# Circular and linear dichroism of proteins

**ARTICLE** *in* PHYSICAL CHEMISTRY CHEMICAL PHYSICS · JUNE 2007

Impact Factor: 4.49 · DOI: 10.1039/b615870f · Source: PubMed



**3 AUTHORS:**

Benjamin Bulheller
Novartis Pharmaceuticals, Wehr, Germany
**14** PUBLICATIONS **288** CITATIONS
SEE PROFILE

Alison Rodger
The University of Warwick
**203** PUBLICATIONS **5,108** CITATIONS
SEE PROFILE

Jonathan D Hirst
University of Nottingham
**165** PUBLICATIONS **3,553** CITATIONS
SEE PROFILE

# Circular and linear dichroism of proteins†‡

**Benjamin M. Bulheller,[a] Alison Rodger[b] and Jonathan D. Hirst*[a]**

***Received 1st November 2006, Accepted 26th January 2007***
***First published as an Advance Article on the web 20th February 2007***
**DOI: 10.1039/b615870f**

Circular dichroism (CD) is an important technique in the structural characterisation of proteins, and especially for secondary structure determination. The CD of proteins can be calculated from first principles using the so-called matrix method, with an accuracy which is almost quantitative for helical proteins. Thus, for proteins of unknown structure, CD calculations and experimental data can be used in conjunction to aid structure analysis. Linear dichroism (LD) can be calculated using analogous methodology and has been used to establish the relative orientations of subunits in proteins and protein orientation in an environment such as a membrane. However, simple analysis of LD data is not possible, due to overlapping transitions. So coupling the calculations and experiment is an important strategy. In this paper, the use of LD for the determination of protein orientation and how these data can be interpreted with the aid of calculations, are discussed. We review methods for the calculation of CD spectra, focusing on semiempirical and *ab initio* parameter sets used in the matrix method. Lastly, a new web interface for online CD and LD calculation is presented.

## 1. Introduction

The theory of optical activity has come a long way since the phenomenon was first observed by Arago[1] in 1811. Jean-Baptiste Biot showed that the polarisation plane of light was altered after it passed through a quartz crystal.[2] Three years later, Biot reported similar rotation of the polarisation plane of linearly polarised light in several liquids, including turpentine and solutions of camphor.[3] Louis Pasteur, in 1848, interpreted these observations at the molecular level, at a time when molecules were not yet understood to be three dimensional.[4] Nevertheless, Pasteur showed that tartaric acid exists in two asymmetric forms, which rotate the polarisation plane of light in different directions.[5] In 1874, Le Bel[6] and van't Hoff[7] related rotatory power to the unsymmetrical arrangement of substituents at a saturated carbon atom, thus identifying the very foundation of stereochemistry. Through the definition of chirality, chemistry was given a mighty tool, which was able to explain the properties of sugars and many other organic compounds and which later led to the development of new analytical methodologies, such as optical rotatory dispersion (ORD)[8] and circular dichroism (CD) spectroscopy.[9]

Although the cause of optical activity was known, the development of a theoretical framework to describe and understand the phenomenon proved to be a complex task. The first adequate theory of optical rotatory power was presented by Born in 1915.[10] It was thoroughly investigated by Kuhn[11] and then reformulated by Rosenfeld in 1928,[12] who introduced the eponymous equation for the calculation of the rotational strength of a transition, which is related to its intensity in the CD spectrum.

Due to the sensitivity of CD and ORD to the secondary structure of proteins, the prediction of the optical spectra of polypeptides was attempted. Fitts and Kirkwood[13] calculated the optical rotation of a helical peptide in 1956 using polarisability theory, while Moffitt[14,15] in the 1950's used exciton theory.[16] Moffitt showed that the coupling of electric dipole allowed $\pi \rightarrow \pi^*$ electronic transitions in a helical arrangement leads to a resultant transition that is an in-phase combination, with a net polarisation parallel to the helix axis, and two transitions that are out-of-phase combinations, with a net polarisation perpendicular to the helix axis. Thus, he correctly predicted the right-handed nature of $\alpha$-helices in proteins years before the first X-ray crystallographic structure of a protein. However, this approach was not readily developed into a quantitative method. In 1961, Doty established the dependence of the ORD on $\alpha$-helical content and identified the electronic transitions of the peptide group as the most likely source of the rotatory power of proteins.[17] Later, he confirmed Moffitt's calculations[15] of the exciton splitting experimentally by resolving the three peptide electronic bands and attributing them to the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions, respectively.[18]

Building on the foundation laid by the aforementioned seminal studies, it has become feasible to calculate the CD spectra of molecules and today is, in fact, fairly routine for small molecules, for example, to determine the absolute configurations of isolated or synthesized compounds.[19–23] The computation of protein CD, however, remains a challenge, due to the size and flexibility of the molecules. For the calculation of the optical spectra of large molecules, such as

[a] *School of Chemistry, University of Nottingham, University Park, Nottingham, UK NG7 2RD. E-mail: jonathan.hirst@nottingham.ac.uk; Fax: +44 115 951 3562; Tel: +44 115 951 3478*
[b] *Department of Chemistry, University of Warwick, Coventry, UK CV4 7AL*

† Contract grant sponsor: EPSRC; Contract grant number: GR/T09224.

‡ The HTML version of this article has been additionally enhanced with colour images.

 

proteins and crystals, several methods have been developed specifically to cope with the size of the systems. The dipole interaction model[24–31] considers atoms and chromophores to act as point dipole oscillators, which interact through mutually induced dipole moments in the presence of an electric field. Another approach is the matrix method,[32] which we discuss in detail below. The following article reviews some of the developments in the calculation of optical properties of proteins, including both CD and the related polarised spectroscopy: linear dichroism (LD).

## 2. Circular dichroism

Among the related methods, ORD,[8] LD,[33] and CD,[9,34] the latter is now by far the most popular type of chiroptical or polarised light spectroscopy. The two energy regimes in current use are vibrational transitions[35–39] and electronic transitions. We focus only on the latter in this paper and refer to it by the abbreviation CD. When circularly polarised light impinges on a protein, the protein's electronic structure gives rise to characteristic bands in specific regions in the CD spectrum, reflecting the electronic excitation energies.[40,41] Secondary structural elements, such as α-helices, β-sheets, β-turns and random coil structures, all induce bands of distinctive shapes and magnitudes in the far-ultraviolet (Fig. 1).[42] For example, in an α-helix, an intense positive band at 190 nm and a negative band at 208 nm arise from the exciton splitting of electronic transitions from the amide non-bonding π orbital, $\pi_{nb}$, to the anti-bonding π orbital, π*, (Fig. 2). A negative band is located at about 220 nm, arising from the electronic transition from an oxygen lone pair orbital, n, to the π* orbital (Fig. 2). Other motifs give other spectroscopic shapes and signs.

The relative proportion of each secondary structure type can be determined by decomposing the far-UV spectrum into a sum of fractional multiples of the reference spectra.[9,43,44] For example, the average fractional helicity, $f_H$, of a peptide consisting of n residues can be determined from the observed mean residue molar ellipticity at 220 nm, $[\theta]_{220}$:

$$f_H = \frac{[\theta]_{220}}{[\theta_{H\infty}]_{220}\left(1 - \frac{k}{N}\right)} \qquad (1)$$

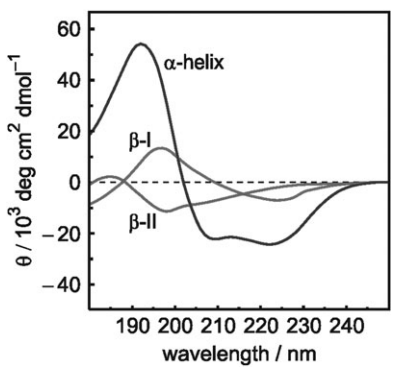


**Fig. 1** Characteristic CD curves of secondary structure elements. The vertical axis shows intensity as the mean residue ellipticity, $\theta$.

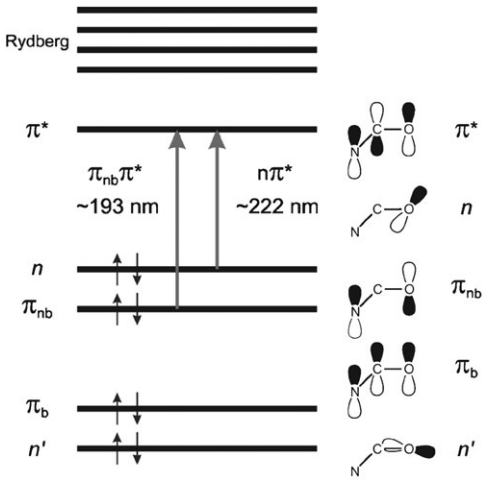


**Fig. 2** Electronic transitions of the amide group in the far-UV region. The molecular orbitals shown are the bonding, nonbonding and antibonding π orbitals ($\pi_b$, $\pi_{nb}$ and π*) and two lone pairs on the oxygen atom (n and n′).

where $[\theta_{H\infty}]_{220}$ is the ellipticity of a completely helical peptide of infinite length and $k$ is an end-effect correction factor of approximately three.[45] Estimates of $[\theta_{H\infty}]_{220}$ range from −37 000 deg cm$^2$ dmol$^{-1}$ [46] to −44 000 deg cm$^2$ dmol$^{-1}$,[47] which corresponds to differential absorbance between −11.2 mol$^{-1}$ dm$^3$ cm$^{-1}$ and −13.3 mol$^{-1}$ dm$^3$ cm$^{-1}$. However, $[\theta]_{220}$ can be influenced by several factors,[48–50] as will be discussed later. The empirical analysis of a spectrum can, therefore, lead to the determination of the protein secondary structure. Moreover, when CD spectroscopy is coupled with time-resolved experiments, protein folding events can be studied[51] and theoretical spectra can be used to interpret the results.[48,49] Although empirical fitting works remarkably well, due to the well-defined nature of the secondary structure motifs of proteins, it is important to proceed beyond empirical data analysis for a number of reasons. Since the conformation of proteins in solution may be (perhaps subtly) different from that in the crystalline environment, calculations can help to uncouple solution phase CD structure analysis from crystallographic structure motifs. Furthermore, the increasing importance of less well-folded or natively unfolded protein domains demands new approaches, since empirical comparison fails in these cases. In contrast to empirical data analysis, theoretical calculations, in principle, can also encompass the population dynamics of a solution of proteins and should be able to cope with new protein folds. The calculation of CD spectra from simulated ensembles of conformations can provide additional information from molecular dynamics (MD) simulations of protein folding.[52] The use of MD simulations, in particular, will be discussed later in more detail, but first we turn to the theory of CD and the associated computational methods.

