

Charge-Transfer Transitions in the Vacuum-Ultraviolet of Protein Circular Dichroism Spectra

Benjamin M. Bulheller,[†] Andrew J. Miles,[‡] B. A. Wallace,[‡] and Jonathan D. Hirst^{*,†}

School of Chemistry, University of Nottingham, University Park, Nottingham, NG7 2RD, United Kingdom, and Birkbeck College, University of London, London, WC1E 7HX, United Kingdom

Received: September 17, 2007; In Final Form: October 26, 2007

Circular dichroism (CD) is widely used in the structural characterization and secondary structure determination of proteins. The vacuum UV region (below 190 nm), where charge-transfer transitions have an influence on the CD spectra, can be accessed using synchrotron radiation circular dichroism (SRCD) spectroscopy. Recently, charge-transfer transitions in a conformationally diverse set of dipeptides have been characterized *ab initio* using complete active space self-consistent field calculations, and the relevant charge distributions have been parametrized for use in the matrix method for calculations of protein CD. Here, we present calculations of the vacuum UV CD spectra of 71 proteins, for which experimental SRCD spectra and X-ray crystal structures are available. The theoretical spectra are calculated considering charge-transfer and side chain transitions. This significantly improves the agreement with experiment, raising the Spearman correlation coefficient between the calculated and the experimental intensity at 175 nm from 0.12 to 0.79. The influence of the conformation on charge-transfer transitions is analyzed in detail, showing that the $n \rightarrow \pi^*$ charge-transfer transitions are most important in α -helical proteins, whereas in β strand proteins the $\pi \rightarrow \pi^*$ charge-transfer transition along the chain in the amino- to carboxy-end direction is most dominant.

1. Introduction

Several advantages have contributed to the success of chiroptical methods for the study of biological systems. Although lacking the remarkable spatial resolution of NMR and X-ray diffraction, electronic circular dichroism (CD), in particular, is popular, since it is neither reliant on high concentrations nor limited by the size of the protein. The measurements are straightforward to perform and subsequent analysis of the CD spectra, usually based on empirical relationships, yields information on the secondary structure. The need for information on protein folding motivates the improvement of the theoretical calculation of CD spectra, in order to understand the relationship between conformation and CD spectrum more quantitatively, without reliance on the availability of reference spectra data on closely related proteins.¹

The origin of protein CD is the chiral nature of the polypeptide backbone, whose electronic transitions give rise to distinctive bands in the far UV. The most important structure type is the helix, to which even β strands belong, in the sense that they are very narrow helices. Moffitt showed that the coupling of electric dipole allowed $\pi \rightarrow \pi^*$ electronic transitions in a helix leads to one in-phase combination with a net polarization parallel to the helix axis and two out-of-phase combinations with a net polarization perpendicular to the axis.² The electric field of circularly polarized light causes a linear displacement of charge when promoting electrons into an excited state of a molecule. In the case of the $\pi \rightarrow \pi^*$ transition, for example, this occurs along the bond. The magnetic field on the other hand induces a circulation of charge.³ These two interac-

tions occur at the same time, and the combination of linear and circular displacement leads to a helical movement of electrons with which left and right circularly polarized light interact differently, leading to the phenomenon of CD. Hence, for the calculation of the theoretical spectrum, the electrostatic field caused by the chromophoric groups is required.

Conventional CD spectrometers can record spectra down to approximately 190 nm and provide information on excitation bands in the near- and far-UV.⁴ Below 190 nm, a low signal-to-noise ratio makes measurements more challenging, because of the absorption by the solvent and the low intensity of light. However, using short path-length cells at high concentrations, it is sometimes possible to measure down to 178 nm.⁵ Synchrotron radiation CD (SRCD)^{6–9} can reach into the vacuum UV (VUV) and reveal spectral features down to 170 nm and below. SRCD spectra provide a better signal-to-noise ratio¹⁰ and contain information about more electronic transitions¹¹ than data collected by conventional spectrometers.^{5,12,13} The additional data in the deep-UV in combination with calculations may provide new insights into protein folding. Recently, SRCD measurements in combination with *ab initio* calculations were used to study the influence of solvent molecules on dialanine.⁹ The CD spectra of the anionic, cationic, and zwitterionic forms show substantial differences around 210 and 170 nm, the latter information only made accessible by SRCD experiments. *Ab initio* calculations indicated that the backbone conformation was not the only factor influencing CD spectra but that solvent effects have to be considered.

In the near-UV (250–350 nm), the transitions of aromatic side chains show an influence,¹⁴ whereas in the far-UV (190–250 nm), the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions of the amide groups are dominant.¹⁵ The origin of the bands in the VUV has not been fully elucidated. The proposition of an $n \rightarrow \sigma^*$ transition¹⁶ was refuted by *ab initio* calculations showing that

* Corresponding author. Phone: +44 115 951 3478. Fax: +44 115 951 3562. E-mail: jonathan.hirst@nottingham.ac.uk.

[†] University of Nottingham.

[‡] University of London.

the σ^* orbital lies at a much higher energy.¹⁷ A possible transition between the bonding and the antibonding π orbitals has been suggested, but this occurs around 140 nm.¹⁷ Low-lying Rydberg states are important in this region for gas-phase spectra but are destabilized in a condensed phase.¹⁷ Today it is believed that charge-transfer transitions between adjacent peptide groups have without doubt an important influence in this region.^{18,19} Calculations on a dipeptide^{18–21} showed that charge-transfer transitions from either the oxygen lone pair or the nonbonding π orbital on one peptide group to the π^* orbital on the next group occur between 120 and 175 nm. The effects might, therefore, be visible in the SRCD spectra. To exploit the higher information content of the experimental SRCD spectra, it is necessary to take charge-transfer transitions into account.

Goldmann et al.²² have calculated the optical spectra of polyalanines using the time-dependent Hartree–Fock technique and the INDO/S Hamiltonian. The transition density matrices suggested that charge-transfer has a considerable influence on the NV₁ band (arising from a local $\pi \rightarrow \pi^*$ transition) and that the Frenkel model (which is the basis of matrix method calculations of protein CD) in its original form, therefore, does not adequately describe the electronic excitations in this band. Woody and Koslowski²³ have noted, however, that INDO/S calculations may exaggerate the extent of mixing between local and charge-transfer transitions as the energy gap between them is underestimated by the semiempirical method. To incorporate the charge-transfer transitions directly into the Hamiltonian matrix, we have extended our ab initio characterization of peptide backbone²⁴ and aromatic side chain chromophores¹⁴ to include the charge-transfer transitions between neighboring peptide groups.^{18,25} The use of these transitions in matrix method calculations has been assessed by comparison with the experimental spectra recorded between 190 and 230 nm for a set of 47 proteins²¹ and another set of 31 proteins, whose spectra were recorded between 161 and 230 nm,^{25–27} using SRCD. The charge-transfer chromophores span two peptide groups with ten different geometries, distributed over the accessible regions of the Ramachandran plot. This will be discussed later in more detail.

