# Circular and Linear Dichroism Spectroscopy of Proteins


Benjamin M. Bulheller

A thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

August 2009

To my mother, my father and Heike

None of this would have been possible
without your support

# Abstract

Circular dichroism (CD) is an important technique in the structural characterization of proteins, and especially for secondary structure determination. The CD of proteins can be calculated from first principles using the matrix method, with an accuracy that is almost quantitative for helical proteins. Thus, for proteins of unknown structure, CD calculations and experimental data can be used in conjunction to aid structure analysis. 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. Calculations of the vacuum-UV CD spectra have been performed for 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, which significantly improves the agreement with experiment, raising the Spearman correlation coefficient between the calculated and experimental intensity at 175 nm from 0.12 to 0.79. The influence of the different conformations used for the calculation of charge-transfer transitions is discussed in detail, focussing on the effect in the vacuum-UV.

Linear dichroism (LD) provides information on the orientation of molecules but is more challenging to analyze than CD. To aid the interpretation of LD spectra, the calculation of protein LD using the matrix method is established and the results compared to experimental data. The orientations of five prototypical proteins are correctly reproduced by the calculations. Using a simplified approach, matrix method parameter sets for the nucleic bases and naphthalenediimide (NDI) have been created and are used to determine DNA/RNA conformations and to study NDI nanotubes. Finally, to make CD and LD calculations available for the scientific community in an easy-to-use fashion, the web interface DichroCalc is introduced.

# Publications and Presentations

## Refereed Publications

- M.T. Oakley, B.M. Bulheller, and J.D. Hirst, First Principles Calculations of Protein Circular Dichroism in the Far-Ultraviolet and Beyond, *Chirality*, 2006, **18**, 340–347.
- B.M. Bulheller, A. Rodger, and J.D. Hirst, Circular and Linear Dichroism of Proteins, *Phys. Chem. Chem. Phys.*, 2007, **9**, 2020–2035.
- B.M. Bulheller, A.J. Miles, B.A. Wallace, and J.D. Hirst, Charge-Transfer Transitions in the Vacuum-Ultraviolet of Protein Circular Dichroism Spectra. *J. Phys. Chem. B*, 2008, **112**, 1866–1874.
- B.M. Bulheller and J.D. Hirst, DichroCalc—Circular and Linear Dichroism Online, *Bioinformatics*, 2009, **25**, 539–540.
- B.M. Bulheller, G.D. Pantoş J.K.M. Sanders, and J.D. Hirst, Electronic Structure and Circular Dichroism Spectroscopy of Oligomeric Naphthalenediimides, *Phys. Chem. Chem. Phys.*, 2009, **11**, 6060–6065.
- B.M. Bulheller, A. Rodger, M.R. Hicks, T.R. Dafforn, L.C. Serpell, K.E. Marshall, E.H.C. Bromley, P.J.S. King, K.J. Channon, D.N. Woolfson, and J.D. Hirst, Flow Linear Dichroism of Some Prototypical Proteins, *J. Am. Chem. Soc.*, *accepted*.

## Poster Presentations

- Calculating Electronic Excitations in Proteins

  $27^{th}$ RSC Graduate Student Meeting, $18^{th}$ April 2007.

- DichroCalc—Circular and Linear Dichroism of Proteins Online

  $11^{th}$ International Conference on Circular Dichroism, Groningen (Netherlands), $2^{nd}$ – $6^{th}$ September 2007.

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

  - $11^{th}$ International Conference on Circular Dichroism, Groningen (Netherlands), $2^{nd}$ – $6^{th}$ September 2007.

    Poster prize by *Chirality*.

  - COSTUK Meeting, Oxford, $10^{th}$ – $12^{th}$ September 2007.

  - $28^{th}$ RSC Graduate Student Meeting, Uni. of Manchester, $16^{th}$ April 2008.

    Poster prize by *PCCP*.

- Circular and Linear Dichroism of Proteins

  $18^{th}$ Annual Northern Universities Meeting on Chemical Physics, Sheffield, $25^{th}$ June 2008.

## Oral Presentations

- Computing Dichroism of Proteins

  $28^{th}$ RSC Graduate Student Meeting, University of Manchester, $16^{th}$ April 2008.

- Circular and Linear Dichroism of Proteins

  $17^{th}$ Annual Northern Universities Meeting on Chemical Physics, University of Leeds, $11^{th}$ July 2007.

# Acknowledgments

First of all, I am most grateful to my supervisor, Prof. Jonathan Hirst, for all his help, guidance and patience.

The project has been part of a collaborative project and I want to thank Alison Rodger, Matthew Hicks and Tim Dafforn for their support. Other collaborators included Bonnie Wallace, Andrew Miles, Jeremy Sanders and Dan Pantoş, who are thanked for the fruitful cooperations. Funding was gratefully received from the Engineering and Physical Sciences Research Council (EPSRC).

The A11 crew is thanked for refuge and the many times of good laughs and raised moods. Craig Bruce has always been a godsend in all matters of computational problems. A big '*Ta*!' goes to Clare Evans for the charm*m*ing structures and for dealing with the beast. I am also grateful to Tim Lillestolen, who, in the absence of energy for writing all the stuff once on the wiki, is always happy to explain it trifold. I thank Mark Oakley for helpful discussions and David Robinson for his help and expertise.

Daniel Barthel helped a great deal in getting the PDB parser and, therefore, DichroCalc on its way. Hours upon hours of impulsive discussions—where two bullheads tried to convince each other of their immutable opinion—have been most productive and I owe him my sincere gratitude.

Tom Struthers is thanked for the great work we did together on the nucleic bases and he will always be remembered in connection with the best conference I've ever attended.

Donald, Leslie, Bram and Steve are acknowledged for all the stress and sweat that was spared on their account during the past four years.

In order to complete a PhD thesis one needs endurance and patience. Loads of them. While these characteristics do also come in handy for the one actually working on the thesis, they are even more important for his loved ones. On the one hand, this means the family, and I am deeply grateful to Heike and Elmar for enduring unwarranted snappy responses during seemingly endless times of stress.

On the other hand—even more critical—this means the partner, who has to bear all moods first-hand, ranging from just plain bad through to foul, all the way to hostile. If there were degrees for such achievements then Susi would deserve *at least* a couple of them. At the end of the day, it was she who had to handle all the piled-up bad moods and stress situations and for all that my deepest love, gratitude, and the biggest *Thank you* are for her.

# Contents

**List of Figures** xi

**List of Tables** xiv

**List of Abbreviations** xv

**1 Introduction** 1

1.1 Calculation of Circular Dichroism . . . . . 2

1.2 Theoretical Methods . . . . . 4

1.3 Circular Dichroism . . . . . 11

1.3.1 Theory of CD . . . . . 14

1.3.2 Different Parametrizations of the Amide Chromophore . 17

1.3.3 Influence of the Side Chains . . . . . 21

1.3.4 Comparing Theory and Experiment . . . . . 23

1.3.5 Computational Tractability . . . . . 31

1.3.6 Amide Transitions in the VUV . . . . . 35

1.3.7 Applications and Other Developments . . . . . 37

1.4 Conclusion . . . . . 38

**2 Charge-Transfer Transitions** **39**

2.1 Introduction . . . 40

2.2 Correlation with Experiment . . . 45

2.3 Influence of Secondary Structure . . . 49

2.4 Individual Contributions of Charge-Transfer Transitions . . . 51

2.5 Conclusion . . . 56

**3 Linear Dichroism** **57**

3.1 Introduction . . . 58

3.2 Molecular Alignment Techniques . . . 64

3.3 Examples of Protein LD Data . . . 68

3.4 Protein LD Calculations . . . 73

3.5 Comparing Theory and Experiment . . . 75

3.6 Conclusion . . . 91

**4 Nucleic Bases** **93**

4.1 Introduction . . . 94

4.2 Methods . . . 96

4.3 Polynucleotide Conformations . . . 104

4.4 Results and Discussion . . . 106

4.4.1 Nucleic Base CD Calculations . . . 107

4.4.1.1 DNA and RNA in the Far-UV and Near-UV . . . 107

4.4.1.2 DNA and RNA in the VUV . . . 110

4.4.1.3 Guanine-Quadruplexes . . . 113

4.4.2 Nucleic Base LD Calculations . . . 114

4.5 Conclusion . . . 117

**5 Naphthalenediimides** **119**

5.1 Introduction . . . . . 119

5.2 Methods . . . . . 123

5.3 Results and Discussion . . . . . 128

5.4 Conclusion . . . . . 133

**6 Online Calculation using DichroCalc** **134**

6.1 Introduction . . . . . 134

6.2 Approach . . . . . 136

6.3 Implementation . . . . . 138

6.4 The PDB Parser . . . . . 142

6.5 Conclusion . . . . . 143

**7 Conclusions** **144**

**A Bibliography** **149**

**B CD Spectra of the CD130 Protein Set** **177**

**C CD Spectra of the SP175 Protein Set** **180**

**D Orientation Search Report** **193**

# List of Figures


1 Parameters of a Gaussian curve . . . . . 10
2 Characteristic CD spectra of secondary structure elements . . . 12
3 Electronic transitions of the amide group in the far-UV . . . . . 13
4 Left and right circularly polarized light, propagating in space . 14
5 Circularly polarized light in time and space . . . . . 15
6 Monopole positions of the *ab initio* amide parameter set . . . . . 18
7 Monopole positions of the semiempirical amide parameter set . 18
8 Contour plots of the amide electrostatic potential . . . . . 20
9 Far-UV and near-UV spectra of cytochrome C . . . . . 22
10 Calculated spectra from $NMA_{initio}$ and $NMA_{semi}$ parameters . . 24
11 Effects of different bandwidths on calculated spectra . . . . . 25
12 Spearman rank correlation for CD190 . . . . . 27
13 Comparison of experimental and calculated intensity (CD190) . 28
14 Spectra calculated from the single chains of proteins . . . . . 33
15 Spearman rank correlation of diagonal matrix calculations . . . 34
16 Spectra using two and four backbone transitions . . . . . 36
17 Spearman rank correlation for SP175 . . . . . 46
18 Improvement in the deep-UV considering CT transitions . . . . 51
19 Spectra using or omitting specific CT conformations . . . . . 54

20 Effect of single CT transitions in a typical $\alpha$-helical protein . . . 55
21 Effect of single CT transitions in a typical $\beta$-sheet protein . . . . 55

22 Schematic illustration of an LD experiment . . . . . 59
23 Polarizations of transitions in $\alpha$-helix, $\beta$-sheet, and $P_{II}$-helix . . 60
24 Couette flow cell . . . . . 65
25 Uniaxial and biaxial orientation . . . . . 66
26 Schematic illustration of an LD spectrum for an $\alpha$-helix . . . . . 69
27 Experimental LD spectrum of F-actin and tubulin . . . . . 69
28 Experimental LD spectra of bacteriorhodopsin . . . . . 72
29 LD spectra of an $\alpha$-helix aligned to two different axes . . . . . 74
30 Orientations of the SAF chosen for the calculations . . . . . 78
31 Absorbance and LD spectra of the SAF . . . . . 78
32 Orientations of tropomyosin chosen for the calculations . . . . . 80
33 Absorbance and LD spectra of tropomyosin . . . . . 80
34 Orientations of FtsZ chosen for the calculations . . . . . 82
35 Absorbance and LD spectra of FtsZ . . . . . 82
36 Orientations of the amyloid fibril chosen for the calculations . . 84
37 Absorbance and LD spectra of the amyloid fibril . . . . . 84
38 Orientations of collagen chosen for the calculations . . . . . 88
39 Absorbance and LD spectra of collagen . . . . . 88

40 Structures of the nucleic bases in DNA and RNA . . . . . 95
41 Parameter set data of adenine and guanine . . . . . 99
42 Parameter set data of cytosine and uracil . . . . . 100
43 Parameter set data of thymine . . . . . 101
44 Characteristic CD spectra of polynucleotide conformations . . . 105
45 DNA·DNA and DNA·RNA duplexes . . . . . 108

46 RNA·DNA and RNA·RNA duplexes . . . . . 108
47 DNA·DNA duplex in B- and Z-form . . . . . 110
48 A- and Z-form RNA·RNA in the VUV . . . . . 111
49 Bulged DNA . . . . . 112
50 Spectra of Guanine quadruplexes . . . . . 114
51 Guanine orientations in FtsZ . . . . . 116
52 LD spectra of FtsZ with different orientations of guanine . . . . 116
53 NDI enantiomer structures . . . . . 120
54 Helical arrangement of an NDI trimer . . . . . 121
55 NDI experimental spectra . . . . . 122
56 Parameter set data of NDI . . . . . 126
57 Absorbance and CD spectra of NDI . . . . . 129
58 Dependency of the CD spectrum on the NDI monomer angle . . 132
59 Length dependency in the NDI CD spectra . . . . . 132
60 Calculated CD spectrum as returned by DichroCalc . . . . . 138

# List of Tables


1 Wavelengths of interest of the CD190 set . . . . . 29
2 Large proteins, which required splitting into single chains . . . 32
3 Wavelengths and oscillator strengths of the CT chromophores . 41
4 PDB codes and names of the SP175 set . . . . . 44
5 Spearman rank correlation coefficients for SP175 . . . . . 46
6 Secondary structure content of the SP175 set . . . . . 47
7 $\beta$-II proteins in the SP175 set . . . . . 50
8 Most frequent conformations of CT chromophores . . . . . 52
9 Electronic transitions, energies and dipole moments . . . . . 62
10 Interaction energies of excitations within a single strand . . . . . 85
11 Interaction energies of excitations located on adjacent strands . 86
12 Parameter set data of the nucleic bases . . . . . 102
13 CASPT2 results for NDI . . . . . 126
14 Franck-Condon transitions used for NDI . . . . . 129
15 Interaction between the $1^1B_{3u}$ and $1^1B_{2u}$ states of NDI . . . . . 131

# List of Abbreviations

| | |
|---|---|
| A | Adenine |
| BR | Bacteriorhodopsin |
| CASSCF | Complete-Active Space Self Consistent Field |
| CASPT2 | Complete-Active Space $2^{nd}$ order Perturbation Theory |
| CD | Circular Dichroism |
| CT | Charge-Transfer |
| C | Cytosine |
| FC | Franck-Condon |
| G | Guanine |
| GTP | Guanosine-triphosphate |
| LD | Linear Dichroism |
| MD | Molecular Dynamics |
| $\vec{m}$ | Magnetic transition dipole moment |
| $\vec{\mu}$ | Electric transition dipole moment |
| NDI | Naphthalenediimide |
| nm | Nanometre |
| NMA | *N*-methylacetamide |
| NMR | Nuclear Magnetic Resonance |
| ORD | Optical Rotatory Dispersion |
| PDB | Protein Data Bank |
| SCF | Self Consistent Field |
| SCRF | Self Consistent Reaction Field |
| SRCD | Synchrotron Radiation Circular Dichroism |
| T | Thymine |
| U | Uracil |
| UV | Ultraviolet |
| VMD | Visual Molecular Dynamics |
| VUV | Vacuum-Ultraviolet |

# Chapter 1

# Introduction

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

## 1.1 Calculation of Circular Dichroism

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

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

Building on the foundation laid by the aforementioned seminal studies, it has become feasible to calculate the CD spectra of molecules and today it is, in fact, fairly routine for small molecules, for example, to determine the absolute configurations of isolated or synthesized compounds.[19–24] The computation of protein CD, however, remains a challenge, due to the size and flexibility of the molecules. For the calculation of the optical spectra of large molecules, such as proteins and crystals, several methods have been developed specifically to cope with the size of the systems. The dipole interaction model[25–32] considers atoms and chromophores as acting as point dipole oscillators, which interact through mutually induced dipole moments in the presence of an electric field. Another approach is the matrix method,[33] which will be discussed in detail.

Conventional CD spectrometers can record spectra down to approximately 190 nm and provide information on excitation bands in the near- and far-UV.[34] Below 190 nm, a low signal-to-noise ratio makes taking measurements more challenging, due to absorption by the solvent and the low intensity of light. Using short path-length cells at high concentrations, it is possible to measure down to 178 nm.[35] Synchrotron radiation CD (SRCD)[36–39] 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 ratio[40] and contain information about more electronic transitions[41] than data collected by conventional spectrometers.[35,42,43] The additional data in the VUV in combination with calculations may provide new insights into protein folding. For example, SRCD measurements in combination with *ab initio* calculations were used to study the influence of solvent molecules on dialanine.[39] 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.

## 1.2 Theoretical Methods

There are several methods to compute the CD spectra of proteins. In 1962, Tinoco adopted a perturbation approach, in which he considered the chromophores of a protein separately[44] and assumed that electrons are localized on a particular chromophore, with Coulombic forces the only means of interaction between different chromophores.[45] The matrix method,[33,46,47] which is derived from the Frenkel exciton model,[16] is commonly used for molecular crystals and chromophore aggregates. It is an improved formulation of the Tinoco[44] method and involves solving the eigenvalue problem via a matrix diagonalization rather than by applying perturbation theory. This approach is hence more accurate (especially for degenerate and near-degenerate states) and easily implemented in computer algorithms. In the matrix method, orbitals on different chromophores are assumed not to overlap so that no inter-chromophore charge-transfer occurs. The matrix method is fairly successful in calculating the CD spectra of proteins. Proteins with a high amount of $\alpha$-helix can be calculated almost quantitatively.[48]

In the matrix method, the protein is split into $M$ independent chromophores, with a monomer wave function $\phi_{is}$ for each chromophoric group $i$ and electronic state $s$. The protein's $k^{\text{th}}$ excited state wavefunction, $\psi^k$, is then written (to a first approximation) as a linear combination of electronic configurations, $\Phi_{ia}$, in which only one chromophoric group, $i$, is in an excited state, $a$, and the others are in the ground state, 0.[49] Thus

$$\Phi_{ia} = \phi_{10} \ldots \phi_{ia} \ldots \phi_{j0} \ldots \phi_{M0} \,, \tag{1}$$

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

$$\psi^k = \sum_i^M \sum_a^{n_i} c_{ia}^k \Phi_{ia} \; . \tag{2}$$

$c_{ia}^k$ are expansion coefficients, which, for example, account for the constructive (in-phase) or destructive (out-of-phase) interaction of states. Similarly, the ground state of the protein is

$$\psi^0 = \phi_{10} \ldots \phi_{i0} \ldots \phi_{j0} \ldots \phi_{M0} \; . \tag{3}$$

Each transition in the protein CD spectrum is characterized by an energy and a rotational strength, which is given by the Rosenfeld equation[12] and is related to the experimental intensity. Therefore, to calculate the CD spectrum, the wave functions $\psi^k$ for each electronic excited state $k$ of the protein are needed (or at least those occurring in the spectral region of interest). The result of such a calculation is simply an intensity and a wavelength of each transition. To produce a spectrum, a bandshape and bandwidth is needed for each transition. This non-trivial issue will be discussed in more detail on Pages 9 and 24.

The first summation in Equation 2 is over the $M$ chromophores; the second is over the $n_i$ electronic excitations of each individual chromophore $i$. For each state, $k$, the expansion coefficients $c_{i,a}^k$ have to be determined. To illustrate what this means in practice, consider a dipeptide with two excited states on each peptide, the lowest energy backbone $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions. $\psi^k$ is then written as

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

$$\text{where} \quad \begin{aligned} \Phi_{1,n\pi^*} &= \phi_{1,n\pi^*} \cdot \phi_{2,0} \\ \Phi_{1,\pi\pi^*} &= \phi_{1,\pi\pi^*} \cdot \phi_{2,0} \\ \Phi_{2,n\pi^*} &= \phi_{1,0} \cdot \phi_{2,n\pi^*} \\ \Phi_{2,\pi\pi^*} &= \phi_{1,0} \cdot \phi_{2,\pi\pi^*} \end{aligned} \quad . \tag{4}$$

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

$$\hat{H}\psi^k = E^k\psi^k \, . \tag{5}$$

The Hamiltonian, $\hat{H}$, describes the dynamic properties of the system and needs to be constructed to solve Equation 5. If the Hamiltonian of a local chromophore, $i$, is denoted as $\hat{H}_i$, then the Hamiltonian of the $M$ independent chromophores is simply the sum of these local chromophore Hamiltonians. When the chromophores are allowed to interact, the Hamiltonian of the protein is the sum of all local Hamiltonians, $\hat{H}_i$, for each chromophore plus the sum of all intergroup potentials, $\hat{V}_{ij}$, of the entire molecule:

$$\hat{H} = \underbrace{\sum_{i=1}^{M} \hat{H}_i}_{\hat{H}_0} + \underbrace{\sum_{i=1}^{M-1}\sum_{j=i+1}^{M} \hat{V}_{ij}}_{\hat{V}} \, . \tag{6}$$

Combining Equations 2, 5 and 6, gives an equation in terms of chromophore wavefunctions and the coefficients $c^k_{i,a}$. It is convenient to convert this into a matrix formalism.

The matrix form of Equation 6 permits the calculation of the energy (the eigenvalues) and the wave functions (eigenvectors, that is, the coefficients $c_{i,a}^{k}$) by solving an eigenvalue problem through matrix diagonalization. The Hamiltonian matrix is diagonalized by a unitary transformation using the matrix $U$:

$$U^{-1} \cdot \hat{H} \cdot U = H_{diag} \, . \tag{7}$$

The diagonal elements of the resulting diagonal matrix (the eigenvalues) are the transition energies (excited state energies, since the ground state is set to be zero) of the interacting system and the eigenvectors, $c_{i,a}^{k}$, form the unitary matrix $U$. The electric and magnetic transition dipole moments of the excitation from the ground state to the $k^{\text{th}}$ excited state are then calculated as follows. The transition dipole moments of the non-interacting chromophores, which are denoted $\vec{\mu}_a^0$ and $\vec{m}_a^0$, can be transformed into the interacting system using the unitary matrix from Equation 7:[49]

$$\begin{aligned} \vec{\mu}_i &= \textstyle\sum_a U_{ai} \vec{\mu}_a^0 \\ \vec{m}_i &= \textstyle\sum_a U_{ai} \vec{m}_a^0 \, . \end{aligned} \tag{8}$$

From these transition dipole moments, the rotational strengths in the interacting system are readily calculated. For example, for a dipeptide and two transitions per group, the Hamiltonian matrix constructed using the Hamiltonian of Equation 6 has the form

$$\hat{H} = \begin{pmatrix} E_{1n\pi^*} & V_{1n\pi^*;\,1\pi\pi^*} & V_{1n\pi^*;\,2n\pi^*} & V_{1n\pi^*;\,2\pi\pi^*} \\ V_{1n\pi^*;\,1\pi\pi^*} & E_{1\pi\pi^*} & V_{1n\pi^*;\,2\pi\pi^*} & V_{1\pi\pi^*;\,2\pi\pi^*} \\ V_{1n\pi^*;\,2n\pi^*} & V_{1\pi\pi^*;\,2n\pi^*} & E_{2n\pi^*} & V_{2n\pi^*;\,2\pi\pi^*} \\ V_{1n\pi^*;\,2\pi\pi^*} & V_{1\pi\pi^*;\,2\pi\pi^*} & V_{2n\pi^*;\,2\pi\pi^*} & E_{2\pi\pi^*} \end{pmatrix} . \tag{9}$$

$V_{ii}$ are interactions between states on the same group and $V_{ij}$ are interactions between different groups.

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

$$V_{i0a;\,j0b} = \int_i \int_j \phi_{i0}\,\phi_{ia}\;\hat{V}_{ij}\;\phi_{j0}\,\phi_{jb}\;d\tau_i d\tau_j\;. \tag{10}$$

If the interaction between the chromophores is regarded as an electrostatic interaction between charge densities $\rho$ of separation $r$, that is $\hat{V}_{ij} = \frac{1}{4\pi\epsilon_0 r_{ij}}$, then Equation 10 becomes:

$$V_{i0a;\,j0b} = \int_{r_i} \int_{r_j} \frac{\rho_{ia0}(r_i)\;\rho_{j0b}(r_j)}{4\pi\varepsilon_0 r_{ij}}\;d\tau_i d\tau_j\;, \tag{11}$$

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

In principle, these matrix elements could be evaluated exactly from the monomer wave functions using the integral evaluation routines in many quantum chemical packages. In practice, an approximation is introduced to make the calculations tractable. In the *monopole-monopole approximation*, the permanent and transition densities are approximated by point charges and the integrals in Equation 11 are re-cast as a sum of the Coulomb interactions of these:[44]

$$V_{i0a;\,j0b} = \sum_{s=1}^{N_s} \sum_{t=1}^{N_t} \frac{q_s q_t}{r_{st}} \,, \tag{12}$$

where $q_s$ and $q_t$ are the point charges on chromophores $i$ and $j$, and $N_s$ and $N_t$ are the number of these charges on the chromophore. Thus, at the core of matrix method calculations are the magnitudes and locations of the monopoles.[46,47] They reflect the orientation and magnitudes of the transition dipole moments and they are critical for deducing the inter-chromophore interactions. Since the CD calculation depends only on these monopoles and their distance and orientation to each other, no ad-hoc definition of the secondary structure is needed.

The implementation of the matrix method as a computer program is quite straightforward. The computation requires the coordinates of the atoms and the positions of the chromophores in the protein. Each chromophore is parametrized by a set of monopoles describing the electrostatic potential and for every group in the protein the respective set of monopoles is superposed on the chromophore's atoms. By calculating the interaction between

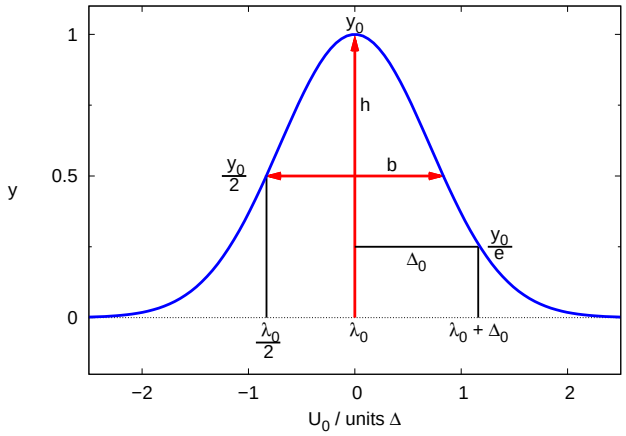



Figure 1: Parameters of a Gaussian curve. The full width at half maximum height is the bandwidth $b$.

all the different electronic excitations, the Hamiltonian matrix is constructed and, as described above, the rotational strengths are determined. This yields a line spectrum, that is, values for the rotational strength of each transition. However, in an experimental spectrum, the transitions are broadened due to the uncertainty principle, unresolved vibronic components, and the interaction of the chromophore with its environment, including other chromophores and the solvent. Thus, an overlay of approximately Gaussian-shaped bands is observed. Since the result of the calculation is a line spectrum, a convolution with a bandshape function is required. Gaussian bandshapes give better results than Lorentzian curves. Hence, all the spectra shown in this work have been created by taking a Gaussian function of the type[50]

$$y = y_0 \cdot e^{-U_0^2} \qquad \text{with} \qquad U_0 = \frac{\lambda - \lambda_0}{\Delta_0} \,. \tag{13}$$

The band width or full width at half maximum height of this function (Figure 1) is set to a single fixed value of 12.5 nm for all transitions. The resulting calculated spectrum can then be compared to the experimental spectrum.

## 1.3 Circular Dichroism

Among the related methods ORD,[8] LD,[51] and CD,[9,52] the latter is now by far the most popular type of chiroptical or polarized light spectroscopy. The two energy regimes in current use are vibrational transitions[53–57] and electronic transitions. In this thesis, vibrational CD is not considered, only electronic CD and henceforth the abbreviation CD refers to the latter. When circularly polarized light impinges on a protein, the protein's electronic structure gives rise to characteristic bands in specific regions in the CD spectrum, reflecting the electronic excitation energies.[50,58] Secondary structural elements, such as $\alpha$-helices, $\beta$-sheets, $\beta$-turns and random coil structures, all induce bands of distinctive shapes and magnitudes in the far-ultraviolet (Figure 2).[59] For example, in an $\alpha$-helix, an intense positive band at 190 nm and a negative band at 208 nm arise from the exciton splitting of electronic transitions from the amide non-bonding $\pi$ orbital, $\pi_{nb}$, to the antibonding $\pi$ orbital, $\pi^*$, (Figure 3). A negative band is located at about 222 nm, arising from the electronic transition from an oxygen lone pair orbital, $n$, to the $\pi^*$ orbital (Figure 3). Other motifs give other spectroscopic shapes and signs.

The relative proportion of each secondary structure type can be determined by decomposing the far-UV spectrum into a sum of fractional multiples of the reference spectra.[9,60,61] For example, the average fractional helicity, $f_H$, of a peptide consisting of $N$ residues can be determined from the observed mean residue molar ellipticity at 222 nm, $[\theta]_{222}$:

$$f_H = \frac{[\theta]_{222}}{[\theta_{H\infty}]_{222} \cdot (1 - \frac{k}{N})}\,, \tag{14}$$

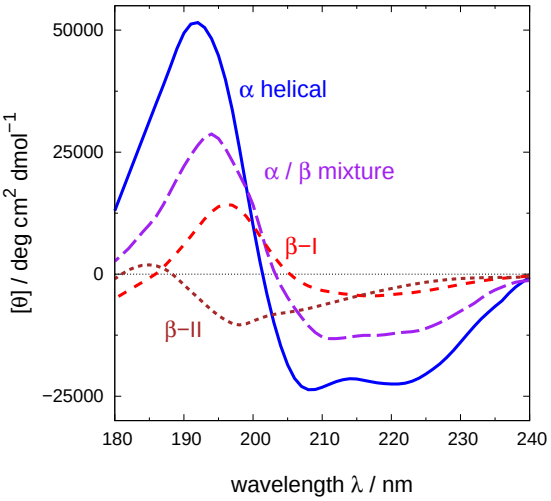



Figure 2: Characteristic CD spectra of secondary structure elements. The vertical axis shows intensity as the mean residue molar ellipticity, $\theta$.

where $[\theta_{H\infty}]_{222}$ is the mean residue molar ellipticity of a completely helical peptide of infinite length and $k$ is an end-effect correction factor of approximately three.[62] Estimates of $[\theta_{H\infty}]_{222}$ range from $-33{,}000$ deg cm$^2$ dmol$^{-1}$ [63] to $-44{,}000$ deg cm$^2$ dmol$^{-1}$, [64] which correspond to differential absorbance between $-11.2$ mol$^{-1}$ dm$^3$ cm$^{-1}$ and $-13.3$ mol$^{-1}$ dm$^3$ cm$^{-1}$. However, $[\theta]_{222}$ can be influenced by several factors,[65–67] as will be discussed later. The empirical analysis of a spectrum can, therefore, lead to the determination of the protein secondary structure. Moreover, when CD spectroscopy is coupled with time-resolved experiments, protein folding events can be studied[68] and theoretical spectra can be used to interpret the results.[65,66] Although empirical fitting works remarkably well, due to the well-defined nature of the secondary structure motifs of proteins, it is important to proceed beyond empirical data analysis for a number of reasons. Since the conformation of proteins in solution may be (perhaps subtly) different from that in the crystalline environment, calculations can help to uncouple the solution-phase CD structure analysis from crys-

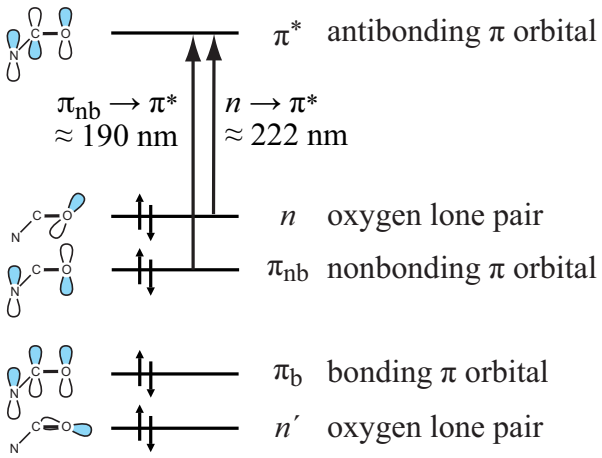



Figure 3: Electronic transitions of the amide group in the far-UV region.

tallographic structure motifs. Furthermore, the increasing importance of less well-folded or natively unfolded protein domains demands new approaches, since empirical comparison fails in these cases. In contrast to empirical data analysis, theoretical calculations, in principle, can also encompass the population dynamics of a solution of proteins and should be able to cope with new protein folds. The calculation of CD spectra from simulated ensembles of conformations can provide additional information from molecular dynamics (MD) simulations of protein folding.[69]

### 1.3.1 Theory of CD

Circularly polarized light can be produced by the superposition of two linearly polarized light beams of equal magnitude that are oscillating perpendicular to each other and propagating with a phase difference of $\frac{\pi}{2}$ radians. The magnitude of the electric field vector of the resulting beam is constant, but rotates about the propagation direction. If the tip of the "rotating" vector forms a right-handed helix, it is right circularly polarized light and vice versa (Figure 4).[4,70] A stationary observer looking towards the light source at the electric field vector of right circularly polarized light sees it rotating in an anti-clockwise sense, if regarding its progression in space. However, more important is the dependence of the field vector on time, that is, the direction of the field at the same position as a function of time. In this case, the vector of right circularly polarized light rotates *clockwise* (Figure 5).[8,56]

A solution of chiral molecules possesses different refractive indices for left and right circularly polarized light, that is, the beams travel at different speeds and are absorbed to different extents at each energy. This means that the extinction coefficients for left and right circularly polarized light are different.

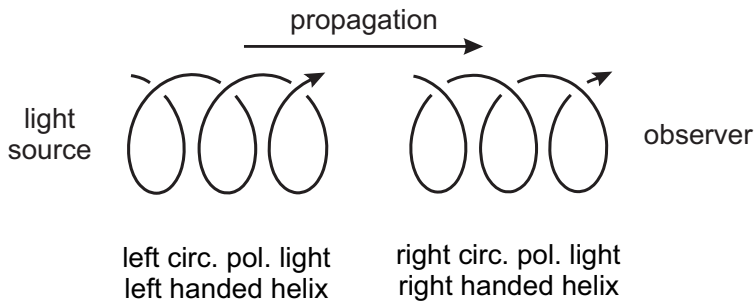



Figure 4: Left and right circularly polarized light, propagating in space.

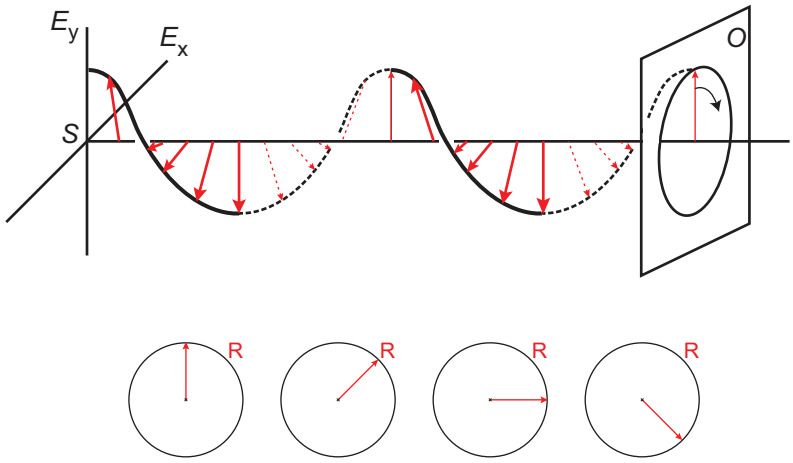



Figure 5: Dependence of right circularly polarized light on the distance from the light source, $S$, (top) and on time (bottom) when the observer at position $O$ is looking towards the source. The helix moves along the propagation direction without rotation, causing the vector of right circularly polarized light to rotate anti-clockwise in space but clockwise in time.

This effect, $\varepsilon_L \neq \varepsilon_R$, is called CD and the differential absorbance, $\Delta\varepsilon$, of left and right circularly polarized light (Equation 15) is plotted against the wavelength $\lambda$ to yield the CD spectrum.[9,49,60]

$$\Delta\varepsilon = \varepsilon_L - \varepsilon_R \,. \tag{15}$$

The integral of $\Delta\varepsilon$ over a wavelength range associated with a particular transition is known as the CD strength or rotational strength of that transition. It is analogous to the oscillator strength of normal absorption. From the quantum electrodynamic viewpoint, the rotational strength, $R^{0k}$, of a transition from the ground state 0 to an electronically excited state $k$ is the product of the electric transition dipole moment, $\vec{\mu}$, and the magnetic transition dipole moment, $\vec{m}$. The probability of a transition $0 \rightarrow k$ is represented by the integral $\langle \psi^0 | \vec{\mu} | \psi^k \rangle$, which can be regarded as an oscillating dipole induced by the light beam.[60]

$\vec{\mu}$ describes a linear displacement, whereas $\vec{m}$ characterizes a circulation of charge and accordingly the integral $\langle \psi^k | \vec{m} | \psi^0 \rangle$ can be understood as a light-induced current loop. Thus, the combination of $\vec{\mu}$ and $\vec{m}$ creates a helical displacement of charge, leading to a different interaction with left and right circularly polarized light. While the operator for $\vec{\mu}$ is a real vector, that for $\vec{m}$ is imaginary, as it describes the rotation of charge in a complex coordinate system,[60] and it involves the linear momentum operator $\vec{p}$:

$$\vec{m} = \frac{e}{2mc} \cdot (\vec{r} \times \vec{p}) \, , \qquad \vec{p} = \frac{\hbar}{i} \nabla \, , \tag{16}$$

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

$$R^{0k} = Im(\langle \psi^0 | \vec{\mu} | \psi^k \rangle \cdot \langle \psi^k | \vec{m} | \psi^0 \rangle) \, , \tag{17}$$

where $\psi^0$ and $\psi^k$ denote the wave functions of the ground state and the excited state, respectively, and $Im$ the imaginary part of the product. Using the Rosenfeld equation, it is possible to obtain the rotational strengths of a molecule and, therefore, calculate its CD spectrum. The required wave functions for the ground and excited states can be approximated using the matrix method by parametrizing the electrostatic fields caused by the involved states.

### 1.3.2 Different Parametrizations of the Amide Chromophore

As discussed in Chapter 1.2, for expedience, the Hamiltonian matrix elements may be estimated using the monopole-monopole approximation, which involves representing the transition densities of a chromophore by a set of point charges. Two modern parameter sets for matrix method calculations, arising from different approaches, are an *ab initio* derived set, $NMA_{initio}$,[71] and the set of Woody and Sreerama, $NMA_{semi}$.[72]

The $NMA_{initio}$ set has been derived entirely from *ab initio* calculations.[73,74] Using *ab initio* techniques, the amide chromophore was parametrized with N-methylacetamide (NMA) as a model compound.[71] The electronic spectrum of NMA in solution was calculated, using the complete active space self-consistent field method implemented within a self-consistent reaction field (CASSCF/SCRF),[75–78] combined with multi-configurational second-order perturbation theory (CASPT2-RF).[75,78] Monopoles for a given state were determined by fitting their electrostatic potential to reproduce the *ab initio* electrostatic potential for that state so that the least-squares difference is minimized (typically within 5%). The parameter set[71] 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, and 20 monopoles for the $\pi \rightarrow \pi^*$ transition, with one charge situated at each atom centre and four around each atom centre at a distance of 0.05 Å (Figure 6). Two higher energy transitions at 129 nm and 123 nm were also parametrized (Section 1.3.6). $NMA_{initio}$ has been tested on a set of 46 proteins, for which data were available between 190 and 250 nm.[79] The set is referred to as CD190. The calculated CD agreed well with experiment, particularly at 222 nm, where the accuracy was nearly quantitative. This wavelength is important, since the intensity of the band in that region correlates with the helical content of the protein.[80]

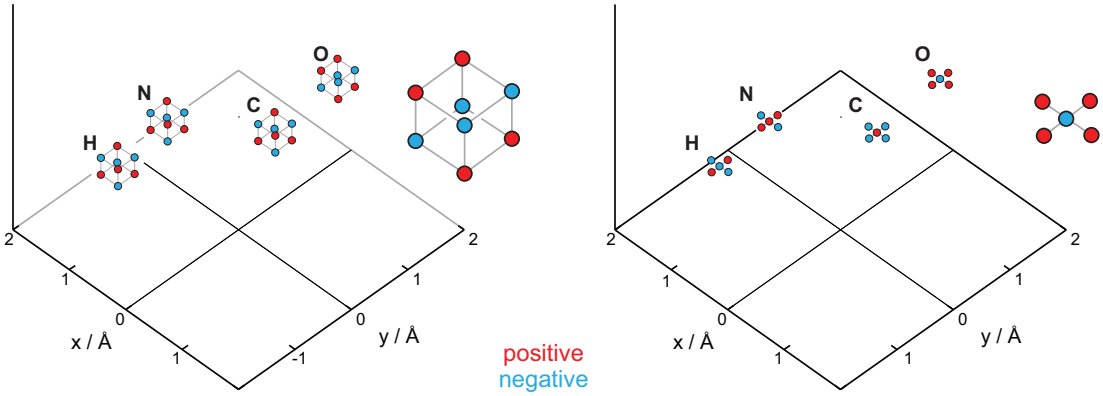



Figure 6: Monopole positions of the *ab initio* amide parameter set, $\mathrm{NMA}_{initio}$, for the $n \rightarrow \pi^*$ transition (left panel) and the $\pi \rightarrow \pi^*$ transition (right panel). Negative charges are represented by blue circles and positive charges by red circles. The distance of the monopoles to the atoms is about 0.1 Å in the left panel and either 0 or 0.05 Å in the right panel.

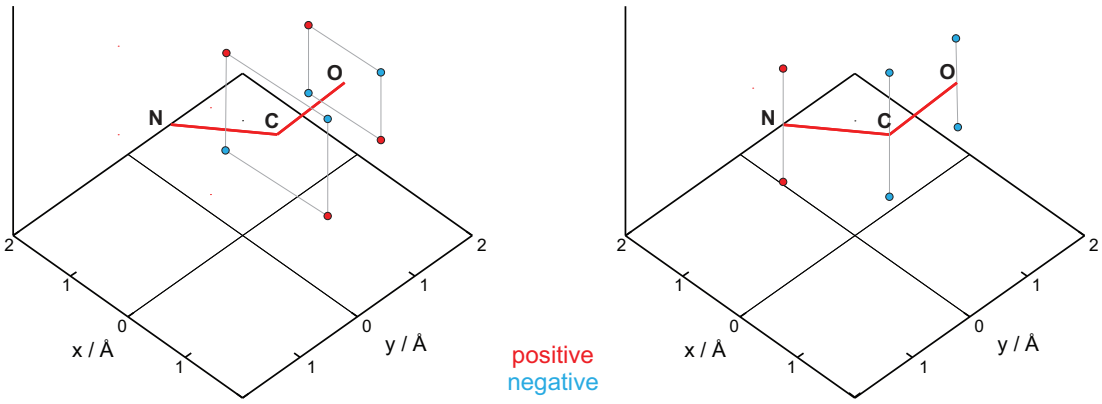



Figure 7: Monopole positions of the semiempirical amide parameter set, $\mathrm{NMA}_{semi}$, for the $n \rightarrow \pi^*$ transition (left panel) and the $\pi \rightarrow \pi^*$ transition (right panel). Negative charges are represented by blue circles and positive charges by red circles. The distance of the monopoles to the atoms is between 0.8 and 1.2 Å.

$NMA_{semi}$ is derived from a combination of semiempirical calculations using the INDO/S method and experimental data. The $n \rightarrow \pi^*$ transition density is represented by four monopoles surrounding each of the carbon and oxygen atoms, whereas for the $\pi \rightarrow \pi^*$ transition density, a total of six charges are located above and below the peptide plane (Figure 7). Compared to the *ab initio* set, it consists of fewer monopoles with smaller charges at greater distance from the atomic centres. However, the electrostatic potential created by the charges is the physically relevant quantity, rather than the particular values and locations of the charges, which are (within certain constraints) arbitrary, as an infinite number of monopole sets give rise to the same electrostatic potential.

For the peptide chromophore, the electrostatic potentials for the two parameter sets are compared in Figure 8, which shows a contour plot of the electrostatic potential arising from the $\pi \rightarrow \pi^*$ transition density 0.5 Å above the plane of the peptide group. In $NMA_{semi}$, the monopoles located around the carbon and oxygen produce a symmetric potential around the CO bond. Due to the additional monopoles at the nitrogen and hydrogen atoms, the electrostatic field generated by $NMA_{initio}$ is asymmetric. The most significant difference between the two parameter sets lies indeed with the $n \rightarrow \pi^*$ transition, as the $\pi \rightarrow \pi^*$ transitions are highly dipolar and look very similar in shape and magnitude. However, there is a difference in the orientation of the $\pi \rightarrow \pi^*$ transition dipole moment of 8.5°, and the importance of this orientation has been discussed in detail in the literature.[48,72]

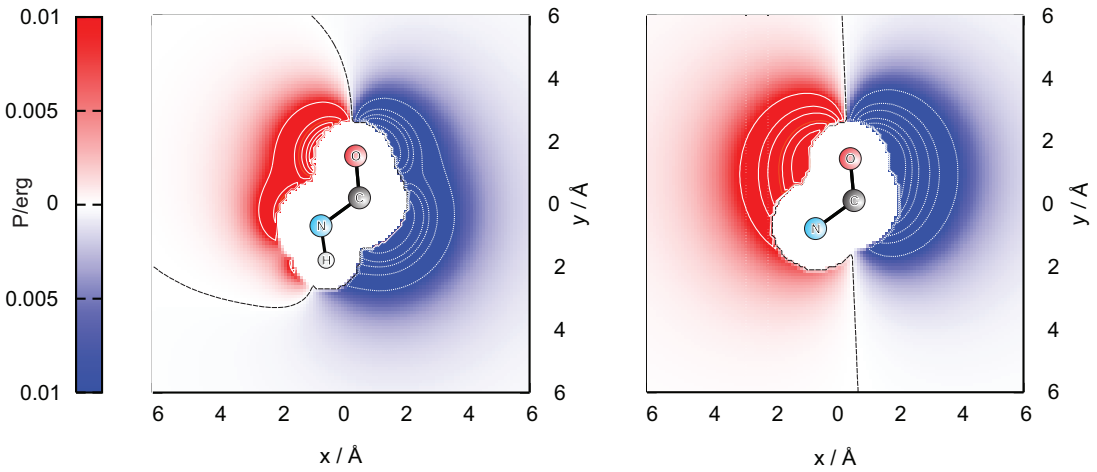



Figure 8: Contour plots of the electrostatic potential $P$ of the $\pi \rightarrow \pi^*$ transition caused by a peptide chromophore for the *ab initio* parameter set, $\mathrm{NMA}_{initio}$ (left panel) and for the semiempirical parameter set, $\mathrm{NMA}_{semi}$ (right panel). Solid lines (left half of the potential, red) denote a positive, dotted lines (right half of the potential, blue) a negative potential. The contour line of the zero-crossing is dashed.

### 1.3.3 Influence of the Side Chains

Apart from the peptide group, there are other chromophores of interest. The chromophoric groups of the side chains dominate the CD spectrum in the near-UV (250 nm – 350 nm). The intensity in this region is about two orders of magnitude lower than in the far-UV in most proteins, reflecting the fact that only a few percent of the residues in any protein contain aromatic groups. For the same reason, the backbone amide transitions dominate the far-UV in most proteins and the aromatic groups are observed clearly only above 250 nm, where the amide group has no electronic transitions. In the near-UV, as in the far-UV, chromophore parameter sets based on *ab initio* or semiempirical calculations on individual chromophores may be constructed. Our approach has been to gain a more accurate description of the valence electronic transitions of the side chain chromophores by performing *ab initio* calculations.[81] Benzene, phenol and indole are the chromophoric groups of the amino acids phenylalanine, tyrosine and tryptophan, respectively. The influence of solvent on the most important transitions of each chromophore has also been investigated.[81] The excited states are by convention labelled as $^1L_b$, $^1L_a$, $^1B_b$ and $^1B_a$, in which the superscript indicates that the state is a singlet state and the subscript indicates the orientation of the electric transition dipole moment as along either (a) the long or (b) the short axis of the ring in phenylalanine and tyrosine.[82] $L$ refers to the low-lying transitions and $B$ to the transitions at high energy. In tryptophan the moments are inclined to the long axis of the indole chromophore at about $+54^\circ$ ($^1L_b$), $-41^\circ$ ($^1L_a$), $+18^\circ$ ($^1B_b$) and $-61^\circ$ ($^1B_a$).[83] The four states of phenylalanine, tyrosine and tryptophan are responsible for the majority of the spectroscopic bands due to side chains in proteins. The $^1B$ states are more intense than the $^1L$ states and are located in the far-UV region.

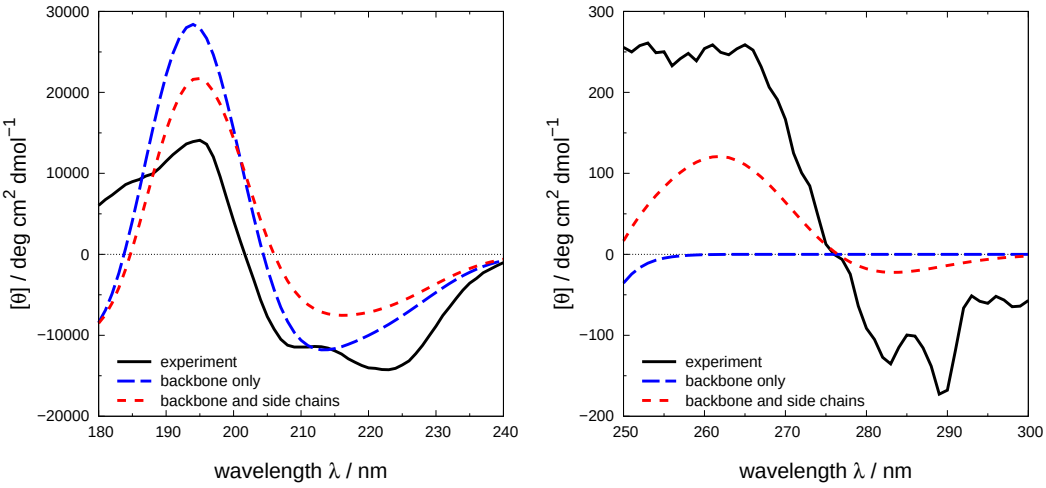



Figure 9: CD spectra of cytochrome C (1hrc) in the far-UV (left) and the near-UV (right).

*Ab initio* calculations reproduced the solvatochromatic shifts of the individual chromophores well, although the oscillator strengths were underestimated for benzene and vibronic coupling was neglected. The derived parameter sets were tested on the near-UV spectra of 30 proteins and on the difference spectra of 20 mutants and in both cases gave a significant improvement on the semiempirically derived parameters.[84]

Cytochrome C (PDB code 1hrc), for example, contains four phenylalanine, one tryptophan and four tyrosine residues of a total of 105 amino acids. The inclusion of the aromatic groups in the calculations correctly reproduces two bands in the near-UV (Figure 9, right). In the far-UV the intensity generally decreases, leading to less agreement with the experimental spectrum between 210 and 230 nm, but an improvement of the band near 190 nm. In the case of cytochrome C, the most influential aromatic chromophores in the far-UV are the tyrosines, whose $^1B$ states interact with some of the peptide chromophores.

### 1.3.4 Comparing Theory and Experiment

Figure 10 (left panel) compares calculated spectra using the $NMA_{initio}$ and $NMA_{semi}$ parameters with the experimental spectrum of a helical protein. The peak around 193 nm is well matched by the *ab initio* parameter set, whereas the peak of $NMA_{semi}$ is slightly shifted to higher energy. In the region between 200 and 230 nm $NMA_{semi}$ indicates the double minimum, which is characteristic for a helical protein, but not resolved for $NMA_{initio}$. For the $\beta$-sheet protein (Figure 10, right panel) the intensity at 195 nm is overestimated by both parameter sets, whereas $NMA_{initio}$ shows a better fit at longer wavelengths, including again a good estimation at 222 nm. In $\beta$-II proteins,[85] the strands are rather short or not aligned in a parallel manner, but often twisted and bent. The resulting spectra are similar to unfolded, random coil proteins (Figure 2, Page 12) and the agreement of the calculations with experiment is poor. Around 222 nm the intensity is reproduced well, but at higher energies, especially at 200 nm where the minimum is not predicted at all, no correlation between theory and experiment is achieved.

Another structure occurring in peptides is the poly(proline)-II-helix ($P_{II}$, $\phi = 75°$, $\psi = +145°$). The $P_{II}$ conformation is the subject of much current research and has been identified as a possible origin for the distinction between $\beta$-I and $\beta$-II in $\beta$-rich proteins. Sreerama and Woody[85] related the observation of different types of spectra for $\beta$-I and $\beta$-II proteins back to the content of $P_{II}$-conformations, that is, proteins possessing a ratio of $P_{II}$ to $\beta$-sheet less than 0.4 appear to belong to the $\beta$-II-type, whereas $\beta$-II-proteins show a ratio of $P_{II}$ to $\beta$-sheet greater than 0.4.[85] The CD spectrum of the $P_{II}$ conformation exhibits a strong negative band at 200 nm, which is not reproduced in calculations.[49] Another structural factor is the number of disulfide bridges.

