chemistry

MDPI

Article

Dynamical Resolution of QM/MM Near-UV Circular Dichroism Spectra of Low-Symmetry Systems

Jérémy Morere [1,*], Tanguy Leyder [2], Catherine Michaux [2], Claude Millot [1], Emmanuelle Bignon [1,*] and Thibaud Etienne [1,*]

[1] Université de Lorraine, CNRS, LPCT, F-54000 Nancy, France; claude.millot@univ-lorraine.fr
[2] Laboratoire de Chimie Physique des Biomolécules (CPB), Chemistry Department, Namur Institute of Structured Matter (NISM), Namur Research Institute for Life Sciences (NARILIS), University of Namur (UNamur), 5000 Namur, Belgium; tanguy.leyder@unamur.be (T.L.); catherine.michaux@unamur.be (C.M.)
* Correspondence: jeremy.morere@univ-lorraine.fr (J.M.); emmanuelle.bignon@univ-lorraine.fr (E.B.); thibaud.etienne@univ-lorraine.fr (T.E.)

Abstract: Near-UV circular dichroism (CD) spectroscopy is a widely used method that provides, among others, information about the tertiary structure of biomolecular systems such as proteins, RNA, or DNA. Experimental near-UV CD spectra of proteins reflect the CD signals averaged over the many conformations that these systems can adopt. Theoretical approaches have been developed to predict such spectroscopic properties and link modeled conformations of complex biosystems to easily accessible experimental data, without having the resort to costly structural biology techniques. However, these predictions are mostly generated on the basis of a single experimental structure, missing the dynamic information reflecting the protein conformational variability. Here, we describe a complete reformulation of the theoretical foundations behind the prediction of CD spectra. We propose a QM/MM-based automated pipeline that generates an average near-UV CD spectrum from a given MD ensemble in a fast manner based on these theoretical considerations and further test it on protein systems. This pipeline has been implemented in an open-source program called `DichroProt`.

Keywords: electronic circular dichroism; multi-scale excited state calculations; matrix method; first-order perturbation theory

check for updates

Academic Editor: Felix Plasser

Received: 31 January 2025
Revised: 14 March 2025
Accepted: 9 April 2025
Published: 16 April 2025

Citation: Morere, J.; Leyder, T.; Michaux, C.; Millot, C.; Bignon, E.; Etienne, T. Dynamical Resolution of QM/MM Near-UV Circular Dichroism Spectra of Low-Symmetry Systems. *Chemistry* **2025**, *7*, 63. https://doi.org/10.3390/chemistry7020063

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Proteins are optically active chiral molecules, meaning that their light absorption depends on polarization. One manifestation of optical activity is electronic circular dichroism (CD). Proteins are structured at different levels—secondary, tertiary and quaternary—each of which has an effect on the CD spectrum. In other words, protein structural change alters the CD spectrum. Depending on the spectral region, we have access to different structural information [1]. The ultraviolet (UV) CD spectrum is divided into three regions. In far-UV, i.e., below 250 nm, peptide bonds are active. This provides information about the arrangement of the backbone and thus about the secondary structure (i.e., proportion of α-helix, β-sheet, etc.). The aromatic side chains and the disulfide bridge appear in the near-UV (250–300 nm) range. Side chain absorption bands are strongly influenced by the tertiary and quaternary structure. The last region is the UV-visible (300–700 nm) range, where non-protein chromophores contribute, such as porphyrins for example.

The circular dichroism measures the difference of extinction coefficient between left and right circularly polarized light at a given wavelength λ. This quantity, noted $\Delta\epsilon(\lambda)$, expressed in L·mol^{-1}·cm^{-1}, reads

$$\Delta\epsilon(\lambda) := \epsilon_L(\lambda) - \epsilon_R(\lambda). \tag{1}$$

In contrast to the UV absorption spectrum, the CD spectrum can exhibit both positive and negative peaks. The CD is positive when the left polarized light is more absorbed and negative otherwise. Other physical quantites can be used, such as the molar ellipticity $\theta(\lambda) := 3298\Delta\epsilon(\lambda)$ [2]. The mean residue ellipticity is obtained by dividing the molar ellipticity by the number of residues in a protein. The integration of a CD spectrum band gives the rotational strength expressed in the CGS system of units (erg·esu·cm·G^{-1}).

$$\mathcal{R}_i := \frac{\hbar c}{16\pi^2 \mathcal{N}_a} \int_{\lambda_{i,1}}^{\lambda_{i,2}} \frac{\Delta\epsilon(\lambda)}{\lambda} \mathrm{d}\lambda, \tag{2}$$

where $\mathcal{N}_a$ is the Avogadro number, $\hbar$ is the reduced Planck constant, and c is the velocity of light in vacuum. The rotational strength in CD spectroscopy is equivalent to the oscillator strength in UV spectroscopy. The first quantum theory of rotational strength was developed by Rosenfeld [3]. If Ψ_1 and Ψ_2 are two real-valued wavefunctions, the Rosenfeld formula gives the rotational strength of a transition from state $|\Psi_1\rangle$ to state $|\Psi_2\rangle$ as

$$\mathcal{R}_{1\to 2} := \mathrm{Im}(\langle\Psi_1|\,\hat{\mu}\,|\Psi_2\rangle \cdot \langle\Psi_2|\,\hat{m}\,|\Psi_1\rangle), \tag{3}$$

where $\hat{m}$ is the magnetic dipole moment operator and $\hat{\mu}$ is the electric dipole moment operator. This equation can be rewritten in two gauge formulations. If the electric moment is expressed with the position operator, a change in coordinate origin induces a change in rotational strength. This formulation, called the length gauge, is not origin gauge-invariant. Otherwise, expressing the electric moment with the gradient operator leads to an independence of the origin choice [4]. This velocity gauge is a simpler alternative to the use of magnetic-field-dependent "gauge-including atomic orbitals" (GIAOs) [5–7]. Pedersen et al. have shown that when using finite basis sets, the velocity gauge retains the origin-invariance but this leads to somewhat large values of the gyration tensor G′ at zero frequency and modifies the velocity gauge by subtracting G′ ($\omega = 0$) from the calculations at other frequencies [8].

The large number of nuclei and electrons in proteins makes conventional wavefunction-based or electron-density-based methods impractical for accurately approximating $|\Psi_1\rangle$ and $|\Psi_2\rangle$. Some approximations have been made to push back the limits to around thousands of atoms [9,10], but this is not enough for huge proteins. Several groups have extended Rosenfeld's approach over the years [11–21]. The Frenkel exciton model has been developed to obtain optical properties of aggregates and solids [22–24]. It is largely used for J-aggregates, H-aggregates, or for molecular crystals [25–27]. In the context of the exciton model, different authors use different underlying assumptions, for example the use of energies of states or excitation energies, or the use of coupling between point electric transition dipoles or between generalized molecular charge densities [28–36]. The Frenkel exciton model was also adapted by Tinoco to polymer systems [37]; this method is known as the first-order perturbation theory. This approach was quickly extended to all orders by considering an eigenvalue problem, also called the matrix method [38,39]. The matrix method offers computational efficiency and ease of implementation but lacks the physical interpretability of the perturbation method. The state of the art presents a variety of formulas to compute rotational strength. One purpose of the present work is to reformulate the problem on a rigorous mathematical basis.

