimation of the CD spectra of the fully adsorbed and dissolved states. The study used DichroCalc³⁶ to compute the CD spectra for the oligopeptide. There was remarkably good agreement between the computational predictions of the helicity at 222 nm and experimental measurements obtained for the 4DAR5 peptide in solution and adsorbed on SiO_2 colloids. The modelling also showed that the assumption of a linear relationship between peptide (fractional) helicity and CD ellipticity is a correct one, although only for the case of macroscopic conformational states (i.e. an ensemble of biomolecular structures). This implied that the fractional helicity of a peptide in solution, as deduced from its CD spectra, should be interpreted as a mean value of a distribution of structures, each with its own - and differing - structural helicity. Moreover, the authors believe that discrepancies between measured and theoretically computed CD spectra arise from a lack of statistical averaging over the correct ensemble of biomolecular structures and not from limitations in the theoretical formalism (or its computational implementation) underlying CD spectroscopy. This is a view in line with the MM/QM study of Kaminsky and co-workers.³⁴

MD simulations were also employed by Ianeselli and co-workers³⁷ to study the folding of canine milk lysozyme protein, which is a 129 residue globular protein with a folding time on the order of seconds. A scheme to analyse changes in protein tertiary structure was developed whereby the protein configurations

and the optical spectra were drawn directly from atomistic calculations. Firstly, an enhanced MD technique was employed (biased Ratchet-and-Pawl MD (rMD)) to elucidate an ensemble of folding trajectories and, secondly, quantum chemical simulations of the excitonic spectra of long-lived configurations were performed (using the exciton model with a Hamiltonian for the $\pi\pi^*$ transitions in aromatic residues Phe, Tyr and Trp). TDDFT calculations on toluene, *p*-cresol and 3-methyl-indole furnished parameters to describe the four $\pi\pi^*$ transitions for each aromatic residue. The study reproduced all the essential features observed in the experimental time-resolved CD measurements (in the range 200 to 300 nm), and showed that the CD signals due to the excitonic coupling of the aromatic residues can be used to reveal extremely small changes in the tertiary structure of the polypeptide chain. An additional and analogous study³⁷ on the protein colicin immunity binding Im7 further validated the approach. Both examples demonstrate that the combination of experimental time-resolved CD spectra with advanced MD path sampling and quantum electronic structure techniques can provide a powerful approach for the characterisation of protein folding pathways.

A general quantum-mechanical “black-box” approach has been developed by Grimme and co-workers³⁸ to compute UV/Vis and CD spectra for a range of (bio)molecular systems. The sTDA-xTB method³⁹ is a semi-empirical quantum mechanical scheme, designed to compute electronic absorption and CD spectra of molecules with more than 1000 atoms and for nearly all chemical species. The approach uses the tight-binding procedure to simplify the Tamm-Dancoff approximation in the computation of vertical electronic transition energies and properties. The authors used the sTDA-xTB method to compute the CD spectra for a small tri-peptide (MeNH-Ala-Gly-Ala-Ac) in seven structural motifs (from folded-helical to unfolded), and a 20-residue peptide (PDB code 12LY) in an α -helical conformation. The calculated spectra correctly feature the characteristic bands in the CD spectrum for the secondary structures of α helices, β sheets and random coils, although the predicted spectrum for 12LY is red-shifted by 0.5 eV (corresponding to \sim 25 nm). The sTDA-xTB method was also used to calculate the CD spectra for myoglobin in the far-UV and for photoactive yellow protein (PYP) in the near-UV and visible region. Myoglobin has around 2500 atoms and is beyond the scope of standard quantum chemistry techniques. The sTDA-xTB method was used to compute CD for 100 snapshots taken from an MD trajectory, which was necessary to suppress spurious CD transitions on flexible side chains that can interfere with the main transitions of the peptide backbone and which would be present in a single structure approach, whereas they are averaged out by the MD simulation. The computed CD spectrum shows all characteristic features of the α -helix rich myoglobin (eight α -helices) and, when red-shifted by 1 eV (\sim 50 nm), is in very good agreement with the experimental spectrum. PYP is a small cytosolic photoreceptor, in which the chromophore is a deprotonated 4-hydroxy cinnamic acid

derivative covalently bound to the apoenzyme via a thioester bond at Cys69. The sTDA-xTB computed absorption and CD spectrum (after a red shift of 0.5 eV) upon comparison to the experiment are in excellent agreement. The computed vertical excitation energy for the bright $\pi\pi^*$ transition (at 3.27 eV) is in accord with values obtained from higher-level QM methods (TD-DFT and SAC-CI) for fragmented (divide and conquer) model systems.

In another QM/MM and exciton model study, Gattuso and co-workers⁴⁰ studied the behaviour of a model 27 amino acid α -helical peptide in water. CD spectra were computed for 80 MD snapshots, taken at random, using the exciton model with parameters for the $n\pi^*$ and $\pi_{nb}\pi^*$ transitions in the peptide backbone from TDDFT calculations for each of the QM regions considered, where the remaining polypeptide and solvent is described at the MM level. They showed that the choice of the QM region (the number of amino acids included), including different possible hydrogen-bonding patterns, can significantly change the shape of the final CD spectrum. Of the six fragmentation schemes considered, the most appropriate QM scheme to study the model peptide was hybrid sets of four amino acids followed by three couples of amino acids, which describes a complete α -helical turn as represented by a quadruplet followed by three consecutive amino acid couples. (This scheme, presumably, involves nine TDDFT calculations to evaluate the transition energies and properties to construct the Hamiltonian for the exciton model). In the same study, Gattuso and co-workers applied their approach to the conformational change in the 27 amino acid peptide when a retinal-like photo-switch is covalently linked to the two cysteine residues. In the *E* conformation of the photo-switch the peptide is α -helical and when the photo-switch is the *Z* conformer the peptide adopts an α -hairpin secondary structure. The approach correctly predicted the reduction in α -helicity of the model peptide upon *E* to *Z* photoisomerisation of the switch (by comparing CD spectra computed for free and *Z*-derived peptide with experimental spectra for *E*- and *Z*-derived peptide).

2.3.2 Machine learning and statistical approaches

More recent computational approaches to the prediction of protein secondary structure from their CD spectra - and vice versa - have utilised machine learning and informatics-based techniques. Louis-Jeune and co-workers⁴¹ employed theoretically derived spectra as a reference set for accurate CD based predictions of protein secondary structure. The authors used DichroCalc³⁶ to compute the theoretical CD spectra for a non-redundant set of structures, representing most proteins in the PDB, and then applied a straightforward approach to predict secondary structure content using the theoretical CD spectra as a reference set. Theoretical CD spectra were computed for 16,050 protein chains from the structures selected from the PDB.

The approach to predict the secondary structure was based on a k-nearest neighbours (k-NN) algorithm (a standard method in pattern recognition). It was assumed that the CD spectra of two proteins, which have similar secondary structural content, will also be very similar. The K2D3 method improved the predictions of secondary structure content over the previous K2D2 method, in particular for the wavelength range between 200 and 240 nm and for β -strand content.

Conversely, a recent empirical-based approach has been developed to generate CD spectra from protein atomic coordinates.⁴² The PDB2CD algorithm uses three different descriptors of structure-based information to generate a protein CD spectrum. The three descriptors are the secondary structure content, the localised topological features of these secondary structure components, and the overall structural similarity between the query - or input - structure and the proteins within the basis dataset used. The results from the application of all three descriptors define proteins in the dataset with similar characteristics to the query protein. Lastly, the PDB2CD algorithm uses a refinement method to remove proteins from those selected whose CD spectra could be considered as outliers relative to the rest of the spectra, yielding a set of proteins with CD spectra that can be used to create the resulting CD spectrum for the input protein. The PDB2CD approach was cross-validated with the SP175 dataset,³¹ and tested on an additional 14 proteins. The approach performed consistently better than the first principles DichroCalc method at replicating known protein spectra in the far-UV region.

Hall and co-workers⁴³ have developed a machine learning approach, named secondary structure neural network (SSNN), to predict protein secondary structure for a protein from its CD spectrum for the range 190 to 240 nm. SSNN is a self organising map (SOM) neural network approach that employs a database of spectra from proteins with known X-ray structures, enabling prediction of secondary structures for new proteins. SSNN has three sequential units: SSNN1 takes spectra for known proteins (the reference set) and clusters them into a map; SSNN2 creates a matching secondary structure map; and SSNN3 places unknown CD spectra on the map and gives them structure vectors to output an estimate of secondary structure and model spectra. The reference data set used to train SSNN was taken from the CDPro website⁴⁴ and is denoted as set CDDATA.48, which contains data on 48 proteins. SSNN offers a slight improvement over similar and previously developed SOM techniques. A major practical advantage of the above three approaches (K2D3, PDB2CD and SSNN) - and BeStSel - is that they are readily accessible via web-based interfaces to enable researchers to upload, straightforwardly, CD spectra (or protein atomic coordinates) to predict protein secondary structure (or CD spectra).

3 Near-ultraviolet CD

3.1 Recent experimental studies

3.1.1 Site-directed mutagenesis and tertiary structure

Site-directed mutagenesis is an established technique that has been used to investigate the contribution of aromatic residues to the near-UV CD spectra of proteins.^{45,46} Two recent studies have employed near-UV CD and site-directed mutagenesis to help optimise TAG (3-methyladenine DNA glycosylase I) homology modelling,⁴⁷ and to characterise the structural and thermodynamical properties of the L94F mutant of horse cytochrome *c*.⁴⁸

Tomar and Peddinti⁴⁷ carried out an *in silico* and experimental study to generate a three dimensional model structure of the TAG enzyme from pathogenic bacterium *Acinetobacter baumannii*. Inhibition of the TAG enzyme is a promising drug design strategy to combat infections caused by *A. baumannii* growth. Near-UV CD spectra were measured for TAG and for TAG bound to the inhibitor 3mA. On titration with the inhibitor 3mA, a significant increase of the positive peaks in the near-UV CD spectrum was observed, in particular in the 270–285 nm region. From the spectral changes, it was inferred that the local tertiary structure of Tyr16 in the TAG protein was affected upon substrate binding. The study also employed far-UV CD to investigate the changes in secondary structure upon substrate binding. Analysis of the ellipticity at 222 nm showed that, after substrate binding, helical content increased in the TAG enzyme. As part of the *in silico* modelling, mutagenesis was employed to compute $\Delta\Delta G$ values for a range of point mutants of the TAG protein binding to the 3mA substrate.

