Theoretical Study of the Circular Dichroism Spectroscopy of Proteins

Zhuo Li

A thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

September 2017

Abstract

Circular dichroism (CD) spectroscopy is an important technique in studying protein structure, especially for protein secondary structures and conformational changes during biological processes. A fully quantitative theory of the relationship between protein conformation and optical spectroscopy would facilitate deeper interpretation and insight into biophysical and simulation studies of protein dynamics and folding. Vibrational structure in the electronic CD spectra of proteins is an important source of information on protein conformation and can be exploited to study structure and folding. We employ the state-averaged complete active space (CAS) method to calculate the *ab initio* electronic ground and excited states of N-methylacetamide (NMA), toluene, *p*-cresol and 3-methylindole (3-MI), which represent chromophores that are significant in the CD spectroscopy of proteins in the far- and near-ultraviolet (UV) regions. The results of these calculations are used to incorporate vibronic levels of the excited states into first principles calculations of CD using an exciton approach.

The far-UV CD spectra of a set of 49 proteins, comprising a range of structural types, are calculated to assess the influence of the vibrational structure. The calculated spectra of α-helical proteins are better resolved

using the vibronic parameters and correlation between the experimental and the calculated intensity of less regular β structure proteins improves over most wavelengths in the far-UV. No obvious improvement is observed in the calculated spectra of regular β-sheet proteins. The near-UV CD spectra of 40 proteins are calculated with the new parameter set and the correlation between the computed and the experimental intensity from 270 to 290 nm is much improved. The contribution of individual chromophores to the CD spectra has been calculated for several mutants and in many cases helps rationalize changes in their experimental spectra. Considering conformational flexibility by using families of NMR structures leads to further improvements for some proteins and illustrates an informative level of sensitivity to side chain conformation. In several cases, the near-UV CD calculations can distinguish the native protein structure from a set of computer-generated misfolded decoy structures.

CD spectra of proteins are better reproduced in both far- and near-UV by considering vibrational structures in electronic transitions of chromophores. This improvement can provide more details in connecting the spectroscopic data to the conformations of proteins and will encourage a broader use of CD in protein studies. Besides shedding light on the importance of vibronic transition, results in this thesis also show other aspects that may further improve CD calculations, such as developing

parameters of disulfide bond, calculating CD using molecular dynamics (MD) trajectories and taking into account the influence of surroundings of chromophores.

Acknowledgements

First of all, I would like to thank my supervisor, Professor Jonathan Hirst, for his help and guidance. It is hard to keep full power on research throughout four years. Jonathan was very patient with me and kept the whole journey smooth.

I am grateful to Dr. Benjamin Bulheller, a former PhD student in Hirst group. I did not meet him in person, but his work, DichroCalc, was used every day. The well-written and organized C^{++} code truly made my life easier.

I thank Dr. David Robinson and Dr. Fouad Husseini for teaching me the calculation skills which are important in finishing this thesis. Dr. John Baker, Pritesh Tailor, Peter Morgan-Tansley and Dr. Colin Bannister are thanked for their supporting in solving problems I met on computers. I also want to thank former and current group members, Sarah Jasim and Dr. Hainam Do, for both scientific discussions and small talk.

Thanks to the University of Nottingham for funding my PhD study and for providing such a beautiful campus to study in. The walk between home and office is the most enjoyable and relaxing time every day. I also want to

thank the computing time provided by our local cluster, university High Performance Computing facilities and also the new launched HPC Midlands Plus, Athena.

I have been fortunate to meet a lot of friends both from School of Chemistry (Dr. Guping He, Dr. Jie Ke, Dr. Xuezhong Sun, Zhimin Liu, Jing Jin, Lingqiao Wu and Xue Wu) and as my roommates (Boyu Zheng, Wei Su, Dr. Xiaoling Kang, Zishang Zhu, Xianfeng Wu and Haoyu Liu). Dr. Yanyi Sun and Runqi Liang are by my side since the first day I arrived in Nottingham and sharing my laughs and tears.

I am grateful to Dr. Hao Feng for appearing in my life.

At last, special thanks go to my parents, for their endless support during my PhD study and also for their love since the very beginning of my life. They are always there for me and trying their best to support my dreams. I would like to dedicate this thesis to them.

Referred Publications and Presentations

Publications

- **Li, Z.** and Hirst, J.D., Quantitative first principles calculations of protein circular dichroism in the near-ultraviolet. *Chem. Sci.*, **8**, 4318–4333 (2017).

- **Li, Z.**, Robinson, D. and Hirst, J.D., Vibronic structure in the far-UV electronic circular dichroism spectra of proteins. *Faraday Discuss.*, **177**, 329–344 (2015).

Oral presentations

- 2017.5 High Performance Computing User Conference, Nottingham

- 2017.4 1st Faraday Joint Interest Group Conference, Coventry

Contents

Abstract i
Acknowledgements iv
Referred Publications and Presentations vi

List of Figures ix
List of Tables xi
List of Abbreviations xii

Chapter 1 Introduction 1
1.1 Principle of spectroscopy 1
1.2 Proteins 11
1.3 Calculation of circular dichroism 25
1.4 Outline of thesis 34

Chapter 2 Methodology 36
2.1 Introduction 36
2.2 The matrix method 37
2.3 Franck-Condon calculation 42
2.4 Statistical 52

Chapter 3 Calculations of Protein Circular Dichroism in the Far-Ultraviolet 56
3.1 Introduction 56
3.2 Methods 59
3.3 Results and Discussion 61
3.4 Conclusions 75

Chapter 4 Calculations of Protein Circular Dichroism in the Near-Ultraviolet 78
4.1 Introduction 78
4.2 Methods 81
4.3 Results and Discussion 88

4.4 Conclusions 128

Chapter 5 Conclusions..... 129

References 138

Appendix A Calculated Geometries and Vibrational Frequencies Assignment for Toluene, *p*-Cresol and 3-Methylindole 163

Appendix B Experimental and Calculated Protein Far-UV CD Spectra..... 178

Appendix C Experimental and Calculated Protein Near-UV CD Spectra... 189

List of Figures

Figure 1.1 Principle of circular dichroism 4

Figure 1.2 Potential energy diagrams of a vertical excitation in the electronic transition 9

Figure 1.3 Model structure of L-amino acid, side chains of aromatic amino acids and disulfide bond 13

Figure 1.4 Fragment of protein structure, peptide bond 14

Figure 1.5 Structures of various protein secondary structures.....15

Figure 1.6 Distinctive CD spectrum features of different protein secondary structures in the far-UV 21

Figure 2.1 Electrons are partitioned into core, acive and virtual orbitals in CASSCF method 45

Figure 3.1 Molecular orbitals of amide structure relevant to electronic transitions in the far-UV region..... 58

Figure 3.2 Calculated geometries and C-N stretching mode of N-methylacetamide 62

Figure 3.3 Absorption spectra of N-methylacetamide 64

Figure 3.4 Experimental and calculated CD spectra of myoglobin.....67

Figure 3.5 Mean absolute errors of calculations without and with vibronic structure for proteins of different secondary structures 68

Figure 3.6 Experimental and calculated CD spectra of deoxyribonuclease-1.....70

Figure 3.7 Experimental and calculated CD spectra of concanavalin A and elastase.....71

Figure 3.8 The Spearman rank correlation coefficient between the experimental and calculated intensities for the non-vibronic and vibronic parameters for proteins of different secondary structures..... 74

Figure 3.9 The Spearman rank correlation coefficient and mean absolute error of 49 proteins in the dataset. 75
Figure 4.1 The Franck-Condon active vibrational modes of toluene.. 91
Figure 4.2 The Franck-Condon active vibrational modes of *p*-cresol. 94
Figure 4.3 The Franck-Condon active C-C stretch mode and benzene C-H bend mode of 3-methylindole..... 96
Figure 4.4 Experimental and calculated near-UV CD spectrum of adenylate kinase..... 99
Figure 4.5 Root mean square error of the intensities between the experimental and calculated near-UV CD spectra computed with and without vibrational structrue for 40 proteins in the dataset..... 100
Figure 4.6 Spearman rank correlation coefficients between the calculated (with and without vibrational structrue) and experimental intensity for 40 proteins. 101
Figure 4.7 Experimental and calculated near-UV CD spectra of calmodulin, subtilisin BPN', insulin and odorant binding protein..... 103
Figure 4.8, Difference spectra (wild-type minus the mutant) of barnase 110
Figure 4.9 Experimental and calculated CD spectra of wild-type bovine pancreatic trypsin inhibitor and three mutants..... 113
Figure 4.10 Difference spectra (wild-type minus the mutant) of RNase A114
Figure 4.11 Difference spectra (wild-type minus the mutant) of human carbonic anhydrase II 116
Figure 4.12 Experimental and calculated spectra with vibrational parameters from the X-ray structure and the NMR structures of apolipophorin III and staphylococcal nuclease 119
Figure 4.13 Spearman rank correlation distributions of the native and the decoy structures..... 127

List of Tables

Table 3.1 Proteins considered in the far-UV study. 60

Table 3.2 Ground and excited state geometries of N-methylacetamide 62

Table 3.3 Franck-Condon overlap integrals of the $\pi_{nb}\pi^*$ transition in N-methylacetamide 65

Table 3.4 Spearman rank correlation coefficients between the calculated and experimental intensity for β-II proteins 73

Table 4.1 Proteins considered in the near-UV study 83

Table 4.2 Calculated vibrational frequencies of Franck-Condon active modes of toluene, *p*-cresol and 3-methylindole. 89

Table 4.3 Transition energies and Franck-Condon integrals of toluene, *p*-cresol and 3-methylindole 92

Table 4.4 Distances and angles between transition dipole moments of residue pairs with strong interactions of barnase. 111

Table 4.5 Proteins studied in decoy calculations. 123

List of Abbreviations

1D	One dimensional
2D	Two dimensional
BPTI	Bovine pancreatic trypsin inhibitor
CASSCF	Complete active space self-consistent field
CASPT2	Complete active space second-order perturbation theory
CSFs	Configuration state functions
CI	Configuration interaction
CIS	Configuration interaction singles
CD	Circular Dichroism
DFT	Density functional theory
DNA	Deoxyribonucleic acid
DSSP	Definition of Secondary Structure of Proteins
ECD	Electronic circular dichroism
GTOs	Gaussian–type orbitals
HCAII	Human carbonic anhydrase II
HF	Hartree-Fock
HPLC	High performance liquid chromatography
ICD	Induced circular dichroism
IR	Infrared
3-MI	3-Methylindole
MAE	Mean absolute error
MB	Methylene blue
MCD	Magnetic circular dichroism
MD	Molecular dynamics
MO	Molecular orbital
NDI	Naphthalenediimide

nm	nanometer
NMR	Nuclear Magnetic Resonance
NMA	N-methylacetamide
ORD	Optical rotatory dispersion
PCDDB	Protein circular dichroism data bank
PDB	Protein Data Bank
Phe	Phenylalanine
QM	Quantum mechanical
RNase A	Ribonuclease A
RMSE	Root mean square error
RNA	Ribonucleic Acid
SCF	Self-consistent field
S/N	Signal-to-noise
SNase	Staphylococcal nuclease
SRCD	Synchrotron radiation circular dichroism
TDDFT	Time-dependent density functional theory
TDM	Transition dipole moment
TRCD	Time-resolved circular dichroism
Trp	Tryptophan
Tyr	Tyrosine
UV	Ultraviolet
VCD	Vibrational circular dichroism
VUV	Vacuum ultraviolet

Chapter 1

Introduction

1.1 Principle of spectroscopy

1.1.1 Principle of absorption spectroscopy

When a beam of light passing through some matter, photon will interact with the molecule in different ways, such as being absorbed, emitted or scattered. Absorption is the simplest form of spectroscopy. The molecule increases its energy by ΔE when it absorbs a photon of frequency ν (eqn. 1.1).

$$\Delta E = h\nu \qquad (1.1)$$

where h is Planck's constant. A photon of UV light, the energy range under consideration in this thesis, will cause the rearrangement of electron distribution in a molecule causing an electronic excitation. The direction and intensity changes of net linear displacement of charge are characterized by the transition dipole moment (TDM), whose length is proportional to the

square root of the absorbance *A*. The absorbance can be calculated by the Beer-Lambert law (eqn. 1.2):

$$A = \varepsilon Cl \qquad (1.2)$$

where ε is the extinction coefficient while *l* is the path length of the light through the sample whose concentration is *C*.

An absorption signal indicates the probability of a transition occurring at the measured energy. We can observe fine structure of simple molecules in the gas phase. The spectrum becomes blurred in solution. The broad envelope of an absorption spectrum without fine structure can be ascribed to: 1. vibrational structures of the electronic states involved in the excitation; 2. different conformations of the same molecule will have slightly different transition energies, as the electron distribution varies, which leads to an averaged spectrum; 3. other effects such as the uncertainty principle or solvent effects.

1.1.2 Principle of circular dichroism spectroscopy

In 1811, Arago discovered the phenomenon of optical activity by observing a spectrum of coloured images when he rotated a polarizer or

a calcite analyzer with a quartz plate inserted in between.[1] Boit, Arago's colleague, showed that this effect is caused by the rotation of the polarization plane after light passed through the quartz plate.[2] He also showed the existence of both dextrorotatory and laevorotatory quartz based on the different rotation directions caused by them. Three years later, the study of optical activity was extended to the liquid state.[3] Pasteur, after over three decades, associated optical activity of matter with its structural dissymmetry.[4] He separated two types of crystals of tartaric acid mechanically under a microscope. These crystals share mirror images of one another while their solutions gave equal but opposite rotations of polarized light. In 1874, Le Bel[5] and van't Hoff[6] announced the theory relating optical activity to an asymmetric carbon atom, *i.e.*, one that adopted a tetrahedral shape and was bonded to four different groups, which laid the foundation of stereochemistry. The interaction of optical activity with molecular structure led to a vast range of studies of chiroptical properties, from synthetic organic compounds to natural products, and also encouraged the development of new techniques, such as optical rotatory dispersion (ORD) and circular dichroism (CD) spectroscopy.

Linear polarized light can be decomposed into two equal magnitude circular polarized components (Figure 1.1). CD is the phenomenon that

optically active chiral molecules absorb non-equivalently left and right circularly polarized light. Namely, the extinction coefficients for the left and right circularly polarized light are different ($\varepsilon_L \neq \varepsilon_R$). The difference $\Delta\varepsilon = \varepsilon_L - \varepsilon_R$ is circular dichroism.

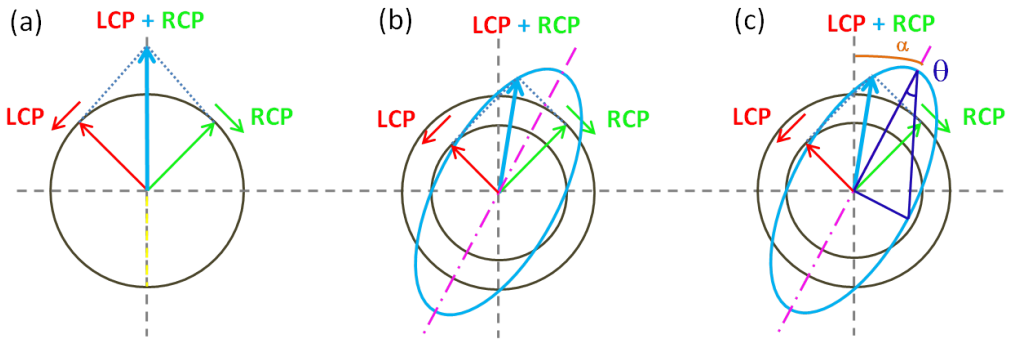


Figure 1.1 (a) Linear polarized light resolved into two circular polarized components, LCP and RCP. The resultant of LCP and RCP whose intensities and phases remain the same will oscillate in a plane. (b) The LCP is absorbed more (and so is less intense) than the RCP, the summation of LCP and RCP describes an elliptical path (blue trace). (c) The angle α corresponds to the rotation of the major axis of the ellipse, which is the optical rotation. The angle θ, which is the angle opposite the semiminor axis in the triangle formed by the semimajor and semiminor axes of the ellipse, is proportional to the circular dichroism.