### 2.1 Theory of CD

Circularly polarised light can be produced by the superposition of two linearly polarised light beams that are oscillating perpendicular to each other and propagating with a phase difference of π/2 radians. The magnitude of the electric field

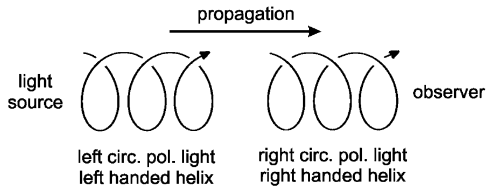


**Fig. 3** Left and right, circularly polarised light, propagating in space.

vector of the resulting beam is constant, but rotates about the propagation direction. If the vector forms a right-handed helix, it is right circularly polarised light and *vice versa* (Fig. 3).[4,53] A stationary observer looking towards the light source at the electric field vector of right circularly polarised light sees it rotating in an anti-clockwise sense, if regarding its progression in space. However, more important is the dependence of the field vector with time, that is the direction of the field at the same position as a function of time. In this case, the vector of right circularly polarised light rotates clockwise (Fig. 4).[8,38]

A solution of chiral molecules possesses different refractive indices for left and right circularly polarised light, that is, the beams travel at different speeds and are absorbed to different extents at each energy. This means that the extinction coefficients for left and right circularly polarised light are different, $\varepsilon_L \neq \varepsilon_R$. This effect is called CD and the differential absorbance, $\Delta\varepsilon$, of left and right circularly polarised light (eqn (2)) is plotted against the wavelength $\lambda$ to yield the CD spectrum.[9,43,54]

$$\Delta\varepsilon = \varepsilon_L - \varepsilon_R \tag{2}$$

The integral of $\Delta\varepsilon$ over a wavelength range associated with a particular transition is known as the CD strength or rotational strength of that transition. It is analogous to the oscillator strength of normal absorption. From the quantum electrodynamic viewpoint, the rotational strength, $R^{0k}$, of a transition from the ground state 0 to an electronically excited state $k$ is the product of the electric transition dipole moment, $\vec{\mu}$, and the magnetic transition dipole moment, $\vec{m}$. The probability of a transition $0 \rightarrow k$ is represented by the integral $\langle\psi^k|\vec{\mu}|\psi^0\rangle$, which can be regarded as an oscillating dipole induced by the incident light beam.[43] $\vec{\mu}$ describes a linear displacement, whereas $\vec{m}$ characterizes a circulation of charge and accordingly the integral $\langle\psi^k|\vec{m}|\psi^0\rangle$ can be understood as a light-induced current loop. Thus, the combination of $\vec{\mu}$ and $\vec{m}$ creates a helical displacement of charge, leading to a different interaction with left and right circularly polarised light. While the operator for $\vec{\mu}$ is a real vector, that for $\vec{m}$ is imaginary, as it describes the rotation of charge in a complex coordinate system,[43] and it involves the linear momentum operator $\vec{p}$:

$$\vec{m} = \frac{e}{2mc}\left(\vec{r} \times \vec{p}\right), \quad \vec{p} = \frac{\hbar}{i}\nabla, \tag{3}$$

where $e$ is the charge, $m$ the mass of the electron, $c$ the speed of light, $\vec{r}$ is the position operator of the electron, $\hbar$ is Planck's constant divided by $2\pi$, $i = \sqrt{-1}$ and $\nabla$ is the gradient operator. The rotational strength is given by the Rosenfeld equation:[12]

$$R^{0k} = \mathrm{Im}(\langle\psi^0|\vec{\mu}|\psi^k\rangle\langle\psi^k|\vec{m}|\psi^0\rangle), \tag{4}$$

where $\psi^0$ and $\psi^k$ denote the wave functions of the ground state and the excited state, respectively, and Im the imaginary part of the product.

Using the Rosenfeld equation, it is possible to obtain the rotational strengths of a molecule and therefore calculate its CD spectrum. However, this requires wave functions for the ground and excited states, which due to computational limitations can only be determined *ab initio* for rather small compounds. Calculations on a protein, comprising hundreds of atoms and thousands of electrons, are thus a challenge, which demands some approximations. In the next section, we describe how these approximate wave functions may be calculated for proteins.

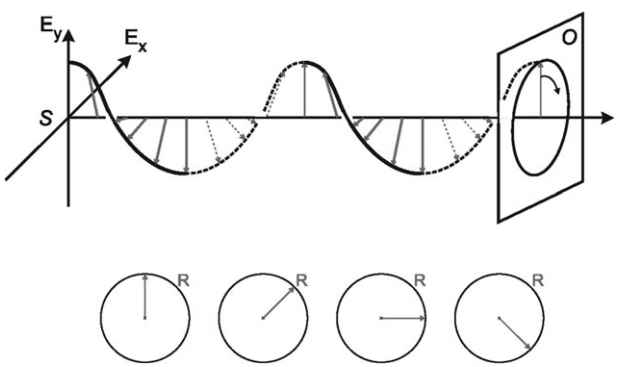


**Fig. 4** Dependence of right circularly polarised light (R) on the distance from the light source, *S*, (top) and on time (bottom) when the observer at position *O* is looking towards the source. The helix moves along the propagation direction without rotation, causing the vector of right circularly polarised light to rotate anticlockwise in space but clockwise in time.

### 2.2 The matrix method

There are several methods to compute the CD spectra of proteins. In 1962, Tinoco adopted a perturbation approach, in which he considered the chromophores of a protein separately[55] and assumed that electrons are localized on a particular chromophore, with Coulombic forces the only means of interaction between different chromophores.[56] The matrix method,[32,57,58] which is derived from the Frenkel exciton model,[16] is commonly used for molecular crystals and chromophore aggregates. It is an improved formulation of the Tinoco[55] method and involves solving the eigenvalue problem *via* a matrix diagonalization rather than by applying perturbation theory and is hence more accurate (especially for degenerate and near-degenerate states) and easily implemented in computer algorithms. In the matrix method, orbitals on different chromophores are assumed not to overlap so that no interchromophore charge transfer occurs. The matrix method is fairly successful in calculating the CD spectra of proteins. Proteins with a high amount of α-helix can be calculated almost quantitatively.[59]

In the matrix method, the protein is split into $M$ independent chromophores, with a monomer wave function $\phi_{is}$ for each chromophoric group $i$ and electronic state $s$. The protein's $kth$ excited state wavefunction, $\psi^k$, is then written (to a

 

first approximation) as a linear combination of electronic configurations, $\Phi_{ia}$, in which only one chromophoric group $i$, is in an excited state $a$ and the others are in the ground state, 0.[54] Thus

$$\Phi_{ia} = \phi_{10} \ldots \phi_{ia} \ldots \phi_{j0} \ldots \phi_{M0} \tag{5}$$

where $\phi_{ia}$ is the wave function of chromophore $i$ after the transition $0 \rightarrow a$, and

$$\psi^k = \sum_{i}^{M} \sum_{a}^{n_i} c_{ia}^k \, \Phi_{ia} \tag{6}$$

The ground state of the protein is, similarly,

$$\psi^0 = \phi_{10} \ldots \phi_{i0} \ldots \phi_{j0} \ldots \phi_{M0} \tag{7}$$

Each transition in the protein CD spectrum is characterized by an energy and a rotational strength, which is given by eqn (4) and is related to the experimental intensity. To calculate the CD spectrum, we therefore need the wave functions $\psi^k$ for each electronic excited state $k$ of the protein (or at least those occurring in the spectral region of interest). The result of such a calculation is simply an intensity and a wavelength of each transition. To produce a spectrum we also need a bandshape and bandwidth for each transition. We return to this non-trivial issue below.

The first summation in eqn (6) is over the $M$ chromophores; the second is over the $n_i$ electronic excitations of each individual chromophore $i$, and $c_{ia}^k$ are (expansion) coefficients, which have to be determined. To illustrate what this means in practice, consider a dipeptide with two excited states on each peptide, the lowest energy backbone n $\rightarrow$ $\pi^*$ and $\pi \rightarrow \pi^*$ transitions. We then write

$$\psi^k = c_{1,\mathrm{n}\pi^*}^k \Phi_{1,\mathrm{n}\pi^*} + c_{1,\pi\pi^*}^k \Phi_{1,\pi\pi^*} + c_{2,\mathrm{n}\pi^*}^k \Phi_{2,\mathrm{n}\pi^*} + c_{2,\pi\pi^*}^k \Phi_{2,\pi\pi^*},$$

$$\text{where} \quad \begin{aligned} \Phi_{1,\mathrm{n}\pi^*} &= \phi_{1,\mathrm{n}\pi^*} \, \phi_{2,0} \\ \Phi_{1,\pi\pi^*} &= \phi_{1,\pi\pi^*} \, \phi_{2,0} \\ \Phi_{2,\mathrm{n}\pi^*} &= \phi_{1,0} \, \phi_{2,\mathrm{n}\pi^*} \\ \Phi_{2,\mathrm{n}\pi^*} &= \phi_{1,0} \, \phi_{2,\pi\pi^*} \end{aligned} \tag{8}$$

For each electronic excited state ($k$ = 1–4) of the four-transition dipeptide, there is a different set of coefficients. The electronic excited states $\psi^k$ of the protein and their corresponding energies can be calculated by solving the Schrödinger equation

$$\hat{H}\psi^k = E^k \psi^k \tag{9}$$

To solve eqn (9) we first need to construct the Hamiltonian, $\hat{H}$, which describes the dynamic properties of the system. If we denote the Hamiltonian of a local chromophore, $i$, as $\hat{H}_i$, then the Hamiltonian of the $M$ independent chromophores is simply the sum of these local chromophore Hamiltonians. When the chromophores are allowed to interact, the Hamiltonian of the protein is the sum of all local Hamiltonians, $\hat{H}_i$, for each chromophore plus the sum of all intergroup potentials, $\hat{V}_{ij}$, of the entire molecule:

$$\hat{H} = \underbrace{\sum_{i=1}^{M} \hat{H}_i}_{\hat{H}_0} + \underbrace{\sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \hat{V}_{ij}}_{\hat{V}} \tag{10}$$

Combining eqns (6), (9) and (10), one gets an equation in terms of chromophore wavefunctions and the coefficients $c_{i,a}^k$. It is convenient to convert this into a matrix formalism.

The matrix form of eqn (10) permits the calculation of the energy (the eigenvalues) and the wave functions (eigenvectors, that is, the coefficients $c_{i,a}^k$) by solving an eigenvalue problem through matrix diagonalization. The Hamiltonian matrix is diagonalised by a unitary transformation using the matrix $U$:

$$U^{-1} \hat{H} U = H_{\mathrm{diag}} \tag{11}$$

The diagonal elements of the resulting diagonal matrix (the eigenvalues) are the transition energies (excited state energies, since we set the ground state to be zero) of the interacting system and the eigenvectors, $c_{i,a}^k$, form the unitary matrix $U$. We then calculate the electric and magnetic transition dipole moments of the excitation from the ground state to the $k$th excited state as follows. The transition dipole moments of the non-interacting chromophores, which we denote $\vec{\mu}_a^0$ and $\vec{m}_a^0$, can be transformed to the interacting system using the unitary matrix from eqn (11):[54]

$$\begin{aligned} \vec{\mu}_i &= \sum_a U_{ai} \vec{\mu}_a^o \\ \vec{m}_i &= \sum_a U_{ai} \vec{m}_a^o \end{aligned} \tag{12}$$

From these transition dipole moments, the rotational strengths in the interacting system are readily calculated. For example, for a dipeptide and two transitions per group, the Hamiltonian matrix constructed using the Hamiltonian of eqn (10) has the form

$$\hat{H} = \begin{pmatrix} E_{1,\mathrm{n}\pi^*} & V_{1\mathrm{n}\pi^*;\,1\pi\pi^*} & V_{1\mathrm{n}\pi^*;\,2\mathrm{n}\pi^*} & V_{1\mathrm{n}\pi^*;\,2\pi\pi^*} \\ V_{1\mathrm{n}\pi^*;\,2\pi\pi^*} & E_{1\pi\pi^*} & V_{1\mathrm{n}\pi^*;\,2\pi\pi^*} & V_{1\pi\pi^*;\,2\pi\pi^*} \\ V_{2\mathrm{n}\pi^*;\,1\mathrm{n}\pi^*} & V_{2\mathrm{n}\pi^*;\,1\pi\pi^*} & E_{2\mathrm{n}\pi^*} & V_{2\mathrm{n}\pi^*;\,2\pi\pi^*} \\ V_{2\mathrm{n}\pi^*;\,1\pi\pi^*} & V_{2\pi\pi^*;\,1\pi\pi^*} & V_{2\mathrm{n}\pi^*;\,2\pi\pi^*} & E_{2\pi\pi^*} \end{pmatrix} \tag{13}$$

The above process is dependent on knowing values for the elements of the Hamiltonian matrix. This in turn requires wavefunctions and intergroup potentials for the independent chromophores. The diagonal elements of the matrix are the transition energies for each transition of each chromophore and the off-diagonal elements, $V_{ij}$, are the interaction energies between different transitions. These interactions are the cause for the dependency of protein CD spectra on secondary (and tertiary) structure. For the transitions $0 \rightarrow a$ on group $i$ and $0 \rightarrow b$ on group $j$, the matrix element, $V_{ij}$, has the form:

$$V_{i0a;\,j0b} = \int_i \int_j \phi_{i0} \, \phi_{ia} \, \hat{V}_{ij} \, \phi_{j0} \, \phi_{jb} \, \mathrm{d}\tau_i \, \mathrm{d}\tau_j \tag{14}$$

If the interaction between the chromophores is regarded as an electrostatic interaction between charge densities $\rho$ of

separation $r$, that is $\hat{V}_{ij} = \frac{1}{4\pi\varepsilon_0 r_{ij}}$, then eqn (14) becomes:

$$V_{i0a;j0b} = \int_{r_i}\int_{r_j} \frac{\rho_{ia0}(r_i)\,\rho_{j0b}(r_j)}{4\pi\,\varepsilon_0\,r_{ij}}\,\mathrm{d}\tau_i\,\mathrm{d}\tau_j \qquad (15)$$

where $\rho_{ia0}(r_i)$ and $\rho_{j0b}(r_j)$ represent the transition electron densities on chromophores $i$ and $j$, $\varepsilon_0$ is the vacuum permittivity and $r_{ij}$ is the distance between the chromophores.

