

Molecular Physics

An International Journal at the Interface Between Chemistry and Physics

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/tmph20>

Electronic circular dichroism of proteins computed using a diabatisation scheme

David M. Rogers, Hainam Do & Jonathan D. Hirst

To cite this article: David M. Rogers, Hainam Do & Jonathan D. Hirst (2022): Electronic circular dichroism of proteins computed using a diabatisation scheme, Molecular Physics, DOI: [10.1080/00268976.2022.2133748](https://doi.org/10.1080/00268976.2022.2133748)

To link to this article: <https://doi.org/10.1080/00268976.2022.2133748>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 27 Oct 2022.



[Submit your article to this journal](#)



Article views: 43



[View related articles](#)



[View Crossmark data](#)

Electronic circular dichroism of proteins computed using a diabatisation scheme

David M. Rogers ^a, Hainam Do ^{b,c} and Jonathan D. Hirst ^a

^aSchool of Chemistry, University of Nottingham, University Park, Nottingham, United Kingdom; ^bDepartment of Chemical and Environmental Engineering and Key Laboratory of Carbonaceous Waste Processing and Process Intensification Research of Zhejiang Province, University of Nottingham Ningbo China, Ningbo, People's Republic of China; ^cNew Materials Institute, University of Nottingham Ningbo China, Ningbo, People's Republic of China

ABSTRACT

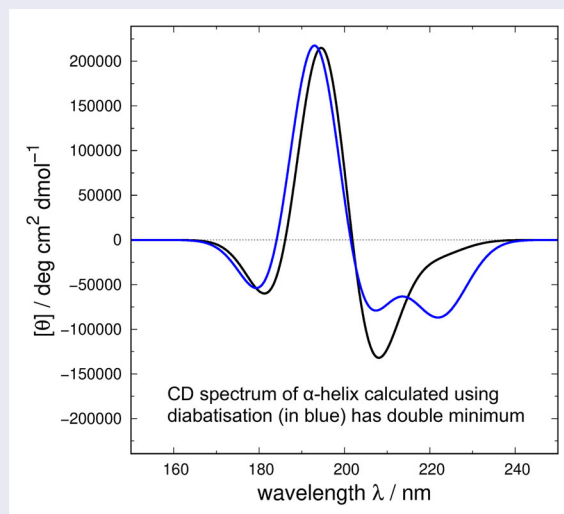
Circular dichroism (CD) spectroscopy is a powerful technique employed to study the structure of biomolecules. More accurate calculation of CD from first principles will aid both computational and experimental studies of protein structure and dynamics. We apply a diabatisation scheme to improve the description of nearest neighbour interactions between two electronic transitions ($n\pi^*$ and $\pi_{nb}\pi^*$) localised on each individual peptide bond (amide group) in a protein. These interactions are incorporated into DichroCalc, an exciton-based computational method to calculate CD, and yield improvements over the standard DichroCalc parameter set, particularly for calculation of CD for important secondary structural elements such as an α helix.

ARTICLE HISTORY

Received 29 July 2022
Accepted 30 September 2022

KEYWORDS



Exciton; CD spectroscopy; diamide; ab initio; far-UV




Introduction

Calculations from first principles of the electronic circular dichroism (CD) spectroscopy of proteins are challenging. It is a topic that has attracted the interest of many researchers, including the late Nick Besley [1], to whose memory this Special Issue is dedicated. In this paper, we present a new theoretical development, but we begin by surveying some of the contributions of Besley to the field, which underpin the current study.

The CD spectra of proteins in the far-ultraviolet (far-UV) have distinct signatures, reflecting the protein secondary structure, which arise from the regular, repeating arrays of backbone amide chromophores. Thus, CD is a widely used biophysical technique for characterising proteins in solution [2]. Often, statistical techniques are used to estimate the secondary structure content from the experimentally determined CD spectrum [3,4]. However, our focus is on the reverse, namely, computing

CONTACT Jonathan D. Hirst  jonathan.hirst@nottingham.ac.uk  School of Chemistry, University of Nottingham, University Park, Nottingham, NG7 2RD United Kingdom

 Supplemental data for this article can be accessed here. <https://doi.org/10.1080/00268976.2022.2133748>

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

the CD spectrum from the atomic coordinates of the protein.

CD spectra can be computed from protein structures in various ways, *e.g.* see [5–10], some of which have been reviewed elsewhere [11]. We focus here on the exciton framework, whereby an effective Hamiltonian is constructed through a consideration of localised chromophores within the protein [12,13]. The peptide bond is the most important contributor to the far-UV CD spectra of proteins. A simple effective Hamiltonian would consider the $n\pi^*$ and $\pi_{nb}\pi^*$ electronic transitions of the amide chromophore, which occur at about 222 and 190 nm, respectively. The π_{nb} , n and π^* molecular orbitals for an example diamide are depicted in Figure 1.

A desire to improve the quality of protein CD calculations that use the exciton framework motivated several

studies of the electronic structure of small amides using *ab initio* multi-configurational methods, including complete active space self-consistent field (CASSCF) calculations on formamide [14] and *N*-methylacetamide [15] in the gas phase. Analogous calculations with various models of solvent showed subtle, but important, changes to the electronic transitions, including, for example, differences in the orientation of the $\pi_{nb}\pi^*$ electric transition dipole moment compared to the gas phase [16,17]. This factor was influential in leading to an improvement in the accuracy of calculations of protein CD spectra based on parameterisation of the *ab initio* descriptions of the amide chromophore [18–20] compared to spectra calculated using parameters from semi-empirical calculations [21]. The computational methodology for protein CD calculations and the chromophore parameters are