There are software programs that calculate protein CD spectra, such as DichroCalc [40], PDB2CD [41], and SESCA [42]; however, these methods use a certain degree of parametrization of their systems. PDMD2CD [43], an extension of PDB2CD, averages CD spectra over a molecular dynamic (MD) trajectory researching the closest similarity with structures of a bank of precalculated CD spectra. Such parametric methods have recently been reviewed [44]. Beyond such approaches and avoiding any parametrization, this study proposes to develop an automated pipeline that, from MD trajectories, computes a near-UV CD spectrum prediction from both methods (matrix method and first-order perturbation theory). As depicted in Figure 1, our pipeline automatically extracts conformations from an input MD trajectory, identifies the chromophores present in the structure, performs time-dependent density-functional theory/molecular mechanics (TD-DFT/MM) calculations separately on each chromophore, and generates an average near-UV CD prediction taking into account the coupling between the photoactive residues.

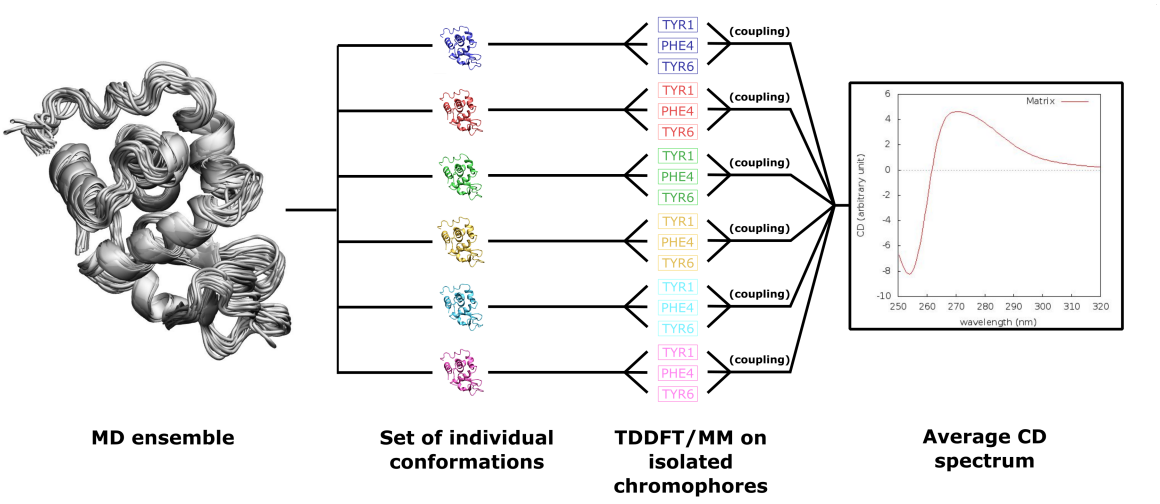


Figure 1. Workflow of the near-UV spectrum prediction pipeline. Conformations are extracted from a given MD ensemble, chromophores are identified in the structure and submitted to TD-DFT calculations separately taking into account their MM environment, and an average CD spectrum is generated using the coupling method as chosen by the user.

We will divide our work into two parts. First, we derive the matrix method and the first-order perturbation method for a general system of M optically active subsystems. Then, in the case of near-UV, we show how to obtain the rotational strength of a protein using both methods and compare a dynamically resolved use of these two methods with experimental CD results.

2. Theoretical Methods for Constructing the CD Spectrum

This section is a summary of the theoretical part of this paper related to the generation of the CD spectrum from electronic excited states' quantum-chemical calculations.

A fully detailed, self-contained reformulation of these methods and of the prerequisites that are necessary for building them is given in Supplementary Information SI, so that the present section is only a brief textual summary of the content of Supplementary Information SI. The reader who is more interested in the concrete and computational aspects of this article may completely ignore it.

2.1. General Considerations and Zero-Order Quantum States

In this contribution, we deal with N-electron complex systems containing M optically active groups of atoms. A first important thing to notice is that in both methods none of the groups share or exchange electrons or nuclei with another one.

Each group of atoms is represented by a clamped-nuclei Hamiltonian including an electronic kinetic energy term, a nucleus–electron potential energy term, a nucleus–nucleus potential energy term, and an electron–electron potential energy term. For every such group Hamiltonian, we consider $(A+1)$ mutually orthogonal, normalized quantum electronic states $(\varphi_{ia})_{a\in[\![0,A]\!]}$, and the corresponding $(A+1)$ real-valued energies $(\varepsilon_{ia} = h\nu_{ia})_{a\in[\![0,A]\!]}$, where ia refers to the ath quantum electronic state of the ith group, and h is the Planck constant.

A very fundamental step in the construction of the two methods is the construction of a reference ground state for the total system, $|\Psi_0^{(0)}\rangle$. For both methods, this reference state is the tensor product of all the ground states of the individual groups—any possible interaction between the groups is neglected for the reference ground state. By extension, zero-order excited states of the system—involving one or more groups being in an excited state—is also a tensor product of the states of the individual groups.

2.2. Coupling Between the Quantum States

The coupling between the groups is then introduced with the form of a $\langle\Psi_1^{(0)}|\hat{V}|\Psi_2^{(0)}\rangle$ quantity, where $\hat{V}$ is the potential energy operator of Coulomb nature, and $|\Psi_1^{(0)}\rangle$ and $|\Psi_2^{(0)}\rangle$ are any two reference states of the total system. In particular, couplings implying two distinct zero-order singly excited states of the total system—i.e., two states of the total system in which distinct groups are in an excited states, with one excited group per state—are of particular interest in this paper. They can be approximately evaluated using the so-called "dipole–dipole" approximation [45], where the coupling value between the $0\to a$ transition of the ith group and the $0\to b$ transition of the jth group, i.e.,

$$J_{ij}(a0,b0) = \frac{1}{4\pi\epsilon_0}\left[\frac{\boldsymbol{\mu}_i^{0\to a}\cdot\boldsymbol{\mu}_j^{0\to b}}{|\mathbf{X}_{ij}|^3} - 3\frac{\left(\mathbf{X}_{ij}\cdot\boldsymbol{\mu}_i^{0\to a}\right)\left(\mathbf{X}_{ij}\cdot\boldsymbol{\mu}_j^{0\to b}\right)}{|\mathbf{X}_{ij}|^5}\right]. \tag{4}$$

solely depends on dot products implying two-group inter-centroid position vectors ($\mathbf{X}_{ij}$), electric transition dipole moment vectors ($\boldsymbol{\mu}_i^{0\to a}$ and $\boldsymbol{\mu}_j^{0\to b}$), and two-group inter-centroid distances ($|\mathbf{X}_{ij}|$).

2.3. The Matrix Method

In the matrix method, an effective Hamiltonian matrix is built with, in its diagonal, the excitation energies of the individual groups and, out of the diagonal, the couplings implying two distinct singly excited states of the total system, evaluated using the dipole–dipole approximation. Eigenvectors of this matrix in the basis of singly excited states of the total system, i.e.,

$$\forall g\in[\![1,M\times A]\!],\ |\Psi_g\rangle = \sum_{i=1}^{M}\sum_{a=1}^{A}\mathbf{U}_{ia,g}\,|\Psi_{ia}^{(0)}\rangle\,, \tag{5}$$

are then interpreted as approximations to the excited states of the fully interacting total system, with corresponding $(h\nu_g)$ energies, i.e., the eigenvalues of the effective Hamiltonian matrix. They are introduced in the Rosenfeld equation, which establishes the rotatory strength of CD as the imaginary part of the dot product of electric and magnetic transition dipole moments.