In principle, these matrix elements could be evaluated exactly from the monomer wave functions using the integral evaluation routines in many quantum chemical packages. In practice, an approximation is introduced to make the calculations tractable. In the *monopole–monopole approximation*, the permanent and transition densities are approximated by point charges and the integrals in eqn (15) are re-cast as a sum of the Coulomb interactions of these monopoles:[55]

$$V_{i0a;j0b} = \sum_{s=1}^{N_s}\sum_{t=1}^{N_t} \frac{q_s\,q_t}{r_{st}} \qquad (16)$$

where $q_s$ and $q_t$ are the point charges on chromophores $i$ and $j$, and $N_s$ and $N_t$ are the number of these charges on the chromophore. Thus, at the core of matrix method calculations are the magnitudes and locations of the monopoles.[57,58] They reflect the orientation and magnitudes of the transition moments and they are critical for deducing the inter-chromophore interactions. Since the CD calculation depends only on these monopoles and their distance and orientation to each other, no *ad-hoc* definition of the secondary structure is needed. The amide group is the most important chromophore in protein far-UV CD and has therefore been a key target for research into protein CD.

### 2.3 Calculation of CD

The implementation of the matrix method as a computer program is quite straightforward. The computation requires the coordinates of the atoms and the positions of the chromophores in the protein. Each chromophore is parameterised by a set of monopoles describing the electrostatic potential and for every group in the protein the respective set of monopoles is superposed on the chromophore's atoms. By calculating the interaction between all the different electronic excitations, the Hamiltonian matrix is constructed and, as described above, the rotational strengths are determined, yielding a line spectrum, that is, values for the rotational strength of each transition. However, in an experimental spectrum, the transitions are broadened due to the uncertainty principle, unresolved vibronic components and the interaction of the chromophore with its environment including other chromophores and the solvent. Thus, an overlay of approximately Gaussian shaped bands is observed. Since the result of the calculation is a line spectrum, a convolution with a lineshape function is required. Gaussian line shapes give better results than Lorentzian curves. Hence, all the spectra shown in this paper have been created by taking a Gaussian function of the type[40]

$$y = y_0 e^{-U_0^2} \quad \text{with} \quad U_0 = \frac{\lambda - \lambda_0}{\Delta_0} \qquad (17)$$

The bandwidth or full width at half maximum height (fwhm) of this function is set to a single fixed value of 12.5 nm for all transitions in this work. The resulting calculated spectrum can then be compared to the experimental spectrum.

### 2.4 Comparison of different parametrisations of the amide chromophore

As discussed above, for expedience the Hamiltonian matrix elements may be estimated using the monopole–monopole approximation, which involves representing the transition densities of a chromophore by a set of point charges. Two modern parameter sets for matrix method calculations are an *ab initio* derived set, NMAabinit,[60] and the set of Woody and Sreerama,[61] NMAsemi. They arise from different approaches and are discussed in detail below.

The NMAabinit set has been derived entirely from *ab inito* calculations. Using *ab initio* techniques, the amide chromophore was parametrised with N-methylacetamide (NMA) as a model compound.[60] The electronic spectrum of NMA in solution was calculated, using the complete-active space self-consistent-field method implemented within a self-consistent reaction field (CASSCF/SCRF),[62–65] combined with multi-configurational second-order perturbation theory (CASPT2-RF).[62,65] Monopoles for a given state were determined by fitting their electrostatic potential to reproduce the *ab initio* electrostatic potential for that state so that the least-squares difference is minimised (typically within 5%). The parameter set[60] consists of 32 monopoles for the amide n → π* transition all at a distance of 0.1 Å from the C, N, O and H atoms, as shown in Fig. 5, and 20 monopoles for the π → π* transition,

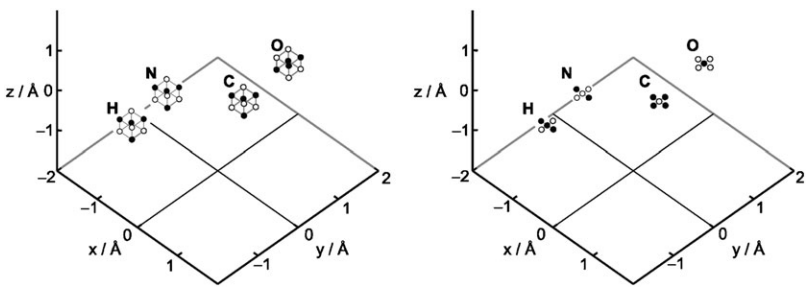


**Fig. 5** Monopole positions of the *ab initio* amide parameter set, NMAabinit, for the n → π* transition (left panel) and the π → π* transition (right panel). Negative charges are represented by black circles and positive charges by white circles. The distance of the monopoles to the atoms is about 0.09 Å in the left panel and either 0 or 0.05 Å in the right panel.

 

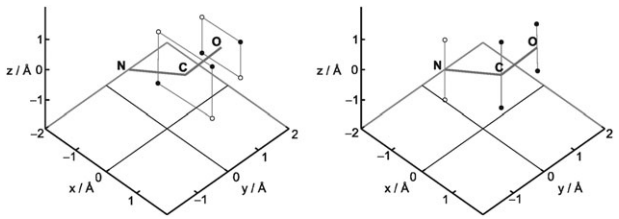


**Fig. 6** Monopole positions of the semiempirical amide parameter set, NMAsemi, for the n → π* transition (left panel) and the π → π* transition (right panel). Negative charges are represented by black circles and positive charges by white circles. The distance of the monopoles to the atoms is between 0.8 and 1.2 Å.

with one charge situated at each atom centre and four around each atom centre at a distance of 0.05 Å (Fig. 5). Two higher energy transitions at 129 and 123 nm were also parametrised. NMAabinit, has been tested on a set of 47 proteins,[66] and the calculated CD agreed well with experiment, particularly at 220 nm, where the accuracy was nearly quantitative. This wavelength is important, since the intensity of the band in that region correlates with the helical content of the protein.[67]

NMAsemi is derived from a combination of semiempirical calculations using the INDO/S method and experimental data. The n → π* transition density is represented by four monopoles surrounding each of the carbon and oxygen atoms, whereas for the π → π* transition density a total of six charges are located above and below the peptide plane (Fig. 6). Compared to the *ab initio* set, it consists of fewer monopoles with smaller charges at greater distance from the atomic centres. However, the electrostatic potential created by the charges is the physically relevant quantity, rather than the particular values and locations of the charges, which are (within certain constraints) arbitrary, as an infinite number of monopole sets give rise to the same electrostatic potential.

For the peptide chromophore, the electrostatic potentials for the two parameter sets are compared in Fig. 7, which shows a contour plot of the electrostatic potential arising from the n → π* transition density 0.5 Å above the plane of the peptide group. In NMAsemi, the monopoles located around the carbon and oxygen produce a symmetric potential around the CO bond. Due to the additional monopoles at the nitrogen and hydrogen atoms, the electrostatic field generated by NMAabinit is asymmetric. The most significant difference

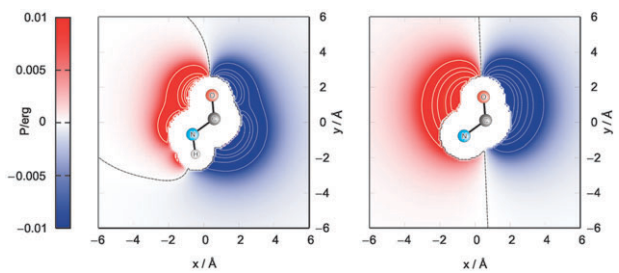


**Fig. 7** Contour plots of the electrostatic potential $P$ of the n → π* transition caused by a peptide chromophore for the *ab initio* parameter set, NMAabinit (left panel) and for the semiempirical parameter set, NMAsemi (right panel). Solid lines denote a positive, dotted lines a negative potential. The contour line of the zero-crossing is dashed.

between the two parameter sets lies indeed with the n → π* transition, as the π → π* transitions are highly dipolar and look very similar in shape and magnitude. However, there is a difference in the orientation of the π → π* transition dipole moment of 8.5°, and the importance of this orientation has been discussed in detail in the literature.[59,61]

### 2.5 Influence of the side chains

In this paper, we concentrate on CD calculations, which consider only the peptide chromophore. However, there are other chromophores of interest. The chromophoric groups of the side chains dominate the CD spectrum in the near-UV (250–350 nm). The intensity in this region is about two orders of magnitude lower than in the far-UV in most proteins, reflecting the fact that only a few percent of the residues in any protein contain aromatic groups. For the same reason, the backbone amide transitions dominate the far-UV in most proteins and we only observe the aromatic groups clearly above 250 nm, where the amide group has no electronic transitions. In the near-UV, as in the far-UV, chromophore parameter sets based on *ab initio* or semiempirical calculations on individual chromophores may be constructed. Our approach has been to gain a more accurate description of the valence electronic transitions of the side chain chromophores by performing *ab initio* calculations.[68] Benzene, phenol and indole are the chromophoric groups of the amino acids phenylalanine, tyrosine and tryptophan, respectively. The influence of solvent on the most important transitions of each chromophore has also been investigated.[68] The excited states are by convention labelled as $^1L_b$, $^1L_a$, $^1B_b$ and $^1B_a$, in which the superscript indicates that the state is a singlet state and the subscript indicates the orientation of the electric transition dipole moment as along either (a) the long or (b) the short axis of the ring in phenylalanine and tyrosine.[69] L refers to the low-lying transitions and B to the transitions at high energy, respectively. In tryptophan the moments are inclined to the long axis of the indole chromophore at about $+54°$ ($^1L_b$), $-41°$ ($^1L_a$), $+18°$ ($^1B_b$) and $-61°$ ($^1B_a$), respectively.[70] The four states of phenylalanine, tyrosine and tryptophan are responsible for the majority of the spectroscopic bands due to side chains in proteins. The $^1B$ states are more intense than the $^1L$ states and are located in the far-UV region. *Ab initio* calculations reproduced the solvatochromatic shifts of the individual chromophores well, although the oscillator strengths were underestimated for benzene and vibronic coupling was neglected. The derived parameter set was tested on the near-UV spectra of 30 proteins and on the difference spectra of 20 mutants and in both cases gave a significant improvement on the semiempirically derived parameters.[71]