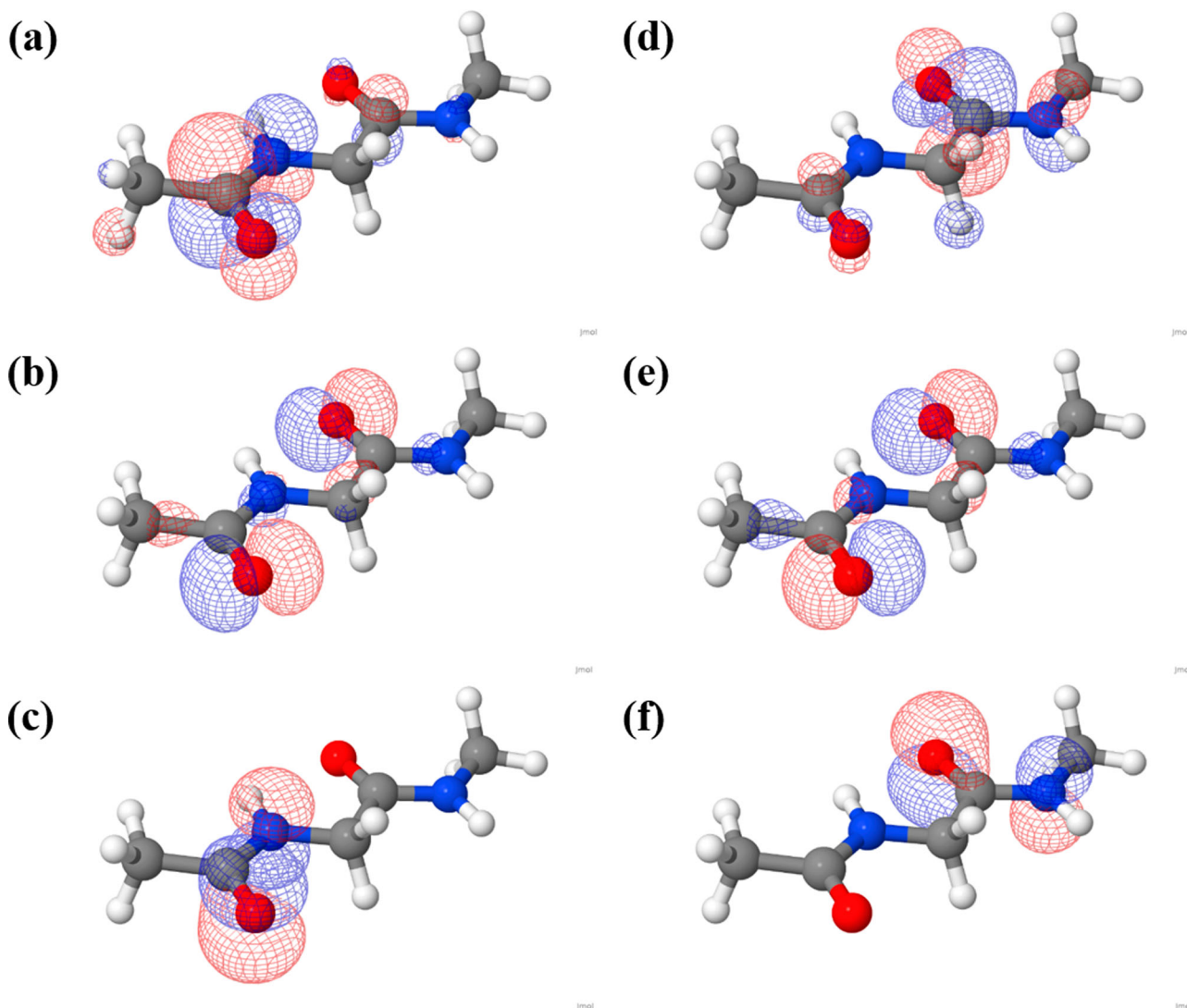


Figure 1. Molecular orbitals for diamide **2a** [$\varphi = 180^\circ$, $\psi = 180^\circ$] with an isosurface cut-off of 0.05. Orbitals from a state-averaged CASSCF calculation as described in reference 32. (a) π^* for amide 1, (b) n for amide 1, (c) π_{nb} for amide 1, (d) π^* for amide 2, (e) n for amide 2 and (f) π_{nb} for amide 2.

available via a web-interface [22], which allows the non-expert simply to upload a Protein Data Bank file with the coordinates of a protein structure and to compute the CD spectrum. The theoretical background and related developments have been summarised elsewhere [11,23] and the approach has been used, sometimes in conjunction with molecular dynamics simulations, to study a diverse range of systems, *e.g.* glycosylated antimicrobial peptides [24], SARS-CoV-2 proteins [25] and chymotrypsin adsorbed on silica [26].

The above discussion has focused on two localised electronic transitions, the $n\pi^*$ and $\pi_{nb}\pi^*$ excitations, on an individual amide chromophore. Whilst the computational approach based on this approximation has been successful, additional insights have been gained through the study of cyclic [27–29] and linear diamides [30–32]. Electronic excitations with higher transition energies may be important [33] and in this region charge transfer transitions between orbitals on neighbouring peptide groups are also observed. Quantum chemical calculations at the CASSCF level on a linear diamide enabled the generation of parameters describing the charge transfer transitions [31]. Protein CD calculations incorporating these transitions gave an encouraging level of agreement with synchrotron radiation CD spectra [34] measured down to 170 nm. The CASSCF calculations on the diamide provide an additional opportunity, which is the subject of our study, namely, to consider the interaction between two covalently bound amide chromophores at the quantum chemical level using a procedure known as diabatisation.

The effective Hamiltonian matrix in the exciton framework comprises diagonal elements corresponding to the energies of monomer (diabatic) states and off-diagonal elements. The latter are inter-state interactions or couplings. Whilst the isolated monomer, *i.e.* *N*-methylacetamide, appears to be a good choice of diabatic state with the couplings computed using a Coulombic model, an alternative approach would be to apply a diabatisation scheme to a full *ab initio* calculation on a diamide model, whereby the eigenstates of the diamide are transformed to a diabatic basis. This method is capable of treating multiple excited states and also taking into account both short- and long-range effects of the excitonic couplings. Thus, an orthogonal transformation is used to connect the adiabatic excited states with a set of diabatic excited states. There is no single ‘ideal’ set of diabatic states [35], but several strategies can be adopted to define the adiabatic-to-diabatic transformation. The fragment excitation difference scheme uses the so-called excitation density [36]. A variant of this and some extensions are available in the EXAT software [37]. An alternative strategy derives the couplings from the

energy splitting between the excited states and the extent of exciton delocalisation between the fragments [38]. Other approaches have maximised the overlap with a reference wavefunction [39] or the similarity with the dipole moment [40]. Diabatisation approaches have been applied to aggregates of anthracene and tetracene [41], to aggregates of propeller-shaped emitters [42] using the fromage software [43], dimers of ethylene and derivatives [44] and to various nucleic acid systems [45,46]. Herein, we apply a diabatisation scheme to compute the CD of some polypeptide structures corresponding to conformations of particular significance, including α -helices, β -strands, polyproline II helices and 3_{10} -helices.

Proteins rich in β -strands can be separated into two distinct classes: β -I and β -II [47]. This classification can be made based on their CD spectra [47,48]. β -I proteins, *e.g.* concanavalin A, contain regular β strands and give rise to CD spectra with a positive band at 195 nm and a negative band at 215–220 nm. β -II proteins, *e.g.* elastase, yield CD spectra reminiscent of unordered polypeptides with a negative band around 198 nm [49,50]. It is thought that β -II proteins may adopt more polyproline II conformation [50], in addition to the possibility that they are more conformationally labile. Accurate computation of the CD spectrum of the polyproline II conformation would, thus, help provide useful insight into the spectra of β -II proteins. Another conformation of interest is the 3_{10} helix, which, despite some structural similarity to the α helix, has a distinct CD spectrum [51].

Methods

Diabatisation

A diabatisation procedure was performed to generate more accurate inter-amide couplings between two electronic transitions ($n\pi^*$ and $\pi_{nb}\pi^*$) localised on each of the two amide monomers when in a dimeric system (a diamide). The procedure follows that described by Aragó and Troisi [41], which is based on orthogonal transformation of the adiabatic states (eigenstates of the total Hamiltonian) to diabatic (noninteracting) states for a two-level dimer system and generalisable to systems with an arbitrary number of interacting states (four in this study). For two monomers (amides) labelled 1 and 2 that combine, via a peptide bond, as a dimer (a diamide), consider four singlet excited states on the dimer and two singlet excited states on each of the monomers. The four adiabatic excited states on the dimer are the $n_1\pi_1^*$ and $\pi_{nb1}\pi_1^*$ localised on amide 1 in the dimer and the $n_2\pi_2^*$ and $\pi_{nb2}\pi_2^*$ localised on amide 2 in the dimer. The two excited states on each of the two isolated monomers are $n\pi^*$ and $\pi_{nb}\pi^*$. The diabatic Hamiltonian \mathbf{H}^D is found

by a transformation of the adiabatic Hamiltonian \mathbf{H}^A , thus:

$$\mathbf{H}^D = \mathbf{C}\mathbf{H}^A\mathbf{C}^\dagger \quad (1)$$

The unitary transformation matrix \mathbf{C} relates the diabatic states (on the dimer) to the adiabatic states (on the dimer). It is desirable that the diabatic wave functions be as close as possible to the initial noninteracting wave functions. The unitary matrix \mathbf{C} is computed by minimising the difference between the adiabatic transition dipole moments (TDMs) on the dimer and the isolated monomer TDMs in the coordinate frame of the dimer [41]. Computation of matrix \mathbf{C} proceeds as follows [41].

The matrix \mathbf{M} is formed from a product of matrices representing the adiabatic TDMs on the dimer, $\boldsymbol{\mu}^A$, and the isolated TDMs on each of the monomers, $\boldsymbol{\mu}^{\text{ISO}}$,

$$\mathbf{M} = (\boldsymbol{\mu}^A)^* \boldsymbol{\mu}^{\text{ISO}} \quad (2)$$

Singular valued decomposition is carried out on matrix \mathbf{M} to find the matrices \mathbf{U} and \mathbf{V}^* . These are used to obtain $\mathbf{C}^{*,\text{best}}$,

$$\mathbf{C}^{*,\text{best}} = \mathbf{U}\mathbf{V}^* \quad (3)$$

where $\mathbf{C}^{*,\text{best}}$ is the transpose of \mathbf{C}^{best} which is defined as the best unitary transformation satisfying the following equation,

$$\mathbf{C}^{*,\text{best}} = \arg \min_{\mathbf{R}} \|\boldsymbol{\mu}^A \mathbf{R} - \boldsymbol{\mu}^{\text{ISO}}\| \quad (4)$$

where $\|\cdot\|$ denotes the Frobenius norm.