CD-based techniques and their applications

Electronic CD (ECD) is one of the simplest spectroscopic techniques in the structural analysis of chiral molecules. It finds many applications. Researchers working on small molecules use it to determine the absolute configurations, while those dealing with macromolecules investigate conformations with ECD. The application of ECD spectroscopy in synthetic and natural products is beyond the scope of this thesis. Thus, experimental applications to these compounds will not be covered. Recently, ECD spectroscopy has been used in characterizing structures of nanostructure, e.g. twisted bilayer graphene[7] and single-wall carbon nanotubes.[8] However, the physical origin of CD is not the same as the CD discussed in this thesis.

ECD spectroscopy is now probably more widely used to study biological systems than the non-biological systems mentioned above. The ECD signal is a sensitive probe of conformational changes in biological molecules. Therefore, ECD spectroscopy is used to study processes such as interactions between biomolecules (e.g. protein-protein, protein-deoxyribonucleic acid (DNA), protein-ligand and DNA-ligand interactions, etc) or determine thermodynamics or kinetics properties of biomolecules. The biological systems of interest are usually

very large and many conclusions gained from CD studies are often qualitative or empirical. Studies focused on proteins will be reviewed in the following section, as the CD of proteins has been heavily studied in this thesis. Another essential biomolecule whose structure has been widely studied with CD spectroscopy is nucleic acids. The chromophores in DNA and RNA are the bases (adenine, guanine, cytosine or thymine (for DNAs or uracil for RNAs)), which are planar themselves while the chirality originates from the ribose sugar unit of the backbone. The usage of CD in DNA and RNA includes determination of GC contents by analysing CD signals appearing at various wavelengths with different intensities[9-12], studying non-classical conformations or modified oligonucleotides for potential therapeutic use[13-20], and providing valuable information on the binding mode[21,22] (single or multiple binding sites), stoichiometry or DNA conformational changes[23-25] in interaction studies between DNA and ligand, DNA and DNA or DNA and protein, etc.

When studying biomolecules, not only their structures but also the time evolution of their conformations is essential in understanding their functions and biological processes in which they are involved. CD spectroscopy has inherent potential as a sensitive method in fast time-resolved measurements. The time resolution of CD spectroscopy was millisecond in the 1950s and 1960s when CD became more common

than ORD with the development of cells and photoelastic modulators.[26] The time regime was improved to nanosecond in 1985 when an ellipsometric approach was introduced for time-resolved CD (TRCD) measurements.[27] The time resolution in the far-ultraviolet (UV) region, the peptide bond absorbing range, was improved from the order of 10^{-7} s to 10^{-9} s when an upconverted Ti:sapphire laser was used instead of xenon flashlamps.[28] TRCD has been demonstrated with studies of heme proteins myoglobin and hemoglobin involving photolysis[27] and ligand binding processes,[29,30] early events of cytochrome *c* folding[31] and mechanism of photo-cross-linking reaction in transfer ribonucleic acid (tRNA).[32]

There are other CD-based techniques available in structure determination such as magnetic CD (MCD) in which the signal is induced by placing the sample, usually achiral, in a magnetic field. The most common use of MCD is to provide interpretation of how many transitions are involved in one absorption envelope, because MCD signals with different signs may indicate different polarizations of transitions. MCD spectra are widely used in studying bioinorganic systems[33], such as iron-sulfur proteins[34], porphyrins and heme proteins[35,36] and non-heme ferrous enzymes[37]. Vibrational CD (VCD) is the analogue of ECD for studying vibrational transitions in a molecule[38,39]. The usage of VCD in solving

stereochemical problems enjoys the advantage of impressive agreement between the experimental measurement and the developed theoretical methodology. VCD only detects vibrational transitions in the electronic ground state, which is easier to calculate compared to ECD (accurate electronic excited state calculations are needed) from a quantum-mechanical point of view. Some biologically related molecules have been calculated for a variety of conformations in different solvent models and compared to their VCD spectra, such as dipeptides with different secondary structures constraints[40], zwitterionic L-alanine[41,42] and N-acetyl-alanine N'-methylamide.[43] There are also X-ray CD[44] which studies transitions localized near the core-excited atom; and nuclear spin-induced CD[45] which can provide information about nuclear magnetization. These techniques share different physical origins as conventional CD, which bring some new directions of using CD in structure determination. In this thesis, we focused on studying ECD only and CD is used to represent ECD spectroscopy specifically.

1.1.3 Franck-Condon principles

In an absorption spectrum, the transitions from the electronic ground state v"=0 to the vibronic excited state form a progression (Figure 1.2). Although all values of v' are allowed by the selection rule, the intensity

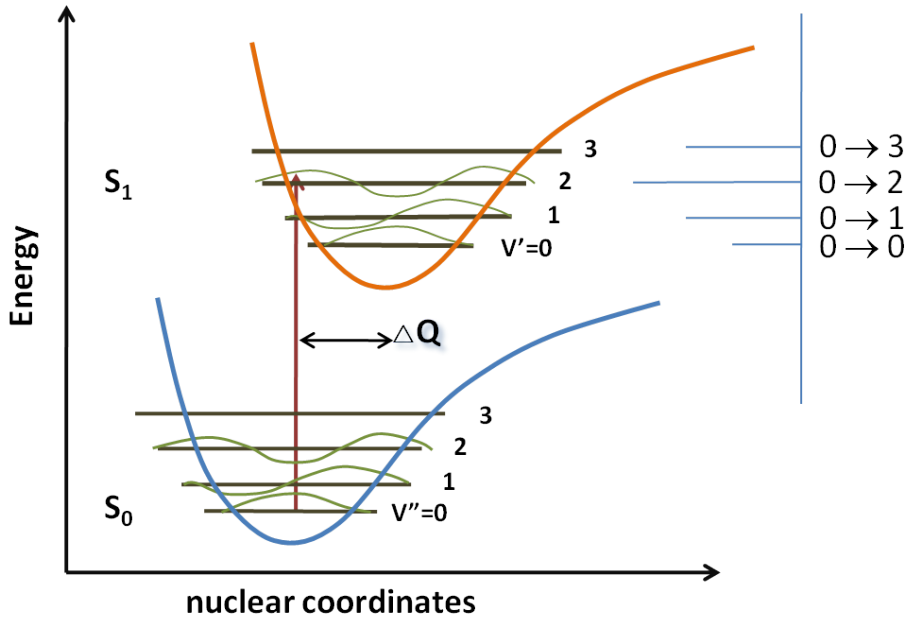


Figure 1.2 Potential energy diagrams of a vertical excitation from the lowest vibrational level in the ground electronic state (S_0) to the third vibrational level in the excited electronic state (S_1). The horizontal lines labeled v = 0, 1, 2, 3 represent the vibrational levels within each state. The curves associated with these horizontal lines represent the vibrational wave function. The potential energy curve is a function of molecular geometry. ΔQ indicates the nuclear movements along the excitation. The possibility of appearance of an electron in which vibrational level in the excited state is represented by the blue lines on the right and governed by the overlap of the wave function of each vibrational level in the ground and excited electronic states.

distribution along the progression is governed by the Franck-Condon principle.[46,47] An electronic transition takes place much more rapidly than the vibrational motion due to the slower reaction of nuclei compared to the instantaneous motion of electrons because electrons are much lighter than nuclei. Consequently, the greatest intensity is excited from the transition starting from the lowest vibrational energy level of the ground state to a vibronic level of the excited state with static atom coordinates between two states. This is called the Franck-Condon principle. The intensity distribution along the progression is proportional to the Franck-Condon factors, and the Franck-Condon factor is equal to the overlap integral of vibrational wave functions of the ground and excited states.

Vibrational fine structure is common in high-resolution gas phase spectra. These fine features may be not obvious in solution, but can contribute to the broadening of bands. The Franck-Condon effect has been considered in previous calculations of the CD spectra of various molecules. The CD spectrum of dimethyloxirane in the gas phase was calculated using TDDFT including the envelopes of several states[48]. Some narrower bands arise from excitations to the 1B, 2A and 3B states, while other broader bands were assigned to transitions to the 4A and 5A states. The CD spectrum was reproduced well once the vibrational structure for the various transitions was considered. The disappearance of a negative band in the experimental

spectrum was rationalised by the calculation: the band was cancelled by other positive bands because of a strong vibrational progression. Vibronic contributions were taken into account in calculations of the CD spectra of NDI dimers[49] and oligomers[50]. The CD spectra of NDI helical nanotubes were calculated with the matrix method considering several vibronic transitions associated with two electronic transitions. Inclusion of the Franck-Condon effect was essential in reproducing the key features of the CD spectra.

1.2 Proteins

1.2.1 Protein structures

As one of the most important biomolecules, the overall shape of a protein molecule is crucial for its biological activity. Thus, any technique that can provide structural information is potentially valuable for understanding proteins' function. Hlasiwetz and Habermann first recognized the units that compose proteins are amino acids in 1873 when they hydrolysed casein with strong acids or alkali.[51] They also found that hydrolysis of different proteins yielded different products. It is difficult to tell exactly how many proteins exist. Despite a wide range of molecular mass and an enormous functional diversity proteins are essentially built up of 20 amino acids. 18 amino acids consist of carbon, hydrogen, nitrogen and oxygen while cysteine and

methionine also have sulfur. The structure of an amino acid (Figure 1.3a) contains a chiral centre, the α-carbon, bonded with a carboxyl group, an amino group, single hydrogen atom and a side chain group (R group), except proline (Figure 1.3b). All natural amino acids are in L absolute configuration, which means they are optically active, except glycine whose side chain is also hydrogen. The side chain of proline forms a pyrrolidine ring bonding to the amide nitrogen. The different properties of amino acids come from the R group. The R groups of particular interest in this thesis are the aromatic side chain of phenylalanine, tyrosine and tryptophan, which are benzene, phenol and indole, respectively (Figure 1.3c). Some discussions in this thesis also touch on the disulfide bond which is formed between two cysteine side chains (Figure 1.3d).

Another milestone in understanding protein structures occurred in 1902 when Hofmeister and Fischer independently proposed models of the peptide bond which connects two amino acids.[52,53] A peptide bond is an amide bond formed by a water elimination reaction between the amino group of one amino acid and the hydroxyl group of the other (Figure 1.4). The peptide bond is planar and the C-N bond length in peptides is shorter than an ordinary C-N bond (1.32 Å compared to 1.45 Å).[54] A peptide bond possesses partial double bond character, due to two resonance states in which the double bond appears at between C=O and C=N. As a result of less mobility

around the C-N bond, two conformations are available, namely two consecutive α-carbons are in the *trans* or *cis* form; the *trans* conformation is more favorable. The primary structure of a protein is the sequence of amino acids joined together via peptide bonds.

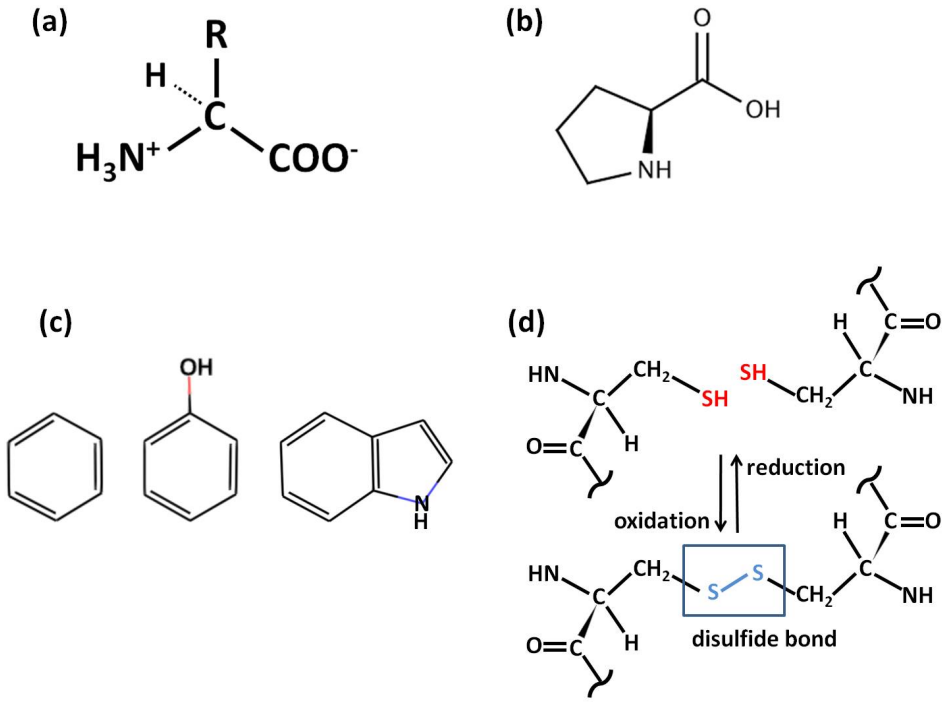


Figure 1.3 (a) Model structure of L-amino acid showing in zwitterionic state, a carbon atom connected by an amino group, a carboxyl group, one hydrogen and a side chain represented by R; (b) structure of L-proline; (c) structures of benzene, phenol and indole, hydrogen atoms are not shown; (d) disulfide bond formation between side chains of two cysteines.

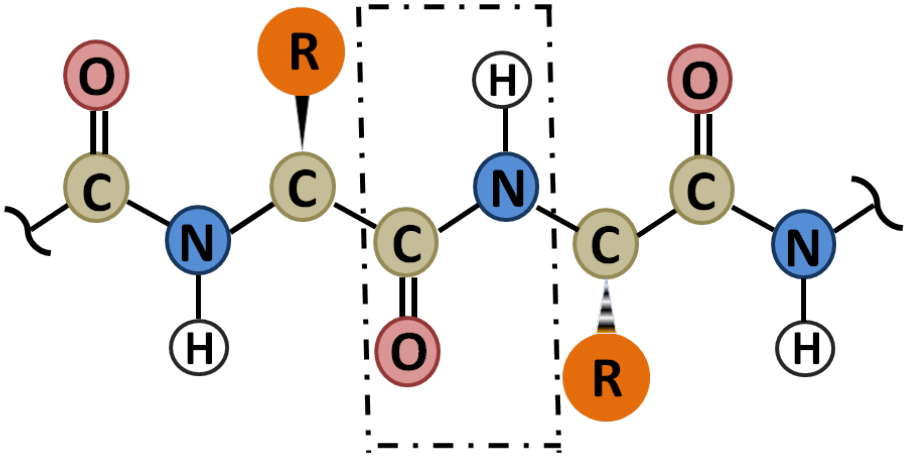


Figure 1.4 Fragment of protein structure. C, N, O, H represent carbon, nitrogen, oxygen and hydrogen atoms, while R represents the side chain of each amino acids. The four atoms in the rectangle form the basic structure of the peptide bond, the backbone structure of a protein.