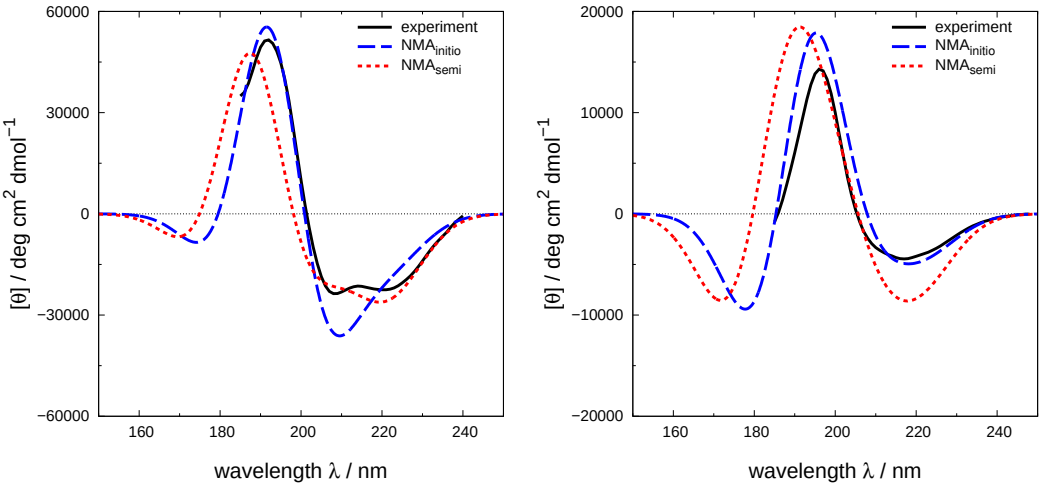



Figure 10: Comparison of the experimental spectra of colicin A (1col, mostly $\alpha$-helical, left panel) and green fluorescent protein (1ema, $\beta$-I class, right panel) with the spectra calculated using two different parameter sets. Experimental (solid), *ab initio* parameter set (dashed), semiempirical parameter set (dotted).

The agreement between calculated and experimental spectra deteriorates with more irregular structure and greater numbers of disulfide bonds.

In order to enhance the quality of the calculations of CD spectra, many aspects have been studied, for example, the inclusion of additional chromophores[81] and investigations into hydrogen bonding.[86] Another parameter that has been examined, is the width of each band at a specific transition energy. In most calculations, this is set to a single fixed value for all wavelengths. Wavelength-dependent bandwidths have also been tested, but can lead to unexpected results. The resulting CD spectra when all transitions were assigned the same bandwidth between 8.5 nm and 13.5 nm are compared in Figure 11. To improve the resolution of the two peaks around 208 nm and 222 nm for helical proteins, bandwidths between 7.5 nm and 15.5 nm have been explored.[79] The minimum of the $n \rightarrow \pi^*$ transition at 222 nm is resolved for values below 12.5 nm, but such narrow bandwidths then exaggerate the

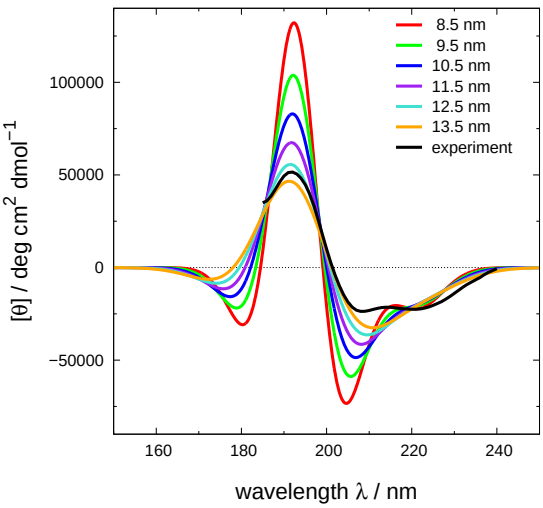



Figure 11: Calculated CD spectra where all transitions have been assigned the same bandwidth, with values ranging from 8.5 to 13.5 nm.

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

To assess the quality of the matrix method calculations, the calculated and experimental spectra of a representative set have to be compared and analyzed in order to get a statistically meaningful result. A useful statistic is the Spearman rank correlation coefficient,[87] between the experimental and calculated intensity at each wavelength for each of the proteins. The coefficient is a measure of the strength of the associations between the experimental and calculated spectra and ranges from 1 (perfect correlation) over 0 (no correlation) to

−1 (inverse correlation). The Spearman rank correlation is a nonparametric statistic, which does not require any assumptions about the probability distribution of the data, thus making it superior to parametric methods, as it is more reliable and robust.

The accuracy of the calculations was assessed on a set of 46 proteins (Figure 12). $\alpha$-Bungarotoxin (2abx) has been removed from the initial set, because an analysis using PROCHECK[88] showed that only 14.8% of the residues lie in the core regions of the Ramachandran plot and 24.6% in disallowed regions, thus questioning the structural integrity of the file. At 190 nm the $NMA_{semi}$ parameters give a correlation of 0.68 compared to 0.85 of $NMA_{initio}$. In this region, the intensity is mostly overestimated by both sets (Figure 13, top), but the overall correlation is positive. Notably, bacteriorhodopsin (2brd) is an outlier, despite being mostly $\alpha$-helical. It is a membrane protein with a high helical content. Its CD spectrum is less intense than the spectra of globular proteins with similar helical content and, thus, the intensity of the spectra is overestimated by both parameter sets.

The point where the CD intensity changes sign, at around 200 nm, is a challenge for both parameter sets. In this region the gradient of the spectrum is greatest, causing a high difference in intensity for even small deviations of the calculated versus the experimental wavelength. Although the absolute error in the wavelength where the intensity is calculated to be zero is below 3 nm (Table 1, Page 29), the intensities may differ up to 10,000 deg $cm^2$ $dmol^{-1}$. Hence, the correlation is almost zero for $NMA_{initio}$ and is actually negative for $NMA_{semi}$. $NMA_{initio}$ tends to overestimate the intensity, that is, the calculated wavelength of the zero point is too large and the spectrum slightly redshifted, whereas the deviations of $NMA_{semi}$ are usually equally distributed.

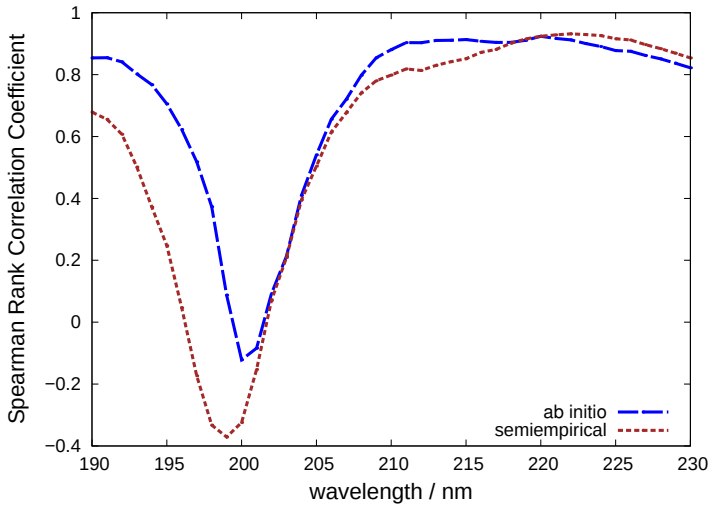



Figure 12: Spearman rank correlation between the calculated and experimental intensity for the 46 proteins comprising the CD190 set. Dashed: using the *ab initio* parameter set ($NMA_{initio}$). Dotted: using the semiempirical parameter set ($NMA_{semi}$).

The correlation steadily increases after the zero-crossing and both parameter sets reach almost constant values of between 0.8 and 0.9 above 210 nm. The band at 208 nm, arising from the exciton splitting of the $\pi \rightarrow \pi^*$ transition, shows a correlation of 0.80 and 0.75 for $NMA_{initio}$ and $NMA_{semi}$, respectively. At 222 nm, which is important for the determination of helical content, both sets give a comparably (high) correlation (Figure 13, bottom).

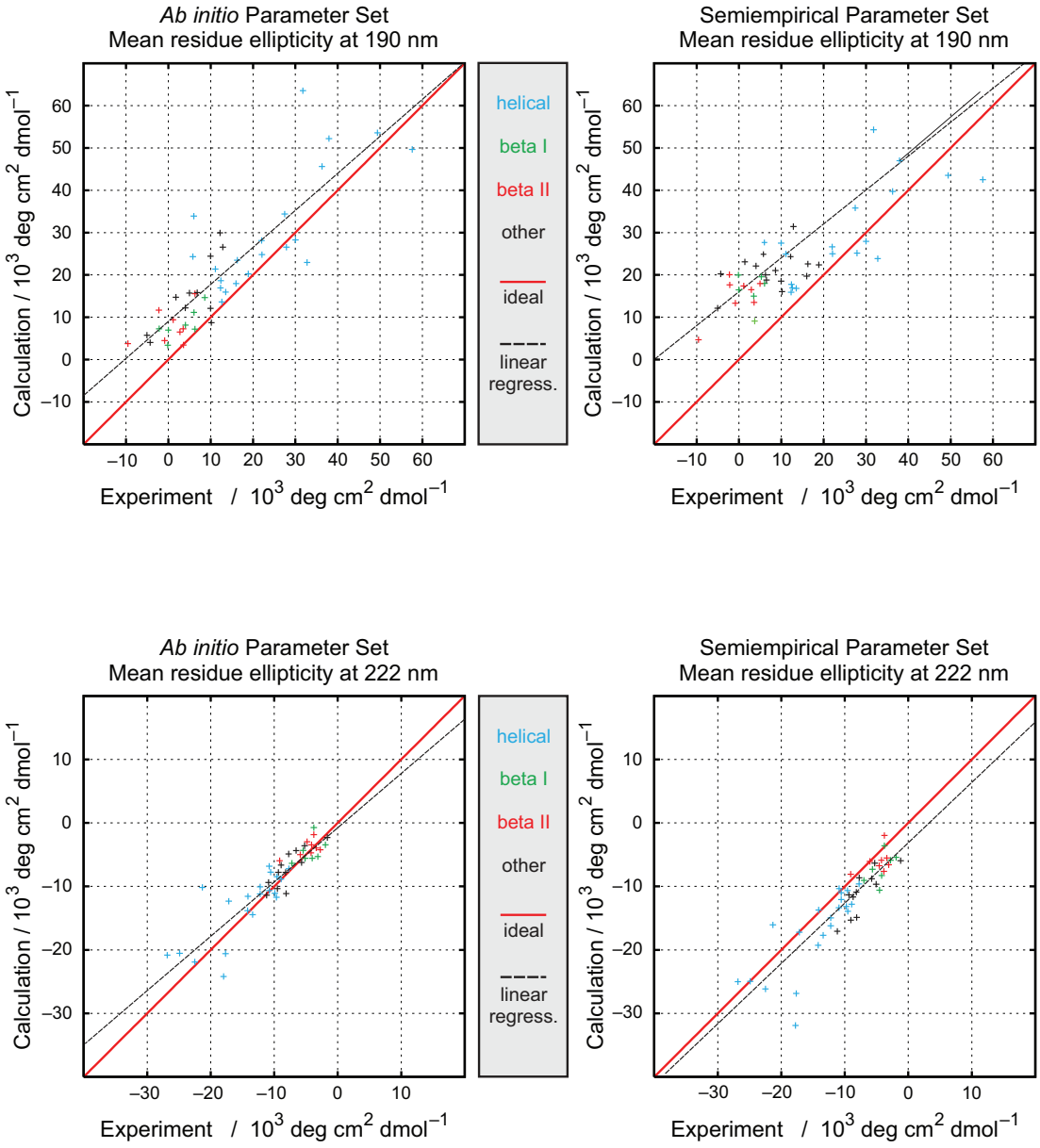



Figure 13: Comparison of experimental and calculated mean residue molar ellipticity at 190 nm (top) and 222 nm (bottom). The solid line corresponds to perfect correlation. The dashed line shows the linear regression through all points.

Table 1: Calculated (using NMA$_{initio}$ parameters) and experimental wavelengths of interest of the CD190 set. The 12 proteins at the top show the least agreement between the calculated and experimental spectra; $\beta$-II proteins are printed in bold. If no value is given, it means that the experimental spectrum does not show a maximum or minimum, or cross the axis, as appropriate.

| **PDB** | $\lambda_{max}$ / [nm] | | $\lambda_{\Delta\varepsilon=0}$ / [nm] | | $\lambda_{min}$ / [nm] | |
|---|---|---|---|---|---|---|
| | **Calc** | **Exp** | **Calc** | **Exp** | **Calc** | **Exp** |
| **1ca2** | 195 | — | 207 | — | 218 | — |
| 1rei | 199 | 202 | 215 | 210 | 223 | 217 |
| 2brd | 191 | 196 | 201 | 205 | 209 | — |
| **2cga** | 196 | — | 209 | — | 219 | — |
| 2rns | 195 | — | 206 | — | 217 | — |
| **2sod** | 199 | 190 | 212 | 196 | 222 | 209 |
| 3ebx | 199 | 201 | 217 | 209 | 223 | 215 |
| **3est** | 197 | — | 211 | — | 221 | — |
| **3ptn** | 198 | — | 212 | — | 222 | — |
| 4gcr | 197 | 203 | 211 | — | 221 | — |
| **5cha** | 197 | — | 211 | — | 220 | — |
| **5pti** | 193 | — | 203 | — | 211 | — |
| 1beb | 194 | 190 | 205 | 201 | 216 | 213 |
| 1col | 192 | 192 | 201 | 201 | 210 | 208 |
| 1e5z | 196 | 195 | 209 | 205 | 219 | 219 |
| 1ema | 195 | 195 | 208 | 206 | 218 | 217 |
| 1eri | 192 | 190 | 202 | 201 | 211 | 208 |
| 1fx1 | 193 | 194 | 204 | 203 | 214 | 214 |
| 1ifc | 195 | 197 | 207 | 204 | 217 | 216 |
| 1lys | 192 | 192 | 202 | 199 | 211 | 208 |

continued on next page

**Table 1 — continued from previous page**

| PDB | $\lambda_{max}$ / [nm] | | $\lambda_{\Delta\varepsilon=0}$ / [nm] | | $\lambda_{min}$ / [nm] | |
|---|---|---|---|---|---|---|
| | **Calc** | **Exp** | **Calc** | **Exp** | **Calc** | **Exp** |
| 1plc | 198 | 200 | 213 | — | 222 | — |
| 1rhd | 194 | 193 | 204 | 201 | 213 | 210 |
| 1sbt | 192 | 192 | 202 | 200 | 212 | 218 |
| 1tnf | 197 | 201 | 211 | 209 | 221 | 221 |
| 2ctv | 198 | 197 | 212 | 209 | 221 | 224 |
| 2hmz | 192 | 194 | 201 | 202 | 210 | 210 |
| 2mhb | 191 | 195 | 201 | 202 | 210 | 210 |
| 2pab | 197 | 195 | 210 | 202 | 221 | 217 |
| 2psg | 195 | 193 | 207 | 200 | 218 | 211 |
| 2sns | 193 | — | 203 | 199 | 211 | 209 |
| 3adk | 193 | 192 | 203 | 202 | 211 | 210 |
| 3gpd | 192 | 192 | 201 | 200 | 211 | 212 |
| 3grs | 194 | 191 | 204 | 199 | 212 | 207 |
| 3pgk | 192 | 191 | 202 | 200 | 211 | 209 |
| 3por | 194 | 195 | 205 | 208 | 214 | 218 |
| 3rn3 | 194 | 193 | 205 | 199 | 215 | 209 |
| 3tim | 192 | 193 | 202 | 202 | 211 | 219 |
| 4ins | 190 | 195 | 199 | 202 | 208 | 209 |
| 4mbn | 192 | 192 | 201 | 201 | 210 | 209 |
| 5adh | 194 | 196 | 205 | 203 | 217 | 211 |
| 5cpa | 193 | 198 | 203 | 204 | 212 | 211 |
| 5cpv | 191 | 195 | 200 | 202 | 209 | 209 |
| 5cyt | 194 | 195 | 205 | 202 | 214 | 210 |
| 6ldh | 193 | 192 | 203 | 202 | 212 | 212 |
| 8tln | 193 | 194 | 203 | 203 | 211 | 211 |
| 9pap | 192 | 188 | 201 | 194 | 211 | 209 |

### 1.3.5 Computational Tractability

Calculations on proteins are computationally demanding due to the size of the compounds, which can contain dozens of chains comprising thousands of atoms. The utilization of the matrix-method facilitates the use of *ab initio* parameters, but three limits have to be considered for large proteins: the maximum number of atom coordinates to define the chromophore positions, the number of chromophoric groups to be processed, and the dimension of the Hamiltonian matrix that can be diagonalized. While, in the past, those limitations have been addressed manually if encountered for a protein, the automation of the calculations for public use on the web interface DichroCalc[89] posed the need for dealing with as many PDB files as possible without manual intervention.

Due to compiler constraints, the maximum number of atoms whose coordinates can be read in is 10,000. However, only the atoms that are used to define the chromophore position and to whose positions monopole charges have been fitted are actually needed. In a peptide bond, for example, these are the atoms designated as C, N and O in the PDB and only a subset of the side chain chromophore atoms. The PDB file is, therefore, filtered in an additional step for all atoms not required in the calculation. This includes, for example, solvent molecules, which have been a reason in the past for reaching the critical atom limit. This reduces the amount of atoms to about 50% and considerably increases the number of proteins that can be processed.

The diagonalization algorithm limits the size of the Hamiltonian matrix. Inclusion of charge-transfer transitions doubles the matrix dimension. Side chain chromophores also enlarge the calculation, but to a smaller extent, which depends on the content of aromatic side chains (usually between 5 and 15%).

Table 2: Large proteins of SP175, which required splitting into single chains to calculate the CD.

| Protein | PDB Code | Chains | Residues | Residues per Chain |
|---|---|---|---|---|
| Pyruvate kinase | 1a49 | 8 | 4152 | 519 |
| Aldolase | 1ado | 4 | 1452 | 363 |
| Peroxidase | 1atj | 6 | 1836 | 306 |
| $\beta$-Galactosidase | 1bgl | 8 | 8168 | 1021 |
| Catalase | 1f4j | 4 | 1928 | 479 / 481 |
| c-Phycocyanin | 1ha7 | 24 | 3996 | 162 / 171 |
| Glutamate dehydrogenase | 1hwx | 6 | 3006 | 501 |
| Immunoglobulin | 1igt | 4 | 1316 | 214 / 444 |
| Ceruloplasmin | 1kcw | 1 | 1017 | 1017 |
| Ovalbumin | 1ova | 4 | 1519 | 386 / 373 |
| Phosphoglucomutase | 3pmg | 2 | 1122 | 561 |

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 calculated. If all single chain spectra are similar, it is most likely that the overall spectrum will look similar, since CD is not orientation dependent and interchain interactions are negligible due to the distance. There are some variations in the intensity, but these differences are relatively small. In the SP175 set there were 11 proteins that were too large to be calculated as a whole, that is, all chains in the oligomer (Table 2). The single chains usually have the same sequence; some of them have different chain termini, but this has almost no effect on the CD spectrum. An exception, however, was immunoglobin (1igt), which consists of four chains, of which chains A and C are equivalent, as are chains B and D.

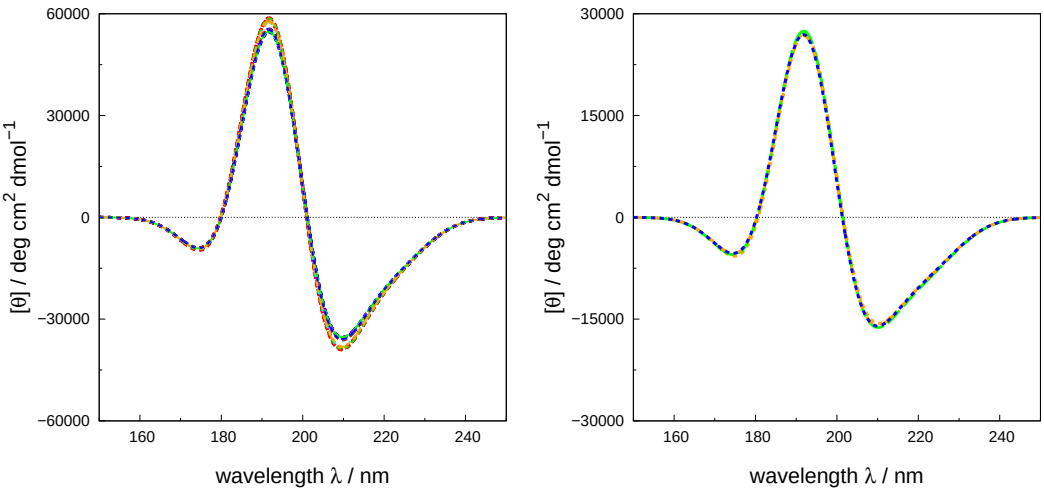



Figure 14: Comparison of spectra calculated from the single chains of proteins for which calculation would otherwise be intractable. Left: c-Phycocyanin (1ha7, 24 chains, 29,916 atoms). Right: Catalase (1f4j, 4 chains, 15,402 atoms).

Rather than calculating each separate chain, the file was split in half, thus creating two equal fractions of the protein. Figure 14 shows some examples of single chain calculations and the minor differences in the intensities. For the proteins $\beta$-galactosidase (1bgl) and ceruloplasmin (1kcw), which contain chains with more than 1000 residues (Table 2), calculations are feasible for the backbone-only calculations, but not if charge-transfer is also considered.

The maximum dimension of the Hamiltonian poses the most critical limit for the calculations, and possibilities to tackle this limitation were explored. About 98% of the off-diagonal interactions are orders of magnitude smaller than the strongest interactions. Removing these matrix elements before diagonalization did not have a substantial effect on the results of the CD190 protein set. Since the interaction of dipoles decreases with the cube of their separation, the vast majority of the influential 2% is due to close neighbour interactions situated at the diagonal of the matrix. Calculations have been performed considering different numbers of residues depending on their separation in the

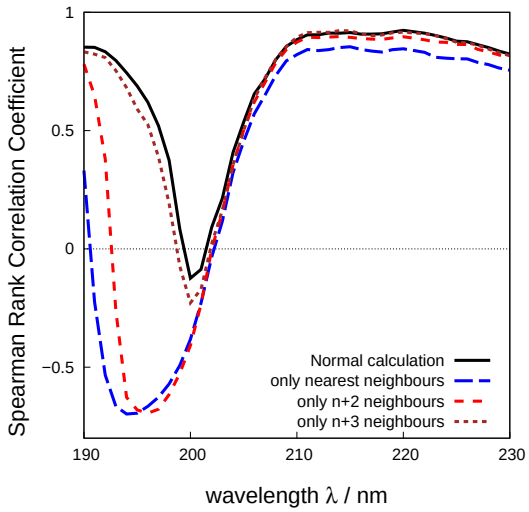



Figure 15: Spearman rank correlation coefficient of the CD190 set for distance-dependent consideration of interactions. The solid line shows the correlation when all interactions are considered. The other graphs show the correlation considering only interactions with the nearest-neighbour (n+1, long dashes, blue), up to two residues (dashed, red) and up to three residues away (dotted, brown).

chain sequence (Figure 15). Taking only the nearest- and n+2 neighbours into account considerably worsens the results, affecting the Spearman rank correlation coefficient especially below 200 nm. However, considering only interactions with chromophores up to three residues away does not visibly affect the correlation graph above 205 nm and only slightly decreases the correlation at shorter wavelengths. Therefore, if the need arose to deal with larger proteins, the square Hamiltonian matrix could be reduced to a band diagonal matrix. The diagonalization routine could then be replaced and algorithms for band matrices can handle larger dimensions.

### 1.3.6 Amide Transitions in the VUV

The origin of a band around 165 nm has prompted some speculation. At first a transition from the higher-energy lone-pair orbital on oxygen to an antibonding $\sigma$ orbital ($n \rightarrow \sigma^*$) was suggested.[90] This was discounted by *ab initio* calculations that showed that the $\sigma^*$ orbital lies at a much higher energy.[91] The involvement of an excitation from the bonding $\pi_b$ orbital to $\pi^*$ ($\pi_b \rightarrow \pi^*$) was suggested,[92] but this transition occurs around 140 nm.[91] An excitation from the second lone pair on oxygen to the antibonding $\pi$ orbital ($n' \rightarrow \pi^*$) has also been considered.

Although charge-transfer transitions have been identified as the cause of the band at 165 nm,[93] the inclusion of higher energy transitions remained an area of investigation, since interaction with these could affect the intensities and positions of the bands in the far-UV. The $\pi_b \rightarrow \pi^*$ (bonding $\pi$ orbital to $\pi^*$) and $n' \rightarrow \pi^*$ (second lone pair on oxygen to $\pi^*$) transitions occur around 140 nm and are thus outside the measured region of 180 – 250 nm. Nevertheless, they could couple with the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions and have an effect on the observed far-UV.[79]

Nesgaard *et al.* have recorded SRCD spectra of 13 proteins, which were dried to allow measurement down to 130 nm.[94] The number of available spectra is too small to calculate significant statistics, but allows one to study a region of the VUV that is hard to measure even when using SRCD. For five of the 13 proteins, the band of the $\pi_b \rightarrow \pi^*$ transition around 140 nm is reproduced with a good match of the band position, but with an overestimated intensity. However, in the near-UV, the dry-phase spectra show a lower intensity than the wet-phase spectra and, therefore, this underestimation could be due to a

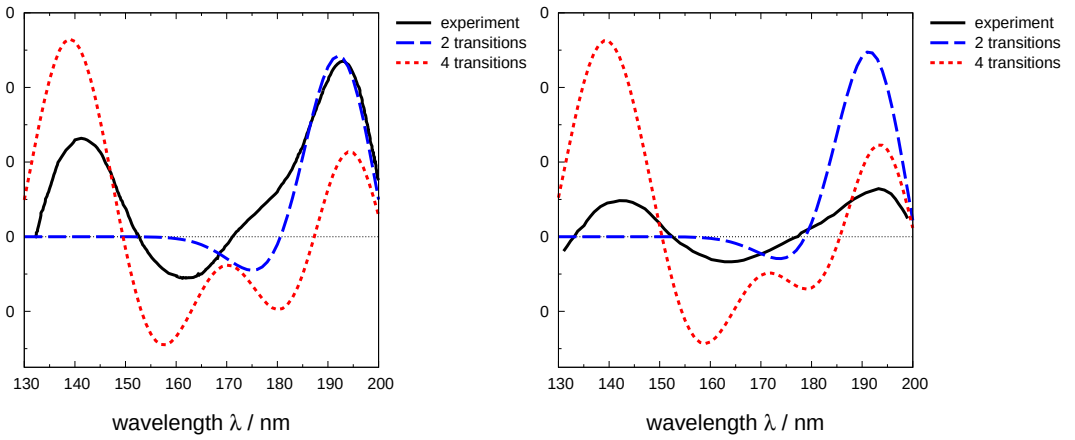



Figure 16: Spectra using two backbone transitions (dashed, blue) and four backbone transitions including charge-transfer (dotted, red), respectively. The experimental spectrum (solid, black) was recorded using SRCD on a dried sample. Left: Acetyl coenzyme A (1hb6). Right: Human serum albumin (1hk1). Both structures are mainly $\alpha$-helical.

lack of solvent effects. Figure 16 shows two examples of the CD130 set (see Appendix B for all spectra). The band around 160 nm originates from backbone charge-transfer transitions and is also apparent in the experimental spectra. However, the negative peak at 182 nm is not reproduced correctly. As observed for the SP175 set (Page 49), $\alpha$-helical proteins show a positive and $\beta$-sheet proteins a negative intensity in this region. The additional backbone transitions always decrease the intensity of this band and worsen the agreement. In the far-UV, the intensity is generally decreased and the bands are shifted to lower energy. In conclusion, the additional transitions show a promising agreement with experiment in the VUV, but worsen the results in the far-UV by decreasing the intensities of the bands.

### 1.3.7 Applications and Other Developments

In protein folding, the transient and short-lived nature of the intermediates makes the direct observation of folding and unfolding exceptionally challenging for techniques such as NMR and X-ray crystallography. However, nanosecond CD spectroscopy is a promising candidate for monitoring the conformational changes, which motivates the combination of these first-principles calculations with MD simulations to study the influence of dynamics on the CD of polypeptides.[69] MD simulations and CD calculations have been performed on concanavalin A, a $\beta$-I- and elastase, a $\beta$-II-type protein, which possess similar structures but show different CD spectra.[69] In the former case, the spectrum did not change substantially and remained close to that calculated for the X-ray crystal structure. For the $\beta$-II protein, the intensity of the calculated positive peak at 195 nm (which is negative in the experiment) decreased considerably, suggesting that relaxation of the protein structure in solution may be a contributing factor to the differences between $\beta$-I and $\beta$-II proteins. The well ordered, aligned sheets in $\beta$-I proteins and the helices in $\alpha$-helical proteins induce a relatively rigid structure supported by main-chain hydrogen bonds. This framework cannot be formed to the same extent in less ordered peptides, such as $\beta$-II-type proteins, which makes them more flexible.

The effect of hydrogen bonding on CD was highlighted by the unexpectedly high negative intensities at 222 nm reported for several peptides, which exceed even the highest estimate of $[\theta_{H\infty}]_{222}$.[62] Calculations suggested that the interaction with highly polar solvent molecules, which bind selectively to the helical conformation of peptides, may strengthen the backbone hydrogen bonds of the peptide. The intensity at 222 nm is sensitive to the main-chain hydrogen bond length[66] and the stabilization (and shortening) of hydrogen bonds could cause an over-estimation of the helicity in polar solvents.

## 1.4 Conclusion

CD spectra provide information on the secondary structure of proteins and the helical content can be determined from the intensity at 222 nm. Matrix method calculations are able to reproduce the features observed in experiment and can, therefore, be used to aid the structural analysis of proteins. The amide group has been parametrized using semiempirical and *ab initio* methods. Calculations from either parameter set show a Spearman rank correlation coefficient above 0.9 around 222 nm for a set of 46 proteins, with the *ab initio* parameters being superior near 210 and below 200 nm. Aromatic side chain chromophores have been parametrized to account for experimental features in the near-UV and two amide transitions below 160 nm extend the region covered by the calculations into the VUV.

Measurements in the VUV are not possible with conventional CD spectrometers and only a small number of experimental spectra containing VUV data have been available. A set of 71 spectra, recorded using synchrotron radiation, has facilitated a more detailed analysis of the calculations in the VUV and of the information such calculations can provide on the secondary structure. The study of the charge-transfer parameters will be the focus of Chapter 2.

# Chapter 2

# Charge-Transfer Transitions

The deep-UV is a region of the spectrum of growing interest for CD. The use of synchrotron radiation has made it possible to extend the wavelength range over which data can be measured into the vacuum-UV (VUV)[36] down to 160 nm to uncover further features of the protein CD spectra.[38,42,43,95,96] There is experimental and theoretical evidence indicating that charge-transfer transitions between neighbouring amide groups are the cause of the band at 165 nm.[97–99] Previously, charge-transfer chromophores have been parametrized for matrix method calculations by Oakley and Hirst[100] and tested on 31 proteins, for which VUV data are available. The SRCD spectra of a diverse set of 71 proteins facilitated a more detailed analysis of parameters with respect to the influence of secondary structure elements and specific charge-transfer transitions on the CD. 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 experimental intensity at 175 nm from 0.12 to 0.79.

## 2.1 Introduction

Goldmann *et al.*[101] 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 $NV_1$ band (arising from a local $\pi \rightarrow \pi^*$ transition) and that the Frenkel model in its original form, which is the basis of matrix method calculations of protein CD, therefore, does not adequately describe the electronic excitations in this band. To extend the model and incorporate the charge-transfer transitions directly into the Hamiltonian matrix, the *ab initio* characterization of peptide backbone[71] and aromatic side chain chromophores[81] has been extended to include the charge-transfer transitions between neighbouring peptide groups.[100,102]

CASSCF/CASPT2 calculations have been performed on the dipeptide *N*-actetylglycine-*N'*-methylamide[58,100,102] with dihedral angles from all accessible areas of the Ramachandran plot (Table 3). These geometries include a planar dimer ($\phi = 180°$, $\psi = 180°$) and several conformations in the main regions of the Ramachandran plot (corresponding to $\alpha$-helical and $\beta$-sheet structures). Mirror images of all structures were generated to describe, for example, left-handed helices and also cover less-favoured but occurring conformations. The charge-transfer between neighbouring groups had a significant impact on the calculated spectrum in a manner dependent on the conformation, that is, the dihedral angles of the two amide groups. For a matrix method calculation, this means that each conformation has to be parametrized. The Hamiltonian matrix (Equation 9, page 8) needs to be extended by the addition of a charge-transfer chromophore, which spans two neighbouring monomers in the chain.

Table 3: Wavelengths and oscillator strengths ($f$) of the charge-transfer chromophores. A mirror image of each geometry was created additionally to construct parameters, for example, for the left-handed helical and sheet regions of the Ramachandran plot.

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

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 consists of 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 of 0.1 Å to the atom centre. For the assignment of the suitable chromophore, the $\Phi$ and $\Psi$ angles of two neighbouring groups are calculated and the closest conformation (Table 3) is determined. Many PDB files contain chains with large gaps due to missing residues; therefore, the distance between the groups needs to be checked before assigning the charge-transfer groups.

The structure assignment of the geometries in Table 3 was done by modelling polypeptides with the respective dihedral angles and dividing them into strand- and helix-forming groups. The type *hybrid* refers to structures that look like wound strands or stretched helices and, therefore, cannot 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 $\beta$-type proteins than in helical ones.

There are four transitions for charge-transfer between adjacent groups, which have been considered in the parameter sets (Table 3): the transfer of an electron from a lone pair in one group to the $\pi^*$ orbital in the other group and transfer between a non-bonding $\pi$ orbital and the other group's $\pi^*$ orbital. The Hamiltonian for a tripeptide, showing only one charge-transfer transition per peptide group for clarity, has the form

$$\hat{H} = \begin{pmatrix} E_{Local_1} & V_{E_1;\,E_2} & V_{E_1;\,E_3} & V_{E_1;\,CT_{12}} & V_{E_1;\,CT_{23}} \\ V_{E_1;\,E_2} & E_{Local_2} & V_{E_2;\,E_3} & V_{E_2;\,CT_{12}} & V_{E_2;\,CT_{23}} \\ V_{E_1;\,E_3} & V_{E_2;\,E_3} & E_{Local_3} & V_{E_3;\,CT_{12}} & V_{E_3;\,CT_{23}} \\ V_{E_1;\,CT_{12}} & V_{E_2;\,CT_{12}} & V_{E_3;\,CT_{12}} & CT_{12} & 0 \\ V_{E_1;\,CT_{23}} & V_{E_2;\,CT_{23}} & V_{E_3;\,CT_{23}} & 0 & CT_{23} \end{pmatrix} . \quad (18)$$

The blocks involving solely local excitations are as defined in Equation 9 (page 8). The zeros in Equation 18 reflect the fact that charge-transfer chromophores sharing a common peptide group are not allowed to interact. In Equation 18, the first charge-transfer chromophore consists of peptide groups one and two, whereas the latter spans groups two and three.

Lees *et al.* have recorded, using synchrotron radiation CD, 71 spectra from 175 nm to 240 nm for the use as a reference dataset for CD analysis.[40] Atomic coordinates for these proteins are available via the Protein Data Bank (PDB).[103] 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.[40] To function as a reference set, the proteins cover a range of secondary structure characteristics. About 18% of the proteins are mainly $\alpha$-helical, 39% mainly $\beta$ and 35% $\alpha\beta$ mixtures. The rest contain multiple classes or only few secondary structures.[40] Since this set (referred to as SP175, Table 4) is considerably larger than the previous set, which provided VUV data[95,96,100] and has been specially selected for structural diversity, it is used to assess the accuracy of CD calculation and the relative importance of different charge-transfer transitions in the VUV.

Table 4: PDB codes and names of the 71 proteins comprising the SP175 set.

| PDB | Protein | PDB | Protein |
|---|---|---|---|
| 193l | Lysozyme | 1m8u | $\gamma$-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 | $\beta$-Lactoglobulin | 1r0i | Rubredoxin |
| 1bgl | $\beta$-Galactosidase | 1rav | Avidin |
| 1blf | Lactoferrin | 1rhs | Rhodanese |
| 1bn6 | Haloalkane dehydrogen | 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 | $\gamma$-D-Crystallin | 1vjs | $\alpha$-amylase |
| 1f4j | Catalase | 1ymb | Myoglobin horse |
| 1fa2 | $\beta$-Amylase | 2bb2 | $\beta$-B2-Crystallin |
| 1gpb | Glycogen phosphorylase B | 2cga | $\alpha$-Chymotrypsinogen |
| 1ha4 | $\gamma$-S-Crystallin (C-term) | 2cts | Citrate synthase |
| 1ha7 | c-Phycocyanin | 2dhq | Dehydroquinase type 2 |
| 1hc9 | $\alpha$-Bungarotoxin | 2fdn | Ferredoxin |
| 1hcb | Carbonic anhydrase I | 2psg | Pepsinogen |
| 1hda | Hemoglobin | 3dni | Deoxyribonuclease I |
| 1hk0 | $\gamma$-D-Crystallin (human) | 3est | Elastase |
| 1hnn | PNMT | 3pgk | Phosphoglycerate kinase |
| 1hrc | Cytochrome C | 3pmg | Phosphoglucomutase |
| 1hwx | Glutamate dehydrogenase | 3rn3 | Ribonuclease A |
| 1igt | Immunoglobulin | 4gcr | $\gamma$-B-Crystallin |
| 1k6j | NmrA | 5cha | $\alpha$-Chymotrypsin |
| 1kcw | Ceruloplasmin | 5cpa | Carboxypeptidase A |
| 1ku8 | Jacalin | 5pti | Aprotinin |
| 1les | Lentil lectin | 7tim | Triose phos. isomerase |
| 1lin | Calmodulin | | |

## 2.2 Correlation with Experiment

Matrix method calculations have been carried out on the 71 PDB files of SP175, considering different combinations of chromophores: all backbone-transitions only and in combination with included charge-transfer and side chain transitions. All calculated spectra compared to experiment are shown in Appendix C. The Spearman rank correlation coefficient[87] between the calculated and experimental intensity at each wavelength has been calculated. 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 17) the 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 parameters, the correlation coefficient at 175 nm is 0.79 (Table 5).

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, as seen previously for the set of 46 proteins (CD190, page 26). After the zero-crossing the correlation reaches values greater than 0.8 above 207 nm. The band at 208 nm, arising from the exciton splitting of the $\pi \rightarrow \pi^*$ transition, shows a correlation of 0.87 if charge-transfer and side chain transitions are included.

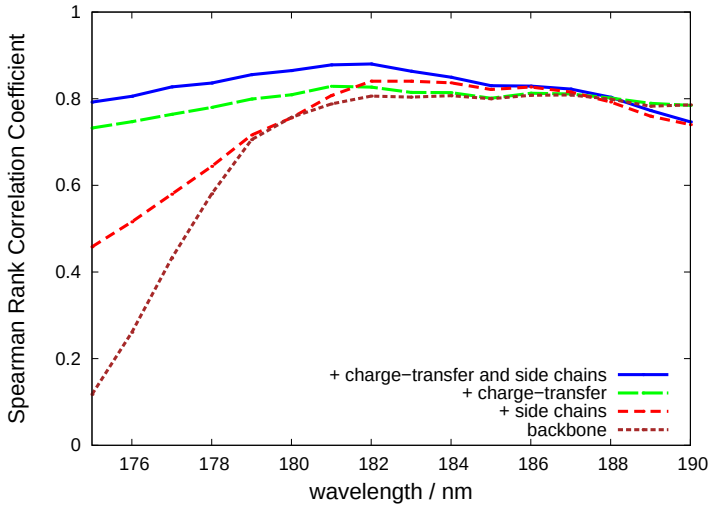



Figure 17: Spearman rank correlation coefficients between the experimental and calculated intensities of the SP175 CD spectra, calculated considering different chromophores.

Table 5: Spearman rank correlation coefficients for the SP175 set 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 |

Table 6: Secondary structure content of the SP175 set. $\alpha \mathrel{\widehat{=}} > 40\%$ helical. $\beta \mathrel{\widehat{=}} > 40\%$ strand. $\alpha\beta \mathrel{\widehat{=}} > 24.5\%$ of both $\alpha$ and $\beta$ content. Other proteins are considered unstructured.

| PDB | Structural type | helix | sheet | unordered |
|---|---|---|---|---|
| 1ado | $\alpha$ | 0.46 | 0.24 | 0.30 |
| 1ed9 | $\alpha\beta$ | 0.31 | 0.32 | 0.37 |
| 1vjs | $\alpha\beta$ | 0.29 | 0.33 | 0.39 |
| 1hc9 | $\beta$ | 0.05 | 0.52 | 0.42 |
| 5cha | $\beta$ | 0.12 | 0.44 | 0.44 |
| 2cga | $\beta$ | 0.14 | 0.47 | 0.40 |
| 5pti | | 0.21 | 0.31 | 0.48 |
| 1rav | $\beta$ | 0.07 | 0.54 | 0.38 |
| 1fa2 | | 0.38 | 0.24 | 0.38 |
| 1bgl | $\beta$ | 0.14 | 0.48 | 0.38 |
| 1b8e | $\beta$ | 0.14 | 0.49 | 0.36 |
| 2bb2 | $\beta$ | 0.07 | 0.46 | 0.47 |
| 4gcr | $\beta$ | 0.09 | 0.54 | 0.37 |
| 1elp | $\beta$ | 0.09 | 0.53 | 0.38 |
| 1m8u | $\beta$ | 0.06 | 0.53 | 0.41 |
| 1ha7 | $\alpha$ | 0.76 | 0.09 | 0.15 |
| 1lin | $\alpha$ | 0.57 | 0.14 | 0.29 |
| 1hcb | | 0.18 | 0.39 | 0.43 |
| 1ca2 | $\beta$ | 0.16 | 0.41 | 0.43 |
| 5cpa | $\alpha\beta$ | 0.38 | 0.28 | 0.34 |
| 1f4j | $\alpha\beta$ | 0.30 | 0.27 | 0.43 |
| 1kcw | $\beta$ | 0.12 | 0.49 | 0.39 |
| 2cts | $\alpha$ | 0.61 | 0.12 | 0.27 |
| 1nls | $\beta$ | 0.04 | 0.58 | 0.38 |
| 1hrc | $\alpha$ | 0.41 | 0.17 | 0.41 |
| 1qfe | $\alpha\beta$ | 0.46 | 0.28 | 0.27 |
| 2dhq | $\alpha\beta$ | 0.44 | 0.26 | 0.30 |
| 3dni | $\alpha\beta$ | 0.29 | 0.34 | 0.37 |
| 3est | $\beta$ | 0.11 | 0.51 | 0.38 |
| 2fdn | | 0.13 | 0.40 | 0.47 |
| 1ha4 | $\beta$ | 0.12 | 0.51 | 0.37 |
| 1cf3 | $\alpha\beta$ | 0.34 | 0.32 | 0.34 |
| 1hwx | $\alpha\beta$ | 0.49 | 0.25 | 0.26 |
| 1gpb | $\alpha\beta$ | 0.50 | 0.25 | 0.25 |
| 1bn6 | $\alpha\beta$ | 0.45 | 0.28 | 0.27 |

continued on next page

**Table 6 — continued from previous page**

| PDB | Structural type | helix | sheet | unordered |
|---|---|---|---|---|
| 1hda | $\alpha$ | 0.77 | 0.09 | 0.14 |
| 1n5u | $\alpha$ | 0.72 | 0.10 | 0.18 |
| 1hk0 | $\beta$ | 0.08 | 0.57 | 0.35 |
| 1igt | $\beta$ | 0.07 | 0.56 | 0.37 |
| 1trz | $\alpha$ | 0.58 | 0.09 | 0.33 |
| 1ku8 | $\beta$ | 0.00 | 0.68 | 0.32 |
| 1blf | $\alpha\beta$ | 0.32 | 0.36 | 0.32 |
| 1les | $\beta$ | 0.04 | 0.59 | 0.37 |
| 1ax8 | $\alpha$ | 0.69 | 0.04 | 0.28 |
| 193l | $\alpha\beta$ | 0.40 | 0.30 | 0.29 |
| 1mol | $\beta$ | 0.17 | 0.62 | 0.21 |
| 1ymb | $\alpha$ | 0.74 | 0.13 | 0.13 |
| 1a6m | $\alpha$ | 0.78 | 0.09 | 0.13 |
| 1k6j | $\alpha\beta$ | 0.37 | 0.28 | 0.35 |
| 1ova | $\alpha\beta$ | 0.31 | 0.45 | 0.24 |
| 1dot | $\alpha\beta$ | 0.29 | 0.32 | 0.39 |
| 1ppn | $\alpha\beta$ | 0.26 | 0.31 | 0.43 |
| 1ofs | $\beta$ | 0.04 | 0.58 | 0.38 |
| 1air | | 0.13 | 0.40 | 0.47 |
| 2psg | $\beta$ | 0.20 | 0.50 | 0.29 |
| 1atj | $\alpha$ | 0.51 | 0.19 | 0.31 |
| 3pmg | $\alpha\beta$ | 0.36 | 0.38 | 0.27 |
| 3pgk | | 0.34 | 0.15 | 0.50 |
| 1une | $\alpha$ | 0.50 | 0.23 | 0.28 |
| 1hnn | $\alpha\beta$ | 0.37 | 0.36 | 0.26 |
| 1a49 | $\alpha\beta$ | 0.39 | 0.30 | 0.31 |
| 1rhs | $\alpha\beta$ | 0.33 | 0.25 | 0.42 |
| 3rn3 | $\beta$ | 0.21 | 0.48 | 0.32 |
| 1r0i | $\beta$ | 0.17 | 0.42 | 0.41 |
| 1avu | $\beta$ | 0.02 | 0.48 | 0.50 |
| 1stp | $\beta$ | 0.08 | 0.60 | 0.31 |
| 1sca | $\alpha\beta$ | 0.30 | 0.34 | 0.37 |
| 1cbj | $\beta$ | 0.04 | 0.55 | 0.40 |
| 1thw | $\beta$ | 0.11 | 0.48 | 0.41 |
| 7tim | $\alpha\beta$ | 0.45 | 0.27 | 0.28 |
| 1ubi | $\alpha\beta$ | 0.25 | 0.46 | 0.29 |

## 2.3 Influence of Secondary Structure

The proteins have been divided into $\alpha$, $\beta$, $\alpha\beta$ and unstructured proteins, based on assignments by DSSP[104] (Table 6, Page 47). Proteins with greater than 40% $\alpha$-helical or more than 40% $\beta$ content were assigned $\alpha$ and $\beta$, 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 assignments by the CATH database.[105,106]

When charge-transfer was considered, in 69% of the cases, the difference at 175 nm between the calculated and experimental intensity was reduced for $\alpha$-helical proteins. On the other hand, only 39% of the $\beta$-type proteins showed an improvement and none of the $\alpha\beta$ proteins. 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 the $\alpha$-helical proteins show a positive intensity at 175 nm, whereas the spectra of $\beta$-sheet proteins are all negative. The $\alpha\beta$ cases 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 worsens the level of agreement.

While the intensity at 175 nm becomes more negative without exception for all $\beta$, $\alpha\beta$ and unstructured proteins, it increases in nine of the 13 $\alpha$-helical cases. A decrease was found for aldolase (1ado), cytochrome c (1hrc), insulin (1trz) and citrate synthase (2cts), whose $\alpha$-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 the proteins with higher $\alpha$-helix contents.

Of the other helical proteins, only calmodulin (1lin), peroxidase (1atj) and phospholipase-A2 (1une) possess an equally low helical content (below 0.6) and 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 $\beta$-II-type protein (Table 7).[85] The agreement between calculation and experiment is worse than for other proteins in these cases, especially between 190 and 210 nm. In $\beta$-II proteins,[85,107] the strands are rather short or not aligned in a parallel manner, but often twisted and bent. The resulting spectra are similar to unfolded, random coil proteins. It may be that conformational dynamics contribute to the unusual shape of the spectrum.[69]

A noticeable improvement was obtained for avidin (1rav), a $\beta$-sheet protein, which shows an intense negative band at 180 nm, with a mean residue molar ellipticity of $-21{,}000$ deg cm$^2$ dmol$^{-1}$. The backbone-only calculation gives a band intensity of $-12{,}000$ deg cm$^2$ dmol$^{-1}$, but the calculated intensity is $-18{,}000$ deg cm$^2$ dmol$^{-1}$, if charge-transfer is taken into account. Similar trends are shown in Figure 18 for an $\alpha$-helical protein (hemoglobin, 1hda) and a $\beta$-sheet protein (lentil lectin, 1les).

Table 7: $\beta$-II proteins in the SP175 set.

| PDB | Protein |
|---|---|
| 1avu | Serine protease inhibitor |
| 1r0i | Rubredoxin |
| 2cga | $\alpha$-Chymotrypsinogen |
| 2fdn | Ferredoxin |
| 3est | Elastase |
| 5cha | $\alpha$-Chymotrypsin |
| 5pti | Aprotinin |

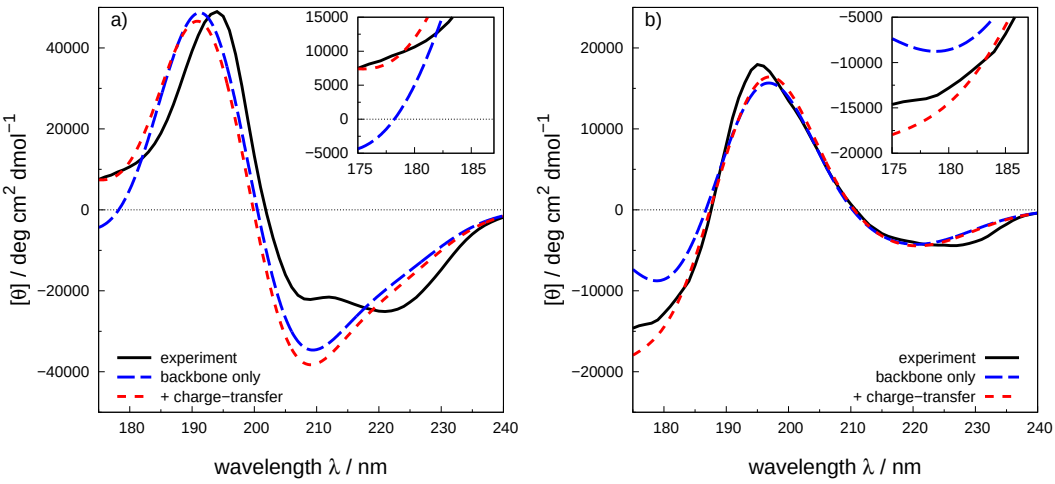



Figure 18: Experimental and calculated CD spectra of hemoglobin (a, 1hda) and lentil lectin (b, 1les), which show a considerable improvement with included charge-transfer transitions. Experimental (solid), backbone-only (dashed) and backbone plus charge-transfer (dotted).

## 2.4 Individual Contributions of Charge-Transfer 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 3, page 41). The sum of all the hybrid conformations makes up about 15% of the charge-transfer conformations in an $\alpha$-helical protein, about 30% for $\beta$-sheet proteins and approximately 20% for an $\alpha\beta$ protein. That is, the hybrid chromophores have more $\beta$ character, as suggested by the location of their dihedral angles in the Ramachandran plot.

Table 8: Most frequent conformations of charge-transfer chromophores.

| **Geometry** | $\phi$/° | $\psi$/° |
|---|---|---|
| $\alpha$-helical | −62 | −41 |
| $3_{10}$-helical | −74 | −4 |
| $\beta$-strand | −120 | 120 |
| | −135 | 135 |
| | 180 | 180 |

Although the amount of helical conformations was usually well above 70% for $\alpha$ proteins, the amount of strand-like conformations in the $\beta$ set was about 45%. In the $\alpha\beta$ group, helical conformations account for approximately 55% of all charge-transfer groups. 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 8. Among the helical chromophores, the ($\phi = -62°$, $\psi = -41°$) group is the most important conformation, representing a typical $\alpha$-helix.[108] It is also the most frequent helical type in $\alpha\beta$ and unstructured proteins. In $\beta$ proteins, the $3_{10}$-helical conformation ($\phi = -74°$, $\psi = -4°$) occurs more frequently than $\alpha$-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 (Figure 19, Page 54). 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 and strand 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 the others 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, due to their combined contribution. Notably, the full charge-transfer spectra are not a linear combination of contributions from single chromophores, due to the interaction terms in the Hamiltonian matrix.

The inclusion of charge-transfer chromophores usually causes the intensity in the VUV to be more negative, except in highly $\alpha$-helical proteins. The average decrease of the band at 175 nm is between 5000 and 15,000 deg cm$^2$ dmol$^{-1}$ with the absolute minimum being predicted around 160 nm. 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 20). 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 $\alpha$-helical conformation, ($\phi = -62^\circ$, $\psi = -41^\circ$), even in those $\beta$ 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 $\beta$ 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 $\beta$ 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 21).