The adiabatic Hamiltonian is diagonal, where the diagonal elements are the transition energies of the excited states in the dimeric system. The output of the diabatisation is the diabatic Hamiltonian with transition energies on the diagonal and the coupling between excited states in the diabatic representation in the off-diagonal matrix elements. The amide is the monomer and the diamide is the dimer. The monomer and the dimer state properties (TDMs) and adiabatic Hamiltonian are listed as matrices in Table S1. The input properties are from DichroCalc parameter sets for the amide *N*-methylacetamide (set NMA4FIT2) [17] and the diamide *N*-acetyl glycine *N'*-methylamide (13 sets for various φ and ψ dihedral angles of the diamide) [31,32]. The TDMs for the two monomers (from the above parameter set) are transformed onto the coordinate frame of the dimer prior to diabatisation. The 13 pairs of φ and ψ angles and the corresponding parameter sets for the diamide are listed in Table 1. The geometries were chosen to span the Ramachandran plot. The diabatic Hamiltonian output from the diabatisation contains off-diagonal matrix elements that are the couplings between the states on each monomer in the system of the

Table 1. The 13 diamides and their secondary structure and associated DichroCalc parameter set.

φ angle (°)	ψ angle (°)	Secondary structure	Diamide in Oakley et al. [32]	DichroCalc parameter set [22]
180	180	Fully extended	2a	CT01009A
-120	180		2b	CT02009B
-60	180		2c	CT03009A
-135	135	β strand	2d	CTBE009A
-120	120		2e	CT05009A
-120	60		2f	CT08009A
-120	0		2g	CT11009A
-60	0		2h	CT12009A
-74	-4	3_{10} helix	2i	CT31009A*
-48	-57	α helix	2j	CTAL009A
-60	-60	α helix	2k	CT15009A
-62	-41	α helix	2l [#]	CTBT009B*
-75	145	Polyproline II	2m [#]	CTP2007A

[#]Notation following Oakley et al. [32] * Set exclusive to Oakley and Hirst [31].

dimer, and it is these quantities that are used to modify the computation of CD (discussed below). The diabatic Hamiltonian matrices are listed in Table S1.

Model peptides

13 model idealised peptide structures, each comprising 20 L-alanine residues, with different repeating φ and ψ dihedral angles to describe differing proteinaceous secondary structure elements were constructed using the molecular editor Avogadro [52,53]. The 13 model peptides were considered based on the availability of DichroCalc parameter sets that describe intra-amide and inter-amide (charge-transfer) electronic transitions in a diamide (*N*-acetyl glycine *N'*-methylamide) of varying φ and ψ angles [31,32], for which the diabatisation procedure was carried out. The diamides are listed in Table 1. The naming of diamides **2l** and **2m** follows the labelling sequence of Oakley et al. [32], who studied diamides **2a** through to **2k** (Table 1). The three α -helical diamide structures **2j**, **2k** and **2l** are, respectively, the Corey-Pauling-Branson helix, an idealised helix and the Barlow-Thornton helix [54]. Figure 2 shows a Ramachandran plot of the φ and ψ dihedral angles for the 13 structures. For clarity, we adopt a notation for each diamide and peptide with the φ and ψ dihedral angles in square brackets, e.g. for **2d** the notation is **2d** [-135° , 135°].

CD calculations

DichroCalc [22] was used to compute the CD spectra for the 13 model peptides. A modified version of the DichroCalc code was used to read a Hamiltonian matrix where, for each peptide, the nearest neighbour inter-amide interactions from the diabatisation procedure replaced the default off-diagonal matrix elements generated by DichroCalc. The theoretical approach utilised

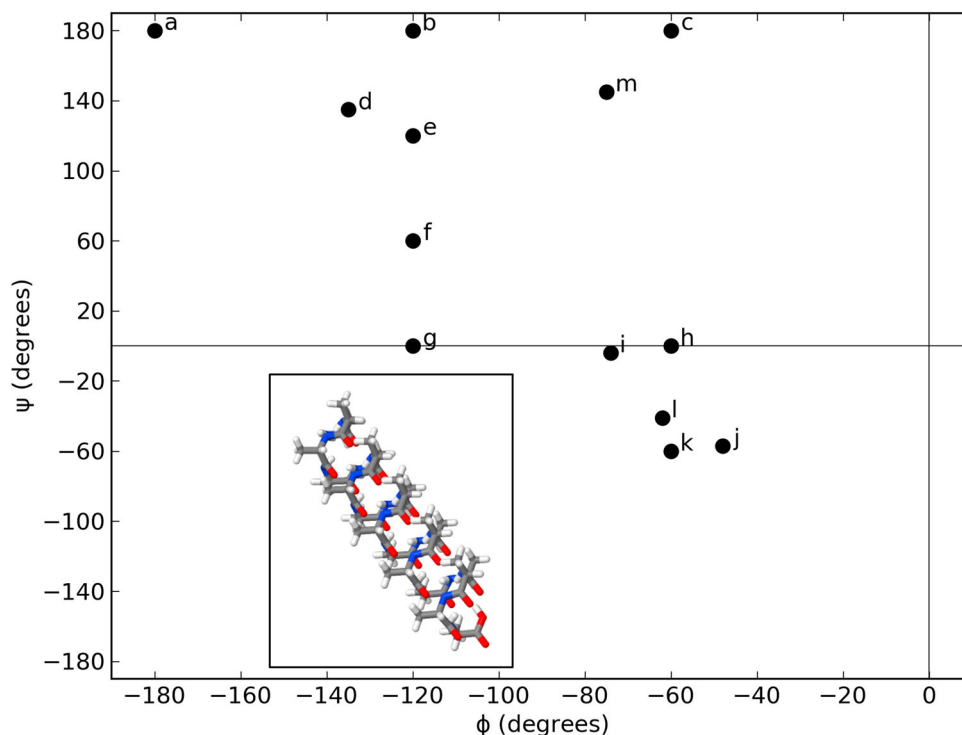


Figure 2. Ramachandran plot of the ϕ and ψ dihedral angles for the 13 diamide and 13 20 L-Ala peptide structures. The scatter points are labelled by the alphabetical (second) character naming each structure (see Table 1). Inset is the structure of peptide **2k** [$\phi = -60^\circ$, $\psi = -60^\circ$].

by DichroCalc, as well as alternative approaches to the computation of CD for biomolecules, has been described in detail in a recent review [11]. Briefly, the DichroCalc software implements the exciton framework, with the transition electron densities associated with excitations of biological chromophores represented by a set of point charges (monopoles) that reproduce the electrostatic potential arising from the electron density. The parameters (transition energies, electric and magnetic transition dipole moments, electric permanent dipole moments and monopole charge sets) for the amide backbone are derived from *ab initio* calculations (parameter set NMA4FIT2) [17] and describe the intra-amide $n\pi^*$ transition (at 220 nm) and the $\pi_{nb}\pi^*$ transition (at 193 nm). To obtain a CD spectrum, a rotational strength line spectrum, computed by DichroCalc, is convoluted with Gaussian functions of full-width half-maximum (FWHM) of typically, e.g. 9.0 or 12.5 nm. Throughout the rest of the paper, results using the un-modified DichroCalc Hamiltonian are termed ‘*ab initio*’ and the results obtained from modifying the off-diagonal elements of the DichroCalc Hamiltonian are termed ‘diabatic’.