The formation of three-dimensional secondary structure from amino acid sequence leads to the functional properties of proteins. There are three basic kinds of regular secondary structure: α-helix, β strand and turns (Figure 1.5). The structure of α-helix was first suggested from model-building studies by Pauling in 1951[55] and the right-handed conformation was confirmed by the crystallographic structure of myoglobin in 1960.[56] The α-helix structure (Figure 1.5a) accounts for approximate one third of the residues in globular proteins.[57] There are 3.6 residues per turn in the regular α-helix structure

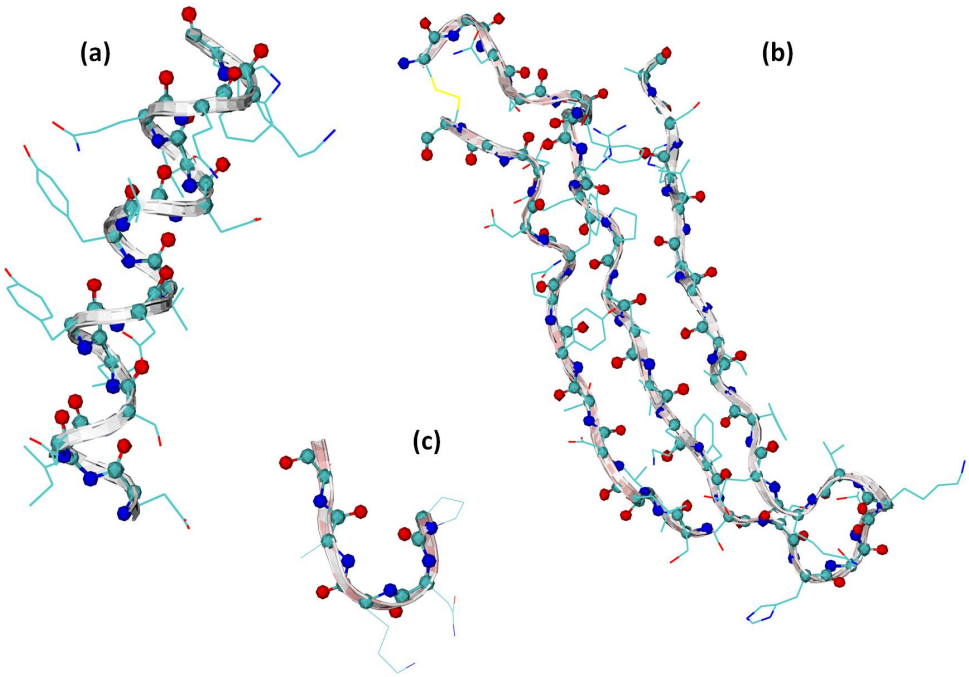


Figure 1.5 Structures of protein secondary structures shown in ribbon diagrams, fragments are from protein Dnase (PDB code: 3DNI): (a) α-helix, residues 140 to 155; (b) β-sheet, also denoted as β-I in this thesis, residues 90 to 133; (c) turn, also denoted as β-II in this thesis, residues 85 to 89. The backbone atoms are represented by ball-and-stick model while the side chain atoms are shown in lines. C, N, O, S atoms are represented by grey, blue, red and yellow, respectively.

which allows formation of backbone hydrogen bonds between the oxygen of one residue (residue *i*) and hydrogen of the other residue four ahead (*i*+4). The backbond hydrogen bond forming between *i* and *i*+3 or *i* and *i*+5 would lead to another two helical structures, namely 3_{10} helix and π helix, respectively. The β strand (Figure 1.5b) is another common motif in proteins. It is the second structure studied by Pauling and Corey through model-building study showing extended conformation.[58] The β strand usually appears as a β-sheet which forms through hydrogen bonding between two or more β strands, because the stability of the structure is significantly increased. There are two kinds of arrangements, parallel and antiparallel, of adjacent strands determined by the directions of the polypeptide chain (N- to C- terminal). Turns (Figure 1.5c) connect and change the direction of β strands. They are named as γ or β turn when three or four residues are involved in forming the structure, respectively. Proteins also have a large amount of random coil structures which do not adopt specific structural features as described above. However, they have distinctive CD spectroscopic signals which were also discussed in this thesis. CD signals of peptides and proteins have been recognized to arise from: (1) exciton coupling between $\pi\pi^*$ transition on two different groups[59]; (2) mixing of $n\pi^*$ and $\pi\pi^*$ transitions from two different groups (μ-m coupling)[60]; and (3) from a single peptide group (one-electron effect)[61].

The overall topology of proteins in their folded state is described by the tertiary structure which defines the spatial arrangements of residues that can be far apart in their primary sequence. Some of the important molecular interactions governing protein structure include disulfide bridges, van der Waals interactions, hydrophobic contacts, hydrogen bonds and electrostatic interactions.

Different techniques and methods have been developed to study protein structures based on various physical principles. The use of electromagnetic radiation at different energy regions may cause a variety of responses. The most common techniques for protein structure determination are X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. Currently, there are over 100,000 X-ray determined protein structures and over 10,000 NMR structures deposited in the Protein Data Bank (PDB).[62] Electron microscopy is also able to visualize the structure of proteins; over 1200 structures are in the PDB, currently. Infrared and Raman spectroscopy provide information of vibrational features of chemical bonds in proteins. These techniques can be supplemented by CD spectroscopy, which has certain advantages compared to them, although CD spectroscopy does not provide atomic level structural information. A CD signal is measured in solution phase. Thus, it is not subject to structural changes which may occur

in the crystallization process. Compared to NMR, CD spectroscopy is an inexpensive technique in both finance and time respects. No high-field instrumentation is required to accumulate high signal-to-noise (S/N) CD spectrum, which may provide qualitative evidence of interested protein events at early stage of researches. CD measurements can study fast changes of local protein structure, such as electron transfer and bond breakage/formation while NMR is limited to pico- to nanosecond protein motions through mechanisms such as nuclear spin relaxation or residual dipolar coupling[63] (10^{-15} s for the absorption of a UV photon compared to 10^{-9} s for radio waves in NMR[64]); a short timescale may also be useful for protein-ligand interaction studies if the ligand has a particularly short residence time[65]. Although no atomic level data can be revealed, changes of bands are easy to monitor in CD measurements; NMR, in other words, provides large amount of information hidden in overlapped signals, which need more effort to interpret. Also, there is no size limitation of proteins in CD spectroscopy while NMR is typically used to determine the structure of proteins with molecular weight smaller than 100 kDa.

1.2.2 Circular dichroism of proteins: principles

The information contained in a CD spectrum depends on the chromophore that absorbs light at the wavelength being measured.

Conventional CD spectroscopy can record excitation bands in the UV region which contain the asymmetric features of the protein backbone in the far-UV (180 – 240 nm) and the orientations of the aromatic side chains of amino acids in the near-UV (250 – 300 nm). The spectrum in the vacuum UV (VUV) can be measured down to 160 nm (limited by absorption by water[66]) through the availability of synchrotron radiation CD (SRCD) to uncover further information on protein structure.[67] The spectral range can be extended down to 130 nm by eliminating water, e.g., protein films or for samples in non-aqueous solvent.[68,69] The bands in this region arise from the charge-transfer transitions between neighbouring amide bonds. Some metallo-proteins which have extrinsic chromophores may also have transitions in the visible region.

The CD signal in the far-UV region mainly corresponds to the amide structure of the protein backbone. Two electronic transitions of the peptide bond are located in this energy range, which are from the amide non-bonding π orbital (π_{nb}) to the anti-bonding π orbital (π^*) ($\pi_{nb}\pi^*$ transition)[70,71] and from the oxygen lone pair n orbital to the π^* orbital ($n\pi^*$ transition)[71,72] with extinction

coefficients $\varepsilon \sim 7000$ and 100 mol^{-1} dm^3 cm^{-1}, respectively. Some side chains of the amino acids also absorb light in this area; however, their signals are covered by the peptide absorption envelope since the amide chromophores are in excess. Different secondary structural elements of proteins induce bands of distinctive shapes and magnitudes (Figure 1.6) in the far-UV region through transitions of the peptide chromophores. For an α-helix, there is an intense positive band at 190 and a negative band at 208 nm ($\pi_{nb}\pi^*$). There is another band arising from $n\pi^*$ transition located at around 222 nm. For other secondary structure types, the bands are more difficult to characterize than for the α-helical proteins. The general features of the β-sheet proteins (denoted as β-I) are a negative band at about 216 nm and a positive band of comparable magnitude near 195 nm.[73] The overall shape of a β-sheet CD spectrum is affected by the parallel/antiparallel conformation with various lengths and widths and also influenced by the twist of the sheet structure[74, 75]. An additional negative band near 175 nm was observed in the VUV region.[76] The random coil (denoted as β-II) has a strong negative signal just below 200 nm. The spectra of β-I and β-II have one of their extrema at similar wavelengths but with opposite sign, which may cause incorrect weighting of these structures in empirical fitting methods. The protein CD spectra in the far-UV are routinely analyzed as a combination of standard spectra of each secondary motif.[77-82]

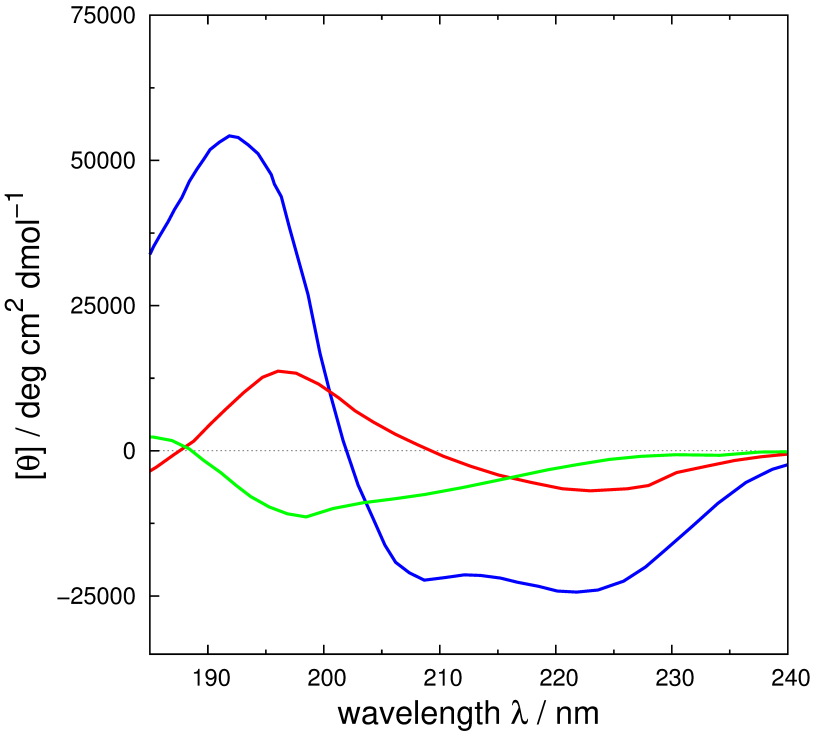


Figure 1.6 Distinctive CD spectrum features of different protein secondary structures: α-Helical structure (blue line) has an intense positive band at 190 and two negative bands at 208 nm and 222 nm; the maximum in the spectrum of the β-sheet (red line) protein appears at about 216 nm and near 195 nm with comparable magnitude but opposite sign; the random coil (green line) has a strong negative signal just below 200 nm.

The spectra in the near-UV arise from the $\pi\pi^*$ electronic excitations on the aromatic rings of tryptophan, phenylalanine and tyrosine which can reflect the conformation of aromatic side chains as well as their surrounding environment. The side chains of aromatic amino acids play an essential role in both protein structure and function. The CD signal in the near-UV is much weaker than that in the far-UV. The extinction coefficient of indole is ε (279 nm) ~5000 mol^{-1} dm^3 cm^{-1} for two or more transitions while only one transition appears in this region for both tyrosine and phenylalanine with extinction coefficient ε (274 nm) ~ 1400 mol^{-1} dm^3 cm^{-1} and ε (258 nm) 190 mol^{-1} dm^3 cm^{-1} for phenol and benzene, respectively. However, the intensity of near-UV CD is limited by the low number (relative to the number of peptide bonds) of aromatic amino acids in proteins.

1.2.3 Circular dichroism of proteins: applications

CD spectroscopy has been used in structural determination of protein since the 1960s. One of the most common applications in structural biology is the determination of secondary structure composition of proteins using the far-UV spectrum through a wide range of empirical algorithm and reference databases.[77-82] Some of these methods are in standard use by experimentalists and provide accurate results especially for highly helical structures.

CD measurements of some designed structures are able to reveal more structure-spectroscopic relationships. The blocked hepta-peptide, BOC-$(X)_7$-OMe, has a shoulder with a higher energy than the main positive band and a crossover at 178 nm in the CD spectrum where *X*=alanine, but there is no such shoulder for the valine heptamer and the crossover is at 192 nm. This is interpreted as formation of parallel and antiparallel β-sheet structures for different oligopeptides.[83]

CD spectroscopy has undergone a major expansion since protein folding attracted more attention in molecular biophysics in the late 1980s. The aromatic amino acids can form key structures with themselves or with other amino acid side chains in the protein folding process[84]. The intermediates of unfolding processes usually show similar far-UV CD spectrum but less intense near-UV CD signals compared to the native protein. This indicates a state where the protein has lost some of its tertiary structure but retains much of the secondary structure.[85,86] The aromatic structures tend to be buried in the interior of proteins as part of the hydrophobic core[87]. The CD spectrum of bovine pancreatic ribonuclease A (RNase A) in the near-UV arises from six tyrosine, three phenylalanine and four disulfide bonds. By analyzing the wavelength of each 0-0 transition, it was concluded that one tyrosine is buried, three tyrosine are exposed while the other two tyrosine

are partially buried based on the energy-shift caused by the solvent effect.[88, 89] Aromatic amino acids are also essential for the function of proteins, such as molecular recognition[90,91] and catalysis. The CD signal can also be used in both the far- and near-UV to follow the structural changes caused by site-directed mutagenesis. The spectroscopic variation is from either the conformational changes, locally or globally; or from the contribution of individual residues. The difference spectrum (wild-type minus mutant) shows the CD spectrum of the mutated residues if the overall protein conformation is not changed which can be observed through far-UV CD spectrum.[92-95] The CD spectrum of single residue makes a good probe for the environment of the chromophore.