### 2.6 Comparing theory and experiment

Fig. 8 (left panel) compares calculated spectra using the NMAabinit and NMAsemi parameters with the experimental spectrum of a helical protein. The peak around 193 nm is well matched by the *ab initio* parameter set, whereas the peak of NMAsemi is slightly shifted to higher energy. In the region between 200 and 230 nm NMAsemi indicates the double minimum which is characteristic for a helical protein, but

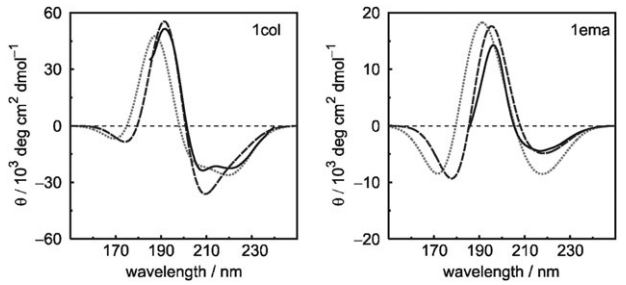


**Fig. 8** Comparison of the experimental spectra of colicin A (1col, mostly α-helical, left panel) and green fluorescent protein (1ema, β-I class, right panel) with the spectra calculated using two different parameter sets. Experimental (solid), *ab initio* parameter set (dashed), semiempirical parameter set (dotted).

which is not resolved for NMAabinit. For the β-sheet protein (Fig. 8, right panel) the intensity at 195 nm is overestimated by both parameter sets, whereas NMAabinit shows a better fit at longer wavelengths, including again a good estimation at 220 nm. In β-II proteins,[72] the strands are rather short or not aligned in a parallel manner, but often twisted and bent. The resulting spectra are similar to unfolded, random coil proteins (Fig. 1) and the agreement of the calculations with experiment is poor. Around 220 nm the intensity is reproduced well, but at higher energies, especially at 200 nm where the minimum is not predicted at all, no correlation between theory and experiment is achieved.

Another structure occurring in peptides is the poly(proline)II-helix ($P_{II}$, $\phi = -75°$, $\psi = +145°$). The $P_{II}$ conformation is the subject of much current research and has been identified as a possible origin for the distinction between β-I and β-II in β-rich proteins. Sreerama and Woody[72] related the observation of different types of spectra for β-I and β-II proteins back to the content of $P_{II}$-conformations, *i.e.* proteins possessing a ratio of $P_{II}$ to β-sheet less than 0.4 appear to belong to the β-II-type, whereas β-II-proteins show a ratio of $P_{II}$ to β-sheet greater than 0.4.[72] The CD spectrum of the $P_{II}$ conformation exhibits an intense negative band at 200 nm, which is not reproduced in calculations.[54] Another influential structural factor is the number of disulfide bridges. The agreement between calculated and experimental spectra deteriorates with more irregular structure and greater numbers of disulfide bonds.

In order to enhance the quality of the calculations of CD spectra, many aspects have been studied, for example, the inclusion of additional chromophores[68] and investigations into hydrogen bonding.[73] Another parameter which has been examined is the width of each band at a specific transition energy. In most calculations, this is set to a single fixed value for all wavelengths. Wavelength-dependent bandwidths have also been tested, but can lead to unexpected results. A comparison of the resulting CD spectra when all transitions were assigned the same bandwidth between 8.5 and 13.5 nm are compared in Fig. 9. To improve resolution of the two peaks around 208 and 220 nm for helical proteins, bandwidths between 7.5 and 15.5 nm have been explored.[66] The minimum of the n → π* transition at 220 nm is resolved for values below 12.5 nm, but such narrow bandwidths then exaggerate the

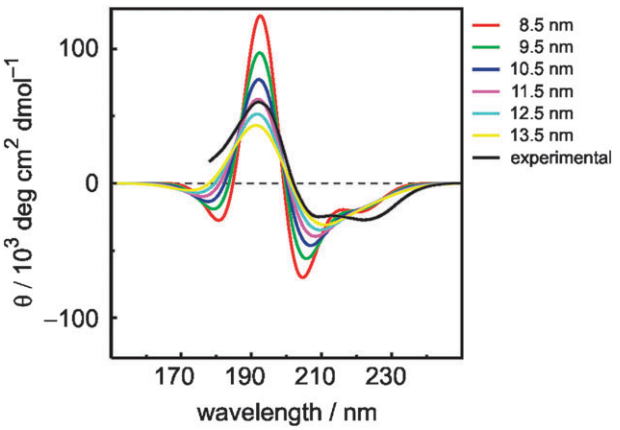


**Fig. 9** Calculated CD spectra where all transitions have been assigned the same bandwidth, with values ranging from 8.5 to 13.5 nm.

intensity maximum of the peak at 190 nm. In reality, different transitions may well have different bandwidths, but, without any justification for a particular choice, there is a danger of introducing a number of empirical parameters which artificially enhance agreement between theory and experiment. In the absence of a rigorous reason for choosing different bandwidths for different transitions, we adopt the best overall value (12.5 nm) in all of our calculations, while recognizing that we are thereby almost certainly worsening the agreement of calculation and experiment.

To assess the two different parameter sets, the Spearman rank correlation[74] between the experimental and calculated intensity has been determined at each wavelength for a set of 46 proteins (Fig. 10). α-Bungarotoxin (2abx) has been removed from the initial set, because an analysis using PROCHECK[75] showed that only 14.8% of the residues lie in the core regions of the Ramachandran plot and 24.6% in disallowed regions, thus questioning the structural integrity of the file. At 190 nm the NMAsemi parameters give a correlation of 0.68, compared to 0.85 of NMAabinit. In this region, the intensity is mostly overestimated by both sets (Fig. 11), but the

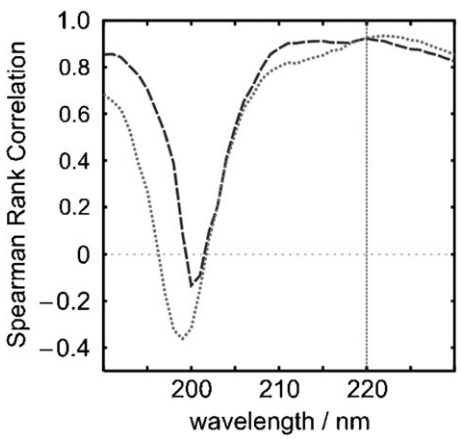


**Fig. 10** Spearman rank correlation between the calculated and experimental intensity at a particular wavelength for a set of 46 proteins for the *ab initio* parameter set (NMAabinit, dashed) and the semiempirical parameter set (NMAsemi, dotted).

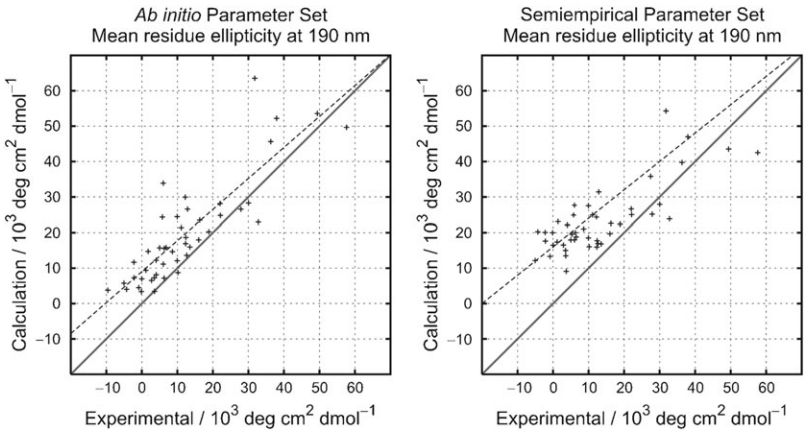


**Fig. 11** Comparison of experimental and calculated mean residue ellipticity at 190 nm. The solid line corresponds to perfect correlation. The dashed line shows the linear regression through all points.

overall correlation is positive. Notably, bacteriorhodopsin (2brd) is an outlier, despite being mostly α-helical. It is a membrane protein with a high helical content. Its CD spectrum is less intense than the spectra of globular proteins with similar helical content and, thus, the intensity of the spectra is overestimated by both parameter sets.

The point where the CD intensity changes sign, at around 200 nm, is a challenge for both parameter sets. In this region the gradient of the spectrum is greatest, causing a high difference in intensity for even small deviations of the calculated *versus* the experimental wavelength. Although the absolute error in the wavelength where the intensity is calculated to be zero is below 3 nm (Table 1), the intensities may differ up to 10 000 deg cm$^2$ dmol$^{-1}$. Hence, the correlation is almost zero for NMAabinit and is actually negative for NMAsemi. NMAabinit tends to overestimate the intensity, *i.e.* the calculated wavelength of the zero point is too large and the spectrum slightly red-shifted, whereas the deviations of NMAsemi are equally distributed. The correlation steadily increases after the zero-crossing and both parameter sets reach almost constant values of between 0.8 and 0.9 above 210 nm. The band at 208 nm, arising from the exciton splitting of the $\pi \rightarrow \pi^*$ transition, shows a correlation of 0.80 and 0.75 for NMAabinit and NMAsemi, respectively. At 220 nm, which is important for the determination of helical content, both sets give a comparably (high) correlation (Fig. 12).

### 2.7 Applications and other developments

In protein folding the transient and short-lived nature of intermediates makes the direct observation of folding and unfolding exceptionally challenging for techniques such as NMR and X-ray crystallography. However, nanosecond CD spectroscopy is a promising candidate for monitoring the conformational changes, which motivates the combination of these first-principles calculations with MD simulations to study the influence of dynamics on the CD of polypeptides.[52] MD simulations and CD calculations have been performed on concanavalin A, a β-I- and elastase, a β-II-type protein, which possess similar structures but show different CD spectra.[52] In the former case, the spectrum did not change substantially and remained close to that calculated for the X-ray crystal structure. For the β-II protein, the intensity of the calculated positive peak at 195 nm (which is negative in the experiment) decreased considerably, suggesting that relaxation of the protein structure in solution may be a contributing factor to the differences between β-I and β-II proteins. The well ordered, aligned sheets in β-I proteins and also the helices in α-helical proteins induce a relatively rigid structure supported by main-chain hydrogen bonds. This framework cannot be formed to the same extend in less ordered peptides such as β-II-type proteins, making them more flexible.

The effect of hydrogen bonding on CD is of interest, and this was highlighted by the unexpectedly high negative intensities at 220 nm reported for several peptides, which exceed even the highest estimate of $[\theta_{H\infty}]_{220}$, the intensity expected for a 100% infinitely long helix.[45] Calculations suggested that the interaction with highly polar solvent molecules, which bind selectively to the helical conformation of peptides, may strengthen the backbone hydrogen bonds of the peptide. The intensity at 220 nm is sensitive to the main-chain hydrogen bond length[49] and the stabilization (and shortening) of hydrogen bonds could thus cause an over-estimation of the helicity in the presence of polar solvents.