Khan and co-workers⁴⁸ studied the folding properties of h-cyt *c* using a range of experimental techniques (optical and thermal) and MD simulations. The near-UV CD spectra helped to show the effect of mutation on the tertiary structure of the protein. The wild-type (WT) protein has one Trp, four Phe and four Tyr residues. The four residues Leu94, Gly6, Phe10 and Tyr97 are involved in a common folding nucleus that guides cyt *c* to fold properly.⁴⁹ The mutation of Leu to Phe at position 94 may perturb the tertiary structure due to the bulkiness of the Phe side chain. The near-UV CD spectra for the L94F mutant in the native buffer featured bands at 282 and 289 nm, characteristic of the native protein, but which were less intense than those for WT h-cyt *c*. The two characteristic negative peaks at 282 and 289 nm, which arise from interaction of Trp59 with one heme propionate, are a signature of natively folded mammalian mitochondrial cytochromes *c*. The near-UV CD spectrum of the L94F mutant lay in between the spectrum of the native WT h-cyt *c* protein and the spectra of the unfolded (denatured) proteins. Hence the near-UV CD measurements

suggested partial loss of tertiary structure of the WT h-cyt *c* protein upon L94F mutation.

3.1.2 Time-resolved measurements

Time-resolved CD in the near-UV can be used to study the dynamics of photochemical events. In an ultrafast time-resolved CD study, Mendonça and co-workers⁵⁰ investigated the motions of the *trans-p*-coumaric acid carbonyl group following the photoexcitation of the R52Q mutant of photoactive yellow protein (PYP) over a time scale of 10 ps. In the near-UV and upon photo-excitation at 332 nm, the time-resolved CD spectra (as a function of rotation angle of the carbonyl group) featured a large negative peak that decayed after approximately 2.1 ps. A quantitative analysis of the CD signals revealed that, upon photo-excitation, the chromophore underwent a fast ($\ll 0.8$ ps) and unidirectional flipping motion of its carbonyl group. For the subset of proteins that did not undergo the photocycle, time-resolved CD provided compelling evidence that the carbonyl group returned to its initial position, leading to the formation of a nonreactive ground-state intermediate of *trans* conformation. The initial ground state is then restored within approximately 3 ps, which could be due to vibrational relaxation of the chromophore (ground-state intermediate) and restoration of the initial hydrogen-bond network. A comparative transient study of R52Q and wild-type PYP provided direct evidence that the absence of Arg52 has no effect on the conformational changes of the chromophore in the excited state.

3.2 Recent computational studies

3.2.1 Aromatic side chains and vibrational structure

Electronic transitions in the aromatic side chains of the amino acids Phe, Tyr and Trp have been incorporated into matrix-method exciton computations (to predict CD in the far and near-UV) by using parameters derived from semi-empirical⁵¹ and *ab initio* calculations.^{52,53} These established and widely-used aromatic side-chain parameter sets do not, however, take into account vibrational fine structure. The semi-empirical⁵⁴ and *ab initio*⁵³ parameters (derived from calculations on benzene) do include a set of parameters with an empirical correction for vibronic coupling in Phe. Electronic transitions to the L states in benzene are symmetry (dipole) forbidden and are accessible through vibronic coupling.

Recently, Li et al. have described an extension of the exciton approach to include vibrational structure in the far-UV.⁵⁵ Vibrational structure is an important, but not the sole, contributor to the bandwidth of transitions. The approach is exactly analogous for the near-UV,^{56,57} where Li and Hirst have developed new parameters for the Phe, Tyr and Trp side chains. For near-UV CD spectra, the main focus of interest is

the 1L_b and 1L_a transitions. For Phe and Tyr, the vibrational structure of only the 1L_b excited states was considered, since their 1L_a states are located at higher energies, but for Trp 1L_a is considered. Toluene was used to represent the aromatic side chain of Phe, *p*-cresol was used to represent Tyr and 3-methylindole was used to represent Trp. The vibrational transitions were included by extending the exciton Hamiltonian⁵⁶ 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 were not considered. The first six to nine transitions in a vibrational progression were included in the parameters. The new parameters can be accessed via the DichroCalc web interface.

Karabancheva-Christova and co-workers⁵⁸ performed MD simulations on wild-type human carbonic anhydrase II (HCAII) along with seven of its Trp mutants and computed the CD spectra, using exciton (matrix method) and TDDFT calculations for comparison with experiment. The study focused on electronic transitions to the aromatic side-chain excited states 1L_b and 1L_a in the near-UV CD; the higher energy aromatic transitions (to 1B_b and 1B_a states) were considered to contribute to the far-UV, where they mix with the large number of peptide backbone transitions. As described for the computational studies in the far-UV that also employed MD simulations to furnish an ensemble of protein structures (discussed above in section 2.3.1), the authors sought to understand how the conformational flexibility of the protein influences the interaction between the chromophores and the CD spectrum as a whole. DichroCalc was used to perform the matrix method calculations using semi-empirical⁵¹ and *ab initio*⁷ parameters for the backbone peptide transitions and *ab initio*⁵² parameters for transitions in the aromatic side chains. The TDDFT calculations used a small subset of the protein atomic coordinates (a restricted structural model): 3-methylindole to describe each of the seven Trp side chains and phenol to describe each of the eight Tyr side chains; these structures were capped with hydrogen atoms with the rest of the protein being described by a continuum solvation model with a dielectric, $\epsilon = 4.0$. MD simulations were performed for 20 ns for the wild-type protein and its seven Trp mutants (W5F, W16F, W97C, W123C, W192F, W209F and W245C). CD spectra were calculated for 40 snapshots along the MD trajectory of each protein. For the wild type, the averaged CD spectrum calculated by the matrix method over the 40 MD snapshots provided almost a two-fold better agreement to the experimental one for the main near-UV spectral feature (the negative peak at 270 nm in the experimental spectrum compared to 263 nm in the averaged spectrum) than using the static X-ray structure alone. The spectrum computed for the X-ray structure was in better agreement with experiment at wavelengths above 267 nm. The improvement in the former was explained by the MD-based structures facilitating stronger

interactions and coupling between the L excited states for clusters of Trp and Tyr residues (which are at different distances and orientated differently than in the static X-ray structure) and the asymmetric field of the protein influencing coupling between L transitions on the same aromatic side chain. The one electron effect (where intra-chromophore transitions mix) was predicted to be the mechanism by which Trp residues generate the strongest CD contributions.

The authors highlight the importance of computing spectra both for MD snapshots and for the X-ray structure to investigate the underlying coupling mechanisms of the chromophores that give rise to the CD spectrum. This was further evidenced by comparing MD-snapshot and X-ray-structure computed CD for the seven Trp mutants with experiment. Three MD-based spectra (W16F, W209F and W245C) had better agreement with experiment, two X-ray based spectra (W97C and W192F) had better agreement experiment, with the remaining two MD- and X-ray-based spectra (W5F and W123C) both comparing favourably with experiment. Differential - or difference - spectra were also computed for the seven mutants. TDDFT was used to compute CD spectra on restricted model structures of the wild type and its seven mutants. The calculations showed sensitivity to the structures and were able to distinguish between the wild-type enzyme and each mutant form. However, they did not reproduce the important spectral features - the positions and magnitudes of the minima and maxima in the experimental CD spectra. Through an analysis that compared the TDDFT-predicted CD spectra with spectra from matrix method calculations that only considered the equivalent seven Trp and eight Tyr residues, it was concluded that, to compute CD spectra reliably, it is of vital importance to include explicitly the protein environment.

3.2.2 Influence of electrostatics

The aromatic chromophores in the near-UV region are sensitive to environmental electrostatic changes, for example, the 1L_a transition of Trp is very sensitive to its surroundings, due to the large dipole moment of this excited electronic state.^{59,60} A couple of studies have considered the three mechanisms outlined in section 2.1 to investigate the influence of conformational changes and the effects of the electrostatic interactions on the near-UV CD spectrum of TEM-1 β -lactamase.^{61,62} More detailed analysis of aromatic contributions of individual chromophores examined which of these mechanisms was important.⁶³⁻⁶⁶ The electrostatic field of the ground state of a protein contributes to the one-electron mechanism (or static-field mixing) in the generation of rotational strength. A computational study of the CD properties of HCAII evaluated the effects of the protein flexibility on the quality of the calculated spectra,⁵⁸ but did not consider the vibronic transitions which can be included in matrix method calculations using new parameters.⁵⁶ We illustrate these

aspects using the CD spectrum of BPTI, which has been calculated by Jasim and co-workers.⁵⁷

Electrostatic potentials of regions around the aromatic residues of BPTI were calculated using APBS^{67,68} (with the CHARMM force field). For a residue of interest, the atomic charge and radius field parameters were set to zero in the APBS input for atoms starting at the C_β carbon atom to the final atom of a side chain, and a script was used to interpolate the potentials. The experimental near-UV spectrum for BPTI⁵⁴ is shown in Figure 4 along with spectra computed for the X-ray structure (PDB code: 5PTI) and three of the NMR (PDB code: 1PIT). This plot is analogous to Figure 1 as presented by Jasim and co-workers.⁵⁷ The near-UV spectrum computed using the X-ray structure underestimates the magnitude of the CD signal found in the experimental spectrum. Two NMR models, frames 3 and 5, are in much better agreement with the experimental spectra, with frame 17 having the least intense negative bands.

To examine the influence of the electrostatic environment and the structure on the predicted near-UV CD spectra, we consider the interpolated electrostatic potential at the atom centres of Tyr23 in NMR frames 3, 5 and 17 and the X-ray structure. Frames 3 and 5 have marginally different CD spectra and were the closest to the experimental spectrum. The atom-centred potentials for these two frames are markedly different: frame 3 has positive values whereas frame 5 has negative values. Frame 17 and the X-ray structure have, respectively, negative and predominantly positive values (data not shown). Figure 5 displays frames 3 and 5 superimposed on one another. Frame 3 features an arginine residue (Arg1) that has one of its amine groups (a positively charged side chain) at a distance of 2.7 Å from the hydroxy group on Tyr23. For frame 5 a methyl group on Arg1 is 5.5 Å from the hydroxy group on Tyr23. This structural difference may give rise to the difference in electrostatic potential experienced by Tyr23 in frames 3 and 5.

Jasim and co-workers⁵⁷ examined the contributions that each aromatic side chain makes to the computed near-UV CD signal of BPTI. For the calculated CD spectrum of frame 3 of the NMR models, the contribution of Tyr10 and Tyr21 is the largest. Tyr23 at this wavelength shows a more modest contribution. The overall rotational strength at 279.7 nm is -34.0×10^{-3} Debye Bohr Magnetron (DBM). For frame 17, however, Tyr23 seems important as the rotational strength of this residue cancels out the oppositely-signed contributions from Tyr10 and Tyr21 which dominated in the rotational strengths for frame 3. For frame 17, the computed rotational strengths for Tyr10, Tyr21 and Tyr23 at 279.7 nm are 10.8×10^{-3} , 23.2×10^{-3} and -48.0×10^{-3} DBM, respectively. The overall rotational strength at 279.7 nm is -12.0×10^{-3} DBM. The appearance of the near-UV CD spectrum itself comes from a combination of the rotational strengths of all aromatic residues - Tyr being the main contributor in this case.