Proteins containing prosthetic groups exhibit CD bands above 300 nm. In many cases, the ligand is either achiral or chiral but has weak CD (e.g. nicotinamide adenine dinucleotide and flavin adenine dinucleotide). However, the CD of the ligand is often quite strong when it binds to the protein, which makes CD spectroscopy a valuable tool in studying protein-ligand binding.[96] A group of chromophores may be optically inactive because the composition of enantiomers leads to a racemic mixture. But the host protein will choose one enantiomer over the other and this protein-ligand binding show optical activity of the ligand while non-bonded molecules remain racemic.[97-99]

As a higher energy light source used, SRCD provides structural information in a wider wavelength range than conventional CD,[67] which means more transitions are measured and the spectra contain more information. CD spectra in the VUV also show more distinguishable features for various secondary structures of proteins. Thus, SRCD can provide more accurate secondary structure composition.[66,100] SRCD needs even less concentrated protein samples as a result of the improved S/N ratios from using a more intense light source. The higher S/N ratio also makes the stopped-flow/SRCD apparatus extensible to a faster time range.[101,102]

1.3 Calculation of circular dichroism

1.3.1 CD calculations: principles

The CD strength or the rotational strength of an excitation is analogous to the oscillator strength of a normal absorption. It is the integral of $\Delta\varepsilon$ over a particular wavelength range for the transition. The rotational strength (R^{0k}) of an excitation from the ground state *0* to the excited state *k* is the product of the electric TDM, $\vec{\mu}$, and the magnetic TDM, $\vec{m}$, given by Rosenfeld equation[103] (eqn. 1.3):

$$R^{0k} = \mathrm{Im}\left(\left\langle \psi^0 \middle| \vec{\mu} \middle| \psi^k \right\rangle \left\langle \psi^k \middle| \vec{m} \middle| \psi^0 \right\rangle\right) \qquad (1.3)$$

where ψ^0 and ψ^k represent the wave functions of the ground state and excited state respectively. *Im* represents the imaginary part of the product. The integral $\left\langle \psi^0 \left| \vec{\mu} \right| \psi^k \right\rangle$ can be understood as a light-induced oscillating dipole and the integral $\left\langle \psi^k \left| \vec{m} \right| \psi^0 \right\rangle$ can be regarded as current loop induced by the incident light beam. The operator of $\vec{\mu}$ is a real vector describing a linear displacement whereas $\vec{m}$ is imaginary vector describing the rotation of charge in a complex coordinate system[104]. The magnetic transition moment $\vec{m}$ involves the linear momentum operator $\vec{p}$:

$$\vec{m} = \frac{e}{2mc}\left(\vec{r} \times \vec{p}\right), \quad \vec{p} = \frac{\hbar}{i}\nabla \qquad (1.4)$$

where *e* is the charge, *m* is the mass of the electron, *c* is the speed of light, $\vec{r}$ is the position operator of the electron, $\hbar$ is Planck's constant divided by 2π, $i = \sqrt{-1}$ and ∇ is the gradient operator.

In practice, a CD spectrum is usually plotted as the mean residue ellipticity $[\theta]$ against the wavelength instead of the rotational strength against the wavelength. The mean residue ellipticity is $[\theta]=3298\Delta\varepsilon$.

As stated, the rotational strength can be calculated with Rosenfeld equation (eqn. 1.3). As a result, any method that can provide the electric and magnetic TDM can be used in principle in CD calculation. CD considers electronic transitions which require the presence of chromophores which means a more or less extended π-electron system. In organic molecules we can distinguish a chromophore from other parts which are not conjugated with the π system. The optical activity is from either a skewed chromophore (intrinsic chirality) or twisted caused by connecting groups.[105] The situation becomes different if we have more than one chromophore in the molecule. In this case the exciton-coupling may override other terms and the structure not conjugated with chromophores may be neglected. The difficulty of calculation for both cases for small molecules would come from the conformational manifolds. CD calculation is meaningful when averaged with weighting factors derived from computed energies for various conformers. CD spectroscopy is mainly used in absolute configuration determination for small molecules. Chemical modifications[106] or homogeneous crystallization can help reduce the conformational freedom and aid the theoretical calculation with a reasonable structure. It is nowadays common practice to perform CD calculation for moderate size molecules with any quantum mechanical (QM) method affording an excited state description. The impact of approximations in different levels of theory in providing an accurate CD calculation is not fully understood. Electron correlation is necessary to be

taken into account, though. Details on the advantages and drawbacks of specific first principles methods for CD calculation are not discussed, as it is beyond the scope of this thesis.

The calculation of protein CD spectra remains a challenge. Approaches[107-111] that can tackle the large size of proteins are needed. The aim of this thesis is to improve the matrix method for CD calculation. Thus, the matrix method will be introduced in detail in the methodology section (chapter 2). Another method that has been used for studying large systems is the dipole interaction method.[107,108] This method is based on classical physics considering chromophores as polarizable groups. The chromophores interact with each other and with light as oscillating dipole moments. In this theory, no mixing of transitions is allowed as the chromophore is affected by the static field of the rest of the system and the interaction described by point dipole is often inaccurate.[112] DeVoe's approach calculated the CD at individual frequencies which means no arbitrary band functions are needed, unlike the matrix method. Another difference is that the dipole-interaction model is equivalent to weak-coupling exciton theory while the matrix method is strong-coupling.[108,112] Recently, a semiempirical QM method has been developed in calculation of biomolecules including oligopeptides, proteins and DNAs without fragmentation.[113, 114] The method is an extension of the simplified Tamm-Dancoff approximated time dependent density functional

theory (TDDFT) with tight binding procedure to allow its usage for large systems. The whole molecule calculated can be treated quantum mechanically in a reasonable time frame using standard desktop computers. The average error, 0.4 eV, is typical of TDDFT theory and this discrepancy can be narrowed down by computing along a molecular dynamics (MD) trajectory.

1.3.2 CD calculation: applications

Alongside the development of the QM methods and the power of computer, the above mentioned methods have been used to compute of the CD of molecules of various sizes to interpret their CD spectra. For example, by comparing with experimental data, the matrix method has been used to determine the absolute configuration of 5H-indolo[1,7-ab][1]benzazepine and coumarin photodimer[115] and aldehyde ester with flexible conformation.[116] Calculations of small molecules aimed at helping to better understand the properties of proteins are discussed in the following section.

Interpretation of spectroscopic data with a theoretical method can greatly improve our understanding of the relationship between the structure of a molecule and its intrinsic optical properties. Moffitt recognized the spectral features arising from in-phase and two out-of-phase coupled

transitions when considering the combination of $\pi\pi^*$ transitions of chromophores in a helical arrangement.[117] He predicted the right-handed helical structure in proteins correctly. Madison and Schellman conducted the first CD calculation on proteins which showed good results for α-helical proteins.[118]

Some simpler molecules were studied as models to help understand the relationship between the conformation of chromophores and their optical activities. A rigid cyclic diamide[119] or other cyclic dipeptides, e.g. cyclo-(L-tyrosine-L-tyrosine)[120] and cyclo-(L-proline-D-tyrosine)[121], can eliminate some degree of conformational mobility. Complete active space self-consistent field (CASSCF) and complete active space second-order perturbation theory (CASPT2) calculations on cyclo-(L-glycine-L-glycine) including a solvent model were compared with experimental infrared (IR), absorption and CD spectra. The results helped to assign the bands to specific transitions.[119] TDDFT calculations combined with MD simulation has been used to study CD and 2 dimensional (2D) UV of cyclo-(L-proline-D-tyrosine) which identified a charge-transfer state between two well reproduced negative peaks.[121] As mentioned in section 1.2.3, far-UV CD spectra of BOC-(alanine)$_7$/(valine)$_7$-OMe have distinguishable features due to different β-sheet conformations.

Manning *et al.* suggested that there is a variation of CD spectra that is attributable to the twist of β-sheets, based on theoretical studies of energy-minimized oligo-alanine and oligo-valine β-sheets.[122] Turn structures make up a quarter of residues in globular proteins, and have been classified into eight types of conformations.[123,124] CD bands have distinct features for different type of β turns.[125] Although the structure is rather well-defined, the α-helix shows various backbone dihedral angles depending on whether it is located at the inner or outer side of globular proteins. The CD band of the ω-helix, which has additional hydrogen bonding to the environment, was predicted to be the strongest by theoretical study.[126] The Soret band and other $\pi\pi^*$ transition of porphyrin in heme proteins have been studied intensively both experimentally[127] and theoretically.[128,129] Calculations using the crystal structure revealed the source of Soret CD to be the coupling between the $\pi\pi^*$ transition on porphyrin and aromatic residues. The CD of another pigment, rhodopsin, arises from the intrinsic chirality of retinal and was investigated by same sort of calculation as above combined with TDDFT.[130]

The calculation of protein CD has been heavily studied in our group and improved through the use of more accurate calculations of chromophores from the first principles.[131-139] Different aspects have been taken into account to improve the accuracy of the calculations, such as (1) hydrogen bonding

with the solvent was considered by using parameters generated from *ab initio* calculation of *N*-methylacetamide (NMA) in a solvent model combining explicit water molecules and a continuum dielectric[140]; (2) contributions of the transitions to the high-lying B states of aromatic side chains have been studied and incorporated into the CD calculations[133,141]; (3) a new algorithm, EHEF, has been combined with the matrix method model which considered the excitation energy fluctuation causing by the surroundings of the chromophore[142]; (4) charge-transfer transitions in VUV region have been incorporated by CASSCF/CASPT2 calculation on different geometries of N-acetylglycin-N'-methylamide, which represent various dipeptide conformations[136,137]. The vibrational features of electronic excited states were considered in the CD calculation of naphthalenediimide (NDI).[50] Additional bands are resolved with relative intensities close to the experiment. This result encourages further study considering the Franck-Condon principle in CD calculations. Multiple vibrational levels in the electronic transitions of NMA[138], toluene, *p*-cresol and 3-methylindole (3-MI)[139] have been calculated and implemented in CD calculations through the matrix method. These improvements form the primary results in this thesis which will be discussed in the chapter 3 and 4.

1.3.3 Dichrocalc

DichroCalc[143] is programmed based on the matrix method, which is a well-established approach for calculating CD spectrum from first principles. A set of small molecules are used to model the electronic excitations of the various chromophores in proteins. Over the years, the parameter sets of DichroCalc have been developed using *ab initio* quantum chemical methods. NMA[144] is used to represent the peptide bond which is responsible for the transitions in the far-UV. Three aromatic molecules (toluene, phenol and indole) were studied to mimic the transitions of the aromatic amino acids[145]. The charge-transfer transitions[146] have been incorporated to calculate the spectrum in the VUV and the associated experimental spectra have been measured using SRCD.

These parameters and the computational methodology have been used by over 1300 different researchers to compute the CD spectra of more than 3.3 million structures via our DichroCalc web-interface[143]. The calculations include all the proteins in the PDB[147], polymers[50] and snapshots from MD simulations. Calculations of amyloid structures[148] with multiple conformations provide a model that takes the conformational diversity in the solution phase into account. The simulated conformations of Mets7 bound and unbound to cis-platin[149] were also calculated to study the important

function of copper influx transporter protein, Ctr1. DichroCalc was also used to aid design mutants of the coiled coil domain of the oncoprotein Bcl-Abl[150]. After incorporation of the charge transfer parameter into DichroCalc, the VUV CD spectrum of human myelin P2 protein has been computed[151] and qualitatively agreed with experimental measurements made using synchrotron radiation. DichroCalc has also been used to study the loss of helicity of peptides binding to the surface of silica[152]. As is evident from the preceding selected examples, DichroCalc is finding utility across a range of challenging problems.

1.4 Outline of thesis

In this thesis, we primarily focused on incorporating the vibrational structure of the excited states of four chromophores to improve the quality of our CD calculations. The conformational diversity is also touched on by computing protein CD with its NMR structures. Currently research is aiming to reveal the information hiding in the structureless band shape. This thesis will address the calculations of CD spectra of proteins in the three subsequent chapters followed by a conclusion chapter. No further detailed background will be introduced in the rest of the text, as most relevant content has been covered in the preceding section. The second chapter describes the theoretical model, the matrix method, in detail and presents

our strategy of implementing the Franck-Condon effect in our CD calculation by extending parameters and the matrix. Chapter 3 and 4 form the predominant part of this thesis and contain the main scientific findings. The *ab initio* calculations on chromophores are described in these two chapters separately based on the energy range being studied (NMA in chapter 3, toluene, *p*-cresol and 3-MI in chapter 4). Detailed calculated geometries and vibrational frequencies are compared with experimental data and other calculations in the Appendix A. Chapter 3 covers theoretical study in the far-UV, while chapter 4 focuses on CD calculations in the near-UV. Two sets of proteins are calculated to illustrate the effectiveness of our new parameters: 49 proteins in chapter 3 and 40 proteins in chapter 4, respectively. Only selected results are shown in the main text, while the rest of the spectra are collected in the Appendices B and C.

Chapter 2

Methodology

2.1 Introduction

Much effort has been made to improve the accuracy of calculations of CD spectra, in part to understand better the relationship between molecular structures and the associated spectra. Fully *ab initio* calculations are feasible and routine for small molecules and can be performed with density functional theory (DFT)[153,154], Hartree-Fock (HF)[155] or coupled cluster theory[156]. High-level calculations have been performed on simple amides and dipeptides[119,132,157]. However, more approximate approaches are necessary for computing the CD of proteins from first principles to cope with the size and the flexibility of the systems. There are several approximate methods for polymeric systems like proteins, such as the dipole interaction model[107-110,158] and the matrix method[59,71,159-161]. The matrix method is a well-established formulation derived from the exciton model, which has proven to be quite successful[144,145]. Exciton and vibronic coupling theories have recently been reviewed in the context of molecular dimers.[162] In the Frenkel

exciton model, the interchromophore charge transfer is omitted and orbitals on different chromophores are not overlapped. There are discussions on the validity of the simplification in this model.[163] A key aspect in the matrix method is the set of parameters used to model the electronic excitations of the various chromophores in proteins and much effort has been devoted to developing parameter sets from *ab initio* QM methods.

2.2 The matrix method

As stated in section 1.3.1, the rotational strength can be calculated through the Rosenfeld equation[103] (eqn. 1.3). Due to the size and the complexity of proteins, it is impractical to obtain the wave functions from *ab initio* calculations. Instead, we calculate protein CD spectra with the matrix method[71,159,160] as implemented in the code DichroCalc.[143] In the matrix method, the proteins are considered as being composed of *M* chromophores. The wave function of protein in its k^{th} excited state, ψ^k, is expressed as a linear combination of electronic configurations, Φ_{ia} (eqn. 2.1). The first summation is over the *M* chromophores while the second one is over the n_i electronic transitions of individual chromophore *i*. The interactions of the states are reflected by the expansion coefficients, c_{ia}^k. In each electronic configuration, only one chromophore *i* is in its excited

state *a,* while the rest of the chromophores are in their ground state *0*, expressed by the wave function of individual chromophore, ϕ_{ia} (eqn. 2.2). No charge transfer transitions are considered in the current calculations.

$$\psi^k = \sum_i^M \sum_a^{n_i} c_{ia}^k \Phi_{ia} \qquad (2.1)$$

$$\Phi_{ia} = \phi_{10} \ldots \phi_{ia} \ldots \phi_{j0} \ldots \phi_{M0} \qquad (2.2)$$

The wave functions ψ^k of the protein and their corresponding energies can be calculated by solving the Schrödinger equation (eqn. 2.3) under the matrix formalism. The total Hamiltonian operator $\hat{H}$ of the protein is constructed with the local Hamiltonian $\hat{H}_i$ of chromophore *i* and the intergroup potential $\hat{V}_{ij}$, between two transitions (eqn. 2.4). The dimension, *M*, of the Hamiltonian matrix will be extended if multiple transitions are taken into account for chromophore *i*.

$$\hat{H}\psi^k = E^k \psi^k \qquad (2.3)$$

$$\hat{H} = \sum_{i=1}^{M} \hat{H}_i + \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \hat{V}_{ij} \qquad (2.4)$$

There are various approaches to exciton-vibrational coupling[164]. Yang[165] described a reduced density–matrix theory for the case where a single vibration mode is predominant in the coupling, with a spectral density used to represent a bath of nuclear degrees of freedom. We take advantage of this model in constructing the system Hamiltonian.