Electric transition dipole moments are origin gauge-invariant, while in general, magnetic transition dipole moments are not. This means that if the origin of the reference frame is shifted, magnetic transition dipole moments may not remain invariant. We show, using the hypervirial theorem, that we can circumvent this problem regarding the rotatory strength, and provide an origin gauge-invariant expression of the rotatory strength for the matrix method:

$$\mathcal{R}_g = \frac{1}{\nu_g} \sum_{i,j=1}^{M} \sum_{a,b=1}^{A} \left[\mathbf{U}_{ia,g} \mathbf{U}_{jb,g} \nu_{ia} \boldsymbol{\mu}_i^{0\to a} \cdot \boldsymbol{m}_j^{b\to 0} + \frac{\pi}{2} \mathbf{U}_{ia,g} \mathbf{U}_{jb,g} \nu_{ia} \nu_{jb} \mathbf{X}_{ij} \cdot \left(\boldsymbol{\mu}_i^{0\to a} \times \boldsymbol{\mu}_j^{0\to b} \right) \right]. \quad (6)$$

In (6), the $\boldsymbol{m}_j^{b\to 0}$ vectors are the groups' intrinsic magnetic transition dipole moments.

2.4. First-Order Non-Degenerate Perturbation Theory

In Tinoco's first-order perturbation theory [37], first-order correction to the reference (zero-order) ground and singly excited states of the total system are computed using $\hat{V}$ and reference ground, singly excited, and doubly excited states of the total system, together with the corresponding total energies. Once these first-order corrections are added to the reference ground and singly excited states of the total system, the obtained first-order states are introduced in Rosenfeld's equation, leading to an expression for the rotatory strength that is readily origin gauge-invariant. Coupling terms are evaluated using the dipole–dipole approximation, giving, for every k in $[\![1, M]\!]$ and for every c in $[\![1, A]\!]$,

$$\begin{aligned} \mathcal{R}_{kc} = \mathrm{Im} \Bigg[& \boldsymbol{\mu}_k^{0\to c} \cdot \boldsymbol{m}_k^{c\to 0} - 2 \sum_{i \in \mathbb{M}_k} \sum_{a=1}^{A} \frac{J_{ki}(c0, a0)}{h(\nu_{ia}^2 - \nu_{kc}^2)} \left(\nu_{kc} \boldsymbol{\mu}_k^{0\to c} \cdot \boldsymbol{m}_i^{a\to 0} + \nu_{ia} \boldsymbol{\mu}_i^{0\to a} \cdot \boldsymbol{m}_k^{c\to 0} \right) \\ & - \sum_{i \in \mathbb{M}_k} \frac{J_{ki}(c0, 00)}{h \nu_{kc}} \left(\boldsymbol{\mu}_k^{c\to c} - \boldsymbol{\mu}_k^{0\to 0} \right) \cdot \boldsymbol{m}_k^{c\to 0} \Bigg] \\ & - 2\pi \sum_{i \in \mathbb{M}_k} \sum_{a=1}^{A} J_{ki}(c0, a0) \frac{\nu_{kc} \nu_{ia}}{h(\nu_{ia}^2 - \nu_{kc}^2)} (\mathbf{X}_i - \mathbf{X}_k) \cdot \left(\boldsymbol{\mu}_i^{0\to a} \times \boldsymbol{\mu}_k^{0\to c} \right) \end{aligned} \quad (7)$$

where

$$J_{ki}(c0, a0) = \langle \Psi_{kc}^{(0)} | \hat{V} | \Psi_{ia}^{(0)} \rangle \text{ and } J_{ki}(c0, 00) = \langle \Psi_{kc}^{(0)} | \hat{V} | \Psi_{i0}^{(0)} \rangle = \langle \Psi_{kc}^{(0)} | \hat{V} | \Psi_{0}^{(0)} \rangle . \quad (8)$$

In (7) above, $\mathbb{M}_k$ is equal to $[\![1, M]\!] \setminus \{k\}$.

3. Materials and Methods

3.1. Molecular Dynamics Simulations

All MD simulations were performed using the Amber suite of programs. Two systems for which we produced the experimental near-UV CD data were used to test our protocol for the prediction of spectra from MD simulations conformational ensembles: the Lysozyme and the Trp-Cage Miniprotein Construct TC5b. The choice of the systems was based on the fact that the Lysozyme and the Trp-Cage exhibit different levels of complexity: the first one is a 129 amino acid protein harboring 16 chromophores including CYS-CYS bridges, and the second one is a 20-residue-long peptide featuring 3 aromatic rings.

The starting structures were downloaded from the PDB (ID 2YVB [46] and 1L2Y [47], for the Lysozyme and the Trp-Cage respectively), and protonation states of the amino acids were assigned with respect to their pKa as predicted by propka3 [48]. Parameters were taken from the Amber ff14SB force field [49] for the protein residues and the system was soaked into a TIP3P [50] water box shaped in a truncated octahedron and using a 15 Å buffer, resulting in systems of 30,879 and 11,363 atoms for the Lysozyme and Trp-Cage,

respectively. The total charge of the system was neutralized by adding Cl counterions. Each system was subjected to a 10,000-step minimization, followed by a 400 ps NVT thermalization run to increase the temperature to 300 K and to equilibrate the solvent, and a 100 ns NPT equilibration to relax the entire system. Production runs were then carried out, with the total sampling for each system amounting to 3 µs. The Langevin thermostat was used to keep the temperature at 300 K along the simulations, and the Particle Mesh Ewald approach [51] was applied for long-range electrostatics. A 4 fs time step was used for the production runs, which was allowed by the use of the Hydrogen Mass Repartitioning algorithm [52].

A cluster analysis was performed with the cpptraj module of Amber for each system based on the RMSD of the protein residues, in order to characterize the main conformations sampled and determine if we could predict the near-UV CD spectrum directly from a representative frame of the main clusters.

3.2. Automated Pipeline from MD Data to Circular Dichroism Spectrum

The pipeline has been implemented in a Fortran90 program called `DichroProt` [53]—see Figure 1. The principle of this program is to run in parallel electrostatic embedding QM/MM calculations on each separate chromophore for each frame, and to generate from this divided information an average theoretical CD spectrum, which can be computed taking into account (or not) the coupling between the chromophore. The user can decide to include the coupling using either the matrix method or the perturbation method. It provides an easy, fast, and automated way to generate a CD spectrum from an MD ensemble of configurations. This program works with the Amber/Gaussian interface and is built in three steps:

(i) First, the protein conformations are extracted from the MD ensemble of configurations, with a user-defined increment. This step is performed by the cpptraj module of Amber, and simply requires compatible topology and MD trajectory files. During the first step, a working directory is created with the input files necessary for the QM/MM calculations. The coordinates of photoactive residues (phenylalanine, tyrosine, tryptophan, and disulfide bridge) are individually extracted from each snapshot and a QM/MM Gaussian09/Amber16 input is generated for each, including only the side chain of the residue in the QM region. The QM and MM frontier is set between the Cα and Cβ of the residue and treated by the link atom approach [54]. In the present work, snapshots were taken every 20 ns, amounting for a total of 150 frames for each system that were used for QM/MM input preparation. The coordinates of 12 aromatic residues and 4 disulfide bridges were extracted for the Lysozyme, and these of 2 aromatic residues were taken for Trp-Cage system.