This ranking of Tyr contributions to the CD spectra could help explain why the electrostatic environment

and the local structure perturbs the electronic transitions on Tyr10, Tyr21 and Tyr23 to a degree that would affect the spectra, as seen in the computed spectra for NMR frames 3 and 5 compared to frame 17. For NMR frame 3, Tyr10 is located between Pro9 which is neutral and Thr11 which has a polar side chain and it is buried in the middle of the protein. While Tyr21 is also buried inside the protein, it is located between Arg20 with a positive charge and Phe22 with a hydrophobic aromatic side chain. These preliminary findings suggest that further work should be directed toward analysis of the electrostatic potentials and their effect on the chromophores.

4 Vacuum-ultraviolet CD

4.1 Recent experimental studies

4.1.1 Synchrotron radiation CD and time-resolved measurements

Vacuum UV spectroscopy is concerned with electronic transitions with energies which fall in the sub 180 nm region of the spectrum. In order to broaden the accessible spectral range, Snyder and Rowe⁶⁹ and Sutherland and co-workers⁷⁰ in 1980 developed CD spectroscopy techniques that use synchrotron radiation. The light source provided by a beamline in a synchrotron does not have spectral emission limits. It can be optimized for any energy range between X-rays and infrared. These first setups were built, respectively, in the Synchrotron Radiation Center (SRC), located in Stoughton, Wisconsin and at the SURF II facility at the National Bureau of Standards. They were optimized for the VUV range. These developments demonstrated that the use of synchrotron radiation significantly enhances the signal-to-noise ratio of the measurement as well as its spectral resolution. In order to produce circularly polarised light, Sutherland et al. used a photo-elastic modulator that circularly polarises the incoming linearly polarised light. The polarisation varies alternately from circularly right and circularly left according the stretching direction of the photo-elastic modulator crystal. The CD signal is determined from the intensity variation measured with a photomultiplier tube. The analogue signal must be amplified and isolated with electronic devices before obtaining the final CD amplitude.

Current synchrotron radiation CD (SRCD) experimental stations in synchrotrons around the world still use the same method. The only difference is that a polariser is usually added before the photo-elastic modulator, to polarise almost perfectly linearly the incoming light. This approach gives good results for steady-state studies. It allows one to obtain an acceptable signal-to-noise ratio with an integration time below 2 seconds.

However, because the photo-elastic modulator only works at a specific wavelength with a given setting, it is necessary to scan the spectral range to obtain a complete spectrum. Thus, a spectrum from 180 nm to 260 nm with a 1 nm increment is recorded in about 2 minutes. The use of synchrotron radiation significantly enhances the signal to noise ratio and gives access to a wider spectral range, providing structural distinction inaccessible with typical desktop CD instrumentation. Steady-state SRCD measurements are used in a range of applications, such as the identification of structural changes of protein synthesised with mutated genes,⁷¹ structure determination, study of the impact of the environment on protein structure and the formation of molecular complexes.⁷²⁻⁷⁵ Many recent studies focus on system dynamics. The speed of the processes studied generally spans between a few nanoseconds to seconds. The acquisition duration of a standard SRCD setup precludes time-resolved broadband measurements in this time range. For monochromatic measurement, the theoretical time resolution limitation of these setups is linked to the modulator frequency. This frequency is usually around 50 kHz; so the resolution limitation of monochromatic measurements is about 20 μ s. The usual way to follow dynamics with a SRCD setup in a wider spectral range is to use a microfluidic device and adapt the flow of the sample or to shift the probing beam to probe different moments after the reaction trigger.⁷⁶ However, this approach consumes a lot of sample and could have significant bias due to the heterogeneity of the sample at the measurement spot.

In order to avoid this limitation, Réfrégiers et al. at the DISCO beamline SRCD endstation⁷⁷ in the synchrotron SOLEIL investigated whether the natural polarisation of the synchrotron radiation⁷⁸ can be used to measure CD. The synchrotron radiation is naturally polarised; the polarisation state of the emitted photons depends on their propagation direction (Figure 6). Photons propagating in the same plane as the electrons orbit are linearly polarised, photons located at the vertical edge of the emission cone are circularly polarised and finally photons emitted between these two planes are elliptically polarised. The polarisation ellipticity approaches unity as the angle between the plane of the electron orbit and the light emission direction increases. The polarisation ellipticity also depends of the wavelength of the emitted photons. Moreover, the direction of the polarisation depends of the emission angle; the polarisation is elliptically right polarised above the plane of the electron orbit and elliptically left polarised below it. So, the polarisation is not perfectly circularly polarised in the whole synchrotron radiation beam. However, elliptical polarisation can be used to determine CD if the left-handed and right-handed probes exhibit the same ellipticity. The only impact of the imperfect circular polarisation will be the decrease of the amplitude of the signal.

The beam provided by the DISCO beam line is a continuum (120 nm to 600 nm) separated in two parts, the upper and the lower parts, which have opposite directions of polarisation. The principle of time-resolved

SRCD is to send this continuum directly through the sample and then measure the intensity of the light using a spectrograph. The 2D detector of the spectrograph measures simultaneously the intensity of the probing light from the two parts of the beam at each wavelength in the whole range of interest. Thus, this setup records the light intensities needed to acquire a CD spectrum in a single measurement.

The temporal resolution of the setup depends of the integration time of the detector. In order to obtain the CD spectrum, Auvray et al.⁷⁹ first calibrated the polarisation of the synchrotron radiation beam and they used clusters of pixels having the same polarisation ellipticity. The initial tests were made on camphor sulfonic acid enantiomers, to check whether the spectra are balanced and to estimate the signal to noise ratio of the setup. They proposed two approaches to highlight the capability of the setup. The first aims to show that using an image intensifier and the temporal distribution of the synchrotron radiation, it is possible to reach temporal resolutions as short as 82 picoseconds. The second showed the capability of the setup to record real-time spectra, following the unfolding process of an azo-benzene peptide with a temporal resolution of 500 μ s. They used a UV sensitive camera and the light was integrated over 40 reactions cycles. The temporal resolution could probably be enhanced, and keep an equivalent signal to noise ratio, by reducing the exposure time of the camera and increasing the number of cycles.

This new setup is a tool that has great potential for the study of the dynamics of biomolecules or molecules exhibiting CD in the UV spectral range. The 82 ps resolution might give access to ultra-fast dynamics occurring during protein folding and unfolding processes. These dynamics can already be studied with MD simulations, and so it could be possible in the near future to combine theoretical approaches with time-resolved SRCD data to enhance the accuracy of the models.

4.1.2 Applications of synchrotron radiation CD

In a far-UV SRCD study, Nesgaard and co-workers⁸⁰ obtained CD spectra of 13 solubilised proteins and dried protein films to assess the advantage of extending the spectroscopic window (down to the 130 to 160 nm range) weighed against the risk of structural rearrangement of proteins due to altered hydration. Dry phase samples are required for studies using synchrotron radiation in the VUV region. They were also concerned with elucidating whether the drying process perturbed the CD spectra to an extent that prevented a meaningful interpretation of the spectra for dried protein samples. The 13 globular model proteins comprised three classes of secondary structure: three mainly α , three mixed α - β and seven mainly β proteins. The observation that the spectral features over the range 170 to 260 nm upon drying for all structural classes were preserved suggested that the protein films did not undergo significant structural

changes compared to the solution state proteins. In addition, there was good correspondence between PDB and CD data to determine the structural class for both solubilised proteins and protein films with the exception of dry phase transthyretin (a mainly β protein). A principal component analysis of wet and dry spectra with secondary structure also confirmed that controlled drying does not significantly alter structure. Drying of the 13 samples expanded the spectroscopic window towards 130 nm in the VUV and revealed new spectral features in the CD spectra of the proteins (Figure 7). The mainly α -helical proteins (ACBP, α -lactalbumin, porcine serum albumin) and mixed α - β proteins (S6 ribosomal subunit) featured low wavelength positive band maxima around 140 to 142 nm; mainly β , sandwich (architecture) proteins (concanavalin A, thaumatin, transthyretin, TNFN3, TII27) had maxima at 149 to 157 nm; mainly β , other (architecture) proteins (carbonic anhydrase, jacalin, trypsin) had positive maxima, for instance, at 147 nm for trypsin; and β -galactosidase had a positive maximum at 166 nm. In addition to the set of 13 proteins, the study obtained spectra of protein aggregates and fibrils of various types in the wet and dry phase to acquire both far-UV and VUV-range structural information. For the dry-phase VUV CD spectra: the lysozyme fibrils featured a positive peak at 145 nm; the α -synuclein fibrils a peak at 151 nm; and the SerADan aggregates a peak at 148 and 149 nm for a pH of 7.5 and 5.0, respectively. The 13 globular model proteins and the fibril and aggregate samples were found to have the expected β -rich characteristics in their CD spectra, but with varying intensities of the low wavelength band compared with the intensity of the positive $\pi_{nb}\pi^*$ band at 190 nm; this was most notable for α -synuclein fibrils.

In a study seeking to understand the conformational properties of amyloid fibrils, Matsuo and co-workers⁸¹ employed VUVCD spectroscopy, MD simulations and CD theory to characterise the intermolecular structures of β_2 -microglobulin (β_{2m}) core fragments in the amyloid fibrils. They considered six models of $\beta_{2m_{21-29}}$ core fragments ($[^{21}\text{NFLNCYVSG}^{29}]$) for MD simulations from which VUV, far-UV and near-UV CD was computed (using the matrix method) for 50 snapshot structures taken every 400 ps from the 20 ns MD trajectories. The CD spectrum of each simulated structure was the average of calculated spectra for the 50 snapshot structures. The matrix method calculations incorporated parameters for three electronic transitions in the peptide backbone ($n\pi^*$, $\pi_{nb}\pi^*$ and $\pi_b\pi^*$) and four transitions in the aromatic side chains of Phe and Tyr (1L_a , 1L_a , 1B_a and 1B_b). Charge-transfer transitions were not considered, although the program employed, PROTPOL,^{82,83} included contributions of the VUV region to transitions in the accessible UV by considering the contributions of polarizabilities. The six model fragments differed in how the constituent β -sheets lined up and on how they stacked in either a parallel or antiparallel manner. It was found that “model 6” (three clusters of “model 2”, two parallel β -sheets stacked in an antiparallel manner

and connected with disulfide bonds between the cysteine side chains, lined up so their sheet surfaces stack in an antiparallel manner) was of most interest. Model 6 had a stable β -sheet conformation during the MD simulation; had the same positions of β -strand structures on the sequence level as found in an FTIR analysis; exhibited the theoretical spectrum similar to the experimentally observed spectrum in the wavelength range from the near-UV (300 nm) to the VUV (178 nm) region; and had the three unique interactions between Phe and Tyr side chains, necessary for explaining the experimentally observed and theoretically calculated CD spectra. Three characteristics of the experimental CD spectrum for β_2m_{21-29} at pH 8.5, with which the computed spectra is in good agreement, are a negative peak around 185 nm, a positive peak around 202 nm and negative peak around 248 nm. The results led the authors to propose that the conformations of β_2m_{21-29} amyloid fibrils consist of parallel β -sheets stacked in an antiparallel fashion, and that the Phe-Tyr interactions are formed amid the parallel β -sheets.