An illustrative Hamiltonian matrix is shown below (eqn. 2.5). Here we consider a two-chromophore (denoted 1 and 2) model with two transitions (denoted *a* and *b*) for each chromophore. Two vibrational levels (denoted v_1 and v_2) of the transition to the excited state *a* are shown.

$$\hat{H} = \begin{pmatrix} E_{1a_v_1} & 0 & V_{1a_v_1;1b} & V_{1a_v_1;2a_v_1} & V_{1a_v_1;2a_v_2} & V_{1a_v_1;2b} \\ 0 & E_{1a_v_2} & V_{1a_v_2;1b} & V_{1a_v_2;2a_v_1} & V_{1a_v_2;2a_v_2} & V_{1a_v_2;2b} \\ V_{1a_v_1;1b} & V_{1a_v_2;1b} & E_{1b} & V_{1b;2a_v_1} & V_{1b;2a_v_2} & V_{1b;2b} \\ V_{1a_v_1;2a_v_1} & V_{1a_v_2;2a_v_1} & V_{1b;2a_v_1} & E_{2a_v_1} & 0 & V_{2a_v_1;2b} \\ V_{1a_v_1;2a_v_2} & V_{1a_v_2;2a_v_2} & V_{1b;2a_v_2} & 0 & E_{2a_v_2} & V_{2a_v_2;2b} \\ V_{1a_v_1;2b} & V_{1a_v_2;2b} & V_{1b;2b} & V_{2a_v_1;2b} & V_{2a_v_2;2b} & E_{2b} \end{pmatrix}$$

(2.5)

where the diagonal element *E* is the energy of a distinct transition of one chromophore, the off-diagonal element *V* is either the intergroup interaction or the static-field mixing of excitations of a single group, and the interaction

between different vibrational levels of a given electronic excited state in a particular chromophore is set to zero.

The interactions between chromophores are considered to be purely electrostatic. The interaction potential $\hat{V}_{ij}$ is then expressed with transition charge densities ρ and the separation r between chromophores. The transition charge densities are represented by a set of monopole charges so that the interactions between transitions can be calculated as the Coulomb interactions between these charges (eqn. 2.6 and 2.7).[166]

$$V_{i0a;j0b} = \iint \frac{\rho_{i0a}(\mathbf{r}_i)\rho_{j0b}(\mathbf{r}_j)}{4\pi\varepsilon_0 r_{ij}} d\mathbf{r}_i d\mathbf{r}_j \qquad (2.6)$$

where ρ is the electronic transition density of either chromophore *i* or *j* from the ground state 0 to the excited state *a* or *b*, respectively; ε_0 is the vacuum permittivity and r_{ij} is the separation of the chromophores.

$$V_{i0a;j0b} = \sum_{s=1}^{N_s}\sum_{t=1}^{N_t} \frac{q_s q_t}{r_{st}} \qquad (2.7)$$

where q is the magnitude of the monopole charges on the chromophores, N is the number of charges and r_{st} is the distance between two charges.

By solving the Schrödinger equation, through diagonalizing the Hamiltonian matrix by a unitary transformation (eqn. 2.8), the energies of the transitions of the protein are the eigenvalues. The coefficients, c_{ia}^{k}, which show the contributions of each excitation of individual chromophore to the system, are the eigenvectors.

$$U^{-1}\hat{H}U = H_{diag} \qquad (2.8)$$

The electric and magnetic TDM of the transitions of the protein can be calculated from the TDMs of individual chromophores with the unitary matrix U_{ai} (eqn. 2.9 and 2.10).

$$\vec{\mu}_i = \sum_a U_{ai}\vec{\mu}_a^0 \qquad (2.9)$$

$$\vec{m}_i = \sum_a U_{ai}\vec{m}_a^0 \qquad (2.10)$$

An experimental spectrum is not a series of lines (*i.e.* is not a line spectrum) due to various effects, such as the uncertainty principle, the Boltzmann distribution of the structural ensemble, the existence of hot bands, the Franck-Condon effect and solvent effects, etc. As a result, a function to expand the calculated line spectrum is needed in order to make the calculation comparable with the experiment. A convolution of the calculated

line spectrum with a Gaussian type function is used to mimic the observed broadened spectrum[166].

2.3 Franck-Condon calculation

2.3.1 First principles method

The electron distribution in the excited state during the excitation can be evaluated based on the Franck-Condon factors. The Franck-Condon factors, which are the square of the overlap integrals of the vibrational wave functions between the ground and the excited states, govern the intensity distribution along the progression in the absorption spectrum. In order to obtain these numbers, an accurate description of the electronic ground and excited states of the chromophore is needed, which requires a considered choice of method. The electronic structures of molecules can be obtained from empirical or semi-empirical methods[167]. However, these computational cheaper methods depend significantly on the parameterization. Solutions of the Schrödinger equation can also be generated without reference to experimental data, namely *ab initio* methods. These methods solve the Schrödinger equation under the Born-Oppenheimer approximation[168], which considers the nuclei and the electrons separately due to their weight differences. One crucial part of obtaining good description of a system is to

deal with the molecular correlation energy (E_{corr}), which is the difference between the nonrelativistic true molecular energy (E_{nonrel}) and the HF energy (E_{HF}). There are two sources of E_{corr}. One is called dynamic correlation, which is from the failure of accounting for the instantaneous correlations between motions of electrons. The other one, static correlation, is from using single determinant or single configuration state functions (CSFs) to poorly represent the system's state. Selection of specific *ab initio* methods depends on the performance against the experimental data and also the physical problems that are of interest. In our cases, excitation of electron from the occupied molecular orbitals (MOs) to the unoccupied MOs is under consideration. Thus, a linear combination of the CSFs is needed to replace single CSF to deal with the static correlation for this open-shell system.

The configuration interaction (CI) [169,170] method is one of the three main methods for computing electron correlation. The other two methods are many-body perturbation theory and coupled cluster, which will not be covered here. In electron correlation methods, a series of CSFs containing excited Slater determinants are generated by replacing the occupied MOs in the HF determinant by virtual (unoccupied) MOs. The more unoccupied MOs considered the more excited determinants can be generated. The total number of the determinants depends on the size of the basis set which is formed by a series of basis functions. A larger basis set would construct a

series of more determinants which will give better results. For a system containing *n* electrons and *b* basis functions, the number of CSFs will about proportional to b^n. Even with techniques to deal with large matrices[171,172], a CI calculation considering all possible CSFs is not computationally achievable for even medium-size system. Thus, approaches[173,174] are designed to accelerate the convergence to the final wave function which can be produced with any set of MOs obtained from the basis set. The multi-configurational SCF (MCSCF) method is one of these approaches.

In MCSCF method, both the expansion coefficients of CSFs and the MOs used to construct the CSFs are varied[175]. The MOs are optimized by varying the expansion coefficients of the basis functions forming the MOs. In this thesis, we employed state-averaged CASSCF[176] method, the most commonly used MCSCF method, to calculate the electronic structures of the ground and excited states of chromophores (NMA were computed using CASPT2[177] based on CASSCF reference). In the CASSCF method, MOs are partitioned into core orbitals which are kept doubly occupied, virtual orbitals which are kept unoccupied and active orbitals (Figure 2.1). A full CI calculation is performed within the active MOs, which will give a good description of the electron structures between the MOs of interest with less computational cost. Our CASSCF calculations are on MOs obtained using HF method. The selection of the active space is crucial in the CASSCF calculation. In general, the active

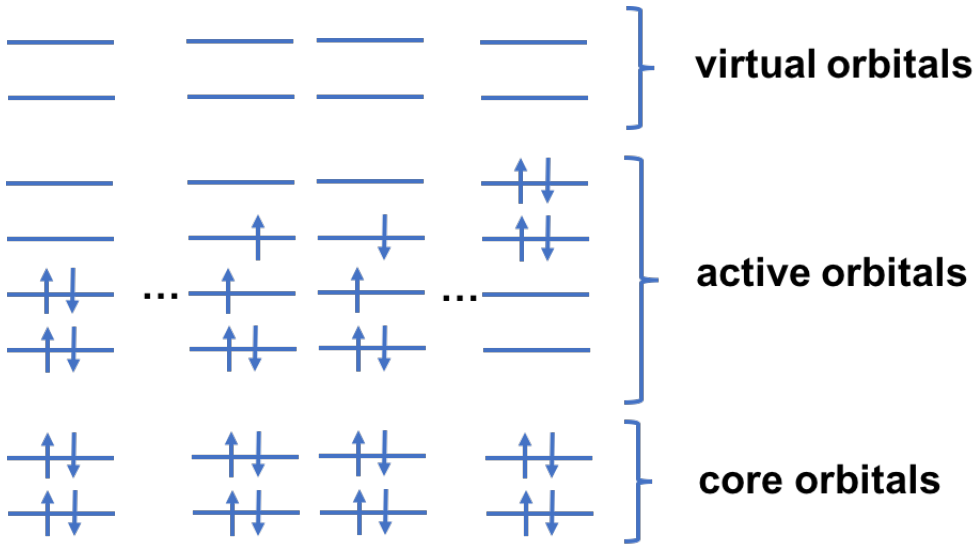


Figure 2.1 MOs are divided as core, active and virtual orbitals in CASSCF method with different electron occupation. Electrons are represented by arrows with arrowheads facing up and down to represent spin state. The core orbitals are kept doubly occupied while the virtual orbitals are kept unoccupied. All the configuration states in the active orbitals are taken into account in the CASSCF calculation.

space should include MOs whose occupation numbers are expected to change in any of the excited states under consideration, which means that all near-degeneracy effects are included. The wave function is then written as a linear combination of all CSFs that derived from all the possible configurations, namely active electrons distributed in all possible ways among the active orbitals. Both electrons and orbitals are limited up to 16 when determining the active space. Because the number of CSFs derived from an active space over this value will beyond the capability of current computing power. As

mentioned above, NMA was calculated with CASSCF/CASPT2 method, which means using a CASSCF wave function as the zeroth-order wave function and applying second-order perturbation theory to it. All remaining static and dynamic correlation, which is covered by CASSCF, are treated at perturbation level.

The selection of the active space of individual chromophores is specified in the chapter 3 and 4. We averaged over the ground and the excited states when optimizing the excited state energy to avoid root-flipping problems. State-averaged calculation would yield a single set of compromise orbitals for the states under consideration. The weights of all states are identical. The Hessian, which is the matrix of the second derivatives of the energy regarding to the vibrational parameters, of all states are checked to confirm the converging at true minimum after the iterative SCF procedure.

We employed the MOLPRO package (Version 2010.1)[178] for the CASSCF calculation with the embedded code MULTI.[179,180] MOLPRO was founded as a high-accuracy wave function quantum chemical program in 1960s. Different levels of theory such as multi-reference correlation treatments, coupled cluster, DFT and many other methods were added along the years. MULTI is a fast converging program combining a fast orbital optimization MCSCF method[181] with determinant-based full CI codes[182].

Another well-developed method is the DFT[183]. In DFT calculation, the electron correlation energy is estimated from the electron density through a density functional whose effectiveness was proved by Hohenberg and Kohn.[184] In contrast to the wave function approaches, DFT calculation requires the electron density, which is the square of the wave function, regardless of the number of electrons in the system. A wave function of an *N* electron system has three spatial and one spin variables per electron while DFT has a fixed number of variables. The computational complexity increases with the number of variables. By using an auxiliary set of orbitals, the exchange-correlation energy is the only unknown functional in the total energy expression, which is a small fraction. Thus, DFT can provide quite accurate results even with crude approximation of the exchange-correlation energy. However, the main problem in DFT is that the calculation cannot be improved systematically as in wave function methods, which means the performance of the calculation depends on the property of interest. The calculation of the excited states can be achieved by extending the DFT theorem to include time-dependent electric potentials, namely TDDFT[185,186]. There is growing interest in the use of TDDFT to study CD. TDDFT has been used to study the absorption and CD spectra of some fairly small molecules[121,187]. However, it was not used as extensively as wave function

methods in previous research on the excited states of chromophores of interest in this thesis.

The other major approximation in *ab initio* calculations is the introduction of a basis set. The approximation comes from using finite basis functions to expand MOs, as a complete basis set is impossible in actual calculations. Selection of the basis set can significantly affect the quality of the QM calculations. Both the size and the type of the basis functions matter when they are used to represent MOs. Fewer basis functions are needed if a single basis function can better reproduce the MOs to achieve the same level of accuracy. Unfortunately, a basis set is not universal for different methods; otherwise it will not be computationally efficient. In this thesis, we use a split valence basis set 6-31G* in our CASSCF calculation.[188] This basis set is one of Gaussian–Type Orbitals (GTOs)[189]. In GTO functions, a basis function is formed by a linear combination of multiple GTOs, namely primitive function, through a contracted function. The exponents are optimized by total energy minimization in GTOs, which means the core orbitals will be described by many functions while few functions will be distributed on valence orbitals which are chemically important. The computational cost of *ab initio* calculations is proportional to the fourth power of the number of basis functions. Thus, a basis set is contracted to minimize the number of functions on the inner-shell orbitals. The notation of 6-31G* is: six primitive GTOs are

contracted for the core orbitals while the valence orbitals are segmented into inner and outer part and are represented by three and one primitive GTOs, respectively. The asterisk represents the added polarization functions[190] for non-hydrogen atoms. Contracting a basis set restricts the number of variational parameters which will increase the energy, but the computational cost is reduced significantly. This basis set has been used extensively in the calculations of similar molecules to those that we are interested in[191-195]. The large number of studies gave us good idea of the accuracy we can obtain.

2.3.2 Franck-Condon Calculation

We calculated the Franck–Condon overlap integrals, f_v, with the freely available software, ezSpectrum.[196] The vibronic TDM is expressed as eqn. 2.11 in the adiabatic approximation:

$$\mu_{w',w} = \langle w' | \langle e' | \mu | e \rangle | w \rangle = \langle w' | \mu_{e',e} | w \rangle \qquad (2.11)$$

where $|w\rangle$ represents the vibrational state of the electronic state $|e\rangle$. Calculation of the Franck-Condon integral is under the harmonic approximation, in which the vibrational Hamiltonian for each electronic state of interest is a sum of *N* one-dimensional (1D) Hamiltonians (*N* is the number of normal coordinates) and the vibrational states are direct products of 1D

states $\left|w_k\right\rangle$ for each mode. Eqn. 2.12 is a linear transformation (Duschinsky transformation) connecting the normal coordinates between different electronic states.

$$\vec{Q}' = J \cdot \vec{Q} + \vec{d} \qquad (2.12)$$

where $\vec{Q}$ is the sets of *N* normal coordinates of one electronic state, *J* is the rotation matrix (Duschinsky matrix) and $\vec{d}$ is the vector of displacement along the normal coordinates between different states. The Duschinsky rotation is needed to obtain multidimensional Franck-Condon factors if the two sets of normal coordinates of different electronic states are significantly non-parallel. The vibrational modes of NMA, toluene, *p*-cresol or 3-MI which dominate the progressions for the absorption have been selected based on the symmetry and guided by previous studies of the vibrational features in the absorption and CD spectra (shown in chapter 3 and 4). Our *ab initio* calculations gave consistent results with previous work in the literature on these predominant vibrational modes, which indicated that there is no severe distortion of the vibrational modes in the electronic ground and excited states, which means the TDM is independent on the nuclear coordinates (Condon approximation). This situation fulfills the so-called parallel approximation in the code ezSpectrum, which assumes that the normal coordinates of the electronic ground and excited states are the same. Thus, the Duschinsky

rotation is not considered in currently model and the multidimensional Franck-Condon integrals are presented by the products of one dimensional Franck-Condon integrals which are computed analytically.