(ii) In the second step, the QM/MM calculations are launched in parallel through the Amber16/Gaussian09 interface, to calculate transition energies and dipole moments of each individual chromophore. According to the residue, the QM region contains between 8 (disulfide bridge) and 18 atoms (tryptophan). The user can define the level of theory of the excited states calculation, and can modify the submission script (SLURM by default). In the present work, TD-DFT was employed as a compromise between computational cost and accuracy. Three broadly used functionals were tested: B3LYP [55], a hybrid functional; CAM-B3LYP [56], which incorporates long-range corrections; and ωB97X-D [57], which combines long-range corrections and empirical dispersion. The 6-311 + G** basis was used for each functional. By default, only the first five transitions were calculated to recover those present in the spectral band from 250 nm to 300 nm. To obtain the permanent electric dipole moments, needed for the perturbation method, one QM/MM calculation per targeted state is launched. Terms containing permanent moments contribute little to the rotational

strength. In this work, these contributions are computed; the user can disable this option and divide the computing cost by the number of transitions.

(iii) The last step merges Gaussian outputs into Formula (6) or (7). Aromatic residues and disulfide bridges contribute with the same weight to the circular dichroism. `DichroProt` generates a stick spectrum for each snapshot, then convolves Lorentzian functions with a width of 0.2 eV [58] with it. Both the matrix and perturbation formulas were implemented in the code for the coupling, and the user can choose to use one, the other, or both. All quantities are manipulated in atomic units; the rotatory strength is converted to Gaussian units (CGS) at the end.

It is worth noting that the transition magnetic moments in the Gaussian output are given in electron magnetic moment and not in atomic units. A factor $-2g_e = 2.0023193043$ is therefore applied, where g_e is the g-factor of the electron, to bring it back to atomic units. This can be directly checked by applying the Rosenfeld equation using the electric and magnetic transition dipole moments given by the Gaussian. It is customary to express the rotational strength in the CGS system of units, as in the output of Gaussian 09 (Revision D.01) software. It is then expressed in 10^{-40} erg·esu·cm·G^{-1}. The CGS system of units differs from the international one regarding the dimensions of certain quantities. In the CGS system, the electric and magnetic dipole moments have the same dimension, and they are expressed in two different units to distinguish them. Since the calculations are carried out in the atomic unit system, the coefficient 471.44 was used to switch the rotational strength to the CGS system.

3.3. Near-UV Circular Dichroism Spectroscopy

In order to increase the number of reference points in the near-UV spectra to test our approach, we performed experimental measurements on two additional systems. The near-UV (250–320 nm range) CD spectra of the hen-egg-white Lysozyme (purchased from Merck, Darmstadt, Germany) and the Trp-Cage mini-protein (purchased from Eurogentec, Seraing, Belgium) were recorded with an MOS-500 spectropolarimeter at 20 °C in 20 mM Tris-HCl pH 8.0, using a 1 mm optical pathlength quartz Suprasil cell (Hellma), with protein concentrations of 1 mg/mL. Four scans were averaged, buffer baselines were subtracted, and corrected spectra were smoothed. The following parameters were used: 15 nm/min scanning rate, 2 nm bandwidth, 0.5 nm data pitch, and 2 s digital integration time.

4. Results and Discussion

4.1. Comparison of the Theoretical and Experimental CD Spectra

Near-UV CD spectra of the Lysozyme and Trp-Cage proteins were predicted from the generated MD ensembles, using the automated pipeline described above. Conformations were extracted with a 20 ns increment and each photoactive moiety (tyrosines, tryptophans, phenylalanines, and disulfide bridges) was individually subjected to a TDDFT calculation taking into account the MM electrostatic embedding. The two coupling methods (matrix and perturbation) were tested and compared to the CD prediction without coupling (called "uncoupled" below) and to the experimental data.

For the three functionals that were tested (B3LYP, CAM-B3LYP, and ωB97X-D), we observed important deviations of the predicted CD signal—see Figures S1 and S2. Of note, we applied the same protocol to two other systems for which near-UV CD spectra were available in the literature (the bovine basic pancreatic trypsin inhibitor and the photoactive yellow protein), that resulted in the same trends—see Figures S3–S5. The CD spectrum of a protein is highly sensitive to geometric variations [45], while the UV spectrum is less affected by small changes in protein geometry. In order to first confirm the robustness of each functional in predicting spectroscopic data, we ran UV spectra predictions revealing

that the B3LYP functional was failing at producing consistent trends from one MD frame to the other, while CAM-B3LYP and ωB97X-D showed much better results exhibiting similar curve shapes with a narrower and more refined global envelope—see Figure 2.

Given that the B3LYP functional showed poor performance in predicting robust trends for the UV spectrum of a protein, only the results obtained with the CAM-B3LYP and ωB97X-D functionals for the prediction of CD spectra are discussed below.

The superposition of the predicted average near-UV CD spectra with the experimental ones reveals strong differences in the shape of the CD signal—see Figure 3. For the Lysozyme protein, the experimental data exhibit negative values under 270 nm and a positive peak around 290 nm.

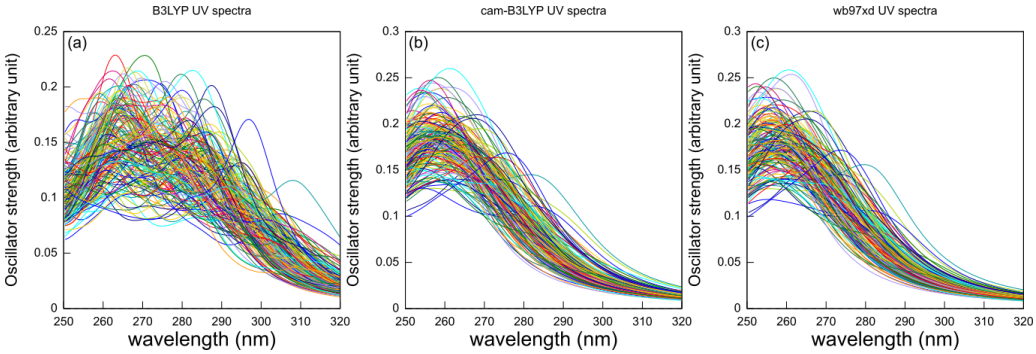


Figure 2. Influence of the functional on the UV spectrum of Lysozyme along the trajectory: superimposed predicted spectra of MD frames using (**a**) B3LYP, (**b**) CAM-B3LYP, and (**c**) ωB97X-D.

Theoretical predictions with both the CAM-B3LYP and ωB97X-D functionals fail to reproduce the positive peak. Concerning the Trp-Cage system, the experimental spectrum shows positive values with peaks at 250 and 280 nm. Here, the theoretical spectra show negative values below 270 nm, followed by a plateau around 0.