Peptides are often studied as model systems for probing interactions that are important in protein stability and folding. Many aspects of the stability of α-helices have been investigated, including the interactions between aromatic and basic side chains.[48] Often a tyrosine residue is introduced into the peptide to facilitate concentration measurements. Calculations of protein CD from first principles indicate that the presence of tyrosine can have a direct effect on the CD and can change the intensity at 222 nm up to ±5000 deg cm$^2$ dmol$^{-1}$, depending on the orientation of the aromatic ring.[48] Further investigation into the influence of tyrosine has shown that the intensity at 220 nm for tyrosine-containing peptides will lead to underestimates of the helical content of the peptide of between 5 and 20%.[50]

The deep-UV is a region of the spectrum of growing interest for CD. The use of synchrotron radiation has made it possible to extend the wavelength range over which data can be measured into the vacuum-UV[76] down to 160 nm to uncover further features of the protein CD spectra.[77–82] The origin of a

**Table 1** Calculated (using NMAabinit parameters) and experimental wavelengths of interest in protein CD. The 13 proteins at the top show the least agreement between the calculated and experimental spectra; β-II proteins are printed in bold

| | | $\lambda_{max}$/nm | | $\lambda_{[\Delta\varepsilon=0]}$/nm | | $\lambda_{min}$/nm | |
|---|---|---|---|---|---|---|---|
| Protein name | Protein PDB code | Calc | Exp | Calc | Exp | Calc | Exp |
| **Carbonic anhydrase** | **1ca2** | 195 | — | 207 | — | 218 | — |
| Bence-Jones immunoglobulin | 1rei | 199 | 202 | 215 | 210 | 223 | 217 |
| Bacteriorhodopsin | 2brd | 191 | 196 | 201 | 205 | 209 | — |
| **Chymotrypsinogen A** | **2cga** | 196 | — | 209 | — | 219 | — |
| Ribonuclease S | 2rns | 195 | — | 206 | — | 217 | — |
| **Cu,Zn Superoxide dismutase** | **2sod** | 199 | 190 | 212 | 196 | 222 | 209 |
| Erabutoxin B | 3ebx | 199 | 201 | 217 | 209 | 223 | 215 |
| **Native elastase** | **3est** | 197 | — | 211 | — | 221 | — |
| **Trypsin** | **3ptn** | 198 | — | 212 | — | 222 | — |
| Gamma-B Crystallin | 4gcr | 197 | 203 | 211 | — | 221 | — |
| **Alpha Chymotrypsin A** | **5cha** | 197 | — | 211 | — | 220 | — |
| **Trypsin inhibitor** | **5pti** | 193 | — | 203 | — | 211 | — |
| Beta-Lactoglobulin | 1beb | 194 | 190 | 205 | 201 | 216 | 213 |
| Colicin A | 1col | 192 | 192 | 201 | 201 | 210 | 208 |
| Azurin | 1e5z | 196 | 195 | 209 | 205 | 219 | 219 |
| Green fluorescent protein | 1ema | 195 | 195 | 208 | 206 | 218 | 217 |
| Eco Ri Endonuclease | 1eri | 192 | 190 | 202 | 201 | 211 | 208 |
| Flavodoxin | 1fx1 | 193 | 194 | 204 | 203 | 214 | 214 |
| Intestinal fatty acid binding protein | 1ifc | 195 | 197 | 207 | 204 | 217 | 216 |
| Lysozyme | 1lys | 192 | 192 | 202 | 199 | 211 | 208 |
| Plastocyanin | 1plc | 198 | 200 | 213 | — | 222 | — |
| Rhodanese | 1rhd | 194 | 193 | 204 | 201 | 213 | 210 |
| Subtilisin | 1sbt | 192 | 192 | 202 | 200 | 212 | 218 |
| Cachectin | 1tnf | 197 | 201 | 211 | 209 | 221 | 221 |
| Concanavalin A (native) | 2ctv | 198 | 197 | 212 | 209 | 221 | 224 |
| Hemerythrin (adizomet) | 2hmz | 192 | 194 | 201 | 202 | 210 | 210 |
| Hemoglobin (horse) | 2mhb | 191 | 195 | 201 | 202 | 210 | 210 |
| Prealbumin (human plasma) | 2pab | 197 | 195 | 210 | 202 | 221 | 217 |
| Pepsinogen | 2psg | 195 | 193 | 207 | 200 | 218 | 211 |
| Staphylococcal nuclease | 2sns | 193 | — | 203 | 199 | 211 | 209 |
| Adenylate kinase | 3adk | 193 | 192 | 203 | 202 | 211 | 210 |
| D-Glyceraldehyde-3-phosphate dehydrogenase | 3gpd | 192 | 192 | 201 | 200 | 211 | 212 |
| Glutathione reductase | 3grs | 194 | 191 | 204 | 199 | 212 | 207 |
| Phosphoglycerate kinase | 3pgk | 192 | 191 | 202 | 200 | 211 | 209 |
| Porin | 3por | 194 | 195 | 205 | 208 | 214 | 218 |
| Ribonuclease A | 3rn3 | 194 | 193 | 205 | 199 | 215 | 209 |
| Triosephosphate isomerase | 3tim | 192 | 193 | 202 | 202 | 211 | 219 |
| Insulin | 4ins | 190 | 195 | 199 | 202 | 208 | 209 |
| Myoglobin | 4mbn | 192 | 192 | 201 | 201 | 210 | 209 |
| Apo-Liver alcohol dehydrogenase | 5adh | 194 | 196 | 205 | 203 | 217 | 211 |
| Carboxypeptidase | 5cpa | 193 | 198 | 203 | 204 | 212 | 211 |
| Calcium-binding parvalbumin B | 5cpv | 191 | 195 | 200 | 202 | 209 | 209 |
| Cytochrome C | 5cyt | 194 | 195 | 205 | 202 | 214 | 210 |
| Apo-Lactate dehydrogenase | 6ldh | 193 | 192 | 203 | 202 | 212 | 212 |
| Thermolysin | 8tln | 193 | 194 | 203 | 203 | 211 | 211 |
| Papain | 9pap | 192 | 188 | 201 | 194 | 211 | 209 |

band around 165 nm has prompted some speculation. At first a transition from the higher-energy lone-pair orbital on oxygen to an antibonding σ orbital ($n \rightarrow \sigma^*$) was suggested.[83] This was discounted by *ab initio* calculations, which showed that the σ* orbital lies at a much higher energy.[84] The involvement of an excitation from the bonding $\pi_b$ orbital to π* ($\pi_b \rightarrow \pi^*$) was suggested,[85] but this transition occurs around 140 nm.[84] An excitation from the second lone pair on oxygen to the anti-bonding π orbital ($n' \rightarrow \pi^*$) has also been considered. Alternatively, there is experimental and theoretical evidence indicating that charge-transfer transitions between neighbouring amide groups are the cause of the band at 165 nm.[86–88]

To investigate this, we have performed CASSCF/CASPT2 calculations on 13 peptide dimers[41,82,89] with dihedral angles from all accessible areas of the Ramachandran plot. The charge-transfer between neighbouring groups had a considerable impact on the calculated spectrum in a manner dependent on the conformation, that is the dihedral angles of the two amide groups. For a matrix method calculation this means that each conformation has to be parametrised. The Hamiltonian matrix needs to be extended by the addition of a charge-transfer chromophore, which spans two neighbouring monomers in the chain.

Another area of investigation is the inclusion of higher energy transitions, since interaction with these could affect the intensities and positions of the bands in the far-UV. The $\pi_b \rightarrow \pi^*$ (bonding π orbital to π*) and $n' \rightarrow \pi^*$ (second lone pair on oxygen to π*) transitions occur around 140 nm and are thus outside the measured region of 180–250 nm. Nevertheless, they

 

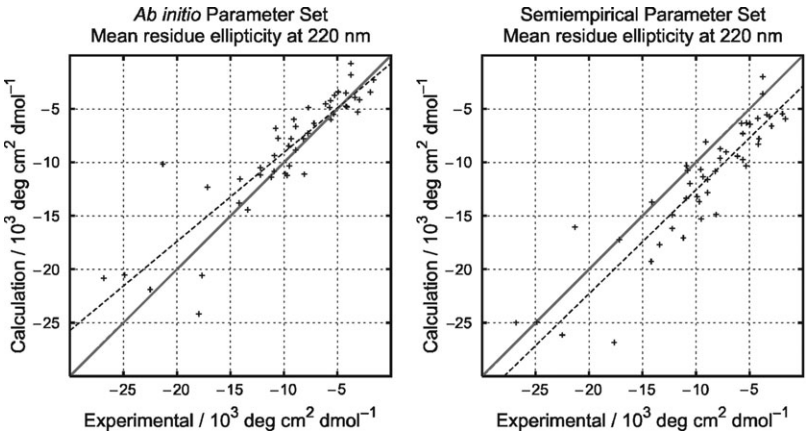


**Fig. 12** Comparison of experimental and calculated mean residue ellipticity at 220 nm. The solid line corresponds to perfect correlation. The dashed line shows the linear regression through all points.

could couple with the n → π* and π → π* transitions and have an effect on the observed far-UV. However, including the two higher energy transitions decreases the intensity of the peak at 193 nm and shifts it to longer wavelengths. The calculations show a negative peak at 182 nm (which is not apparent in experimental spectra) and the inclusion of the two additional transitions even increased its intensity. The bands around 208 nm and 220 nm are practically unaffected and the correlation did not improve in this region. Thus, the consideration of the two higher energy transitions did not lead to an improvement of the agreement.[66]

## 3. Linear dichroism

### 3.1 Introduction to LD

Linear dichroism (LD) is also a differential absorption technique, but in this case it is the difference in absorption of light linearly polarised parallel and perpendicular to an *orientation* axis (Fig. 13):[90]

$$\mathrm{LD} = A_{\parallel} - A_{\perp} \tag{18}$$

LD is related to CD in that both require the difference between the absorbances of different polarised light beams to be measured and CD spectropolarimeters can be adapted to produce the required alternating beams of linearly polarised light for LD. However, LD measurements are performed on systems that are either intrinsically oriented or are oriented during the experiment.

In order to understand LD, one has to consider what happens when a molecule absorbs a photon: absorption can be pictorially viewed as either the electric field or the magnetic field (or both) of the radiation pushing the initial electron density to a final arrangement of higher energy. The direction of net linear displacement of charge is known as the *polarisation of the transition*. The polarisation and intensity of a transition are dependent on the electric transition dipole moment $\langle\psi^0|\vec{\mu}|\psi^k\rangle$, which may be regarded as the antenna by which the molecule absorbs light. Each transition thus has its own antenna and the maximum probability of absorbing light is obtained when the antenna and the electric field of the light are parallel. Now imagine we have a linearly polarised light beam (Fig. 13) and a sample of molecules all oriented in exactly the same way. If we first measure the normal absorption with the light polarised so that it is parallel to the direction of orientation of the sample and then measure it when the light is polarised perpendicular to this direction, the difference of these two spectra is the LD spectrum. The two extreme situations we might encounter are:

(1) If the polarisation of the transition we are probing is perfectly *parallel* to the orientation direction, then

$$\mathrm{LD} = A_{\parallel} - A_{\perp} = A_{\parallel} > 0 \tag{19}$$

(2) If the polarisation of the transition we are probing is *perpendicular* to the orientation direction, then

$$\mathrm{LD} = A_{\parallel} - A_{\perp} = -A_{\perp} < 0 \tag{20}$$

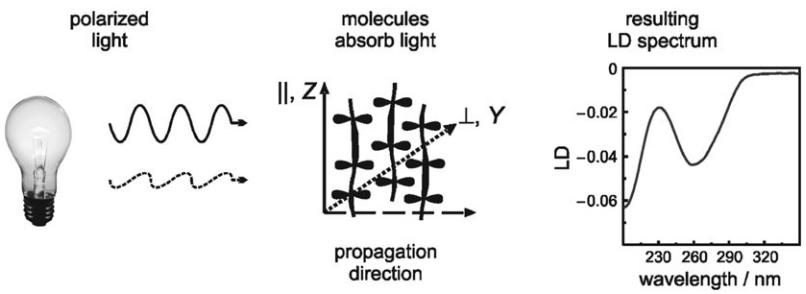


**Fig. 13** Schematic illustration of an LD experiment.

For intermediate polarisations, the LD is between these cases and for a perfectly oriented system:[90]

$$LD^r = \frac{LD}{A} = \frac{A_{\parallel} - A_{\perp}}{A} = \frac{3}{2}(3\cos^2\alpha - 1) \qquad (21)$$

where the reduced LD, $LD^r$, is defined to be the ratio of the LD over the isotropic absorbance, $A$. ($LD^r$ is independent of sample pathlength and concentration) and $\alpha$ is the angle between the transition moment and the orientation axis. Thus, we can determine the polarisation of a given transition from its LD spectrum if we know how the molecule is oriented; conversely, we can use LD as a probe of molecular orientation if we know the polarisation of a transition moment within the molecule.