The dimension of the expanded matrix of CD calculation depends on the number of the vibrational levels considered for chromophores of interest. The tractability of diagonalization of very large matrices limits us and we take only the dominant vibronic transitions into account. Transitions were considered to be excited from the lowest vibrational level of the ground electronic state, *i.e.*, no hot bands were taken into account, but combination bands were allowed which means the electron is allowed to be excited to a vibrational state with multiple normal modes excited. Six vibrational levels were calculated for individual modes in the excited state. Franck-Condon integrals with magnitude greater than 0.1 were considered as sufficiently significant to be incorporated into the CD calculations. The Franck-Condon integrals were normalized using eqn. 2.13.

$$F_{\nu} = \frac{f_{\nu}}{\sqrt{\sum_{i}^{N} f_{\nu_i}^2}} \quad (2.13)$$

The vibrational transitions are incorporated by modifying the electric transition dipole moments and the monopole charges representing the

transition charge densities by scaling by the normalized Franck–Condon overlap integral. No interaction between the vibrational levels of the same chromophore is considered in the CD calculation. We consider transitions in a vibrational progression with the precise number determined based on a threshold of 0.1 for the Franck-Condon integral (below which transitions are taken to make a negligible contribution). The new parameters are accessible via the DichroCalc web interface at:

http://comp.chem.nottingham.ac.uk/dichrocalc.

2.4 Statistical measurements

In statistical, a variety of methods are used to evaluate the bias, precision or accuracy of a developed model by calculating the errors with different equations. Errors are difference of the estimated value and the true value obtained from the experiment. Bias tells the distribution (over- or under-estimated) of errors while precision measure the variability, namely the appearance of random error. Since we are interested in a better prediction of the spectrum, an unbiased distribution of errors or the precision of the model which is completely independent of the true value are out of our concern. In order to assess the accuracy (overall distance between the estimation to the true value) of our CD calculation model considering the Franck-Condon effect, we used statistical measurements including mean

absolute error (MAE), mean relative error (MRE), root mean square error (RMSE) and Spearman rank correlation coefficient to compare the predicted spectra with different parameter sets to the experiment. Taking square or absolute value of errors will only consider the magnitude of the difference and eliminate the directions.

MAE measures the accuracy of a model by calculating the average absolute difference between the predicted and the measured values. The calculated CD spectra with various parameter sets are compared with the experiment separately via eqn. 2.14 to tell how far (average distance) our prediction is from the observation.

$$MAE = \frac{\sum_{i=1}^{n} |I_{cal_i} - I_{\exp_i}|}{n} \quad (2.14)$$

where I_{cal_i} and I_{exp_i} represent the calculated and the experimental CD intensities at each integer wavelength, respectively; n is the number of data point considered, which represents the number of integer wavelength covered in the experimental CD spectrum in this thesis.

MRE (eqn. 2.15) is a scaled version of MAE, which is used in the near-UV CD calculation. This is used to make the statistical measurements comparable within the protein dataset. Because, unlike in the far-UV, the

band shapes and intensities of CD spectra are highly structure-dependent in the near-UV.

$$MRE = \frac{\sum_{i=1}^{n}\left|\frac{I_{cal_i}-I_{\exp_i}}{I_{\exp_i}}\right|}{n} \qquad (2.15)$$

where I_{cal_i}, I_{exp_i} and n represent the same variables as in eqn. 2.14.

RMSE (eqn. 2.16) is another common measure to evaluate the accuracy of the differences between the calculation and the observation. The squared operation will dominant the outlier (far away from the true value), namely give higher weight to the larger error, generated by the model.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(I_{cal_i}-I_{\exp_i})^2}{n}} \qquad (2.16)$$

where I_{cal_i}, I_{exp_i} and n represent the same variables as in eqn. 2.14.

Since the vibrational parameters are expected to improve the prediction of the fine structure in the protein CD spectra, we need to assess (1) the effectiveness of the parameters in predicting signals of the specific chromophore; since the excitation of the chromophore will appear at approximate the same energy (vary slightly according to their unique environment); (2) the similarity of the overall shape of the calculated

spectrum comparing to the experiment, which will tell the performance of the parameters for specific protein. We employed the Spearman rank correlation coefficient creatively to measure these two aspects. The variable measured for the first case are the rank of intensities at specific integer wavelength of proteins in the dataset while the intensities of all calculated wavelength are ranked for single protein and compared to the experiment for the second case. Compare to the Pearson product-moment correlation which measures the linear relationship between two variables, the Spearman rank correlation determines the monotonic relationship between two variables. A monotonic relationship is as the value of one variable increases, the value of the other variable either increases or decreases along with it. The Spearman rank correlation is less restrictive than the linear relationship measured by the Pearson correlation.

Chapter 3

Calculations of Protein Circular Dichroism in the Far-Ultraviolet[1]

3.1 Introduction

CD spectroscopy in the far-UV is a widely used technique for studying proteins in terms of their secondary structure contents and their changes in processes such as protein folding[197,198] and binding[199]. These studies can benefit from an accurate calculation from first principles of the CD spectrum. The calculations are almost, but not fully, quantitative for proteins with a high amount of α-helical structure; the spectra for proteins with other secondary types are less well reproduced. In this chapter, vibrational structures of electronic excitations of the protein backbone are considered to obtain more accurate calculated spectra.

The parameterization in the matrix method is from *ab initio* calculation of *trans* NMA, which represents the chromophoric group of the

[1] Work in this chapter has been published in ref 138.

protein backbone[200]. Figure 3.1 shows the two electronic transitions of NMA located in the far-UV region, which are from the amide non-bonding π orbital (π_{nb}) to the anti-bonding π orbital (π^*) and from the oxygen lone pair orbital n to the π^* orbital. The parameters used in the CD calculation describe the charge distributions of the different electronic states. Deviation from planarity of the peptide bond will have some influence on the electronic structure[201], but this is not the focus of the current study. Hydrogen bonding may also be important, but previous work on the explicit consideration of this did not find any significant improvement for protein calculations[140,202].

An early study considering vibronic coupling in proteins investigated absorption in the far-UV of an idealized α-helix[203] and poly-L-lysine as β-sheet[204] and calculated the spectrum, focusing on the C-N stretch mode in the $\pi_{nb}\pi^*$ band. The predominant role of the C-N stretch in this transition has been confirmed by resonance Raman experiments[205]. No combination transition along the amide II' overtones was observed in D_2O solution. Although the role of protic solution is not clear yet, the band of the carbonyl stretch, which is the primary mode in gas phase and in acetonitrile, was not apparent in aqueous solution.

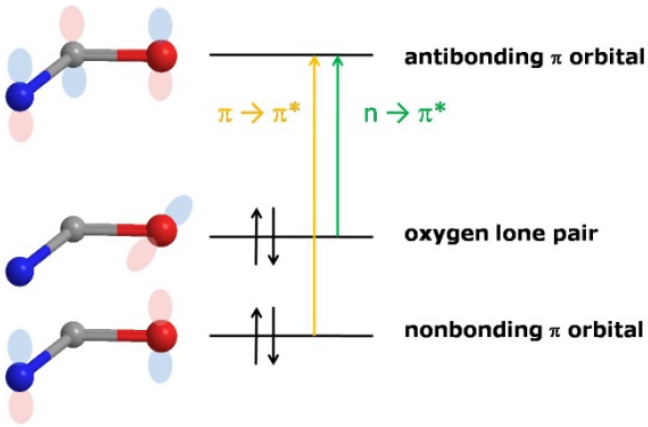


Figure 3.1 Molecular orbitals of amide structure relevant to electronic transitions in the far-UV region. The ball and stick model represents amide bond (blue-nitrogen, grey-carbon, red-oxygen).

In this chapter, we calculate the vibronic structure of the $\pi_{nb}\pi^*$ transition in NMA, which represents the backbone chromophore of proteins, and incorporate this into the calculations of protein CD spectra. A set of 49 proteins, comprising a range of structural types, are calculated to assess the influence of the vibrational structure. Considering the vibronic structure of peptide bond has provided some further insight into the physical origins of the nature of protein CD spectra in the far-UV.

3.2 Methods

We have considered the *trans* form of NMA, which is much more common in proteins than the *cis* form[206] (0.34% peptides are *cis* in our dataset). The electronic ground and $\pi_{nb}\pi^*$ excited state geometries and vibrational frequencies of NMA were computed using multi-reference CASPT2[177] based on a CASSCF reference and the 6-31G* basis set with the MOLPRO package (Version 2010.1)[178]. A multi-configurational method is necessary for a proper description of the electronic structure of amides, because of a significant interaction between the ground and $\pi_{nb}\pi^*$ states. The active space we used is four electrons in five active orbitals (two occupied and three virtual), which, although minimal, suffices for our purposes. The excited state optimization used state-averaged CASPT2, averaged over the ground, $n\pi^*$ and $\pi_{nb}\pi^*$ states, which should give MOs that are not biased toward any particular electronic state. Although the ground state of NMA is planar[207], we calculated both the ground and excited states with no symmetry imposed.

CD spectra were calculated for a set of 49 proteins (Table 3.1) with the approach described in section 2.2. The crystal structures of these proteins are in the PDB. These proteins are taken from the database SP175 of the

Table 3.1 Proteins considered in this study.

Class	Protein (PDB entry)
α-helix	Lysozyme (193L), Myoglobin whale (1A6M), Leptin (1AX8), c-Phycocyanin (1HA7)*[a], Hemoglobin (1HDA), Cytochrome C (1HRC), Calmodulin (1LIN), Insulin (1TRZ), Phospholipase A2 (1UNE), Myoglobin horse (1YMB)
mixed-α,β	Aldolase (1ADO)*, Haloalkane dehydrogenase (1BN6), Carbonic anhydrase I (1HCB), PNMT (1HNN)*, NmrA (1K6J)*, Monellin (1MOL), Papain (1PPN), Dehydroquinase type 1 (1QFE), Rhodanese (1RHS), Ubiquitin (1UBI), Dehydroquinase type 2 (2DHQ), Deoxyribonuclease-1 (3DNI), RNase A (3RN3), Carboxypeptidase A (5CPA), Triosephosphate isomerase (7TIM)
β-I	Pectate lyase C (1AIR), β-Bactoglobulin (1B8E), γ-D-crystallin (1ELP), γ-S-crystallin (C-term) (1HA4), γ-D-crystallin (human) (1HK0), Jacalin (1KU8)*, Lentil lectin (1LES), γ-E-crystallin (1M8U), Concanavalin A (1NLS), Pea lectin (1OFS), Avidin (1RAV), Thaumatin (1THW), β-B2-crystallin (2BB2), Pepsinogen (2PSG), γ-B crystallin (4GCR)
β-II	Cu,Zn Superoxide dismutase (1CBJ), α-Bungarotoxin (1HC9), Rubredoxin (1R0I), Streptavidin (1STP), α-Chymotrypsinogen (2CGA), Ferredoxin (2FDN), Elastase (3EST), α-Chymotrypsin (5CHA), Aprotinin (5PTI)

[a] For computational tractability, only the unique chains of multi-meric proteins (denoted by asterisk) containing more than 500 residues were considered. The validity of this approximation has been established elsewhere.[146]

Protein Circular Dichroism Data Bank[208] (PCDDB) in which the experimental spectra are available. To assess our calculations, the CD spectra of these proteins were calculated and compared with the experimental data in the far-UV region.

3.3 Results and Discussion

The active space employed in the CASPT2 calculation comprises two occupied and three virtual orbitals. The first transition is the $n\pi^*$ transition, followed by the $\pi_{nb}\pi^*$ transition. The lowest energy conformers of ground and $\pi_{nb}\pi^*$ excited states of NMA are shown in Figure 3.2 and the optimized geometric parameters are summarized in Table 3.2. The structure of NMA has been previously studied experimentally[209-212] and theoretically[213-215]. Our calculation suggests a planar symmetry for the ground state which is consistent with X-ray crystallography[216] and gas electron diffraction experiments[217]. The $\pi_{nb}\pi^*$ excited state has been less well studied, due to the challenges of characterising the geometries of electronically excited states. The geometry distortion and the displacements involved in the excited states are known mainly from Raman experiments[205,215]. Chen *et al*.[215] concluded that C-N elongation is the largest displacement of NMA in the aqueous phase, which agrees with Mayne and Hudson's analysis[205] of

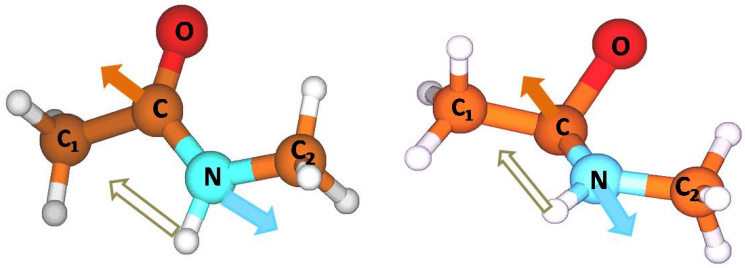


Figure 3.2 Calculated geometries and C-N stretching mode of NMA in the ground state (left) and the $\pi_{nb}\pi^*$ excited state (right).

Table 3.2 Ground and excited state geometries of NMA, optimized at the CASPT2 level, along with the frequencies of C-N stretch in both states. Experimental results[216,218] are listed as a reference.

Geometry parameter	**Exp.**	**CASPT2**	
	g.s.[216]	g.s.	$\pi_{nb}\pi^*$ e.s.
C-N / Å	1.386	1.368	1.326
C=O / Å	1.224	1.231	1.587
N-H / Å	1.106	1.010	1.021
$C\text{-}C_1$ / Å	1.520	1.517	1.485
$N\text{-}C_2$ / Å	1.469	1.450	1.457
O=C-N / °	121.8	123.2	108.4
C-N-H / °	110.0	118.9	120.2
$N\text{-}C\text{-}C_1$ / °	114.1	115.2	117.1
$C\text{-}N\text{-}C_2$ / °	119.6	122.2	121.3
H-N-C=O / °		-179.9	138.6
C-N stretch[218] / cm^{-1}	1581	1610	1696

the resonance Raman spectrum that the C-N stretch is the predominant mode in water. They estimated the vibrational frequency of the C-N stretch to be 1400 cm^{-1}, with which they obtained a good fit to the absorption spectrum of NMA. Li *et al*.[213] calculated, using the configuration interaction between all singly excited configurations (CIS) approach, that the excited state of NMA possesses a pyramidal carbonyl and a planar amine conformation. Our CASPT2 calculation suggests similar amide geometry. However, our calculation indicated a shorter C-N bond in the excited state structure compared to the ground state geometry, while this bond appeared to be elongated in the Raman experiments. This discrepancy may have arisen from the neglect of solvent effects in our calculation. Nevertheless, the absorption spectrum of NMA in water (Figure 3.3) can be reproduced quite well by the Franck-Condon factors generated with these *ab initio* data. We incorporate the vibronic coupling into the CD calculations with the Franck-Condon overlap integrals calculated from the geometries and frequencies of the current results. The calculated vibrational frequencies of the C-N stretch in the ground and excited states are 1610 cm^{-1} and 1696 cm^{-1}, respectively. Compared to the experimental frequency of the ground state, which is 1581 cm^{-1}, the calculated ground state frequency is within the known systematic errors.[219]

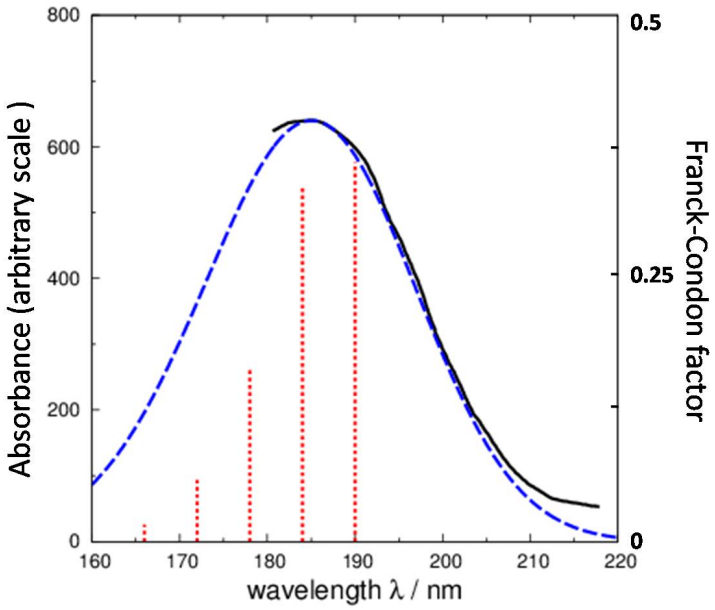


Figure 3.3 Absorption spectra. The black solid line is experimental spectrum, the dashed line is the calculated spectrum and the dotted vertical lines are proportional to the Franck-Condon factors.