4.2 Recent computational studies

4.2.1 Charge transfer transitions

Oakley and Hirst⁸⁴ employed the matrix method to calculate CD spectra for 31 proteins in the far-UV and in the VUV (that is, over the range 160 to 240 nm). To describe charge-transfer transitions between two adjacent peptide bonds, they employed parameter sets derived from *ab initio* calculations on *N*-acetylglycine-*N'*-methylamide, a diamide species, at various geometries (ϕ and ψ angles) to describe α -helical and β -sheet structures. Their calculations significantly improved the accuracy of calculated protein CD spectra between 170 and 190 nm. They also gave the first assignment of an inter-peptide charge-transfer band in a CD spectrum. The matrix method calculations performed less well at higher energies and did not predict the positive bands at approximately 160 nm in α -helices and β -sheets. The authors considered the possibility that these bands are caused by non-nearest-neighbour charge-transfer transitions, such as those across hydrogen bonds. An alternative suggestion is that the bands at 160 nm could be caused by coupling between charge-transfer transitions and the high-energy local $n'\pi^*$ and $\pi_b\pi^*$ transitions.

In a related study, Bulheller and co-workers⁷³ calculated the VUV CD spectra of 71 proteins, for which experimental SRCD spectra and X-ray crystal structures were available (the SP175 reference data set³¹). CD spectra were calculated for the 170 to 240 nm region using the matrix method by considering charge-transfer transitions between two adjacent peptides (four transitions), local peptide transitions (two transitions) and electronic transitions in the aromatic side chains Phe, Tyr and Trp (four transitions on each). As in the study by Oakley and Hirst,⁸⁴ the agreement between calculated and experimental VUV CD spectra was improved

when charge-transfer transitions were incorporated; this agreement was improved further when side chains were included. In the majority of the calculated spectra, the inclusion of charge-transfer transitions increased the negative band around 170 nm considerably, which improved the agreement with experiment. The effects in the far-UV region are minimal and slightly worsened the correlation with experiment. The positions of the bands at 190 and 208 nm were not affected for most of the 71 proteins: the CD intensity changes are minuscule in the far-UV region. In addition, Bulheller and co-workers analysed the influence of the protein conformation on charge-transfer transitions. The analysis showed that the $n\pi^*$ charge-transfer transitions are most important in α -helical proteins; in contrast, in β -strand proteins the $\pi_{nb}\pi^*$ charge-transfer transition along the chain in the N- to C-terminus direction is most dominant.

5 Protein dynamics and CD

As discussed in the preceding sections, CD can be used to assess the temporal changes in secondary and tertiary structure of proteins. In the far-UV the ellipticity at 222 nm, $[\Theta]_{222}$, is routinely used to determine the α -helical content of a protein and may be measured - or computed (from, e.g., MD snapshots and matrix method calculations) - over a time period of interest. Figure 1 displays the far-UV CD spectrum of a predominantly α -helical protein and features an intense negative band at 222 nm, which is absent in the CD spectrum of a β -sheet or a random coil. The negative ellipticity at 222 nm is due to the $n\pi^*$ transition in the amide chromophore when it adopts an α -helical geometry (Figure 2). As an example, $[\Theta]_{222}$ was employed by Meißner and co-workers³⁵ to analyse MD simulations to monitor the conformational change of an α -helical oligopeptide (4DAR5) adsorbing on an anionic amorphous silica (SiO_2) surface (see section 2.3.1). In the near-UV the ellipticity is sensitive to the spatial coupling of the L electronically excited states ($\pi\pi^*$ transitions) in the aromatic side chains of Phe, Tyr and Trp residues. As an example, Karabancheva-Christova and co-workers⁵⁸ performed MD simulations and CD calculations on HCAII (see section 3.2.1) to investigate the coupling mechanisms, outlined in section 2.1, that contribute to the near-UV CD for this protein and the sensitivity of the coupling strengths to the dynamics of the tertiary structure of the protein.

6 Conclusion and outlook

This review has described both experimental and theoretical studies that have highlighted the utility of modern CD spectroscopy for proteins and its complementarity with other optical and structural techniques. The theoretical studies, using a range of approximations to compute either CD spectra from protein structure

or protein structure from CD spectra, have illustrated the need for accurate sampling of the conformational ensemble of biomolecules and the requirement to account for the influence of the environment in which the biomolecule and its chromophores reside on the electronic transitions in the chromophoric groups. Work remains to be done before a fully quantitative and complete first principles model is achieved. Nonetheless, this review has documented computational approaches that are leading the way toward this ambitious goal. We hope that readers of this article are encouraged to deploy CD spectroscopy in their experimental studies (for example, to investigate processes that change local conformation of a protein) and to pursue supportive molecular modelling studies (for example, to analyse structural snapshots from MD simulations) to help interpret their CD experiments.

7 Acknowledgements

We thank the University of Nottingham Propulsion Futures Beacon for funding towards this research. S.B.J. thanks The Iraqi Ministry of Higher Education and Scientific Research. Beamtime at Synchrotron SOLEIL is acknowledged under project #20161257. F.A. is the recipient of a scholarship from the University of Nottingham and Synchrotron SOLEIL.

Electrostatic potential data created during this research are openly available from the Figshare data repository at <https://doi.org/10.6084/m9.figshare.8124227>.

8 Author Contributions

Conceptualization, J.D.H.; Methodology, J.D.H., S.B.J., and D.M.R.; Software, D.M.R. and N.T.D.; Investigation, S.B.J., N.T.D., and D.M.R.; Writing - Original Draft, D.M.R., F.A., N.T.D., and S.B.J.; Writing - Review & Editing, J.D.H., D.M.R., and M.R.; Supervision, J.D.H. and M.R.; Funding Acquisition, J.D.H. and M.R.

9 Declaration of Interests

The authors declare no competing interests.

References

- [1] Kelly, S.M., Jess, T.J., and Price, N.C. (2005). How to study proteins by circular dichroism. *BBA Proteins Proteom.* *1751*, 119-139.
- [2] Woody, R.W., and Tinoco, I. (1967). Optical rotation of orientated helices. III. Calculation of the rotatory dispersion and circular dichroism of the alpha- and 310-helix. *J. Chem. Phys.* *46*, 4927-4945.
- [3] Bayley, P.M., Nielsen, E.B., and Schellman, J.A. (1969). Rotatory properties of molecules containing two peptide groups: theory. *J. Phys. Chem.* *73*, 228-243.
- [4] Condon, E.U., Altar, W., and Eyring, H. (1937). One-electron rotatory power. *J. Chem. Phys.* *5*, 753-775.
- [5] Rosenfeld, L. (1928). Quantenmechanische theorie der naturlichen optischen aktivitat von flussigkeiten und gasen. *Z. Phys.* *52*, 161-174.
- [6] Goux, W.J., and Hooker, T.M. (1980). Chiroptical properties of proteins. I. Near-ultraviolet circular dichroism of ribonuclease S. *J. Am. Chem. Soc.* *102*, 7080-7087.
- [7] Besley, N. A., and Hirst, J. D. (1999). Theoretical studies toward quantitative protein circular dichroism calculations. *J. Am. Chem. Soc.* *121*, 9636-9644.
- [8] Schellman, J.A. (1968). Symmetry rules for optical rotation. *Acc. Chem. Res.* *1*, 144-151.
- [9] Ferrone, F.A., Hopfield, J.J., and Schnatterly, S.E. (1974). Measurement of transient circular dichroism: A new kinetic technique. *Rev. Sci. Instrum.* *45*, 1392-1396.
- [10] Bayley, P.M., and Anson, M. (1974). Stopped-flow circular dichroism: a new fast-kinetic system. *Biopolymers* *13*, 401-405.
- [11] Anson, M., Martin, S.R., and Bayley, P.M. (1977). Transient CD measurements at submillisecond time resolution - application to studies of temperature-jump relaxation of equilibria of chiral biomolecules. *Rev. Sci. Instrum.* *48*, 953-962.
- [12] Luchins, J., and Beychook, S. (1978). Far-Ultraviolet Stopped-Flow Circular Dichroism. *Science* *199*, 425-426.
- [13] Akiyama, S., Takahashi, S., Ishimori, K. and Morishima, I. (2000). Stepwise formation of α -helices during cytochrome c folding. *Nat. Struct. Biol.* *7*, 514-520.

- [14] Xie, X.L., and Simon, J.D. (1989). Picosecond time-resolved circular dichroism spectroscopy: experimental details and applications. *Rev. Sci. Instrum.* *60*, 2614–2627.
- [15] Dartigalongue, T., and Hache, F. (2003). Precise alignment of a longitudinal Pockels cell for time-resolved circular dichroism experiments. *J. Opt. Soc. Am. B.* *20*, 1780–1787.
- [16] Niezborala, C., and Hache, F. (2006). Measuring the dynamics of circular dichroism in a pump–probe experiment with a Babinet–Soleil Compensator. *J. Opt. Soc. Am.* *23*, 2418–2424.
- [17] Hache, F. (2009). Application of time-resolved circular dichroism to the study of conformational changes in photochemical and photobiological processes. *J. Photochem. Photobiol.* *204*, 137–143.
- [18] Lewis, J.W., Tilton, R.F., Einterz, C.M., Milder, S.J., Kuntz, I.D., and Kliger, D.S. (1985). New technique for measuring circular dichroism changes on a nanosecond time scale - Application to (carbonmonoxy)myoglobin and (carbonmonoxy)hemoglobin. *J. Phys. Chem.* *89*, 289–294.
- [19] Zhang, C.F., Lewis, J.W., Cerpa, R., Kuntz, I.D., and Kliger, D.S. (1993). Nanosecond circular dichroism spectral measurements - Extension to the far-ultraviolet region. *J. Phys. Chem.* *97*, 5499–5505.
- [20] Wen, Y.X., Chen, E.F., Lewis, J.W., and Kliger, D.S. (1996). Nanosecond time-resolved circular dichroism measurements using an upconverted Ti:sapphire laser. *Rev. Sci. Instrum.* *67*, 3010–3016.
- [21] Oppermann, M., Bauer, B., Rossi, T., Zinna, F., Helbing, J., Lacour, J., and Chergui, M. (2019). Ultrafast broadband circular dichroism in the deep ultraviolet. *Optica* *6*, 56–60.
- [22] Moffitt, W., Fitts, D.D., and Kirkwood, J.G. (1957). Critique of the Theory of Optical Activity of Helical Polymers. *Proc. Natl. Acad. Sci. U.S.A.* *43*, 723–730.
- [23] Moffitt, W. (1956). Optical Rotatory Dispersion of Helical Polymers. *J. Chem. Phys.* *25*, 467–478.
- [24] Chen, F.Y., Lee, M.T., and Huang, H.W. (2002). Sigmoidal concentration dependence of antimicrobial peptide activities: A case study on alamethicin. *Biophys. J.* *82*, 908–914.
- [25] Bürck, J., Wadhvani, P., Fanghänel, S., and Ulrich, A.S. (2016). Oriented Circular Dichroism: A Method to Characterize Membrane-Active Peptides in Oriented Lipid Bilayers. *Acc. Chem. Res.* *49*, 184–192.
- [26] Whitmore, L., Miles, A.J., Mivridis, L., Janes, R.W., and Wallace, B.A. (2017). PCDDDB: new developments at the Protein Circular Dichroism Data Bank. *Nucleic Acids Res.* *45*, D303–D307.