Results

Figure 3 displays computed CD spectra (using Gaussian functions of FWHM of 9.0 nm) for the six most

physically relevant of the 13 model peptide geometries **2d** [-135° , 135°], **2i** [-74° , -4°], **2j** [-48° , -57°], **2k** [-60° , -60°], **2l** [-62° , -41°] and **2m** [-75° , 145°]. Figure S1 displays rotational strength line spectra and CD spectra for all 13 peptides. For seven of the 13 model peptides, there is a marked difference for the rotational strength line spectra and for the CD spectra when the diabatic inter-amide interactions are incorporated in the Hamiltonian. The seven peptides are: **2a** [180° , 180°], **2b** [-120° , 180°], **2c** [-60° , 180°], **2d** [-135° , 135°], **2f** [-120° , 60°], **2i** [-74° , -4°] and **2m** [-75° , 145°]. For the other six peptides, more subtle differences are found between the spectra. For the three α -helical peptides (**2j** [-48° , -57°], **2k** [-60° , -60°] and **2l** [-62° , -41°]) the diabatic CD spectra feature the double minimum (characteristic of an α -helical spectrum) in the 200–230 nm region; and the band maxima and minima are in similar regions of the spectrum for peptides **2e** [-120° , 120°], **2g** [-120° , 0°] and **2h** [-60° , 0°] but with slightly differing band shape characteristics. Reproduction of the double minimum between 200 and 230 nm represents a qualitative improvement over the *ab initio* calculated spectra.

Table S1 displays, for each of the 13 diamides, the three input matrices for the diabatisation procedure: the monomer TDMs, the dimer TDMs and the adiabatic Hamiltonian of the dimer. Table S1 also shows the

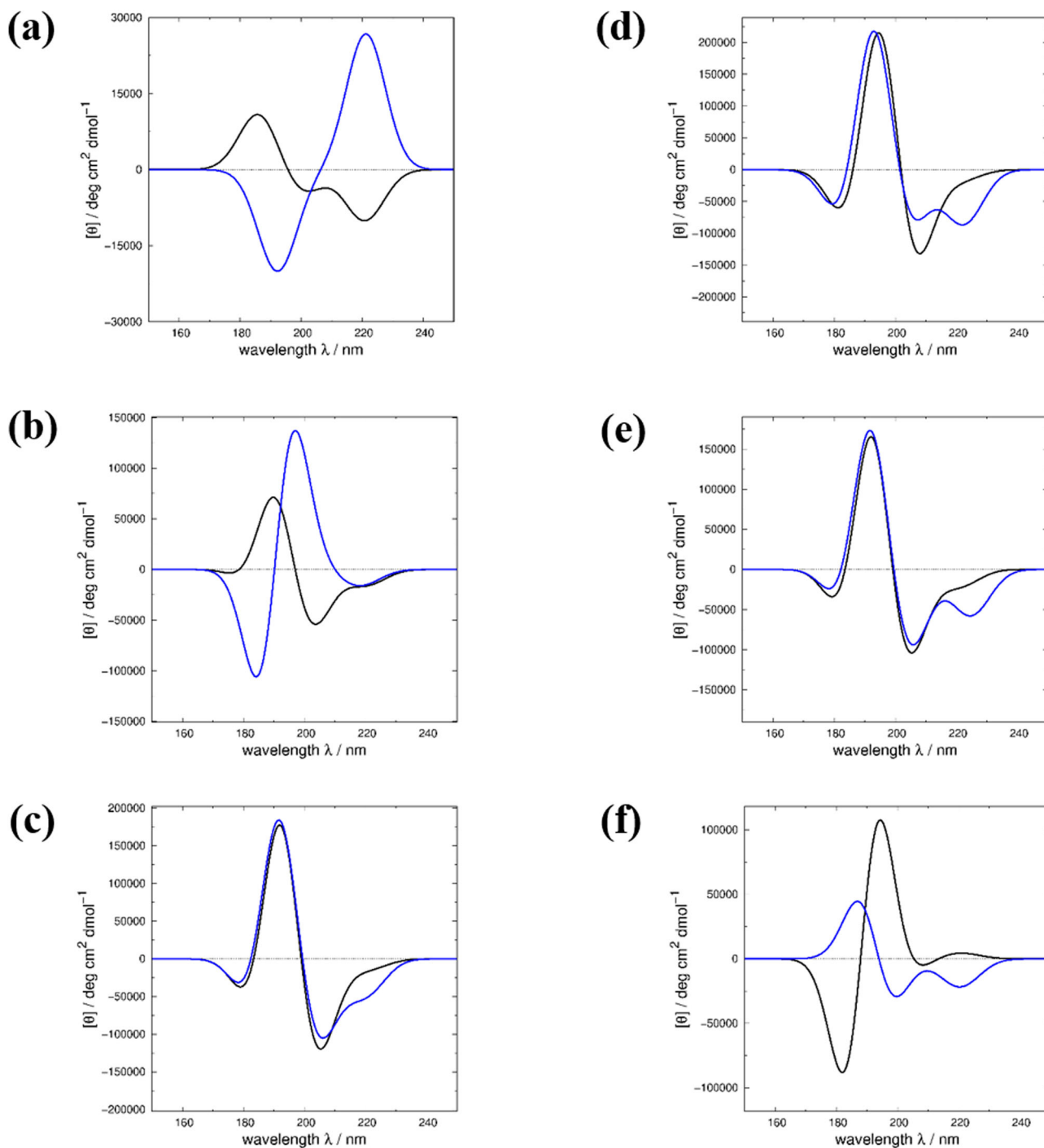


Figure 3. CD spectra for the model peptides comprising 20 L-Ala residues and named as the diamides. CD spectra with band shapes from convolution with Gaussian functions of FWHM of 9.0 nm. Spectra from DichroCalc *ab initio* Hamiltonian (black line) and diabatic Hamiltonian (blue line). (a) **2d** [-135° , 135°], (b) **2i** [-74° , -4°], (c) **2j** [-48° , -57°], (d) **2k** [-60° , -60°], (e) **2l** [-62° , -41°] and (f) **2m** [-75° , 145°].

outputted diabatic Hamiltonian, where the off-diagonal matrix elements 1–2 and 3–4 are intra-amide coupling interactions (for $n_1\pi_1^*$ with $\pi_{nb1}\pi_1^*$ and $n_2\pi_2^*$ with $\pi_{nb2}\pi_2^*$ transitions) and where matrix elements 1–3, 1–4, 2–3 and 2–4 are inter-amide coupling interactions (for $n_1\pi_1^*$ with $n_2\pi_2^*$, $n_1\pi_1^*$ with $\pi_{nb2}\pi_2^*$, $\pi_{nb1}\pi_1^*$

with $n_2\pi_2^*$ and $\pi_{nb1}\pi_1^*$ with $\pi_{nb2}\pi_2^*$ transitions, respectively).

The *ab initio* and the diabatic inter-amide interactions (off-diagonal elements of the Hamiltonian matrix) are shown in Table 2 for six of the 13 diamide geometries (**2d** [-135° , 135°], **2i** [-74° , -4°], **2j** [-48° , -57°], **2k**

Table 2. Inter-amide interactions. Off-diagonal elements of the Hamiltonian (cm^{-1}), DichroCalc *ab initio* matrix elements in the first row and diabatic matrix elements in the second row for each diamide geometry. The *ab initio* parameters have $n\pi^*$ and $\pi_{nb}\pi^*$ transition energies at 220 nm (45455 cm^{-1}) and 193 nm (51813 cm^{-1}), respectively.