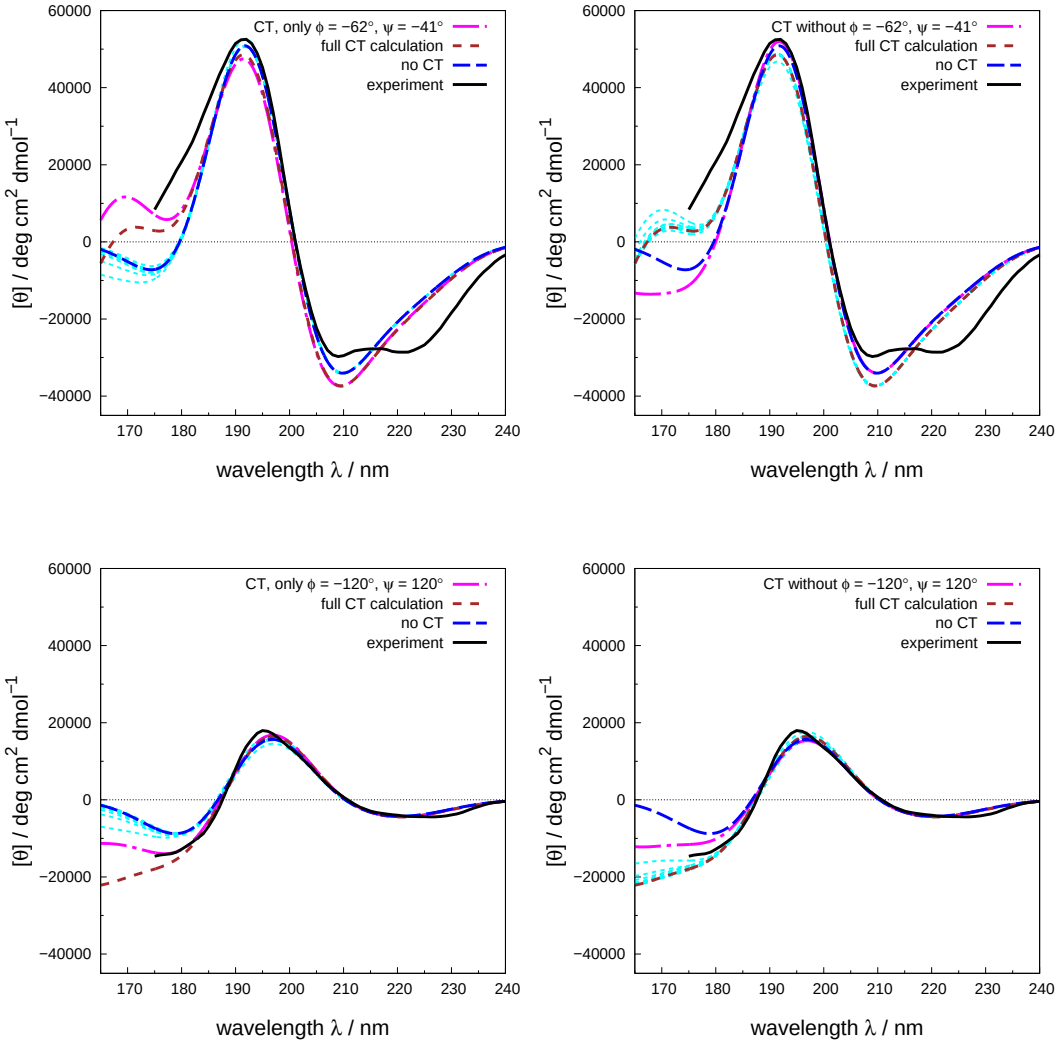



Figure 19: Top: Spectra of serum albumin (1n5u, mainly $\alpha$). Bottom: Spectra of lentil lectin (1les, mainly $\beta$).
All CT calculations consider four CT transitions. Experimental (solid, black), backbone-only (long dashes, blue), full CT calculation (all chromophores, shorter dashes, brown). The short-dashed, turquoise curves in the left panels show calculations using only one CT conformation. The short-dashed, turquoise curves in the right panels show calculations leaving out one CT conformation and using all others.
Left: The pink, dot-dashed curves show the calculation only with the most frequent set for the respective secondary structure type. Right: The pink, dot-dashed curves show calculations without the most frequent set for the respective secondary structure type, but with all other sets.

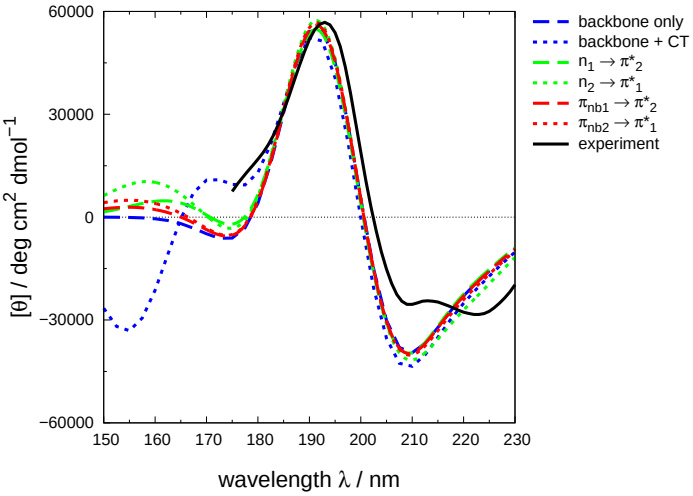



Figure 20: Spectra of whale myoglobin (1a6m) showing the effect of single charge-transfer transitions in a typical $\alpha$-helical protein.

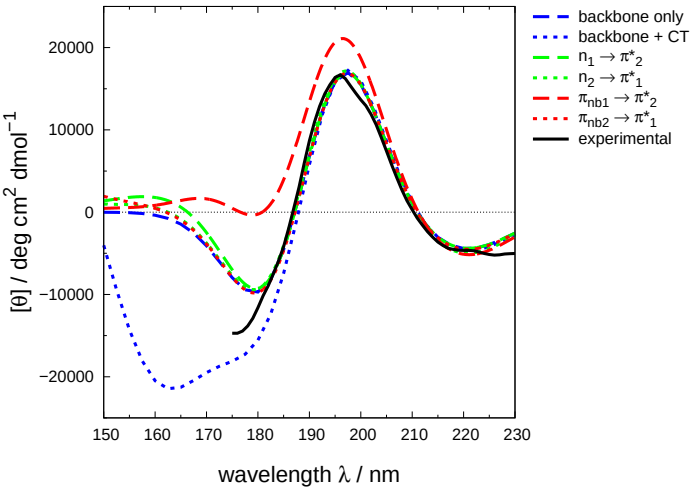



Figure 21: Spectra of pea lectin (1ofs) showing the effect of single charge-transfer transitions in a typical $\beta$-sheet protein. Most significant is the $\pi_1 \rightarrow \pi_2^*$ transition, which leads to a considerable diminishment of the band at 180 nm in $\beta$ and unstructured proteins as well as to an increased intensity at 190 nm.

## 2.5 Conclusion

The SP175 protein set made it possible to assess the influence of the charge-transfer transitions in the VUV, 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 experimental intensity if only local excitations are considered. However, $r$ rises to 0.73, if charge-transfer transitions are taken into account and 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 two of 19 charge-transfer conformations, namely, the main $\alpha$-helical and $\beta$-strand types. However, the combination of all other chromophores contributes just as much and, therefore, cannot be neglected. The change of intensity in the VUV could not be attributed to a specific charge-transfer transition, but is rather caused by interactions among all four types. These results show that the inclusion of charge-transfer transitions enhances the theoretical CD calculations in the VUV region.

The previous chapters have introduced the theory behind the calculation of protein CD and shown the performance of the available parameters using two different sets of proteins. The next chapter will focus on taking dichroism calculations one step further to cover a related spectroscopy method, linear dichroism.

# Chapter 3

# Linear Dichroism

Linear dichroism (LD) spectroscopy, a differential absorption technique closely related to circular dichroism (CD) spectroscopy, is attracting renewed attention.[109] LD experiments require the samples to be aligned in some way and with polymers the alignment is typically provided by two methods, electric field[110,111] and flow. Studies of the dichroism of polymers[110,112,113] and proteins[114,115] oriented in a flow have an extensive history; early investigations focused on the use of double refraction of flow to characterize myosin.[116,117] The focus in this work is on flow LD (henceforth referred to as LD). LD spectra have provided information on how molecules such as gramicidin in membranes, $\alpha_1$-antitrypsin and the amyloid-related protein A$\beta_{1-42}$ align in shear flow.[118] LD is being used to detect and characterize fibril formation by amyloid peptides[119] and can furnish insight into how agents dock to enzymes or fibres.[120–122] The technique has been used to study the polymerization of protein fibres such as G-actin into F-actin,[118] the orientation of bacteriorhodopsin in liposomes,[123] the binding of myosin to actin,[124] and to measure the kinetics of insertion of peptides into lipid membranes.[125,126]

To exploit the information available from LD measurements fully, one needs to understand the relationship between protein conformation and LD spectra and, ideally, one would like to be able to construct structural models and use these to calculate the electronic structure directly.[118] In this chapter, the utility of LD calculation from first principles is shown on several proteins exemplifying prototypical secondary structural classes. This has been made possible by the adaptation of the matrix method, which is used for calculating CD spectra, to the calculation of LD spectra.

## 3.1 Introduction

LD measurements require the difference in absorption of light polarized parallel ($A_{||}$) and perpendicular ($A_{\perp}$) with respect to an orientation direction:

$$\mathrm{LD} = A_{||} - A_{\perp} \; . \tag{19}$$

In order to understand LD, it is necessary to consider what happens when a molecule absorbs a photon: absorption can be pictorially viewed as either the electric field or the magnetic field (or both) of the radiation pushing the initial electron density to a final arrangement of higher energy. The direction of the net linear displacement of charge is known as the polarization of the transition. The polarization and intensity of a transition are dependent on the electric transition dipole moment, $\vec{\mu}$, which may be regarded as the antenna by which the molecule absorbs light. Each transition thus has its own antenna and the maximum probability of absorbing light is obtained when the antenna and the electric field of the light are parallel. Figure 22 shows a linearly polarized light beam and a sample of molecules all oriented in exactly the same way.

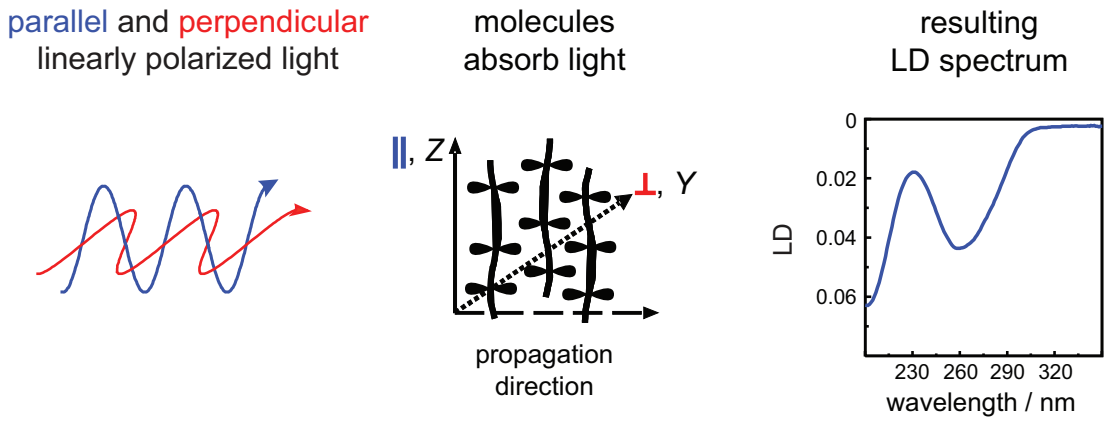



Figure 22: Schematic illustration of an LD experiment.

The normal absorption is measured with the light polarized so that it is parallel to the direction of orientation of the sample and a second absorbance spectrum is measured when the light is polarized perpendicular to this direction. The difference of these two spectra is the LD spectrum. The two extreme situations, when measuring the parallel and perpendicular absorbance, are:

1. If the polarization of the probed transition is perfectly *parallel* to the orientation direction, then

$$\mathrm{LD} = A_{\parallel} - A_{\perp} = A_{\parallel} > 0\,. \tag{20}$$

2. If the polarization of the probed transition is *perpendicular* to the orientation direction, then

$$\mathrm{LD} = A_{\parallel} - A_{\perp} = A_{\parallel} < 0\,. \tag{21}$$

For intermediate polarizations, the LD is between these cases and for a perfectly oriented system:[34]

$$\mathrm{LD}^r = \frac{\mathrm{LD}}{A} = \frac{A_{\parallel} - A_{\perp}}{A} = \frac{3}{2}(3\cos^2\alpha - 1)\,, \tag{22}$$

where the reduced LD, $LD^r$, is defined to be the ratio of the LD over the isotropic absorbance, $A$, and $\alpha$ is the angle between the transition moment and the orientation axis. $LD^r$ is independent of sample pathlength and concentration. Thus, it is possible to determine the polarization of a given transition from its LD spectrum if the orientation of the molecule is known. Conversely, one can use LD as a probe of molecular orientation if the polarization of a transition moment within the molecule is known.

Equation 19 indicates that transitions polarized parallel to the fibre axis will lead to a positive band in the LD spectrum, whereas transitions with a perpendicular polarization will lead to a negative band. One can, therefore, anticipate the qualitative features of the LD spectra of regular secondary structure elements by considering the allowed combinations of coupled local $\pi \rightarrow \pi^*$ transitions, which are polarized along the carbonyl bonds, to determine the net polarization for these transitions. For an $\alpha$-helix it is known from coupled oscillator analysis that the $\sim$195 nm amide $\pi \rightarrow \pi^*$ transitions are split into two components, one polarized parallel to the helix axis at $\sim$210 nm and another polarized perpendicular at $\sim$190 nm (Figure 23).

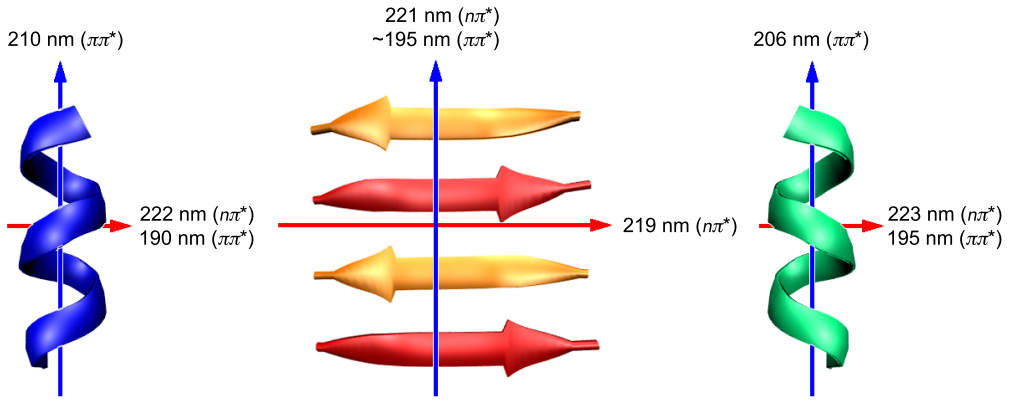



Figure 23: Polarizations of transitions in an $\alpha$-helix, a $\beta$-sheet, and a $P_{II}$-helix.

An analogous consideration of $n \rightarrow \pi^*$ transitions has been developed through calculations on peptides with structures built into ideal, regular helical and strand conformations. The $n \rightarrow \pi^*$ transition at ~222 nm is polarized perpendicular to the helix axis.[118] For a perfectly parallel helix (such as in a coiled coil, e.g., tropomyosin) one would, therefore, expect negative bands around 190 nm and ~222 nm in the LD spectrum. These regions are dominated by transitions polarized perpendicular to the helix axis, resulting in a negative LD signal according to Equation 19. Conversely, the parallel component of the $\pi \rightarrow \pi^*$ transition would result in a positive band at ~210 nm (though in reality it is usually dominated by the two bands on either side of it so appears as a small minimum). In $\beta$-sheets the $\pi \rightarrow \pi^*$ transition is polarized parallel to the long axis of a single strand (perpendicular to the sheet), due to the lack of the parallel component at ~210 nm.[51] In actuality, about 50% of residues in proteins are not in regular secondary structure motifs and proteins often comprise a complex combination of secondary structures. Therefore, a more detailed computational approach was adopted, which deals with the experimentally diverse range of protein conformations.

The electronic excitations of the chromophores are described by parameters derived from *ab initio* calculations,[71,81] which have been tested on several sets of proteins.[79,127] There are chromophore parameters available for the peptide[71] and aromatic side chain transitions.[81] The calculations consider two local, peptide backbone transitions, the $n \rightarrow \pi^*$ transition (at 222 nm) and the $\pi \rightarrow \pi^*$ transition (at 193 nm), and electronic transitions associated with the side chains of tryptophan, phenylalanine and tyrosine. The transitions and their associated properties are summarized in Table 9. Inclusion of higher energy local peptide transitions and charge-transfer transitions made little quantitative difference and the calculations reported here do not include them.

Table 9: Electronic transitions included in the calculations of the LD spectra, and their corresponding transitions energies and electric transition dipole moments.

| Transition | Energy / nm | Electric transition dipole moments[a] ($x$, $y$, $z$) / Debye | | |
|---|---|---|---|---|
| | | **Backbone** | | |
| $n \rightarrow \pi^*$ | 220 | ( 0.0000, | 0.0000, | 0.2260 ) |
| $\pi \rightarrow \pi^*$ | 193 | ( 2.0946, | 2.0200, | 0.0000 ) |
| | | **Tryptophan** | | |
| $^1B_a$ | 196 | ( 3.6291, | 2.1261, | 0.0000 ) |
| $^1B_b$ | 204 | ( 3.6321, | −2.4661, | 0.0000 ) |
| $^1L_a$ | 263 | ( −1.0939, | 1.2255, | 0.0000 ) |
| $^1L_b$ | 283 | ( 0.7198, | 0.6912, | 0.0000 ) |
| | | **Phenylalanine** | | |
| $^1B_b$ | 190 | ( −0.0009, | 5.7320, | 0.0000 ) |
| $^1B_a$ | 191 | ( −5.7153, | −0.0025, | 0.0000 ) |
| $^1L_a$ | 209 | ( −0.0018, | 0.0000, | 0.0000 ) |
| $^1L_b$ | 263 | ( 0.0000, | 0.0171, | 0.0000 ) |
| | | **Tyrosine** | | |
| $^1B_b$ | 193 | ( −0.0220, | −5.3279, | 0.0000 ) |
| $^1B_a$ | 196 | ( 5.7956, | 0.0015, | 0.0000 ) |
| $^1L_a$ | 216 | ( 1.3859, | −0.1582, | 0.0000 ) |
| $^1L_b$ | 274 | ( −0.0563, | −0.6972, | 0.0000 ) |

[a] Defined with respect to coordinate systems provided in the Supplementary Information of earlier publications.[71,93]

The $n \rightarrow \pi^*$ band is evident in several of the experimental absorption and LD spectra. The calculations do not reproduce this band, despite the reasonable agreement observed between experimental and calculated CD spectra. The discrepancy in the case of the LD spectra indicates that the calculated transitions at ∼222 nm do not have a sufficiently large electric transition dipole moment, which in turn suggests that the mixing between the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ states is significantly under-estimated for the conformations in question.

However, investigative calculations, using other parameter sets, where the coupling between these states is stronger,[72] did not improve matters (data not shown).

While the use of LD to gain information on protein orientation has increased over the past years, the related technique of CD spectroscopy is more common and widely used for secondary structure determination of proteins.[9] In a CD spectrometer a sample is irradiated with left and right circularly polarized light and, due to the chirality of the structural motifs, each type of polarized light is absorbed to a different extent. The differential absorption is wavelength dependent and makes it possible to distinguish between different protein folds and to estimate the helical content.[52] Due to this and other advantages, CD spectrometers are widely available and much effort has been put into understanding the theoretical basis of the technique.[128]

Since circularly polarized light consists of two linearly polarized beams travelling with optical retardation, any CD spectrometer can be used to measure LD. Linearly polarized light may be produced by adding a quarter-wave plate or increasing the voltage across the photoelastic modulator.[34] Despite the close relationship of the techniques, the measurement of LD is more challenging than for CD. The latter can be measured for a solution in a cuvette, but in order to record an LD spectrum the sample has to be oriented.[34]

## 3.2 Molecular Alignment Techniques

Proteins may be oriented by a number of methods depending on their nature. Long polymers, such as fibrous proteins, may be oriented by the viscous drag caused when a solution is flowed between the annular gap of a cylindrical cuvette with a rotating quartz rod in the middle, a Couette flow cell (Figure 24).[124,129,130] Depending on the cell design, the light is then propagated either along the flow direction or perpendicular to it. This technique is commonly used for LD studies of DNA. In the 1960s and 70s it was used for a few fibrous protein studies.[131,132] More recently, Rodger and co-workers have been developing it for a wide range of fibrous proteins including actin, tubulin, FtsZ, collagen and amyloid fibres.[109,118,133,134] The advantage of the Couette cell over a flow-through system is that the same volume of sample can be measured indefinitely. Improved cell designs have reduced the amount of sample required to as little as 30 $\mu$L, making the technique valuable even on a biochemical volume scale.[135]

Couette flow orientation of proteins has been extended to membrane proteins and peptides following Ardhammar, Mikati and Nordén discovering[136] that spherical liposomes were distorted in a Couette cell and could be used to orient small molecules bound to the liposomes. It has been shown that any protein or peptide that is bound to a liposome could be similarly oriented.[51,137] Due to the defined symmetry of the liposome, for such a system Equation 22 is replaced by

$$\mathrm{LD}^r = \frac{3S}{4}(1 - 3\cos^2\beta)\,, \tag{23}$$

where $\beta$ is the angle that the transition moment of interest makes with the normal to the liposome surface (that is, parallel to the lipid long axis).

$S$ is the orientation factor that denotes the fraction of the liposome that is oriented as a cylinder perfectly parallel to the flow orientation. It may be viewed as a scaling factor defining the efficiency of macroscopic orientation. $S = 1$ for perfect orientation (assumed in Equation 20) and $S = 0$ for random (i.e., no) orientation. If the sample is macroscopically as well as molecularly uniaxial, such as molecules oriented in a polymer film drawn in one direction or polar molecules in an electric field, there is a simple relation between $A_{\parallel}$, $A_{\perp}$, and $A$ which makes it unnecessary to measure all three quantities (Figure 25).

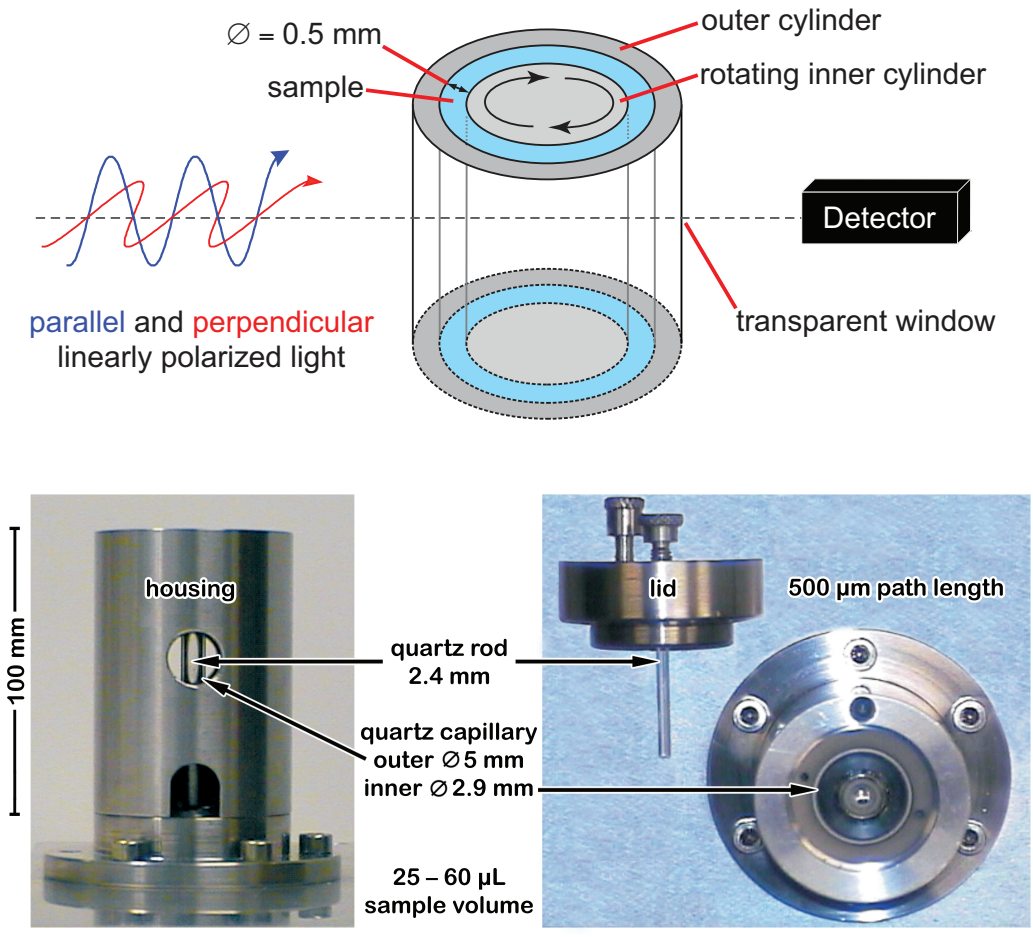



Figure 24: Top: Schematic illustration of a Couette flow cell. Bottom: Micro-volume capillary LD cell indicating how the components are assembled.

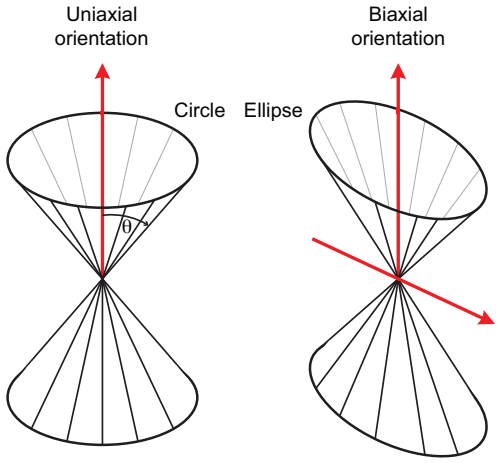



Figure 25: Uniaxial and biaxial orientation. Uniaxial orientation requires all orientations on a cone about the orientation axis to be statistically equally probable. This may be due to static or dynamic factors.[34]

Globular proteins are usually soluble in aqueous solution and can be crystallized. This facilitates the use of X-ray crystallography to aid the structure and orientation determination. However, fibrous proteins, such as FtsZ[122] and tubulin,[124,134] often do not crystallize. Alignment in Couette flow relies on high aspect ratios of the molecules, which is precisely what DNA and fibrous proteins have. LD is, thus, a valuable tool to obtain information about orientation of such molecules, especially if structural information is available about the monomer units (the DNA bases or the units of the protein fibre). For most protein fibres, fibre formation allows an LD measurement, even though the monomers are too small to be aligned individually, because of the regular order of the monomers in a fibre. Thus, LD can be used to follow fibre formation, as only the fibres give a signal. LD has also been used in the past in the context of electric dichroism to probe the structure and electrical properties of polymers.[110,113,138]

The interpretation of LD spectra is challenging, since overlapping transitions lead to the cancellation of bands and complicate the analysis of experimental spectra beyond what is possible by qualitative inspection. This is especially true if the protein contains a mixture of different secondary structural elements. Calculations allow one to study both single and multiple transitions as a function of geometry and orientation and the combination of experiment and calculation can lead to mechanistic and structural insights, which are not possible to derive from either in isolation. Previously,[51] the transition polarizations and oscillator strengths of ideal (i.e. perfectly regular) $\alpha$-helical and $\beta$-sheet model peptides have been calculated from first principles. In the following, the utility of the approach is shown on several proteins exemplifying prototypical secondary structural classes. For each protein of interest one considers the LD calculated for different proposed structures and uses the experiment/calculation comparison to select the experimental fibre geometry. For the protein fibres used in this work, atomic-level structures are available for the monomeric units and in some cases for dimers or higher oligomers, which facilitates the process. If no information is available for how the fibre units combine to make a polymer then a large series of calculations is required.

## 3.3 Examples of Protein LD Data

Figure 26 shows the LD spectrum one might expect to measure in the absence of light scattering artefacts for an $\alpha$-helix oriented parallel to the sample orientation direction. This might, for example, be on the surface of a liposome or a coiled coil $\alpha$-helical fibre. The 220 nm region is dominated by transitions polarized perpendicular to the helix and therefore shows a negative signal; the 208 nm region is polarized parallel to the helix and so it has a positive signal. However, it is sandwiched between two negative signals, so it may or may not actually end up looking like a positive maximum as the signals overlay and cancel each other out. This is an illustration of why it is important to have electronic level calculations of LD spectra, as discussed below, in order to deconvolute spectra in structural terms.

Figure 27, left, shows a near- and far-UV LD spectrum of F-actin over a range of concentrations.[124] In the 12 $\mu$M spectrum one can see the components of the $\pi \rightarrow \pi^*$ transition and the $n \rightarrow \pi^*$ transition of the peptide backbone, comparable to the schematic of Figure 26, but of opposite sign. The lower energy (higher wavelength) component of the $\pi \rightarrow \pi^*$ band is evident as the 215 nm dip in the spectrum. The LD spectrum therefore indicates that, on average, the $\alpha$-helices in the actin fibre are oriented more perpendicular than parallel to the fibre axis. This is in accord with literature data for actin. However, if no literature data were available, a less qualitative description of the structure would be preferred, which is simply not possible without considering all the transitions in the protein as discussed below.

It should be noted that care must be taken in collecting low wavelength data for a system such as actin where light scattering is significant (Figure 27, left).

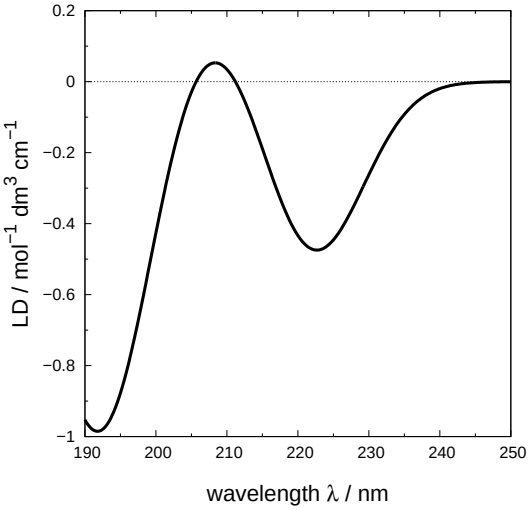



Figure 26: A schematic illustration of the LD spectrum one might expect for an $\alpha$-helix oriented parallel to the sample orientation direction. $\alpha$-helices have an $n \rightarrow \pi^*$ transition polarized perpendicular to the helix axis at 220 nm;[51] a component of the $\pi \rightarrow \pi^*$ transition parallel to the helix at ∼210 nm and a component of the $\pi \rightarrow \pi^*$ transition perpendicular to the helix at ∼190 nm.[51,109]

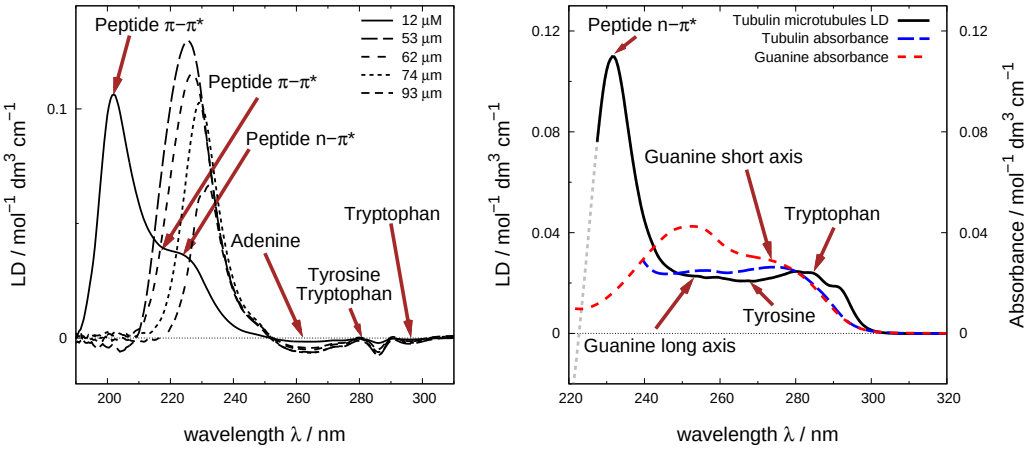



Figure 27: Left: Near- and far-UV LD spectra of F-actin in a Couette flow cell showing the apparent shift to shorter wavelength of the maximum signal as the concentration of F-actin is reduced. The 12 $\mu$M spectrum (solid line) is 'true'.[124] Right: Near-UV absorbance and LD data for tubulin (microtubules include tubulin and guanine).[134]

In that case, more concentrated samples have an apparent wavelength shift of the maximum LD signal of ∼30 nm to longer wavelengths. When the dilution effect is simply to reduce the signal intensity according to the Beer-Lambert law, then it can be concluded that the spectrum is real and one might expect a calculation to be able to give the same result if the geometry is correct. Matrix method calculations (see below) will never reproduce artefacts in spectra.

Figure 27, right, shows the near-UV LD and absorbance spectra of tubulin.[134] Due to the different wavelengths of guanine, tryptophan and tyrosine transitions, comparisons of calculated spectra with experimental data should enable one to determine, for example, the orientation of the guanine (which is essential for polymerization) at all stages of tubulin polymerization/depolymerization. Thus, calculations on tubulin (and indeed the near-UV region of actin) will require accurate parametrization of the NTP/NDP involved in the polymerization as well as the protein chromophores. This is true for any additional chromophore in any protein, both in CD (as discussed above) and LD.

In the case of bacteriorhodopsin (BR), the retinal chromophore is a very useful probe of the extent of protein orientation. The absorption, CD and LD spectra of BR in liposomes are given in Figure 28.[123] The absorption spectra (Figure 28, top) are consistent with those reported in the literature for the initial state of BR.[139] The absorption maximum at ∼570 nm is due to a long axis polarized transition of the retinal chromophore. The broad peak in the near-UV region (260 – 290 nm) is due to the transitions of the protein aromatic side chains phenylalanine, tyrosine and tryptophan, and the peak observed in the far-UV region (215 – ∼230 nm) is due to the peptide $n \rightarrow \pi^*$ transition of the amide groups. The backbone CD spectrum of BR (Figure 28, bottom) is

in accord with the CD spectrum for a highly $\alpha$-helical protein, although the intensity is weaker than for globular proteins, as discussed above.[140,141] The $\sim$570 nm LD signal is consistent with a retinal chromophore oriented at $\sim$70° to the lipids of a lipid bilayer if $S \sim 0.05$. The aromatic region (260 – 280 nm) of the LD spectrum is dominated by the indole chromophore of the tryptophan residues.[142] Contributions from $L_a$ (270 nm) and $L_b$ (287 nm) transitions are apparent, both showing positive LD and suggesting the average tryptophan is tilted so that the normal to the plane of the indole is $\sim$40° from the average lipid and the long and short axes are both at an angle of 60 – 65° to the lipids. This is consistent with the fact that the retinal is sandwiched by tryptophan residues.[140]

The protein backbone LD spectrum shows a maximum at 220 nm ($n \rightarrow \pi^*$) and a minimum at $\sim$213 nm ($\pi \rightarrow \pi^*$). It follows that the $n \rightarrow \pi^*$ transition (which is polarized perpendicular to the $\alpha$-helix long axis) is at $\sim$58° from the average lipid direction. The success of this analysis is dependent on having a value for $S$ (in this case from the retinal chromophore) and the protein being almost totally helical. With more varied protein structures, calculations to determine the net protein LD will be essential.

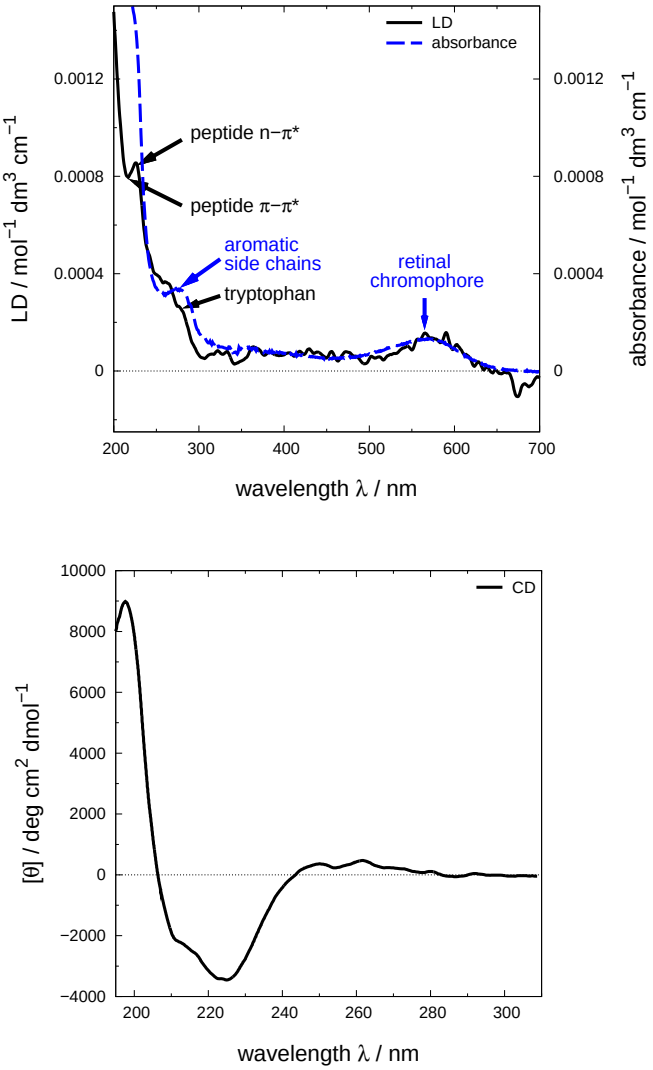



Figure 28: Spectra of bacteriorhodopsin (0.2 mg/mL) added to a soya bean liposome solution (0.5 mg/mL).[123] Top: Absorption (dashed line, baseline: liposome absorption spectrum) and LD (solid line, baseline: LD spectrum of sample without rotation spectrum). Bottom: CD (baseline: liposome CD spectrum).

## 3.4 Protein LD Calculations

It is apparent from the above discussion that LD data are consistent with proteins of known structure, and a qualitative level of structural interpretation is possible. This can lead to mechanistic insights not possible from other techniques, as in the case of FtsZ.[133] The level of interpretation that is possible, however, is restricted due to the cancellation of LD signals by overlapping bands, as illustrated in Figure 26 (Page 69). To get beyond this, it is crucial to know what each transition is doing and to be able to calculate spectra as a function of geometry until a good match between experiment and calculation is obtained.

A successful methodology for computing the CD spectrum of a protein from its structure is the (exciton) matrix method.[33,46,47] The atomic Cartesian coordinates required for the calculation can be downloaded, for example, from the RCSB protein data bank (PDB).[103] The matrix method uses monopole charges, which can be obtained from quantum mechanical *ab initio* calculations, to represent the electrostatic potential of the electronic transitions of the chromophores within the protein.[71] The Hamiltonian matrix needed to calculate the wave function is constructed from the Coulombic interactions between those charges.

When calculating a CD spectrum, values for the components of the electric transition dipole moments are generated in the process. In the case of a single, perfectly oriented molecule and using the notation of Chapter 1.2, for Equation 19 can be assumed:

$$\mathrm{LD}_{molecule} = (A_{\parallel} - A_{\perp}) \propto \left\{ |\langle \psi^k | \mu_Z | \psi^0 \rangle|^2 - |\langle \psi^k | \mu_Y | \psi^0 \rangle|^2 \right\} , \qquad (24)$$

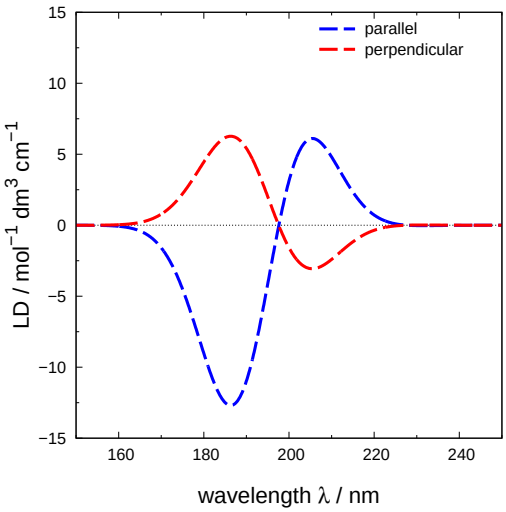



Figure 29: LD spectra of an $\alpha$-helix aligned to two different axes ($z$ being defined as orientation axis).

where $Z$ is the orientation axis and $Y$ is perpendicular to $Z$. In the case of a perfectly oriented molecule, $Z$ is usually the longest axis of the molecule (denoted $z$), and $Y$ is a rotational average over the molecular axes $x$ and $y$. So the LD can be calculated using the same matrix methodology as used for CD, with an expectation of a more accurate match to experiment as LD depends only on the electric moments. Figure 29 shows example calculations for two perpendicularly oriented helices.

To calculate the LD, the absorbance in parallel and perpendicular direction is calculated from the electric transition dipole moments. The $z$-axis is defined as the orientation axis (parallel); an average over the absorption, $A$, in the $x$ and $y$ direction accounts for the perpendicular absorption:

$$A_{||} = A_z \qquad A_{\perp} = \frac{1}{2}(A_x + A_y). \tag{25}$$

LD is the difference of these absorbances (Equation 24).

The electronic excitations of the chromophores are described by parameters derived from *ab initio* calculations,[71,81] which have been tested on several sets of proteins.[79,127] There are chromophore parameters available for the peptide[71] and aromatic side chain transitions.[81] The calculated far-UV spectra consider only the peptide chromophores.

## 3.5 Comparing Theory and Experiment

The intensity observed in experiment depends on the orientation factor, $S$, of the sample. If the sample is not oriented at all the orientation factor is zero and no LD spectrum is observed. If all proteins were perfectly oriented (as is the case, for example, in a crystal), $S$ would be unity and the intensity greatest. Adachi *et al.*[119] suggested that $S \sim 0.1$ for the amyloid fibrils that they studied. However, the orientation factor is difficult to estimate a priori, and there is no solid basis for doing so in the current study. The degree of orientation is related to the velocity gradient between the walls of the rotating Couette cylinders. The angular velocity has to be chosen carefully to maintain a laminar flow (or Couette flow). Turbulence will occur at too high a speed, leading to increased scattering and loss of light.[135] Since the velocity is, therefore, limited and the alignment of the sample depends on it, a perfect orientation of the molecules cannot be achieved, thereby diminishing the LD signal. In addition, samples subjected to shear flow are more prone to form bubbles than stationary samples in a cuvette, increasing the experimental challenge. In contrast, the calculations are carried out on a single defined orientation of a protein, with $S = 1$, by definition. In the absence of a rigorously defined scaling factor for the calculations, the experimental LD spectra were scaled to the calculated intensity for the purpose of comparison.

For each protein the calculated LD spectrum for the monomers assembled into fibres in accord with literature data and aligned perfectly parallel to the flow direction are given. For comparison, the LD of a fibre assembled to be equivalent to the same fibre oriented perpendicular to the flow direction are shown. The $z$-axis (vertical, blue) represents the parallel orientation with respect to the fibre long axis in the Couette flow; $x$ (red) and $y$ equivalently represent a perpendicular orientation. The colours of the structures indicate parallel (blue, aligned to $z$) and perpendicular (red, aligned to $x$) orientation and are used accordingly for calculated spectra throughout the study. Since the $x$ and $y$ coordinates are taken into account as a rotational average (Equation 25), alignment to either axis is equivalent. All 3D plots have been created using VMD.[143] Next to each plot showing LD spectra for the parallel and perpendicular orientation of the protein, the calculated absorbance spectra that were used to generate the LD spectra are shown. In the top half of the absorbance plot, the spectra for the parallel orientation are given, and the results for perpendicular alignment are shown in the bottom half. For each orientation, two calculated absorbance spectra are required: the absorption in the parallel direction ($A_{||}$, along the $z$-axis), and the absorption in the perpendicular direction ($A_{\perp}$, the rotational average according to Equation 25). Additionally, the full absorbance spectrum for each orientation is given for comparison with experiment.

## Experimental Methods

Couette flow experiments were performed by collaborators using a Jasco J815 spectropolarimeter modified for LD measurements (Jasco UK, Great Dunmow, UK). Spectra were measured over the appropriate wavelength range using a bandwidth of 2 nm, a data pitch of 0.2 nm, scanning speed of 100 nm·min$^{-1}$

and a response time of 1 s. Measurements were carried out at room temperature (21°C) and baseline spectra of the buffer were subtracted from the sample spectra. The rotation speed responsible for the alignment of the samples was ∼1000 revolutions per minute, which gives rise to a velocity gradient of ∼1200 $s^{-1}$. The micro-volume Couette cell used had a pathlength of 0.5 mm and was built in-house; equivalent models are commercially available (Kromatek, Great Dunmow, UK). Tropomyosin was obtained from Sigma-Aldrich and FtsZ was supplied by Xi Cheng and R. Pacheco Gomez (University of Warwick). The collagen was guinea pig type I, a heterotrimer comprising two $\alpha$1 chains and one $\alpha$2 chain, supplied by R. Visse (Imperial College, London).

## SAF

Self-assembling fibres (SAFs) have well-understood behaviour in the Couette flow. They are designed peptides that assemble into regular fibres, developed to investigate the fibrillogenesis of proteins.[144] Two synthetic peptides, each 28 residues long, assemble into a regular coiled coil (Figure 30, in parallel and perpendicular orientation). The coiled coil is staggered leaving "sticky-ends", which allow the coils to assemble longitudinally into extended fibres. Lateral bundling of the coiled coils into fibres creates stiff rods tens of microns in length. The rigidity of the fibres ensures minimal deformation in shear flow. This, combined with the known orientation of coiled coils in fibres,[145] makes SAFs a useful model case. If the $\alpha$-helices are aligned to the $z$-axis, parallel to the flow direction, the calculations result in a positive maximum around 205 nm (Figure 31, right), while a perpendicular orientation leads to a minimum at this wavelength. Comparison with experiment indicates that the $\alpha$-helices of the SAF are indeed oriented parallel to the fibre axis.

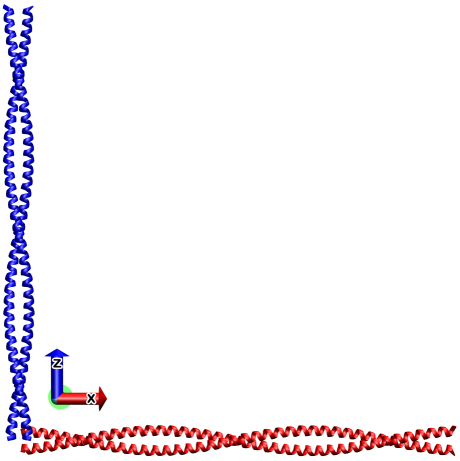



Figure 30: Orientations of the SAF used for the calculations. The colours indicate parallel (blue) and perpendicular (red) orientation.

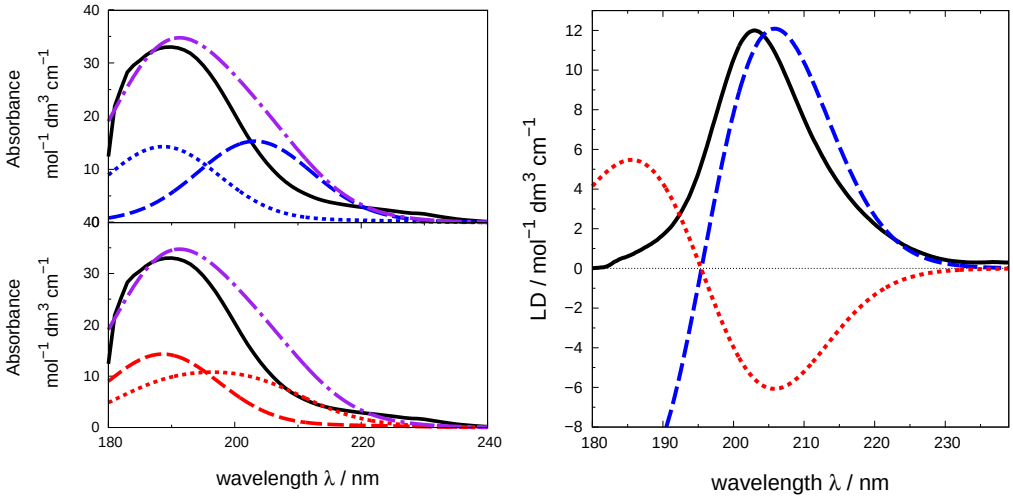



Figure 31: Left: Experimental absorbance spectrum of a SAF (solid line) and calculated spectra for parallel (top, blue) and perpendicular orientation (bottom, red). Parallel absorption (dashed), perpendicular absorption (dotted), and full absorption (dot-dashed).
Right: Experimental LD spectrum (solid line) of a SAF aligned in a Couette flow cell and calculated spectra of parallel (dashed) and perpendicular (dotted) orientations. The experimental spectra are scaled to the calculated intensity for comparison.

## Tropomyosin

Tropomyosin (PDB code 2d3e) is a pure $\alpha$-helical protein consisting of four chains, forming two coiled coils (Figure 32).[146] It has a helical content of 92% according to the DSSP algorithm.[104] Filaments are formed by the interactions of overlapping ends of the coils, thereby extending the helix longitudinally.[147] The left-handed superhelix created by the consecutive coiled coils is about 38 – 40 nm long, with a diameter of less than 2 nm.[148] The rod-shaped constitution makes it a good candidate for LD experiments. The near-UV and $n \rightarrow \pi^*$ region of the far-UV LD spectrum of tropomyosin has been characterized previously[144] and LD has also proven useful in the investigation of tropomyosin binding to actin.[124] While the SAFs form thick stiff rods, tropomyosin is a natural polypeptide with greater flexibility and the rods of tropomyosin are orders of magnitude thinner than for SAF fibres. For calculations, conformational flexibility potentially poses an additional challenge. Calculations were carried out with tropomyosin aligned to the $z$-axis (perfectly parallel to the orientation direction), or aligned to the $x$-axis (perpendicular to the orientation direction, Figure 33). The position of the band caused by the $\pi \rightarrow \pi^*$ transition is reproduced well and shows a positive sign for the protein oriented parallel to the orientation axis. This is in agreement with the experimental spectrum and the expected orientation of tropomyosin in the Couette flow.

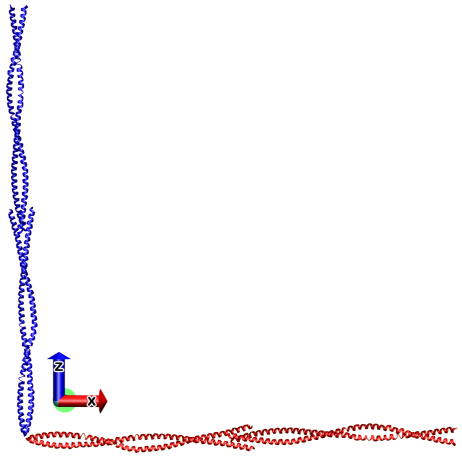



Figure 32: Orientations of tropomyosin used for the calculations. The colours indicate parallel (blue) and perpendicular (red) orientation.

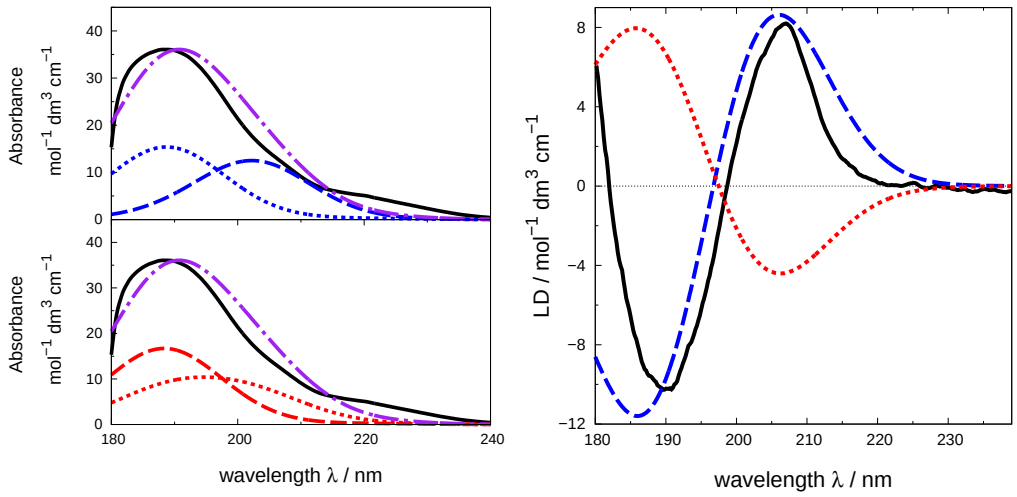



Figure 33: Left: Experimental absorbance spectrum of tropomyosin (solid line) and calculated spectra for parallel (top, blue) and perpendicular orientation (bottom, red). Parallel absorption (dashed), perpendicular absorption (dotted), and full absorption (dot-dashed).
Right: Experimental LD spectrum (solid line) of tropomyosin (left) aligned in a Couette flow cell and calculated spectra for parallel (dashed) and perpendicular (dotted) orientation. The experimental spectrum is scaled to the calculated intensity for comparison.