All matrix method parameters used in this paper are freely available for download and use in calculations. DichroCalc, a web interface for online CD calculation, provides the possibility to upload one or more Protein Data Bank (PDB) files or to retrieve files from the RCSB repository²⁸ via their PDB code. The user can choose between several options, for example, whether to include aromatic side chain and/or charge-transfer transitions, the curve type for the bandshapes, and the bandwidth. If an experimental spectrum is provided for the respective PDB file, a comparison plot is automatically generated. The DichroCalc web interface can be accessed at <http://comp.chem.nottingham.ac.uk>.¹⁵

Recently, 71 SRCD spectra recorded from 175 to 240 nm have been described (Table 1) and used as a reference dataset for CD analysis.¹⁰ Atomic coordinates for these precise proteins are available via the PDB.²⁸ In order to be included in the reference dataset, the protein data had to meet certain standards of purity, crystallographic quality, structural characteristics, and spectral quality.¹⁰ Most importantly for these studies, the spectral magnitudes were very accurately determined by careful calibration of cell path length and accurate determination of concentration. To function as a reference set, the proteins cover a range of secondary structure characteristics and fold architectures. About 18% of the proteins are mainly α -helical, 39% mainly

TABLE 1: PDB Codes and Names of the 71 Proteins Comprising the SP175 Set

PDB	protein	PDB	protein
193l	lysozyme	1m8u	γ -E-crystallin
1a49	pyruvate kinase	1mol	monellin
1a6m	myoglobin whale	1n5u	serum albumin
1ado	aldolase	1nls	concanavalin A
1air	pectate lyase C	1ofs	pea lectin
1atj	peroxidase	1ova	ovalbumin
1avu	serine protease inhibitor	1ppn	papain
1ax8	leptin	1qfe	dehydroquinase type 1
1b8e	β -lactoglobulin	1r0i	rubredoxin
1bgl	β -galactosidase	1rav	avidin
1blf	lactoferrin	1rhs	rhodanese
1bn6	haloalkane dehydrogenase	1sca	subtilisin A
1ca2	carbonic anhydrase II	1stp	streptavidin
1cbj	superoxide dismutase	1thw	thaumatin
1cf3	glucose oxidase	1trz	insulin
1dot	ovotransferrin	1ubi	ubiquitin
1ed9	alkaline phosphatase	1une	phospholipase-A2
1elp	γ -D-crystallin	1vjs	alpha-amylase
1f4j	catalase	1ymb	myoglobin horse
1fa2	β -amylase	2bb2	β -B2-crystallin
1gpb	glycogen phosphorylase-b	2cga	α -chymotrypsinogen
1ha4	γ -S-crystallin (C-term)	2cts	citrate synthase
1ha7	c-phycoyanin	2dhq	dehydroquinase-type2
1hc9	α -bungarotoxin	2fdn	ferredoxin
1hcb	carbonic anhydrase I	2psg	pepsinogen
1hda	hemoglobin	3dni	deoxyribonuclease-I
1hk0	γ -D-crystallin (human)	3est	elastase
1hnn	PNMT	3pgk	phosphoglycerate kinase
1hrc	cytochrome c	3pmg	phosphoglucomutase
1hwx	glutamate dehydrogenase	3rn3	ribonuclease A
1igt	immunoglobulin	4gcr	γ -B-crystallin
1k6j	NmrA	5cha	α -chymotrypsin
1kcw	ceruloplasmin	5cpa	carboxypeptidase A
1ku8	jacalin	5pti	aprotinin
1les	lentil lectin	7tim	triose phosphate isomerase
1lin	calmodulin		

β , and 35% $\alpha\beta$ mixtures. The rest contain multiple classes or only few secondary structures.¹⁰ Since this set is considerably larger than the previous set which provided VUV data²⁵ and has been specially selected for structural diversity, we use it to assess the accuracy of CD calculation and the relative importance of different charge-transfer transitions in the VUV.

2. Theoretical Methods

The most important chromophore in proteins is the peptide bond, whose transitions cause the intense bands around 190 nm ($\pi \rightarrow \pi^*$, the NV₁ band) and 222 nm ($n \rightarrow \pi^*$, the W band). To calculate the electronic transitions in a polypeptide, the electrostatic field which is caused by the peptide chromophores and interacts with the incident light has to be determined. This has been done using ab initio techniques by parametrizing the amide chromophore with *N*-methylacetamide (NMA) as a model compound,²⁴ using the complete-active space self-consistent-field method implemented within a self-consistent reaction field (CASSCF/SCRF),^{29–32} combined with multi-configurational second-order perturbation theory (CASPT2-RF).^{29,32} Monopoles for each transition between electronic states were determined by fitting their electrostatic potential to reproduce the ab initio electrostatic potential for that transition so that the least-squares difference is minimized (typically within 5%). The parameter set²⁴ consists of 32 monopoles for the amide $n \rightarrow \pi^*$ transition all at a distance of 0.1 Å from the C, N, O, and H atoms (forming a cube around each atom), and 20 monopoles for the $\pi \rightarrow \pi^*$ transition, with one charge situated at each atom center and four around each center at a distance of 0.05 Å (all in the peptide plane).

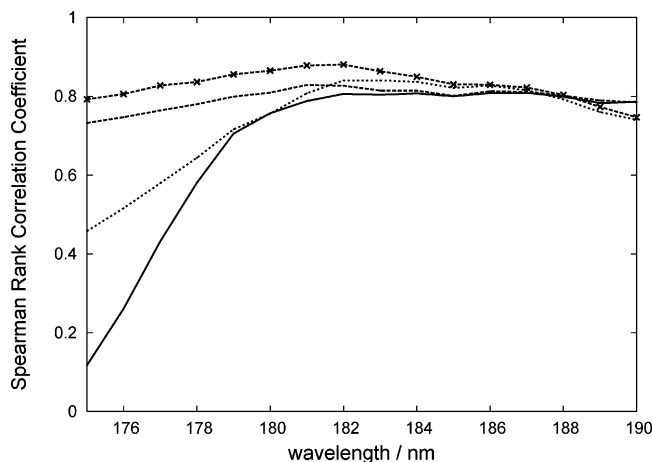


Figure 1. Spearman rank correlation coefficients between the experimental and the calculated intensities of the SP175 CD spectra, calculated with different types of transitions: backbone only (solid line), with aromatic side chains (dotted), with charge transfer (dashed), with side chains and charge transfer (dashed line with crosses).

To calculate the electric and magnetic transition dipole moments of electronic transitions in proteins, the matrix method^{33–36} is commonly used. It is an improved formulation of the Tinoco method³⁷ and is derived from the Frenkel exciton model.³⁸ In the matrix method, the protein is regarded as M independent chromophores, with a monomer wave function ϕ_{is} for each chromophoric group i and electronic state s . To a first approximation, the k^{th} excited-state wavefunction, ψ^k , of the protein is described as a linear combination of electronic configurations, Φ_{ia} , in which only one chromophoric group i is in an excited state a , and the others are in the ground state 0.³ Hence,

$$\Phi_{ia} = \phi_{i0} \cdots \phi_{ia} \cdots \phi_{j0} \cdots \phi_{M0} \quad (1)$$

where ϕ_{ia} is the wave function of chromophore i after the excitation $0 \rightarrow a$, and ψ^k is the sum of all n_i transitions on the M groups:

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

To determine the expansion coefficients c_{ia}^k and thus the wave functions ψ^k , the Schrödinger equation is solved for the protein with M independent chromophores. The Hamiltonian, \hat{H} , is taken to be the sum of the local Hamiltonians, \hat{H}_i , plus the sum of all intergroup potentials, \hat{V}_{ij} , between adjacent groups i and j :

$$\hat{H} = \underbrace{\sum_{i=1}^M \hat{H}_i}_{\hat{H}_0} + \underbrace{\sum_{i=1}^{M-1} \sum_{j=i+1}^M \hat{V}_{ij}}_{\hat{V}} \quad (3)$$