Matrix element index	H_{13}	H_{14}	H_{23}	H_{24}
Diamide (φ, ψ)	$n_1\pi_1^*:\pi_2\pi_2^*$	$n_1\pi_1^*:\pi_{nb2}\pi_2^*$	$\pi_{nb1}\pi_1^*:\pi_2\pi_2^*$	$\pi_{nb1}\pi_1^*:\pi_{nb2}\pi_2^*$
2d ($-135^\circ, 135^\circ$)	72	-20	292	-907
	109	-279	-244	-89
2i ($-74^\circ, -4^\circ$)	15	-36	161	-662
	289	-148	230	284
2j ($-48^\circ, -57^\circ$)	-18	-17	-131	-272
	97	-808	0	-191
2k ($-60^\circ, -60^\circ$)	-42	-29	-180	-83
	-21	-1707	267	173
2l ($-62^\circ, -41^\circ$)	-17	-37	-52	-292
	-359	-660	-131	-363
2m ($-75^\circ, 145^\circ$)	31	-7	16	-971
	81	356	145	29

[$-60^\circ, -60^\circ$], **2l** [$-62^\circ, -41^\circ$] and **2m** [$-75^\circ, 145^\circ$]. Table S2 lists the couplings for all 13 diamides. For the diamide in the β strand arrangement (**2d** [$-135^\circ, 135^\circ$]), there are significant differences in the three matrix elements for the 1-4, 2-3 and 2-4 interactions (H_{14} , H_{23} and H_{24}). The sign is opposite for 2-3 (positive for the *ab initio* and negative for the diabatic) and the order of magnitude is different for 1-4 (one order larger for the diabatic) and 2-4 (one order larger for the *ab initio*). The diamide **2i** [$-74^\circ, -4^\circ$] (3_{10} helix) has differences in all the four off-diagonal matrix elements, with H_{24} ($\pi_{nb1}\pi_1^*$ with $\pi_{nb2}\pi_2^*$) having the largest difference between the *ab initio* and diabatic Hamiltonians.

For the three α -helical diamides (**2j** [$-48^\circ, -57^\circ$], **2k** [$-60^\circ, -60^\circ$] and **2l** [$-62^\circ, -41^\circ$]), the largest difference between the *ab initio* and the diabatic matrix elements are for the $n_1\pi_1^*$ with $\pi_{nb2}\pi_2^*$ interaction (H_{14}), where the diabatic interactions are one order of magnitude greater than the *ab initio* for diamides **2j** [$-48^\circ, -57^\circ$] and **2l** [$-62^\circ, -41^\circ$] and two orders of magnitude for diamide **2k** [$-60^\circ, -60^\circ$], with all these matrix elements having the same sign (negative). This is clearly an important factor leading to the distinctive double minimum in the α -helical spectra. The diamide representing a polyproline II helical conformation (diamide **2m** [$-75^\circ, 145^\circ$]), has large differences between *ab initio* and diabatic Hamiltonians for matrix elements H_{14} , H_{23} and H_{24} , with the largest differences between H_{14} and H_{24} .

The Coulombic coupling of two electronic transitions is related to the dot product of the transition dipole moments located on monomers A and B [55],

$$V_{dd} = \frac{\mu_A \mu_B}{R_{AB}^3} - \frac{3(\mu_A \vec{R}_{AB})(\mu_B \vec{R}_{AB})}{R_{AB}^5} \quad (5)$$

V_{dd} is the coupling between the two dipoles μ_A and μ_B , and R_{AB} is the distance between the two monomers A

and B. If the diabatic coupling is different from the *ab initio* coupling and the dot product of the transition dipole moments is similar in magnitude, then the coupling may have a short-range non-Coulombic interaction that is not described by the interaction of two dipole moments (or two sets of monopole charges) from a monomeric calculation. The diabatic coupling is from diabatisation of the adiabatic states on a complete dimer, including all the electrons in the system, using a basis set (for the diabatic states) of the monomer transition dipole moments. For cases where the *ab initio* and diabatic coupling differs and the magnitude of the dot product of the transition dipole moments differ, the diabatic states may contain a more accurate description of Coulombic interactions between the two transitions, due to the more accurate wave function of the adiabatic states, i.e. for the complete dimer.

Table 3 displays the dot products of the un-normalised transition dipole moments for the monomer (in the arrangement of the dimer) and the dimer for the subset of six diamides (**2d** [$-135^\circ, 135^\circ$], **2i** [$-74^\circ, -4^\circ$], **2j** [$-48^\circ, -57^\circ$], **2k** [$-60^\circ, -60^\circ$], **2l** [$-62^\circ, -41^\circ$] and **2m** [$-75^\circ, 145^\circ$]). Data for all 13 diamides are in Table S5.

Discussion

α helices

The experimental spectrum for an α helix possesses an intense positive band peak at 190 nm and a double minimum band with the negative band peaks at 208 and 220 nm. The CD spectra computed for the α -helical peptides (**2j** [$-48^\circ, -57^\circ$], **2k** [$-60^\circ, -60^\circ$] and **2l** [$-62^\circ, -41^\circ$]) using the *ab initio* Hamiltonians and the diabatic Hamiltonians all feature a positive band peak around 190 nm (Figure 3), with very little difference between the *ab initio* and diabatic spectral bands in this region. Differences in the computed spectra are apparent at wavelengths greater

Table 3. Dot products of the un-normalised transition dipole moments (D^2) for six of the 13 diamide structures (Table S1). Upper triangle of Hamiltonian H_{ij} . Intra-amide interactions are 1–2 and 3–4. Inter-amide interactions are 1–3, 1–4, 2–3 and 2–4. Transition dipole moments are not normalised.

Transitions	1–2	1–3	1–4	2–3	2–4	3–4
Diamide (φ, ψ)	Monomer					
2d ($-135^\circ, 135^\circ$)	0.0	-0.032	-0.350	-0.342	-2.037	0.0
2i ($-74^\circ, -4^\circ$)	0.0	-0.013	-0.454	0.615	-2.985	0.0
2j ($-48^\circ, -57^\circ$)	0.0	-0.002	-0.595	0.593	1.312	0.0
2k ($-60^\circ, -60^\circ$)	0.0	0.009	-0.563	0.578	3.020	0.0
2l ($-62^\circ, -41^\circ$)	0.0	-0.003	-0.547	0.629	0.984	0.0
2m ($-75^\circ, 145^\circ$)	0.0	-0.002	0.005	-0.619	-2.865	0.0
Diamide (φ, ψ)	Dimer					
2d ($-135^\circ, 135^\circ$)	0.284	-0.036	0.407	-0.399	2.842	-0.052
2i ($-74^\circ, -4^\circ$)	0.065	0.011	-0.280	-0.730	3.175	-0.512
2j ($-48^\circ, -57^\circ$)	-0.445	-0.040	-0.020	0.759	1.140	-0.127
2k ($-60^\circ, -60^\circ$)	-0.155	-0.040	-0.296	0.390	-5.282	0.529
2l ($-62^\circ, -41^\circ$)	-0.389	-0.035	-0.116	0.618	-3.997	0.434
2m ($-75^\circ, 145^\circ$)	-0.953	0.048	0.543	-0.610	-4.660	0.637

than 200 nm where the double minimum is expected. For peptide **2j** [$-48^\circ, -57^\circ$] the negative peak at 205 nm in the *ab initio* spectrum is slightly less intense and red shifted to 206 nm in the diabatic spectrum, and the negative band shoulder around 220 nm in the *ab initio* spectrum is both more pronounced and intense in the diabatic spectrum. At wavelengths greater than 200 nm for the peptide **2k** [$-60^\circ, -60^\circ$], the *ab initio* spectrum has a negative band peak at 208 nm and a shoulder around 220 nm. The diabatic spectrum for this peptide displays the characteristic double minimum with negative peaks at 207 and 222 nm, with the latter peak being slightly more intense than the former. For peptide **2l** [$-62^\circ, -41^\circ$] and in the *ab initio* spectrum, the negative band is at 205 nm and the band shoulder is around 220 nm, whereas the diabatic spectrum features a double minimum with negative peaks at 205 and 224 nm, with the former peak being more intense than the latter peak (vice versa to peptide **2k** [$-60^\circ, -60^\circ$]).