## FtsZ

FtsZ (PDB code 1fsz)[149] is an $\alpha\beta$-protein and the bacterial homologue of tubulin.[150] The formation of fibrous FtsZ is a key step in cytokinesis. FtsZ contains two mainly parallel $\beta$-sheets surrounded by $\alpha$-helices. The secondary structure content, estimated using the DSSP algorithm,[104] is 44% helical and 25% $\beta$-sheet. The protein comprises one chain with guanosine diphosphate bound in a polar pocket. This study focuses on the orientation of the FtsZ protein itself and the nucleotide was neglected in the calculations. FtsZ forms fibres and the orientation of the monomers within the long axis of the fibre is not unambiguously resolved.[151] The mechanism of polymerization has been studied using LD[122,133] and such studies would benefit from a direct comparison between structural models and LD spectra. In contrast to SAF and tropomyosin, which both have high aspect ratios, FtsZ is globular and the definition of a parallel or perpendicular orientation is ambiguous if based on the monomer structure alone. Therefore, the protein has been aligned to the $z$- and $x$-axis, respectively, according to the supposed orientation in the fibre determined from X-ray diffraction measurements on a dimer by Löwe and Amos (Figure 34).[151] The orientation suggested by Löwe and Amos proved not to be unique in producing the correct spectrum, since the intensity could not be taken into account. However, in a truly unknown case, the vast majority of orientations eliminated by the LD experiment and calculation would enable a limited molecular modelling approach to be undertaken to establish the most likely fibre structure. The top structure (blue) in Figure 34 depicts the Löwe/Amos orientation within the fibre. The blue monomer is, therefore, oriented parallel in the Couette flow. The calculated LD spectra show a positive maximum at 206 nm for the parallel orientation and a negative band for the structure aligned perpendicular to the orientation axis (Figure 35).

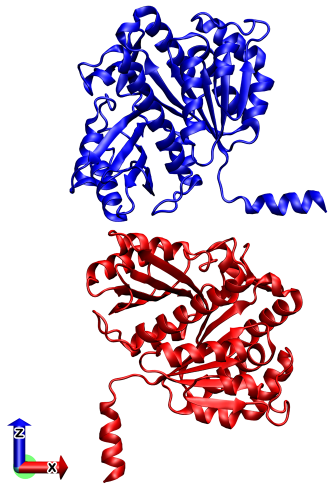



Figure 34: Orientation of FtsZ used for the calculations. The colours indicate parallel (blue) and perpendicular (red) orientation.

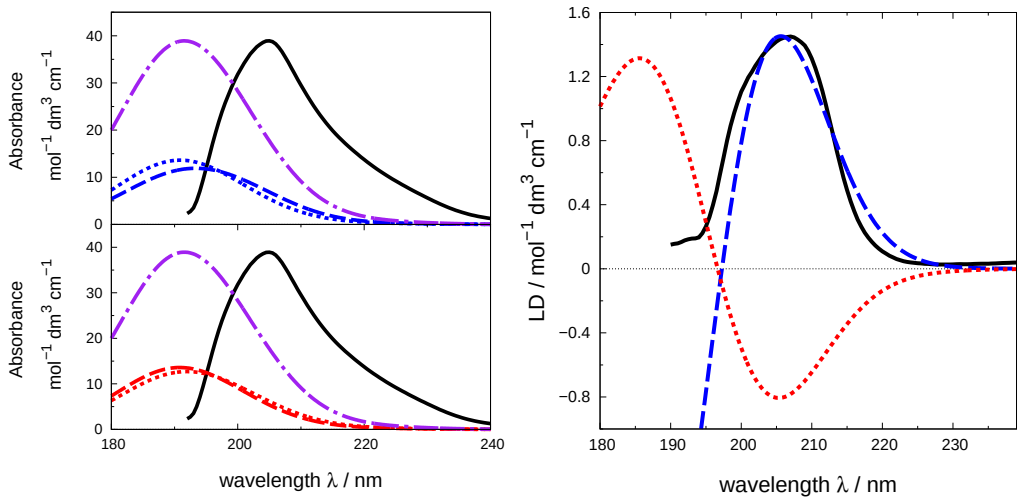



Figure 35: Left: Experimental absorbance spectrum of FtsZ (solid line) and calculated spectra for parallel (top, blue) and perpendicular orientation (bottom, red). Parallel absorption (dashed), perpendicular absorption (dotted), and full absorption (dot-dashed).
Right: Experimental LD spectrum (solid line) of FtsZ aligned in a Couette flow cell and calculated spectra for parallel (dashed) and perpendicular (dotted) orientation. The experimental spectrum is scaled to the calculated intensity for comparison.

The positive band observed in experiment in this region agrees with the calculation of the parallel structure and, therefore, confirms the orientation of FtsZ fibres is indeed in accord with that extrapolated from the experimental dimer structure.

## Amyloid Fibre

In contrast to helical structures, the theoretical origins of the electronic structure of $\beta$-sheet structures are less well understood, even at a qualitative level. For example, Nesloney and Kelly[152] remarked in 1996 that a standard $\beta$-sheet CD spectrum has yet to be agreed upon and that the textbook spectrum, with a maximum at 195 nm and a minimum at 218 nm, is usually that of poly-L-Lys in an aggregated $\beta$-sheet. Previous theoretical studies of $\beta$-sheet CD have reported significant sensitivity to the precise structure and regularity of the protein chain and also to the location of the transition dipole moment on the peptide chromophore.[153] More recent literature[49] suggests that exciton coupling theory predicts a negative band at 175 nm and a positive band 195 nm.

Amyloid fibres have been linked to conditions such as Parkinson's and Alzheimer's disease and are, therefore, a focus of current research.[154–156] A common feature in amyloid-like fibrils is the cross-$\beta$ pattern, which is characterized by pairs of $\beta$-sheets that are parallel to the long axis of the fibril. The fibril-forming segment GNNQQNY has been studied using X-ray diffraction analysis and the structure and orientation of the fibrous crystals in the fibrils is known (PDB code 1yjp).[157,158] A model containing 15 $\beta$-strands in parallel sheets has been created according to the experimental results of the X-ray diffraction results. The orientation of the parallel segment (Figure 36,

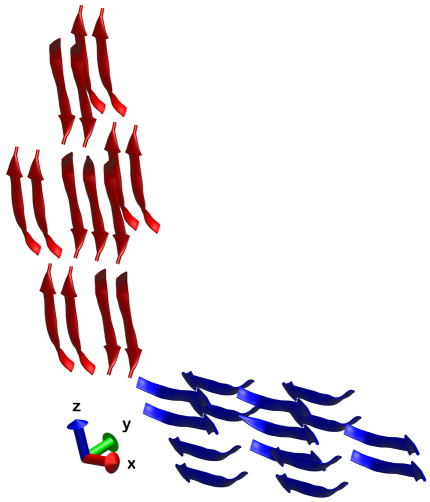



Figure 36: Orientation of the amyloid fibril used for the calculations. The colours indicate parallel (blue) and perpendicular (red) orientation.

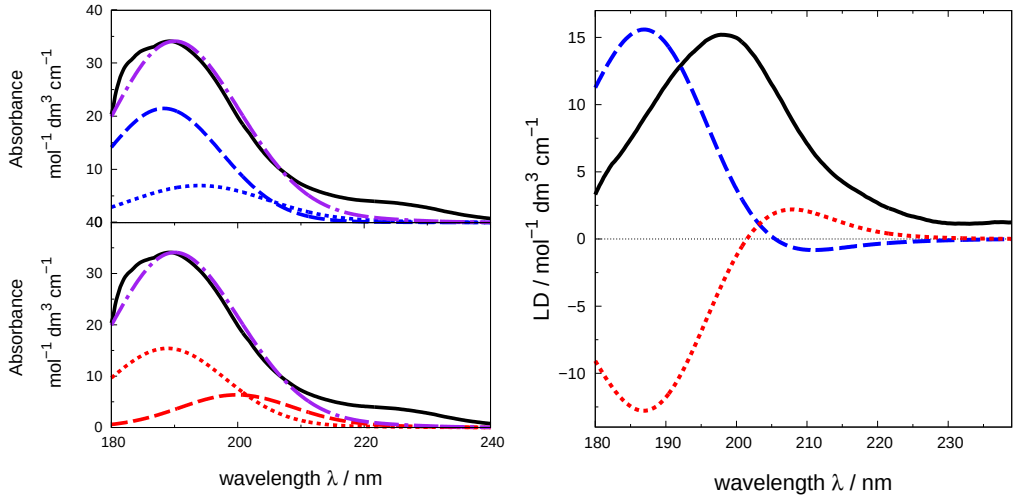



Figure 37: Left: Experimental absorbance spectrum of an amyloid fibril (solid line) and calculated spectra for parallel (top, blue) and perpendicular orientation (bottom, red). Parallel absorption (dashed), perpendicular absorption (dotted), and full absorption (dot-dashed).
Right: Experimental LD spectrum (solid line) of the amyloid fibril aligned in a Couette flow cell and calculated spectra for parallel (dashed) and perpendicular (dotted) orientation. The experimental spectrum is scaled to the calculated intensity for comparison.

Table 10: Interaction energies for each peptide $i$ between electronic excitations within a single strand.

| | | **Interaction energy / cm$^{-1}$** | | | | |
|---|---|---|---|---|---|---|
| $i$ | $r_{i,i+1}$ / Å | $\pi\pi_i^* n\pi_{i+1}^*$ | $\pi\pi_i^* \pi\pi_{i+1}^*$ | $n\pi_i^* n\pi_{i+1}^*$ | $n\pi_i^* \pi\pi_i^*$ | $\pi\pi_i^* \pi\pi_{i+2}^*$ |
| 1 | 3.00 | 92 | −846 | 26 | 76 | 11 |
| 2 | 3.52 | 336 | −934 | 81 | 67 | 59 |
| 3 | 3.63 | 409 | −762 | 46 | −26 | 16 |
| 4 | 3.44 | 357 | −966 | 73 | 67 | 84 |
| 5 | 3.41 | 613 | −930 | 114 | −15 | n/a |

blue) was, therefore, defined to be consistent with its orientation in the X-ray data for fibrils: the $\beta$-sheets are oriented parallel and the single strands perpendicular to the orientation axis (in contrast to the long axis of the fragment in the PDB, which coincides with the axis of the strand). The sheet extends parallel to the shown strands of each of the fragments and the short axis of the fragment is, therefore, the long axis of the sheet. The experimental spectrum shows an intense band at 198 nm, which is due to the $\pi \rightarrow \pi^*$ transition polarized parallel to the sheet axis (Figure 23). The calculated band of the parallel sheet is about 10 nm blue-shifted (Figure 37). However, the $\pi \rightarrow \pi^*$ band of the perpendicular sheet orientation (Figure 37) is negative and disagrees with experiment.

Every strand in the amyloid fibre has the same structure and the seven residues in each strand correspond to six peptide bonds. Calculations on individual strands or pairs of strands from the amyloid structure gave similar absorption and LD spectra. Thus, the electronic structure of $\beta$-sheets appears, from these calculations, to be dominated by nearest-neighbour interactions along individual strands. The largest coupling between electronic excitations is the coupling between $\pi \rightarrow \pi^*$ transitions located on neighbouring pep-

Table 11: Interaction energies between $\pi \rightarrow \pi^*$ excitations located on adjacent strands within a parallel $\beta$-sheet.

| **Peptide groups** | **Interaction energy / cm$^{-1}$** | | | |
|---|---|---|---|---|
| $i, j$ | $\pi\pi_i^*\pi\pi_{j-1}^*$ | $\pi\pi_i^*\pi\pi_j^*$ | $\pi\pi_i^*\pi\pi_{j+1}^*$ | $\pi\pi_i^*\pi\pi_{j+2}^*$ |
| 1 | n/a | 340 | −29 | −29 |
| 2 | −107 | −450 | 159 | 67 |
| 3 | 91 | −374 | 79 | −97 |
| 4 | 32 | −271 | 79 | 70 |
| 5 | 98 | −406 | 131 | n/a |
| 6 | 36 | −455 | n/a | n/a |

tide groups within a single strand (Table 10). This interaction is on the order of 850 cm$^{-1}$ $\pm$ 100 cm$^{-1}$ and gives rise to the splitting of $\sim$5 nm between the parallel and perpendicular components in the calculated absorption spectrum. If the splitting were significantly larger, it would lead to a couplet in the LD spectrum. Calculations using a semi-empirical parameter set,[72] where the orientation of the $\pi \rightarrow \pi^*$ electric transition dipole moment is different, lead to nearest-neighbour $\pi\pi^* - \pi\pi^*$ interactions almost twice as large as the calculations give, and to a correspondingly larger exciton splitting. A couplet is not observed in the experimental LD spectrum, indicating that the calculated interaction is, at least, qualitatively correct. Other interactions between excitations within the strand are noticeably smaller (Table 10). However, the coupling between the $\pi \rightarrow \pi^*$ transition on one peptide group and the $n \rightarrow \pi^*$ transition on the following peptide group is significant. Interactions below 50 cm$^{-1}$ have little quantitative effect on the computed spectra. Equivalent atoms in the two strands comprising the parallel sheet are separated by 4.86 Å, which is further than the distance between neighbouring peptide groups within a strand. Thus, the interactions between excitations on different strands (Table 11) are weaker and are also less regular.

## Collagen

Collagen (PDB code 1clg)[159] is rich in glycine, proline and proline hydroxyl (hydroxyproline), usually in this sequence (GPO). The GPO content leads to the formation of extended left-handed helices of the poly(proline)-II ($P_{II}$) type. The helices are stabilized by the steric interaction of the proline side chain groups, while the small glycine residues facilitate a very narrow helix diameter. Three $P_{II}$-helices supercoil to form a triple helix with a length of about 9 nm (Figure 38). The $P_{II}$ structure is still a challenge for matrix method CD calculations and the characteristic negative maximum around 200 nm has not been reproduced.[85,127] Mandel and Holzwarth[160] measured the LD of samples of calfskin collagen and poly-L-proline which had been oriented by flow through a set of parallel capillary tubes. They deconvolved the absorption spectrum of collagen into Gaussian bands, resulting in a parallel band at 200 nm with 40% of the $\pi \rightarrow \pi^*$ intensity and a perpendicular band at 189 nm.

The LD spectrum of collagen in Couette flow has recently been reported[118] and a positive band around 201 nm suggests from estimates of transition polarizations that collagen is probably oriented with its long axis parallel within the fibre. Our calculations were based on the PDB structure 1clg, a synthetic sequence, which is 33% glycine and 66% proline, similar in composition to the experimental sample. The dihedral backbone angles show a small spread around the typical $P_{II}$ values of ($\phi = -75^\circ$, $\psi = 160^\circ$). Calculations on a single helix, in contrast to the triple helical structure, yielded similar LD spectra, indicating that the inter-helical interactions are limited. The calculated LD spectrum (Figure 39) for the parallel orientation of collagen agrees well with experiment. Since the rod-shaped structure is similar to SAF and tropomyosin, the orientation of collagen is assumed to be parallel in the fibre.

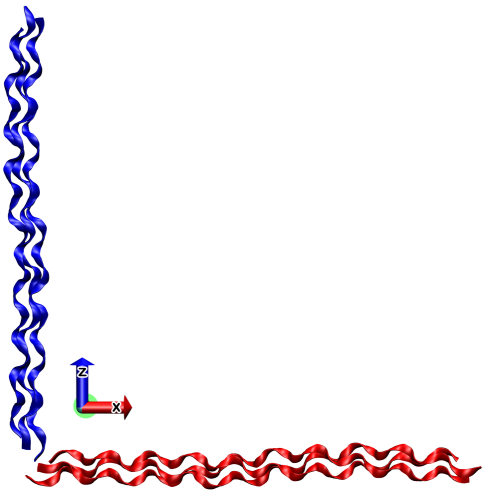



Figure 38: Orientation of collagen used for the calculations. The colours indicate parallel (blue) and perpendicular (red) orientation.

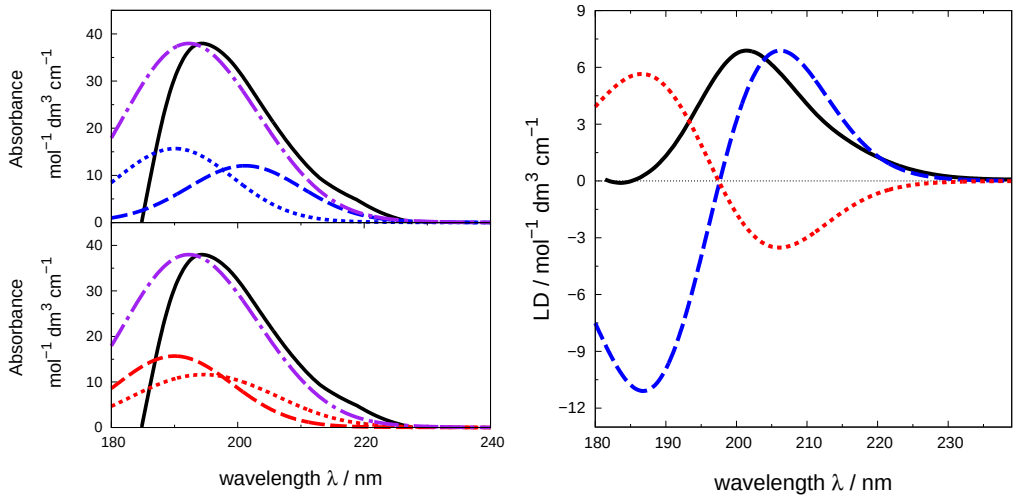



Figure 39: Left: Experimental absorbance spectrum of collagen (solid line) and calculated spectra for parallel (top, blue) and perpendicular orientation (bottom, red). Parallel absorption (dashed), perpendicular absorption (dotted), and full absorption (dot-dashed).
Right: Experimental LD spectrum (solid line) of collagen aligned in a Couette flow cell and calculated spectra for parallel (dashed) and perpendicular (dotted) orientation. The experimental spectrum is scaled to the calculated intensity for comparison.

In the calculations, the parallel polarized component of the $\pi \rightarrow \pi^*$ transition lies at higher wavelength (201 nm) than the perpendicular polarized component (190 nm), but it is slightly less intense. The agreement between the experimental spectrum and the calculated absorption spectrum is good. These components are 5 nm blue-shifted, compared to previous suggestions[118] of 195 nm and 206 nm, but give a similar splitting. Despite this blue shift, the calculated LD band at 206 nm is actually red-shifted by 5 nm compared to the experimental spectrum, suggesting that the perpendicular polarized component of the $\pi \rightarrow \pi^*$ transition is actually more intense than in the presented calculations.

The negative band in the calculated LD spectrum located at 187 nm has no experimental counterpart. Inclusion of charge transfer transitions in the deep-UV[93] did not qualitatively change the calculated LD spectrum. Woody[161,162] has incorporated the coupling of higher energy transitions into the calculation of the CD spectrum of the $P_{II}$ conformation, using a polarizability term, and has suggested that this may account for the non-conservative nature of the CD in the region 180 – 250 nm. The additional negative band in the calculated LD spectrum may be another manifestation of this effect. LD depends only on the electric transition dipole moments, while CD is equally dependent on the magnetic transition dipole moments. The promising results of the LD calculations, reproducing the experimental features of a $P_{II}$ structure, suggest that the challenge for CD is related to the magnetic transition dipole moments.

## Discussion

The aim of this work was to determine whether it was possible to calculate the LD spectra of fibrous proteins and gain information on the orientation of the component units. In contrast to a CD spectrum, an LD spectrum depends only on the electric transition dipole moments. There is, therefore, no sensitivity to the magnetic transition dipole moments, which can be challenging to calculate accurately. This aspect manifests itself in the successful calculation of the LD spectrum of collagen, a structure for which the calculation of the CD spectrum remains problematic. The matrix method LD calculations produced the correct result for five proteins with known orientation in the fibre. In all the examples of fibrous proteins studied in this work, there was a clear indication from the literature of how to define the orientation axis (parallel) from which the definition of perpendicular orientation follows. This was either given by a high aspect ratio or by experimental data in the cases of FtsZ and the amyloid-like fibril. However, more generally such data may not be available. It is, therefore, desirable to be able to perform calculations for any given protein orientation between the perfectly parallel and perpendicular alignment. Once the matrix method methodology for LD had been established, this was the main challenge faced in this work. The LD of an ideal helix is invariant with respect to a rotation about the long helix axis, which can be assumed to be the orientation axis. More complicated folding motifs require consideration of the rotation about the molecular axis for each orientation. This analysis is crucial if LD calculations are to be used to elucidate unknown orientations instead of confirming a previously determined orientation. The functionality is available on DichroCalc[89] and allows one to calculate the LD spectrum for every possible orientation of the monomer unit within the fibre.

The *orientation search* defines the $z$-axis as the parallel direction or orientation axis. The LD is invariant with respect to rotation about this axis, because of the rotational average over the perpendicular axes (Equation 25). In order to cover all possible orientations, the protein is tilted about one of the perpendicular axes at angles between 0 and 90° with respect to the orientation axis. At each tilt angle, the molecule is rotated about its internal molecular axis. For symmetry reasons, orientations with angles greater than 90° with respect to the orientation axis do not have to be considered. An LD spectrum is calculated for every orientation, making it possible to then identify the orientation that gives a calculated LD spectrum in closest agreement with experiment. For each orientation only a single monomer is considered, i.e. different orientations of monomers within the fibre are not yet taken into account. A detailed report is generated, which presents the spectra for each orientation and helps to analyze the results of the calculation. An example report of an orientation search is given in Appendix D.

## 3.6 Conclusion

For all proteins, in the calculated LD spectrum the couplet around 200 nm is reproduced with a good match of the band position. For fibres with a high helical content, the theoretical LD spectra agree very well with the expected orientation of SAF, tropomyosin and collagen with their long axes oriented parallel to the flow direction. For FtsZ, containing $\alpha$-helices and $\beta$-sheets, the results support the orientation determined from X-ray diffraction analyses of a dimer. For the $\beta$-sheet example, the amyloid-like fibril, the calculations agree with experimental evidence of the sheets being oriented parallel to the orientation axis.

Deviations from a perfectly parallel orientation would be apparent in the intensity of the bands, which decreases until the magic angle of 54.7° is reached, at which point no LD signal is observed. Since the experimental and calculated intensities cannot be quantitatively compared (as the orientation parameter is unknown in the experiments), it can only be concluded that the inclination from the orientation axis is smaller than 54.7°, but this does rule out a perpendicular alignment in the flow direction.

In conclusion, matrix method calculations provide insight into protein orientation in shear flow for all major secondary structure classes. The calculations have been made available online to the scientific community. Improvements to the experimental technique or determination of a defined scaling factor taking into account the orientation factor in experiment could allow a quantitative comparison of the experimental and calculated intensity, which would facilitate an even more accurate estimate of the orientation of the protein in solution.

# Chapter 4

# Nucleic Bases

The matrix method for the calculation of CD spectra is particularly successful for proteins. While one might wish to calculate as much as possible *ab initio*, this is computationally prohibitive for whole proteins. However, the independent chromophores considered in the matrix method are of a feasible size for high-level *ab initio* calculations. For the parametrization of the most important chromophore in proteins, N-methylacetamide (NMA) has been used as a model compound representing a single peptide bond.[71] To account for the aromatic side chain groups occurring in proteins, the chromophores of phenylalanine, tyrosine, and tryptophan have been parametrized.[81] Furthermore, parameters for the side chain chromophores of asparagine, aspartic acid, glutamine, and glutamic acid are available. While *ab initio* calculations are possible for chromophores of such sizes, they are computationally expensive and often the determination of the correct active space to use is challenging and may require multiple calculations and analyses.

Due to the popularity of CD and the increasing usage of LD spectroscopy, a multitude of compounds has been studied using these methods. Therefore, there is a demand to parametrize additional chromophores. The complexity of *ab initio* calculations requires some consideration and not all chromophores are equally suited for the exciton matrix method. For the determination of absolute configurations, for example, it has to be established that the exciton Cotton effects indicate the chirality.[163] Therefore, a mechanism has been developed to perform preliminary matrix method calculations on a new chromophore in order to evaluate the possible approach, prior to using expensive *ab initio* methods. A second motivation is that such a routine could make use of already published results for the chromophore in question, since for many molecules the results of semiempirical as well as *ab initio* calculations are available. Therefore, a method has been devised to create parameters for matrix method calculations using only the electric transition dipole moments and excitation energies of the chromophores. For several examples of chromophores with different sizes, the performance of such simplified parameter sets in CD and LD calculations was tested by comparison to experimental spectra.

## 4.1 Introduction

CD spectroscopy can be used to gain insight into DNA and RNA conformation[164] and the ability of polynucleotides to align in Couette flow cells also makes them a target for LD spectroscopy.[124] While the peptide bond covers over 90% of the chromophores present in proteins, DNA and RNA contain five different nucleic bases, which are connected by sugars and phosphate groups.

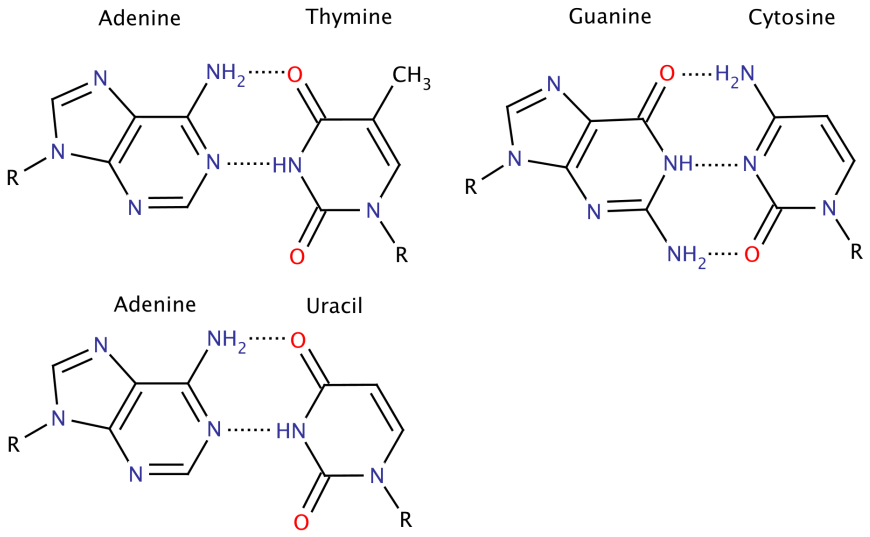



Figure 40: Structures of the nucleic bases in DNA and RNA forming hydrogen bonds with their respective counterparts. Thymine is only present in DNA and is replaced by uracil in RNA.

Adenine (A) and guanine (G) contain a purine chromophore, whereas cytosine (C), thymine (T), and uracil (U) all contain a pyrimidine group. In the helices formed by DNA and RNA, guanine forms three hydrogen bonds to cytosine (Figure 40). In DNA, adenine forms two H-bonds to thymine, while in RNA the latter base is replaced with uracil. The phosphate groups of the polynucleotides show electronic transitions in the deep-UV and the sugar residues have weak absorption bands around 190 nm.[9] The main chromophores of interest in polynucleotides are the aromatic nucleic bases, which have strong absorption bands between 190 and 300 nm.

## 4.2 Methods

In 1962, Tinoco used perturbation theory to develop a method for the computation of protein CD.[44] It was superseded by the matrix method, which improved the original procedure by solving the eigenvalue problem via a matrix diagonalization. This is more accurate than perturbation theory and can be easily implemented in computer programs. The general approach of both methods is to regard the chromophores of the system separately. The electrons are thus localized on the individual groups and may be excited only into higher energetic states of the same chromophore.

The calculations determine the dipole-dipole interactions of the electric transition dipole moments. These dipole moments are defined by the chromophore's electronic transition charge densities, which may be represented by monopole charges positioned around the atoms of the group. For the calculation of the associated electrostatic potential as well as the fitting of the charges, *ab initio* techniques have been used for the amide group and side chain chromophores present in proteins.[71,81] In the absence of *ab initio* data, even with only the magnitudes and directions of the electric transition dipole moments and the excitation energies available, matrix method parameters may be created by representing each dipole moment by only two monopole charges. This is a simplified representation and neglects the interactions of states on the same group. Therefore, the Hamiltonian matrix (Equation 9, page 8) of a dipeptide considering two transitions contains the local excitations, $E_i$, and the interactions between states on different groups, $V_{ij}$, becomes:

$$\hat{H} = \begin{pmatrix} E_{1n\pi^*} & 0 & V_{1n\pi^*;\,2n\pi^*} & 0 \\ 0 & E_{1\pi\pi^*} & V_{1n\pi^*;\,2\pi\pi^*} & V_{1\pi\pi^*;\,2\pi\pi^*} \\ V_{1n\pi^*;\,2n\pi^*} & V_{1\pi\pi^*;\,2n\pi^*} & E_{2n\pi^*} & 0 \\ V_{1n\pi^*;\,2\pi\pi^*} & V_{1\pi\pi^*;\,2\pi\pi^*} & 0 & E_{2\pi\pi^*} \end{pmatrix}. \quad (26)$$

In 1990, Petke *et al.* published *ab initio* results on adenine and guanine.[165] Two years later, they presented calculations on uracil and cytosine.[166] Multireference configuration interaction (MRCI) was used on the ground-state self-consistent field (SCF) orbitals and random phase approximation (RPA) calculations were performed to obtain singlet-singlet transition energies and intensities. The results of the four bases allow calculations on RNA strands. However, data are missing for thymine in DNA.

A series of three publications by Fülscher *et al.* provided calculations on adenine and guanine,[167] uracil and thymine,[168] and cytosine.[169] The results were obtained using the complete active space (CASSCF) and multiconfigurational second-order perturbation theory (CASPT2) methods. The level of theory is higher than for the MRCI/RPA approach used by Petke *et al.*, and the results provided for thymine facilitate the study of DNA.

The publications present the results for the singlet and the triplet excited states, reporting the energies, oscillator strengths and orientations of the electric transition dipole moments, $\vec{\mu}$. Data for the permanent moment of the groups are usually provided. However, the magnetic dipole transition moments, $\vec{m}$, are not given. The effect of $\vec{m}$ on the $\pi \rightarrow \pi^*$ transitions is negligible, since it is a linear displacement of charge. The $n \rightarrow \pi^*$ transition, however, is electric dipole forbidden and the consideration of $\vec{m}$ is crucial.

Due to the lack of magnetic transition dipole moments, only $\pi \rightarrow \pi^*$ transitions were taken into account. For each transition, the magnitude of the transition dipole moment, $\mu$, was calculated from the oscillator strength, $f$, and the transition energy, $\Delta e$ using

$$\mu = \sqrt{\frac{3f}{2\Delta e}} \,. \tag{27}$$

The position of the chromophores in the coordinate system is crucial for the correct placement and orientation of the transition dipole moments. To be used in matrix method calculations, the PDB atom labels need to be assigned to the structures. Each dipole moment is represented by two monopoles of equal but opposite charge with the negative monopole being located at the origin. For a given dipole moment, the separation of the monopoles (that is, the length of the moment) is a function of the magnitude of the charges. However, the charges must not be located outside the chromophore's circumference, as this could lead to interference with neighbouring groups. The transition vector length was, therefore, defined as 1 Å and the charges were determined accordingly to obtain the required dipole moment. A negative charge was placed at the centre of mass and the equivalent positive charge 1 Å apart from it in the direction of the moment angle. The position of the chromophores, atom labels, and the orientations of the electric transition dipole moments used to generate the parameter sets are shown in Figures 41 – 43.

For guanine, Fülscher *et al.* have also provided results considering solvent effects. Using these results improved the agreement of the calculated DNA and RNA spectra with experiment and, therefore, these data have been used in the calculations and are shown in Figure 41.

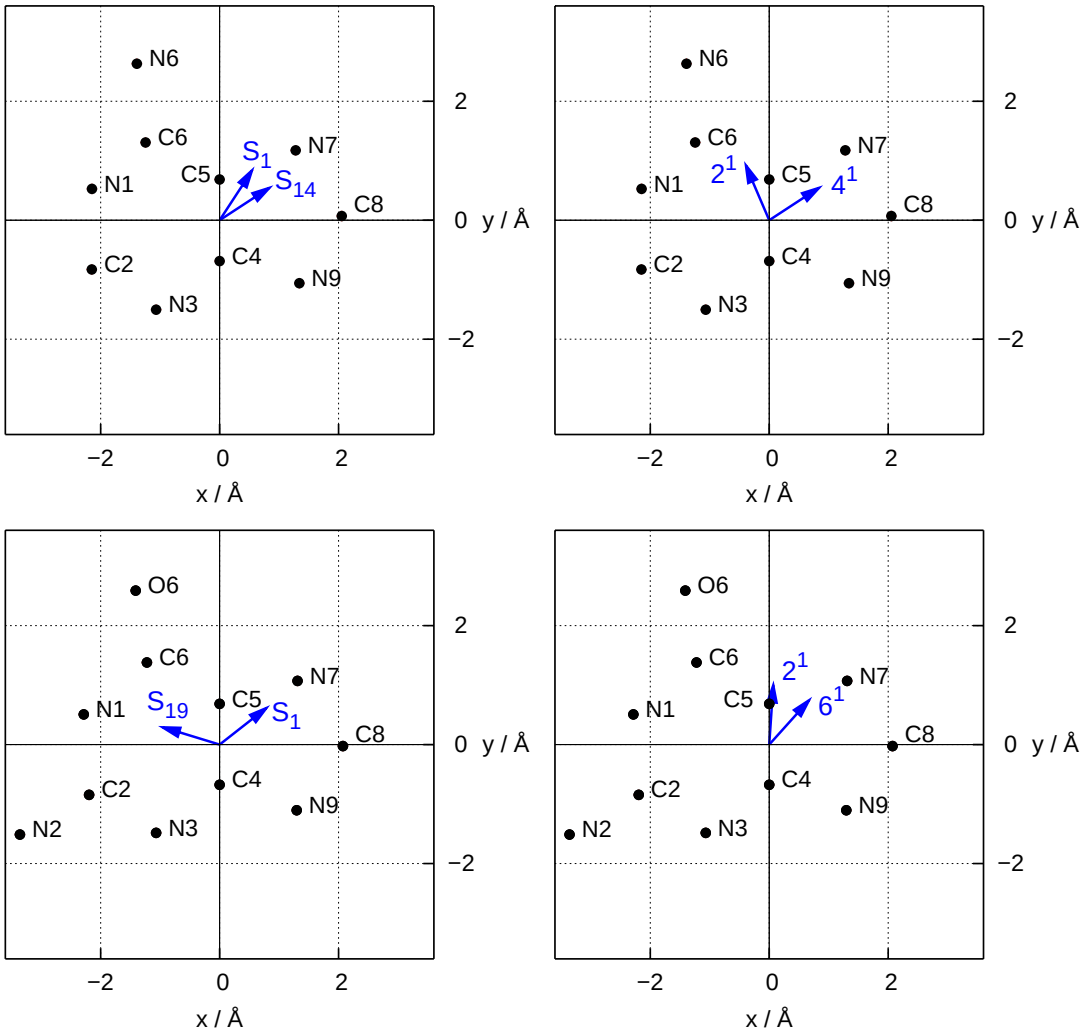



Figure 41: Positions of adenine (top panels) and guanine (bottom panels) in the parameter sets and orientation of the electric transition dipole moments. Left: Using data of Petke *et al.*[165] Right: Using data of Fülscher *et al.*[167]

The results of the two published studies, which use different theoretical approaches, often differ significantly regarding the transition energies, dipole moments and the direction of the latter. For consistency, only parameters derived from the same approach have been used in a calculation. Calculations were only performed if a complete set of parameters for all chromophores contained in the nucleotide was available. This ruled out the MRCI/RPA parame-

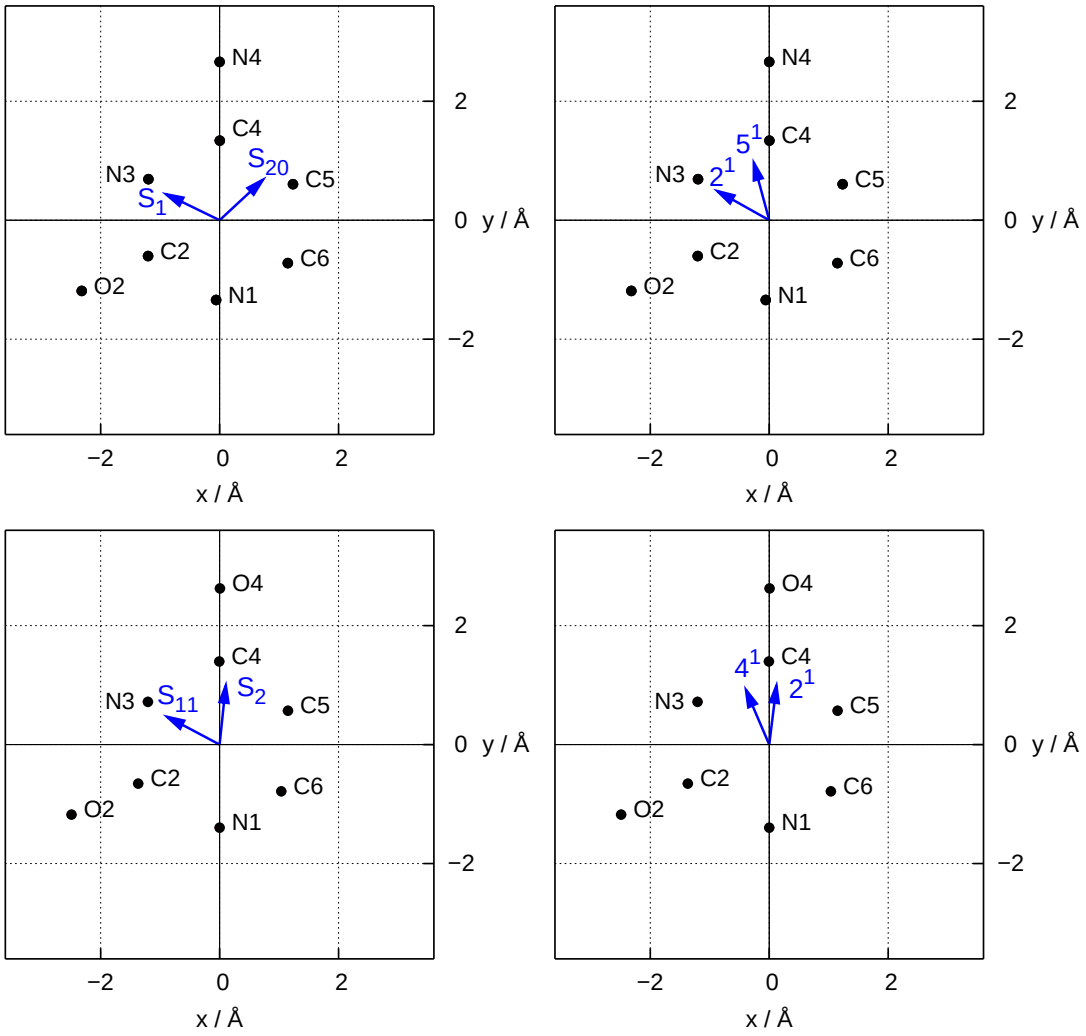



Figure 42: Positions of cytosine (top panels) and uracil (bottom panels) in the parameter sets and orientation of the electric transition dipole moments. Left: Using data of Petke *et al.*[166] Right: Using data of Fülscher *et al.*[168,169]

ters for DNA sequences containing thymine. Calculations were performed using a consistent number of one to four transitions for all nucleic bases present in the sequence.

In their later publication on uracil and cytosine,[166] Petke *et al.* note that the polarizations of uracil agree with experiment, while theoretical evidence sup-

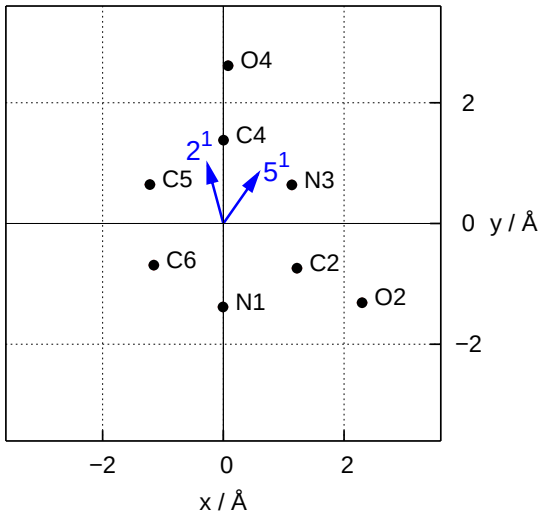



Figure 43: Positions of the thymine atoms and orientation of the electric transition dipole moments in the parameter set created from data of Fülscher *et al.*[168]

ports the deviations from experiment of cytosine, proving the accepted experimental polarizations wrong. This suggestion was later refuted by Fülscher *et al.*,[169] who found that basis set deficiencies are responsible for spurious states created by incorrect Rydberg-valence state mixing. Between 20 and 25 singlet states resulted from the MRCI/RPA calculations of each base, compared to less than 10 states presented from the CASSCF calculations. However, due to the need for testing the dipole-dipole-interaction approach, the small number of experimental spectra, and unexpected results when considering multiple transitions, the MRCI/RPA parameter sets have proven useful (and sometimes superior) and some of the results are shown for comparison.

In the following, parameters derived from the CASSCF/CASPT2 results are referred to as $\mathrm{NB}_{CAS}$ and parameters based on the MRCI/RPA calculations as $\mathrm{NB}_{MRCI}$. For all presented spectra the lowest energy transition and one transition near 190 nm were considered (Table 12). For guanine using the $\mathrm{NB}_{CAS}$ set, the data considering solvent effects were used.

Table 12: Data used to generate the parameter sets of the nucleic bases. Transition energy $\Delta e$ ($cm^{-1}$), wavelength $\lambda$ (nm), oscillator strength $f$, angle of the electric transition dipole moment $\Theta$ (degrees), electric transition dipole moment $\mu$ (Debye).

| | **Petke *et al.* (NB$_{MRCI}$)**[165,166] | | | | | **Fülscher *et al.* (NB$_{CAS}$)**[167–169] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | State | $\Delta e$ | $\lambda$ | $f$ | $\Theta$ | $\mu$ | State | $\Delta e$ | $\lambda$ | $f$ | $\Theta$ | $\mu$ |
| **Adenine** | $S_1$ | 36310 | 275.4 | 0.001 | $-33$ | 0.24 | $2^1A'$ | 41376 | 241.7 | 0.070 | 23 | 1.90 |
| | $S_{14}$ | 50180 | 199.3 | 0.075 | $-57$ | 1.78 | $4^1A'$ | 50329 | 198.7 | 0.370 | $-57$ | 6.00 |
| **Guanine** | $S_1$ | 35790 | 279.4 | 0.239 | $-52$ | 3.77 | $2^1A'$ | 38150 | 262.1 | 0.154 | $-4$ | 2.93 |
| | $S_{19}$ | 51580 | 193.9 | 0.228 | 73 | 3.07 | $6^1A'$ | 53152 | 188.1 | 0.183 | $-42$ | 2.71 |
| **Cytosine** | $S_1$ | 37200 | 268.8 | 0.072 | 64 | 3.95 | $2^1A'$ | 35408 | 282.4 | 0.061 | 60.6 | 1.91 |
| | $S_{20}$ | 51000 | 196.1 | 0.608 | $-47$ | 7.23 | $4^1A'$ | 54362 | 184.0 | 0.147 | 14.9 | 2.40 |
| **Uracil** | $S_2$ | 38000 | 263.2 | 0.369 | $-6$ | 5.95 | $2^1A'$ | 40328 | 248.0 | 0.19 | $-7$ | 3.17 |
| | $S_{11}$ | 50920 | 196.4 | 0.083 | 62 | 5.07 | $4^1A'$ | 52104 | 191.9 | 0.29 | 23 | 3.44 |
| **Thymine** | | | | | | | $2^1A'$ | 39360 | 254.1 | 0.17 | 15 | 3.03 |
| | | | | | | | $5^1A'$ | 57507 | 173.9 | 0.85 | $-35$ | 5.61 |

The calculations were performed using only single and up to four transitions of each considered base. The single transition calculations reproduced features of the characteristic spectra. However, using three or four transitions decreased the agreement considerably. This effect was found for both the $NB_{CAS}$ and $NB_{MRCI}$ parameters. In particular, considering two states with close excitation energies led to a multitude of bands not observed in experiment. For all the chromophores, solely the lowest energy transition ($S_1$ state or $2^1A'$, respectively) is required for the features in the region between 260 and 290 nm. A second transition, chosen for being the closest to 190 nm, accounts for the features in the higher energy region. The interaction between these transitions does not negatively affect the produced bands, which would be the case if a third transition around 200 or 250 nm were included. This effect may be caused by the proximity of strong chromophores, which are much closer in polynucleotides than in proteins. The hydrogen bonding places two strong chromophores with aligned $\pi$-systems close together and the vertical stacking of the base pairs adds equally strong groups above and below the first pair. Test calculations were carried out using only a single base pair and considering four transitions for each base. The adverse effect when using multiple transitions in a complete strand was not found for the two base examples, suggesting that the excessive, anomalous coupling is caused by close vertically stacked systems. The two transitions chosen for each nucleic base (Table 12) are not necessarily the ones with the largest dipole moments. Calculations using specifically the two most intense transitions were tested, but did not improve the agreement compared to parameter sets choosing two transitions depending on the wavelength.

## 4.3 Polynucleotide Conformations

Three main conformations are adopted by regular DNA and RNA strands. The B-helix is the most common conformation, as suggested by Watson and Crick in 1953.[170] This is a right-handed double helix with 10.5 base pairs per turn and a diameter of about 20 Å. The A-form is a wider, right-handed helix, with a diameter of 23 Å and 11 base pairs per turn. This type is adopted, for example, by RNA·DNA and RNA·RNA duplexes and at low humidity.[171] The Z-form is less common than the A and B types and is a left-handed helix with a diameter of 18 Å and 12 base pairs per turn.[172] Although the Z-helix is thermodynamically less favoured, it is of biological significance and the B-Z-transition can be triggered *in vitro* by methylation of the cysteine residues or under certain experimental conditions.[164,173] The possibility of other left-handed types was ruled out by Watson and Crick;[174] however, about a decade later a left-handed B-helix was shown to exist.[175] Nevertheless, due to considerable strain on the sugar-phosphate backbone, the Z-type is the preferred left-handed conformation.[176] Polynucleotide CD spectra are often more sensitive to the orientation of the nucleic bases than on the actual sequence.[34] The CD is dependent on the sequence, but to a lesser extent than it is the case for proteins, where a different sequence causes a change in secondary structure.

Figure 44 shows the characteristic CD spectra of the three main polynucleotide conformations. The spectrum of the A-form features intense maxima around 190 and 260 nm and a minimum is usually visible near 210 nm. B-helices show a minimum around 240 nm and a maximum around 275 nm, with the zero crossing near 260 nm. The Z-form causes a distinct minimum at 290 nm and a positive band around 260 nm.

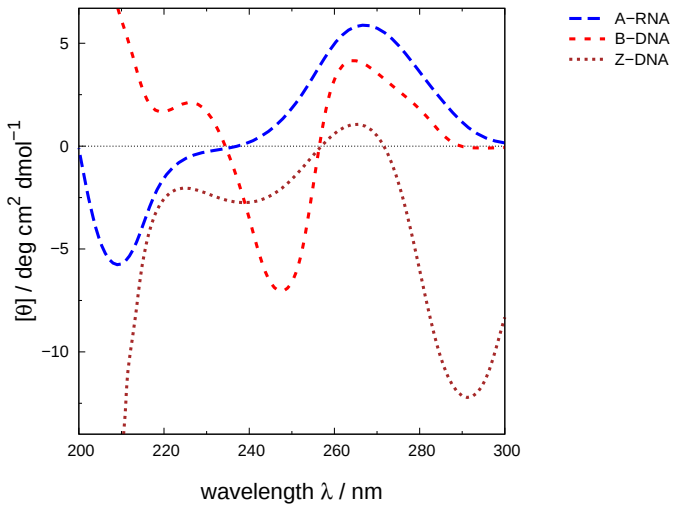



Figure 44: Characteristic CD spectra of polynucleotides in A- (long dashes), B- (short dashes) and Z-form (dotted).[34]

To identify the conformation of the helix, the region between 210 and 250 nm can be inconclusive and is to some extent sequence-dependent. The maximum of the Z-form near 260 nm can easily be mistaken as indicating an A-helix, which also possesses a positive band near this wavelength and may show a minimum around 290 nm. Distinguishing between the A- and Z-form by relying on the long-wavelength bands can, therefore, be problematic. A more significant indication of the Z-form is a strong minimum between 190 and 200 nm, where A-helices show a maximum.[9] The minimum of A-helices in this region is above 200 nm, closer to 210 nm.

## 4.4 Results and Discussion

Several experimental spectra were digitized from literature and the structures created using the Make-NA web server,[177] which uses the Amber program[178] to model the polynucleotides. The publications note the conformation of the duplexes; however, for comparison, the structures were created in different forms. This allowed comparison of the match with the actual helix type and probing to determine whether the results facilitate the distinction between the various conformations. A particularly interesting example was published by Ferreira *et al.*,[164] who provided CD spectra for the same sequence in the B- and Z-form.

In CD calculations, using the full-scale *ab initio* parameters, the intensity of the experimental spectra is reasonably well reproduced. At 222 nm the Spearman rank correlation coefficient between the calculated and experimental intensity is 0.91 for a set comprising 71 proteins with diverse conformations. However, the parameters created using the dipole approach result in much higher intensities than in experiment. A factor contributing to this effect may again be the small separation of strong chromophores in polynucleotides. For the purpose of comparison, the experimental spectra in the presented results were scaled to match the calculation.

The results presented in the following were obtained using the $\mathrm{NB}_{CAS}$ parameters generated from the CASSCF/CASPT2 results by Fülscher *et al.*,[167–169] unless otherwise noted. The calculations using the MRCI/RPA results ($\mathrm{NB}_{MRCI}$) show a comparable agreement, but were omitted due to the lack of thymine data.

### 4.4.1 Nucleic Base CD Calculations

#### 4.4.1.1 DNA and RNA in the Far-UV and Near-UV

For a study on the effect of phosphorothioate modification on the structure and stability of polynucleotides, Clark *et al.* have published the spectra of several DNA·DNA, RNA·RNA, and mixed duplexes.[179] Figure 45, left, shows the experimental spectrum of a $d(AC)_{12}$·$d(GT)_{12}$ duplex compared to the $NB_{CAS}$ calculations. DNA·DNA duplexes adopt the B-form, which is evident by a broad minimum between 230 and 250 nm with a zero crossing at 260 nm and the indicated maximum below 200 nm. The experimental features are qualitatively reproduced by the calculations using the $NB_{CAS}$ parameters. However, a definite distinction between the A- and B-form based solely on these calculations would be ambiguous.

Figure 45, right, shows the spectra for the duplex $r(AC)_{12}$·$r(GU)_{12}$. Polynucleotides containing RNA strands usually exist in the A-form as apparent in the experimental spectrum by the intense maximum at 260 nm. The $NB_{CAS}$ calculation for the A-helix shows a good agreement of the peak positions on the bands at 205, 220 and 233 nm, although the relative intensities of the bands are not reflected. Since no thymine is present the calculations were also performed using the $NB_{MRCI}$ parameters. The positive band at 280 nm, apparent as a distinct (positive) shoulder in experiment, is reproduced with the wrong sign for both sets; however, the $NB_{MRCI}$ calculations show an exact match of the band positions at 260 and 205 nm, while the latter band is not shown for the B-helix using the $NB_{MRCI}$ parameters (data not shown). This allows an unambiguous assignment of the A-helix type.

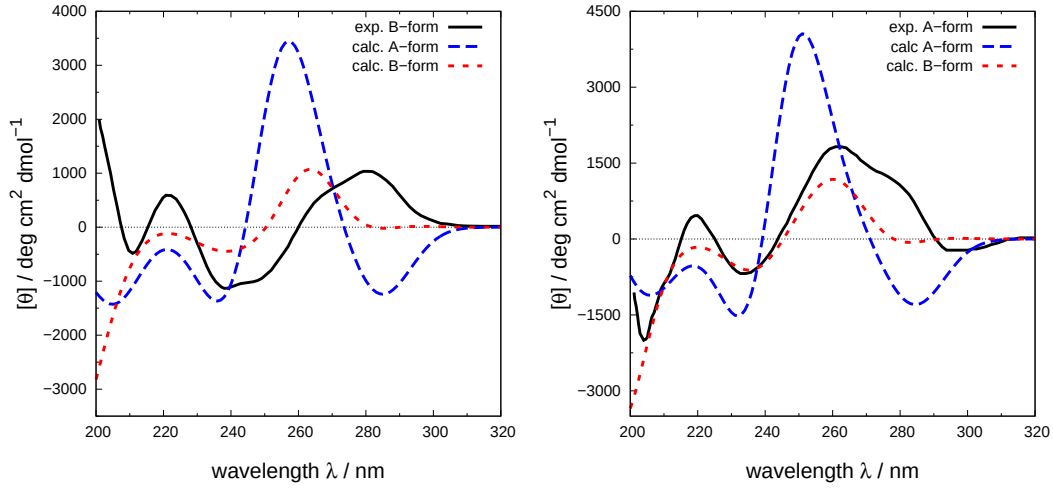



Figure 45: Polynucleotide duplexes calculated using the $NB_{CAS}$ parameters.[179]
Left: DNA·DNA duplex $d(AC)_{12}$·$d(GT)_{12}$, B-form.
Right: DNA·RNA duplex $d(AC)_{12}$·$r(GU)_{12}$, A-form.