In order to understand why the theoretical predictions are inconsistent with the experimental data despite the in-depth review of the underlying theory described above, we further scrutinized the influence of the functional, the coupling method, and the protein geometry on the CD predictions.

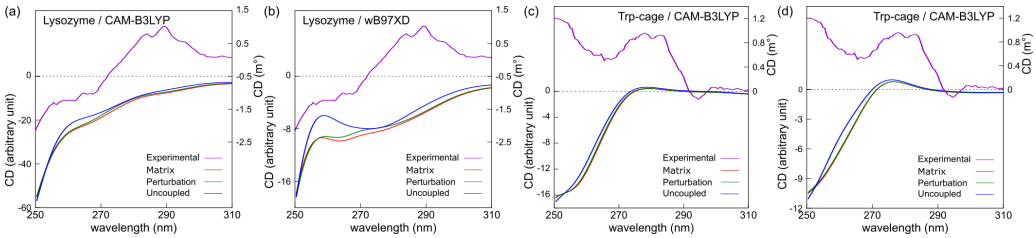


Figure 3. Comparison of the average theoretical and experimental spectra for the Lysozyme with CAM-B3LYP (**a**) or ωB97X-D (**b**), and for the Trp-Cage with CAM-B3LYP (**c**) or ωB97X-D (**d**). Left axis: theoretical CD values obtained using the matrix coupling method (red), the perturbation coupling method (green), and using the Gaussian output without coupling (blue). Right axis: experimental spectrum values.

4.2. The Functional and the Coupling Method Do Not Have a Critical Influence on the Predictions

The theoretical spectra obtained for the Lysozyme system show that, while deviating from the experimental spectrum, the theoretical CD signals exhibit the same global trend with both CAM-B3LYP or ωB97X-D functionals, with only a few local differences—see Figure 3a,b. Looking at the uncoupled spectra, the ωB97X-D results exhibit a clear minimum at 272 nm, unlike CAM-B3LYP, which exhibits a shoulder in the curve at the same wavelength. With CAM-B3LYP, the matrix and perturbation methods provide similar

trends that solely diverge from the uncoupled spectrum by being a bit more intense between 260 and 270 nm. With ωB97X-D, a more pronounced difference between the coupled and uncoupled spectra is noted, showing a stronger correlation between residues with this functional than with CAM-B3LYP, but this is not a tendency observed for all systems. The matrix and perturbation coupling methods give a similar intensity band and shape spectrum, but the minima disappear into a shoulder.

The Trp-Cage theoretical spectra are all extremely similar for both functionals. They all have a plateau from 300 nm to 270 nm and a small maximum around 280 nm, with the maximum being more pronounced in positive values with ωB97X-D. For this system, the coupled and uncoupled spectra are much more alike. Again, theoretical spectra diverge from the experimental spectrum. No feature from the experimental spectrum can be found in the theoretical ones, except for the final plateau at null CD values found above 295 nm wavelengths.

In the matrix method, the transition energies are computed as eigenvalues of the effective Hamiltonian. For a small coupling between residues, the off-diagonal elements of the matrix are very small compared to the diagonal element; this means that the eigenvalues are close to the isolated transition energy. For a stronger coupling, the eigenvalues will shift away from the isolated system. Unlike the matrix method, the perturbation method does not add correction to the energy and uses the transition energy from the Gaussian directly. For all proteins studied, the wavelength shift between matrix and perturbation methods is less that 1 nm. We can expect the shift to be larger for non-protein systems. Even weak coupling can lead to significant spectral changes, despite minor variations in transition energy—see Figure 4d.

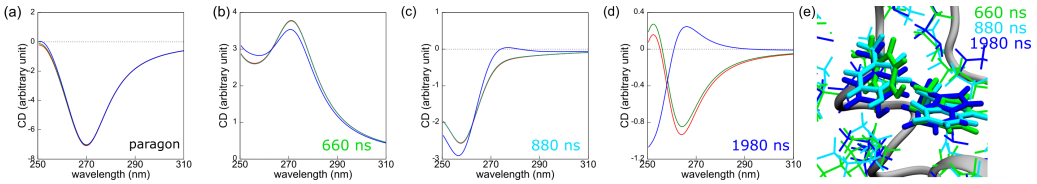


Figure 4. Paragon near-UV CD spectra for the major cluster identified for the Trp-Cage conformations (**a**); spectra of three snapshots at 660 (**b**), 880 (**c**), and 1980 ns (**d**); and geometry comparisons of TYR3 (residue on the left) and TRP6 (residue on the right) for the corresponding snapshots (**e**). For each snapshot, a superposition of spectra obtained with the matrix coupling method (red), the perturbation coupling method (green), and from the Gaussian output without coupling (blue) are showed.

Overall, the choice of functional and coupling method does not drastically influence the shape average spectra for the two considered systems. The same results are found for the BPTI and PYP additional systems tested—see Supplementary Information. However, a closer look at the evolution of the theoretical spectra generated along the MD trajectory reveals that the average results hide a huge diversity of CD signals and can show strong divergences depending on the coupling method used, as described in the following section.

4.3. Slight Conformational Deviations Drastically Re-Shape the CD Signal Prediction

In order have a better understanding of the effect of geometry on the predicted CD spectrum, we more closely assessed the presence of deviations among the different geometries sampled along the MD simulations. We discuss below what we observed for the CAM-B3LYP functional on the Trp-Cage system, which harbors only two photoactive residues: a tyrosine (TYR3) and a tryptophan (TRP6). We scrutinized the major conformation identified by clustering of the MD ensemble based on the RMSD of the protein, and the CD spectra of its representative structure (called below "paragon"). Due to their spatial proximity, a strong coupling between TYR3 and TRP6 was expected.

The identified major cluster accounts for 91% of all conformations in the MD ensemble. The CD spectrum corresponding to its representative structure (or paragon) was generated with and without coupling. Surprisingly, all theoretical spectra exhibit the same shape with an intense negative peak at 270 nm—see Figure 4a. Both coupled methods give the same spectra, with almost no correction to the intensity. Nevertheless, this paragon spectrum poorly represents the diversity of spectra in its cluster. Analyzing three randomly picked snapshots belonging to the major cluster reveals significant variability in produced CD signals within the same cluster—see Figure 4b–d .

In order to show the diversity of the CD signals computed for conformations belonging to the same cluster, we discuss in the following the results obtained for three different snapshots. The first one, corresponding to the geometry sampled after 660 ns of production, matches very well the experimental spectra, exhibiting positive values, and a minima at 260 nm followed by a peak at 270 nm—see Figure 4b. This conformation is relatively close to the crystallographic conformation, with an RMSD of 1.48 Å for the entire protein and of 0.39 for the aromatic residues (without the hydrogen atoms). However, for the second selected snapshot, corresponding to the geometry sampled at 880 ns, the spectra diverge hugely from the experimental reference, featuring strictly negative values with a peak at 256 nm. This conformation is slightly more different from the experimental structure with an RMSD of 1.75 Å for the protein and of 1.16 Å for the aromatic residues—see Figure 4c. Another interesting case is observed for the geometry sampled at 1980 ns snapshot. For this conformation, which exhibits an RMSD of 1.57 Å for the protein and 1.0 Å for the aromatic residues, the influence of the coupling is dramatic—see Figure 4d. Indeed, for this particular snapshot, the matrix and perturbation methods add a drastic correction to the uncoupled spectrum. Corrections change the sign and intensity of two transitions. The positive peak at 261 nm becomes negative and more intense, while the negative peak at 254 nm becomes positive and less intense. Despite the fact that the matrix and perturbation approaches are built with different objects, their results are similar, even with stronger coupling between residues. This drastic change comes from two nearly degenerate transitions, which exacerbate the intensity in the final formula of matrix and perturbation methods.