If eq 3 is recast in matrix form, the energy (the eigenvalues) and the wave functions (eigenvectors, that is, the coefficients c_{ia}^k) can be calculated by solving an eigenvalue problem through matrix diagonalization. From the unitary matrix which effects the diagonalization, the electric and magnetic transition dipole moments can be derived. The Hamiltonian of eq 3 for a dipeptide considering two transitions per group has the form

$$\hat{H} = \begin{pmatrix} E_1^{\text{local}} & V_{1;2} \\ V_{1;2} & E_2^{\text{local}} \end{pmatrix} \quad (4)$$

$$\text{where } E_i^{\text{local}} = \begin{pmatrix} E_{in\pi^*}^{\text{local}} & V_{in\pi^*;i\pi\pi^*}^{\text{local}} \\ V_{in\pi^*;i\pi\pi^*}^{\text{local}} & E_{i\pi\pi^*}^{\text{local}} \end{pmatrix} \quad (5)$$

$$\text{and } V_{ij} = \begin{pmatrix} V_{in\pi^*;jn\pi^*} & V_{i\pi\pi^*;jn\pi^*} \\ V_{in\pi^*;j\pi\pi^*} & V_{i\pi\pi^*;j\pi\pi^*} \end{pmatrix} \quad (6)$$

The diagonal elements represent the transition energies for each transition of each chromophore. The off-diagonal elements, V_{ij} , are the interaction energies between different transitions in the protein. These interactions are the origin of the dependency of protein CD spectra on secondary (and tertiary) structure. If the interaction between the transitions is regarded as an electrostatic interaction between charge densities ρ of separation r , that is, $\hat{V}_{ij} = (1/4\pi\epsilon_0 r_{ij})$, then V_{ij} have the form

$$V_{i0a;j0b} = \int_{r_i} \int_{r_j} \frac{\rho_{ia0}(r_i)\rho_{j0b}(r_j)}{4\pi\epsilon_0 r_{ij}} d\tau_i d\tau_j \quad (7)$$

where $\rho_{ia0}(r_i)$ and $\rho_{j0b}(r_j)$ are the transition electron densities on chromophores i and j , ϵ_0 is the vacuum permittivity, and r_{ij} is the distance between the chromophores. If the permanent and transition densities are approximated by point charges and the integrals in eq 7 are re-cast as a sum of the Coulomb interactions of these monopoles (the monopole–monopole approximation), then eq 7 becomes³⁷

$$V_{i0a;j0b} = \frac{1}{4\pi\epsilon_0} \sum_{s=1}^{N_s} \sum_{t=1}^{N_t} \frac{q_s q_t}{r_{st}} \quad (8)$$

where q_s and q_t are the point charges on chromophores i and j , and N_s and N_t are the number of these charges on the respective chromophore. From the diagonalization of the Hamiltonian matrix, the wavelength of each transition and the electric and magnetic transition dipole moments are derived, from which the rotational strength at the respective wavelength can be calculated. Since in a real spectrum, the transitions are broadened, and each transition in the line spectrum needs to be superimposed with a band shape. Gaussian shapes with a fixed bandwidth of 12.5 nm were used to produce the presented results.

To account for charge-transfer between adjacent groups, chromophores spanning two peptide bonds had to be parameterized. The calculations were carried out ab initio on *N*-acetyltyrosine-*N'*-methylamide in 10 different conformations (Table 2).²⁵ These geometries include a planar dimer ($\phi = 180^\circ$, $\psi = 180^\circ$) and several conformations in the main regions of the Ramachandran plot (corresponding to α -helical and β -sheet structures). Mirror images of all structures were generated to describe, for example, left-handed helices and also cover less-favored but occurring conformations. This gives a total of 19 conformations (the planar dimer is its own mirror image).

For each chromophore, four transitions have to be taken into account. $n_1 \rightarrow \pi_2^*$ and $\pi_{nb1} \rightarrow \pi_2^*$ describe a charge-transfer from peptide group 1 (at the amino end) to group 2, whereas $n_2 \rightarrow \pi_1^*$ and $\pi_{nb2} \rightarrow \pi_1^*$ account for transitions in the opposite direction. Every charge-transfer transition is described by 64 monopoles, which are located around the positions of the C, N, O, and H atoms of the peptide bond. Around each atom, there are 8 charges arranged at the corners of a cube with a distance

TABLE 2: Wavelengths and Oscillator Strengths (f) of the Charge-Transfer Chromophores^a

$\Phi/^\circ$	$\Psi/^\circ$	$n_1 \rightarrow \pi_2^*$		$n_2 \rightarrow \pi_1^*$		$\pi_{nb1} \rightarrow \pi_2^*$		$\pi_{nb2} \rightarrow \pi_1^*$		structure
		λ/nm	f	λ/nm	f	λ/nm	f	λ/nm	f	
180	180	155.0	0.000	132.6	0.042	171.0	0.040	131.8	0.042	strand
-135	135	154.0	0.001	133.5	0.045	174.4	0.168	122.9	0.006	strand
-120	120	155.7	0.009	135.2	0.111	173.6	0.202	126.4	0.008	strand
-120	180	153.2	0.029	142.1	0.056	167.0	0.015	138.5	0.047	hybrid
-120	60	166.7	0.215	134.9	0.021	159.9	0.056	135.2	0.005	hybrid
-60	180	156.1	0.004	135.9	0.023	170.9	0.051	136.4	0.058	hybrid
-74	-4	138.2	0.022	162.9	0.091	153.1	0.002	145.9	0.010	helical
-62	-41	163.5	0.121	159.2	0.068	155.3	0.011	156.1	0.042	helical
-48	-57	156.9	0.123	162.1	0.013	164.4	0.014	152.7	0.015	helical
-60	-60	170.7	0.162	154.5	0.051	163.1	0.020	145.1	0.013	helical

^a A mirror image of each geometry was created additionally to construct parameters, e.g., for the left-handed helical and sheet regions of the Ramachandran plot.

of 0.1 Å to the atom center. For the assignment of the suitable chromophore, the Φ and Ψ angles of two neighboring groups are calculated, and the closest conformation (Table 2) is determined. Because of missing residues, for example, many PDB files contain large gaps within the chains which are not detected by simply calculating the dihedral angles. Therefore, the distance between the groups needs to be checked before assigning the charge-transfer groups.

The geometries in Table 2 may be divided into strand- and helix-forming groups. The type hybrid refers to structures that look like wound strands or stretched helices and could therefore not be classified unambiguously. The set with the angles -60° and 180° is listed as hybrid, because, although it forms a helix when using only these angles, it lacks hydrogen bonds to stabilize the structure. This type also appears much more often in β -type proteins than in helical ones.

There are four transitions involving charge-transfer between adjacent groups, which have been considered in the parameter sets. These are the excitation of an electron from a lone pair in one group to the π^* orbital in the other group (Table 2) and excitation from a nonbonding π orbital to the other group's π^* orbital, respectively. The Hamiltonian for a tripeptide, showing only one charge-transfer transition per peptide group for clarity, has the form

$$\hat{H} = \begin{pmatrix} E_{local_1} & V_{E_1:E_2} & V_{E_1:E_3} & V_{E_1:CT_{12}} & V_{E_1:CT_{23}} \\ V_{E_1:E_2} & E_{local_2} & V_{E_2:E_3} & V_{E_2:CT_{12}} & V_{E_2:CT_{23}} \\ V_{E_1:E_3} & V_{E_2:E_3} & E_{local_3} & V_{E_3:CT_{12}} & V_{E_3:CT_{23}} \\ V_{E_1:CT_{12}} & V_{E_2:CT_{12}} & V_{E_3:CT_{12}} & CT_{12} & 0 \\ V_{E_1:CT_{23}} & V_{E_2:CT_{23}} & V_{E_3:CT_{23}} & 0 & CT_{23} \end{pmatrix} \quad (9)$$

The blocks involving solely local excitations are as defined in eq 5. In eq 9, the first charge-transfer chromophore consists of peptide groups one and two, whereas the latter spans groups two and three. The zeros in eq 9 reflect that charge-transfer chromophores sharing a common peptide group are not allowed to interact. Neglect of this interaction is a necessary practicability, as overlapping sets of monopoles would lead to catastrophic Coulomb interactions. The constitution of the off-diagonal elements for the charge-transfer transitions is more complicated than for the case of local transitions only (eq 7), because, in addition to the interactions between ground states and local excited states, charge-transfer transitions may be localized on a single or on different dimers. For completeness, expressions for the various off-diagonal elements involving charge-transfer transitions are given in the Appendix.