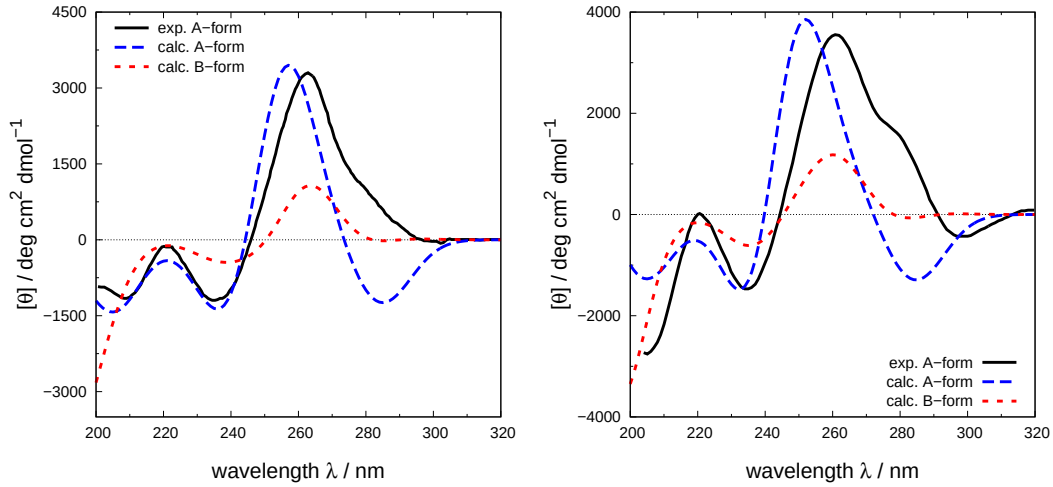



Figure 46: Polynucleotide duplexes calculated using the $NB_{CAS}$ parameters.[179]
Left: RNA-DNA duplex $r(AC)_{12}$·$d(GT)_{12}$, A-form.
Right: RNA·RNA $r(AC)_{12}$·$r(GU)_{12}$, A-form.

The two duplexes r(AC)$_{12}$·d(GT)$_{12}$ and r(AC)$_{12}$·r(GU)$_{12}$ assume the A-form due to the RNA strands (Figure 46). The RNA·DNA duplex (left) shows a very good agreement of relative intensities and band positions for the A-form and, therefore, agrees with experiment. The RNA·RNA assignment would be ambiguous judging solely from the NB$_{CAS}$ results, since the three low-energy bands agree well with the B-form calculation and only the band at 205 nm correlates with the A-helix spectrum. However, the NB$_{MRCI}$ calculations support the A-form assignment (data not shown), with an exact match of the band positions at 205 and 260 nm, while the former band is not shown for the B-helix calculation. The NB$_{MRCI}$ parameters do not reproduce the couplet centred at 225 nm, but given that the excitations between 190 and 260 nm are not considered this is to be expected.

A particular polynucleotide sequence usually adopts a specific conformation. Using special conditions such as a low or high humidity, and with certain reagents, it is possible to trigger a conformation change, which can be monitored by CD. For the DNA·DNA duplex d(C)$_4$·d(G)$_4$ the transition from B- to Z-form under the influence of high concentrations of sodium and cobalt-III-hexamine has been studied.[164] The model of the B-form was built using the Make-NA server[177] Since the server only supports building the A- and B-type, for the Z-helix the PDB file 1zna was used.[180] The NB$_{CAS}$ calculation for the B-form (Figure 47, left) is in good agreement with experiment for the two bands at 260 and 280 nm. A small additional couplet is produced around the zero-crossing at 270 nm. This artefact is not observed for the NB$_{MRCI}$ calculations, which are in similar good agreement (data not shown). The calculation of the Z-form (Figure 47, right) reproduces the inverted couplet with a red shift of 10 nm. The assignment of the Z-type is, however, unambiguously possible.

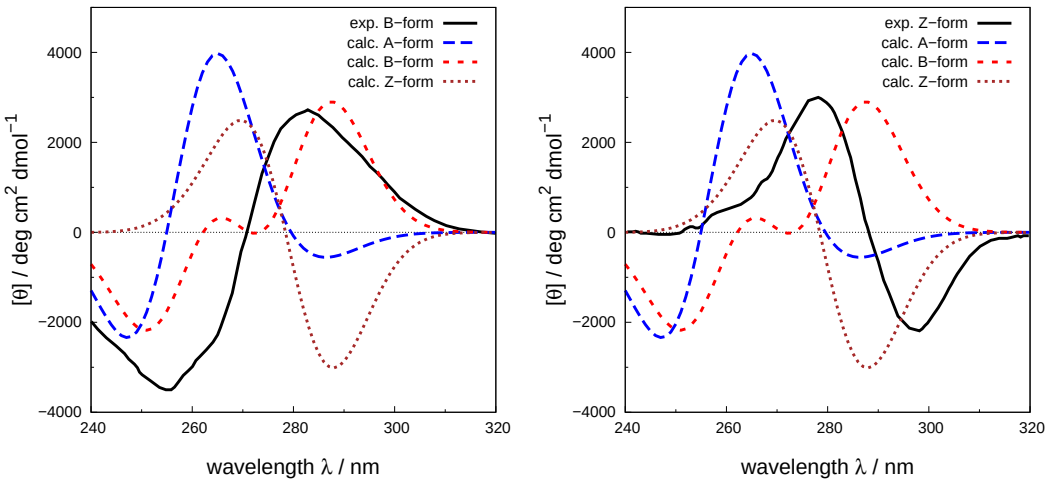



Figure 47: Polynucleotide duplexes calculated using the $NB_{CAS}$ parameters.
Left: DNA·DNA duplex, $d(C)_4d(G)_4$, B-form.
Right: DNA·DNA duplex, $d(C)_4d(G)_4$, Z-form.[164]

#### 4.4.1.2 DNA and RNA in the VUV

As noted in Section 4.3, Page 105, distinguishing between different conformations can be challenging if only one region of the spectrum is examined. The Z-type shows a minimum near 290 nm; however, a similar band may also be observed for A-type structures. Furthermore, Figure 47, right, shows a Z-helix spectrum with a more intense maximum at 280 nm than the negative band around 300 nm and, therefore, the former band can be ambiguous as well. Z-RNA duplexes present an additional complication, as their features in the near-UV are even more similar to A-RNA.[181] For a definite assignment, the features in the far-UV down to 190 nm may be necessary, and the VUV region, below 190 nm, provides additional information.[42]

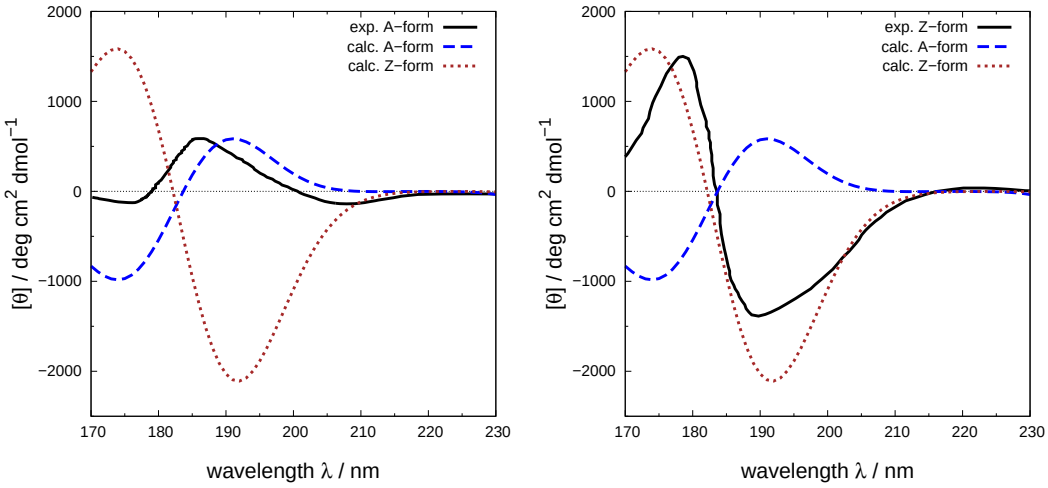



Figure 48: VUV spectra of r(GC)$_7$·r(GC)$_7$ duplexes in A-form (left) and Z-form (right), calculated using the NB$_{CAS}$ parameters.[173]

Figure 48 shows the far- and VUV CD spectra of the duplex r(GC)$_7$·r(GC)$_7$ in A- and Z-form. Using the Make-NA server,[177] the A-form was created; for the Z-form, the PDB 1d39 was used. The calculations using the NB$_{CAS}$ parameters reproduce the couplets with a good agreement of the band positions and thus facilitate a definite identification of the conformation. A comparable agreement was obtained for NB$_{MRCI}$ calculations. Notably, for this duplex an assignment of the conformation using only the near-UV data would not be definitive, and therefore, the VUV data are crucial.

An exception of the fairly regular polynucleotide structures are overhangs and bulges in the sequence. Unpaired nucleic bases within a strand cause distortions of the otherwise uniform helix and can, like overhangs at the end of the sequence, act as binding sites for small molecules.[182,183] These deviations pose an additional challenge for the calculations. The interaction between chiral compounds and bulged DNA has been studied[184] using the sequence

```
5'-GTC CGA   TGC GTG T T
3'-CAG GCT G ACG CAC T
```

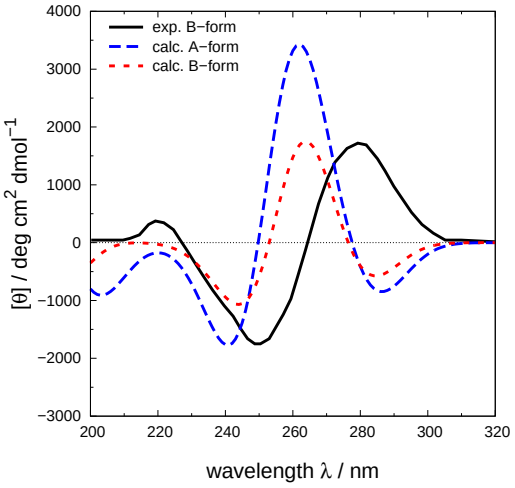



Figure 49: Bulged DNA (B-form) calculated using the NB$_{CAS}$ parameters compared to experiment.[184]

The Make-NA server[177] allows overhangs and the bulge was approximated by removing the complementary base of the guanine residue in the bottom half. The NB$_{CAS}$ calculations (Figure 49) show a good agreement with experiment at 220 and 245 nm, while the maximum at 280 nm is 20 nm blue-shifted. The latter may be caused by the approximation of the bulge.

#### 4.4.1.3 Guanine-Quadruplexes

The ability of guanine-rich nucleic acid motifs to form hydrogen-bonded macrocycles, which can further associate into oligomeric stacks, is of interest due to the chemical diversity of the structures.[185] Guanine possesses self-complementary hydrogen bond donors and acceptors, leading to the formation of G-quartets in parallel or antiparallel orientation of the monomers. The highly dipolar and polarizable aromatic surface of guanine facilitates the stacking of the G-quartets, forming a negatively charged centre in which cations can be bound while the surface of the complex is lipophilic.[186] G-quadruplexes are studied using CD spectroscopy but the analysis of the spectra remains a challenge. The two chiral quartet configurations can be stacked in different orientations with respect to the central axis and lead to diastereomers. Furthermore, the CD of the stacked quartets is influenced by the other chromophores of the quadruplex-forming sequence. Theoretical CD analyses are, therefore, a valuable tool to support the characterization of the G-quadruplexes.[185–187]

Figure 50 shows the experimental spectra of a parallel and an antiparallel G-quadruplex compared to the results derived from the $NB_{CAS}$ parameters. For the parallel example (left), the relative intensity of the band at 200 nm is greatly underestimated; however, the band itself is reproduced. The couplet centred around 245 nm is obtained with the correct relative intensities, although with a red-shift of about 15 nm. For the antiparallel structure, the calculations do not reproduce the bands at 210 and 290 nm. However, the couplet centred at 260 nm shows a positive agreement and the parameters are, therefore, capable of distinguishing between the two conformations of the G-quadruplexes.

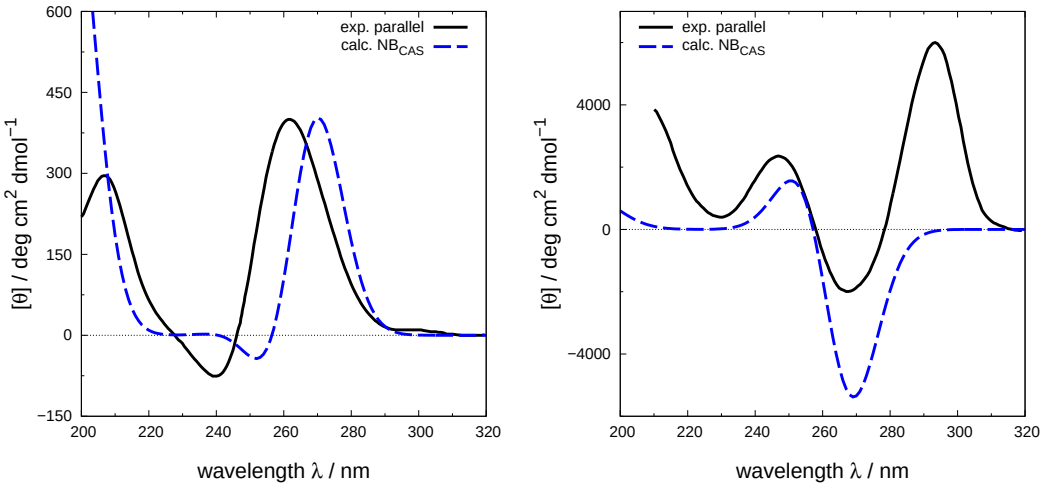



Figure 50: CD spectra of parallel (left) and antiparallel (right) G-quadruplexes calculated using the $NB_{CAS}$ parameters.[185]

### 4.4.2 Nucleic Base LD Calculations

The formation of FtsZ fibres can be monitored via the increasing LD signal since only the fibres align and, therefore, show an LD signal. In the near-UV region, the LD spectrum is dominated by the guanine chromophore of one guanosine-triphosphate residue (GTP) in a cavity of the protein. The bundling of FtsZ, that is, the lateral association of the fibres, can be triggered by calcium ions and shows a distinct change of the guanine LD in the near-UV. The simultaneous change of the guanine orientation and the fibre bundling was explained by the exchange of GTP at the active site with $Ca^{2+}$ and the reorientation of GTP. From the positive LD signal it was deduced that the guanine chromophore aligns in a parallel orientation on the FtsZ surface and facilitates the lateral association of the fibres.[133]

To investigate this mechanism, FtsZ (PDB code 1fsz) was aligned parallel to the orientation axis ($z$). The orientation of the monomers in the fibre is known from X-ray diffraction measurements[151] and was confirmed by LD calculations (Chapter 3, Page 57). The guanine chromophore was moved away from its initial position in the cavity to the surface of the protein at a distance close enough to interact with the neighbouring chromophores. Two different structures were created with a displaced guanine residue, with the chromophore aligned exactly parallel to the orientation direction and oriented perpendicular to it (Figure 51). The results obtained with parameters using the MRCI/RPA or the CASSCF/CASPT2 data, respectively, are equivalent. For polynucleotides, the consideration of solvent effects for guanine improved the agreement of the calculations with experiment. However, for FtsZ, containing only a single nucleic base residue, the $NB_{CAS}$ parameters of guanine using the gas-phase data performed better and these results are shown in Figure 52.

The calculated LD spectrum in the far-UV agrees with the previous calculations in Chapter 3. The experimental spectrum shown in Figure 52 (left) was measured after the addition of calcium and fibre bundling. The $n \rightarrow \pi^*$ band is more pronounced than in the spectrum without calcium (Figure 35, Page 82) and is underestimated by the calculations. Without calcium, the experiment shows a broad, slightly negative band in the near-UV between 240 and 280 nm. This is reproduced by the calculation with the guanine in its initial position. After the addition of calcium, the near-UV LD signal increases significantly and a strong positive maximum near 260 nm emerges with a negative band 280 nm. The calculation with the displaced guanine chromophore in parallel orientation reproduces the positive signal, while predicting a strong minimum if the residue is aligned perpendicular. These results support the suggested mechanism of the reorientation of GTP during fibre association.[133]

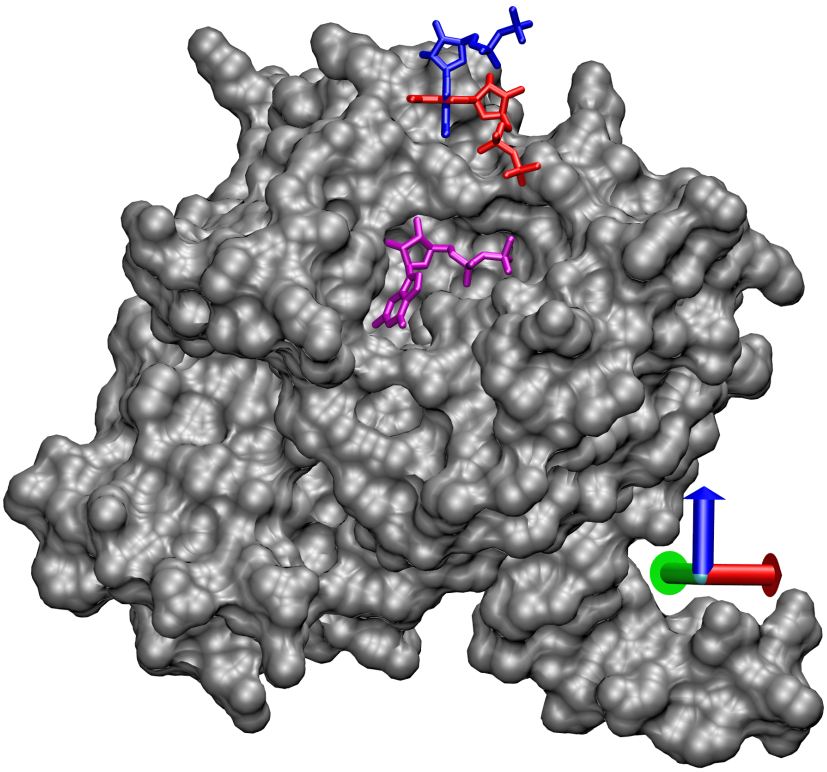


Figure 51: The guanine chromophore in FtsZ in its original orientation in the cavity (purple), and oriented parallel (blue) and perpendicular (red) to the $z$-axis outside the cavity.

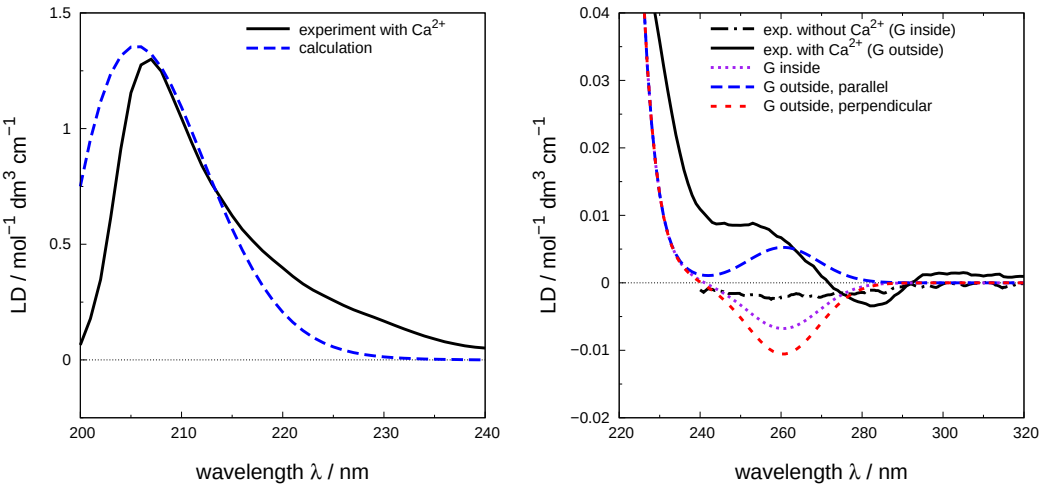



Figure 52: Left: LD spectrum in the far-UV for the parallel orientation of FtsZ. Right: LD spectra of FtsZ with different orientations of guanine. With the original orientation in the cavity (purple, dotted), and with parallel (blue, long dashes) and perpendicular (red, short dashes) orientation outside the cavity.

## 4.5 Conclusion

The spectra calculated using the simplified matrix method parameter sets successfully reproduce features observed in experiment. Due to the small separation between the chromophores, multiple transitions with very close excitation energies lead to excessive coupling and reduce the agreement with experiment. Selective inclusion of the lowest energy transition around 260 nm and one transition near 190 nm avoided the effect, and the important bands necessary for the identification of the nucleotide conformation were reproduced. The parameters were derived from *ab initio* calculations at two different levels of theory. While the CASSCF/CASPT2 method is superior to MRCI/RCA, both approaches showed a positive agreement with experiment. The dipole-dipole-interaction approach, therefore, can be used to create matrix method parameter sets from the dipole moments and excitation energies of chromophores.

The calculations of uniform and mixed DNA and RNA duplexes allowed the identification of the duplex conformation. Distinguishing between the A- and B-forms is a challenge, but the Z-form could be reliably assigned. Features in the near-UV, far-UV and the VUV were shown to agree with experiment. For an unambiguous identification of the helix type, the spectral features in the far- and VUV can be crucial. Overhangs and DNA bulges are a challenge for the calculation and the distinction between A- and B-form was less clear. The calculations of guanine quadruplexes allow an unambiguous distinction between the parallel and antiparallel orientations of the guanine residues. For FtsZ fibre bundling, using LD calculations a suggested mechanism involving a reorientation of the guanine chromophore could be strengthened.

The parameters derived from the CASSCF/CASPT2 calculations by Fülscher *et al.*[167–169] were implemented in DichroCalc. The nucleic bases can be considered individually to study the contribution of particular bases and the effects of interaction between the bases.

The agreement of the calculations based on the simplified matrix method parameters is promising and illustrates the potential of CD and LD calculations for the study of nucleic acids. The presented data suggest that full-scale *ab initio* parametrization of the bases could be beneficial. This may improve the match of the CD intensity with experiment and facilitate the consideration of additional electronic transitions.

# CHAPTER 5

# NAPHTHALENEDIIMIDES

## 5.1 Introduction

Carbon nanotubes have been a focus of research since 1993[188,189] and a vast range of possible applications has been devised for these covalent structures.[190] The concept of molecules forming tubular structures is neither limited to only carbon nor to covalent bonds between the monomers. A class of $\alpha$-amino acid functionalized naphthalenediimides (NDIs, Figure 53) has been synthesized from 1,4,5,8-naphthalenetetracarboxylic anhydride and the respective $\alpha$-amino acid, using microwave dielectric heating. In both nonpolar solution and solid state, the NDIs form hydrogen-bonded helical nanotubes,[191] which have a propensity to bind guest molecules in the core.[192]

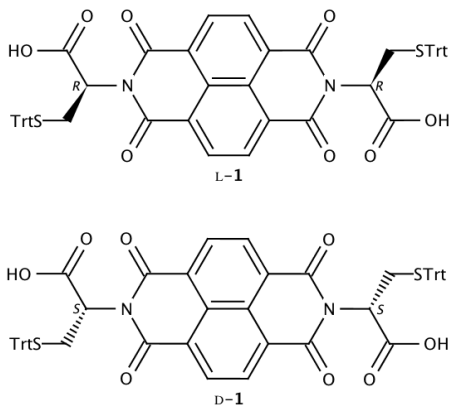



Figure 53: Cysteine derivatives of NDIs forming helical nanotubes.

NDIs in general are chemically robust and versatile. The diimide nitrogens and the naphthalene core can be functionalized to obtain, for example, semiconductor materials,[193] dyes with variable absorption and emission properties,[194] or molecules capable of self-organization via $\pi - \pi$-stacking, hydrogen bonds and van der Waals interactions.[195] Sterzel *et al.* have studied the NDI monomer using CASSCF/CASPT2 calculations to determine Franck-Condon parameters for the $1^1B_{3u}$ and $1^1B_{2u}$ states.[196] The absorption and emission spectra could be reproduced and the substituents of two N-substituted derivatives showed little effect on the low-energy states of the chromophore. The latter finding allows one to transfer results obtained for the optical properties of the NDI monomer to derivatives of the chromophore such as **1**. The same methods have been applied to a covalent NDI dimer and it was found that a $5\pi[4n]5\pi^*$ active space describes the electronic structure sufficiently well to reproduce the most prominent features in the absorption and CD spectrum.[197]

The absolute configuration of the two stereocentres of L-**1** is *R*,*R* for cysteine derivatives (Figure 53). In the solid state L-**1** assembles into *P*-helical nanotubes, that is, right-handed helices, with an average inner diameter of 12.4 Å.

The helices contain three residues per turn and are stabilized by the carboxyl groups on each monomer, which form four strong hydrogen bonds with two neighbours (Figure 54). The NDI planes of successive facing monomers are at an angle of 60°, while the naphthalene moieties of $i$ and $i+3$ monomers are coplanar, stacked along the helix long axis. Each residue forms two weak hydrogen bonds to the $i+3$ monomer, and two to the $i-3$ monomer, which reinforces the structure further.

CD is a popular spectroscopy method, because of its sensitivity to the chirality of molecules in solution. The spectra of two enantiomers are mirror images. Therefore, CD can be used to determine the absolute configuration of chiral molecules.[19–23] A chiral system in solution absorbs the left and right circularly polarized light to a different extent and the difference of these absorptions yields the CD spectrum.

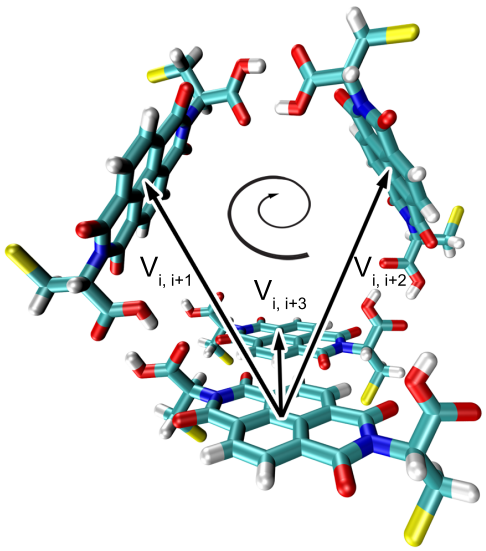



Figure 54: Helical arrangement of three naphthalenediimide residues with a 60° angle between the NDI planes. An additional monomer would form hydrogen bonds to one of the free carboxyl groups with the naphthalene plane parallel above or below the respective $i+3$ monomer.

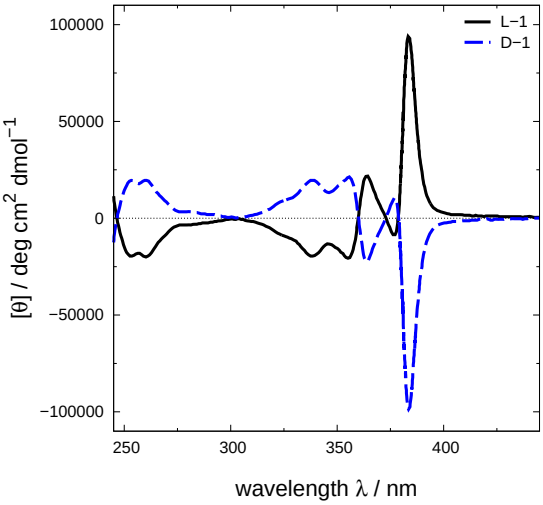



Figure 55: Experimental CD spectra of L-**1** (solid line) and D-**1** (dashed line); solutions in $CHCl_3$, $4.6 \cdot 10^{-5}$ M, recorded at 21°C. Experimental data reproduced from Pantoş *et al.*[191]

In solution, the nanotubes possess a distinct electronic CD spectrum between 250 and 400 nm (Figure 55). The origin of the bands is attributed to the chiral supramolecular structure of the nanotubes, which is formed uniformly, but with opposite chirality from the *R*, *R* (L-**1**) and *S*, *S* (D-**1**) enantiomers. In aprotic solvents, such as $CHCl_3$, the nanotubes can incorporate guest molecules, such as the otherwise poorly soluble fullerene $C_{60}$.[192] The fullerenes sense the chirality of the nanotube, as can be seen in the CD spectrum. After uptake of the fullerenes, additional features appear in the CD, which are mirror-inverted for nanotubes created from different enantiomers of the monomers. Mixtures of the NDI enantiomers self-segregate to form *P*- and *M*-helical nanotubes, allowing uptake of fullerenes at the same ratio of NDI:$C_{60}$ = 3.9:1 as an enantiopure sample.

## 5.2 Methods

Methods for the theoretical calculation of CD have been intensively investigated in order to aid the analysis of CD spectra. A successful theoretical framework for polymers, particularly for proteins, is the matrix method,[33] which is based on exciton theory.[16,198] It is an improved version of the model devised by Tinoco[44] and involves the diagonalization of a Hamiltonian matrix to calculate the interaction between chromophores. In a system with $M$ chromophoric groups each group, $i$, is regarded separately and excitations are only allowed into higher electronic states of the same chromophore. The wavefunction $\psi^k$ is required to calculate the excitations for the state $k$. $\psi^k$ is described as a linear combination of electronic excitations, $\Phi_{ia}$, from the ground state 0 into the excited state $a$:

$$\psi^k = \sum_i^M \sum_a^{n_i} c_{ia}^k \Phi_{ia} \ , \tag{28}$$

where $c_{ia}^k$ are expansion coefficients that reflect the interactions of the states. The Hamiltonian, $\hat{H}$, of the system is a sum of the local Hamiltonians, $\hat{H}_i$, of the independent groups. The terms $V_{ij}$ describe the interactions between two groups $i$ and $j$:

$$\hat{H} = \sum_{i=1}^{M} \hat{H}_i + \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \hat{V}_{ij} \ . \tag{29}$$

Using the monopole-monopole approximation, the interactions between chromophores are assumed to be purely of electrostatic nature. Monopole charges are fitted to the atoms of the group to reproduce the *ab initio* electrostatic potential. The Coulomb interactions between these monopoles then account for the interaction between groups:

$$V_{i0a;j0b} = \frac{1}{4\pi\varepsilon_0} \sum_{s=1}^{N_s} \sum_{t=1}^{N_t} \frac{q_s q_t}{r_{st}} \ , \tag{30}$$

where $q_s$ and $q_t$ are monopole charges on groups $i$ and $j$, $N$ the number of charges on the respective group and $r_{st}$ the distance between the monopoles. Diagonalization of the Hamiltonian matrix yields the energy of each transition (eigenvalues) and the expansion coefficients (eigenvectors), $c_{ia}^{k}$, which are used to derive the electric and magnetic transition dipole moments of the interacting system from the initial dipole moments of the single groups.

The parametrization of a chromophore for use in matrix method calculations requires the electrostatic potential of each electronic transition that is to be considered. Monopole charges have to be fitted to the atoms to reproduce the potential. For small chromophores, the relevant electrostatic potentials can be calculated *ab initio* using, for example, the complete active space self-consistent field (CASSCF) method.[199,200]

Exciton theory is particularly successful for the determination of the absolute configuration of molecules. Exciton matrix method calculations have been used to study a wide range of macromolecular systems. Porphyrins are powerful chromophores, showing exciton coupling over distances of up to 50 Å.[201] Zwitterionic meso-tetra(4-sulfonatophenyl)porphyrin assembles into tubular aggregates and using an exciton approach, considering the four most dominant excited states, the experimental CD spectra could be reproduced.[202] Moreover, calculation of the LD yielded information about the orientation of the monomers forming the tube.[202] Coulombic coupling of nucleotide bases can be exploited to provide insight into DNA structure.[203] Peptidomimetics, such as oligoureas, have also been studied using exciton theory.[204] The axial configuration of oligonaphthalenes was elucidated by introducing additional excitonic chromophores into the structure and analyzing the resulting Cotton effects.[205] The applicability of the exciton chirality method to the NDI chromophore has been demonstrated using several bis-diimides.[163]

Calculations of the absorption spectra using the ZINDO/S Hamiltonian qualitatively reproduced the experimental spectra. The exciton Cotton effects in the CD spectra allowed the determination of the absolute configuration of the diimides. The closely related chromophore 1,4,5,8-naphthalenetetracarboxy dianhydride was studied using time-dependent DFT and the Franck-Condon analysis of the $1^1B_{2u}$ and $1^1B_{3u}$ states provided insight into solvent effects that are not easily deducible from experiment.[206] The exciton chirality method also facilitated investigation of conformation changes of helically stacked NDI chromophores.[207,208]

The success of matrix method calculations suggested the same approach may clarify the origin of the bands in the CD spectrum of the NDI nanotubes and provide useful information about the solution state geometries. Since the spectrum is affected by guest molecules within the tube, theoretical calculations of the CD spectra could aid the analysis of the experimental spectra for these complex host-guest systems.

In this study, matrix method parameters for NDI were created using data presented by Sterzel *et al.*,[209] which were derived from CASSCF calculations[199,200] combined with multi-configurational second-order perturbation theory (CASPT2).[210] The minimal active space needed to describe the electronic structure of the NDI moiety includes the five occupied and five unoccupied $\pi$-orbitals of the naphthalene core and four lone pair orbitals of the carbonyl groups. The $5\pi[4n]5\pi$ active space contains 14 electrons and electronic transitions between 200 and 420 nm arise from seven states (Table 13). The unsubstituted diimide possesses $D_{2h}$ symmetry and the notation referring to this point group is used to denote the states of NDI. A given dipole moment was represented by two monopoles of an equal but opposite charge at the head and tail of the respective vector (Figure 56).

Table 13: CASPT2 results for NDI using a $5\pi[4n]5\pi$ active space. $e_x$, $e_y$ and $e_z$ denote the axis along which the transition dipole moment is oriented (Figure 56).[209]

| State | Energy (vertical) / nm | Energy adiabatic / nm | Oscillator Strength | Dipole D |
|---|---|---|---|---|
| $1B_{1u}$ | 414.5 | 531.0 | 0.00212 | $-0.432\ e_z$ |
| $1B_{3u}$ | 356.0 | 373.6 | 0.02397 | $-1.347\ e_x$ |
| $1B_{2u}$ | 347.7 | 391.3 | 0.15454 | $3.380\ e_y$ |
| $2B_{1u}$ | 251.8 | 286.9 | 0.00146 | $-0.280\ e_z$ |
| $2B_{3u}$ | 244.1 | 251.0 | 1.13510 | $7.676\ e_x$ |
| $2B_{2u}$ | 207.3 | 220.4 | 0.12671 | $2.364\ e_y$ |
| $3B_{3u}$ | 195.6 | 215.9 | 0.00039 | $-0.127\ e_x$ |

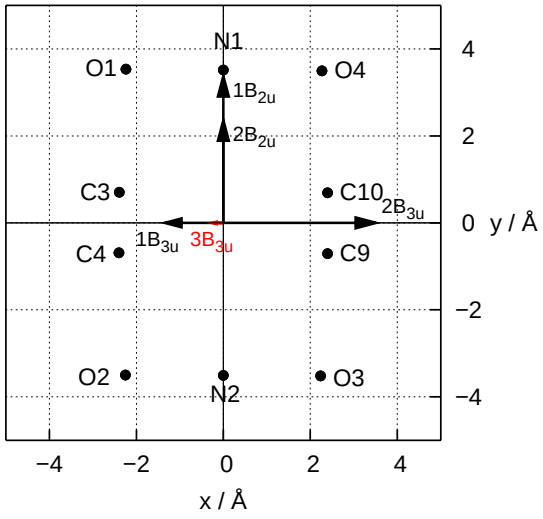



Figure 56: Position of the NDI atoms in the parameter set and the orientation of the electric transition dipole moments. The atom labels refer to the designators used for an NDI moiety in PDB format.

The length of each vector (between 0.1 and 3.4 Å) was chosen to stay within the molecule circumference, to avoid overlapping dipoles with neighbouring groups.

The calculations result in the excitation energy, that is, the wavelength, for each transition on each group and the associated oscillator strengths and rotational strengths. Since, in experiment, the transitions are broadened due to interactions of the molecules with their environment, including the solvent, Gaussian-shaped bands are observed. Therefore, the line spectrum of the gas phase calculation has to be convoluted with a Gaussian lineshape function to be able to compare it directly with experiment.[50] The full width at half maximum (bandwidth) is empirically chosen. A bandwidth of 6 nm was used for all calculated spectra presented herein. Broader bandwidths did not resolve the features visible in the experimental spectra.

For the carboxyl and trityl groups, the calculated spectra show bands in the far-UV and the influence on the NDI bands around 380 nm is negligible. The presented spectra, therefore, were calculated considering only the NDI chromophore. Electronic transitions cause distortions of the molecular orbitals and subsequent vibrations of the atoms as the nuclei oscillate between the initial and the final geometry. Due to the slow reaction of the heavy nuclei to the instantaneous motion of electrons, an electronic transition occurs vertically, that is, with static atom coordinates. Franck-Condon states deviate from the equilibrium excited state geometry and can have a significant effect on the CD spectrum. The band positions for calculations considering only vertical excitations suggested that this is the case for NDI, and vibrational transitions have to be taken into account to reproduce the features in the experimental spectrum.

## 5.3 Results and Discussion

The available data (Table 13, Page 126) provide the energies and dipole moments for vertical and adiabatic excitations.[209] Matrix method parameters have been created for both sets. However, the features of the experimental spectrum could not be reproduced with these data alone. Two of the seven states, $1^1B_{2u}$ and $1^1B_{3u}$, show transitions in the region of interest between 320 and 420 nm. Other transitions have no effect on the bands in this region and do not need to be considered. Using only the vertical excitation energies fails to reproduce the bands in the experimental spectrum, showing only a single couplet centred at 347 nm. The adiabatic excitation energies improve the agreement by shifting the couplet to 391 nm and also resolving a weak positive band at 374 nm. However, for the most intense band at 383 nm, there is a red-shift of 12 nm in the calculated spectrum and the intensity ratios of the bands are not in accord with experiment.

The failure of the vertical and adiabatic transitions to reproduce experimental features indicates that additional information about the electronic structure needs to be taken into account. Consideration of the contribution of vibronic effects improves the description of electronic processes. In addition to the vertical and adiabatic transitions for each of the two electronic states, the parameters of 13 Franck-Condon transitions have been estimated. The most intense, with a Franck-Condon factor greater than 0.7, have been used instead of the vertical or adiabatic excitation energies of the states (Table 14). Calculations using all 13 Franck-Condon transitions did not add visible features to the spectrum (data not shown). The main features in the experimental absorbance spectrum were reproduced (Figure 57, top) using the most intense Franck-Condon transitions given in Table 14.

Table 14: Franck-Condon transitions used to take vibronic effects into account. The most intense transitions with a factor greater than 0.70 reproduce the main features in the absorbance spectrum (Figure 57). Adiabatic energy ($E_{adia}$), vibrational frequency ($\nu_i$), energy of the Franck-Condon transition ($E_{vib}$), wavelength ($\lambda$), Franck-Condon factor and transition dipole moment ($\mu$).

| State | $E_{adia}$ / cm$^{-1}$ | $\nu_i$ / $cm^{-1}$ | $E_{vib}$ / cm$^{-1}$ | $\lambda$ / nm | Factor | $\mu$ / D |
|---|---|---|---|---|---|---|
| $1^1B_{2u}$ | 25559 | 333 | 25892 | 386.2 | 0.90 | 2.224 |
| | | 1471 | 27030 | 370.0 | 1.03 | 2.546 |
| $1^1B_{3u}$ | 26764 | 1471 | 28235 | 354.2 | 1.03 | −0.932 |
| | | 1535 | 28299 | 353.4 | −0.79 | 0.715 |
| | | 1112 | 27876 | 358.7 | −0.73 | 0.660 |

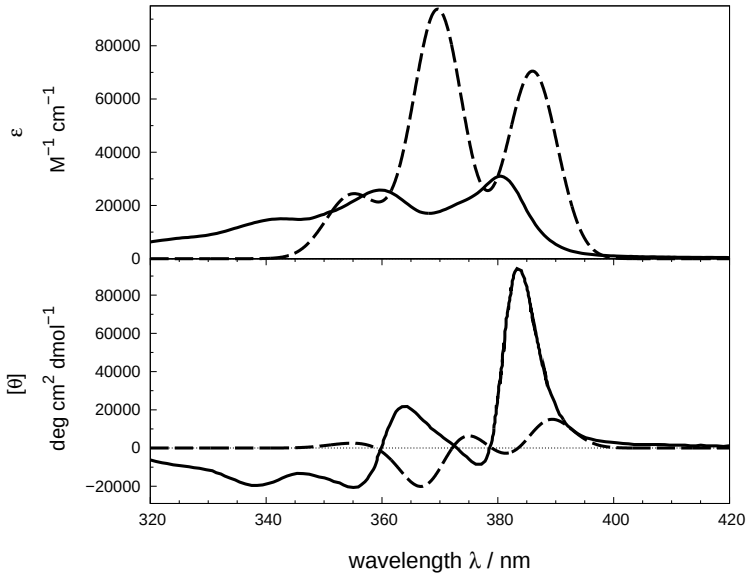



Figure 57: Experimental spectrum from solid-state structure (solid line) and calculated spectrum of a heptamer (dashed line) using Franck-Condon transitions with a factor greater than 0.7 (Table 14). Upper panel: Absorbance spectrum. Lower panel: CD spectrum. The experimental spectra are scaled to the calculated intensity for comparison. Experimental data reproduced from Pantoş *et al.*[192]

For each of these transitions, the dipole moment of the respective adiabatic excitation is relevant, but needs to be weighted appropriately. For $N$ transitions (here $N = 2$ for $1^1B_{2u}$; $N = 3$ for $1^1B_{3u}$), the dipole moment of the transition with the Franck-Condon factor $\varphi$ was multiplied by a weighting factor $\omega$:

$$\omega = \frac{\varphi}{\sqrt{\sum_i^N \varphi_i^2}} . \tag{31}$$

Taking vibronic contributions into account considerably improves the calculated CD spectrum. Two additional bands are resolved and the intensity ratios are reproduced better than in the case for calculations using the vertical and adiabatic energies only. The bands of the calculated spectrum are red-shifted about 9 nm on average. If the calculated spectrum is shifted 9 nm to higher energy, the five bands of the experimental spectrum between 346 and 383 nm can be matched with the calculation. The negative band at 338 nm corresponds to 29586 $cm^{-1}$ and is visible in the absorbance spectrum as a slight shoulder (Figure 57). It is due to a Franck-Condon transition of the $1^1B_{2u}$ state with very small intensity and a Franck-Condon parameter of $-0.22$. It was, therefore, below the cut-off of 0.70 and neglected. The positive band at 355 nm in the calculated spectrum is caused by the three $1^1B_{3u}$ transitions, which create one couplet that is about two orders of magnitude lower in intensity than the following bands. All features above 360 nm originate from the $1^1B_{2u}$ state.

The interaction of dipoles decreases as the cube of their separation. Due to the helical arrangement (Figure 54, Page 121), the $i + 2$ residue is 1 Å further away than the $i + 1$ NDI and the magnitude of the interaction decreases significantly. Apart from the distance, the angle between the dipole moment vectors affects the coupling (Table 15). The helix contains three residues per turn and the $i + 3$ NDI is coplanar with the first one and $\sim$3 Å closer than $i + 1$.

Table 15: Interaction between the most intense transitions of the $1^1B_{3u}$ and $1^1B_{2u}$ states on different residues, $i$. For the arrangement of the residues see Figure 54. The given angle is between the electric transition dipole moment vectors.

| State | $\lambda$ / nm | Interaction | Distance / Å | Angle | Coupling / $cm^{-1}$ |
|---|---|---|---|---|---|
| $1^1B_{2u}$ | 370.0 | $V_{i,\,i+1}$ | 13.3 | 98° | −18 |
| | | $V_{i,\,i+2}$ | 14.3 | 98° | −2 |
| | | $V_{i,\,i+3}$ | 9.0 | 0° | 15 |
| $1^1B_{3u}$ | 354.2 | $V_{i,\,i+1}$ | 13.3 | 50° | 2 |
| | | $V_{i,\,i+2}$ | 14.3 | 50° | −1 |
| | | $V_{i,\,i+3}$ | 9.0 | 0° | −8 |

Figure 58 shows the dependence of the calculated CD spectrum on the angle between two adjacent NDI planes. The intensity decreases uniformly for greater angles, until no interaction is observed for parallel chromophores at an angle of 180°. If the residues allowed a further increase of the angle, this would change the sense of the helix formed by the monomers and invert the signs of the bands.

Calculations have been performed on several oligomers in order to analyze the effect of the nanotube length on the CD spectrum (Figure 59, left). In general, the intensity of the spectra increases with the number of monomers added to the structure. The distance between the chromophore centres across the tube core is about 12 Å and the interaction between such monomers is responsible for the large increase going from a dimer to a trimer. As expected, as further monomers are added, the increase in intensity gradually becomes less pronounced. Figure 59, right, shows how the relative change in intensity depends on the length of the nanotube for the three most intense calculated bands.

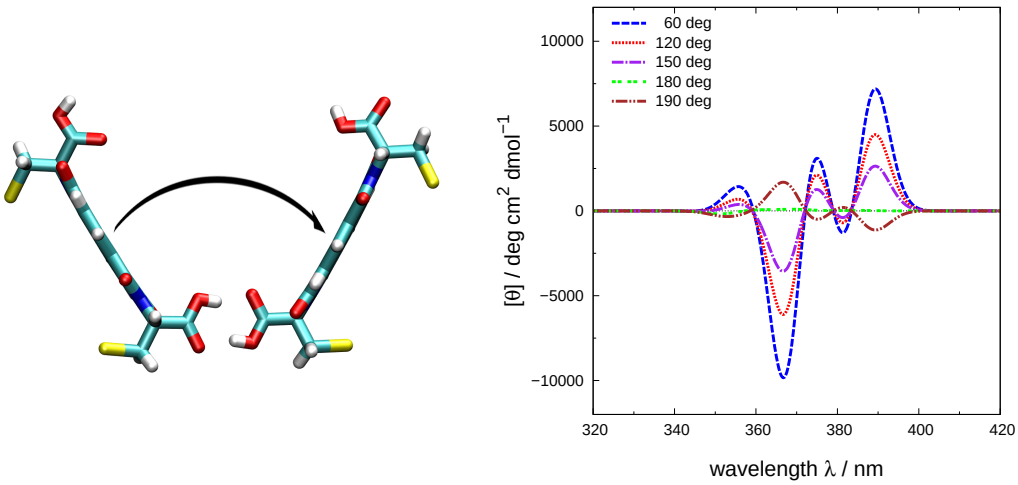



Figure 58: Dependence of the calculated CD spectrum on the angle between the NDI planes of two monomers.

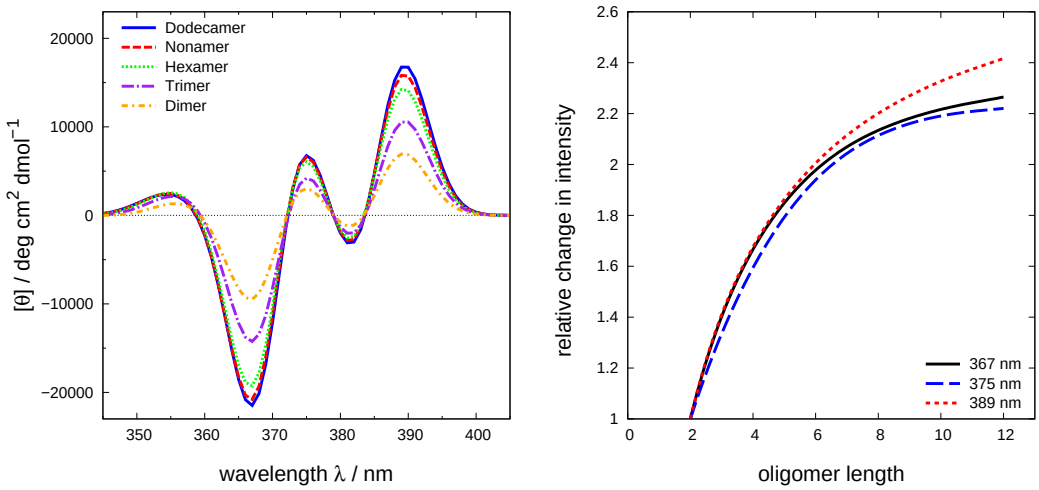



Figure 59: Left: Spectra calculated for oligomers of NDI. Right: Change in intensity of the three most intense bands in the calculated spectra depending on the length of the oligomer.

## 5.4 Conclusion

Using matrix method calculations, with parameters derived from *ab initio* calculations, it was possible to reproduce five of six bands in the experimental spectrum between 320 and 420 nm. The parameters can be used on the DichroCalc web server at http://comp.chem.nottingham.ac.uk/dichrocalc to perform CD and linear dichroism calculations.[89] Consideration of vibronic contributions is crucial and neither the vertical nor the adiabatic excitation energies are sufficient without taking Franck-Condon transitions into account. The examined region in the spectrum is dominated by transitions to the $1^1B_{2u}$ state with weak contributions from the $1^1B_{3u}$ transitions. The calculated spectrum shows a red-shift of approximately 9 nm. This may be due to the representation of each transition dipole moment by only two charges and further investigations are in progress. The agreement of experimental CD and calculation is evidence that the solid state and solution phase structures are equivalent and establishes the utility of solution phase CD to characterize these structures. The calculations show a clear dependence of the spectrum on the oligomer length, which may be useful for estimating the nanotube length once a reference spectrum for a tube of known length could be measured.

Application to multi-coloured NDI chromophores with different core substituents would be possible but would require parametrization of the chromophore for substituents extending or changing the $\pi$ system. Sterzel *et al.*[196] have shown that aliphatic substituents on the nitrogens do not affect the low-energy states of NDI (hence the transfer of these results to the studied structures). However, Würther *et al.*[194] have studied NDIs with electron-donating substituents on the naphthalene core and observed bathochromic shifts of up to 200 nm in the absorbance spectrum. Such structures would require a reparametrization to account for the influence of the substituents.

# Chapter 6

# Online Calculation using DichroCalc

## 6.1 Introduction

The extensive use of the internet nowadays has given rise to a large number of online services for protein analyses, e.g., for the interpretation of mass spectra,[211,212] investigations into protein similarity[213] and the prediction of secondary structure.[214] Sensitivity to secondary structure and ease of use have made circular dichroism (CD) spectroscopy a popular method among protein scientists. A spectrum can be quickly recorded and allows the main secondary structure type and the approximate helical content to be identified.[34,127] Specifically concerning CD, the Protein Circular Dichroism Data Bank (PCDDB) aims to build a freely available database of CD spectra,[215] and the web interface DICHROWEB provides analyses of CD data using several open source algorithms.[216] DICHROWEB currently supports the programs SELCON3,[217,218] CONTINILL,[219,220] CDSSTR,[221–223] VARSLC,[221–223] and K2D,[224] and combines seven different reference databases to choose from.

Using these tools, it calculates the secondary structure from the experimental spectrum, a goodness-of-fit parameter for the analyses, and several overviews of calculation and experiments.

The theory behind the origin of CD is well understood and much effort has been devoted to the theoretical prediction of the spectra based on the 3D structure of the proteins. The most successful framework is the matrix method,[33,46] for which parameters have been derived from *ab initio* calculations. These parameters include the peptide chromophore[71] in the far-UV, charge-transfer[100] between neighbouring peptide groups in the deep-UV and aromatic side chain chromophores[81] with transitions in the near-UV. For CD, the parameters have been tested on two different sets of proteins, showing an almost quantitative accuracy at 222 nm.[93,127] The Spearman rank correlation coefficient[87] between the calculated and experimental intensity at 222 nm is 0.9. The calculations can, therefore, be used to compare the theoretical spectrum of an assumed structure with experiment in order to aid the elucidation of the actual conformation.

Although it is much more common for researchers to wish to extract structural information from a CD spectrum, it is sometimes of interest to generate a CD spectrum from a structure. If it is suspected that a protein contains a novel structure, which the structure-fitting databases do not include, then this approach is the only option. It can be particularly helpful for conformationally flexible structures, as it is possible to determine an ensemble average of a series of spectra, computed from structures generated by molecular dynamics simulations. Calculations of CD using the matrix method have been reported for decades, but accessibility to the necessary programs and parameters was limited. LD is closely related to CD and provides information about how proteins align in the sample or how agents dock to fibrous structures.[51,225]

The analysis of experimental LD spectra is a challenge due to overlapping transitions, which can lead to disappearing bands. The calculation of the LD spectra, therefore, can aid assignment. To make such calculations available to non-theoreticians, DichroCalc was developed, a freely accessible web interface implementing the matrix method.