Although we expected some variability in the CD signals calculated from one frame to the other, such dramatic changes of the spectrum trend was surprising, especially when the RMSD values remain rather low. The spatial position of TRP6 and TYR3 do not change much between the selected snapshots discussed—see Figure 4e. Our results show that minor changes in the geometry result in substantial spectral variations, without any clear correlation between the tertiary structure and the spectrum. This behavior is also observed for the Lysozyme system and the BPTI and PYP proteins.

The near-UV spectra reflects the tertiary structure of a protein. Here, we show that although the Trp-Cage protein conformation stays rather stable along the MD trajectory, the CD signal changes drastically from one snapshot to the next, which results in a huge diversity of calculated spectra. Even though some individual conformations can lead to a theoretical CD spectrum that matches the experimental trends, the final average spectrum diverges a lot from the experimental reference. The failure may be due to the non-respect of one of the hypotheses made in the theoretical development. In particular, we can suppose that the proximity of the photoactive residues induces an exchange of electron density between them. If this is true, the framework used in this paper is no longer adequate. Overall, and despite being based on a solid theoretical framework, the proposed pipeline fails to recover the experimental near-UV spectrum of the studied systems as the CD signal calculation is too sensitive to the slightest conformational change in the protein. This important first observation is a crucial step towards future improvements that

could rely on the use of short runs of QM/MM-MD to further improve the chromophore structural description of each snapshot, on the use of a polarizable force field, or even on a parametrization refinement of the force field. While these options will necessitate further time-consuming efforts, and remain out of the scope of this contribution, this may constitute a major perspective of work in a second iteration.

5. Conclusions

In this work, we describe a complete reformulation of the theoretical framework for the prediction of CD spectra and we propose an automated pipeline for the QM/MM-based prediction of proteins' near-UV CD spectra from classical MD ensembles, based on these theoretical considerations.

To obtain an accurate theoretical near-UV CD spectrum of a molecular system, two approaches have been completely re-derived: the matrix method and the first-order perturbation method. The development has been carried out in order to be as general as possible, in other words to be practical with any system composed of optically active groups.

In the case of proteins, these methods have been implemented in an automated pipeline. This pipeline extracts a series of snapshots from a molecular dynamics trajectory, starts a batch of QM/MM calculations on isolated photoactive amino acids, and then introduces residue–residue coupling using one of the two methods. Each snapshot gives a near-UV theoretical spectrum, and averaging over all spectra would allow one to recover the dynamical effect in the spectrum. Probing the performances of this approaches revealed that, even though for a few individual snapshots the experimental spectrum could be reproduced, the variation in the computed CD signals is dramatic even for two very close conformations. The CD values are extremely sensitive to the slightest geometric variation, which prevents us from producing any accurate averaged spectrum. Investigations of the classical force field limitations for the prediction of correct near-UV spectra will constitute the next steps of this work, for example exploring how to integrate QM/MM-MD runs in the protocol.

The automated code we developed to compute and average near-UV spectra from MD simulations works without any need for parametrization. It takes into account all quantum effects within each photoactive residue, and the conformational variability information provided by MD data for any protein system. While the performances of near-UV CD predictions for proteins remain poor with this approach, this code can be used as a building block for further work to extend the spectral band for CD spectrum calculation outside the near-UV or to be used on other systems.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/chemistry7020063/s1, Supporting Information SI—Reformulation of the theory; Supporting Information SII—MD simulations of the BPTI and PYP systems: Table S1: Details of the MD simulations for the BPTI and PYP proteins; Figure S1: Trp-cage theoretical spectra for B3LYP, CAM-B3LYP and ωB97X-D; Theoretical CD values obtained using the matrix coupling method, the perturbation coupling method, and using the Gaussian output without coupling; Figure S2: Lysozyme theoretical spectra for B3LYP, CAM-B3LYP and ωB97X-D; Theoretical CD values obtained using the matrix coupling method, the perturbation coupling method, and using the Gaussian output without coupling; Figure S3: PYP theoretical spectra for B3LYP, CAM-B3LYP and ωB97X-D; Theoretical CD values obtained using the matrix coupling method, the perturbation coupling method, and using the Gaussian output without coupling; Figure S4: BPTI theoretical spectra from the 1pit NMR structure, for B3LYP, CAM-B3LYP and ωB97X-D; Theoretical CD values obtained using the matrix coupling method, the perturbation coupling method, and using the Gaussian output without coupling; Figure S5: BPTI theoretical spectra from the 5pti NMR structure, for B3LYP, CAM-B3LYP and ωB97X-D; Theoretical CD values obtained using the matrix coupling method, the

perturbation coupling method, and using the Gaussian output without coupling. References [59–64] are cited in the supplementary materials.