For the discussion we focus on the third α -helical dimer **2l** [$-62^\circ, -41^\circ$], the Barlow-Thornton helix whose geometry is representative of structures in the Protein Data Bank. For dimer **2l** [$-62^\circ, -41^\circ$] the H_{13} , H_{14} , H_{23} and H_{24} couplings are -17, -37, -52 and -292 cm^{-1} for the *ab initio* and -359, -660, -131 and -363 cm^{-1} for the diabatic calculation. The dot product of the un-normalised TDMs for these couplings are, respectively, -0.003, -0.547, 0.629 and 0.984 D^2 for the monomer and -0.035, -0.116, 0.618 and -3.997 D^2 for the dimer. Using the rationale outlined in the Methods section, H_{13} , H_{14} , H_{23} and H_{24} may have non-Coulombic interactions in the diabatic representation, with H_{14} having the largest non-Coulombic interaction, and a significant Coulombic interaction, in the diabatic representation. Coupling 2–4 may also have a significant Coulombic interaction in the diabatic representation. These coupling and dot

product values are close to those for dimer **2k** [$-60^\circ, -60^\circ$], reflecting the closeness of the secondary structure of dimers **2k** [$-60^\circ, -60^\circ$] and **2l** [$-62^\circ, -41^\circ$] (both have a φ angle of -60°), except for couplings 1–3 and 1–4. For dimer **2l** [$-62^\circ, -41^\circ$] the extent of non-Coulombic interaction in the H_{14} coupling may be less than for dimer **2k** [$-60^\circ, -60^\circ$], with the H_{13} coupling in dimer **2l** [$-62^\circ, -41^\circ$] having more non-Coulombic interaction than in dimer **2k** [$-60^\circ, -60^\circ$].

β strand

The experimental CD spectrum for a β strand features a positive band at around 195 nm and a negative band at around 215–220 nm. The **2d** [$-135^\circ, 135^\circ$] peptide has a computed CD spectrum using the diabatic inter-amide interactions that does not possess this feature in the 215–220 nm region, whereas the CD spectrum using the *ab initio* inter-amide interactions displays a degree of the correct character in this region (Figure 3) as well as a positive band at 186 nm.

Due to the arbitrariness of the phase of the wave functions, the signs of the excitonic coupling from the diabatization are not well defined. A bimodal distribution (plus or minus) for the excitonic coupling is often observed (see, e.g. Figure 14 from Jiang et al. [56]). Thus, we usually select the signs of the diabatic coupling such that the computed diabatic CD spectra best agree with experimental and/or CD spectra computed using DichroCalc. To examine the differences between the two predicted spectra, the sign and/or magnitude of the three coupling interactions discussed above (1–4, 2–3 and 2–4; respectively, $n_1\pi_1^*$ with $\pi_{nb2}\pi_2^*$, $\pi_{nb1}\pi_1^*$ with $n_2\pi_2^*$ and $\pi_{nb1}\pi_1^*$ with $\pi_{nb2}\pi_2^*$) in the diabatic Hamiltonian were varied so that the negative band at 215–220 nm could be observed in the computed CD spectra (Figure 4). Table 4

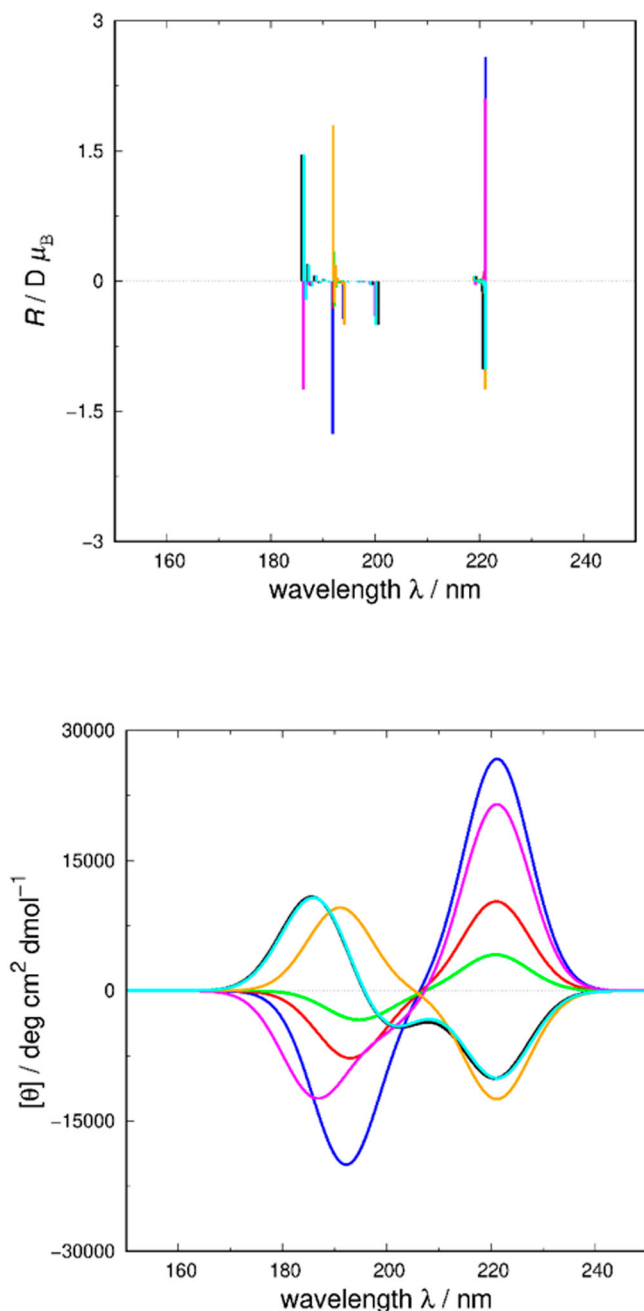


Figure 4. Rotational strength line spectra (upper panel) and CD spectra with band shapes from convolution with Gaussian functions of FWHM of 9.0 nm (lower panel) for peptide **2d** [$\varphi = -135^\circ$, $\psi = 135^\circ$]. Spectra computed using the *ab initio* (black line), diabatic (blue), diabatic with H_{14} modified (red), diabatic with H_{23} modified (green), diabatic with H_{24} modified (magenta), diabatic with H_{14} and H_{23} modified (yellow) and diabatic with H_{14} , H_{23} and H_{24} modified (cyan) Hamiltonians.

lists the *ab initio*, diabatic and the matrix elements of the diabatic Hamiltonian that were modified to compute the spectra for peptide **2d** [-135° , 135°] shown in Figure 4. Modifying H_{14} only gives a less intense CD spectrum (compared to the diabatic), centred around 206 nm and where the negative band at 193 nm and the positive

band at 221 nm are present (Figure 4, red line). Modifying H_{23} yields a spectrum similar to that computed from the modification of H_{14} but less intense with the negative band slightly red shifted to 194 nm (Figure 4, green line). Modifying H_{24} only gives a spectrum that is like that of the diabatic spectrum but where the negative band is blue shifted (from 192 to 187 nm), and the positive band (at 221 nm) is less intense (Figure 4, magenta line). A positive band at 191 nm and a negative band at 221 nm are computed when both H_{14} and H_{23} are modified (Figure 4, yellow line). When, in addition, H_{24} is modified the computed spectrum resembles that of the spectrum computed using the *ab initio* Hamiltonian (Figure 4, cyan and black lines).

H_{13} , H_{14} , H_{23} and H_{24} have *ab initio* couplings of 72, -20 , 292 and -907 cm^{-1} , and diabatic couplings of 109, -279 , -244 and -89 cm^{-1} (Table 2). The dot product of the un-normalised transition dipole moments for these couplings are, respectively, -0.032 , -0.350 , -0.342 and -2.037 D^2 for the monomer and -0.036 , 0.407 , -0.399 and 2.842 D^2 for the dimer (Table 3). Using the rationale outlined in the Methods section, H_{14} , H_{23} and H_{24} may have non-Coulombic interactions in the diabatic representation, with H_{24} having significant non-Coulombic interactions and additional Coulombic interactions in the diabatic representation. As there are little differences in the coupling values and dot products for the H_{13} coupling, this is likely a Coulombic interaction.

310 helix

The two computed spectra for diamide **2i** [-74° , -4°] both feature an intense couplet centred around 197 nm for the *ab initio* spectrum and centred around 190 nm for the diabatic spectrum (Figure 3), which has the relatively more intense couplet of the two spectra. For the *ab initio* spectrum, the couplet has a positive peak at 189 nm and a negative peak at 203 nm, which has a shoulder region around 220 nm. The couplet in the diabatic spectrum consists of a negative peak at 184 nm and a positive peak at 197 nm. This spectrum also features a relatively shallow and broad negative band around 218 nm (where the shoulder in the *ab initio* spectrum is situated).