## 6.2 Approach

The concept of using the matrix method to help correlate experimental CD spectra with protein structure is very attractive. However, until the development of DichroCalc, this was largely the preserve of the theoreticians, who seldom had access to experimental data. Therefore, DichroCalc was designed to enable non-theoreticians with little theoretical background to benefit from this approach. The theoretical framework has been reviewed in detail,[127] but this knowledge is not required to make use of the calculations. The interface expects PDB files as input. They can be uploaded one by one, compressed in one or more archives or by giving their PDB codes. The parameters available on DichroCalc for the backbone are the peptide chromophore and backbone charge-transfer transitions between neighbouring peptide groups. A choice is provided between the semi-empirical and the *ab initio* peptide parameters, the latter allow additionally the consideration of the $\pi_b \rightarrow \pi^*$ and the $n' \rightarrow \pi^*$ transitions. To include aromatic side chain transitions, the chromophores of phenylalanine, tyrosine and tryptophan can be added. For consideration of aliphatic side chains, the chromophores of asparagine, aspartic acid, glutamine, glutamic acid, and peptide bonds between lysine and aspartic acid are available. All mentioned parameters can be chosen cumulatively.

Optionally, a full CD analysis performs calculations for all possible combinations of the selected parameters and the routine generates a report with comparison plots of the different spectra. For spectra of multiple conformations the average CD spectrum can be requested.

CD or LD calculations can be selected individually and each calculated spectrum is provided as $xy$-data for use in spreadsheet applications and also plotted in postscript format (Figure 60). In addition to the chosen range of calculation and plot the user may define a secondary range to zoom in on a special region in a second plot. This is useful, for example, to analyze the contributions of side chain transitions since the intensity in the near-UV is orders of magnitude smaller than in the far-UV.

A complication with LD as compared to CD is that the orientation of the protein in the coordinate system also has to be considered. An LD experiment is usually carried out to determine the orientation of a protein; theoretical calculations can be of help in estimating the expected sign of bands in the LD spectrum for a given orientation. Therefore, DichroCalc offers the option to calculate a series of spectra for different orientations from one PDB input file. The results are presented in a report that also contains the experimental spectrum in all plots for each orientation. An example report of an orientation search is given in Appendix D.

The results of the calculations are sent to the user via email as an archive or as a download link if the archive size prohibits email. In the first two years since the first beta release, over 100 users have submitted more than 1000 jobs calculating spectra for over 54,000 PDB files.

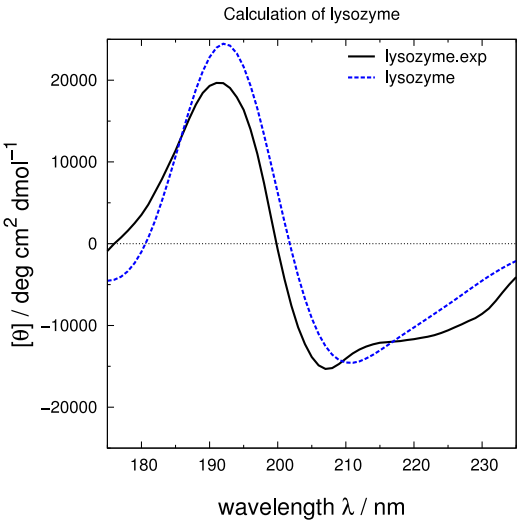
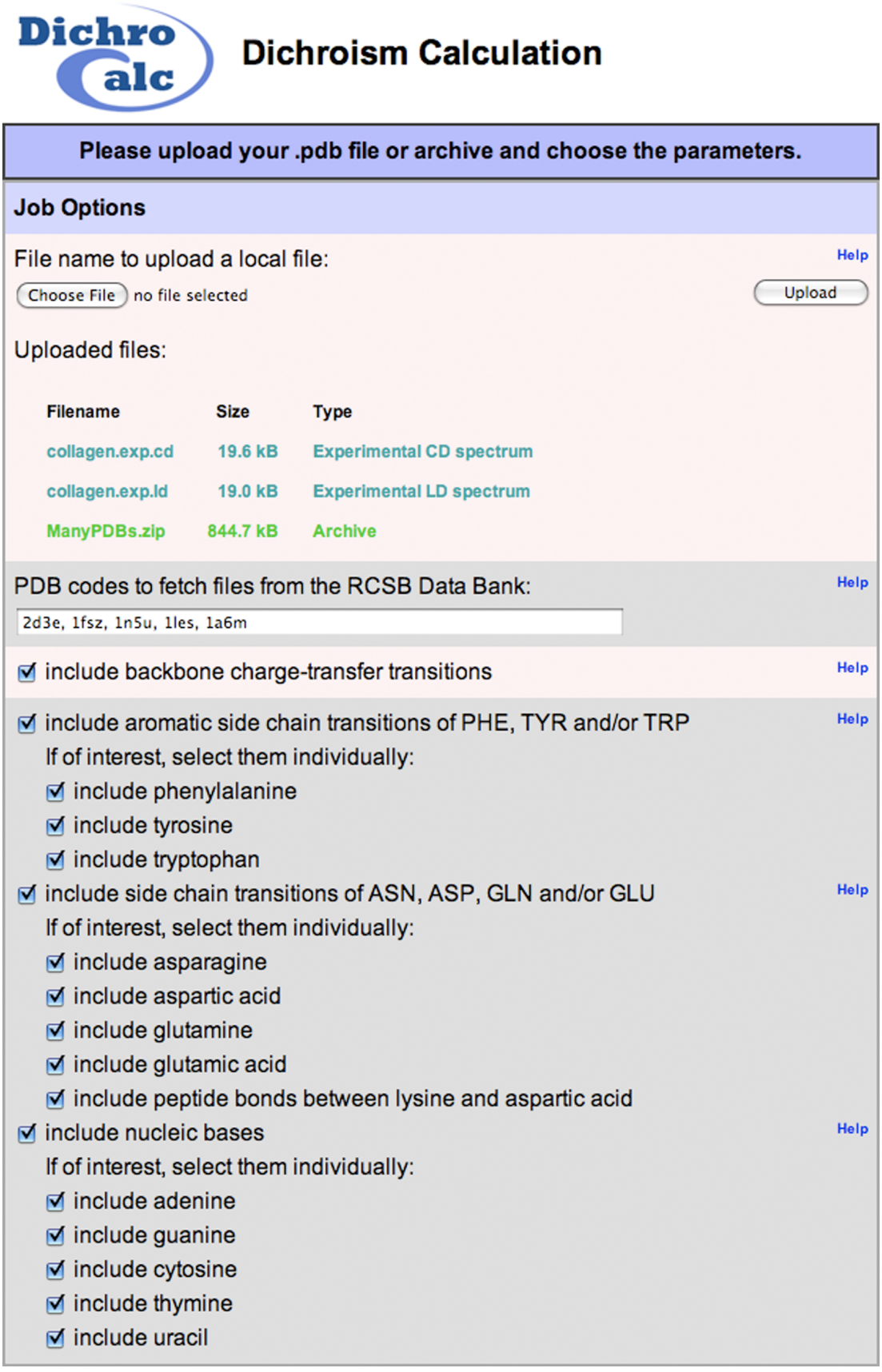
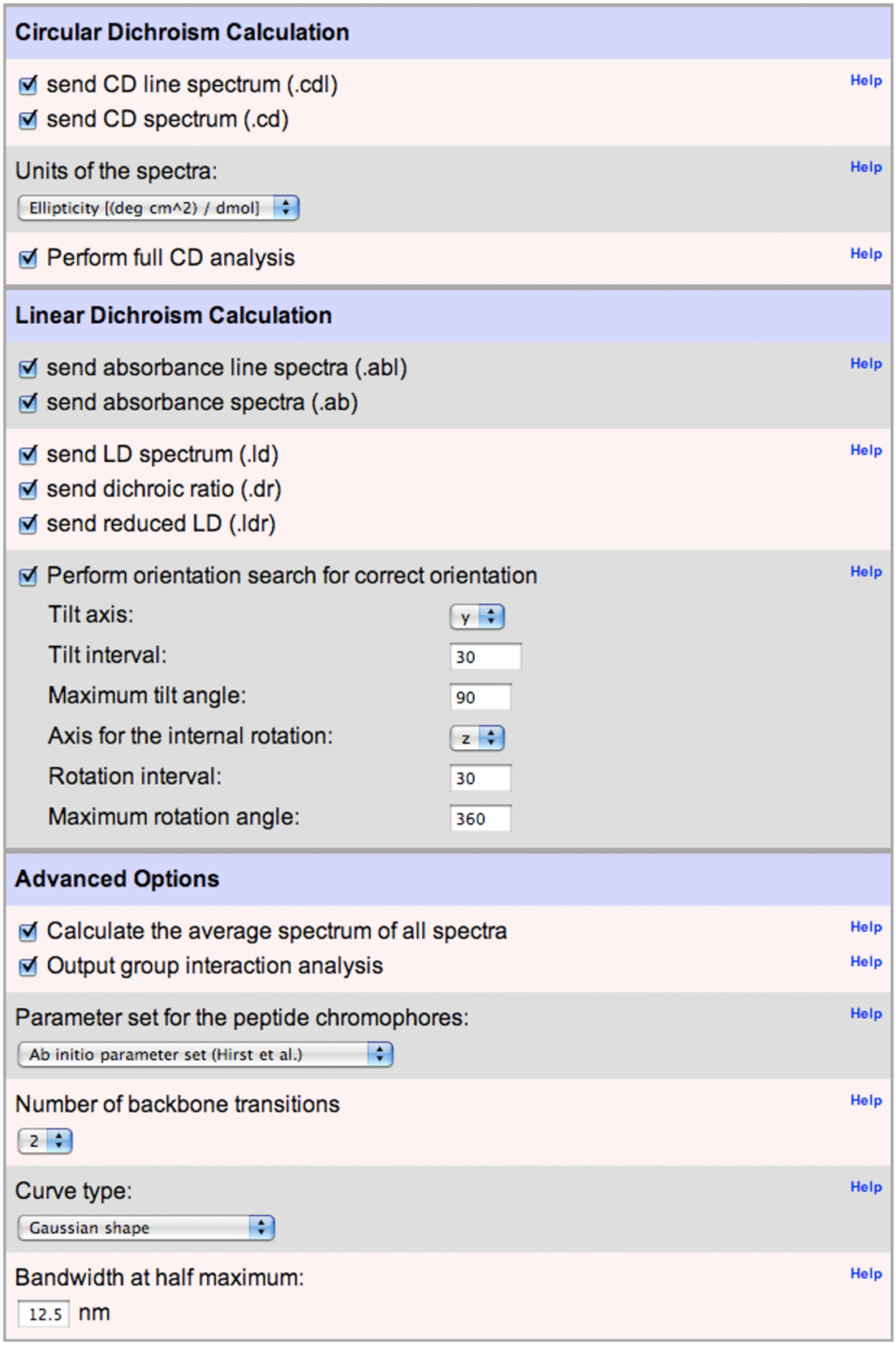


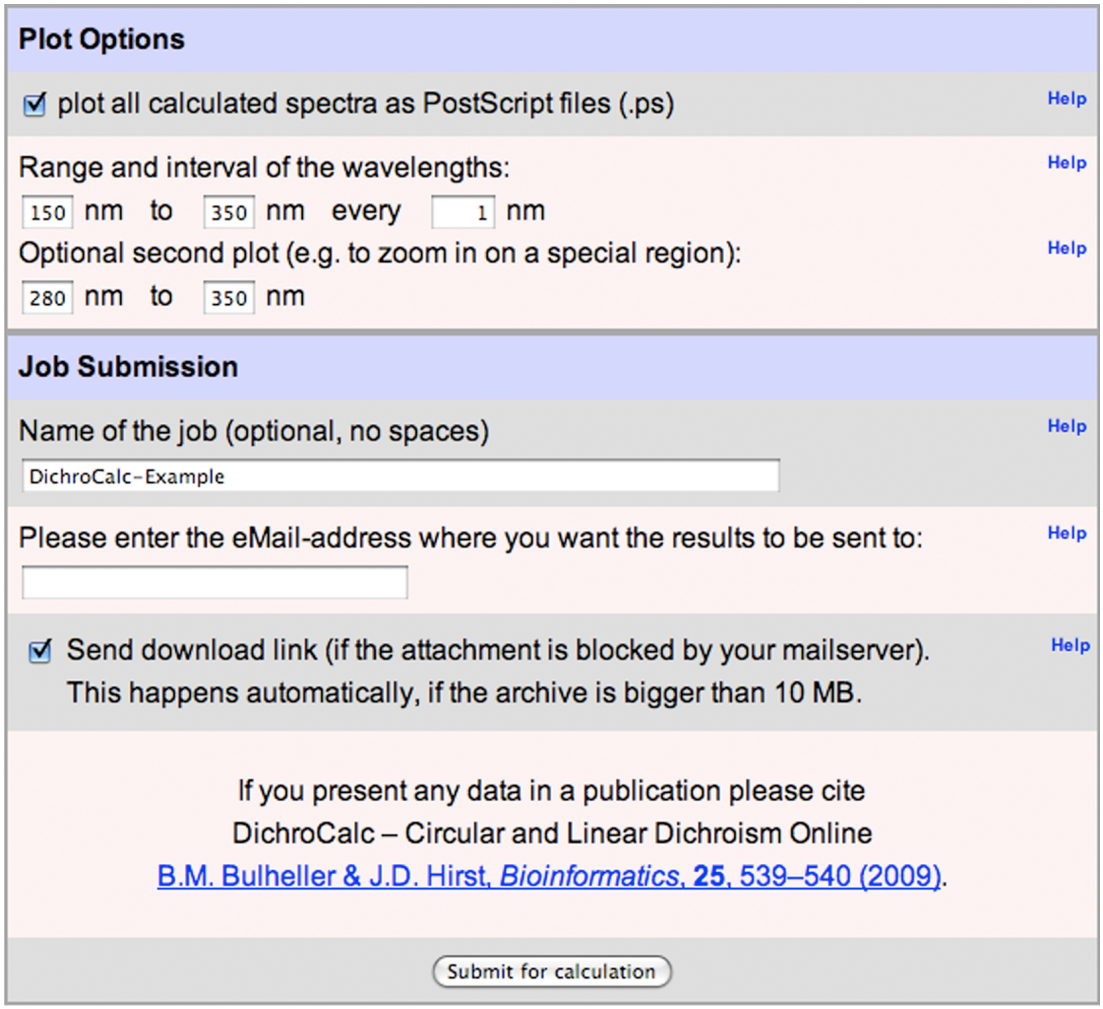

Figure 60: Comparison of a calculated CD spectrum with experiment as returned to the user by DichroCalc.

## 6.3 Implementation

DichroCalc is written in Perl with a CGI interface, whilst the computationally expensive dichroism calculations are carried out in FORTRAN.[89] Choosing a web interface approach over a stand-alone program was important both for usability reasons and to make the calculations available for as many users as possible. The combination of FORTRAN and Perl components results in complicated installation routines, which would have needed to be prepared for the three major operating systems (including different versions of these). The interface ensures cross-platform compatibility, provides instant access without installation for new users, and updates can be administered in one place. All major browsers are supported. The web interface can be reached at http://comp.chem.nottingham.ac.uk/dichrocalc . The following pages show DichroCalc with all possible options selected and sub-menus expanded (Pages 139 – 141).


Dichro Calc
Dichroism Calculation
Please upload your .pdb file or archive and choose the parameters.
Job Options
File name to upload a local file:
Help
Choose File
no file selected
Upload
Uploaded files:
Filename
Size
Type
collagen.exp.cd
19.6 kB
Experimental CD spectrum
collagen.exp.ld
19.0 kB
Experimental LD spectrum
ManyPDBs.zip
844.7 kB
Archive
PDB codes to fetch files from the RCSB Data Bank:
Help
2d3e, 1fsz, 1n5u, 1les, 1a6m
include backbone charge-transfer transitions
Help
include aromatic side chain transitions of PHE, TYR and/or TRP
Help
If of interest, select them individually:
include phenylalanine
include tyrosine
include tryptophan
include side chain transitions of ASN, ASP, GLN and/or GLU
Help
If of interest, select them individually:
include asparagine
include aspartic acid
include glutamine
include glutamic acid
include peptide bonds between lysine and aspartic acid
include nucleic bases
Help
If of interest, select them individually:
include adenine
include guanine
include cytosine
include thymine
include uracil


Circular Dichroism Calculation
send CD line spectrum (.cdl)
send CD spectrum (.cd)
Help
Units of the spectra:
Ellipticity [(deg cm^2) / dmol]
Help
Perform full CD analysis
Help
Linear Dichroism Calculation
send absorbance line spectra (.abl)
send absorbance spectra (.ab)
Help
send LD spectrum (.ld)
send dichroic ratio (.dr)
send reduced LD (.ldr)
Help
Perform orientation search for correct orientation
Help
Tilt axis:
y
Tilt interval:
30
Maximum tilt angle:
90
Axis for the internal rotation:
z
Rotation interval:
30
Maximum rotation angle:
360
Advanced Options
Calculate the average spectrum of all spectra
Help
Output group interaction analysis
Help
Parameter set for the peptide chromophores:
Help
Ab initio parameter set (Hirst et al.)
Number of backbone transitions
Help
2
Curve type:
Help
Gaussian shape
Bandwidth at half maximum:
Help
12.5
nm




### Interface Statistics

| | |
|---|---|
| Single components: | 24 |
| Lines of code: | ∼ 26,000 |

### Usage Statistics (July 2009)

| | |
|---|---|
| Users: | 137 |
| Jobs: | 1016 |
| Files: | 54,917 |
| Jobs using CD: | 1012 |
| Jobs using LD: | 326 |

## 6.4 The PDB Parser

Since the PDB format is forgiving and various augmented versions exist, it was necessary to deal with the files in a robust way and allow for several deviations from the current definition. This was important because users are able to upload their own files, probably created by different programs that do not necessarily obey the PDB standards. Since the code foundation of DichroCalc is Perl-based, the first choice was BioPerl,[226] a powerful package of Perl tools for genomics and bioinformatics. However, changes to the package would have been restrictive for subsequent updates to new versions of the suite. Therefore, a tailor-made parser for PDB files was created instead.

The Perl library ParsePDB.pm provides an object-oriented interface to the fully-parsed content of the PDB file. The library was designed to access all atoms or an arbitrary subset of atoms in the protein, depending on the provided keywords to filter the returned content. In a single command the PDB can be filtered to retrieve atoms in a certain model, chain or residue and/or possessing a particular atom type (for example, "CA") or residue type (for example, "ALA"). Models, chains and residues can be renumbered while retrieving the atoms. The content can be retrieved either as a single string per record as it is read from the PDB file or readily parsed so that each field of a record, such as the atom number or Cartesian coordinates, can be directly accessed. This also applies to $\phi$ and $\psi$ backbone dihedral angles, which may be calculated and are then added to each parsed residue for direct access. For the retrieved content, regardless of whether it is a complete model, chain, residue or filtered parts of such units, the mass as well as the centre of mass can be returned.

The retrieved filtered and renumbered data can be written in PDB or FASTA format and an additional routine allows PDB files to be generated for use with the CHARMM molecular dynamics program.[227] Inserted residues can be handled as well as alternative atom locations and the parser provides functions to access these individually or remove them if necessary, thereby detecting and fixing emerging chain breaks. The PDB parser has been developed for the DichroCalc and the ProCKSi[213] web interface for secondary structure determination.

The Perl library can be downloaded at
http://comp.chem.nottingham.ac.uk/parsepdb .

## 6.5 Conclusion

DichroCalc is the first web interface that makes CD and LD calculations available for use by non-theoreticians, thus considerably improving the access to such calculations. PDB files can be processed by the hundreds, and calculated spectra are sent to the user directly.

The development of the interface spawned the Perl library ParsePDB.pm, which provides quick access to the parsed content of a PDB file in just one compact library.

# Chapter 7

# Conclusions

Matrix method calculations, based on parameters derived from *ab initio* approaches, are able to reproduce features observed in experimental CD spectra. The CD spectra of highly $\alpha$-helical proteins can be calculated almost quantitatively. Since the helix is the most common secondary structure type, the quantitative understanding of its contribution to the CD is a substantial achievement. The Spearman rank correlation coefficient between the experimental and calculated intensities for a diverse set of proteins is 0.92 at 222 nm. Parameters of aromatic side chains reproduce features in the near-UV and the $\pi_b \rightarrow \pi^*$ and $n' \rightarrow \pi^*$ amide transitions are parametrized and show bands in the deep-UV. It was found that consideration of only the four nearest neighbours does not worsen the result of the CD calculations. This finding may be used in the future to facilitate the calculations of significantly larger proteins by reducing the Hamiltonian to a block diagonal matrix. The effects of random coil and particularly $\beta$-II-structures on the observed CD spectra have to be better understood to achieve the level of agreement that can be reached for $\alpha$-helical proteins.

It is generally the case that, apart from the $\beta$-II-class proteins, the *ab initio* set of parameters tends to overestimate CD intensities, especially in the region around 190 nm.

The SP175 set, containing 71 proteins with diverse secondary structures, has made it possible to assess the influence of charge-transfer transitions in the vacuum-UV, a region that cannot be reached with conventional spectrometers. If only local excitations are considered, there is almost no correlation between the calculated and experimental intensity at higher energy; the Spearman rank correlation coefficient, $r$, is 0.12 at 175 nm. However, $r$ rises to 0.73, if charge-transfer transitions are taken into account and to 0.79 if, in addition, the side chain chromophores are considered. 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 two of the 19 charge-transfer conformations, namely the main $\alpha$-helical and $\beta$-strand types. 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 is rather caused by interactions among all four types. These results show that the inclusion of charge-transfer enhances the theoretical CD calculations in the vacuum-UV region. Combined with the computational analysis of the CD spectrum, the spectrum calculated from the 3D structure of the protein could provide new insights into protein folding.

The significant progress that has been made in calculating CD spectra over the last few years has been transferred over to LD spectroscopy. Protein LD data are becoming more available and it has become clear that they contain plenty of structural information. However, due to overlapping transitions, it is often impossible to extract this information without understanding the spectroscopy of the system. The matrix method is directly applicable to LD data interpretation and, as LD depends only on the electric moments of the transition, spectral shapes are easier to calculate. The orientations of five fibrous proteins, comprising the main secondary structure types, have been correctly reproduced by the calculations. A method has been developed to calculate the LD spectra of the protein in all possible orientations for direct comparison with experiment.

A mechanism has been implemented to create simplified matrix method parameters without performing full-scale *ab initio* calculations on the chromophore. It requires only previously published dipole moments and the excitation energies of chromophores. The calculations on DNA and RNA examples were able to reproduce experimental features and identify correctly the conformation of the polynucleotide. This was shown for parameters derived from two different theoretical approaches. For the naphthalenediimide chromophore, five of six bands in the experimental spectrum between 320 and 420 nm could be reproduced. Consideration of vibronic contributions is crucial and neither the vertical nor the adiabatic excitation energies are sufficient without taking Franck-Condon transitions into account. The calculations show a clear dependence of the spectrum on the oligomer length, which may be useful for estimating the nanotube size once a reference spectrum for a tube of known length could be measured.

The web interface DichroCalc offers CD and LD calculations online and has proven to be robust and stable. Several thousands of spectra have been calculated over the past years by dozens of users. With the addition of LD calculations, the interface will hopefully gain additional attention due to the increasing use of LD in current research. DichroCalc is the first web interface that makes CD and LD calculations available for the use of non-theoreticians, thus considerably improving the access to such calculations. PDB files can be processed by the hundreds and calculated spectra are sent to the user directly. The development of the interface spawned the Perl library ParsePDB.pm, which provides quick access to the parsed content of a PDB file in just one compact library.

# Appendix

# Appendix A

# Bibliography


[1] F.J.D. Arago, Mémoire sur une modification remarquable qu'éprouvent les rayons lumineux dans leur passage à travers certains corps diaphanes, et sur quelques autres nouveaux phénomènes d'optique (Note on a Remarkable Phenomenon Shown by Light Rays on Their Passage Through Certain Transparent Materials, and on Other New Optical Phenomena), *Mem. Sci. Math. Phys. Inst.*, **1811**, *12*, 93–134.

[2] J.B. Biot, Mémoire sur un nouveau genre d'oscillation que les molécules de lumière éprouvent en traversant certains cristaux (Note on a New Type of Oscillation that Molecules of Light Show on Traversing Certain Crystals), *Mem. Sci. Math. Phys. Inst.*, **1812**, *13*, 1–372.

[3] J.B. Biot, Phénomènes de polarisation successive, observés dans des fluides homogènes, *Bull. Soc. Phil.*, **1815**, 190–192.

[4] G.D. Fasman, *Circular Dichroism and the Conformational Analysis of Biomolecules*, Springer, Berlin, **1996**.

[5] L. Pasteur, Mémoire sur la relation qui peut exister entre la forme cristalline et la composition chimique, et sur la cause de la polarisa-

tion rotatoire. (Note on the Relationship of Crystalline Form to Chemical Composition, and on the Cause of Rotatory Polarization), *C. R. Acad. Sci.*, **1848**, *26*, 535–538.

[6] J.A. LeBel, On the Relations Which Exist Between the Atomic Formulas of Organic Compounds and the Rotatory Power of Their Solutions, *Bull. Soc. Chim. Fr.*, **1874**, *22*, 337–347.

[7] J.H. van't Hoff, *La Chimie dans L'Espace*, Bazendijk, Rotterdam, **1875**.

[8] G. Snatzke, Circular Dichroism and Optical Rotatory Dispersion. Principles and Application to the Investigation of the Stereochemistry of Natural Products, *Angew. Chem. Int. Ed.*, **1968**, *7*, 14–25.

[9] N. Berova, K. Nakanishi, and R.W. Woody, *Circular Dichroism: Principles and Applications*, Wiley-VCH, New York, 2nd edition, **2000**.

[10] M. Born, The Natural Optical Activity of Liquids and Gases, *Physikal. Z.*, **1915**, *16*, 251–258.

[11] W. Kuhn, Quantitative Verhältnisse und Beziehungen bei der Natürlichen Optischen Aktivität [Quantitative Ratio and Connections in Natural Optical Activity], *Z. Phys. Chem. B*, **1929**, *4*, 14–36.

[12] L. Rosenfeld, Quantenmechanische Theorie der natürlichen optischen Aktivität von Flüssigkeiten und Gasen [Quantum Mechanical Theory of the Optical Activity of Liquids and Gases], *Z. Phys.*, **1928**, *52*, 161–174.

[13] D.D. Fitts and J.G. Kirkwood, The Optical Rotatory Power of Helical Molecules, *Proc. Natl. Acad. Sci. U. S. A.*, **1956**, *42*, 33–36.

[14] W. Moffitt, Optical Rotatory Dispersion of Helical Polymers, *J. Chem. Phys.*, **1956**, *25*, 467–478.

[15] W. Moffitt, The Optical Rotatory Dispersion of Simple Polypeptides (2), *Proc. Natl. Acad. Sci. U. S. A.*, **1956**, *42*, 736–746.

[16] A.S. Davydov, *Theory of Molecular Excitations*, Springer, Berlin, **1971**.

[17] P. Urnes and P. Doty, Optical Rotation and the Conformation of Polypeptides and Proteins, *Adv. Protein Chem.*, **1961**, *16*, 401–544.

[18] G. Holzwarth and P. Doty, The Ultraviolet Circular Dichroism of Polypeptides, *J. Am. Chem. Soc.*, **1965**, *87*, 218–228.

[19] F. Furche, R. Ahlrichs, C. Wachsmann, E. Weber, A. Sobanski, F. Vögtle, and S. Grimme, Circular Dichroism of Helicenes Investigated by Time-Dependent Density Functional Theory, *J. Am. Chem. Soc.*, **2000**, *122*, 1717–1724.

[20] K.M. Specht, J. Nam, D.M. Ho, N. Berova, R.K. Kondru, D.N. Beratan, P. Wipf, R.A. Pascal, and D. Kahne, Determining Absolute Configuration in Flexible Molecules: A Case Study, *J. Am. Chem. Soc.*, **2001**, *123*, 8961–8966.

[21] M. Schreiber, R. Vahrenhorst, V. Buss, and M.P. Fülscher, *Ab Initio* (CASPT2) Excited State Calculations, Including Circular Dichroism, of Helically Twisted Cyanine Dyes, *Chirality*, **2001**, *13*, 571–576.

[22] P. Butz, G.E. Tranter, and J.P. Simons, Molecular Conformation in the Gas Phase and in Solution, *PhysChemComm*, **2002**, *5*, 91–93.

[23] K. Tanaka, Y. Itagaki, M. Satake, H. Naoki, T. Yasumoto, K. Nakanishi, and N. Berova, Three Challenges Toward the Assignment of Absolute Configuration of Gymnocin-B, *J. Am. Chem. Soc.*, **2005**, *127*, 9561–9570.

[24] B. Clark, R.J. Capon, E. Lacey, S. Tennant, J.H. Gill, B.M. Bulheller, and G. Bringmann, Gymnoascolides A-C: Aromatic Butenolides from an

Australian Isolate of the Soil Ascomycete *Gymnoascus reessii*, *J. Nat. Prod.*, **2005**, *68*, 1226–1230.

[25] H. DeVoe, Optical Properties of Molecular Aggregates. 1. Classical Model of Electronic Absorption and Refraction, *J. Chem. Phys.*, **1964**, *41*, 393–400.

[26] H. DeVoe, Optical Properties of Molecular Aggregates. 2. Classical Theory of the Refraction, Absorption, and Optical Activity of Solutions and Crystals, *J. Chem. Phys.*, **1965**, *43*, 3199–3208.

[27] J. Applequist, A Full Polarizability Treatment of the $\pi\pi^*$ Absorption and Circular Dichroic Spectra of $\alpha$-Helical Polypeptides, *J. Chem. Phys.*, **1979**, *71*, 4332–4338.

[28] J. Applequist, K.R. Sundberg, M.L. Olson, and L.C. Weiss, A Normal Mode Treatment of Optical Properties of a Classical Coupled Dipole Oscillator System with Lorentzian Band Shapes, *J. Chem. Phys.*, **1979**, *70*, 1240–1246.

[29] B.K. Sathyanarayana and J. Applequist, Theoretical $\pi\pi^*$ Absorption and Circular Dichroic Spectra of Cyclic Dipeptides, *Int. J. Pept. Protein Res.*, **1985**, *26*, 518–527.

[30] K.A. Bode and J. Applequist, Improved Theoretical $\pi\pi^*$ Absorption and Circular Dichroic Spectra of Helical Polypeptides Using New Polarizabilities of Atoms and NC'O Chromophores, *J. Phys. Chem.*, **1996**, *100*, 17825–17834.

[31] K.L. Carlson, S.L. Lowe, M.R. Hoffmann, and K.A. Thomasson, Theoretical UV Circular Dichroism of Cyclo(L-Proline-L-Proline), *J. Phys. Chem. A*, **2006**, *110*, 1925–1933.

[32] S.L. Lowe, R.R. Pandey, J. Czlapinski, G. Kie-Adams, M.R. Hoffmann, K.A. Thomasson, and K.S. Pierce, Dipole Interaction Model Predicted $\pi\pi^*$ Circular Dichroism of Cyclo(L-Pro)$_3$ Using Structures Created by Semi-Empirical, *ab initio*, and Molecular Mechanics Methods, *J. Pept. Res.*, **2003**, *61*, 189–201.

[33] P.M. Bayley, E.B. Nielsen, and J.A. Schellman, The Rotatory Properties of Molecules Containing Two Peptide Groups, *J. Phys. Chem.*, **1969**, *73*, 228–243.

[34] A. Rodger and B. Nordén, *Circular Dichroism and Linear Dichroism*, Oxford University Press, Oxford, 1st edition, **1997**.

[35] B.A. Wallace, Conformational Changes by Synchrotron Radiation Circular Dichroism Spectroscopy, *Nature Struct. Biol.*, **2000**, *7*, 708–709.

[36] J.C. Sutherland, E.J. Desmond, and P.Z. Takacs, Versatile Spectrometer for Experiments Using Synchrotron Radiation at Wavelengths Greater Than 100 nm, *Nuc. Inst. Meth.*, **1980**, *172*, 195–199.

[37] B.A. Wallace, Synchrotron Radiation Circular Dichroism Spectroscopy as a Tool for Investigating Protein Structures, *J. Synchrotron Radiat.*, **2000**, *7*, 289–295.

[38] R.W. Janes, Bioinformatics Analyses of Circular Dichroism Protein Reference Databases, *Bioinformatics*, **2005**, *21*, 4230–4238.

[39] J. Šebek, B. Gyurcsik, J. Šebestík, Z. Kejík, L. Bednárová, and P. Bouř, Interpretation of Synchrotron Radiation Circular Dichroism Spectra of Anionic, Cationic, and Zwitterionic Dialanine Forms, *J. Phys. Chem. A*, **2007**, *111*, 2750–2760.

[40] J.G. Lees, A.J. Miles, F. Wien, and B.A. Wallace, A Reference Database for Circular Dichroism Spectroscopy Covering Fold and Secondary Structure Space, *Bioinformatics*, **2006**, *22*, 1955–1962.

[41] A. Toumadje, S.W. Alcorn, and W.C. Johnson, Jr., Extending CD Spectra of Proteins to 168 nm Improves the Analysis for Secondary Structures, *Anal. Biochem.*, **1992**, *200*, 321–331.

[42] B.A. Wallace and R.W. Janes, Synchrotron Radiation Circular Dichroism Spectroscopy of Proteins: Secondary Structure, Fold Recognition and Structural Genomics, *Curr. Opin. Chem. Biol.*, **2001**, *5*, 567–571.

[43] A.J. Miles and B.A. Wallace, Synchrotron Radiation Circular Dichroism Spectroscopy of Proteins and Applications in Structural and Functional Genomics, *Chem. Soc. Rev.*, **2006**, *35*, 39–51.

[44] I. Tinoco, Theoretical Aspects of Optical Activity. 2. Polymers, *Adv. Chem. Phys.*, **1962**, *4*, 113–160.

[45] J.G. Kirkwood, On the Theory of Optical Rotatory Power, *J. Chem. Phys.*, **1937**, *5*, 479–491.

[46] R.W. Woody and I. Tinoco, Optical Rotation of Oriented Helices. 3. Calculation of Rotatory Dispersion and Circular Dichroism of the $\alpha$- and $3_{10}$-Helix, *J. Chem. Phys.*, **1967**, *46*, 4927–4945.

[47] R.W. Woody, Improved Calculation of the $n\pi^*$ Rotational Strength in Polypeptides, *J. Chem. Phys.*, **1968**, *49*, 4797–4806.

[48] J.D. Hirst and N.A. Besley, Response to "Comment on 'Improving Protein Circular Dichroism Calculations in the Far-Ultraviolet Through Reparameterizing the Amide Chromophore' " [J. Chem. Phys. 111, 2844 (1999)], *J. Chem. Phys.*, **1999**, *111*, 2846–2847.

[49] N. Sreerama and R.W. Woody, Computation and Analysis of Protein Circular Dichroism Spectra, *Methods Enzymol.*, **2004**, *383*, 318–351.

[50] P.M. Bayley, The Analysis of Circular Dichroism of Biomolecules, *Prog. Biophys. Mol. Biol.*, **1973**, *27*, 1–76.

[51] A. Rodger, J. Rajendra, R. Marrington, M. Ardhammar, B. Nordén, J.D. Hirst, A.T.B. Gilbert, T.R. Dafforn, D.J. Halsall, C.A. Woolhead, C. Robinson, T.J.T. Pinheiro, J. Kazlauskaite, M. Seymour, N. Perez, and M.J. Hannon, Flow Oriented Linear Dichroism to Probe Protein Orientation in Membrane Environments, *Phys. Chem. Chem. Phys.*, **2002**, *4*, 4051–4057.

[52] S.M. Kelly, T.J. Jess, and N.C. Price, How to Study Proteins by Circular Dichroism, *Biochim. Biophys. Acta, Proteins Proteomics*, **2005**, *1751*, 119–139.

[53] L.A. Nafie, J.C. Cheng, and P.J. Stephens, Vibrational Circular Dichroism of 2,2,2-Trifluoro-1-Phenylethanol, *J. Am. Chem. Soc.*, **1975**, *97*, 3842–3843.

[54] L.A. Nafie and M. Diem, Optical Activity in Vibrational Transitions: Vibrational Circular Dichroism and Raman Optical Activity, *Acc. Chem. Res.*, **1979**, *12*, 296–302.

[55] T.A. Keiderling, Vibrational Circular Dichroism, *Appl. Spectrosc. Rev.*, **1981**, *17*, 189–226.

[56] L.D. Barron, *Molecular Light Scattering and Optical Activity*, Cambridge University Press, Cambridge, 1st edition, **1982**.

[57] G. Holzwarth, E.C. Hsu, H.S. Mosher, T.R. Faulkner, and A. Moscowitz, Infrared Circular Dichroism of Carbon-Hydrogen and Carbon-Deuterium Stretching Modes. Observations, *J. Am. Chem. Soc.*, **1974**, *96*, 251–252.

[58] M.T. Oakley, B.M. Bulheller, and J.D. Hirst, First-Principles Calculations of Protein Circular Dichroism in the Far-Ultraviolet and Beyond, *Chirality*, **2006**, *18*, 340–347.

[59] S. Brahms and J. Brahms, Determination of Protein Secondary Structure in Solution by Vacuum Ultraviolet Circular Dichroism, *J. Mol. Biol.*, **1980**, *138*, 149–178.

[60] C.R. Cantor and P.R. Schimmel, *Biophysical Chemistry*, W.H. Freeman, San Francisco, **1980**.

[61] W.C. Johnson, Jr., Protein Secondary Structure and Circular Dichroism: A Practical Guide, *Proteins: Struct., Funct., Genet.*, **1990**, *7*, 205–214.

[62] P. Wallimann, R.J. Kennedy, and D.S. Kemp, Large Circular Dichroism Ellipticities for *N*-Templated Helical Polypeptides Are Inconsistent with Currently Accepted Helicity Algorithms, *Angew. Chem. Int. Ed.*, **1999**, *38*, 1290–1292.

[63] J.T. Yang, C.S.C. Wu, and H.M. Martinez, Calculation of Protein Conformation from Circular Dichroism, *Methods Enzymol.*, **1986**, *130*, 208–269.

[64] P. Luo and R.L. Baldwin, Mechanism of Helix Induction by Trifluoroethanol: A Framework for Extrapolating the Helix-Forming Properties of Peptides from Trifluoroethanol/Water Mixtures Back to Water, *Biochemistry*, **1997**, *36*, 8413–8421.

[65] C.D. Andrew, S. Bhattacharjee, N. Kokkoni, J.D. Hirst, G.R. Jones, and A.J. Doig, Stabilizing Interactions between Aromatic and Basic Side Chains in $\alpha$-Helical Peptides and Proteins. Tyrosine Effects on Helix Circular Dichroism, *J. Am. Chem. Soc.*, **2002**, *124*, 12706–12714.

[66] Z. Dang and J.D. Hirst, Short Hydrogen Bonds, Circular Dichroism, and Over-Estimates of Peptide Helicity, *Angew. Chem. Int. Ed.*, **2001**, *40*, 3619–3621.

[67] S. Bhattacharjee, G. Tóth, S. Lovas, and J.D. Hirst, Influence of Tyrosine on the Electronic Circular Dichroism of Helical Peptides, *J. Phys. Chem. B*, **2003**, *107*, 8682–8688.

[68] K.W. Plaxco and C.M. Dobson, Time-Resolved Biophysical Methods in the Study of Protein Folding, *Curr. Opin. Struct. Biol.*, **1996**, *6*, 630–636.

[69] J.D. Hirst, S. Bhattacharjee, and A.V. Onufriev, Theoretical Studies of Time-Resolved Spectroscopy of Protein Folding, *Faraday Discuss.*, **2002**, *122*, 253–267.

[70] E.L. Eliel, S.H. Wilen, and L.N. Mander, *Stereochemistry of Organic Compounds*, Wiley, New York, **1994**.

[71] N.A. Besley and J.D. Hirst, Theoretical Studies toward Quantitative Protein Circular Dichroism Calculations, *J. Am. Chem. Soc.*, **1999**, *121*, 9636–9644.

[72] R.W. Woody and N. Sreerama, Comment on "Improving Protein Circular Dichroism Calculations in the Far-Ultraviolet Through Reparametrizing the Amide Chromophore" [J. Chem. Phys. 109, 782, (1998)], *J. Chem. Phys.*, **1999**, *111*, 2844–2845.

[73] G. Karlström, R. Lindh, P.Å. Malmqvist, B.O. Roos, U. Ryde, V. Veryazov, P.O. Widmark, M. Cossi, B. Schimmelpfennig, P. Neogrády, and L. Seijo, MOLCAS: A Program Package for Computational Chemistry, *Comp. Mat. Sci.*, **2003**, *28*, 222–239.

[74] H.J. Werner, P.J. Knowles, R. Lindh, F.R. Manby, M. Schütz, P. Celani, T. Korona, G. Rauhut, R.D. Amos, A. Bernhardsson, A. Berning, D.L.

Cooper, M.J.O. Deegan, A.J. Dobbyn, F. Eckert, C. Hampel, G. Hetzer, A.W. Lloyd, S.J. McNicholas, W. Meyer, M.E. Mura, A. Nicklass, P. Palmieri, R. Pitzer, U. Schumann, H. Stoll, A.J. Stone, R. Tarroni, and T. Thorsteinsson, Molpro – A Package of *Ab Initio* Programs, http://www.molpro.net, **2002**.

[75] A. Bernhardsson, R. Lindh, G. Karlström, and B.O. Roos, Direct Self-Consistent Reaction Field with Pauli Repulsion: Solvation Effects on Methylene Peroxide, *Chem. Phys. Lett.*, **1996**, *251*, 141–149.

[76] G. Karlström, A New Approach to the Modeling of Dielectric Media Effects in *Ab Initio* Quantum Chemical Calculations, *J. Phys. Chem.*, **1988**, *92*, 1315–1318.

[77] G. Karlström, Electronic Structure of $HF^-$ and $HCl^-$ in Condensed Phases Studied by a CASSCF Dielectric Cavity Model, *J. Phys. Chem.*, **1989**, *93*, 4952–4955.

[78] L. Serrano-Andrés, M.P. Fülscher, and G. Karlström, Solvent Effects on Electronic Spectra Studied by Multiconfigurational Perturbation Theory, *Int. J. Quantum Chem.*, **1997**, *65*, 167–181.

[79] J.D. Hirst, K. Colella, and A.T.B. Gilbert, Electronic Circular Dichroism of Proteins from First-Principles Calculations, *J. Phys. Chem. B*, **2003**, *107*, 11813–11819.

[80] Y.H. Chen, J.T. Yang, and H.M. Martinez, Determination of the Secondary Structures of Proteins by Circular Dichroism and Optical Rotatory Dispersion, *Biochemistry*, **1972**, *11*, 4120–4131.

[81] D.M. Rogers and J.D. Hirst, *Ab Initio* Study of Aromatic Side Chains of Amino Acids in Gas Phase and Solution, *J. Phys. Chem. A*, **2003**, *107*, 11191–11200.

[82] J.R. Platt, Classification of Spectra of Cata-Condensed Hydrocarbons, *J. Chem. Phys.*, **1949**, *17*, 484–495.

[83] Y. Yamamoto and J. Tanaka, Polarized Absorption Spectra of Crystals of Indole and Its Related Compounds, *Bull. Chem. Soc. Jpn.*, **1972**, *45*, 1362–1366.

[84] D.M. Rogers and J.D. Hirst, First-Principles Calculations of Protein Circular Dichroism in the Near Ultraviolet, *Biochemistry*, **2004**, *43*, 11092–11102.

[85] N. Sreerama and R.W. Woody, Structural Composition of $\beta$I- and $\beta$II-Proteins, *Protein Sci.*, **2003**, *12*, 384–388.

[86] N.A. Besley and J.D. Hirst, Hydrogen Bonding in Protein Circular Dichroism Calculations, *J. Mol. Struct. – THEOCHEM*, **2000**, *506*, 161–167.

[87] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 2nd edition, **1992**.

[88] R.A. Laskowski, M.W. MacArthur, D.S. Moss, and J.M. Thornton, PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures, *J. Appl. Crystallogr.*, **1993**, *26*, 283–291.

[89] B.M. Bulheller and J.D. Hirst, DichroCalc – Circular and Linear Dichroism Online, *Bioinformatics*, **2009**, *25*, 539–540.

[90] M.B. Robin, *Higher Excited States of Polyatomic Molecules*, Academic Press, New York, **1975**.

[91] N.A. Besley and J.D. Hirst, *Ab Initio* Study of the Effect of Solvation on the Electronic Spectra of Formamide and *N*-Methylacetamide, *J. Phys. Chem. A*, **1998**, *102*, 10791–10797.

[92] A.P. Demchenko, *Ultraviolet Spectroscopy of Proteins*, Springer, Berlin, **1986**.

[93] B.M. Bulheller, A.J. Miles, B.A. Wallace, and J.D. Hirst, Charge-Transfer Transitions in the Vacuum-Ultraviolet of Protein Circular Dichroism Spectra, *J. Phys. Chem. B*, **2008**, *112*, 1866–1874.

[94] L.W. Nesgaard, S.V. Hoffmann, C.B. Andersen, A. Malmendal, and D.E. Otzen, Characterization of Dry Globular Proteins and Protein Fibrils by Synchrotron Radiation Vacuum UV Circular Dichroism, *Biopolymers*, **2008**, *89*, 779–795.

[95] K. Matsuo, R. Yonehara, and K. Gekko, Secondary-Structure Analysis of Proteins by Vacuum-Ultraviolet Circular Dichroism Spectroscopy, *J. Biochem.*, **2004**, *135*, 405–411.

[96] K. Matsuo, R. Yonehara, and K. Gekko, Improved Estimation of the Secondary Structures of Proteins by Vacuum-Ultraviolet Circular Dichroism Spectroscopy, *J. Biochem.*, **2005**, *138*, 79–88.

[97] R. Weinkauf, P. Schanen, D. Yang, S. Soukara, and E.W. Schlag, Elementary Processes in Peptides: Electron Mobility and Dissociation in Peptide Cations in the Gas Phase, *J. Phys. Chem.*, **1995**, *99*, 11255–11265.

[98] W.G. McGimpsey, L. Chen, R. Carraway, and W.N. Samaniego, Singlet-Singlet and Triplet-Triplet Energy Transfer in Bichromophoric Peptides, *J. Phys. Chem. A*, **1999**, *103*, 6082–6090.

[99] L. Serrano-Andrés and M.P. Fülscher, Charge Transfer Transitions in Neutral and Ionic Polypeptides: A Theoretical Study, *J. Phys. Chem. B*, **2001**, *105*, 9323–9330.

[100] M.T. Oakley and J.D. Hirst, Charge-Transfer Transitions in Protein Circular Dichroism Calculations, *J. Am. Chem. Soc.*, **2006**, *128*, 12414–12415.

[101] E. Goldmann, S.A. Asher, and S. Mukamel, Electronic Excitations of Polyalanine; Test of the Independent Chromophore Approximation, *Phys. Chem. Chem. Phys.*, **2001**, *3*, 2893–2903.

[102] A.T.B. Gilbert and J.D. Hirst, Charge-Transfer Transitions in Protein Circular Dichroism Spectra, *J. Mol. Struct. – THEOCHEM*, **2004**, *675*, 53–60.

[103] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, The Protein Data Bank, *Nucleic Acids Res.*, **2000**, *28*, 235–242.

[104] W. Kabsch and C. Sander, Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features, *Biopolymers*, **1983**, *22*, 2577–2637.

[105] C.A. Orengo, A.D. Michie, S. Jones, D.T. Jones, M.B. Swindells, and J.M. Thornton, CATH – A Hierarchic Classification of Protein Domain Structures, *Structure*, **1997**, *5*, 1093–1108.

[106] F. Pearl, A.K. Todd, I. Sillitoe, M. Dibley, O. Redfern, T. Lewis, C. Bennett, R. Marsden, A. Grant, D. Lee, A. Akpor, M. Maibaum, A. Harrison, T. Dallman, G. Reeves, I. Diboun, S. Addou, S. Lise, C. Johnston, A. Sillero, J.M. Thornton, and C.A. Orengo, The CATH Domain Structure Database and Related Resources Gene3D and DHS Provide Comprehensive Domain Family Information for Genome Analysis, *Nucleic Acids Res.*, **2005**, *33*, D247–D251.

[107] P. Manavalan and W.C. Johnson, Jr., Sensitivity of Circular Dichroism to Protein Tertiary Structure Class, *Nature*, **1983**, *305*, 831–832.

[108] D.J. Barlow and J.M. Thornton, Helix Geometry in Proteins, *J. Mol. Biol.*, **1988**, *201*, 601–619.

[109] T.R. Dafforn and A. Rodger, Linear Dichroism of Biomolecules: Which Way is Up?, *Curr. Opin. Struct. Biol.*, **2004**, *14*, 541–546.

[110] T.C. Troxell and H.A. Scheraga, Electric Dichroism and Polymer Conformation. I. Theory of Optical Properties of Anisotropic Media, and Method of Measurement, *Macromolecules*, **1971**, *4*, 519–527.

[111] B. Nordén, Applications of Linear Dichroism Spectroscopy, *Appl. Spectrosc. Rev.*, **1978**, *14*, 157–248.

[112] A. Wada and S. Kozawa, Instrument for the Studies of Differential Flow Dichroism of Polymer Solutions, *J. Pol. Sci. A*, **1964**, *2*, 853–864.

[113] T.C. Troxell and H.A. Scheraga, Electric Dichroism and Polymer Conformation. II. Theory of Electric Dichroism, and Measurements on Poly (n-butyl isocyanate), *Macromolecules*, **1971**, *4*, 528–539.

[114] R. Cerf and H.A. Scheraga, Flow Birefringence in Solutions of Macromolecules, *Chem. Rev.*, **1952**, *51*, 185–261.

[115] J. Hofrichter and W.A. Eaton, Linear Dichroism of Biological Chromophores, *Annu. Rev. Biophys. Bioeng.*, **1976**, *5*, 511–560.

[116] A.L. von Muralt and J.T. Edsall, Studies in the Physical Chemistry of Muscle Globulin. IV. The Anisotropy of Myosin and Double Refraction of Flow, *J. Biol. Chem.*, **1930**, *89*, 351–386.

[117] J.T. Edsall and J.W. Mehl, The Effect of Denaturing Agents on Myosin. II. Viscosity and double Refraction of Flow, *J. Biol. Chem.*, **1940**, *133*, 409–429.

[118] T.R. Dafforn, J. Rajendra, D.J. Halsall, L.C. Serpell, and A. Rodger, Protein Fiber Linear Dichroism for Structure Determination and Kinetics in a Low-Volume, Low-Wavelength Couette Flow Cell, *Biophys. J.*, **2004**, *86*, 404–410.

[119] R. Adachi, K. Yamaguchi, H. Yagi, K. Sakurai, H. Naiki, and Y. Goto, Flow-Induced Alignment of Amyloid Protofilaments Revealed by Linear Dichroism, *J. Biol. Chem.*, **2007**, *282*, 8978–8983.

[120] K. Morimatsu and M. Takahashi, Structural Analysis of RecA Protein-DNA Complexes by Fluorescence-Detected Linear Dichroism: Absence of Structural Change of Filament for Pairing of Complementary DNA Strands, *Anal. Biochem.*, **2006**, *358*, 192–198.

[121] E.K. Esbjörner, C.E.B. Caesar, B. Albinsson, P. Lincoln, and B. Nordén, Tryptophan Orientation in Model Lipid Membranes, *Biochem. Biophys. Res. Commun.*, **2007**, *361*, 645–650.

[122] E. Small, R. Marrington, A. Rodger, D.J. Scott, K. Sloan, D. Roper, T.R. Dafforn, and S.G. Addinall, FtsZ Polymer-Bundling by the *Escherichia Coli* ZapA Orthologue, YgfE, Involves a Conformational Change in Bound GTP, *J. Mol. Biol.*, **2007**, *369*, 210–221.

[123] J. Rajendra, A. Damianoglou, M. Hicks, P.J. Booth, P.M. Rodger, and A. Rodger, Quantitation of Protein Orientation in Flow-Oriented Unilamellar Liposomes by Linear Dichroism, *Chem. Phys.*, **2006**, *326*, 210–220.

[124] A. Rodger, R. Marrington, M.A. Geeves, M. Hicks, L. de Alwis, D.J. Halsall, and T.R. Dafforn, Looking at Long Molecules in Solution: What Happens When They Are Subjected to Couette Flow?, *Phys. Chem. Chem. Phys.*, **2006**, *8*, 3161–3171.

[125] M.R. Hicks, A. Damianoglou, A. Rodger, and T.R. Dafforn, Folding and Membrane Insertion of the Pore-Forming Peptide Gramicidin Occur as a Concerted Process, *J. Mol. Biol.*, **2008**, *383*, 358–366.

[126] S.M. Ennaceur, M.R. Hicks, C.J. Pridmore, T.R. Dafforn, A. Rodger, and J.M. Sanderson, Peptide Adsorption to Lipid Bilayers: Slow Processes

Revealed by Linear Dichroism Spectroscopy, *Biophys. J.*, **2009**, *96*, 1399–1407.