### 3.2 Molecular alignment techniques

In addition to the requirement of linearly polarised light for LD, LD experiments are technically demanding because of the need to orient the sample. In order to appreciate how we can use LD to study proteins, it is important to understand what data can be collected. This depends on how samples can be aligned. Proteins may be oriented by a number of methods depending on their nature. Long polymers, such as fibrous proteins, may be oriented by the viscous drag caused when a solution is flowed between narrow walls.[91–93] Depending on the cell design, the light is then propagated either along the flow direction or perpendicular to it. This technique is commonly used for LD studies of DNA. In the 1960's and 70's it was used for a few fibrous protein studies.[94,95] More recently we have been developing it for a wide range of fibrous proteins including actin, tubulin, FtsZ, collagen and amyloid fibres.[96–99] The most successful flow cell has proved to be a cylindrical Couette flow cell, where the solution containing the long molecules is subjected to a constant gradient over the annular gap between two coaxial cylinders one of which is rotating. The advantage of the Couette cell over a flow through system is that the same volume of sample can be measured indefinitely. For proteins a significant step forward has been the recent development of micro-volume cells which only require 30 μL of sample.[100]

Couette flow orientation of proteins has been extended to membrane proteins and peptides following Ardhammar, Mikati and Nordén discovering[101] that spherical liposomes were distorted in a Couette cell and could be used to orient small molecules bound to the liposomes. We later showed that any protein or peptide that bound to a liposome could be similarly oriented.[33,102] Due to the defined symmetry of the liposome, for such a system eqn (21) is replaced by:

$$LD^r = \frac{3S}{4}(1 - 3\cos^2\beta) \qquad (22)$$

where $\beta$ is the angle the transition moment of interest makes with the normal to the liposome surface (*i.e.* parallel to the *lipid* long axis) and $S$ is the orientation factor that denotes the fraction of the liposome that is oriented as a cylinder perfectly parallel to the flow orientation.

### 3.3 Examples of protein LD data

Fig. 14 shows the LD spectrum one might expect to measure in the absence of light scattering artefacts for an α-helix oriented parallel to the sample orientation direction. This might, for example, be on the surface of a liposome or a coiled-coil α-helical fibre. The 220 nm region is dominated by transitions polarised perpendicular to the helix and therefore shows a negative signal; the 208 nm region is polarised parallel to the helix and so it has a positive signal. However, it is sandwiched between two negative signals so it may or may not actually end up looking like a positive maximum as the signals overlay and cancel. This is an illustration of why we need to have electronic level calculations of LD spectra, as discussed below, in order to deconvolute spectra in structural terms.

Fig. 15 shows a near- and far-UV LD spectrum of actin over a range of concentrations.[93] In the 12 μM spectrum one can see the components of the $\pi \rightarrow \pi^*$ transition and the $n \rightarrow \pi^*$ transition of the peptide backbone, comparable to the schematic of Fig. 14, but of opposite sign. The lower energy (higher wavelength) component of the $\pi \rightarrow \pi^*$ band is evident as the 215 nm dip in the spectrum. The LD spectrum therefore indicates that, on average, the α-helices in the actin fibre are oriented more perpendicular than parallel to the fibre axis. This is in accord with literature data for actin. However, if we did not have literature data it is almost certain we would want a less qualitative description of the structure, which is simply not possible without considering all the transitions in the protein as discussed below.

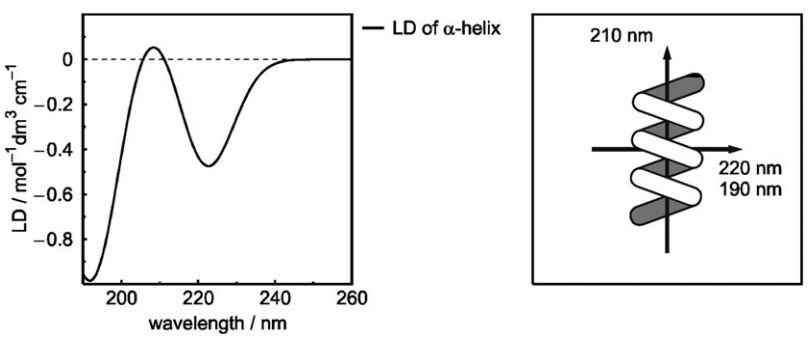


**Fig. 14** A schematic illustration of the LD spectrum one might expect for an α-helix oriented parallel to the sample orientation direction. α-helices have an $n \rightarrow \pi^*$ transition polarised perpendicular to the helix axis at 220 nm;[33] a component of the $\pi \rightarrow \pi^*$ transition parallel to the helix at ~210 nm and a component of the $\pi \rightarrow \pi^*$ transition perpendicular the helix at ~190 nm.[33,97]

 

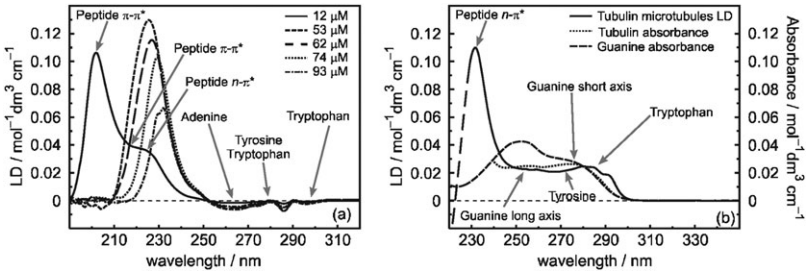


**Fig. 15** (a) Near and far-UV LD spectra of F-actin in a Couette flow cell showing the apparent shift to shorter wavelength of the maximum signal as the concentration of F-actin is reduced. The 12 μM spectrum (solid line) is 'true'. (b) Near-UV absorbance and LD data for tubulin (microtubules include tubulin and guanine).

It should be noted that care must be taken in collecting low wavelength data for a system such as actin where light scattering is significant, as illustrated in Fig. 15. In that case more concentrated samples have an apparent wavelength shift of the maximum LD signal of ~30 nm to longer wavelengths. When the dilution effect is simply to reduce the signal intensity according to the Beer Lambert law, then we can conclude the spectrum is real and one might expect a calculation to be able to give the same result if the geometry is correct. Matrix method calculations (see below) will never reproduce artefacts in spectra.

Fig. 15b shows the near-UV LD and absorbance spectra of tubulin.[99] Due to the different wavelengths of guanine, tryptophan and tyrosine transitions, comparisons of calculated spectra with experimental data should enable one to determine, for example, the orientation of the guanine (which is essential for polymerisation) at all stages of tubulin polymerisation/depolymerisation. Thus, calculations on tubulin (and indeed the near-UV region of actin) will require accurate parametrisation of the NTP/NDP involved in the polymerisation as well as the protein chromophores. This is true for any additional chromophore in any protein, both in CD (as discussed above) and LD.

In the case of bacteriorhodopsin (BR), the retinal chromophore is a very useful probe of extent of protein orientation. The absorption, CD and LD spectra of BR in liposomes are given in Fig. 16.[103] The absorption spectra (Fig. 16a) are consistent with those reported in the literature for the initial state of BR.[104] The absorption maximum at ~570 nm is due to a long axis polarised transition of the retinal chromophore. The broad peak in the near-UV region (260–290 nm) is due to the transitions of the protein aromatic side chains phenylalanine, tyrosine and tryptophan and the peak observed in the far-UV region (215 –~230 nm) is due to the peptide n → π* transition of the amide groups. The backbone CD spectrum of BR (Fig. 16b) is in accord with the CD spectrum for a highly α-helical protein, although the intensity is weaker than for globular proteins as discussed above.[105,106] The ~570 nm LD signal is consistent with a retinal chromophore oriented at ~70° to the lipids of a lipid bilayer if $S$ ~0.05. The aromatic region (260–280 nm) of the LD spectrum is dominated by the indole chromophore of the tryptophan residues.[107] Contributions from $L_a$ (270 nm) and $L_b$ (287 nm) transitions are apparent, both showing positive LD and suggesting the average tryptophan is tilted so that the normal to the plane of the indole is ~40° from the average lipid and the long and short axes are both at an angle of 60–65° to the lipids. This is consistent with the fact that the retinal is sandwiched by tryptophan residues in the X-ray structure.[105]

The protein backbone LD spectrum shows a positive maximum at 220 nm (n → π*) and a negative maximum at ~213 nm (π → π*). It follows that the n → π* transition (which is polarised perpendicular to the α-helix long axis) is at ~58° ($LD^r_{220\ nm}$ = 0.006) from the average lipid direction. The success of this analysis is dependent on having a value for $S$ (in this case from the retinal chromophore) and the protein being almost totally helical. With more varied protein structures, calculations to determine the net protein LD will be essential.

### 3.4 Protein LD calculations

It is apparent from the above discussion that LD data are consistent with proteins of known structure and a qualitative level of structural interpretation is possible. This can lead to mechanistic insights not possible from other techniques as in the case of FtsZ.[98] The level of interpretation that is possible, however, is restricted due to the cancellation of LD signals by overlapping bands as illustrated in Fig. 14. To get beyond this we need to know what each transition is doing and to be able to calculate spectra as a function of geometry until a good match between experiment and calculation is obtained.

If one is able to calculate the CD spectrum (see above), values for the components of the electric dipole transition moments are generated in the process. In the case of a single, perfectly oriented molecule we may write, using the notation of section 2.2

$$\begin{aligned} \mathrm{LD}_{\mathrm{molecule}} &= (A_{\parallel} - A_{\perp})_{\mathrm{molecule}} \\ &\propto \left\{ |\langle \psi^k | \mu_Z | \psi^0 \rangle|^2 - |\langle \psi^k | \mu_Y | \psi^0 \rangle|^2 \right\} \end{aligned} \qquad (23)$$

where $Z$ is the orientation axis and $Y$ is perpendicular to $Z$. For the case of a perfectly oriented molecule $Z$ is usually the longest axis of the molecule (denoted $z$), and $Y$ is a rotational average over the molecular axes $x$ and $y$. So the LD can be calculated using the same matrix methodology as used for CD, with an expectation of a more accurate match to experiment as LD depends only on the electric moments.

Unfortunately, samples are seldom if ever perfectly oriented and then we need to determine how to relate the components of the electric dipole transition moments calculated in the molecular axis system to those in the macroscopic system.

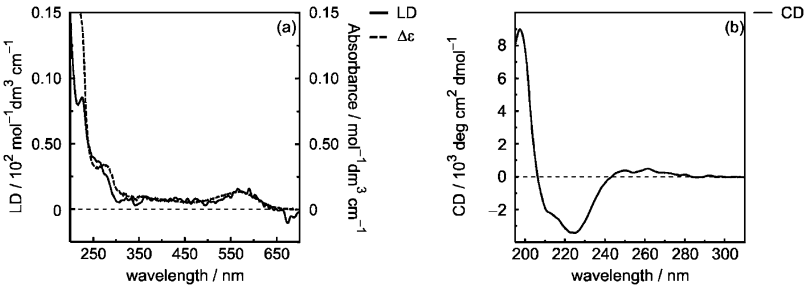


**Fig. 16** Spectra of bacteriorhodopsin (0.2 mg/mL) added to a soya bean liposome solution (0.5 mg/mL). (a) Absorption (dashed line, 1 mm pathlength, baseline: liposome absorption spectrum) and LD (solid line, 0.5 mm pathlength, baseline: LD spectrum of sample without rotation spectrum); (b) CD (1 mm pathlength, baseline: liposome CD spectrum).