Author Contributions: Conceptualization, J.M., E.B., C.M. (Claude Millot), T.E.; methodology (theory), J.M., C.M. (Claude Millot), E.B., T.E.; methodology (experiments), T.L., C.M. (Catherine Michaux); software, J.M.; formal analysis, J.M., C.M. (Claude Millot), E.B.; resources (experimental), C.M. (Catherine Michaux); investigation, J.M., T.L., C.M. (Catherine Michaux), C.M. (Claude Millot), E.B., T.E.; writing—original draft preparation, J.M., C.M. (Claude Millot), E.B., T.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: T.L. and Ca.M. thank the Belgian National Fund for Scientific Research (F.R.S.-FNRS) for their FRIA (Fund for Research training in Industry and Agriculture) Doctoral grant and Senior Research Associate position, respectively.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Sreerama, N.; Woody, R.W. Computation and Analysis of Protein Circular Dichroism Spectra. *Methods Enzymol.* **2004**, *383*, 318–351. [CrossRef] [PubMed]
2. Woody, R.W. Theory of Circular Dichroism of Proteins. In *Circular Dichroism and the Conformational Analysis of Biomolecules*; Fasman, G.D., Ed.; Springer: Boston, MA, USA, 1996; pp. 25–67. [CrossRef]
3. Rosenfeld, L. Quantenmechanische Theorie der Natürlichen Optischen Aktivität von Flüssigkeiten und Gasen. *Z. Phys.* **1928**, *52*, 161–174. [CrossRef]
4. Pedersen, T.B.; Koch, H.; Boman, L.; Sánchez De Merás, A.M.J. Origin Invariant Calculation of Optical Rotation Without Recourse to London Orbitals. *Chem. Phys. Lett.* **2004**, *393*, 319–326. [CrossRef]
5. Bak, K.L.; Hansen, A.E.; Ruud, K.; Helgaker, T.; Olsen, J.; Jorgensen, P. Ab Initio Calculation of Electronic Circular Dichroism for Trans-Cyclooctene Using London Atomic Orbitals. *Theor. Chim. Acta* **1995**, *90*, 441–458. [CrossRef]
6. Cheeseman, J.R.; Frisch, M.J.; Devlin, F.J.; Stephens, P.J. Hartree-Fock and Density Functional Theory Ab Initio Calculation of Optical Rotation Using GIAOs: Basis Set Dependence. *J. Phys. Chem. A* **2000**, *104*, 1039–1046. [CrossRef]
7. Krykunov, M.; Autschbach, J. Calculation of Optical Rotation with Time-Periodic Magnetic-Field-Dependent Basis Functions in Approximate Time-Dependent Density-Functional Theory. *J. Chem. Phys.* **2005**, *123*, 114103. [CrossRef]
8. Autschbach, J.; Nitsch-Velasquez, L.; Rudolph, M. Time-Dependent Density Functional Response Theory for Electronic Chiroptical Properties of Chiral Molecules. In *Electronic and Magnetic Properties of Chiral Molecules and Supramolecular Architectures*; Naaman, R., Beratan, D.N., Waldeck, D., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; Volume 298, pp. 1–98. [CrossRef]
9. Grimme, S.; Bannwarth, C. Ultra-Fast Computation of Electronic Spectra for Large Systems by Tight-Binding Based Simplified Tamm-Dancoff Approximation (sTDA-xTB). *J. Chem. Phys.* **2016**, *145*, 054103. [CrossRef]
10. Seibert, J.; Bannwarth, C.; Grimme, S. Biomolecular Structure Information from High-Speed Quantum Mechanical Electronic Spectra Calculation. *J. Am. Chem. Soc.* **2017**, *139*, 11682–11685. [CrossRef]
11. Condon, E.U. Theories of Optical Rotatory Power. *Rev. Mod. Phys.* **1937**, *9*, 432–457. [CrossRef]
12. Condon, E.U.; Altar, W.; Eyring, H. One-Electron Rotatory Power. *J. Chem. Phys.* **1937**, *5*, 753–775. [CrossRef]
13. Kirkwood, J.G. On the Theory of Optical Rotatory Power. *J. Chem. Phys.* **1937**, *5*, 479–491. [CrossRef]
14. Kauzmann, W.J.; Walter, J.E.; Eyring, H. Theories of Optical Rotatory Power. *Chem. Rev.* **1940**, *26*, 339–407. [CrossRef]
15. Moffitt, W.; Fitts, D.D.; Kirkwood, J.G. Critique of the Theory of Optical Activity of Helical Polymers. *Proc. Natl. Acad. Sci. USA* **1957**, *43*, 723–730. [CrossRef]
16. Tinoco, I. Optical and Other Electronic Properties of Polymers. *J. Chem. Phys.* **1960**, *33*, 1332–1338. Erratum in *J. Chem. Phys.* **1961**, *34*, 1067–1067. https://doi.org/10.1063/1.1731643. [CrossRef]
17. Moffitt, W.; Woodward, R.B.; Moscowitz, A.; Klyne, W.; Djerassi, C. Structure and the Optical Rotatory Dispersion of Saturated Ketones. *J. Am. Chem. Soc.* **1961**, *83*, 4013–4018. [CrossRef]
18. Abramavicius, D.; Palmieri, B.; Voronine, D.V.; Šanda, F.; Mukamel, S. Coherent Multidimensional Optical Spectroscopy of Excitons in Molecular Aggregates; Quasiparticle Versus Supermolecule Perspectives. *Chem. Rev.* **2009**, *109*, 2350–2408. [CrossRef]
19. Schellman, J.A.; Oriel, P. Origin of the Cotton Effect of Helical Polypeptides. *J. Chem. Phys.* **1962**, *37*, 2114–2124. [CrossRef]