[127] B.M. Bulheller, A. Rodger, and J.D. Hirst, Circular and Linear Dichroism of Proteins, *Phys. Chem. Chem. Phys.*, **2007**, *9*, 2020–2035.

[128] J.A. Schellman, Circular Dichroism and Optical Rotation, *Chem. Rev.*, **1975**, *75*, 323–331.

[129] M. Couette, Études sur le frottement des liquides, *Ann. Chim. Phys.*, **1890**, *6*, 433–510.

[130] A. Mallock, Determination of the Viscosity of Water, *Proc. R. Soc. London*, **1888**, *45*, 126–132.

[131] S. Higashi, M. Kasai, F. Oosawa, and A. Wada, Ultraviolet Dichroism of F-Actin Oriented by Flow, *J. Mol. Biol.*, **1963**, *7*, 421–430.

[132] A. Wada, Dichroic Spectra of Biopolymers Oriented by Flow, *Appl. Spectrosc. Rev.*, **1972**, *6*, 1–30.

[133] R. Marrington, E. Small, A. Rodger, T.R. Dafforn, and S.G. Addinall, FtsZ Fiber Bundling is Triggered by a Conformational Change in Bound GTP, *J. Biol. Chem.*, **2004**, *279*, 48821–48829.

[134] R. Marrington, M. Seymour, and A. Rodger, A New Method for Fibrous Protein Analysis Illustrated by Application to Tubulin Microtubule Polymerisation and Depolymerisation, *Chirality*, **2006**, *18*, 680–690.

[135] R. Marrington, T.R. Dafforn, D.J. Halsall, J.I. MacDonald, M. Hicks, and A. Rodger, Validation of New Microvolume Couette Flow Linear Dichroism Cells, *Analyst*, **2005**, *130*, 1608–1616.

[136] M. Ardhammar, N. Mikati, and B. Nordén, Chromophore Orientation in Liposome Membranes Probed with Flow Linear Dichroism, *J. Am. Chem. Soc.*, **1998**, *120*, 9957–9958.

[137] A. Rodger, J. Rajendra, R. Mortimer, T. Andrews, J.D. Hirst, A.T.B. Gilbert, R. Marrington, T.R. Dafforn, D.J. Halsall, M. Ardhammar, B. Nordén, C.A. Woolhead, C. Robinson, T.J.T. Pinheiro, J. Kazlauskaite, M. Seymour, N. Perez, M.J. Hannon, R.H. Templer, and R. Leatherbarrow, *Flow Oriented Linear Dichroism to Probe Protein Orientation in Membrane Environments*, 3–19, Biophysical Chemistry: Membranes and Proteins, Cambridge, **2002**.

[138] T.C. Troxell and H.A. Scheraga, Use of Electric Dichroism to Study Polymer Conformation, *Biochem. Biophys. Res. Commun.*, **1969**, *35*, 913–119.

[139] J.P. Cartailler and H. Luecke, X-Ray Crystallographic Analysis of Lipid-Protein Interactions in the Bacteriorhodopsin Purple Membrane, *Annu. Rev. Biophys. Biomol. Struct.*, **2003**, *32*, 285–310.

[140] W. Meijberg and P.J. Booth, The Activation Energy for Insertion of Transmembrane $\alpha$-helices is Dependent on Membrane Composition, *J. Mol. Biol.*, **2002**, *319*, 839–853.

[141] D. Oesterhelt and L. Schuhmann, Reconstitution of Bacteriorhodopsin, *FEBS Lett.*, **1974**, *44*, 262–265.

[142] B. Albinsson, M. Kubista, B. Nordén, and E.W. Thulstrup, Near-Ultraviolet Electronic Transitions of the Tryptophan Chromophore: Linear Dichroism, Fluorescence Anisotropy, and Magnetic Circular Dichroism Spectra of Some Indole Derivatives, *J. Phys. Chem.*, **1989**, *93*, 6646–6654.

[143] W. Humphrey, A. Dalke, and K. Schulten, VMD: Visual Molecular Dynamics, *J. Mol. Graph.*, **1996**, *14*, 33–38.

[144] M.J. Pandya, G.M. Spooner, M. Sunde, J.R. Thorpe, A. Rodger, and D.N. Woolfson, Sticky-End Assembly of a Designed Peptide Fiber Provides Insight into Protein Fibrillogenesis, *Biochemistry*, **2000**, *39*, 8728–8734.

[145] D. Papapostolou, A.M. Smith, E.D.T. Atkins, S.J. Oliver, M.G. Ryadnov, L.C. Serpell, and D.N. Woolfson, Engineering Nanoscale Order into a Designed Protein Fiber, *Proc. Natl. Acad. Sci.*, **2007**, *104*, 10853–10858.

[146] Y. Nitanai, S. Minakata, K. Maeda, N. Oda, and Y. Maéda, Crystal Structures of Tropomyosin: Flexible Coiled-Coil, *Adv. Exp. Med. Biol.*, **2007**, *592*, 137–151.

[147] G.N. Phillips, Jr., J.P. Fillers, and C. Cohen, Tropomyosin Crystal Structure and Muscle Regulation, *J. Mol. Biol.*, **1986**, *192*, 111–131.

[148] K.A. Bode and J. Applequist, Helix Bundles and Coiled Coils in $\alpha$-Spectrin and Tropomyosin: A Theoretical CD Study, *Biopolymers*, **1997**, *42*, 855–860.

[149] A.K.W. Leung, E. Lucile White, L.J. Ross, R.C. Reynolds, J.A. DeVito, and D.W. Borhani, Structure of *Mycobacterium tuberculosis* FtsZ Reveals Unexpected, G Protein-Like Conformational Switches, *J. Mol. Biol.*, **2004**, *342*, 953–970.

[150] E. Nogales, K.H. Downing, L.A. Amos, and J. Löwe, Tubulin and FtsZ Form a Distinct Family of GTPases, *Nature Struct. Biol.*, **1998**, *5*, 451–458.

[151] J. Löwe and L.A. Amos, Tubulin-Like Protofilaments in $Ca^{2+}$-Induced FtsZ Sheets, *EMBO J.*, **1999**, *18*, 2364–2371.

[152] C.L. Nesloney and J.W. Kelly, Progress Towards Understanding $\beta$-Sheet Structure, *Bioorg. Med. Chem.*, **1996**, *4*, 739–766.

[153] R.W. Woody, The Circular Dichroism of Oriented $\beta$ Sheets: Theoretical Predictions, *Tetrahedron: Asymmetry*, **1993**, *4*, 529–544.

[154] P. Westermark, Aspects on Human Amyloid Forms and Their Fibril Polypeptides, *FEBS J.*, **2005**, *272*, 5942–5949.

[155] P. Westermark, M.D. Benson, J.N. Buxbaum, A.S. Cohen, B. Frangione, S.I. Ikeda, C.L. Masters, G. Merlini, M.J. Saraiva, and J.D. Sipe, Amyloid: Toward Terminology Clarification. Report from the Nomenclature Committee of the International Society of Amyloidosis, *Amyloid*, **2005**, *12*, 1–4.

[156] J.Q. Trojanowski and V.M.Y. Lee, Parkinson's Disease and Related Synucleinopathies are a New Class of Nervous System Amyloidoses, *Neurotoxicology*, **2002**, *23*, 457–460.

[157] R. Nelson, M.R. Sawaya, M. Balbirnie, A.Ø. Madsen, C. Riekel, R. Grothe, and D. Eisenberg, Structure of the Cross-$\beta$ Spine of Amyloid-Like Fibrils, *Nature*, **2005**, *435*, 773–778.

[158] M.R. Sawaya, S. Sambashivan, R. Nelson, M.I. Ivanova, S.A. Sievers, M.I. Apostol, M.J. Thompson, M. Balbirnie, J.J.W. Wiltzius, H.T. McFarlane, A.Ø. Madsen, C. Riekel, and D. Eisenberg, Atomic Structures of Amyloid Cross-$\beta$ Spines Reveal Varied Steric Zippers, *Nature*, **2007**, *447*, 453–457.

[159] J.M. Chen, C.E. Kung, S.H. Feairheller, and E.M. Brown, An energetic evaluation of a "Smith" collagen microfibril model, *J. Protein Chem.*, **1991**, *10*, 535–552.

[160] R. Mandel and G. Holzwarth, Ultraviolet Circular Dichroism of Polyproline and Oriented Collagen, *Biopolymers*, **1973**, *12*, 655–674.

[161] R.W. Woody, The Exciton Model and the Circular Dichroism of Polypeptides, *Monatsh. Chem.*, **2005**, *136*, 347–366.

[162] R.W. Woody, Circular Dichroism Spectrum of Peptides in the Poly(Pro)II conformation, *J. Am. Chem. Soc.*, **2009**, *131*, 8234–8245.

[163] J. Gawroński, M. Brzostowska, K. Kacprzak, H. Kołbon, and P. Skowronek, Chirality of Aromatic Bis-Imides from Their Circular Dichroism Spectra, *Chirality*, **2000**, *12*, 263–268.

[164] J.M. Ferreira and R.D. Sheardy, Enthalpy of the B-to-Z Conformational Transition of a DNA Oligonucleotide Determined by Isothermal Titration Calorimetry, *Biophys. J.*, **2006**, *91*, 3383–3389.

[165] J.D. Petke, G.M. Maggiora, and R.E. Christoffersen, *Ab Initio* Configuration Interaction and Random Phase Approximation Calculations of the Excited Singlet and Triplet States of Adenine and Guanine, *J. Am. Chem. Soc.*, **1990**, *112*, 5452–5460.

[166] J.D. Petke, G.M. Maggiora, and R.E. Christoffersen, *Ab Initio* Configuration Interaction and Random Phase Approximation Calculations of the Excited Singlet and Triplet States of Uracil and Cytosine, *J. Phys. Chem.*, **1992**, *96*, 6992–7001.

[167] M.P. Fülscher, L. Serrano-Andrés, and B.O. Roos, A Theoretical Study of the Electronic Spectra of Adenine and Guanine, *J. Am. Chem. Soc.*, **1997**, *119*, 6168–6176.

[168] J. Lorentzon, M.P. Fülscher, and B.O. Roos, Theoretical Study of the Electronic Spectra of Uracil and Thymine, *J. Am. Chem. Soc.*, **1995**, *117*, 9265–9273.

[169] M.P. Fülscher and B.O. Roos, Theoretical Study of the Electronic Spectrum of Cytosine, *J. Am. Chem. Soc.*, **1995**, *117*, 2089–2095.

[170] J.D. Watson and F.H.C. Crick, Genetical Implications of the Structure of Deoxyribonucleic Acid, *Nature*, **1953**, *171*, 964–967.

[171] E. Passarge, *Color Atlas of Genetics*, Georg Thieme Verlag, Stuttgart, 3rd edition, **2007**.

[172] R.S. Verma, *Advances in Genome Biology*, Elsevier Science, New York, **1998**.

[173] J.H. Riazance, W.A. Baase, W.C. Johnson, Jr., K. Hall, P. Cruz, and I. Tinoco, Evidence for Z-form RNA by Vacuum UV Circular Dichroism, *Nucleic Acids Res.*, **1985**, *13*, 4983–4989.

[174] F.H.C. Crick and J.D. Watson, The Complementary Structure of Deoxyribonucleic Acid, *Proc. Roy. Soc. A*, **1954**, *223*, 80–96.

[175] W. Fuller, M.H. Wilkins, H.R. Wilson, and L.D. Hamilton, The Molecular Configuration of Deoxyribonucleic Acid. IV. X-Ray Diffraction Study of the A Form, *J. Mol. Biol.*, **1965**, *12*, 60–76.

[176] N. Pattabiraman, S.N. Rao, K. Scott, R. Langridge, and P.A. Kollman, Molecular Mechanical Studies on Left- and Right-Handed B-DNA, *Biopolymers*, **1987**, *26*, 403–414.

[177] J. Stroud, The Make-NA Server, http://structure.usc.edu/make-na/server.html, **2004**.

[178] D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, R.C. Walker, W. Zhang, K.M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, and P.A. Kollman, *Amber*, University of California, San Francisco, **2008**.

[179] C.L. Clark, P.K. Cecil, D. Singh, and D.M. Gray, CD, Absorption and Thermodynamic Analysis of Repeating Dinucleotide DNA, RNA and Hybrid Duplexes $[d/r(AC)]_{12}\cdot[d/r(GT/U)]_{12}$ and the Influence of Phosphorothioate Substitution, *Nucleic Acids Res.*, **1997**, *25*, 4098–4105.

[180] H. Drew, T. Takano, S. Tanaka, K. Itakura, and R.E. Dickerson, High-Salt d(CpGpCpG), a Left-Handed Z' DNA Double Helix, *Nature*, **1980**, *286*, 567–573.

[181] K. Hall, P. Cruz, I. Tinoco, Jr., T.M. Jovin, and J.H. van de Sande, 'Z-RNA' – A Left-Handed RNA Double Helix, *Nature*, **1984**, *311*, 584–586.

[182] D.H. Turner, Bulges in Nucleic Acids, *Curr. Opin. Struct. Biol.*, **1992**, *2*, 334–337.

[183] D.M.J. Lilley, Kinking of DNA and RNA by Base Bulges, *Proc. Natl. Acad. Sci. U. S. A.*, **1995**, *92*, 7140–7142.

[184] Z. Xi, D. Ouyang, and H.T. Mu, Interaction of Bulged DNA with Leucine-Containing Mimics of NCS-Chrom, *Bioorg. Med. Chem. Lett.*, **2006**, *16*, 1185–1190.

[185] D.M. Gray, J.D. Wen, C.W. Gray, R. Repges, C. Repges, G. Raabe, and J. Fleischhauer, Measured and Calculated CD Spectra of G-Quartets Stacked with the Same or Opposite Polarities, *Chirality*, **2008**, *20*, 431–440.

[186] J.T. Davis, G-Quartets 40 Years Later: From 5'-GMP to Molecular Biology and Supramolecular Chemistry, *Angew. Chem. Int. Ed.*, **2004**, *43*, 668–698.

[187] S. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, and S. Neidle, Quadruplex DNA: Sequence, Topology and Structure, *Nucleic Acids Res.*, **2006**, *34*, 5402–5415.

[188] S. Iijima and T. Ichihashi, Single-Shell Carbon Nanotubes of 1-nm Diameter, *Nature*, **1993**, *363*, 603–605.

[189] D.S. Bethune, C.H. Kiang, M.S. de Vries, G. Gorman, R. Savoy, J. Vazques, and R. Beyers, Cobalt-Catalysed Growth of Carbon Nanotubes with Single-Atomic-Layer Walls, *Nature*, **1993**, *363*, 605–607.

[190] E. Bichoutskaia, M.I. Heggie, Y.E. Lozovik, and A.M. Popov, Multi-Walled Nanotubes: Commensurate-Incommensurate Phase Transition and NEMS Applications, *Fullerenes*, **2006**, *14*, 131–140.

[191] G.D. Pantoş, P. Pengo, and J.K.M. Sanders, Hydrogen-Bonded Helical Organic Nanotubes, *Angew. Chem. Int. Ed.*, **2007**, *46*, 194–197.

[192] G.D. Pantoş, J.L. Wietor, and J.K.M. Sanders, Filling Helical Nanotubes with $C_{60}$, *Angew. Chem. Int. Ed.*, **2007**, *46*, 2238–2240.

[193] H.E. Katz, A.J. Lovinger, J. Johnson, C. Kloc, T. Siegrist, W. Li, Y.Y. Lin, and A. Dodabalapur, A Soluble and Air-Stable Organic Semiconductor with High Electron Mobility, *Nature*, **2000**, *404*, 478–481.

[194] F. Würthner, S. Ahmed, C. Thalacker, and T. Debaerdemaeker, Core-Substituted Naphthalene Bisimides: New Fluorophors with Tunable Emission Wavelength for FRET Studies, *Chem. Eur. J.*, **2002**, *8*, 4742–4750.

[195] P. Mukhopadhyay, Y. Iwashita, M. Shirakawa, S.I. Kawano, N. Fujita, and S. Shinkai, Spontaneous Colorimetric Sensing of the Positional Isomers of Dihydroxynaphthalene in a 1D Organogel Matrix, *Angew. Chem. Int. Ed.*, **2006**, *45*, 1592–1595.

[196] M. Sterzel, M. Pilch, M.T. Pawlikowski, and P. Skowronek, The Absorption and Fluorescence Study of 1,4,5,8-Naphthalenetetracarboxy Diimides in their Low-Energy 1 $^1B_{2u}$ and 1 $^1B_{3u}$ Electronic States. The

Franck–Condon Analysis in Terms of CASSCF Method, *Chem. Phys. Lett.*, **2002**, *362*, 243–248.

[197] M. Sterzel, M. Pilch, M.T. Pawlikowski, and J. Gawroński, The Circular Dichroism (CD) and Absorption Studies of 1,4,5,8-Naphthalene Tetracarboxydiimide Dimer in Terms of Vibronic Coupling Theory, *Chem. Phys.*, **2003**, *291*, 251–260.

[198] N. Harada and K. Nakanishi, *Circular Dichroic Spectroscopy: Exciton Coupling in Organic Stereochemistry*, University Science Books, Oxford University Press, Oxford, **1983**.

[199] B.O. Roos, The Complete Active Space SCF Method in a Fock-Matrix-Based Super-CI Formulation, *Int. J. Quant. Chem. Quant. Chem. Sym.*, **1980**, *S14*, 175–189.

[200] B.O. Roos and K. Andersson, Multiconfigurational Perturbation Theory with Level Shift – The $Cr_2$ Potential Revisited, *Chem. Phys. Lett.*, **1995**, *245*, 215–223.

[201] S. Matile, N. Berova, K. Nakanishi, J. Fleischhauer, and R.W. Woody, Structural Studies by Exciton Coupled Circular Dichroism over a Large Distance: Porphyrin Derivatives of Steroids, Dimeric Steroids, and Brevetoxin B, *J. Am. Chem. Soc.*, **1996**, *118*, 5198–5206.

[202] S.M. Vlaming, R. Augulis, M.C.A. Stuart, J. Knoester, and P.H.M. van Loosdrecht, Exciton Spectra and the Microscopic Structure of Self-Assembled Porphyrin Nanotubes, *J. Phys. Chem. B*, **2009**, *113*, 2273–2283.

[203] D. Nachtigallová, P. Hobza, and H.H. Ritze, Electronic Splitting in the Excited States of DNA Base Homodimers and -trimers: An Evaluation of Short-Range and Coulombic Interactions, *Phys. Chem. Chem. Phys.*, **2008**, *10*, 5689–5697.

[204] M.T. Oakley, G. Guichard, and J.D. Hirst, Calculations on the Electronic Excited States of Ureas and Oligoureas, *J. Phys. Chem. B*, **2007**, *111*, 3274–3279.

[205] K. Tsubaki, Synthesis and Properties of the Chiral Oligonaphthalenes, *Org. Biomol. Chem. Org Biomol*, **2007**, *5*, 2179–2188.

[206] P. Zazakowny, M.T. Pawlikowski, and M. Sterzel, The Vibronic Structures of Absorption and Magnetic Circular Dichroism (MCD) in the Low Energy $1^1B_{2u}$ and $1^1B_{3u}$ states of 1,4,5,8-Naphthalenetetracarboxy dianhydride. The Analysis in Terms of DFT and CASSCF Methods, *Chem. Phys. Lett.*, **2009**, *472*, 55–59.

[207] P. Talukdar, G. Bollot, J. Mareda, N. Sakai, and S. Matile, Ligand-Gated Synthetic Ion Channels, *Chem. Eur. J.*, **2005**, *11*, 6525–6532.

[208] N. Sakai, P. Talukdar, and S. Matile, Use of the Exciton Chirality Method in the Investigation of Ligand-Gated Synthetic Ion Channels, *Chirality*, **2006**, *18*, 91–94.

[209] M. Sterzel, M. Andrzejak, M.T. Pawlikowski, and J. Gawroński, Absorption and Magnetic Circular Dichroism (MCD) Studies of 1,4,5,8-Naphthalenetetracarboxy Diimides in Terms of CASSCF Method and FC Theory, *Chem. Phys.*, **2004**, *300*, 93–105.

[210] L. Serrano-Andrés and M.P. Fülscher, Theoretical Study of the Electronic Spectroscopy of Peptides. 1. The Peptidic Bond: Primary, Secondary, and Tertiary Amides, *J. Am. Chem. Soc.*, **1996**, *118*, 12190–12199.

[211] B.T. Chait, PROWL – Mass Spectrometry Analysis Online Tools, http://prowl.rockefeller.edu/prowl/prowl.html.

[212] R.J. Chalkley, K.C. Hansen, M.A. Baldwin, P.R. Baker, K.F. Medzihradszky, A.J. Lynn, A.L. Burlingame, K.R. Clauser, and

L. Huang, ProteinProspector – Proteomics Tools for Mining Sequence Databases in Conjunction with Mass Spectrometry Experiments, http://prospector.ucsf.edu, **2008**.

[213] D. Barthel, J.D. Hirst, J. Błażewicz, and N. Krasnogor, ProCKSI: A Decision Support System for Protein (Structure) Comparison, Knowledge, Similarity and Information, *BMC Bioinf.*, **2007**, *8*, 416–437.

[214] C. Cole, J.D. Barber, and G.J. Barton, The Jpred 3 Secondary Structure Prediction Server, *Nucleic Acids Res.*, **2008**, *36*, W197–W201.

[215] B.A. Wallace, L. Whitmore, and R.W. Janes, The Protein Circular Dichroism Data Bank (PCDDB): A Bioinformatics and Spectroscopic Resource; http://pcddb.cryst.bbk.ac.uk, *Proteins: Struct., Funct., Bioinf.*, **2006**, *63*, 1–3.

[216] L. Whitmore and B.A. Wallace, DICHROWEB, an Online Server for Protein Secondary Structure Analyses from Circular Dichroism Spectroscopic Data; http://www.cryst.bbk.ac.uk/cdweb, *Nucleic Acids Res.*, **2004**, *32*, W668–W673.

[217] N. Sreerama, S.Y. Venyaminov, and R.W. Woody, Estimation of the Number of $\alpha$-helical and $\beta$-strand Segments in Proteins using Circular Dichroism Spectroscopy, *Protein Sci.*, **1999**, *8*, 370–380.

[218] N. Sreerama and R.W. Woody, A Self-Consistent Method for the Analysis of Protein Secondary Structure from Circular Dichroism, *Anal. Biochem.*, **1993**, *209*, 32–44.

[219] S.W. Provencher and J. Glöckner, Estimation of Globular Protein Secondary Structure from Circular Dichroism, *Biochemistry*, **1981**, *20*, 33–37.

[220] I.H.M. van Stokkum, H.J.W. Spoelder, M. Bloemendal, R. van Grondelle, and F.C.A. Groen, Estimation of Protein Secondary Structure and Error

Analysis from Circular Dichroism Spectra, *Anal. Biochem.*, **1990**, *191*, 110–118.

[221] L.A. Compton and W.C. Johnson, Jr., Analysis of Protein Circular Dichroism Spectra for Secondary Structure using a Simple Matrix Multiplication, *Anal. Biochem.*, **1986**, *155*, 155–167.

[222] P. Manavalan and W.C. Johnson, Jr., Variable Selection Method Improves the Prediction of Protein Secondary Structure from Circular Dichroism Spectra, *Anal. Biochem.*, **1987**, *167*, 76–85.

[223] N. Sreerama and R.W. Woody, Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set, *Anal. Biochem.*, **2000**, *287*, 252–260.

[224] M.A. Andrade, P. Chácon, J.J. Merelo, and F. Morán, Evaluation of Secondary Structure of Proteins from UV Circular Dichroism Spectra Using an Unsupervised Learning Neural Network, *Protein Eng.*, **1993**, *6*, 383–390.

[225] M. Ardhammar, P. Lincoln, and B. Nordén, Ligand Substituents of Ruthenium Dipyridophenazine Complexes Sensitively Determine Orientation in Liposome Membrane, *J. Phys. Chem. B*, **2001**, *105*, 11363–11368.

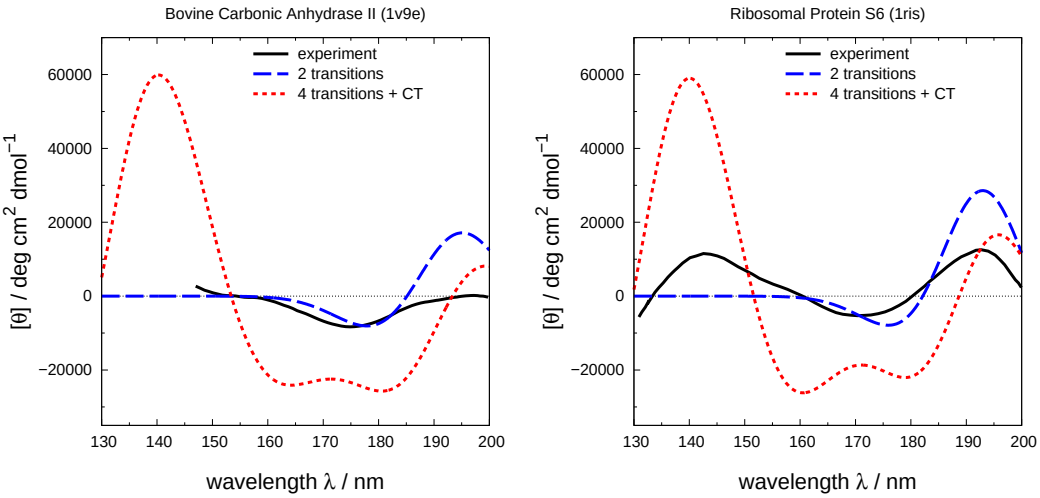
[226] J.E. Stajich, D. Block, K. Boulez, S.E. Brenner, S.A. Chervitz, C. Dagdigian, G. Fuellen, J.G.R. Gilbert, I. Korf, H. Lapp, H. Lehväslaiho, C. Matsalla, C.J. Mungall, B.I. Osborne, M.R. Pocock, P. Schattner, M. Senger, L.D. Stein, E. Stupka, M.D. Wilkinson, and E. Birney, The BioPerl Toolkit: Perl Modules for the Life Sciences, *Genome Res.*, **2002**, *12*, 1611–1618.

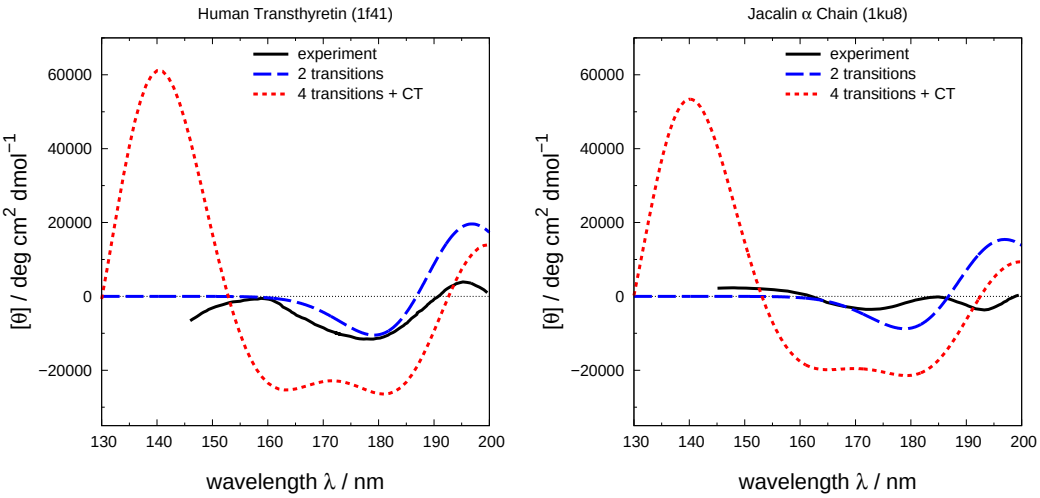
[227] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus, CHARMM – A Program for Macromolecular Energy, Minimization, and Dynamics Calculations, *J. Comput. Chem.*, **1983**, *4*, 187–217.

# Appendix B

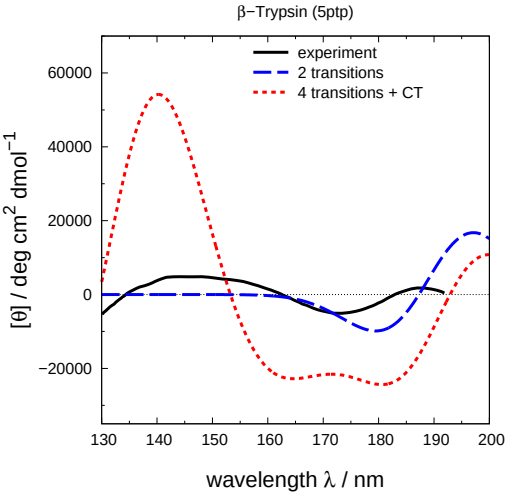
# CD Spectra of the CD130 Protein Set

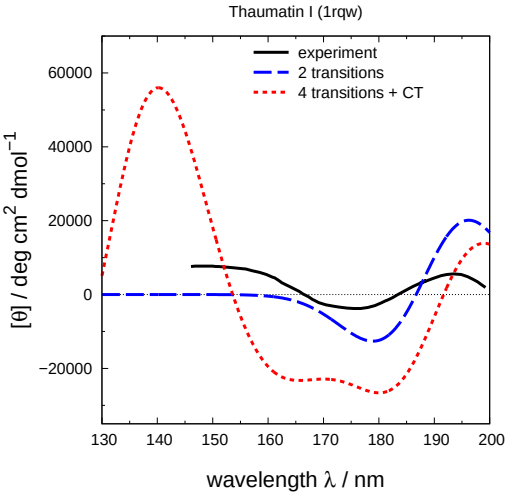
| | |
|---|---|
| 2 transitions | 2 backbone transitions<br>$n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ |
| 4 transitions + CT | 4 backbone transitions and charge-transfer<br>$n \rightarrow \pi^*$, $\pi \rightarrow \pi^*$, $\pi_b \rightarrow \pi^*$ and $n' \rightarrow \pi^*$ |

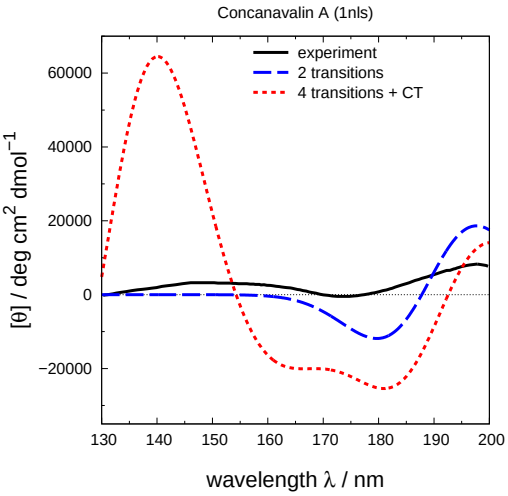


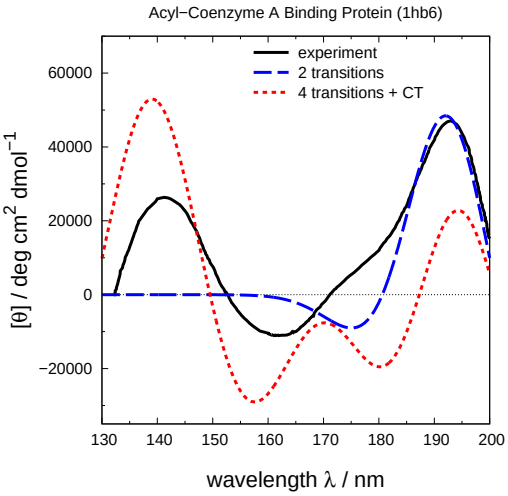

Human Transthyretin (1f41)
experiment
2 transitions
4 transitions + CT
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm

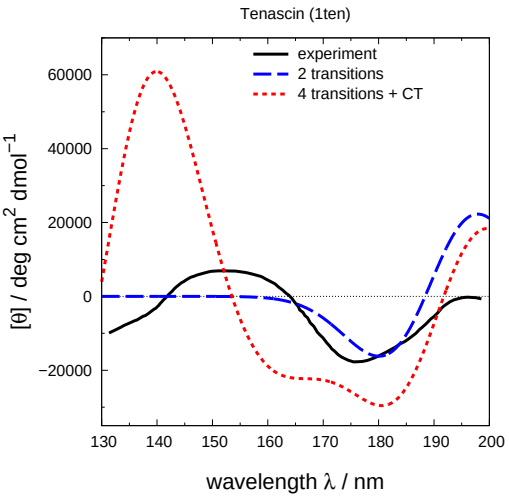

Jacalin α Chain (1ku8)
experiment
2 transitions
4 transitions + CT
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm

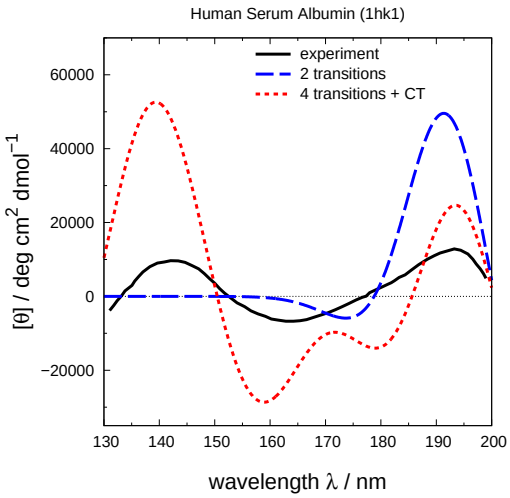

β−Trypsin (5ptp)
experiment
2 transitions
4 transitions + CT
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm

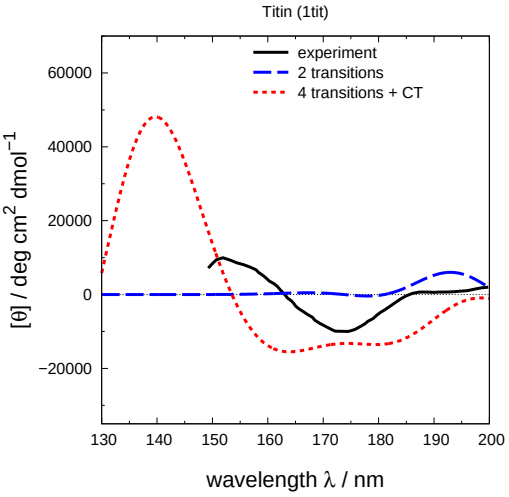

Thaumatin I (1rqw)
experiment
2 transitions
4 transitions + CT
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm

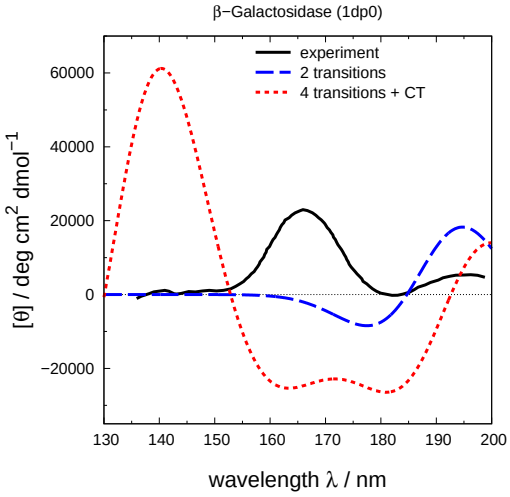

Concanavalin A (1nls)
experiment
2 transitions
4 transitions + CT
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm

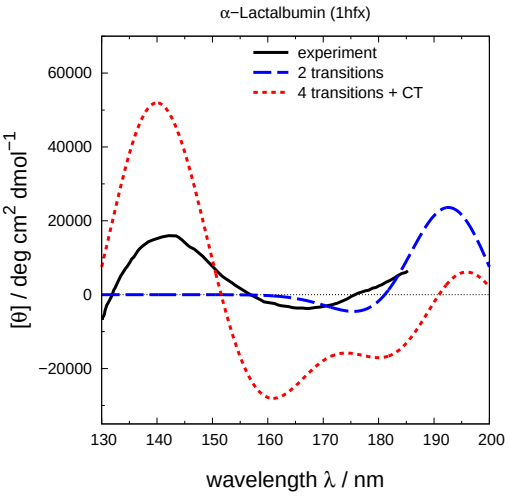

Acyl−Coenzyme A Binding Protein (1hb6)
experiment
2 transitions
4 transitions + CT
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm


Tenascin (1ten)
experiment
2 transitions
4 transitions + CT
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm


Human Serum Albumin (1hk1)
experiment
2 transitions
4 transitions + CT
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm


Titin (1tit)
experiment
2 transitions
4 transitions + CT
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm

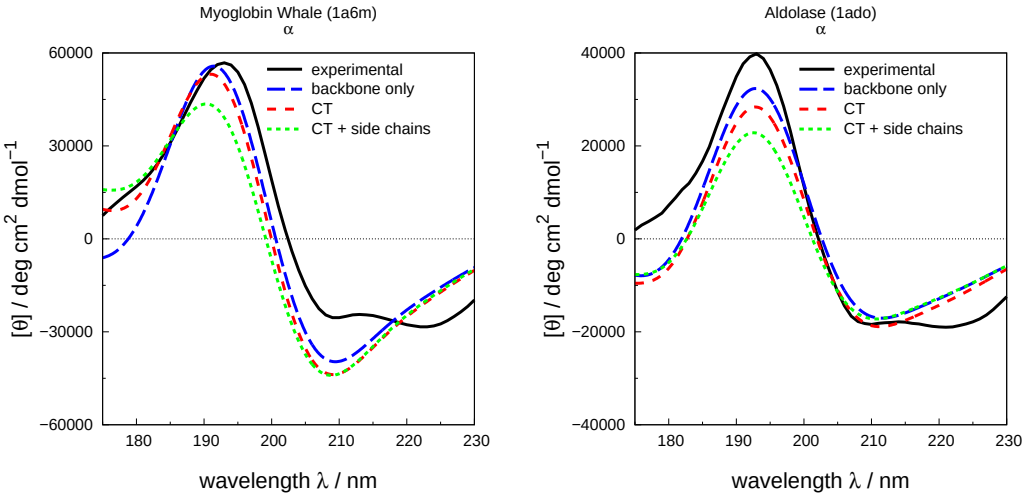

β−Galactosidase (1dp0)
experiment
2 transitions
4 transitions + CT
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm

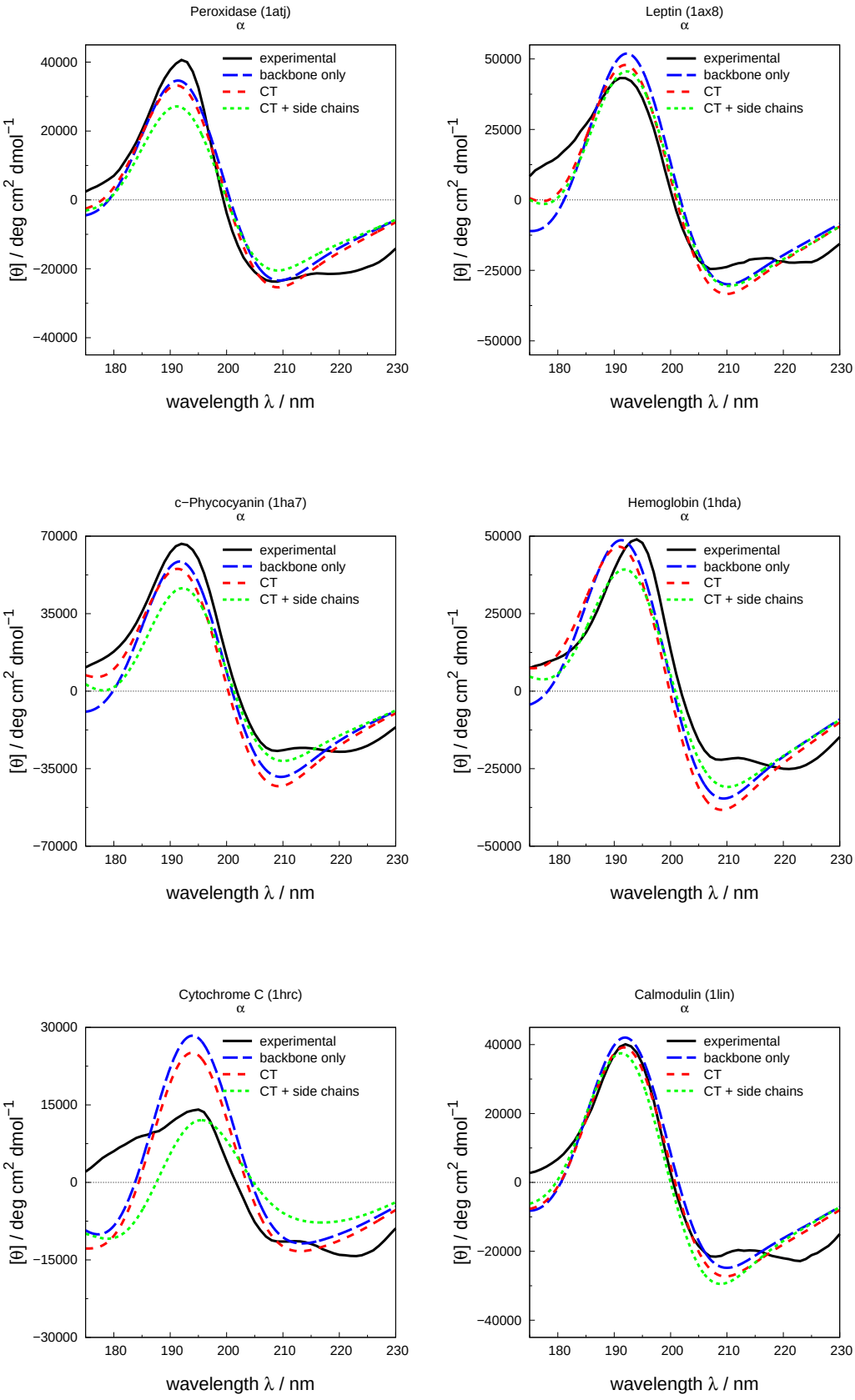

α−Lactalbumin (1hfx)
experiment
2 transitions
4 transitions + CT
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm

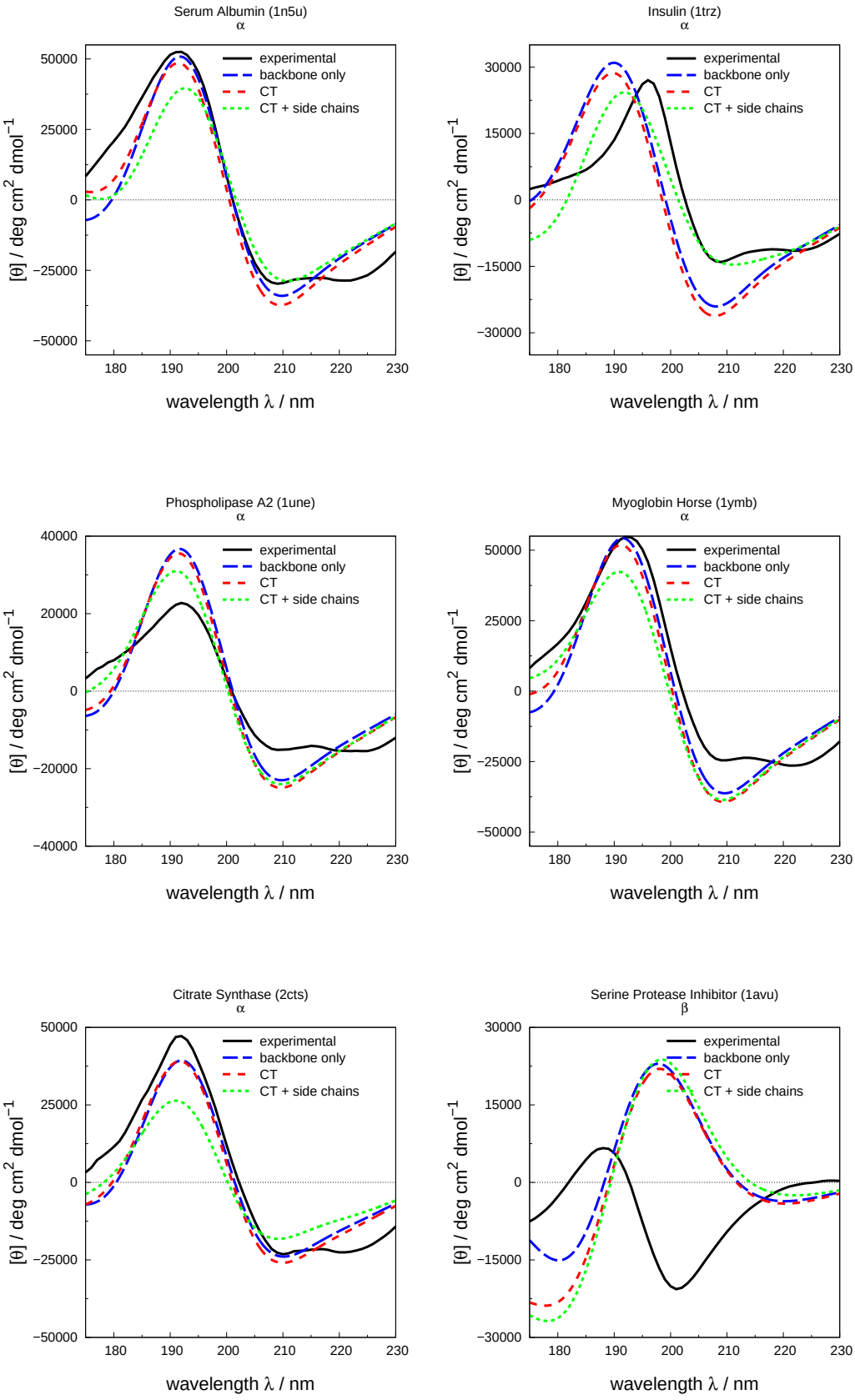
# Appendix C

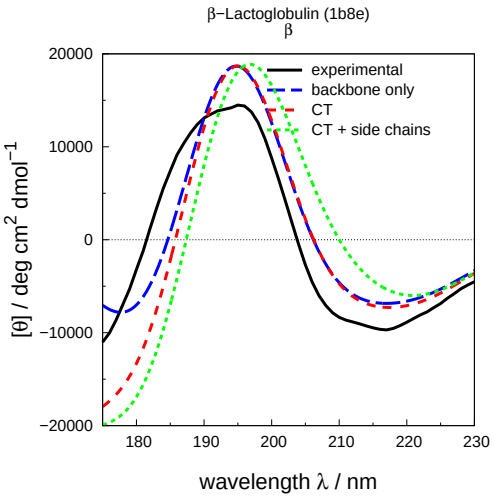
# CD Spectra of the SP175 Protein Set

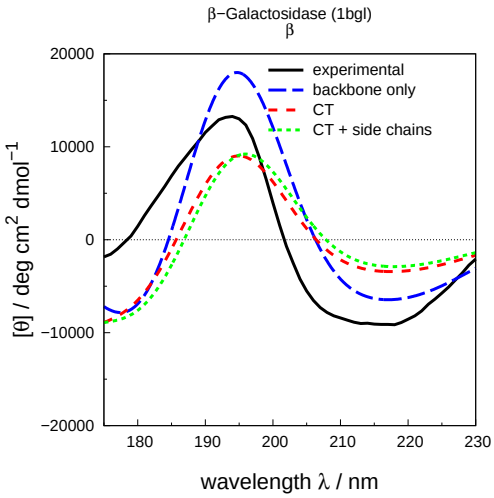
CT = backbone and charge-transfer transitions.