- [27] 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.
- [28] Platt, J.R. (1949). Classification of spectra of cata-condensed hydrocarbons. *J. Chem. Phys.* *17*, 484-495.
- [29] Micsonai, A., Wien, F., Kernya, L., Lee, Y.-H., Goto, Y., Réfrégiers, M., and Kardos, J. (2015). Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc. Nat. Acad. Sci. USA* *112*, E3095-E3103.
- [30] Kabsch W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* *22*, 2577–637.
- [31] Lees, J.G., Miles, A.J., Wien, F., and Wallace, B.A. (2006). A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics* *22*, 1955-1962.
- [32] Dawson, N.L., Lewis, T.E., Das, S., Lees, J.G., Lee, D., Ashford, P., Orengo, C.A., and Sillitoe, I. (2017). CATH: an expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res.* *45*, D289–D295.
- [33] Micsonai, A., Wien, F., Bulyáki, É., Kun, J., Moussong, É., Lee, Y.H., Goto, Y., Réfrégiers, M., and Kardos, J. (2018). BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res.* *46*, W315–W322.
- [34] Kaminsky, J., Kubelka, J., and Bour, P. (2011). Theoretical modeling of peptide α -helical circular dichroism in aqueous solution. *J. Phys. Chem. A* *115*, 1734-1742.
- [35] Meißner, R.H., Schneider, J., Schiffels, P., and Ciacchi, L.C. (2014). Computational prediction of circular dichroism spectra and quantification of helicity loss upon peptide adsorption on silica. *Langmuir* *30*, 3487-3494.
- [36] Bulheller, B.M., and Hirst, J.D. (2009). DichroCalc – circular and linear dichroism online. *Bioinformatics* *25*, 539–540.
- [37] Ianeselli, A., Orioli, S., Spagnoli, G., Faccioli, P., Cupellini, L., Jurinovich, S., and Mennucci, B. (2018). Atomic detail of protein folding revealed by an ab initio reappraisal of circular dichroism. *J. Am. Chem. Soc.* *140*, 3674-3682.

- [38] Seibert, J., Bannwarth, C., and Grimme, S. (2017). Biomolecular structure information from high-speed quantum mechanical electronic spectra calculation. *J. Am. Chem. Soc.* *139*, 11682-11685.
- [39] Grimme, S., and Bannwarth, C. (2016). Ultra-fast computation of electronic spectra for large systems by tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB). *J. Chem. Phys.* *145*, 054103.
- [40] Gattuso, H., Garcia-Iriepa, C., Sampredo, D., Monari, A., and Marazzi, M. (2017). Simulating the electronic circular dichroism spectra of photoreversible peptide conformations. *J. Chem. Theory Comput.* *13*, 3290-3296.
- [41] Louis-Jeune, C., Andrade-Navarro, M.A., and Perez-Iratxeta, C. (2012). Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins: Struct. Func. Bioinform.* *80*, 374-381.
- [42] Mavridis, L., and Janes, R.W. (2017). PDB2CD: A web-based application for the generation of circular dichroism spectra from protein atomic coordinates. *Bioinformatics* *33*, 56-63.
- [43] Hall, V., Nash, A., Hines, E., and Rodger, A. (2013). Elucidating protein secondary with circular dichroism and a neural network. *J. Comp. Chem.* *34*, 2774-2786.
- [44] CDPro: A Software Package for Analyzing Protein CD Spectra. <https://sites.bmb.colostate.edu/sreeram/CDPro/>
- [45] Freskgard, P.O., Martensson, 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.
- [46] Woody, A.Y.M., and Woody, R.W. (2003). Individual tyrosine side-chain contributions to circular dichroism of ribonuclease. *Biopolymers* *72*, 500-513.
- [47] Tomar, J.S., and Peddinti, R.K. (2016). Optimized method for TAG protein homology modeling: *In silico* and experimental structural characterization. *Int. J. Biol. Macromol.* *88*, 102-112.
- [48] Khan, S.H., Kumar, A., Prakash, A., Taneja, B., Islam, A., Hassan, I., and Ahmad, F. (2016). Structural and thermodynamic characterisation of L94F mutant of horse cytochrome *c*. *Int. J. Biol. Macromol.* *92*, 202-212.

- [49] Ptitsyn, O.B. (1998). Protein folding and protein evolution: common folding nucleus in different sub-families of c-type cytochromes? *J. Mol. Biol.* *278*, 655-666.
- [50] Mendonça, L., Hache, F., Changenet-Barret, P., Plaze, P., Chosrowjan, H., Taniguchi, S., and Y. Imamoto, Y. (2013). Ultrafast Carbonyl Motion of the Photoactive Yellow Protein Chromophore Probed by Femtosecond Circular Dichroism. *J. Am. Chem. Soc.* *135*, 14637-14643.
- [51] Woody, R.W., and Sreerama, N. (1999). Comment on “Improving protein circular dichroism calculations in the far-ultraviolet through reparametrizing the amide chromophore” *J. Chem. Phys.* *109*, 782, (1998). *J. Chem. Phys.* *111*, 2844-2845.
- [52] 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.
- [53] Rogers, D.M., and Hirst, J.D. (2004). First-principles calculations of protein circular dichroism in the near ultraviolet. *Biochemistry* *43*, 11092-11102.
- [54] 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.
- [55] Li, Z., Robinson, D., and Hirst, J.D. (2015). Vibronic structure in the far-UV electronic circular dichroism spectra of proteins. *Faraday Discuss.* *177*, 329-344.
- [56] Li, Z., and Hirst, J.D. (2017). Quantitative first principles calculations of protein circular dichroism in the near-ultraviolet. *Chem. Sci.* *8*, 4318-4333.
- [57] Jasim, S.B., Li, Z., Guest, E.E., and Hirst, J.D. (2018). DichroCalc: improvements in computing protein circular dichroism spectroscopy in the near-ultraviolet. *J. Mol. Biol.* *430*, 2196-2202.
- [58] Karabancheva-Christova, T.G., Carlsson, U., Balali-Mood, K., Black, G.W., and Christov, C.Z. (2013). Conformational effects on the circular dichroism of Human Carbonic Anhydrase II: A multilevel computational study. *PLoS ONE* *8*, e56874.
- [59] Kurapkat, G., Krüger, P., Wollmer, A., Fleischhauer, J., Kramer, B., Zobel, E., Koslowski, A., Botterweck, H., and Woody, R.W. (1997). Calculations of the CD spectrum of bovine pancreatic ribonuclease. *Biopolymers* *41*, 267-287.

- [60] Song, P.S., and Kurtin, W.E. (1969). Photochemistry of the model phototropic system involving flavines and indoles. III. A spectroscopic study of the polarized luminescence of indoles. *J. Am. Chem. Soc.* *91*, 4892-4906.
- [61] Christov C., and Karabancheva, T. (2004). Mechanisms of generation of rotational strengths in TEM-1 β -lactamase. Part I: theoretical analysis of the influence of conformational changes in the near-UV. *Chem. Phys. Lett.* *396*, 282-287.
- [62] Christov C., Kantardjiev A., Karabancheva T., and Tielens, F. (2004). Mechanisms of generation of the rotational strengths in TEM-1 β -lactamase. Part II: theoretical study of the effects of the electrostatic interactions in the near-UV. *Chem. Phys. Lett.* *400*, 524-530.
- [63] Christov C.Z., Karabancheva T.G., and Lodola, A. (2008). Relationship between chiroptical properties, structural changes and interactions in enzymes: A computational study on β -lactamases from class A. *Comp. Bio. Chem.* *32*, 167-175.
- [64] Christov C.Z., Karabancheva T.G., and Lodola, A. (2008). Aromatic interactions and rotational strengths within protein environment: An electronic structural study on β -lactamases from class A. *Chem. Phys. Lett.* *456*, 89-95.
- [65] Karabancheva T.G., Donev R., Balali-Mood K., and Christov, C.Z. (2010). Individual contributions of the aromatic chromophores to the near-UV Circular Dichroism in class A β -lactamases: A comparative computational analysis. *Biophys. Chem.* *151*, 39-45.
- [66] Christov C.Z., Karabancheva, T. G. (2011). Computational insight into protein circular dichroism: detailed analysis of contributions of individual chromophores in TEM-1 β -lactamase. *Theor. Chem. Acc.* *128*, 25-37.
- [67] Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* *98*, 10037-10041.
- [68] Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L.E., Brookes, D.H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D.W., Dolinsky, T., Konecny, R., Koes, D.R., Nielsen, J.E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G.W., Holst, M.J., McCammon, J.A., and Baker, N.A. (2018). Improvements to the APBS biomolecular solvation software suite. *Prot. Sci.* *27*, 112-128.