20. Tinoco, I.; Woody, R.W. Optical Rotation of Oriented Helices. II. Calculation of the Rotatory Dispersion of the Alpha Helix. *J. Chem. Phys.* **1960**, *32*, 461–467. [CrossRef]
21. Höhn, E.G.; Weigang, O.E. Electron Correlation Models for Optical Activity. *J. Chem. Phys.* **1968**, *48*, 1127–1137. [CrossRef]
22. Davydov, A.S. Theory of Absorption Spectra of Molecular Crystals. *Zhur. Eksptl. Teoret. Fiz.* **1948**, *18*, 210–218.
23. Davydov, A.S. The Theory of Molecular Excitons. *Sov. Phys. Uspekhi* **1964**, *7*, 145–178. [CrossRef]
24. Kasha, M. Energy Transfer Mechanisms and the Molecular Exciton Model for Molecular Aggregates. *Radiat. Res.* **1963**, *20*, 55–71. [CrossRef] [PubMed]
25. Spano, F.C. The Spectral Signature of Frenkel Polarons in H- and J-Aggregates. *Acc. Chem. Res.* **2010**, *43*, 429–439. [CrossRef] [PubMed]
26. Morrison, A.F.; You, Z.Q.; Herbert, J.M. Ab Initio Implementation of the Frenkel-Davydov Exciton Model: A Naturally Parallelizable Approach to Computing Collective Excitations in Crystals and Aggregates. *J. Chem. Theory Comput.* **2014**, *10*, 5366–5376. [CrossRef]
27. Morrison, A.F.; Herbert, J.M. Analytic Derivative Couplings and First Principles Exciton/Phonon Coupling Constants for an Ab Initio Frenkel-Davydov Exciton Model: Theory, Implementation, and Application to Compute Triplet Exciton Mobility Parameters for Crystalline Tetracene. *J. Chem. Phys.* **2017**, *146*, 224110. [CrossRef]
28. Prokhorenko, V.I.; Steensgaard, D.B.; Holzwarth, A.R. Exciton Dynamics in the Chlorosomal Antennae of the Green Bacteria Chloroflexus Aurantiacus and Chlorobium Tepidum. *Biophys. J.* **2000**, *79*, 2105–2120. [CrossRef]
29. Prokhorenko, V.I.; Steensgaard, D.B.; Holzwarth, A.R. Exciton Theory for Supramolecular Chlorosomal Aggregates: 1. Aggregate Size Dependence of the Linear Spectra. *Biophys. J.* **2003**, *85*, 3173–3186. [CrossRef]
30. Minami, T.; Tretiak, S.; Chernyak, V.; Mukamel, S. Frenkel-Exciton Hamiltonian for Dendrimeric Nanostar. *J. Lumin.* **2000**, *87–89*, 115–118. [CrossRef]
31. Malyshev, A.V.; Malyshev, V.A. Statistics of Low Energy Levels of a One-Dimensional Weakly Localized Frenkel Eexciton: A Numerical Study. *Phys. Rev. B* **2001**, *63*, 195111. [CrossRef]
32. Ambrosek, D.; Köhn, A.; Schulze, J.; Kühn, O. Quantum Chemical Parametrization and Spectroscopic Characterization of the Frenkel Exciton Hamiltonian for a J-Aggregate Forming Perylene Bisimide Dye. *J. Phys. Chem. A* **2012**, *116*, 11451–11458. [CrossRef]
33. Plötz, P.A.; Niehaus, T.; Kühn, O. A New Efficient Method for Calculation of Frenkel Exciton Parameters in Molecular Aggregates. *J. Chem. Phys.* **2014**, *140*, 174101. [CrossRef] [PubMed]
34. Gattuso, H.; Assfeld, X.; Monari, A. Modeling DNA Electronic Circular Dichroism by QM/MM Methods and Frenkel Hamiltonian. *Theor. Chem. Acc.* **2015**, *134*, 36. [CrossRef]
35. Schröter, M.; Ivanov, S.D.; Schulze, J.; Polyutov, S.P.; Yan, Y.; Pullerits, T.; Kühn, O. Exciton–Vibrational Coupling in the Dynamics and Spectroscopy of Frenkel Excitons in Molecular Aggregates. *Phys. Rep.* **2015**, *567*, 1–78. [CrossRef]
36. Kaiser, A.; Daoud, R.E.; Aquilante, F.; Kühn, O.; De Vico, L.; Bokarev, S.I. A Multiconfigurational Wave Function Implementation of the Frenkel Exciton Model for Molecular Aggregates. *J. Chem. Theory Comput.* **2023**, *19*, 2918–2928. [CrossRef] [PubMed]
37. Tinoco, I. Theoretical Aspects of Optical Activity Part Two: Polymers. In *Advances in Chemical Physics*, 1st ed.; Prigogine, I., Ed.; Wiley: Hoboken, NJ, USA, 1962; Volume 4, pp. 113–160. [CrossRef]
38. Tinoco, I.; Woody, R.W.; Bradley, D.F. Absorption and Rotation of Light by Helical Polymers: The Effect of Chain Length. *J. Chem. Phys.* **1963**, *38*, 1317–1325. [CrossRef]
39. Bayley, P.M.; Nielsen, E.B.; Schellman, J.A. Rotatory Properties of Molecules Containing Two Peptide Groups: Theory. *J. Phys. Chem.* **1969**, *73*, 228–243. [CrossRef] [PubMed]
40. Bulheller, B.M.; Hirst, J.D. DichroCalc—Circular and Linear Dichroism Online. *Bioinformatics* **2009**, *25*, 539–540. [CrossRef]
41. Mavridis, L.; Janes, R.W. PDB2CD: A Web-Based Application for the Generation of Circular Dichroism Spectra from Protein Atomic Coordinates. *Bioinformatics* **2017**, *33*, 56–63. [CrossRef]
42. Nagy, G.; Igaev, M.; Jones, N.C.; Hoffmann, S.V.; Grubmüller, H. SESCA: Predicting Circular Dichroism Spectra from Protein Molecular Structures. *J. Chem. Theory Comput.* **2019**, *15*, 5087–5102. [CrossRef]
43. Drew, E.D.; Janes, R.W. PDBMD2CD: Providing Predicted Protein Circular Dichroism Spectra from Multiple Molecular Dynamics-Generated Protein Structures. *Nucleic Acids Res.* **2020**, *48*, W17–W24. [CrossRef]
44. Miles, A.J.; Janes, R.W.; Wallace, B.A. Tools and Methods for Circular Dichroism Spectroscopy of Proteins: A Tutorial Review. *Chem. Soc. Rev.* **2021**, *50*, 8400–8413. [CrossRef] [PubMed]
45. May, V.; Kühn, O. *Charge and Energy Transfer Dynamics in Molecular Systems*, 3rd ed.; John Wiley & Sons, Incorporated: Hoboken, NJ, USA, 2011.
46. Sauter, C.; Otálora, F.; Gavira, J.A.; Vidal, O.; Giegé, R.; García-Ruiz, J.M. Structure of Tetragonal Hen Egg-White Lysozyme at 0.94 Å from Crystals Grown by the Counter-Diffusion Method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2001**, *57*, 1119–1126. [CrossRef] [PubMed]
47. Neidigh, J.W.; Fesinmeyer, R.M.; Andersen, N.H. Designing a 20-Residue Protein. *Nat. Struct. Biol.* **2002**, *9*, 425–430. [CrossRef]

48. Bas, D.C.; Rogers, D.M.; Jensen, J.H. Very Fast Prediction and Rationalization of pKa Values for Protein-Ligand Complexes. *Proteins Struct. Funct. Genet.* **2008**, *73*, 765–783. [CrossRef]
49. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [CrossRef] [PubMed]
50. Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954–9960. [CrossRef]
51. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N· log (N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [CrossRef]
52. Hopkins, C.W.; Le Grand, S.; Walker, R.C.; Roitberg, A.E. Long-Time-Step Molecular Dynamics Through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11*, 1864–1874. [CrossRef]
53. Morere, J. `DichroProt`. 2025. Available online: https://github.com/Jeremy-Morere/DichroProt (accessed on 10 April 2025).
54. Monari, A.; Rivail, J.L.; Assfeld, X. Theoretical modeling of large molecular systems. Advances in the local self consistent field method for mixed quantum mechanics/molecular mechanics calculations. *Acc. Chem. Res.* **2013**, *46*, 596–603. [CrossRef]
55. Becke, A.D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652. [CrossRef]
56. Yanai, T.; Tew, D.P.; Handy, N.C. A New Hybrid Exchange–Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51–57. [CrossRef]
57. Chai, D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom–Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615. [CrossRef]
58. Scott, M.; Rehn, D.R.; Coriani, S.; Norman, P.; Dreuw, A. Electronic Circular Dichroism Spectra Using the Algebraic Diagrammatic Construction Schemes of the Polarization Propagator Up To Third Order. *J. Chem. Phys.* **2021**, *154*, 064107. [CrossRef]
59. Borucki, B.; Otto, H.; Meyer, T.E.; Cusanovich, M.A.; Heyn, M.P. Sensitive circular dichroism marker for the chromophore environment of photoactive yellow protein: Assignment of the 307 and 318 nm bands to the n→ π* transition of the carbonyl. *J. Phys. Chem. B* **2005**, *109*, 629–633. [CrossRef] [PubMed]
60. Sreerama, N.; Manning, M.C.; Powers, M.E.; Zhang, J.X.; Goldenberg, D.P.; Woody, R.W. Tyrosine, Phenylalanine, and Disulfide Contributions to the Circular Dichroism of Proteins: Circular Dichroism Spectra of Wild-Type and Mutant Bovine Pancreatic Trypsin Inhibitor. *Biochemistry* **1999**, *38*, 10814–10822. [CrossRef]
61. Berndt, K.D.; Güntert, P.; Orbons, L.P.M.; Wüthrich, K. Determination of a High-Quality Nuclear Magnetic Resonance Solution Structure of the Bovine Pancreatic Trypsin Inhibitor and Comparison with Three Crystal Structures. *J. Mol. Biol.* **1992**, *227*, 757–775. [CrossRef]
62. Wlodawer, A.; Walter, J.; Huber, R.; Sjölin, L. Structure of Bovine Pancreatic Trypsin Inhibitor: Results of Joint Neutron and X-Ray Refinement of Crystal Form ii. *J. Mol. Biol.* **1984**, *180*, 301–329. [CrossRef] [PubMed]
63. Borgstahl, G.E.O.; DeWight, R.W.; Getzoff, E.D. 1.4. Ang. Structure of Photoactive Yellow Protein, a Cytosolic Photoreceptor: Unusual Fold, Active Site, and Chromophore. *Biochemistry* **1995**, *34*, 6278–6287. [CrossRef]
64. Bulheller, B.M.; Hirst, J.D. Circular and Linear Dichroism of Proteins. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2020–2035. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.