The parallel approximation in ezSpectrum was used to calculate the Franck-Condon overlap integral of NMA. Only the C-N displacement was considered in this study and the Franck-Condon overlap integrals of C-N stretch mode are summarized in Table 3.3. The transitions are denoted as [excited vibronic state ← ground state]. Figure 3.3 compares the experimental absorption spectrum[220] with the simulated absorption intensity from our *ab initio* calculation. The origin of the transition (0←0) was placed at 190 nm (52632 cm^{-1}) and the energy gap between different

vibrational levels is set to 1696 cm^{-1}, taken from the excited state frequency calculation. The intensities, which are represented by the vertical lines, are proportional to the Franck-Condon factors. A Gaussian function with a bandwidth of 15 nm was applied to simulate other broadening effects.

Table 3.3 Franck-Condon overlap integrals of the $\pi_{nb}\pi^*$ transition in NMA, considering only the C-N stretching mode.

Transition	0←0	1←0	2←0	3←0	4←0
FC integral	0.62	0.60	0.42	0.25	0.13

In order to assess the CD calculations, we used the Spearman rank correlation coefficient and MAE. For each integer wavelength, the calculated and experimental intensities of all proteins considered are ranked separately. The Spearman rank correlation coefficient measures how the rank of calculated intensities correlated with the rank of experimental intensities at a specific wavelength (1 corresponds to a perfect correlation, zero to no correlation, and -1 to a perfect inverse correlation). The CD calculations are discussed below based on various secondary structure types. The proteins are grouped (Table 3.1) based on assignments by DSSP[123]. Proteins containing more than 40% α-helical or

β structures are classified as the α-helix and β-sheet groups, respectively. The β type is further divided into β-I and β-II subtype based on the nature of the β structure. The β-I type proteins have regular sheets while the β-II type contains more β-bulges and irregular strands. The rest of the proteins in the dataset are categorized as mixed-α,β.

Figure 3.4 compares the experimental and calculated spectra of myoglobin (PDB code: 1YMB), which with 73% helix is a representative of the α-helix subset. The main features of the CD spectra of α-helical proteins are an intense positive band at around 190 nm and two negative bands at 208 nm and 222 nm. The bands at 190 nm and 208 nm arise from the $\pi_{nb}\pi^*$ transition, while the band at 222 nm is due to the $n\pi^*$ transition. For myoglobin, the intensity of peaks at 190 nm and 208 nm are reproduced equally well by the calculations with and without the vibronic structure considered with bandwidths of 10 nm and 12.5 nm, respectively. The double negative bands are better reproduced when the vibronic structure is included. CD spectra calculated with a 10 nm bandwidth for the non-vibronic-parameter are also shown in Figure 3.4, 3.6 and 3.7 to examine the influence of the bandwidth in the calculations. The 208 and 222 nm peaks are distinct with a narrower bandwidth without considering vibrational structure. However, both peaks arising from the $\pi_{nb}\pi^*$ transition are overestimated. This is because a narrower bandwidth

is balanced by the vibronic coupling which broadens the peak. Calculations with vibronic parameters and other bandwidth are not shown since those combinations generate worse results in terms of intensity than current settings. Calculating a spectrum simply using a narrower bandwidth without incorporating any other broadening effect in the absorption process is, thus, less satisfactory. However, the MAE between 210 nm and 240 nm is a little higher for the vibronic calculation and the correlation is a little lower between the calculated and experimental intensity (Figure 3.5a) around 215 nm and 235 nm.

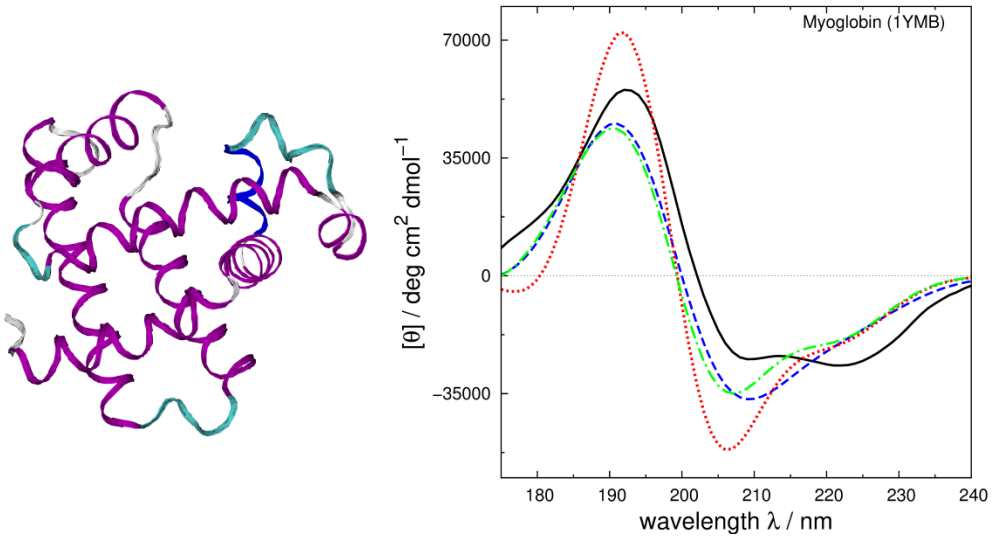


Figure 3.4 Experimental and calculated CD spectra of α-helical protein myoglobin: black solid line, experiment; blue dashed line, non-vibronic calculation with a bandwidth, Δ = 12.5 nm; red dotted line, non-vibronic calculation with Δ = 10 nm; green dash-dotted line, vibronic calculation with Δ = 10 nm. Crystal structure of myoglobin (PDB code: 1YMB) is shown in ribbon, in which purple and cyan represents helical and turn structures while silver represents the sequences that no secondary structures are assigned.

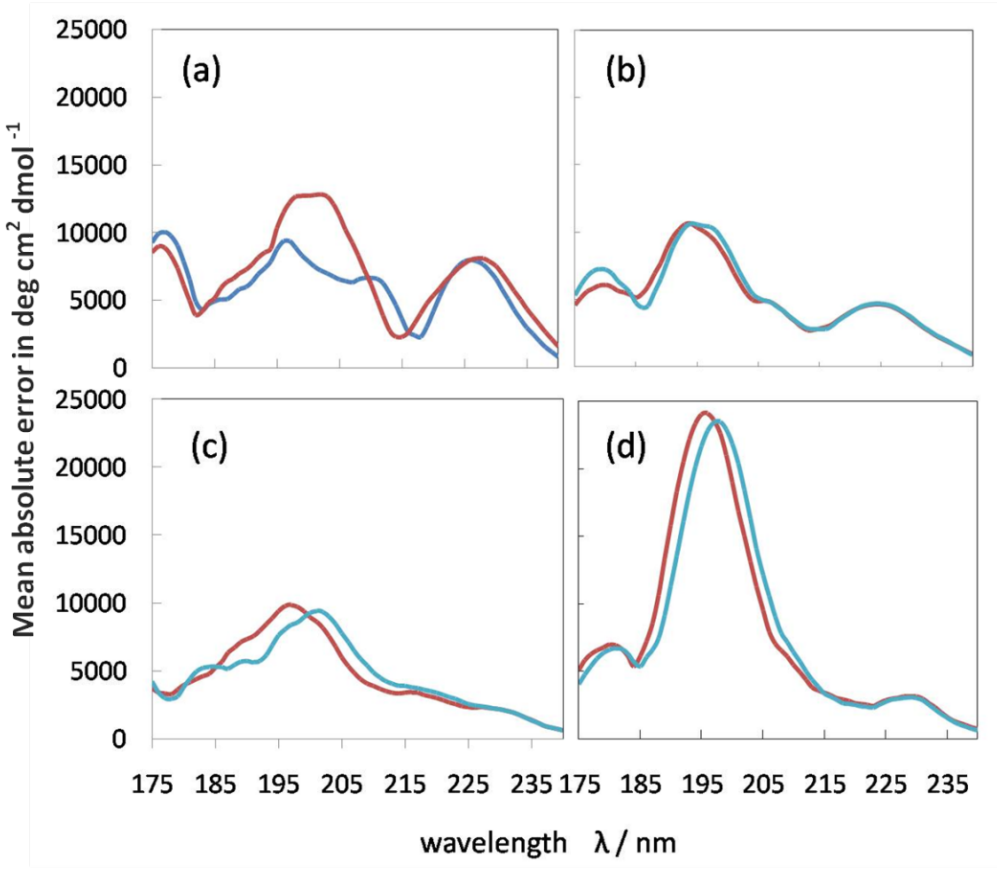


Figure 3.5 MAE of calculations without (blue line) and with vibronic structure (red line). (a) α-helix; (b) mixed-α,β; (c) β-I; (d) β-II.

Although the spectra are reproduced well for the intense positive peak, there is poor correlation around 200 nm, which is the descending edge of the most intense peak in the spectrum of α-helical proteins. This is the point at which the intensity changes from positive to negative with a large gradient, which means a small error in the wavelength calculation

will lead to a large difference in intensity. In the α-helical proteins subset, six out of ten calculated spectra are blue-shifted for both non-vibronic and vibronic calculations. This zero-crossing issue may also explain the difference in the MAEs (Figure 3.5a). Compared to the non-vibronic calculations, those with vibronic structure are slightly shifted to higher energy due to the position of the origin transition. This deviation causes a larger difference when the CD intensity changes sign.

The features of the mixed-α,β type (Figure 3.6) are quite similar to the spectra of the α-helical proteins in most cases. However, the positive peak positions and the intensities, the zero crossing points and the ratio of the double negative bands at 208 and 220 nm all fluctuate, due to the variation in secondary structure content. The Spearman rank correlation coefficient for the mixed-α,β proteins is similar to that of the α-helical group between 175 and 240 nm. The MAE also seems comparable. However, for 10 out of 15 mixed-α,β proteins the peaks at 190 nm are overestimated, while these peaks in the α-helical proteins are either reproduced well or underestimated. The differences between calculated and experimental spectra for mixed-α,β proteins with greater β content are much larger than the other proteins in this subset. There are also double negative bands in this subtype between 205 nm and 225 nm. The vibronic calculation tends to capture this feature in the mixed-α,β group,

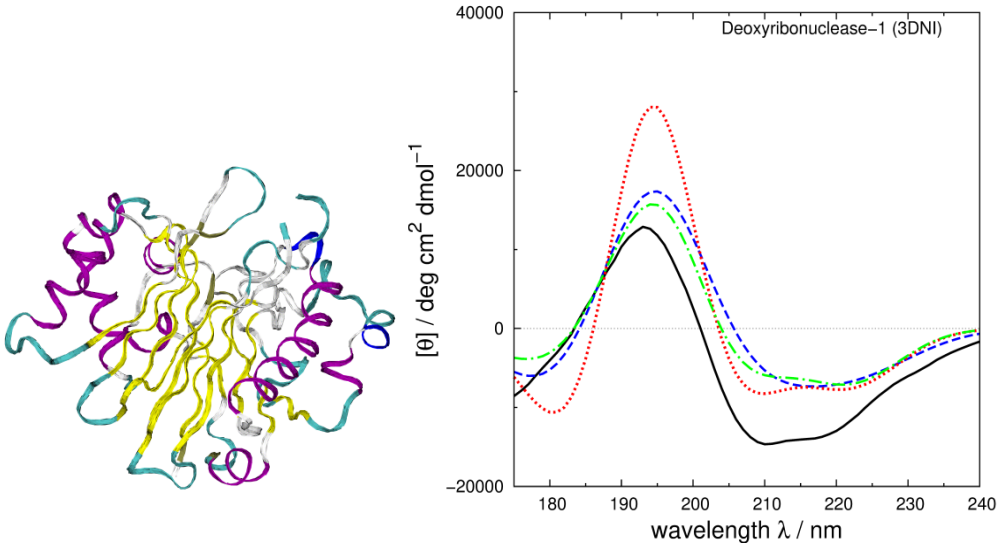


Figure 3.6 Experimental and calculated CD spectra of mixed-α,β protein deoxyribonuclease-1: black solid line, experiment; blue dashed line, non-vibronic calculation with a bandwidth, Δ = 12.5 nm; red dotted line, non-vibronic calculation with Δ = 10 nm; green dash-dotted line, vibronic calculation with Δ = 10 nm. Crystal structure of deoxyribonuclease-1 (PDB code: 3DNI) is shown in ribbon, in which purple, yellow and cyan represents helical, β-strand and turn structures respectively while silver represents the sequences that no secondary structures are assigned.

although the ratio of these negative peaks is not well-reproduced, as is the case with the α-helical subset.