TABLE 3: Spearman Rank Correlation Coefficients between the Calculated and Experimental Intensity at Several Important Wavelengths for Calculations Considering Different Electronic Transitions

chromophores considered	correlation coefficient		
	175 nm	190 nm	222 nm
peptide backbone	0.12	0.79	0.91
+ side chains	0.46	0.74	0.89
+ charge-transfer	0.73	0.78	0.89
+ side chains + charge-transfer	0.79	0.75	0.88

3. Results and Discussion

3.1. Comparison with Experiment. Matrix method calculations have been carried out using the 71 protein structures of the SP175 dataset, considering different combinations of chromophores: all backbone-transitions only and in combination with charge-transfer and side chain transitions. The transitions of the aromatic groups of tyrosine, tryptophan, and phenylalanine comprised the side chain transitions and were modeled by parameters derived from CASSCF/CASPT2 calculations.³⁹ Comparisons of the experimental and calculated spectra are provided as supporting material. The experimental spectra have been described and discussed in earlier work,¹⁰ but they have not previously been published in their entirety. The agreement between calculation and experiment is clearly not fully quantitative. For example, in the far-UV, the balance between the calculated intensities of the bands at 208 nm and those at 222 nm in helical proteins is not correct with the parameters leading to an underestimation of the intensity at 222 nm. Nevertheless, many trends across different proteins in relative intensities are captured, and this can be quantified by the Spearman rank correlation coefficient⁴⁰ between the calculated and the experimental intensity at each wavelength. In the backbone-only calculations at 222 nm, the coefficient reaches 0.91, an almost quantitative correlation at the crucial wavelength for the determination of helical content, whereas at 175 nm (Figure 1), the Spearman rank correlation coefficient is 0.12, indicating virtually no agreement. The inclusion of side chain transitions improves the correlation coefficient at 175 nm to 0.46. Consideration of charge-transfer transitions on the other hand leads to an increase to 0.73, and in combination with the side chain transitions, the correlation coefficient at 175 nm is 0.79 (Table 3).

The Spearman rank correlation coefficient of 0.79 at 175 nm shows that inclusion of the charge-transfer and side chain transitions qualitatively reproduces the features of the experimental spectra. However, the agreement at 200 nm, the point where the CD intensity changes sign, is a challenge. In this region, the gradient of CD spectra is often greatest, causing a

TABLE 4: Proteins Showing a Minimum at 200 nm, Indicating a β -II-Type Protein

PDB code	protein
1avu	serine protease inhibitor
1r0i	rubredoxin
2cga	α -chymotrypsinogen
2fdn	ferredoxin
3est	elastase
5cha	α -chymotrypsin
5pti	aprotinin

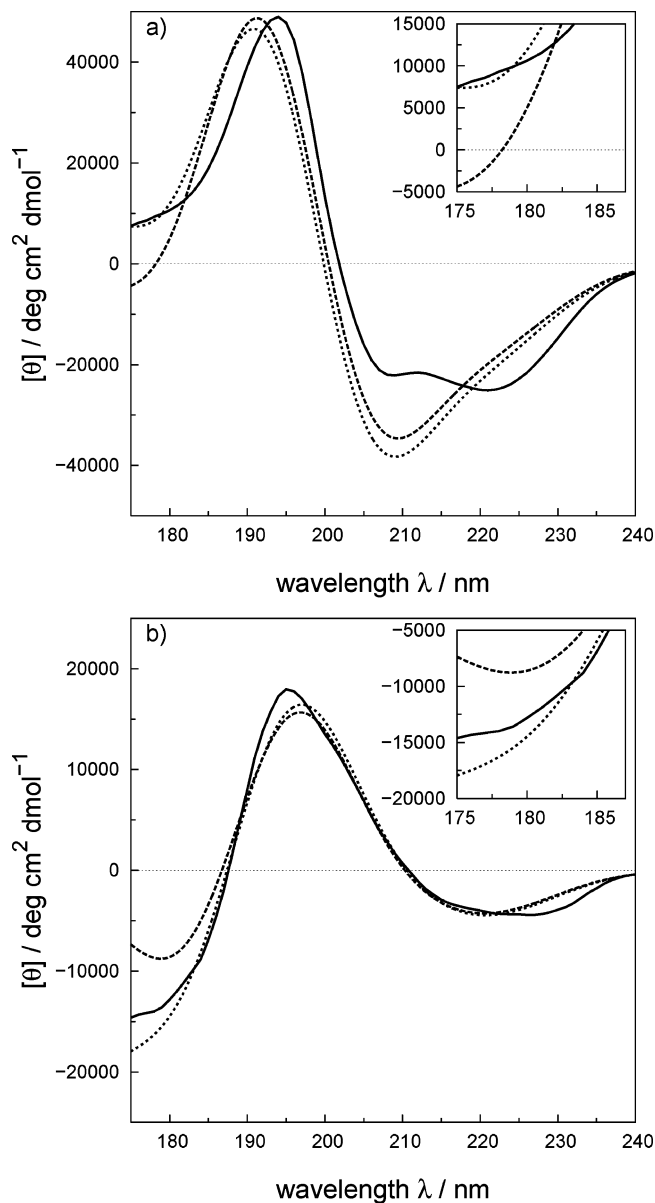
large difference in intensity for even small errors in the calculated wavelength. Although the absolute error in the wavelength where the intensity is calculated to be zero is less than 3 nm, the calculated intensities may differ from experiment by up to 10 000 deg cm² dmol⁻¹. The calculations tend to overestimate the intensity; that is, the calculated wavelength of the zero point is too large, and the spectrum is slightly red shifted. The correlation steadily increases after the zero-crossing and reaches almost constant values between 0.8 and 0.9 above 210 nm.

3.2. Influence of Secondary Structure. The proteins have been divided into α , β , $\alpha\beta$, and unstructured proteins, on the basis of assignments by DSSP⁴¹ (see Supporting Information, Table S1). Proteins with greater than 40% α -helical or more than 40% β content were assigned α and β , respectively. The remainder was assigned $\alpha\beta$ in the case of more than 24.5% of both structure types or unstructured if below that. This classification was consistent with the CATH database.^{42,43}

When charge-transfer was considered, in 69% of the cases, the difference at 175 nm between the calculated and the experimental intensity was reduced for α -helical proteins. On the other hand, only 39% of the β -type proteins showed an improvement, and none of the $\alpha\beta$ proteins did. This was due to an overestimation of the intensity of $\alpha\beta$ proteins in the VUV, because the main consequence of the charge-transfer transitions is a more intense band in this region. The experimental spectra of all α -helical proteins show a positive intensity at 175 nm, whereas the spectra of β -sheet proteins are all negative. The $\alpha\beta$ cases, however, lie between these extremes, either with zero or with small negative intensities. For these proteins, the experimental spectra agree quite well with the backbone-only calculations, which also possess a low intensity in this region and including charge-transfer transitions for these proteins worsens the level of agreement.