Further, we need to consider an ensemble of molecules with their $z$ axes distributed about the macroscopic orientation axis $Z$. In many cases, either the averaging inherent in the orientation method or the nature of the sample means that we can assume local uniaxial orientation (Fig. 17) in which case:

$$\mathrm{LD} = \frac{\mu^2}{2} S(3\cos^2\alpha - 1) \qquad (24)$$

or

$$\mathrm{LD}^r = \frac{\mathrm{LD}}{A} = \frac{A_\parallel - A_\perp}{A} = \frac{3}{2} S(3\cos^2\alpha - 1) \qquad (25)$$

where $S$ may be viewed as a scaling factor defining the efficiency of macroscopic orientation. $A_\parallel$, $A_\perp$, and $A$ all have $C\ell$ as a common factor. Hence, when using $\mathrm{LD}^r$ we do not need to know the sample concentration $C$ or pathlength $\ell$. Thus, $S = 1$ for perfect orientation (assumed in eqn (19)) and $S = 0$ for random (*i.e.* no) orientation. If the sample is macroscopically as well as molecularly uniaxial, such as molecules oriented in a polymer film drawn in one direction or polar molecules in an electric field, there is a simple relation between $A_\parallel$, $A_\perp$, and $A$ which makes it unnecessary to measure all three quantities (Fig. 17).

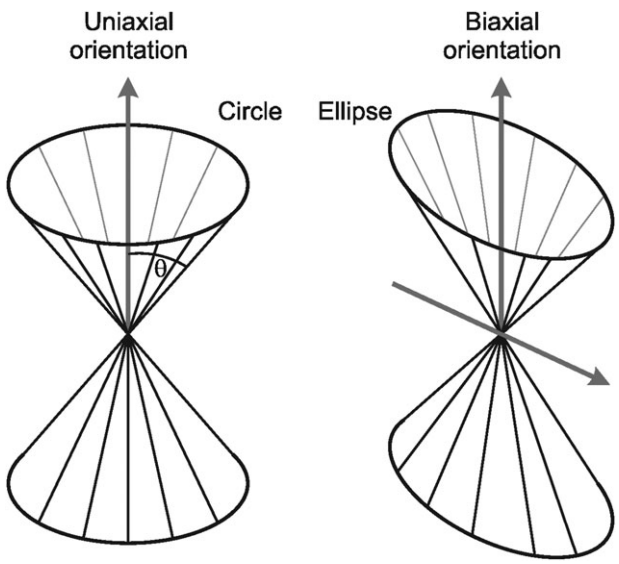


**Fig. 17** Uniaxial and biaxial orientation. Uniaxial orientation requires all orientations on a cone about the orientation axis to be statistically equally probable. This may be due to static or dynamic factors.[90]

The challenge from a computational point of view is to integrate calculations of LD spectra for monomeric proteins of known structure with proposed polymerisation geometries and orientation distributions to enable LD experiment plus calculations to be used to determine fibre geometries. This may well involve molecular dynamics simulations to enable an estimate of $S$. To date LD calculations using the matrix method have been limited to calculating the expected LD for secondary structural motifs or chromophores.[90,102] The next stage is to tackle monomeric units of fibrous proteins, oriented as in a fibre or membrane proteins in liposomes, whose actual orientation cannot be easily determined in experiment. In general, the orientation dependence of an LD experiment is an additional parameter which needs to be taken into account in the calculation. Since even the best aligned liquid sample will never be perfectly oriented in reality (as this would be the case in a crystal), these deviations need to be considered in the calculation to get a more accurate simulation. In practice, this is done by calculating the perfectly aligned protein, including rotational averaging about its long axis. It is then tilted about a small angle and the LD spectra calculated. Comparison of spectra computed for different inclination angles with the experimental spectrum should allow the assignment of the most likely orientation of the protein. This methodology is currently being applied to several problems.

## 4. Online calculation using DichroCalc

The extensive use of the internet nowadays has prompted a large number of online services for protein analyses, *e.g.* for the interpretation of mass spectra, investigations into protein similarity and the prediction of secondary structure. Specifically concerning CD, the Protein Circular Dichroism DataBank (PCDDB) aims to build a freely available database of CD spectra[108] and the web interface DichroWeb provides analyses of CD data using several open source algorithms.[109] Dichroweb currently supports the programmes SELCON3,[110,111] CONTINLL,[112,113] CDSSTR,[114–116] VARSLC[114–116] and K2D[117] and combines seven different reference databases to choose from. Using these tools, it calculates the secondary structure from the experimental spectrum, a goodness-of-fit parameter for the analyses and several overviews of calculation and experiments.

 

Although it is much more common for researchers to wish to extract structural information from a CD spectrum, it is sometimes of interest to generate a CD spectrum from a structure. If one suspects a protein contains a novel structure which the structure fitting databases do not include, then this approach is the only option. Alternatively, if the protein structure is dynamic and one wishes more information than a net average, then calculated spectra for different conformers are attractive. We have developed a freely accessible web interface implementing the matrix method. The required input is a PDB file to define the protein geometry. Files can also be retrieved from the RCSB Protein Data Bank[118] *via* their PDB code. An experimental spectrum for comparison may be included and a number of archive types can be uploaded to run calculations on several proteins.

After uploading a single PDB file or an archive containing many PDB files and/or giving PDB codes to fetch files from the RCSB, the user can choose whether to include the side chain contributions, charge transfer transitions, the curve type of the band shapes and the bandwidth used for them. Units of ellipticity or absorption may be used. If the NMAabinit parameter set is chosen, the user can also include three or four backbone transitions instead of the two selected by default. In addition to the CD spectrum, DichroCalc can also produce the LD spectrum, thus covering all the possibilities described in the preceding sections. After submitting the job, the results are returned by email and contain the coordinate files as well as the comparison plots of calculation and experimental data (if provided). Both the NMAabinit and NMA semi parameter set are available for use in the calculation. The interface can be reached at http://comp.chem.nottingham.ac.uk/dichrocalc/. We hope that it will be of use to researchers undertaking studies similar to those described in section 2.6.

## 5. Conclusion

At the moment, the quality of CD spectra calculated by the matrix method depends very much on the structure of the protein in question. The CD spectra of highly α-helical proteins can be calculated with a Spearman rank correlation of 0.92 between the experimental and calculated intensity at 220 nm. Since the α-helix is the most common secondary structure type, the quantitative understanding of its contribution to the CD is a substantial achievement. The effects of random coil and particularly β-II structures on the observed CD spectra have to be better understood to achieve the level of agreement that can be reached for α-helical proteins. It is generally the case that, apart from the β-II class proteins, the *ab initio* set of parameters tends to overestimate CD intensities, especially in the region around 190 nm. The role of the various parameters in the calculations and the underlying assumptions of the matrix methodology are being further investigated. Furthermore, we are attempting to improve the details of the shape of the calculated spectra, *e.g.* to achieve the two minima of the α-helix between 208 and 220 nm without arbitrary choices of parameters to achieve the goal.

The significant progress which has been made in calculating CD spectra over the last few years is now being transferred over to LD spectroscopy. Protein LD data are becoming more available and it has become clear that they contain a great deal of structural information. However, due to overlapping transitions this information is often impossible to extract without understanding the spectroscopy of the system. The matrix method is directly applicable to LD data interpretation, as LD depends only on the electric moments of the transition, spectral shapes are easier to calculate. The issue of degree of sample orientation remains a challenge both for the experiment and the data interpretation.

## Acknowledgements

We thank the Engineering and Physical Sciences Research Council (EPSRC) for funding (grant number GR/T09224) and Dr Daniel Barthel, University of Nottingham, for useful discussions.

## References

1 D.-F.-M. Arago, *Mem. Inst. Fr.*, 1811, **12**, 93.
2 J.-B. Biot, *Mem. Inst. Fr.*, 1812, **13**, 1–372.
3 J.-B. Biot, *Bull. Soc. Philomath.*, 1815, 190.
4 G. D. Fasman, *Circular Dichroism and the conformational analysis of biomolecules*, Plenum Press, New York, 1996.
5 L. Pasteur, *C. R. Acad. Sci. Paris*, 1848, **26**, 535–539.
6 J. A. Le Bel, *Bull. Soc. Chim. Fr.*, 1874, **22**, 337–347.
7 J. H. van't Hoff, *La Chimie dans L'Espace*, 1875.
8 G. Snatzke, *Angew. Chem., Int. Ed. Engl.*, 1968, **7**, 14–25.
9 N. Berova, K. Nakanishi and R. W. Woody, *Circular Dichroism: Principles and Applications*, John Wiley and Sons, New York, 2nd edn, 2000.
10 M. Born, *Phys. Z.*, 1915, **16**, 251–258.
11 W. Kuhn, *Z. Phys. Chem. B*, 1929, **4**, 14–36.
12 L. Rosenfeld, *Z. Phys.*, 1928, **52**, 161–174.
13 D. D. Fitts and J. G. Kirkwood, *Proc. Natl. Acad. Sci. U. S. A.*, 1956, **42**, 33–36.
14 W. Moffitt, *J. Chem. Phys.*, 1956, **25**, 467–478.
15 W. Moffitt, *Proc. Natl. Acad. Sci. U. S. A.*, 1956, **42**, 736–746.
16 A. S. Davydov, *Theory of molecular excitations*, Plenum Press, New York, 1971.
17 P. Urnes and P. Doty, *Adv. Protein Chem.*, 1961, **16**, 401–544.
18 G. Holzwarth and P. Doty, *J. Am. Chem. Soc.*, 1965, **87**, 218–228.
19 K. M. Specht, J. Nam, D. M. Ho, N. Berova, R. K. Kondru, D. N. Beratan, P. Wipf, R. A. Pascal and D. Kahne, *J. Am. Chem. Soc.*, 2001, **123**, 8961–8966.
20 K. Tanaka, Y. Itagaki, M. Satake, H. Naoki, T. Yasumoto, K. Nakanishi and N. Berova, *J. Am. Chem. Soc.*, 2005, **127**, 9561–9570.
21 P. Butz, G. E. Tranter and J. P. Simons, *PhysChemComm*, 2002, **5**, 91–93.
22 F. Furche, R. Ahlrichs, C. Wachsmann, E. Weber, A. Sobanski, F. Vogtle and S. Grimme, *J. Am. Chem. Soc.*, 2000, **122**, 1717–1724.
23 M. Schreiber, R. Vahrenhorst, V. Buss and M. P. Fulscher, *Chirality*, 2001, **13**, 571–576.
24 H. DeVoe, *J. Chem. Phys.*, 1964, **41**, 393–400.
25 H. DeVoe, *J. Chem. Phys.*, 1965, **43**, 3199–3208.
26 J. Applequist, *J. Chem. Phys.*, 1979, **71**, 4332–4338.
27 J. Applequist, K. R. Sundberg, M. L. Olson and L. C. Weiss, *J. Chem. Phys.*, 1979, **70**, 1240–1246.
28 B. K. Sathyanarayana and J. Applequist, *Int. J. Pept. Protein Res.*, 1985, **26**, 518–527.
29 K. A. Bode and J. Applequist, *J. Phys. Chem.*, 1996, **100**, 17825–17834.
30 K. L. Carlson, S. L. Lowe, M. R. Hoffmann and K. A. Thomasson, *J. Phys. Chem. A*, 2006, **110**, 1925–1933.
31 S. L. Lowe, R. R. Pandey, J. Czlapinski, G. Kie-Adams, M. R. Hoffmann, K. A. Thomasson and K. S. Pierce, *J. Pept. Res.*, 2002, **61**, 189–201.
32 P. M. Bayley, E. B. Nielsen and J. A. Schellman, *J. Phys. Chem.*, 1969, **73**, 228–243.