The couplings listed in Table 2 show a difference for all four between the *ab initio* and the diabatic, where H_{13} , H_{14} , H_{23} and H_{24} are 15, -36 , 161 and -662 cm^{-1} for the former and 289, -148 , 230 and 284 cm^{-1} for the latter. The dot product of the un-normalised transition dipole moments for these couplings are, respectively, -0.013 , -0.454 , 0.615 and -2.985 D^2 for the monomer and 0.011 , -0.289 , -0.730 and 3.175 D^2 for the dimer. Thus, H_{14} , H_{24} , H_{23} and H_{24} may have non-Coulombic interactions in the diabatic representation.

Table 4. Inter-amide interactions (cm^{-1} rounded to integer values) for diamide **2d** [$\varphi = -135^\circ$, $\psi = 135^\circ$] and used to compute the spectra in Figure 2.

Matrix element index	H_{13}	H_{14}	H_{23}	H_{24}
Hamiltonian	$n_1\pi_1^*:n_2\pi_2^*$	$n_1\pi_1^*:\pi_{nb2}\pi_2^*$	$\pi_{nb1}\pi_1^*:n_2\pi_2^*$	$\pi_{nb1}\pi_1^*:\pi_{nb2}\pi_2^*$
<i>Ab initio</i>	72	-20	292	-907
Diabatic	109	-279	-244	-89
Modified diabatic	109	-3	244	-894

The experimental CD spectrum of an octameric sequence of L amino acids in a right-handed 3_{10} -helical structure (with mean φ and ψ angles of -54° and -32° , respectively) displays a negative band at 207 nm with a shoulder centred near 222 nm [51]. In addition to a slight positive band at 195 nm and negative band at 184 nm. The *ab initio* spectrum presented here is closest to the experimental spectrum. We note that the octamer has a structure close to diamide **2j** [-48° , -57°] (discussed above) and not diamide **2i** [-74° , -4°], and that the diabatic CD spectrum for **2j** [-48° , -57°] is closer to the experimental CD spectrum of the 3_{10} -helical octamer.

Polyproline II

The *ab initio* spectrum for diamide **2m** [-75° , 145°] features an intense couplet centred around 188 nm with a negative band at 181 nm and a positive band at 194 nm (Figure 3). The spectrum also has a shallow negative band at 208 nm and a weak and broad positive band at 220 nm. In contrast, the diabatic spectrum possesses a positive band at 186 nm and two negative bands at 199 and 220 nm with the latter negative band being broader and shallower than the former.

The coupling values for H_{13} , H_{14} , H_{23} and H_{24} are, respectively, 31, -7 , 16 and -971 cm^{-1} for the monomer and 81, 356, 145 and 29 cm^{-1} for the dimer (Table 2). The dot product of the un-normalised TDMs for these couplings are, respectively, -0.002 , 0.005 , -0.619 and -2.865 D^2 for the monomer and 0.048 , 0.543 , -0.610 and -4.660 D^2 for the dimer (Table 3). As noted for the β strand dimer (**2d** [-135° , 135°]), H_{14} , H_{23} and H_{24} may have non-Coulombic interactions in the diabatic representation, with H_{24} having significant non-Coulombic interactions and additional Coulombic interactions in the diabatic representation. As there are little differences in the coupling values and dot products for the H_{13} coupling, this is likely a Coulombic interaction.

The experimental CD spectrum of a polyproline II structure (poly-L-proline) features a negative peak in the region around 200 nm and a positive peak at 220 nm [57]. The computed diabatic CD spectrum features a negative band at 199 nm, in agreement with experiment, and a negative band at 220 nm, in disagreement with experiment.

Future work and outlook

Future work will involve derivation of an empirical correction term (for inter-amide couplings) to incorporate into the DichroCalc NMA4FIT2 parameter set, based on the diabatisation work presented here. The empirical correction will follow the work of Scholes and Ghiggino [58], who defined a distance-based exponential term for short-range, non-Coulombic interactions between two chromophores, and Voityuk [55], who derived expressions to compute the electronic coupling based on vertical transition energies, transition dipole moments (of the system and of individual chromophores) and the relative orientation of the chromophores or individual transition dipole moments. The correction term will account for the short-range, non-Coulombic interactions between excited states on neighbouring amides, an interaction currently neglected in the NMA4FIT2 parameter set and which has been shown, in this study, to be a significant contributor toward computing more accurate CD spectra, e.g. the double minimum for an α -helical polypeptide.

Acknowledgements

We are grateful for access to the University of Nottingham's high-performance computer. J. H. is supported by the Royal Academy of Engineering under the Chairs in Emerging Technologies scheme. H.D. is supported by the NSFC Research Fund for International Young Scientists (grant number 21850410456).

Data availability statement

The DichroCalc code is available at <https://github.com/dmrogers75/DichroCalc>. Data, example files and other code used in this study are available in a folder named 'diabatisation'.



Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by National Natural Science Foundation of China: [Grant Number 21850410456]; Royal Academy of Engineering: [Grant Number CiET2021_17].

ORCID

David M. Rogers  <http://orcid.org/0000-0003-2167-113X>
 Hainam Do  <http://orcid.org/0000-0003-4239-6157>
 Jonathan D. Hirst  <http://orcid.org/0000-0002-2726-0983>