While the calculated intensity at 175 nm upon inclusion of charge-transfer transitions becomes more negative without exception for all β , $\alpha\beta$, and unstructured proteins, it increases in nine of the 13 α -helical cases. The four α -helical proteins showing a decrease were aldolase (1ado), cytochrome c (1hrc), insulin (1trz), and citrate synthase (2cts), whose α -helix contents are comparatively low, ranging between 0.41 (1hrc) and 0.61 (2cts). In all of those spectra, the influence of the charge-transfer chromophores is much smaller than for proteins with more α -helices. Of the other helical proteins, only calmodulin (1lin), peroxidase (1atj), and phospholipase-A2 (1une) are less than 60% helical, and they show the same weak influence of the charge-transfer transitions.

Among the 71 proteins, there are seven experimental spectra with a minimum at 200 nm, indicating a β -II-type protein (Table 4).^{44,45} The agreement between calculation and experiment is worse than for other proteins in these cases, especially between 190 and 210 nm. In β -II proteins,^{44,45} the strands are rather short or not aligned in a parallel manner but often twisted and bent. The resulting spectra are similar to unfolded, random coil

**Figure 2.** Experimental and calculated CD spectra of hemoglobin (a, 1hda) and lentil lectin (b, 1les), which show a considerable improvement with included charge-transfer transitions. Experimental (solid), backbone only (dashed), and backbone plus charge transfer (dotted).**TABLE 5: Most Frequent Conformations of Charge-Transfer Chromophores**

geometry	$\phi/^\circ$	$\psi/^\circ$
α -helical	-62	-41
3_{10} -helical	-74	-4
β -strand	-120	120
	-135	135
	180	180

proteins. It may be that conformational dynamics contribute to the unusual shape of the spectrum.¹

A noticeable improvement was obtained for avidin (1rav), a β -sheet protein, which shows an intense negative band at 180 nm, with a mean residue ellipticity of -21 000 deg cm² dmol⁻¹. The backbone-only calculation gives a band intensity of -12 000 deg cm² dmol⁻¹, but the calculated intensity is -18 000 deg cm² dmol⁻¹, if charge-transfer is taken into

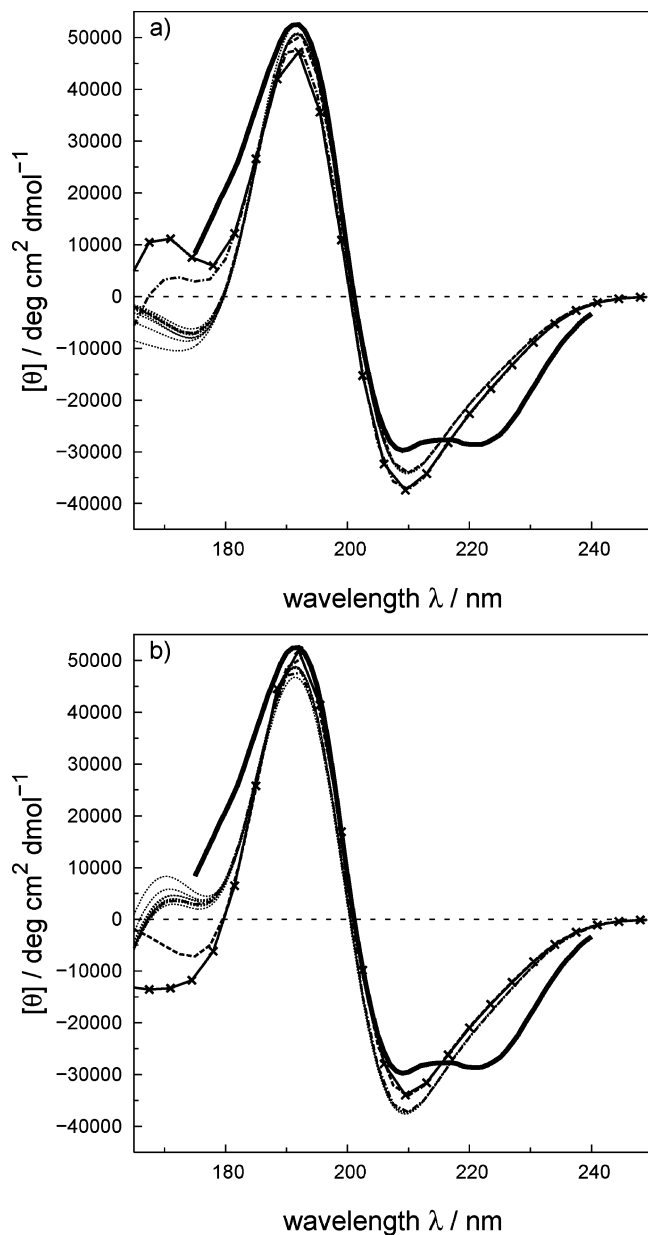


Figure 3. Spectra of serum albumin (1n5u, mainly α). Experimental (solid), backbone-only (dashed), full charge-transfer calculation (all chromophores, dot-dashed). (a) The thin, dotted curves show calculations using only one charge-transfer conformation. The crossed curves show the calculation only with the ($\phi = -62^\circ$, $\psi = -41^\circ$) set. If a conformation was included then all four charge-transfer transitions ($n_1 \rightarrow \pi_2^*$, $n_2 \rightarrow \pi_1^*$, $\pi_1 \rightarrow \pi_2^*$, and $\pi_2 \rightarrow \pi_1^*$) were considered for that chromophore. (b) The thin, dotted curves show calculations leaving out one charge-transfer conformation and using all others. The crossed curves show the calculation without the ($\phi = -62^\circ$, $\psi = -41^\circ$) set, but with all other sets.

account. Similar trends are shown in Figure 2 for an α -helical protein (hemoglobin, 1hda) and a β -sheet protein (lentil lectin, 1les).

3.3. Individual Contributions of Transitions. In order to investigate the influence of specific charge-transfer chromophores on the computed CD spectra, the distribution of the conformations of the charge-transfer chromophores has been analyzed, dividing the conformations into helix- and strand-forming geometries and some hybrid types (Table 2). The sum of all hybrid conformations makes up about 15% of the conformations in an α -helical protein, about 30% for β -sheet proteins, and approximately 20% for an $\alpha\beta$ protein. That is,