- [69] Snyder, P.A., and Rowe, E.M. (1980). The first use of synchrotron radiation for vacuum ultraviolet circular dichroism measurements. *Nuclear Instruments and Methods* 172, 345-349.
- [70] Sutherland, J.C., Desmond, E.J., and Takacs, P.Z. (1980). Versatile spectrometer for experiments using synchrotron radiation at wavelengths greater than 100 nm. *Nuclear Instruments and Methods*. 172, 195-199.
- [71] Evans, P., Wyatt, K., Wistow, G.J., Bateman, O.A., Wallace, B.A., and Slingsby, C. (2004). The P23T cataract mutation causes loss of solubility of folded γ D-crystallin. *J. Mol. Biol.* 343, 435-444.
- [72] Miles, A.J., and Wallace, B.A. (2006). Synchrotron radiation circular dichroism spectroscopy of proteins and applications in structural and functional genomics. *Chem. Soc. Rev.* 35, 39-51.
- [73] Bulheller, B.M., Miles, A.J., Wallace, B.A., and Hirst, J.D. (2008). Charge-transfer transitions in the vacuum-ultraviolet of protein circular dichroism spectra. *J. Phys. Chem. B.* 112, 1866-1874.
- [74] Matsuo, K., Yonehara, R., and Gekko, K. (2004). Secondary-structure analysis of proteins by vacuum-ultraviolet circular dichroism spectroscopy. *J. Biochem.* 135, 405-411.
- [75] Wallace, B.A. (2009). Protein characterisation by synchrotron radiation circular dichroism spectroscopy. *Q Rev Biophys* 42, 317-370.
- [76] Kane, A.S., Hoffmann, A., Baumgärtel, P., Seckler, R., Reichardt, G., Horsley, D.A., Schuler, B., and Bakajin, O. (2008). Microfluidic Mixers for the Investigation of Rapid Protein Folding Kinetics Using Synchrotron Radiation Circular Dichroism Spectroscopy. *Anal. Chem.* 80, 9534-9541.
- [77] Refregiers, M., Wien, F., Ta, H.P., Premvardhan, L., Bac, S., Jamme, F., Rouam, V., Lagarde, B., Polack, F., Giorgetta, J.L., Ricaud, J.P., Bordessoule, M., and Giuliani, A. (2012). Synchrotron radiation circular dichroism endstation at SOLEIL. *J. Sync. Rad.* 19, 831-835.
- [78] Heinzmann, U., Osterheld, B., and Schafers, F. (1982). Measurements and calculations of the circular polarization and of the absolute intensity of synchrotron radiation in the wavelength range from 40 to 100 nm. *Nucl. Instrum. Methods Phys. Res* 195, 395-398.
- [79] Auvray, F., Denetiere, D., Giuliani, A., Jamme, F., Wien, F., Nay, B., Zirah, S., Polack, F., Meneglier, C., Lagarde, B., Hirst, J.D., and Réfrégiers, M. Published online 02.07.2019. Time resolved transient circular dichroism spectroscopy using synchrotron natural polarisation. <https://doi.org/10.26434/chemrxiv.8427188.v1>

- [80] Nesgaard, L.W., Hoffmann, S.V., Andersen, C.B., Malmendal, A., and Otzen, D.E. (2008). Characterization of dry globular proteins and protein fibrils by synchrotron radiation vacuum UV circular dichroism. *Biopolymers* *89*, 779-795.
- [81] Matsuo, K., Hiramatsu, H., Gekko, K., Namatame, H., Taniguchi, M., and Woody, R.W. (2014). Characterization of Intermolecular Structure of β_2 -Microglobulin Core Fragments in Amyloid Fibrils by Vacuum-Ultraviolet Circular Dichroism Spectroscopy and Circular Dichroism Theory. *J. Phys. Chem. B* *118*, 2785-2795.
- [82] Sreerama, N., Woody, R. W. (2004). Computation and analysis of protein circular dichroism spectra. *Methods Enzymol.* *383*, 318-351.
- [83] Woody, R.W. (2009). Circular dichroism spectrum of peptides in the poly (Pro) II conformation. *J. Am. Chem. Soc.* *131*, 8234-8245.
- [84] Oakley, M.T., and Hirst, J.D. (2006). Charge-Transfer Transitions in Protein Circular Dichroism Calculations. *J. Am. Chem. Soc.* *128*, 12414-12415.

10 Figures

Figure 1: Illustrative CD spectra for three secondary structures.

CD spectrum for a predominantly α -helical protein (red line), β -sheet protein (green line) and a random coil protein (blue line).

Figure 2: Diamide in an α -helical conformation.

Molecular structure of the α -helical diamide with dihedral angles $\phi = -48^\circ$ and $\psi = -57^\circ$.

Figure 3: Computed CD spectrum for a diamide in an α -helical conformation.

CD spectrum in the far-UV for an α -helical diamide computed by the matrix method.

Figure 4: Experimental and computed near-UV CD spectra for BPTI.

CD spectrum in the near-UV from experiment (black line) and computed by the matrix method for the X-ray structure (green line) and three NMR models, frame 3 (blue line), frame 5 (magenta line) and frame 17 (red line).

Figure 5: Frame 3 (blue) and frame 5 (pink) of the NMR structure of BPTI (PDB code: 1PIT). Tyr23 is drawn as a liquorice model in each structure and is coloured by electrostatic potential of the environment calculated at each atom. Arg1 in each structure is also drawn as a liquorice model and is coloured corresponding to the colour of the Tyr of the same structure. Calculated nearest atom distances (\AA) are shown for each structure. This figure was drawn using PyMOL.

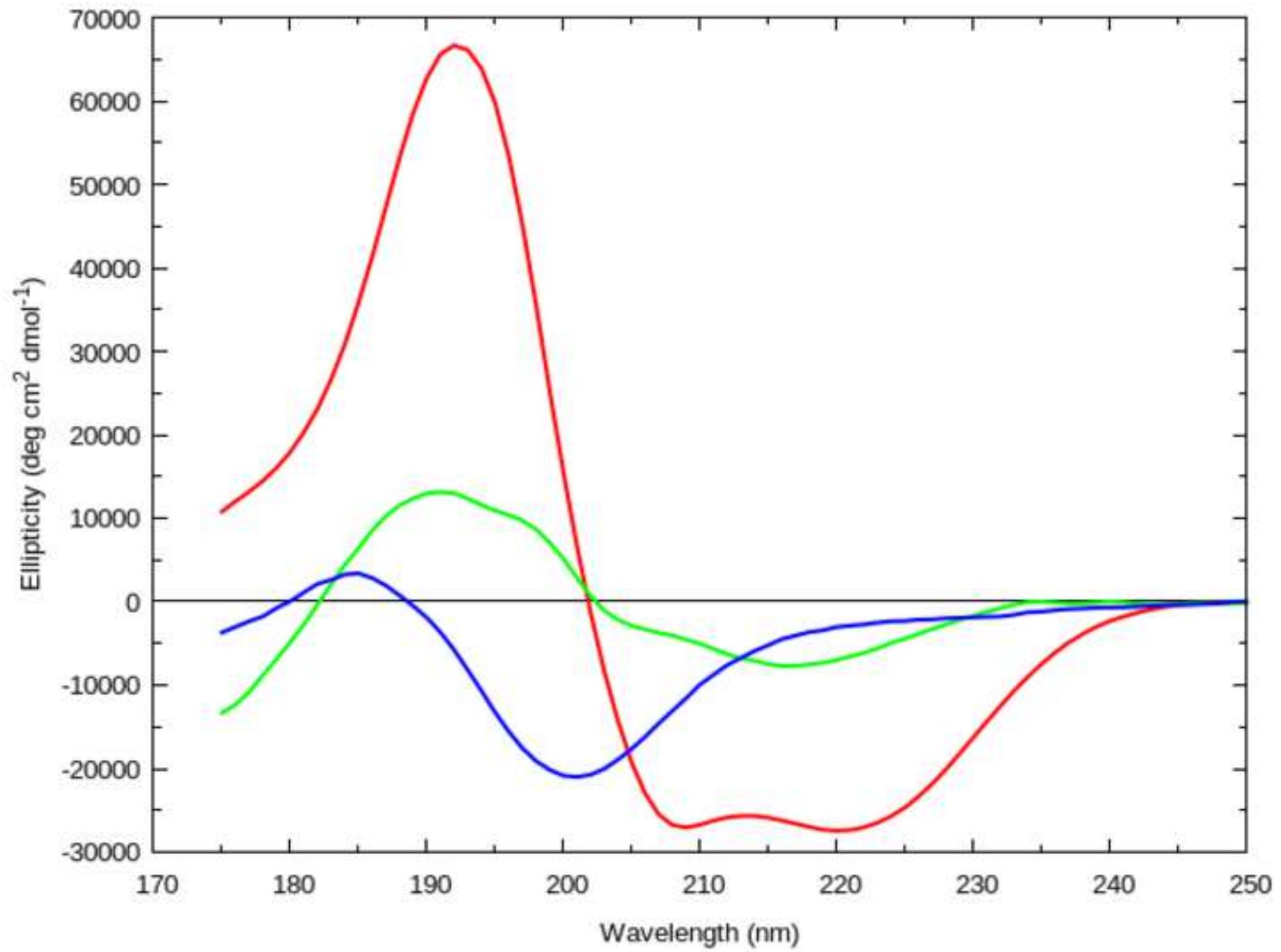
Figure 6: Schematic of synchrotron radiation emission.

The electron bunches (blue dot) emit light while they are radially accelerated when they pass through the bending magnet. The radiation is elliptically right polarised above the electrons orbit and elliptically left polarised below it.

Figure 7: Spectral features at low wavelengths correspond to secondary structure.

(A) Mainly α -helical and α - β -proteins. (B) Sandwich architectures. (C) Mainly β -sheet proteins with more complex architectures. Note that while carbonic anhydrase is classed as α - β , it contains mainly β -sheet. Adapted with permission from reference 80. Copyright 2008 John Wiley and Sons.