References

- [1] J.D. Hirst, *J. Phys. Chem. A*, **125**, 8345–8346 (2021). doi:10.1021/acs.jpca.1c07974.
- [2] N.J. Greenfield, *Nat. Protoc.* **1**, 2876–2890 (2006). doi:10.1038/nprot.2006.202.
- [3] N. Sreerama and R.W. Woody, *Anal. Biochem.* **287**, 252–260 (2000). doi:10.1006/abio.2000.4880.
- [4] S.E.F. Spencer and A. Rodger, *Anal. Meth.* **13**, 359–368 (2021). doi:10.1039/d0ay01645d.
- [5] A. Micsonai, F. Wien, L. Kernya, Y.H. Lee, Y. Goto, M. Réfrégiers and J. Kardos, *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3095–E3103 (2015). doi:10.1073/pnas.1500851112.
- [6] L. Mavridis and R.W. Janes, *Bioinformatics.* **33**, 56–63 (2017). doi:10.1093/bioinformatics/btw554.
- [7] G. Nagy, M. Igaev, N.C. Jones, S.V. Hoffmann and H. Grubmuller, *J. Chem. Theory Comput.* **15**, 5087–5102 (2019). doi:10.1021/acs.jctc.9b00203.
- [8] J. Kaminský, J. Kubelka and P. Bour, *J. Phys. Chem. A*, **115**, 1734–1742 (2011). doi:10.1021/jp110418w.
- [9] J. Seibert, C. Bannwarth and S. Grimme, *J. Am. Chem. Soc.* **139**, 11682–11685 (2017). doi:10.1021/jacs.7b05833.
- [10] L.Y. Zhao, J.X. Zhang, Y.L. Zhang, S. Ye, G.Z. Zhang, X. Chen, B. Jiang and J. Jiang, *JACS Au*, **1**, 2377–2384 (2021). doi:10.1021/jacsau.1c00449.
- [11] D.M. Rogers, S.B. Jasim, N.T. Dyer, F. Auvray, M. Réfrégiers and J.D. Hirst, *Chem.* **5**, 2751–2774 (2019). doi:10.1016/j.chempr.2019.07.008.
- [12] R.W. Woody, *J. Chem. Phys.* **49**, 4797–4806 (1968). doi:10.1063/1.1669962.
- [13] P.M. Bayley, E.B. Nielsen and J.A. Schellman, *J. Phys. Chem.* **73**, 228–243 (1969). doi:10.1021/j100721a038.
- [14] J.D. Hirst, D.M. Hirst and C.L. Brooks III, *J. Phys. Chem.* **100**, 13487–13491 (1996). doi:10.1021/jp960597y.
- [15] J.D. Hirst, D.M. Hirst and C.L. Brooks III, *J. Phys. Chem. A*, **101**, 4821–4827 (1997). doi:10.1021/jp970675x.
- [16] N.A. Besley and J.D. Hirst, *J. Phys. Chem. A*, **102**, 10791–10797 (1998). doi:10.1021/jp982645f.
- [17] N.A. Besley and J.D. Hirst, *J. Am. Chem. Soc.* **121**, 8559–8566 (1999). doi:10.1021/ja990064d.
- [18] J.D. Hirst, *J. Chem. Phys.* **109**, 782–788 (1998). doi:10.1063/1.476617.
- [19] J.D. Hirst and N.A. Besley, *J. Chem. Phys.* **111**, 2846–2847 (1999). doi:10.1063/1.479563.
- [20] N.A. Besley and J.D. Hirst, *J. Am. Chem. Soc.* **121**, 9636–9644 (1999). doi:10.1021/ja990627l.
- [21] R.W. Woody and N. Sreerama, *J. Chem. Phys.* **111**, 2844–2845 (1999). doi:10.1063/1.479562.
- [22] B.M. Bulheller and J.D. Hirst, *Bioinformatics.* **25**, 539–540 (2009). doi:10.1093/bioinformatics/btp016.
- [23] B.M. Bulheller, A. Rodger and J.D. Hirst, *Phys. Chem. Chem. Phys.* **9**, 2020–2035 (2007). doi:10.1039/b615870f.
- [24] B. Villavicencio, R. Ligabue-Braun and H. Verli, *J. Chem. Inf. Model.* **62**, 927–935 (2022). doi:10.1021/acs.jcim.1c01001.
- [25] Z. Li and J.D. Hirst, *Chem. Phys. Lett.* **758**, 137935 (2020). doi:10.1016/j.cplett.2020.137935.
- [26] N. Hildebrand, M. Michaelis, N. Wurzler, Z. Li, J.D. Hirst, A. Micsonai, J. Kardos, A. Gil-Ley, G. Bussi, S. Koppen, M. Delle Piane and L.C. Ciacchi, *ACS Biomater. Sci. Eng.* **4**, 4036–4050 (2018). doi:10.1021/acsbiomaterials.8b00819.
- [27] J.D. Hirst and B.J. Persson, *J. Phys. Chem. A*, **102**, 7519–7524 (1998). doi:10.1021/jp982423h.
- [28] N.A. Besley, M.-J. Brienne and J.D. Hirst, *J. Phys. Chem. B*, **104**, 12371–12377 (2000). doi:10.1021/jp0024524.
- [29] M.-P. Gaigeot, N.A. Besley and J.D. Hirst, *J. Phys. Chem. B*, **115**, 5562–5535 (2011). doi:10.1021/jp111140f.
- [30] A.T.B. Gilbert and J.D. Hirst, *J. Mol. Struct. (THEOCHEM)*, **675**, 53–60 (2004). doi:10.1016/j.theochem.2003.12.038.
- [31] M.T. Oakley and J.D. Hirst, *J. Am. Chem. Soc.* **128**, 12414–12415 (2006). doi:10.1021/ja0644125.
- [32] M.T. Oakley, B.M. Bulheller and J.D. Hirst, *Chirality*, **18**, 340–347 (2006). doi:10.1002/chir.20264.
- [33] R.W. Woody, *Chirality*, **22**, E22–E29 (2010). doi:10.1002/chir.20857.
- [34] B.M. Bulheller, A.J. Miles, B.A. Wallace and J.D. Hirst, *J. Phys. Chem. B*, **112**, 1866–1874 (2008). doi:10.1021/jp077462k.
- [35] C.A. Mead and D.G. Truhlar, *J. Chem. Phys.* **77**, 6090 (1982). doi:10.1063/1.443853.
- [36] C.-P. Hsu, Z.-Q. You and H.-C.H. Chen, *J. Phys. Chem. C*, **112**, 1204–1212 (2008). doi:10.1021/jp076512i.
- [37] S. Jurinovich, L. Cupellini, C.A. Guido and B. Mennucci, *J. Comput. Chem.* **39**, 279–286 (2018). doi:10.1002/jcc.25118.
- [38] F. Plasser, A.J.A. Aquino, W.L. Hase and H. Lischka, *J. Phys. Chem. A*, **116**, 11151–11160 (2012). doi:10.1021/jp304725r.
- [39] H. Tamura, I. Burghardt and M. Tsukada, *J. Phys. Chem. C*, **115**, 10205 (2011). doi:10.1021/jp203174e.
- [40] R.J. Cave and M.D. Newton, *Chem. Phys. Lett.* **249**, 15 (1996). doi:10.1016/0009-2614(95)01310-5.
- [41] J. Aragón and A. Troisi, *Phys. Rev. Lett.* **114**, 026402 (2015). doi:10.1103/PhysRevLett.114.026402.
- [42] L. Stojanović and R. Crespo-Otero, *ChemPhotoChem.* **3**, 907 (2019). doi:10.1002/cptc.201900075.
- [43] M. Rivera, M. Dommert, A. Sidat, W. Rahim and R. Crespo-Otero, *J. Comput. Chem.* **41**, 1045–1058 (2020). doi:10.1002/jcc.26144.
- [44] A. Carreras, O. Uranga-Barandiaran, F. Castet and D. Casanova, *J. Chem. Theory Comput.* **15**, 2320–2330 (2019). doi:10.1021/acs.jctc.9b00019.
- [45] L. Blancafort and A.A. Voityuk, *J. Chem. Phys.* **140**, 095102 (2014). doi:10.1063/1.4867118.
- [46] J.A. Green, H. Asha, F. Santoro and R. Improta, *J. Chem. Theory Comput.* **17**, 405–415 (2021). doi:10.1021/acs.jctc.0c01100.
- [47] P. Manavalan and W.C. Johnson Jr, *Nature*, **305**, 831–832 (1983). doi:10.1038/305831a0.
- [48] J. Wu, J.T. Yang and C.-S.C. Wu, *Anal. Biochem.* **200**, 359–364 (1992). doi:10.1016/0003-2697(92)90479-Q.
- [49] J.D. Hirst, S. Bhattacharjee and A.V. Onufriev, *Farad. Discuss.* **122**, 253–267 (2003). doi:10.1039/B200714B.
- [50] N. Sreerama and R.W. Woody, *Protein Sci.* **12**, 384–388 (2003). doi:10.1110/ps.0235003.

- [51] C. Toniolo, A. Polese, F. Formaggio, M. Crisma and J. Kamphuis, *J. Am. Chem. Soc.* **118**, 2744–2745 (1996). doi: [10.1021/ja9537383](https://doi.org/10.1021/ja9537383).
- [52] Avogadro: an open-source molecular builder and visualization tool. Version 1.2.0. <http://avogadro.cc/>
- [53] M.D. Hanwell, D.E. Curtis, D.C. Lonie, T. Vandermeersch, E. Zurek and G.R. Hutchison, *J. Cheminf.* **4**, 17 (2012). doi: [10.1186/1758-2946-4-17](https://doi.org/10.1186/1758-2946-4-17).
- [54] D.J. Barlow and J.M. Thornton, *J. Mol. Biol.* **201**, 601–619 (1988). doi: [10.1016/0022-2836\(88\)90641-9](https://doi.org/10.1016/0022-2836(88)90641-9).
- [55] A.A. Voityuk, *J. Phys. Chem. C.* **118**, 1478–1483 (2014). doi: [10.1021/jp410802d](https://doi.org/10.1021/jp410802d).
- [56] L. Jiang, D.M. Rogers, J.D. Hirst and H. Do, *J. Chem. Theory Comput.* **16**, 5150–5162 (2020). doi: [10.1021/acs.jctc.0c00399](https://doi.org/10.1021/acs.jctc.0c00399).
- [57] J.L.S. Lopes, A.J. Miles, L. Whitmore and B.A. Wallace, *Prot. Sci.* **23**, 1765–1772 (2014). doi: [10.1002/pro.2558](https://doi.org/10.1002/pro.2558).
- [58] G.D. Scholes and K.P. Ghiggino, *J. Phys. Chem.* **98**, 4580–4590 (1994). doi: [10.1021/j100068a017](https://doi.org/10.1021/j100068a017).