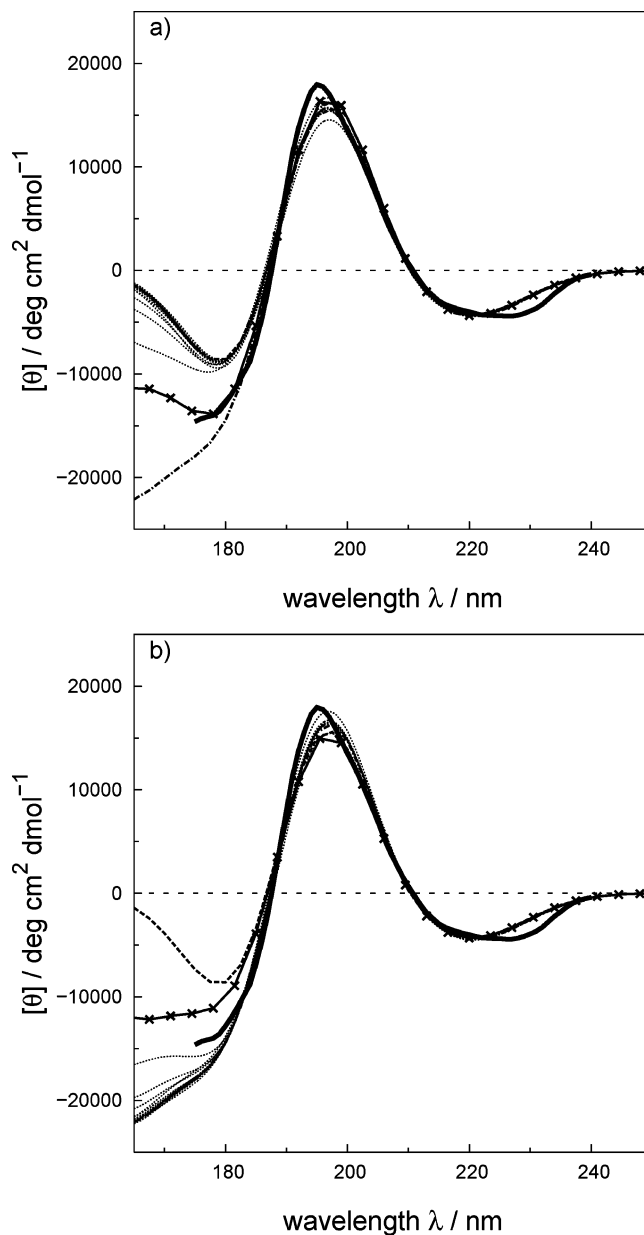


Figure 4. Spectra of lentil lectin (1les, mainly β). Experimental (solid), backbone-only (dashed), full charge-transfer calculation all chromophores, dot-dashed. (a) The thin, dotted curves show calculations using only one charge-transfer conformation. The crossed curves show the calculation only with the ($\phi = -120^\circ$, $\psi = 120^\circ$) set. If a conformation was included then all four charge-transfer transitions ($n_1 \rightarrow \pi_2^*$, $n_2 \rightarrow \pi_1^*$, $\pi_1 \rightarrow \pi_2^*$, and $\pi_2 \rightarrow \pi_1^*$) were considered for that chromophore. (b) The thin, dotted curves show calculations leaving out one charge-transfer conformation and using all others. The crossed curves show the calculation without the ($\phi = -120^\circ$, $\psi = 120^\circ$) set, but with all other sets.

the hybrid chromophores have more β character, as suggested by the location of their dihedral angles in the Ramachandran plot.

Although the amount of helical conformations was usually well above 70% for α proteins, the amount of strand-like conformations in the β set was about 45%. In the $\alpha\beta$ group, helical conformations account for approximately 55% of all charge-transfer chromophores. For the unstructured proteins, there was a fairly even distribution of strand, hybrid, and helical charge-transfer chromophores, with a slight preponderance of the latter. The most important (that is, the most frequent) conformations are given in Table 5. Among the helical chro-

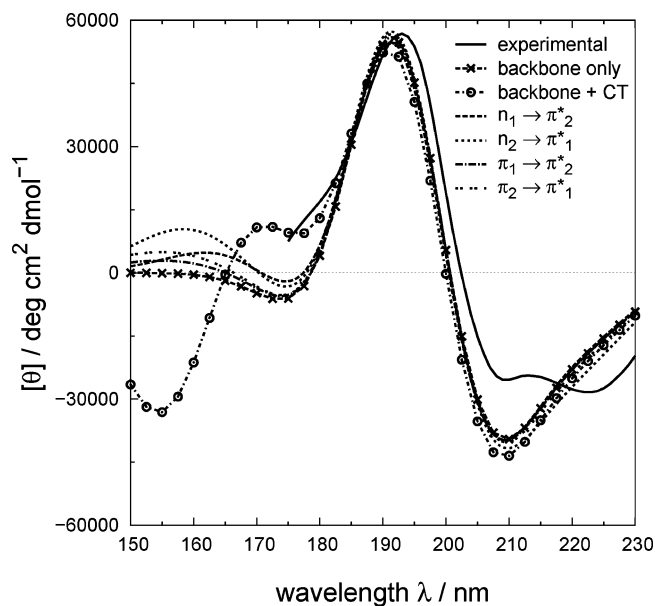


Figure 5. Spectra of whale myoglobin (1a6m) showing the effect of single charge-transfer transitions in a typical α helical protein.

TABLE 6: Large Proteins Which Required Splitting into Single Chains to Calculate the CD

name of the protein	PDB code	number of chains	number of residues	residues per chain
pyruvate kinase	1a49	8	4152	519
aldolase	1ado	4	1452	363
peroxidase	1atj	6	1836	306
β -galactosidase	1bgl	8	8168	1021
catalase	1f4j	4	1928	479/481
c-phycoyanin	1ha7	24	3996	162/171
glutamate dehydrogenase	1hwx	6	3006	501
immunoglobulin	1igt	4	1316	214/444
ceruloplasmin	1kcv	1	1017	1017
ovalbumin	1ova	4	1519	386/373
phosphoglucosmutase	3pmg	2	1122	561

mophores, the ($\phi = -62^\circ$, $\psi = -41^\circ$) group is, as expected, the most important conformation, representing a typical α -helix.⁴⁶ It is also the most frequent helical type in $\alpha\beta$ and unstructured proteins. In β proteins, the 3_{10} -helical conformation ($\phi = -74^\circ$, $\psi = -4^\circ$) occurs more frequently than α -helical conformations.

To assess the individual contribution of a specific conformation, two sets of calculations have been carried out: one using every charge-transfer chromophore on its own and the other omitting each group in turn. A particular peptide group was always assigned the same charge-transfer chromophore throughout the calculations or was ignored in the case of this specific conformation being left out. This showed that the most significant contributions are made by the two most frequent helical (Figure 3) and strand (Figure 4) chromophores but that there is, with a comparable magnitude, a combined contribution from all other groups. Including one of the most frequent groups with dihedral angles ($\phi = -62^\circ$, $\psi = -41^\circ$) or ($\phi = -120^\circ$, $\psi = 120^\circ$) has a large influence on the calculated spectrum. Conversely, leaving one of these out (with all of the other chromophores included) shows a contribution to the same extent. Chromophores with less frequently occurring geometries have little influence on their own but should not be ignored, because of their combined contribution. Notably, the full charge-transfer spectra are not a linear combination of contributions from single chromophores, because of the interaction terms in the Hamiltonian matrix.

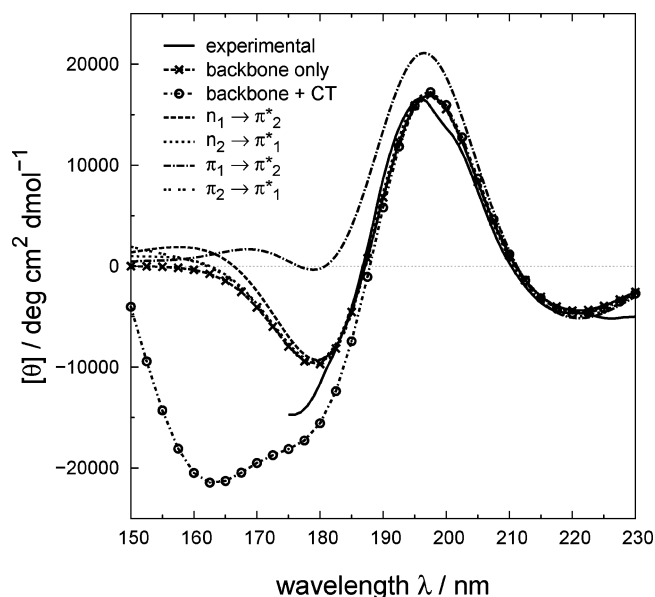


Figure 6. Spectra of pea lectin (1ofs) showing the effect of single charge-transfer transitions in a typical β protein. Most significant is the $\pi_1 \rightarrow \pi_2^*$ transition, which considerably diminishes the band at 180 nm in β and unstructured proteins and increases intensity at 190 nm.