Figure 1



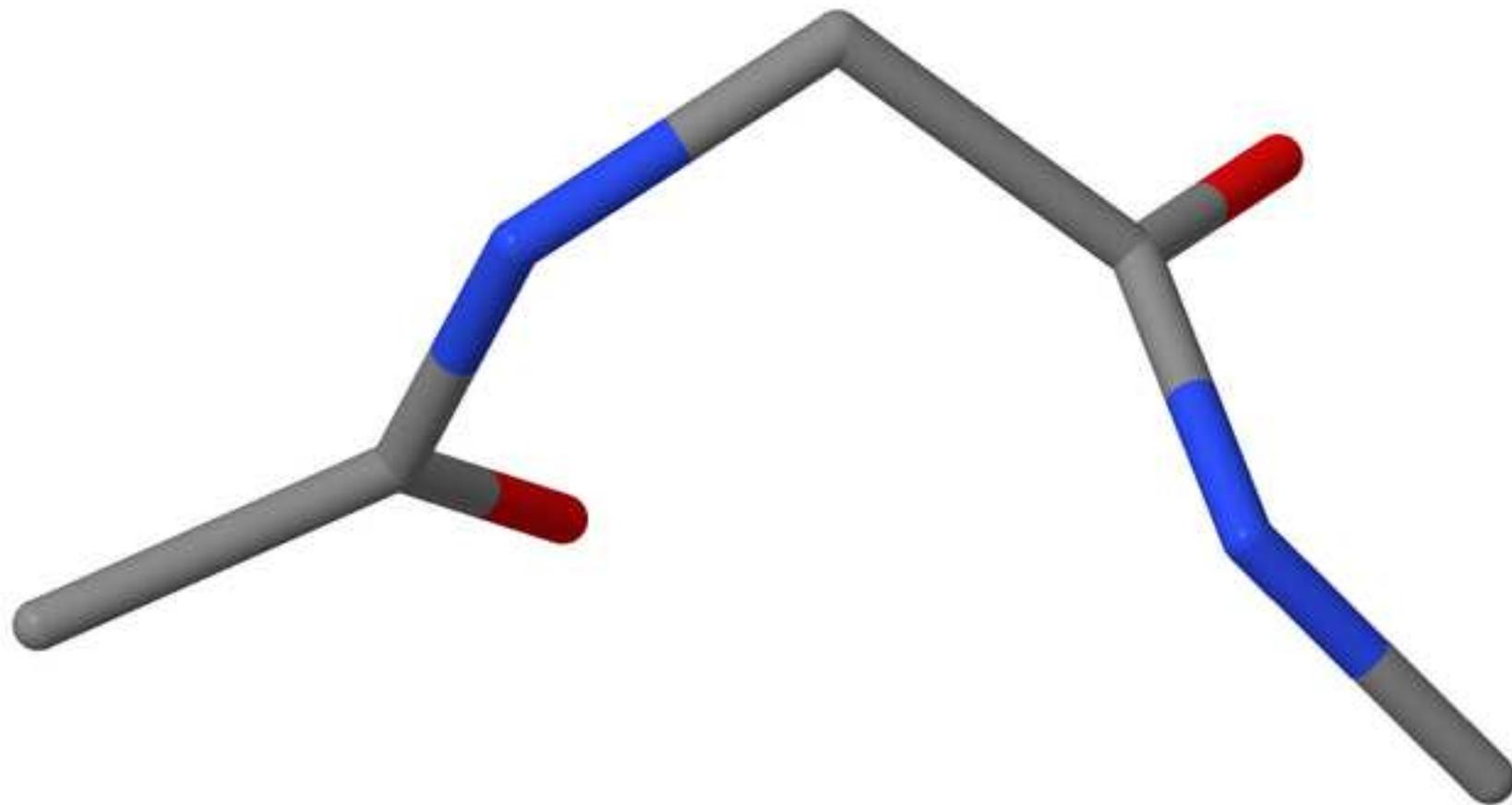


Figure 3

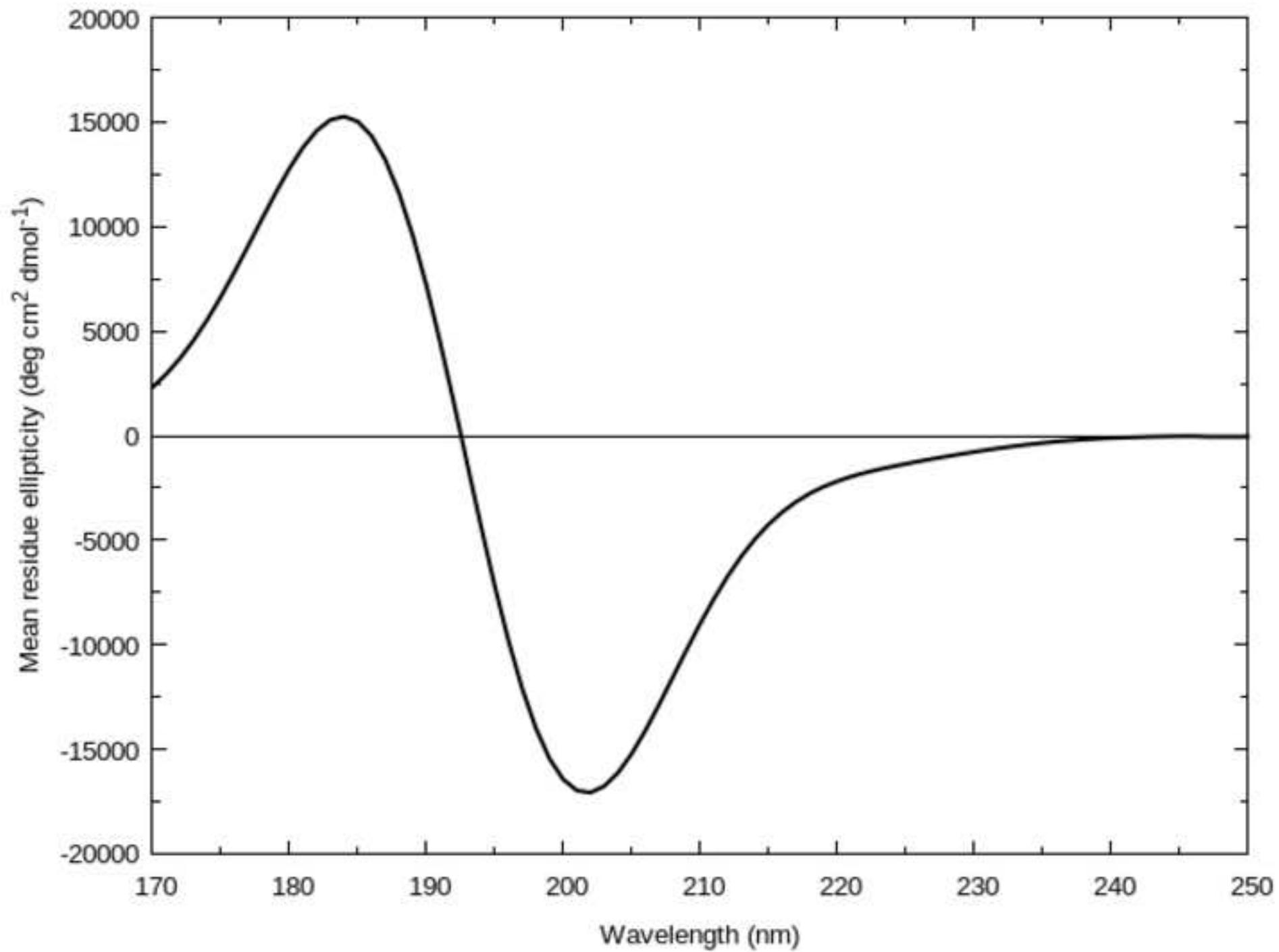


Figure 4

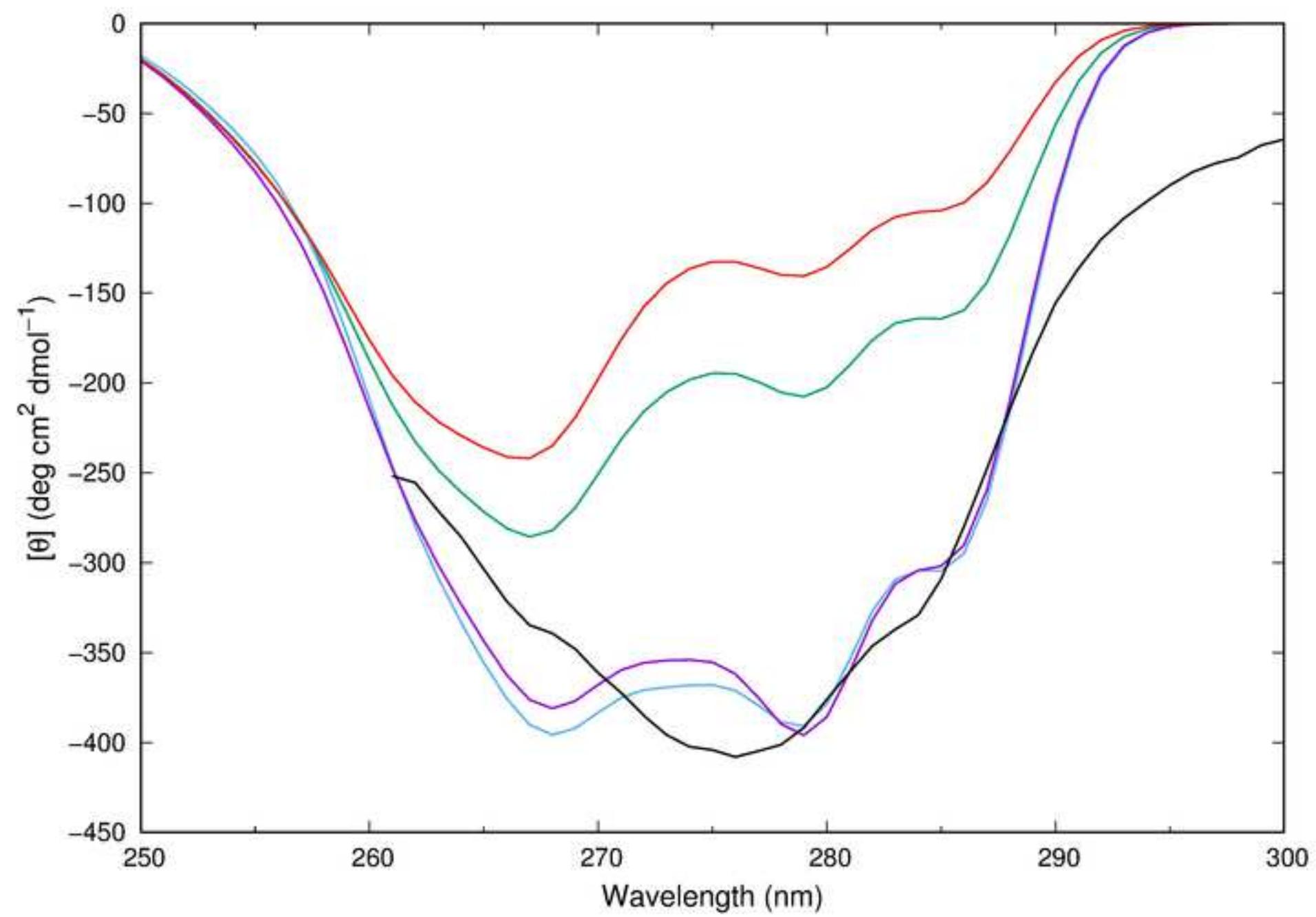
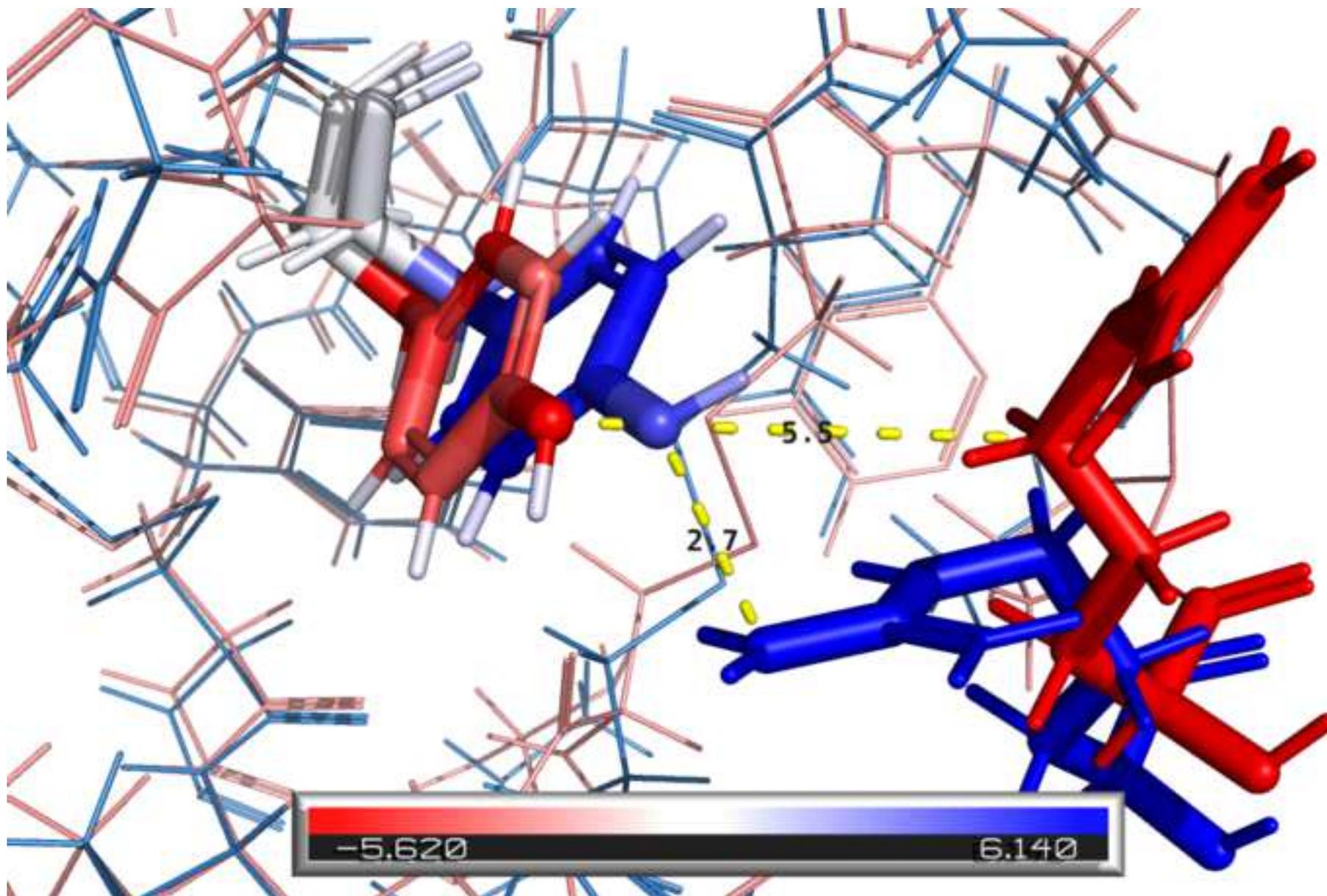
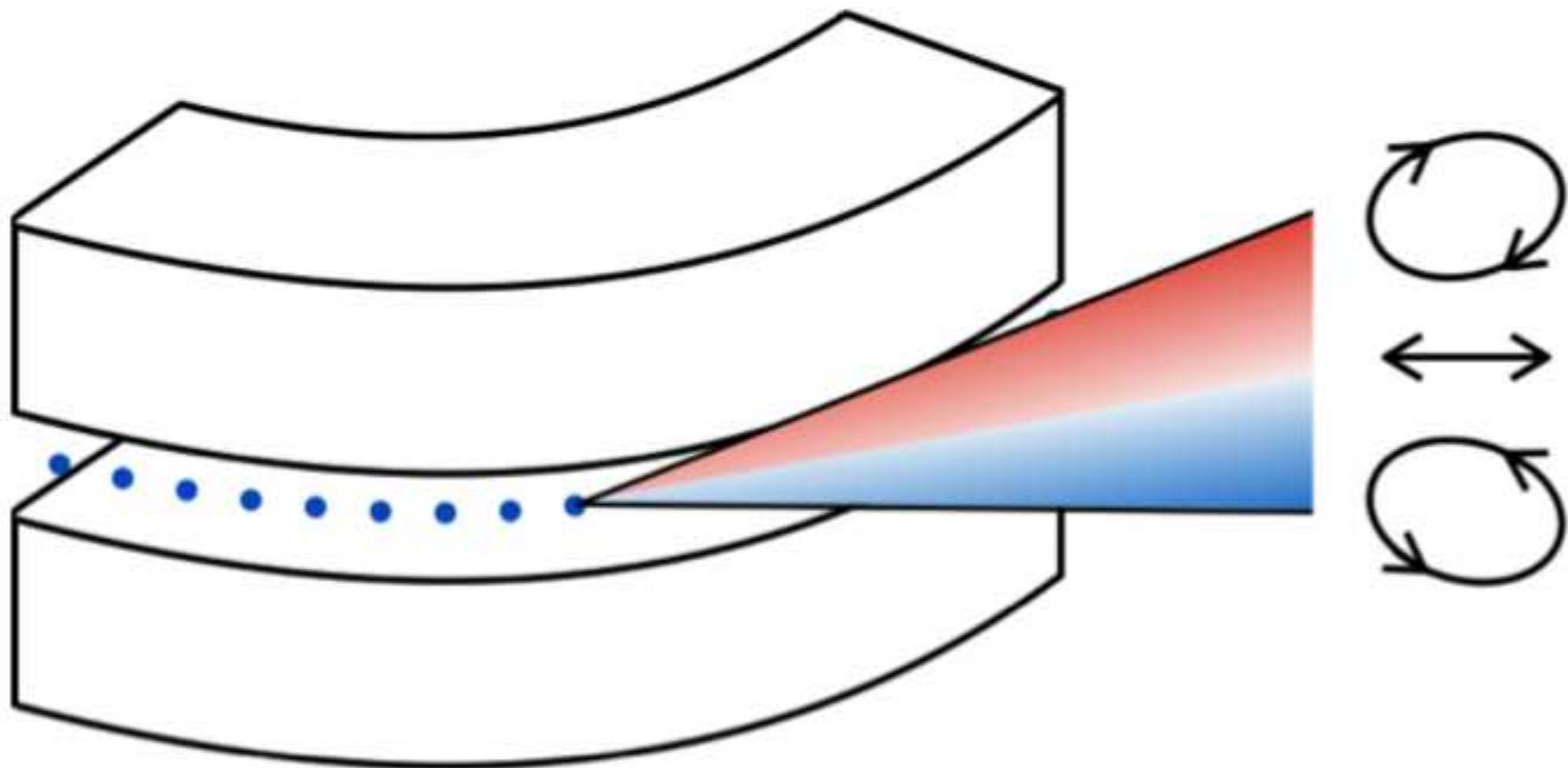