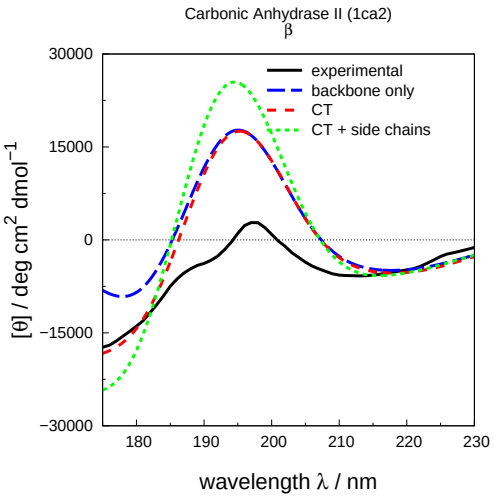


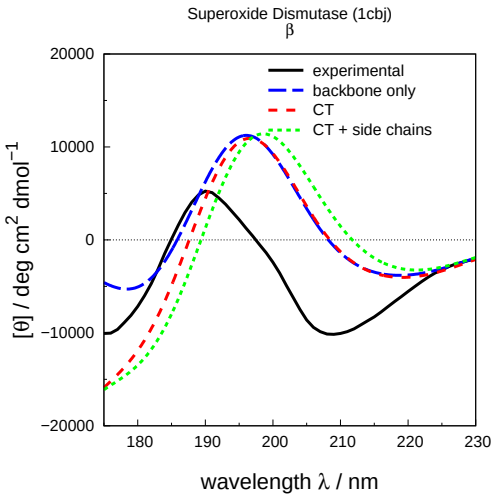

Peroxidase (1atj)
α
Leptin (1ax8)
α
c−Phycocyanin (1ha7)
α
Hemoglobin (1hda)
α
Cytochrome C (1hrc)
α
Calmodulin (1lin)
α
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

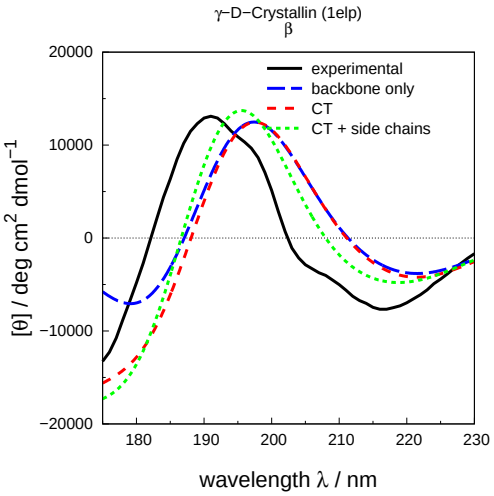

Serum Albumin (1n5u)
α
Insulin (1trz)
α
Phospholipase A2 (1une)
α
Myoglobin Horse (1ymb)
α
Citrate Synthase (2cts)
α
Serine Protease Inhibitor (1avu)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

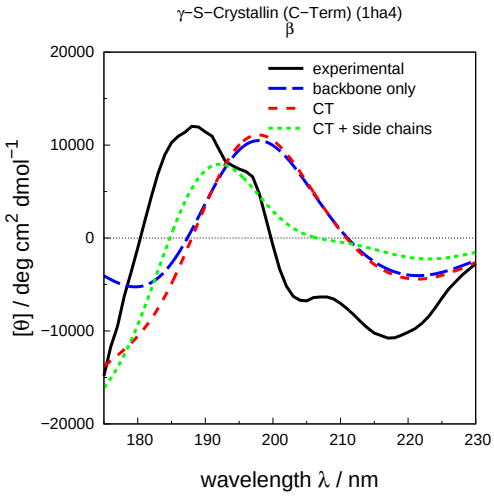

β−Lactoglobulin (1b8e)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

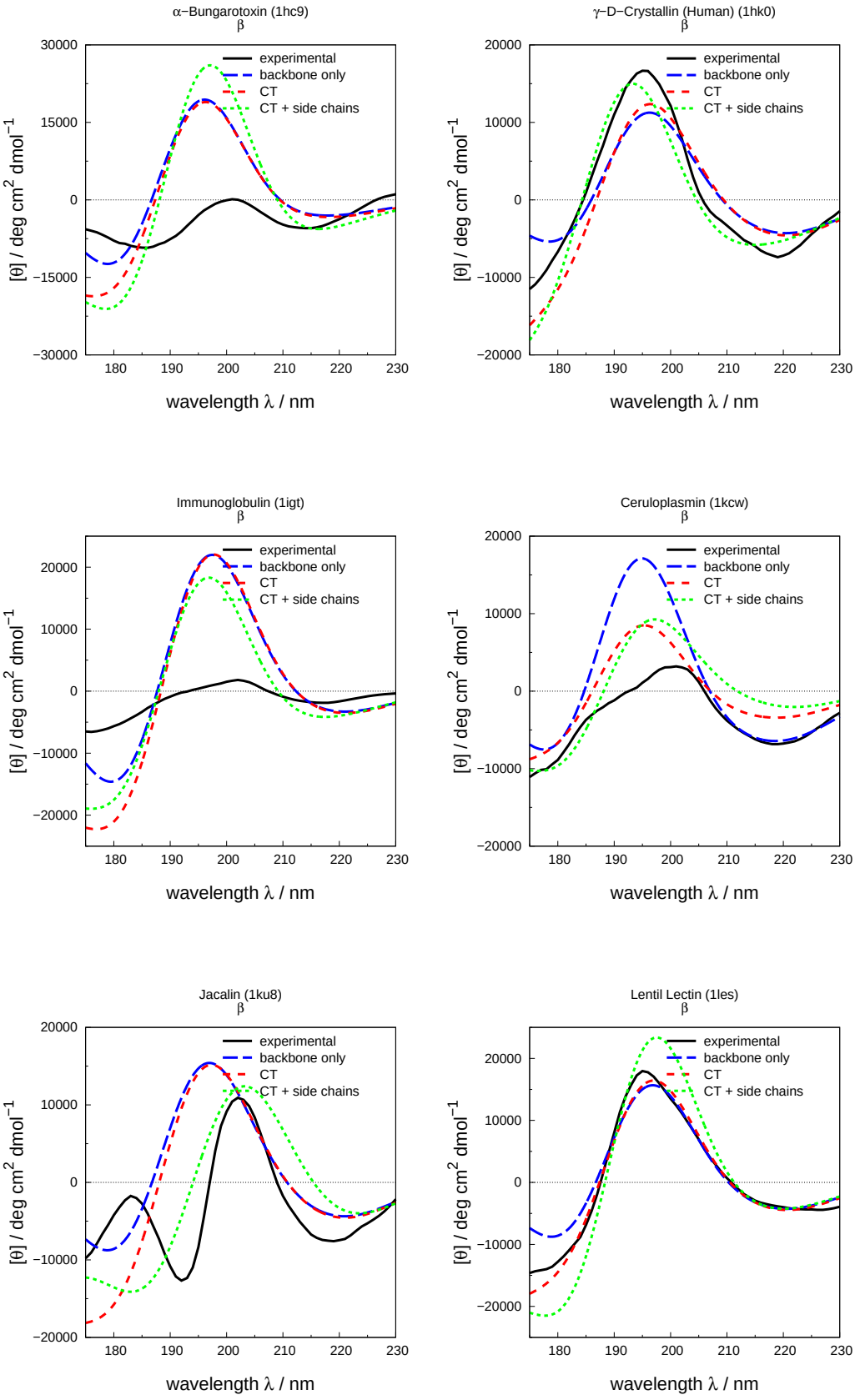

β−Galactosidase (1bgl)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

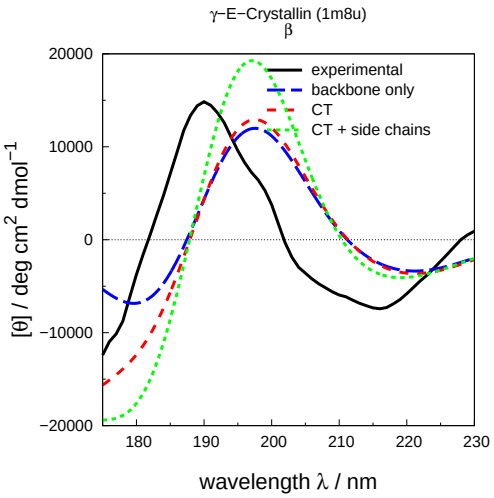

Carbonic Anhydrase II (1ca2)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

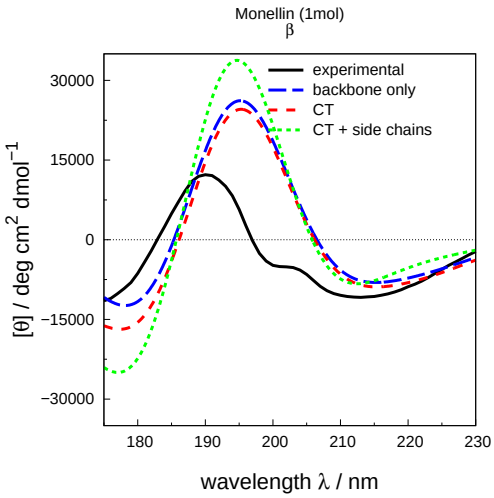

Superoxide Dismutase (1cbj)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

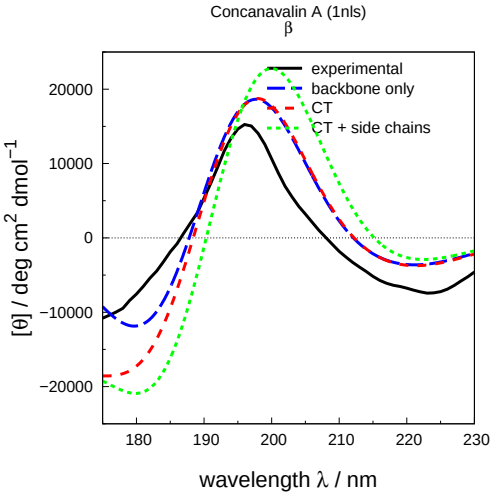

γ−D−Crystallin (1elp)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

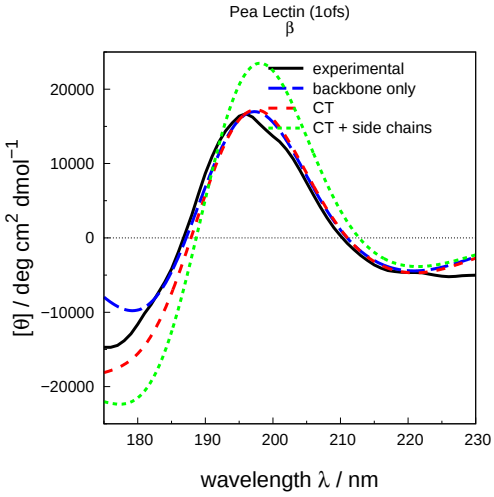

γ−S−Crystallin (C−Term) (1ha4)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

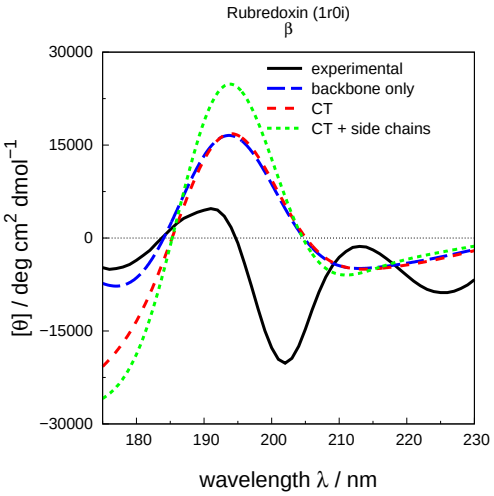

α−Bungarotoxin (1hc9)
β
γ−D−Crystallin (Human) (1hk0)
β
Immunoglobulin (1igt)
β
Ceruloplasmin (1kcw)
β
Jacalin (1ku8)
β
Lentil Lectin (1les)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

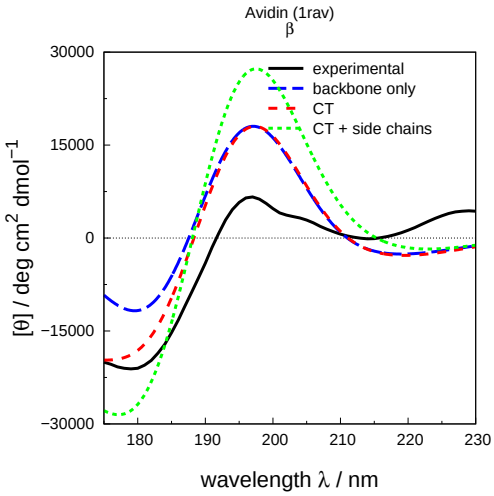

γ−E−Crystallin (1m8u)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

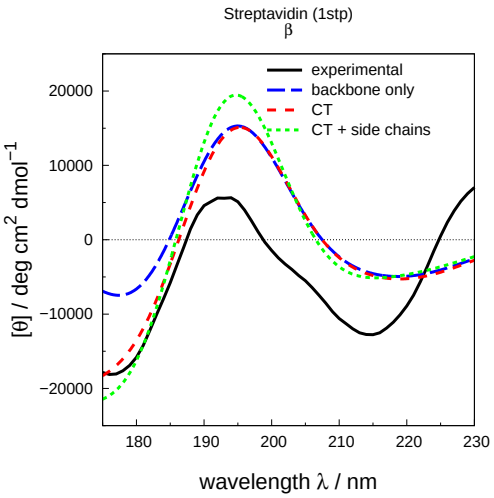

Monellin (1mol)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

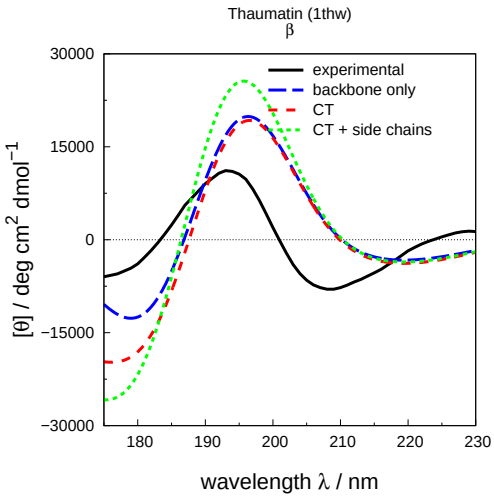

Concanavalin A (1nls)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

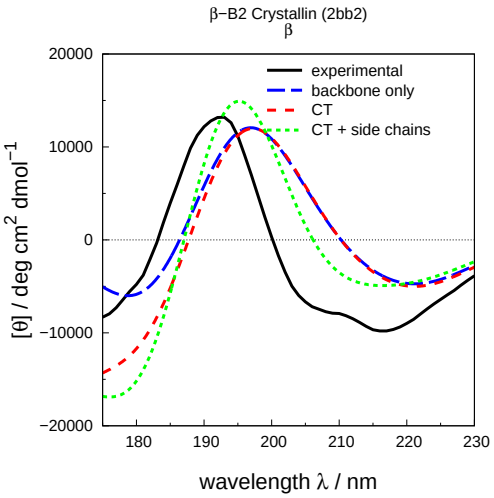

Pea Lectin (1ofs)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

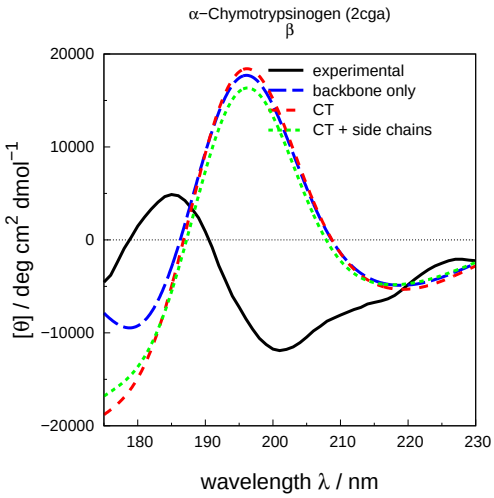

Rubredoxin (1r0i)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

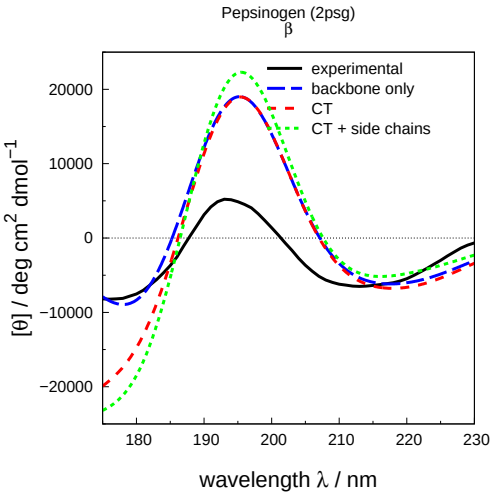

Avidin (1rav)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

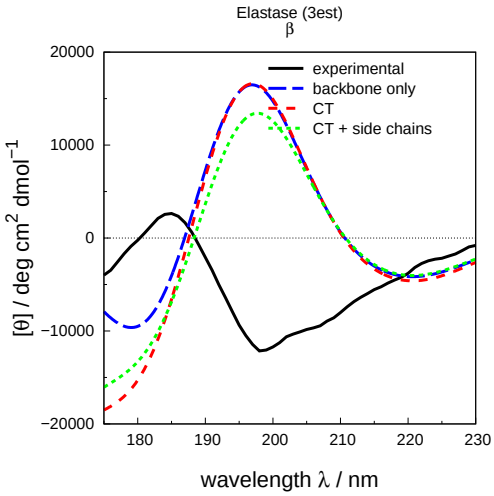

Streptavidin (1stp)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm

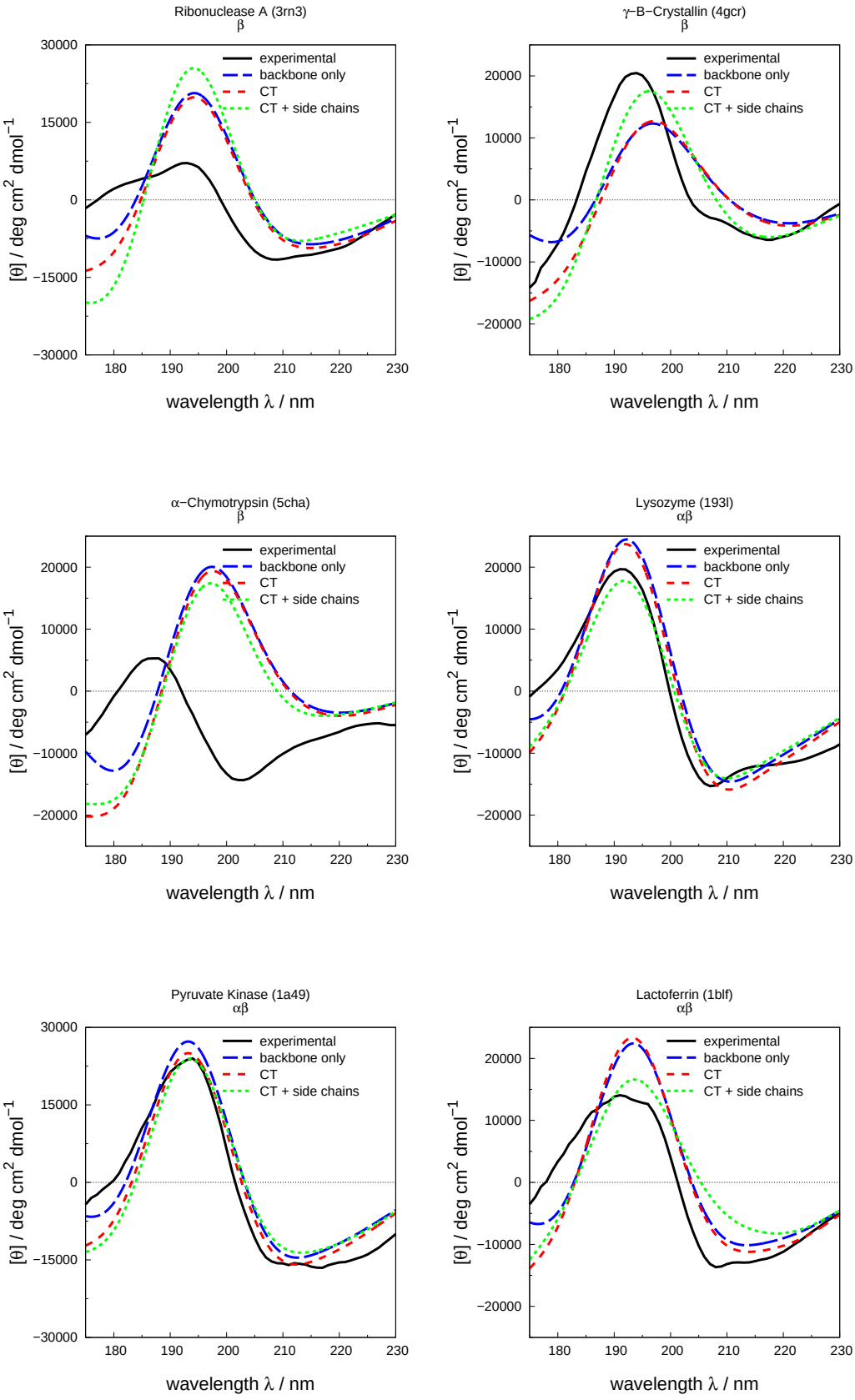

Thaumatin (1thw)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm

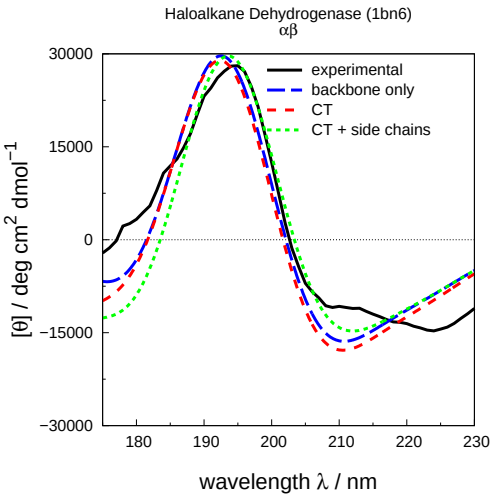

β−B2 Crystallin (2bb2)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm

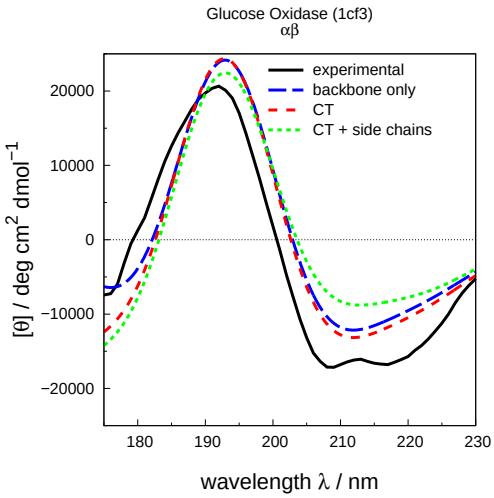

α−Chymotrypsinogen (2cga)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm

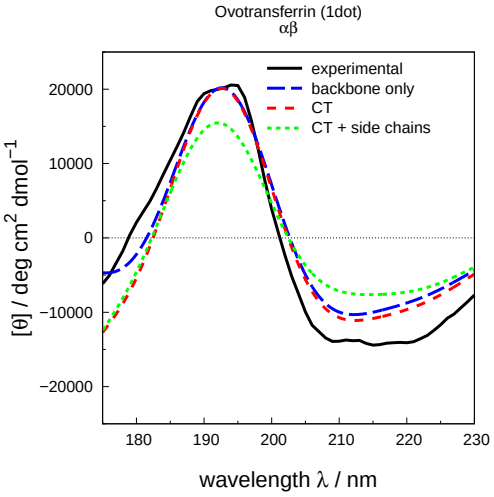

Pepsinogen (2psg)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm

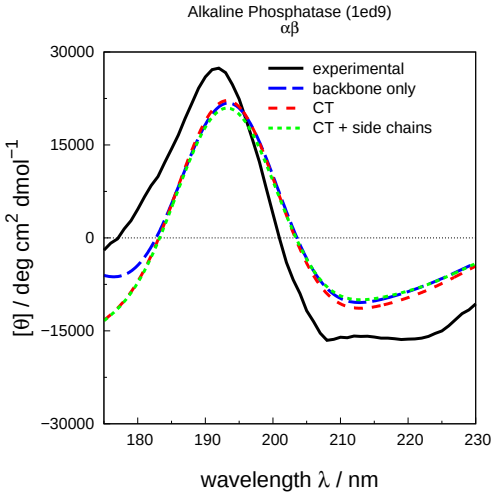

Elastase (3est)
β
experimental
backbone only
CT
CT + side chains
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm

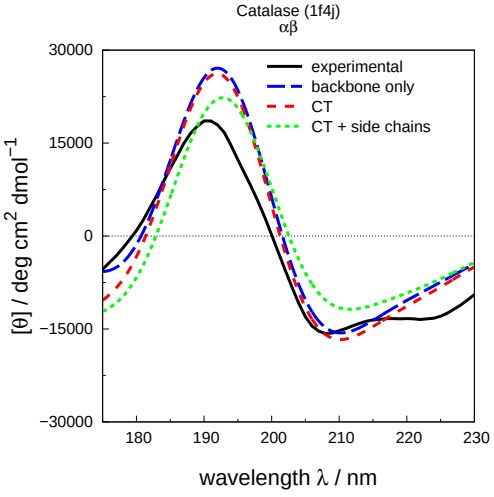

Ribonuclease A (3rn3)
β
γ−B−Crystallin (4gcr)
β
α−Chymotrypsin (5cha)
β
Lysozyme (193l)
αβ
Pyruvate Kinase (1a49)
αβ
Lactoferrin (1blf)
αβ
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

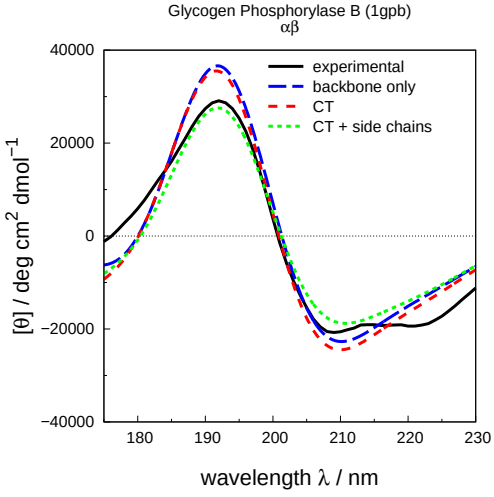

Haloalkane Dehydrogenase (1bn6)
αβ
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

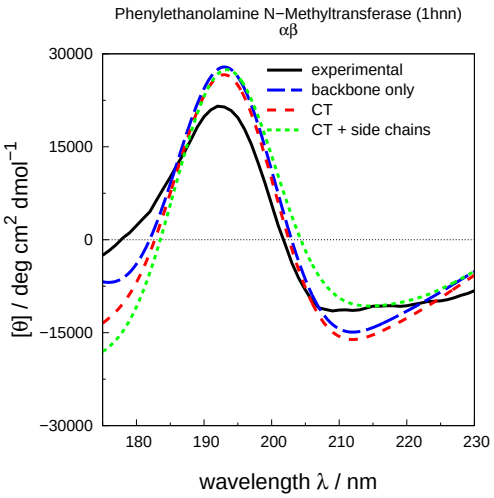

Glucose Oxidase (1cf3)
αβ
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

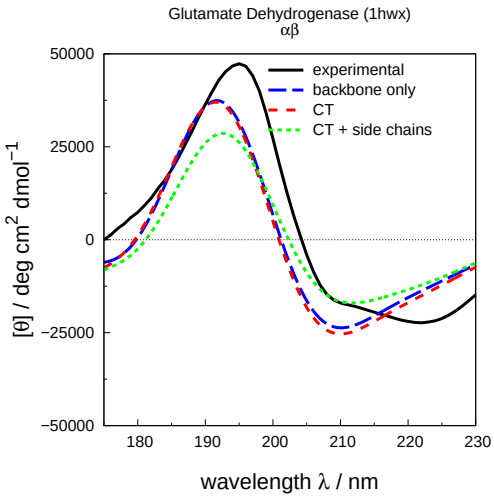

Ovotransferrin (1dot)
αβ
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

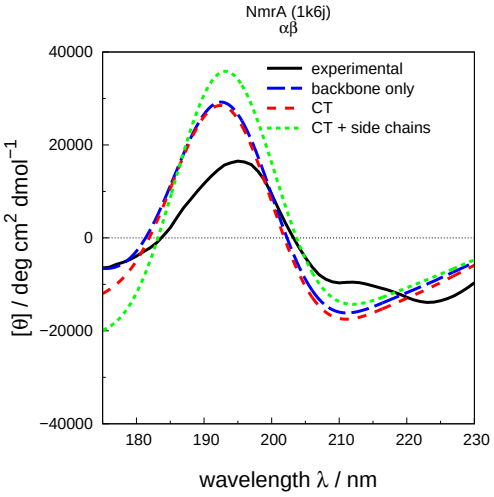

Alkaline Phosphatase (1ed9)
αβ
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

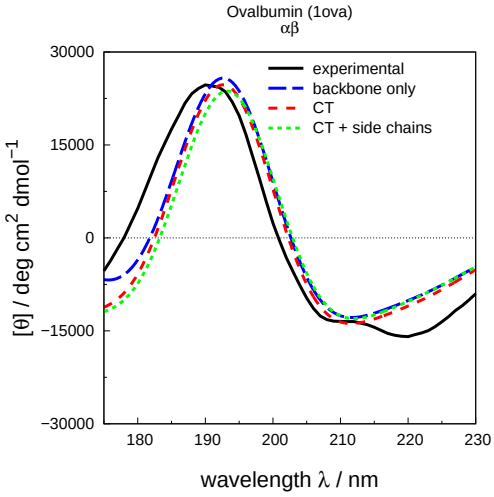

Catalase (1f4j)
αβ
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

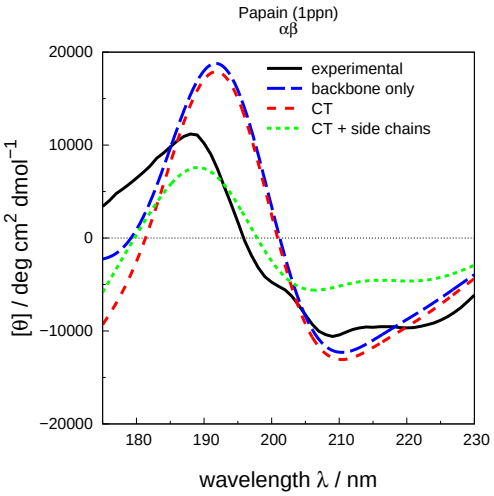

Glycogen Phosphorylase B (1gpb)
αβ
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

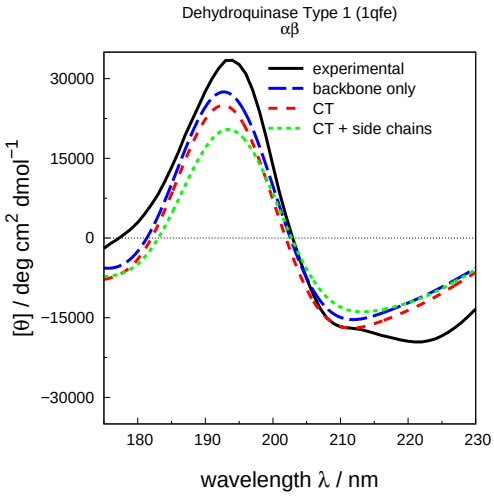

Phenylethanolamine N−Methyltransferase (1hnn)
αβ
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

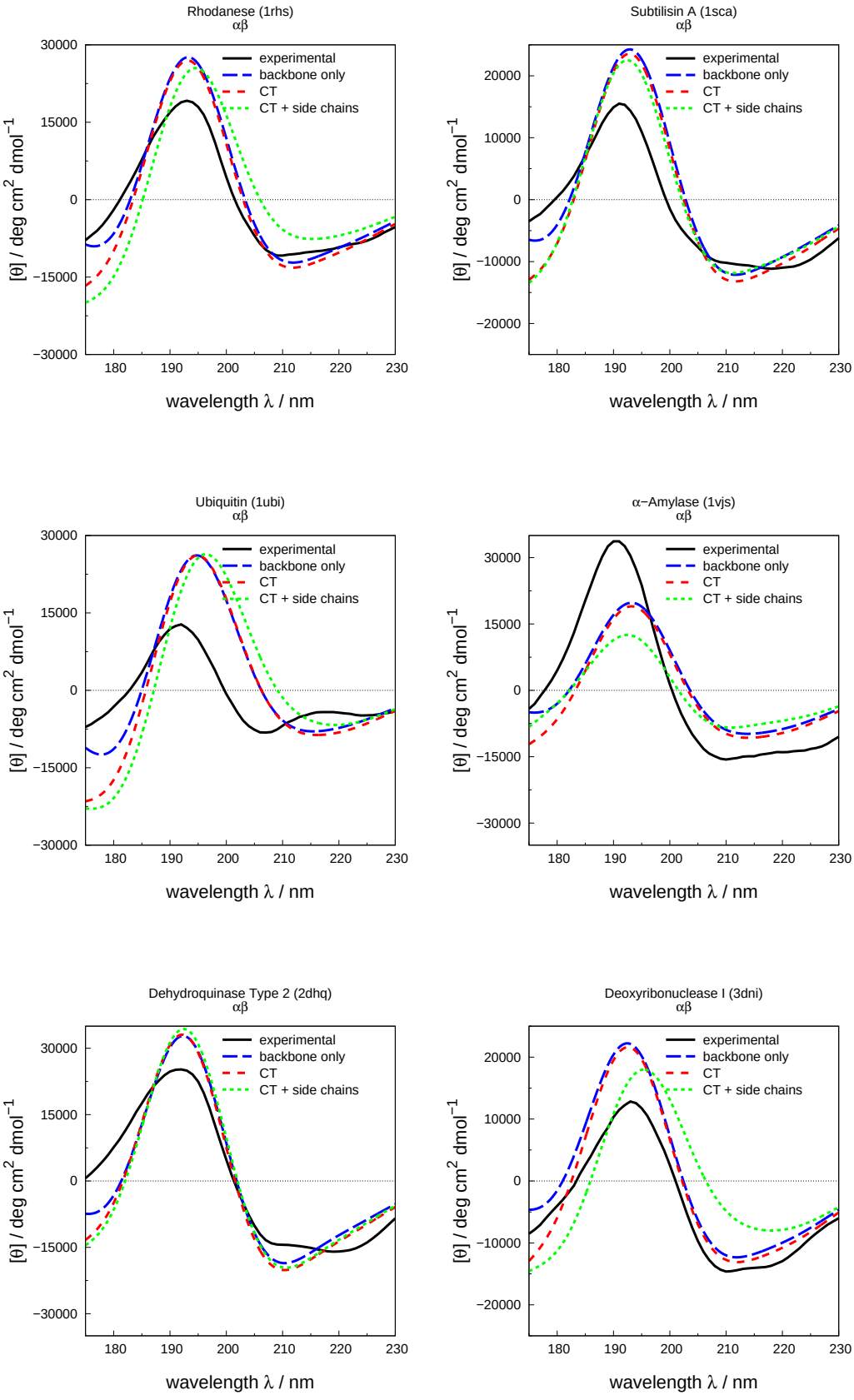

Glutamate Dehydrogenase (1hwx)
αβ
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

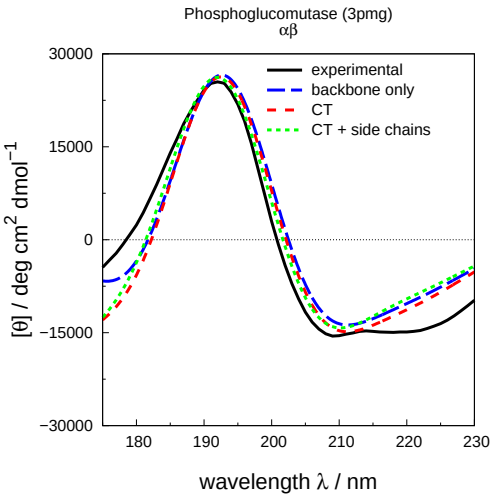

NmrA (1k6j)
αβ
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

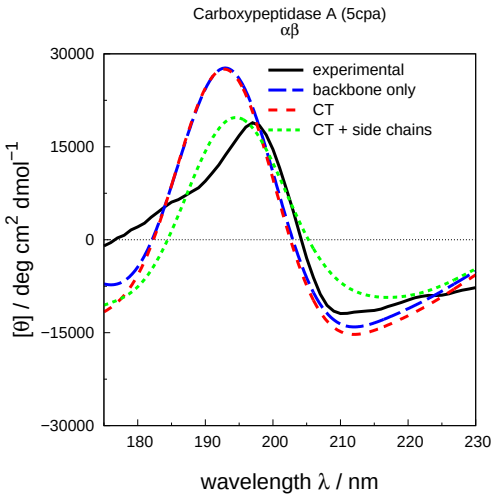

Ovalbumin (1ova)
αβ
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

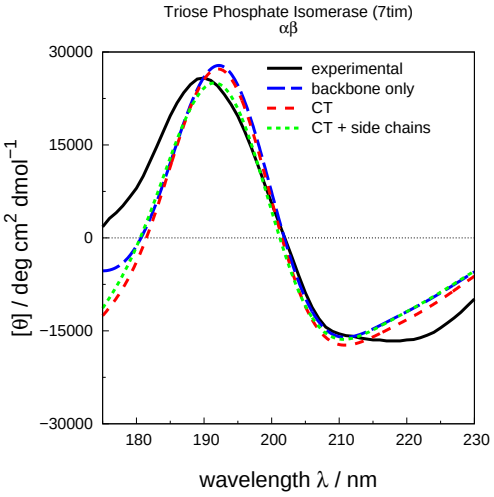

Papain (1ppn)
αβ
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

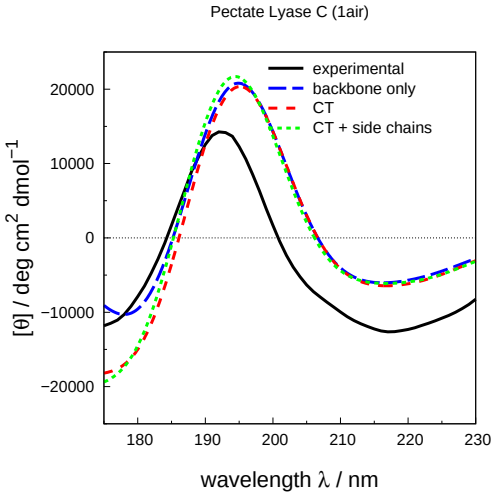

Dehydroquinase Type 1 (1qfe)
αβ
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

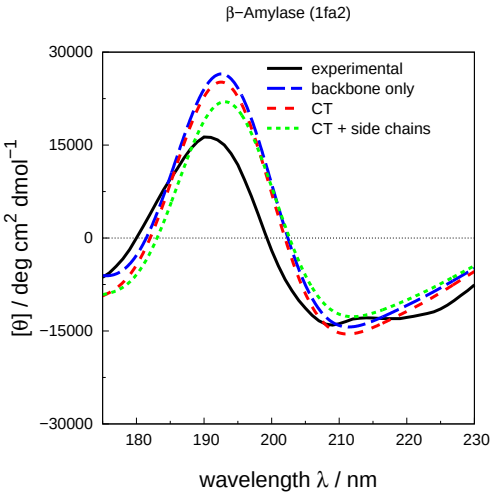

Rhodanese (1rhs)
αβ
Subtilisin A (1sca)
αβ
Ubiquitin (1ubi)
αβ
α−Amylase (1vjs)
αβ
Dehydroquinase Type 2 (2dhq)
αβ
Deoxyribonuclease I (3dni)
αβ
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

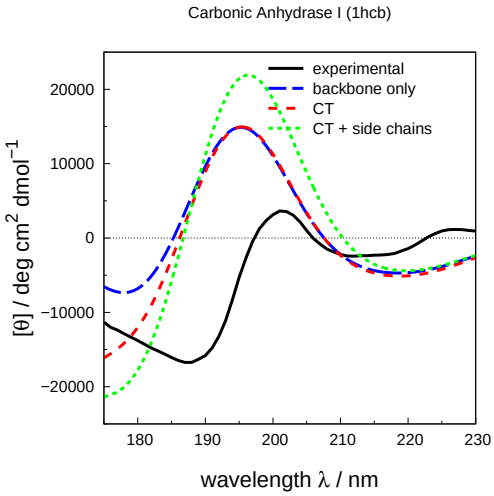

Phosphoglucomutase (3pmg)
αβ
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

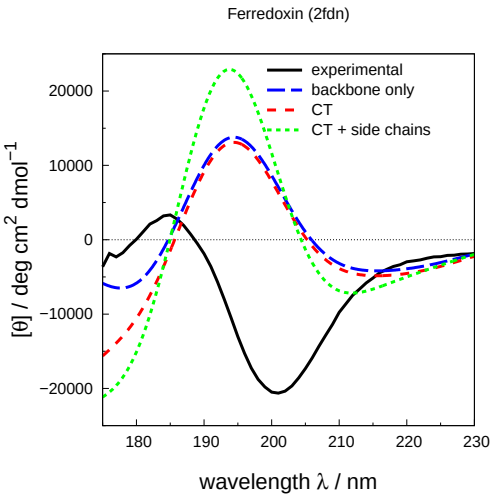

Carboxypeptidase A (5cpa)
αβ
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

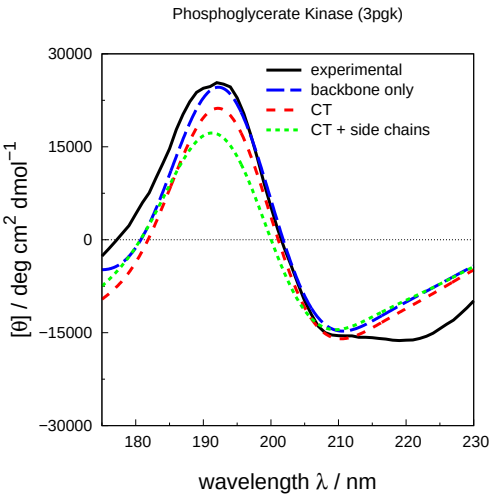

Triose Phosphate Isomerase (7tim)
αβ
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm

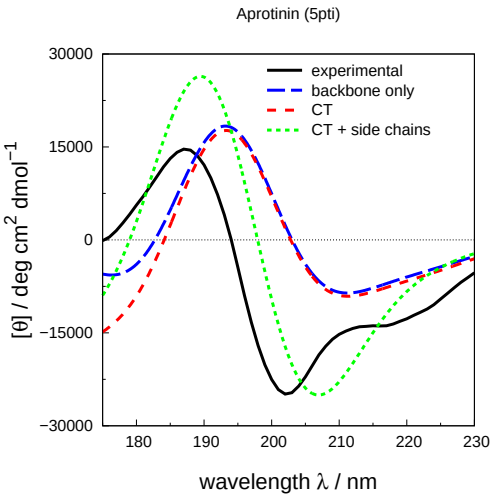

Pectate Lyase C (1air)
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm


β−Amylase (1fa2)
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm


Carbonic Anhydrase I (1hcb)
experimental
backbone only
CT
CT + side chains
[θ] / deg cm2 dmol−1
wavelength λ / nm


Ferredoxin (2fdn)
experimental
backbone only
CT
CT + side chains
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm


Phosphoglycerate Kinase (3pgk)
experimental
backbone only
CT
CT + side chains
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm


Aprotinin (5pti)
experimental
backbone only
CT
CT + side chains
[θ] / deg cm$^2$ dmol$^{-1}$
wavelength λ / nm

# Appendix D

# Orientation Search Report

The orientation search calculates a series of LD spectra for a given PDB file for all possible orientations of the protein. Depending on the chosen intervals and maximum angles for the tilt and internal rotation, this process may generate dozens of spectra. For the analysis, it is crucial to be able to relate the spectra to the corresponding structure easily. The protein is rotated about two axes and the definition of these, as well as the initial orientation of the protein, has to be considered. To aid the analysis, especially for new and casual users, the results of an orientation search are summarized in a report; this enables the user to find the corresponding PDB and *xy* data for a spreadsheet application. The font and graphics sizes had to be slightly reduced to maintain the layout of the report within the page margins of this work.

Actin was not included in the main discussion because it was uncertain whether the underestimation of the intense $n \rightarrow \pi^*$ band between 210 and 230 nm depends on the experimental conditions (if, for example, the sample concentration affects the fibre properties).

# LD Calculation of Actin

The $z$ axis is defined as *parallel* orientation, $x$ and $y$ as *perpendicular*. The protein is rotated with a certain rotation interval about the *rotation axis* until a full rotation has been carried out. For each rotation the LD spectrum is determined. After this, the protein and its rotation axis are tilted away from the initial position around the *tilt axis* about a given tilt interval and rotated again. This is repeated until a maximum tilt angle is reached.

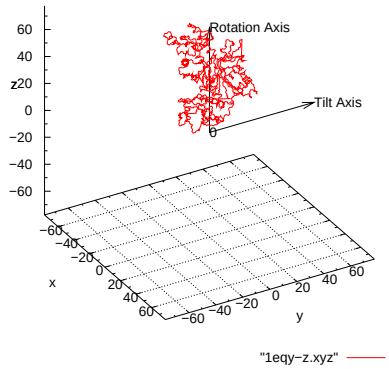
The spectra show the LD, calculated using the formula

$$LD = A_{\parallel} - A_{\perp} = \mu_z^2 - \frac{1}{2}(\mu_x^2 + \mu_y^2)$$

## Parameters for this calculation

| | | | |
|---|---|---|---|
| Tilt Axis: | $y$ | Rotation Axis: | $z$ |
| Tilt Interval: | 45° | Rotation Interval: | 45° |
| Tilt Angle: | 90° | Rotation Angle: | 180° |

## Original Orientation

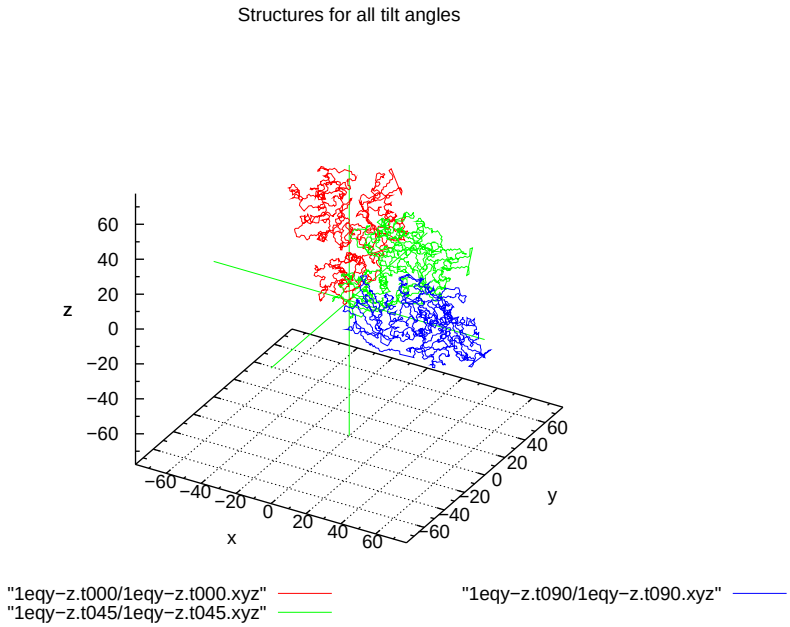



This calculation was done using DichroCalc
http://comp.chem.nottingham.ac.uk/dichrocalc
B.M. Bulheller & J.D. Hirst, DichroCalc — Circular and Linear Dichroism Online.
*Bioinformatics*, **25**, 539–540 (2009).

# Plot of the Protein Orientations at All Angles

## Initial Orientation at All Tilt Angles

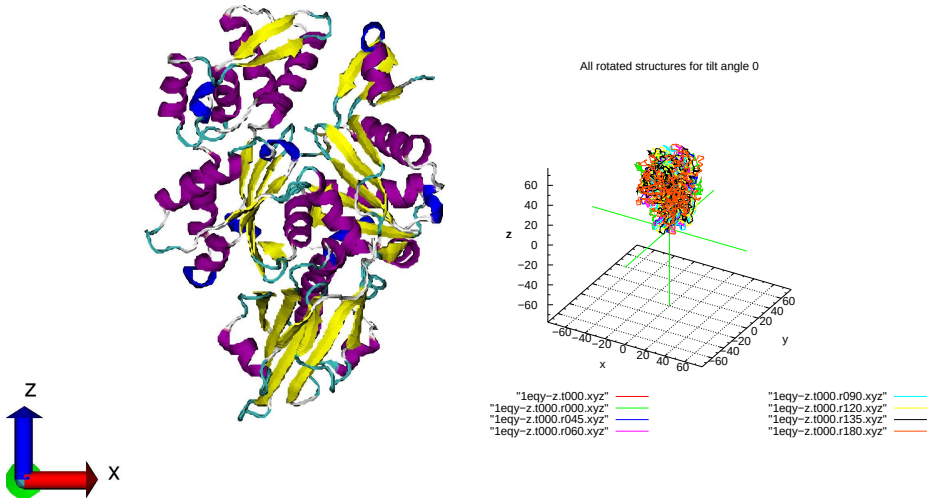
The protein is rotated about the tilt axis using the tilt interval. Each of the structures shown in the next figure is then rotated about the rotation axis. The initial rotation axis is $z$ and is tilted about the tilt axis with the protein.



# Plot of the Rotations at All Tilt Angles

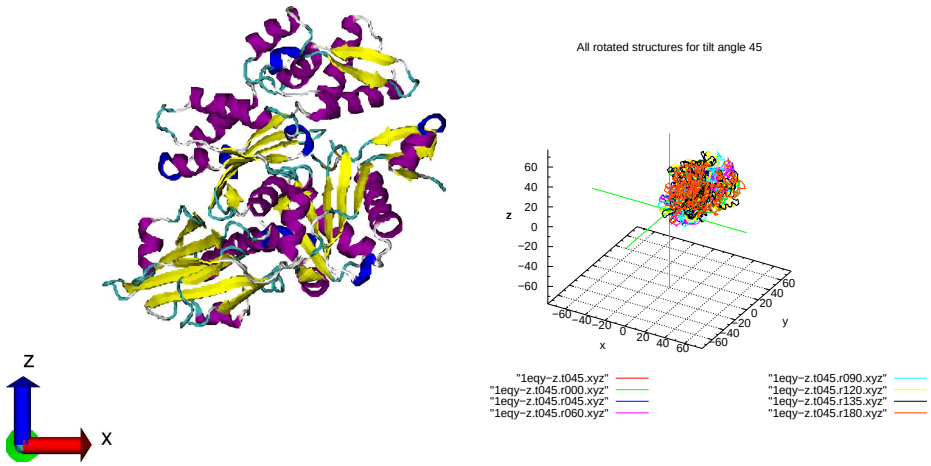
## Tilt Angle 0°

After the protein has been tilted 0° about the tilt axis $y$ it is rotated in 45° intervals around the rotation axis until 180° are reached.



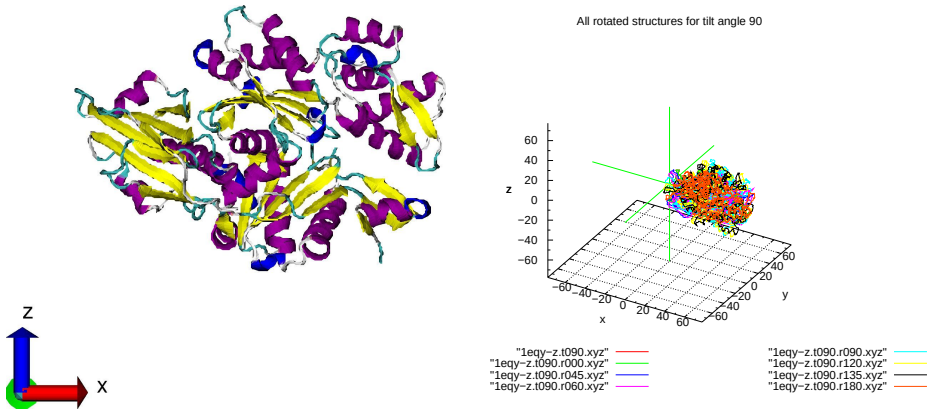

## Tilt Angle 45°

After the protein has been tilted 45° about the tilt axis $y$ it is rotated in 45° intervals around the rotation axis until 180° are reached.



### Tilt Angle 90°

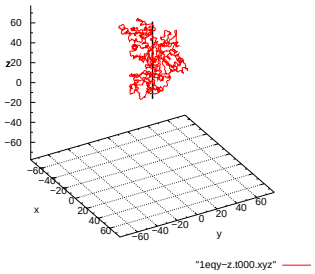
After the protein has been tilted 90° about the tilt axis $y$ it is rotated in 45° intervals around the rotation axis until 180° are reached.

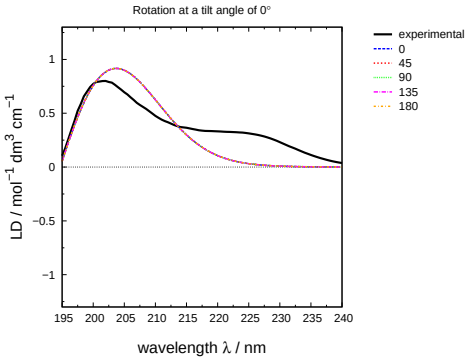


# Tilt Angle 0°

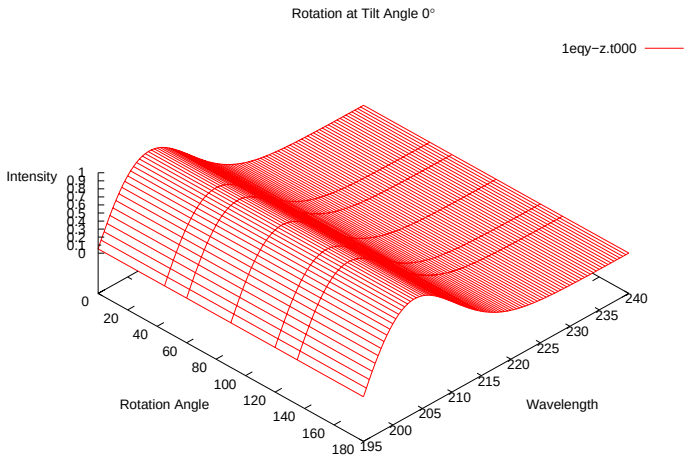
Directory: 1eqy-z.t000
Rotation Axis (black): $x = 0.000$
$y = 0.000$
$z = 1.000$

## Single Spectra of Rotations



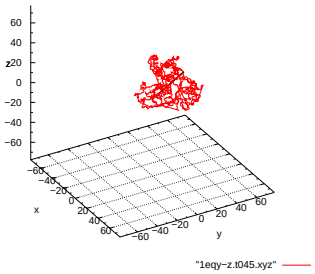

## 3D View

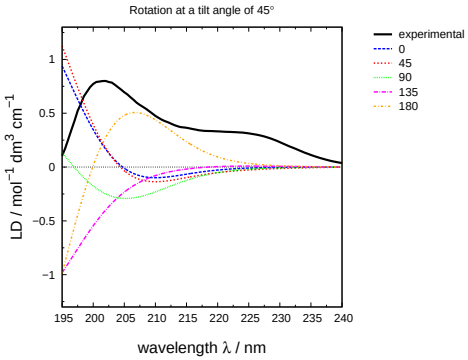


# Tilt Angle 45°

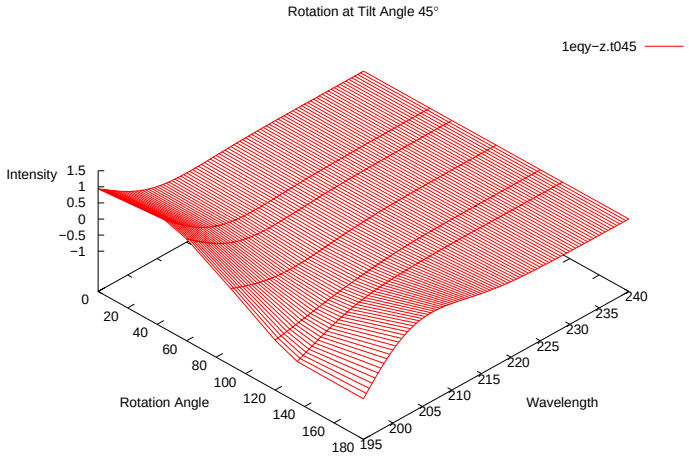
Directory: 1eqy-z.t045
Rotation Axis (black): $x = 0.707$
$y = 0.000$
$z = 0.707$

## Single Spectra of Rotations



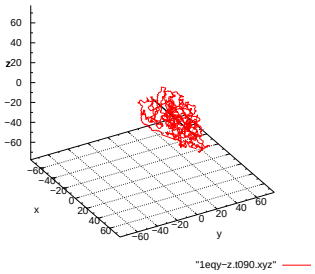

## 3D View

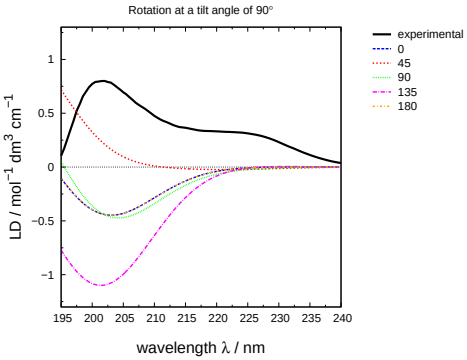


# Tilt Angle 90°

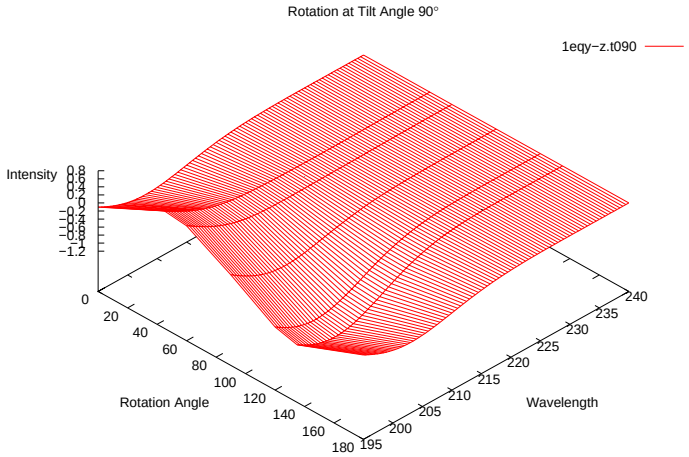
Directory: 1eqy-z.t090
Rotation Axis (black): $x = 1.000$
$y = 0.000$
$z = 0.000$

## Single Spectra of Rotations




## 3D View




Made on a Mac