The inclusion of charge-transfer chromophores usually predicts a minimum around 160 nm and causes the intensity in the deep UV to be more negative (by between 5000 and 15 000 $\text{deg cm}^2 \text{dmol}^{-1}$ at 175 nm), except in highly α helical proteins. This large effect is mainly due to interactions between the charge-transfer transitions themselves, rather than being caused by a particular transition. Consideration of a single charge-transfer transition indicates only a weak influence at 175 nm (Figure 5). The largest effect at this wavelength results from the $n_1 \rightarrow \pi_2^*$ transition. At higher energies around 160 nm, the $n_2 \rightarrow \pi_1^*$ transition becomes more intense. Almost all noticeable effects of single transitions are caused by the α helical conformation, ($\phi = -62^\circ$, $\psi = -41^\circ$), even in those β proteins where it is contained, while the hybrid chromophores showed no influence at all. The charge-transfer transitions of strand-like conformations only show an effect in β proteins and only the $\pi_1 \rightarrow \pi_2^*$ transition changes the intensity, while the other three transitions have no influence whatsoever. The sole inclusion of the $\pi_1 \rightarrow \pi_2^*$ transition in β and unstructured proteins causes a considerable increase of intensity between 170 and 200 nm. In many cases, the negative band around 175 nm (which is also produced in backbone-only calculations) disappears completely (Figure 6).

3.4. Computational Tractability. Inclusion of charge-transfer transitions roughly doubles the size of the Hamiltonian matrix. Side chain chromophores also enlarge the calculation but by a smaller extent, which depends on the content of aromatic side chains (typical values are between 5 and 15%). Considering two backbone transitions, our current algorithm and compiler allow the calculation of the CD spectrum of a protein comprising about 2000 residues for backbone-only calculations and 1000 residues if charge-transfer is taken into account. In cases where the limit is reached, the proteins are split into single chains, and their individual spectra are calculated. If all single chain spectra are similar, it is most likely that the overall spectrum will be similar to them. There are some variations of the intensity, but these differences are relatively small. Sometimes proteins are a combination of identical or almost identical chains. These all yield the same spectrum, because of the orientation

independence of CD, if there is negligible interchain interaction. In the SP175 set, there were 11 proteins which were too large to be calculated as a whole, that is, all chains in the oligomer (Table 6). The single chains have usually the same sequence; some of them have slightly different chain termini, but this has almost no effect on the CD spectrum. An exception, however, was immunoglobulin (ligt), which consists of four chains, of which the chains A and C are equivalent, as are chains B and D. Rather than calculating each separate chain, the file was split in half, thus creating two equal fractions of the protein. For the proteins β -galactosidase (1bgl) and ceruloplasmin (1kcw), which contain chains with more than 1000 residues (Table 6), calculations are feasible for the backbone-only calculations but not if charge-transfer is also considered.

4. Conclusion

The SP175 protein set has allowed us to assess the influence of the charge-transfer transitions in the vacuum UV, a region which cannot be reached by conventional spectrometers. At 175 nm, there is almost no correlation (the Spearman rank correlation coefficient, r , is 0.12) between the calculated and the experimental intensity if only local excitations are considered. However, r rises to 0.73 if charge-transfer transitions are taken into account and rises to 0.79 if the side chain chromophores are considered in addition. In most of the calculated spectra, the inclusion of charge-transfer transitions increases the negative band around 170 nm considerably, improving the agreement, while the effects in the far UV region are mostly marginal and slightly worsen the correlation. The positions of the bands at 190 and 208 nm are usually not affected: the intensity changes are minute in the far-UV.

The major contributions come from just 2 of 19 charge-transfer conformations, namely, the main α -helical and β -strand types (Figures 3 and 4). However, the combination of all other chromophores contributes just as much and, therefore, cannot be neglected. The change of intensity in the deep-UV could not be attributed to a specific charge-transfer transition but, rather, is caused by interactions among all four types. These results show that the inclusion of charge-transfer transitions enhances the theoretical CD calculations in the vacuum UV region. Ultimately, we anticipate that CD measurements in combination with molecular simulations and the calculations of CD spectra from the conformations sampled in the simulations will provide new insights into protein folding.

Acknowledgment. We thank Drs. Mark Oakley and Andrew Gilbert for helpful discussions, the Engineering and Physical Sciences Research Council (EPSRC, Grant GR/T09224) and the BBSRC (to B.A.W.) for funding. We thank the University of Nottingham for access to High Performance Computing facilities.

Appendix

The off-diagonal matrix elements are two-electron integrals involving electronic configurations, Φ_{ia} , in which group i is in an excited state a , as described previously in eq 1. Consideration of the products comprising the configurations Φ_{ia} and recognizing that only terms involving the two electrons indexed in the Coulomb operator persist, leads to the following expressions.

Here, we give explicit expressions for the off-diagonal elements of the Hamiltonian matrix involving charge-transfer transitions. Lower case letters, like a and b , refer to localized transitions, whereas letters with a caret, like \hat{c} and \hat{d} , are charge-

transfer transitions. (kl) refers to a single chromophore made up of the monomers k and l .

The interaction between the ground state and a single charge-transfer transition involves a sum of the ground state charge distributions of every chromophore (except the charge-transfer chromophore itself) interacting with the transition density of the charge-transfer transition:

$$\begin{aligned} V_{0;(kl)\hat{c}} &= \int \Phi_0 \hat{V} \Phi_{(lk)\hat{c}} d\tau \\ &= \sum_{i=1}^M \sum_{j>i}^M \int \int \phi_{10} \dots \phi_{(kl)0} \dots \phi_{M0} V_{ij} \phi_{10} \dots \phi_{(kl)\hat{c}} \dots \\ &\quad \phi_{M0} d\mathbf{r}_i d\mathbf{r}_j \\ &= \sum_{i \neq (kl)}^M \int \int \phi_{i0} \phi_{(kl)0} V_{i(kl)} \phi_{i0} \phi_{(kl)\hat{c}} d\mathbf{r}_i d\mathbf{r}_{(kl)} \\ &= \sum_{i \neq (kl)}^M V_{i00;(kl)0\hat{c}} \end{aligned} \quad (10)$$

Similarly, the interaction between an excited-state and a charge-transfer transition where both transitions are localized on a single dimer ($m = k$ or l) is

$$\begin{aligned} V_{m\alpha;(kl)\hat{c}} &= \int \Phi_{m\alpha} \hat{V} \Phi_{(lk)\hat{c}} d\tau \\ &= \sum_{i=1}^M \sum_{j>i}^M \int \int \phi_{10} \dots \phi_{(kl)a} \dots \phi_{M0} V_{ij} \phi_{10} \dots \phi_{(kl)\hat{c}} \dots \\ &\quad \phi_{M0} d\mathbf{r}_i d\mathbf{r}_j \\ &= \sum_{i \neq (kl)}^M \int \int \phi_{i0} \phi_{(kl)a} V_{i(kl)} \phi_{i0} \phi_{(kl)\hat{c}} d\mathbf{r}_i d\mathbf{r}_{(kl)} \\ &= \sum_{i \neq (kl)}^M V_{i00;(kl)a\hat{c}} \end{aligned} \quad (11)$$

The interaction between two different charge-transfer transitions localized on a single dimer has an analogous form. The transition a would be replaced by \hat{d} . The interaction between an excited-state and a charge-transfer transition where both transitions are not localized on a single dimer (i.e., $m \neq k, l$) is exactly analogous to the interaction between two single excitations on different chromophores.