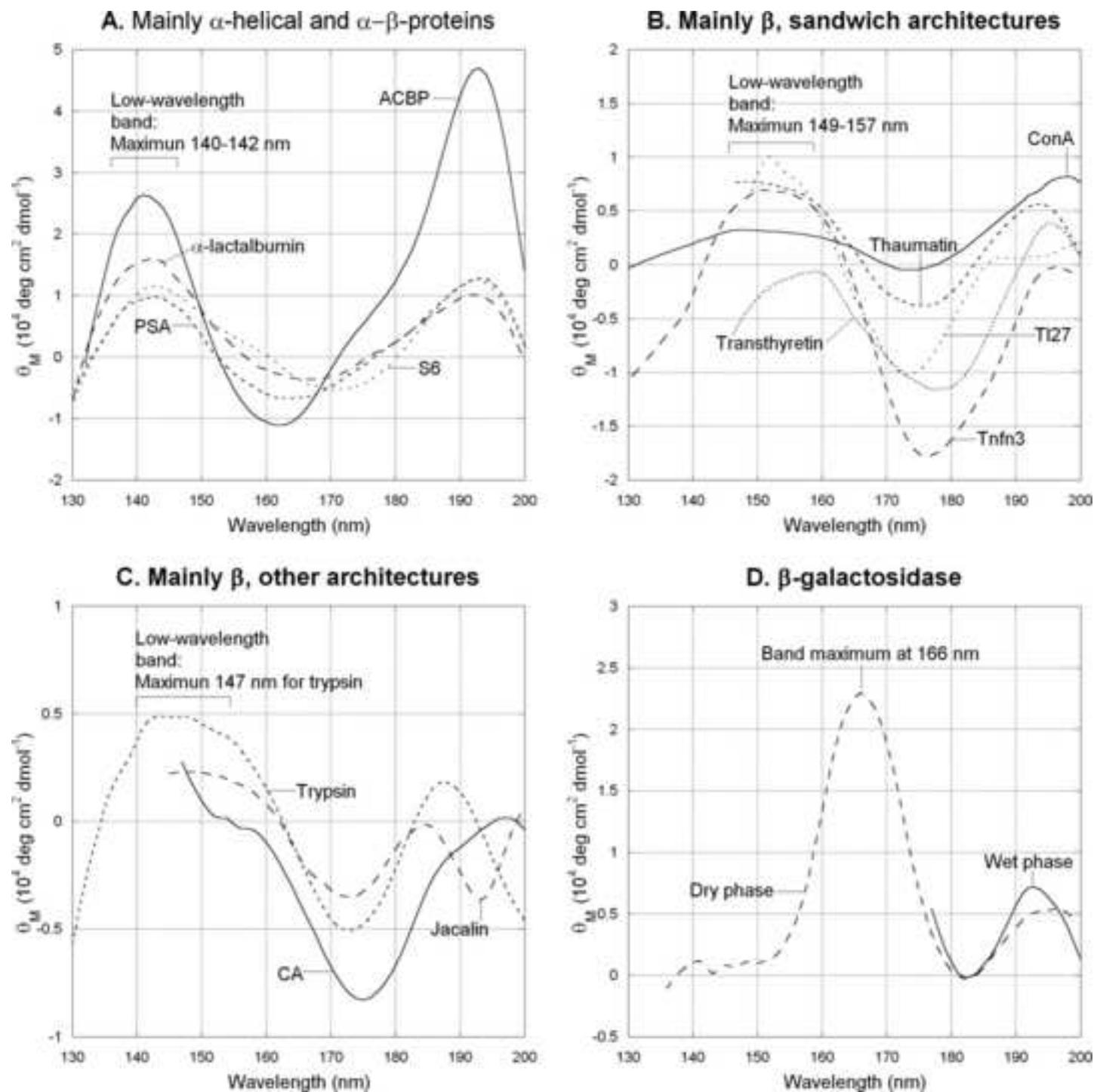


Figure 5

[Click here to access/download;Figure;1pit_frm_3_5.png](#)









[Click here to access/download](#)

ZIP File

[Chem_CD_review_paper_revised.zip](#)