33 A. Rodger, J. Rajendra, R. Marrington, M. Ardhammar, B. Norden, J. D. Hirst, A. T. B. Gilbert, T. R. Dafforn, D. J. Halsall, C. A. Woolhead, C. Robinson, T. J. T. Pinheiro, J. Kazlauskaite, M. Seymour, N. Perez and M. J. Hannon, *Phys. Chem. Chem. Phys.*, 2002, **4**, 4051–4057.
34 S. M. Kelly, T. J. Jess and N. C. Price, *Biochim. Biophys. Acta*, 2005, **1751**, 119–139.
35 L. A. Nafie, J. C. Cheng and P. J. Stephens, *J. Am. Chem. Soc.*, 1975, **97**, 3842–3843.
36 L. A. Nafie and M. Diem, *Acc. Chem. Res.*, 1979, **12**, 296–302.
37 T. A. Keiderling, *Appl. Spectrosc. Rev.*, 1981, **17**, 189–226.
38 L. D. Barron, *Molecular Light Scattering and Optical Activity*, Cambridge University Press, 1982.
39 G. Holzwarth, E. C. Hsu, H. S. Mosher, T. R. Faulkner and A. Moscowitz, *J. Am. Chem. Soc.*, 1974, **96**, 251–252.
40 P. M. Bayley, *Prog. Biophys. Mol. Biol.*, 1973, **27**, 1–76.
41 M. T. Oakley, B. M. Bulheller and J. D. Hirst, *Chirality*, 2006, **18**, 340–347.
42 S. Brahms and J. Brahms, *J. Mol. Biol.*, 1980, **138**, 149–178.
43 C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry*, W. H. Freeman, San Francisco, 1980.
44 W. C. Johnson, *Proteins—Structure Function and Genetics*, 1990, **7**, 205–214.
45 P. Wallimann, R. J. Kennedy and D. S. Kemp, *Angew. Chem., Int. Ed. Engl.*, 1999, **38**, 1290–1292.
46 J. T. Yang, C. S. C. Wu and H. M. Martinez, *Meth. Enzymol.*, 1986, **130**, 208–269.
47 P. Z. Luo and R. L. Baldwin, *Biochemistry*, 1997, **36**, 8413–8421.
48 C. D. Andrew, S. Bhattacharjee, N. Kokkoni, J. D. Hirst, G. R. Jones and A. J. Doig, *J. Am. Chem. Soc.*, 2002, **124**, 12706–12714.
49 Z. Dang and J. D. Hirst, *Angew. Chem., Int. Ed.*, 2001, **40**, 3619–3621.
50 S. Bhattacharjee, G. Toth, S. Lovas and J. D. Hirst, *J. Phys. Chem. B*, 2003, **107**, 8682–8688.
51 K. W. Plaxco and C. M. Dobson, *Curr. Opin. Struct. Biol.*, 1996, **6**, 630–636.
52 J. D. Hirst, S. Bhattacharjee and A. V. Onufriev, *Faraday Discuss.*, 2003, **122**, 253–267.
53 E. L. Eliel, S. H. Wilen and L. N. Mander, *Stereochemistry of Organic Compounds*, Wiley, New York, 1994.
54 N. Sreerama and R. W. Woody, *Meth. Enzymol.*, 2004, **383**, 318–351.
55 I. Tinoco, *Adv. Chem. Phys.*, 1962, **4**, 113–160.
56 J. G. Kirkwood, *J. Chem. Phys.*, 1937, **5**, 479–491.
57 R. W. Woody, *J. Chem. Phys.*, 1968, **49**, 4797–4806.
58 R. W. Woody and I. Tinoco, *J. Chem. Phys.*, 1967, **46**, 4927–4945.
59 J. D. Hirst and N. A. Besley, *J. Chem. Phys.*, 1999, **111**, 2846–2847.
60 N. A. Besley and J. D. Hirst, *J. Am. Chem. Soc.*, 1999, **121**, 9636–9644.
61 R. W. Woody and N. Sreerama, *J. Chem. Phys.*, 1999, **111**, 2844–2845.
62 A. Bernhardsson, R. Lindh, G. Karlstrom and B. O. Roos, *Chem. Phys. Lett.*, 1996, **251**, 141–149.
63 G. Karlstrom, *J. Phys. Chem.*, 1989, **93**, 4952–4955.
64 G. Karlstrom, *J. Phys. Chem.*, 1988, **92**, 1315–1318.
65 L. Serrano-Andrés, M. P. Fülscher and G. Karlstrom, *Int. J. Quantum Chem.*, 1997, **65**, 167–181.
66 J. D. Hirst, K. Colella and A. T. B. Gilbert, *J. Phys. Chem. B*, 2003, **107**, 11813–11819.
67 Y. H. Chen, J. T. Yang and H. M. Martinez, *Biochemistry*, 1972, **11**, 4120–4131.
68 D. M. Rogers and J. D. Hirst, *J. Phys. Chem. A*, 2003, **107**, 11191–11200.
69 J. R. Platt, *J. Chem. Phys.*, 1949, **17**, 484–495.
70 Y. Yamamoto and J. Tanaka, *Bull. Chem. Soc. Jpn.*, 1972, **45**, 1362–1366.
71 D. M. Rogers and J. D. Hirst, *Biochemistry*, 2004, **43**, 11092–11102.
72 N. Sreerama and R. W. Woody, *Protein Sci.*, 2003, **12**, 384–388.
73 N. A. Besley and J. D. Hirst, *J. Mol. Struct. (THEOCHEM)*, 2000, **506**, 161–167.
74 W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, Cambridge University Press, Cambridge, 1992.
75 R. A. Laskowski, M. W. Macarthur, D. S. Moss and J. M. Thornton, *J. Appl. Crystallogr.*, 1993, **26**, 283–291.
76 J. C. Sutherland, E. J. Desmond and P. Z. Takacs, *Nucl. Instrum. Meth.*, 1980, **172**, 195–199.
77 R. W. Janes, *Bioinformatics*, 2005, **21**, 4230–4238.
78 B. A. Wallace and R. W. Janes, *Curr. Opin. Chem. Biol.*, 2001, **5**, 567–571.
79 K. Matsuo, R. Yonehara and K. Gekko, *J. Biochem.*, 2004, **135**, 405–411.
80 K. Matsuo, R. Yonehara and K. Gekko, *J. Biochem.*, 2005, **138**, 79–88.
81 A. J. Miles and B. A. Wallace, *Chem. Soc. Rev.*, 2006, **35**, 39–51.
82 M. T. Oakley and J. D. Hirst, *J. Am. Chem. Soc.*, 2006, **128**, 12414–12415.
83 M. B. Robin, *Higher Excited States of Polyatomic Molecules*, Academic Press, New York, 1975.
84 N. A. Besley and J. D. Hirst, *J. Phys. Chem. A*, 1998, **102**, 10791–10797.
85 A. P. Demchenko, *Ultraviolet Spectroscopy of Proteins*, Springer, Berlin, 1986.
86 R. Weinkauf, P. Schanen, D. Yang, S. Sonkara and E. W. Schlag, *J. Phys. Chem.*, 1995, **99**, 11255–11265.
87 W. G. McGimpsey, L. Chen, R. Carraway and W. N. Samaniego, *J. Phys. Chem. A*, 1999, **103**, 6082–6090.
88 L. Serrano-Andrés and M. P. Fülscher, *J. Phys. Chem. B*, 2001, **105**, 9323–9330.
89 A. T. B. Gilbert and J. D. Hirst, *J. Mol. Struct. (THEOCHEM)*, 2004, **675**, 53–60.
90 A. Rodger and B. Nordén, *Circular Dichroism & Linear Dichroism*, Oxford University Press, Oxford, 1997.
91 M. Couette, *Ann. Chim. Phys.*, 1890, **6**, 433–510.
92 A. Mallock, *Proc. R. Soc. London*, 1888, **45**, 126.
93 A. Rodger, R. Marrington, M. A. Geeves, M. Hicks, L. de Alwis, D. J. Halsall and T. R. Dafforn, *Proc. R. Soc. London*, 2006, **8**, 3161–3171.
94 S. Higashi, M. Kasai, F. Oosawa and A. Wada, *J. Mol. Biol.*, 1963, **7**, 421–430.
95 A. Wada, *Appl. Spectrosc. Rev.*, 1972, **6**, 1–30.
96 T. R. Dafforn, J. Rajendra, D. J. Halsall, L. C. Serpell and A. Rodger, *Biophys. J.*, 2004, **86**, 404–410.
97 T. R. Dafforn and A. Rodger, *Curr. Opin. Struct. Biol.*, 2004, **14**, 541–546.
98 R. Marrington, E. Small, A. Rodger, T. R. Dafforn and S. Addinall, *J. Biol. Chem.*, 2004, **47**, 48821–48829.
99 R. Marrington, M. Seymour and A. Rodger, *Chirality*, 2006, in press.
100 R. Marrington, T. R. Dafforn, D. J. Halsall, J. I. MacDonald, M. Hicks and A. Rodger, *Analyst*, 2005, **130**, 1608–1616.
101 M. Ardhammar, N. Mikati and B. Nordén, *J. Am. Chem. Soc.*, 1998, **120**, 9957–9958.
102 A. Rodger, J. Rajendra, R. Marrington, R. Mortimer, T. Andrews, J. B. Hirst, A. T. B. Gilbert, D. Halsall, T. Dafforn, M. Ardhammar, B. Nordén, C. A. Woolhead, C. Robinson, T. Pinheiro, K. J. M. Seymour, N. Perez and M. J. Hannon, in *Biophysical Chemistry: Membranes and Proteins*, ed. R. H. Templer and R. Leatherbarrow, The Royal Society of Chemistry, Cambridge, Editon edn., 2002, pp. 3–19.
103 J. Rajendra, A. Damianoglou, M. Hicks, P. Booth, P. M. Rodger and A. Rodger, *Chem. Phys.*, 2005.
104 J. P. Cartailler and H. Luecke, *Annu. Rev. Biophys. Biomol. Struct.*, 2003, **32**, 285–310.
105 W. Meijberg and P. J. Booth, *J. Mol. Biol.*, 2002, **319**, 839–853.
106 D. Oesterhelt and L. Schuhmann, *FEBS Lett.*, 1974, **44**, 262–265.
107 B. Albinsson, M. Kubista, B. Nordén and E. W. Thulstrup, *J. Phys. Chem.*, 1989, **93**, 6646–6654.
108 L. Whitmore, B. A. Wallace and R. W. Janes, *The Protein Circular Dichroism DataBank (PCDDB)*; http://pcddb.cryst.bbk.ac.uk/, 2006.
109 L. Whitmore and B. A. Wallace, *Nucleic Acids Res.*, 2004, **32**, W668–W673.


 

110 N. Sreerama, S. Y. Venyaminov and R. W. Woody, *Protein Sci.*, 1999, **8**, 370–380.
111 N. Sreerama and R. W. Woody, *Anal. Biochem.*, 1993, **209**, 32–44.
112 S. W. Provencher and J. Glockner, *Biochemistry*, 1981, **20**, 33–37.
113 I. H. M. Vanstokkum, H. J. W. Spoelder, M. Bloemendal, R. Vangrondelle and F. C. A. Groen, *Anal. Biochem.*, 1990, **191**, 110–118.
114 L. A. Compton and W. C. Johnson, *Anal. Biochem.*, 1986, **155**, 155–167.
115 P. Manavalan and W. C. Johnson, *Anal. Biochem.*, 1987, **167**, 76–85.
116 N. Sreerama and R. W. Woody, *Anal. Biochem.*, 2000, **287**, 252–260.
117 M. A. Andrade, P. Chacon, J. J. Merelo and F. Moran, *Protein Eng.*, 1993, **6**, 383–390.
118 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–242.