Elements describing the interaction between two charge-transfer transitions not localized on a single dimer (i.e., $m \neq k, l; n \neq k, l$) take the form:

$$\begin{aligned} V_{(mn)\hat{d};(kl)\hat{c}} &= \int \Phi_{(mn)\hat{d}} \hat{V} \Phi_{(lk)\hat{c}} d\tau \\ &= \sum_{i=1}^M \sum_{j>i}^M \int \int \phi_{10} \dots \phi_{(mn)\hat{d}} \dots \phi_{M0} V_{ij} \phi_{10} \dots \phi_{(kl)\hat{c}} \dots \\ &\quad \phi_{M0} d\mathbf{r}_i d\mathbf{r}_j \\ &= \int \int \phi_{(mn)\hat{d}} \phi_{(kl)0} V_{(mn)(kl)} \phi_{(mn)0} \phi_{(kl)\hat{c}} d\mathbf{r}_{(mn)} d\mathbf{r}_{(kl)} \\ &= V_{(mn)0\hat{d};(kl)0\hat{c}} \end{aligned} \quad (12)$$

Supporting Information Available: Comparisons of the experimental and calculated spectra. This information is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Hirst, J. D.; Bhattacharjee, S.; Onufriev, A. V. *Faraday Disc.* **2003**, *122*, 253–267.
- (2) Moffitt, W. *Proc. Natl. Acad. Sci. U.S.A.* **1956**, *42*, 736–746.
- (3) Sreerama, N.; Woody, R. W. *Methods Enzymol. Num. Comp. Methods D* **2004**, *383*, 318–351.
- (4) Rodger, A.; Nordén, B. *Circular Dichroism & Linear Dichroism*; Oxford University Press: Oxford, 1997.
- (5) Wallace, B. A. *Nat. Struct. Biol.* **2000**, *7*, 708–709.
- (6) Sutherland, J. C.; Desmond, E. J.; Takacs, P. Z. *Nucl. Instrum. Methods* **1980**, *172*, 195–199.
- (7) Wallace, B. A. *J. Synchrotron Radiat.* **2000**, *7*, 289–295.
- (8) Janes, R. W. *Bioinformatics* **2005**, *21*, 4230–4238.
- (9) Šebek, J.; Gyurcsik, B.; Šebestík, J.; Kejik, Z.; Bednárová, L.; Bouř P. *J. Phys. Chem. A* **2007**, *111*, 2750–2760.
- (10) Lees, J. G.; Miles, A. J.; Wallace, B. A. *Bioinformatics* **2006**, *22*, 1955–1962.
- (11) Toumadje, A.; Alcorn, S. W.; Johnson, W. C. *Anal. Biochem.* **1992**, *200*, 321–331.
- (12) Wallace, B. A.; Janes, R. W. *Curr. Opin. Chem. Biol.* **2001**, *5*, 567–571.
- (13) Miles, A. J.; Wallace, B. A. *Chem. Soc. Rev.* **2006**, *35*, 39–51.
- (14) Rogers, D. M.; Hirst, J. D. *J. Phys. Chem. A* **2003**, *107*, 11191–11200.
- (15) Bulheller, B. M.; Rodger, A.; Hirst, J. D. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2020–2025.
- (16) Robin, M. *Higher Excited States of Polyatomic Molecules*; Academic Press: New York, 1975.
- (17) Besley, N. A.; Hirst, J. D. *J. Phys. Chem. A* **1998**, *102*, 10791–10797.
- (18) Gilbert, A. T. B.; Hirst, J. D. *J. Mol. Struct. (THEOCHEM)* **2004**, *675*, 53–60.
- (19) Serrano-Andrés, L.; Fülischer, M. P. *J. Am. Chem. Soc.* **1998**, *120*, 10912–10920.
- (20) Serrano-Andrés, L.; Fülischer, M. P. *J. Phys. Chem. B* **2001**, *105*, 9323–9330.
- (21) Oakley, M. T.; Bulheller, B. M.; Hirst, J. D. *Chirality* **2006**, *18*, 340–347.
- (22) Goldmann, E.; Asher, S. A.; Mukamel, S. *Phys. Chem. Chem. Phys.* **2001**, *3*, 2893–2903.
- (23) Woody, R. W.; Koslowski, A. *Biophys. Chem.* **2002**, *101–102*, 535–551.
- (24) Besley, N. A.; Hirst, J. D. *J. Am. Chem. Soc.* **1999**, *121*, 9636–9644.
- (25) Oakley, M. T.; Hirst, J. D. *J. Am. Chem. Soc.* **2006**, *128*, 12414–12415.
- (26) Matsuo, K.; Yonehara, R.; Gekko, K. *J. Biochemistry* **2004**, *135*, 405–411.
- (27) Matsuo, K.; Yonehara, R.; Gekko, K. *J. Biochemistry* **2005**, *138*, 79–88.
- (28) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (29) Bernhardsson, A.; Lindh, R.; Karlström, G.; Roos, B. O. *Chem. Phys. Lett.* **1996**, *251*, 141–149.
- (30) Karlström, G. *J. Phys. Chem.* **1988**, *92*, 1315–1318.
- (31) Karlström, G. *J. Phys. Chem.* **1989**, *93*, 4952–4955.
- (32) Serrano-Andrés, L.; Fülischer, M. P.; Karlström, G. *Int. J. Quantum Chem.* **1997**, *65*, 167–181.
- (33) Bayley, P. M.; Nielsen, E. B.; Schellman, J. A. *J. Phys. Chem.* **1969**, *73*, 228–243.
- (34) Woody, R. W.; Tinoco, I. *J. Chem. Phys.* **1967**, *46*, 4927–4945.
- (35) Woody, R. W. *J. Chem. Phys.* **1968**, *49*, 4797–4806.
- (36) Goux, W. J.; Hooker, T. M. *Biopolymers* **1980**, *19*, 2191–2208.
- (37) Tinoco, I. *Adv. Chem. Phys.* **1962**, *4*, 113–160.
- (38) Davydov, A. S. *Theory of molecular excitations*; Plenum Press: New York, 1971.
- (39) Rogers, D. M.; Hirst, J. D. *Biochemistry* **2004**, *43*, 11092–11102.
- (40) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing*; Cambridge University Press: Cambridge, 1992.
- (41) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.
- (42) Orengo, C. A.; Michie, A. D.; Jones, S.; Jones, D. T.; Swindells, M. B.; Thornton, J. M. *Structure* **1997**, *5*, 1093–1108.
- (43) Pearl, F.; Todd, A.; Sillitoe, I.; Dibley, M.; Redfern, O.; Lewis, T.; Bennett, C.; Marsden, R.; Grant, A.; Lee, D.; Akpor, A.; Maibaum, M.; Harrison, A.; Dallman, T.; Reeves, G.; Diboun, I.; Addou, S.; Lise, S.; Johnston, C.; Sillero, A.; Thornton, J.; Orengo, C. *Nucleic Acids Res.* **2005**, *33*, 247–251.
- (44) Sreerama, N.; Woody, R. W. *Protein Sci.* **2003**, *12*, 384–388.
- (45) Manavalan, P.; Johnson, W. C. *Nature* **1983**, *305*, 831–832.
- (46) Barlow, D. J.; Thornton, J. M. *J. Mol. Biol.* **1988**, *201*, 601–619.