Figure 3.7 compare the spectra for β-I and β-II type proteins, respectively. The general features of β-I spectra are a negative band at

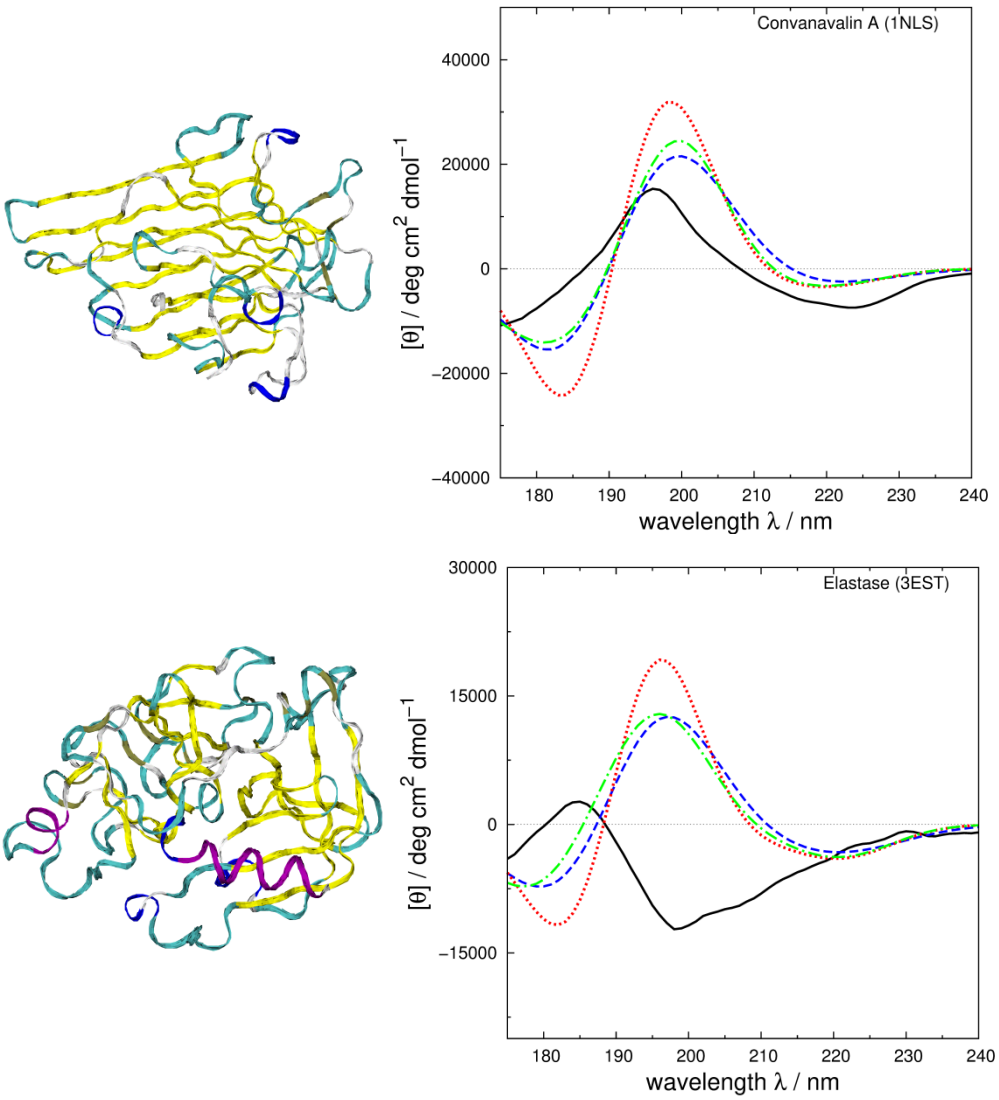


Figure 3.7 Experimental and calculated CD spectra of β-I protein concanavalin A (upper panel) and β-II protein elastase (lower panel): black solid line, experiment; blue dashed line, non-vibronic calculation with a bandwidth, Δ = 12.5 nm; red dotted line, non-vibronic calculation with Δ = 10 nm; green dash-dotted line, vibronic calculation with Δ = 10 nm. Crystal structure of concanavalin A (PDB code: 1NLS) and elastase (PDB code: 3EST) are shown in ribbon, in which purple, yellow and cyan represents helical, β-strand and turn structures respectively while silver represents the sequences that no secondary structures are assigned.

about 216 nm and a positive band with comparable magnitude near 195 nm. When there are irregular β structures, the main features are a minimum at about 200 nm and a positive band in the range of 185 to 190 nm. However, the band positions vary dramatically for different β-sheet proteins. For concanavalin A (Figure 3.7, 1NLS), the intensity of the negative peak was reproduced better with the vibronic structure considered, though both calculations (with and without vibronic structure) show a red shift. The vibronic parameters tend to give more intense bands around 195 nm for β-I subtype proteins. This causes a larger error (Figure 3.5c) between 185 nm and 200 nm; both calculations overestimate the positive bands at 195 nm in most cases in this protein group. For 10 out of 15 β-I proteins, the calculated intensity of the negative peak at 216 nm is too low.

The β-II type proteins have less regular structure and possibly greater conformational dynamics[221] than β-I proteins and their CD spectra are usually poorly reproduced. The MAE is up to 15,000 deg cm^2 $dmol^{-1}$ greater in the region 185 to 210 nm, compared to other secondary structure types (Figure 3.5). There is an improvement in the Spearman rank correlation (Figure 3.8) at most wavelengths for the β-II proteins using the vibronic parameters; the correlations at the two peak positions, 187 and 201 nm, are summarized in Table 3.4. The coefficients are

averaged over a range of ±2 nm to cover the appropriate peak locations. Both peaks listed in Table 3.4 arise from the $\pi_{nb}\pi^*$ transitions. Calculations considering vibronic coupling of the $\pi_{nb}\pi^*$ transition have a positive effect for β-II proteins. Although the correlation coefficient is over 0.75 between 200 and 205 nm, six out of nine proteins in this type, have calculated spectra with a band of the opposite sign to experiment in this region. Both parameter sets overestimate the positive peaks and underestimate the negative ones.

Table 3.4 Calculations with vibronic parameters, compared to calculations without vibronic structures, improve the Spearman rank correlation coefficients between the calculated and experimental intensity at wavelength around 187±2 and 201±2 nm for β-II proteins. The selected wavelengths cover the peak position of β-II proteins we considered in our dataset.

Parameter	**Correlation Coefficient**	
	187±2 nm	201±2 nm
non-vibronic	0.30	0.69
vibronic	0.36	0.81

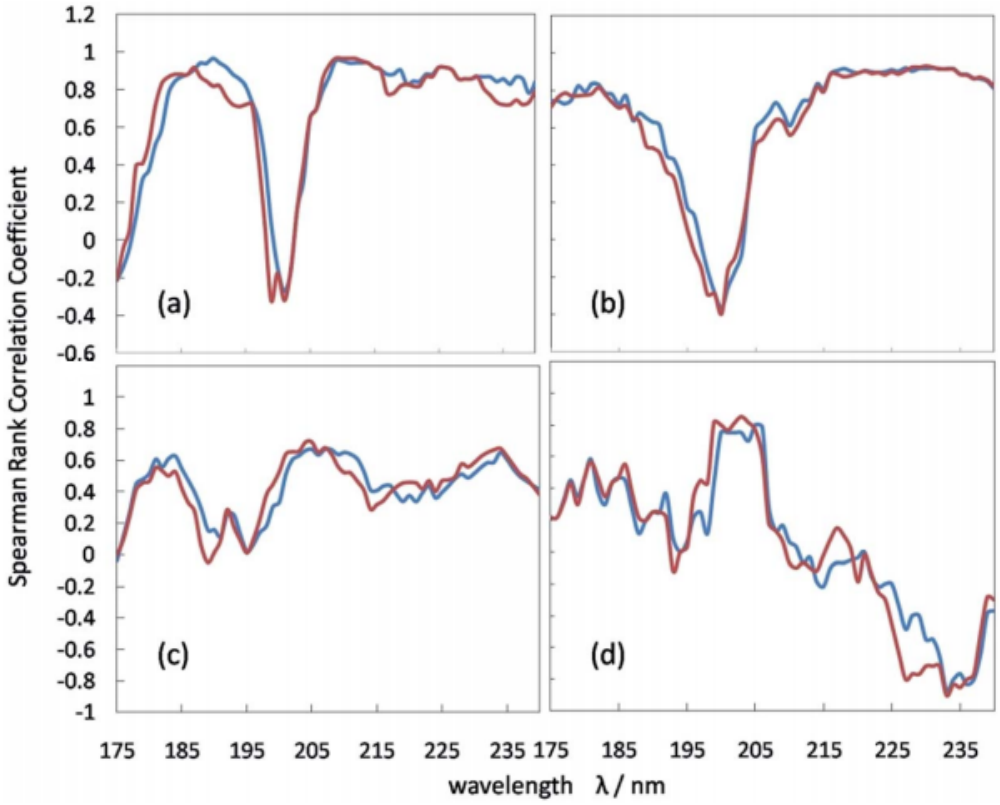


Figure 3.8 The Spearman rank correlation coefficient between the experimental and calculated intensities for the non-vibronic (blue line) and vibronic parameters (red line) of different structural types of proteins: (a) α-helix; (b) mix-α,β; (c) β-I; (d) β-II.

A comparison of the Spearman rank correlation coefficient and MAEs of the entire protein data set for the two sets of calculations (Figure 3.9) reveals no obvious improvement for most regions in the calculated CD spectra with vibronic parameters. However, there is a smaller error around 208 nm for α-helix, β-I and β-II groups. Including the vibronic

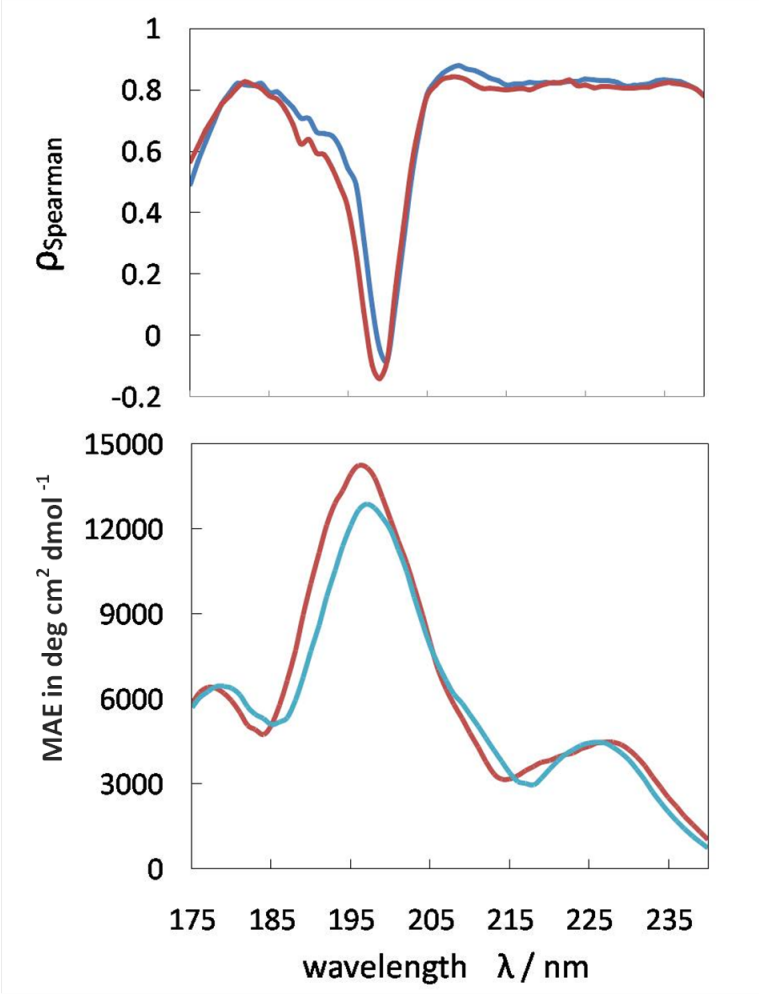


Figure 3.9 The Spearman rank correlation coefficient (upper) and MAE (lower) of all 49 proteins in the dataset. The blue lines correspond to the non-vibronic parameters and the red lines to the vibronic parameters.

structure of the $\pi_{nb}\pi^*$ transition may enhance the coupling between the $\pi_{nb}\pi^*$ transition and the $n\pi^*$ transition. The MAE difference between 185 and 200 nm is mainly from the wavelength where the origin was set. The large error between 190 and 210 nm comes predominantly from the β-II group.

3.4 Conclusions

To incorporate the vibronic coupling in the matrix method, Franck-Condon overlap integrals have been calculated using the ground and excited state geometries and frequencies of NMA. CASPT2 calculations gave a planar geometry of the ground state; in the $\pi_{nb}\pi^*$ excited state the carbonyl is pyramidal and the amine is planar. The conformations are consistent with the previous experimental and theoretical results. Five vibronic transitions in the $\pi_{nb}\pi^*$ excited state were taken into account in the matrix method calculation.

A set of 49 proteins, whose experimental spectra is available, allows us to assess the influence of incorporation of vibrational structure in the CD calculations. The negative peaks between 205 and 225 nm are reproduced with better resolution in α-helical proteins using vibronic parameters. No obvious improvement is observed for mixed-α,β and β-I

proteins. For β-II proteins, whose structures are less regular and whose spectra are difficult to reproduce, the Spearman correlation coefficients between the experimental and the calculated intensities show improvement over most wavelengths in the far-UV region when taking the vibrational structure into account. It is possible that the calculation may be further improved by considering the conformational dynamics and flexibility of β-II proteins. Incorporation of the vibrational structure of NMA in the matrix method calculations provides us with a template to consider Franck-Condon effects in the near-UV[161], which is studied in the following chapter.

Chapter 4

Calculations of Protein Circular Dichroism in the Near-Ultraviolet[2]

4.1 Introduction

CD spectroscopy is a widely used technique which shows the differential absorption of left-handed and right-handed circularly polarized light by chiral molecules, for example, proteins.[222] The side chains of aromatic amino acids play a key role, through hydrophobic packing, in the structure of proteins[87], hydrogels[223] and many non-biological systems, such as conjugated polymers and supramolecular assemblies with aromatic side groups. They can also be integral to protein function, often mediating molecular recognition processes and catalysis. Thus, biophysical measurements to probe aromatic-aromatic interactions can furnish new insight into fundamental aspects of protein behaviour.

[2] Work in this chapter has been published in ref 139.

Bands in the electronic CD spectra of proteins in the near-UV region (240 to 320 nm) provide useful information on the environment and interactions of aromatic side chains. The measurement of the near-UV CD of proteins has been instrumental, for example, in advancing our understanding of compact non-native states of proteins.[224] The $\pi\pi^*$ electronic excitations on the aromatic rings of phenylalanine, tyrosine and tryptophan are sensitive to the local surroundings and can couple with one another. The separation and relative orientation of chromophores varies depending on the structure of the protein. Thus, the near-UV CD spectra of proteins reflect differences in conformation and can be used to probe the structure and dynamics of protein folding. Over many years, there have been applications of near-UV CD spectroscopy for biochemical and biophysical investigations. For example, site-directed mutagenesis has been used to investigate how aromatic residues contribute to the near-UV spectra.[92-95] CD spectroscopy has been used to monitor and study binding processes,[222,225] conformational changes induced by sequence modification,[95,226] and folding and unfolding processes under various conditions,[31,227,228] sometimes on ultrafast timescales.[229] However, the relationship between structure and the near-UV CD spectrum is still not understood in a quantitative fashion, although some progress has been made.

The vibrational structure in the near-UV absorption and CD spectra of aromatic amino acids varies significantly depending on solvent.[88,230-233] The sensitivity of the electronic transitions of aromatic amino acids to molecular environment leads to a richness of information in the near-UV CD spectra of proteins. The vibrational progressions of individual chromophores have been obtained by examining the difference spectra of the wild-type proteins and their mutants.[224,234,235] For various complexes between calmodulin and target peptides, deconvolution of CD spectra into a well-resolved ^1L_b component and a broad ^1L_a component has been used to analyse the characteristic contributions of these transitions in other complexes.[236]

In the previous chapter, the vibrational structure of $\pi_{nb}\pi^*$ transitions of the backbone chromophore has been incorporated in the calculation of the far-UV CD spectra of a set of 49 proteins.[237] The features in the experimental spectra of α-helical proteins are better resolved with parameters considering the vibrational structure. For more irregular β structure proteins, there is a better correlation between the calculated and the experimental intensity in the far-UV.

In this chapter, we consider the Franck-Condon effect in calculations of the near-UV CD of proteins. The vibrational structure of the electronic transitions of aromatic amino acids is obtained from *ab initio* calculations on

toluene, *p*-cresol and 3-MI, which have been widely studied both experimentally[238-245] and theoretically[246-250] as representing the chromophoric part of phenylalanine, tyrosine and tryptophan. The effectiveness of the parameter sets accounting for vibrational structure are assessed by calculations on a set of 40 proteins. We assess the contribution of individual aromatic amino acids by considering changes in spectra due to mutation. Calculations with NMR-determined structures are used to investigate the influence of the conformational diversity and turn out to further improve the accuracy of the calculations of the majority of cases studied. Calculations on native structures of proteins are compared with computer-generated misfolded structures to assess the utility our new parameters in contributing to structure prediction and determination.

4.2 Methods

The electronic ground and ^1L_b excited states of toluene, *p*-cresol and 3-MI and also the ^1L_a state in the case of 3-MI (the ^1L_a states of toluene and *p*-cresol occur in the far-UV at ~208 nm and ~216 nm, respectively) were calculated using the MOLPRO package[178] version 2010.1 employing the state-averaged CASSCF method[176] with the 6-31G* basis set for various active spaces: six electrons in seven orbitals for toluene (six valence π orbitals and one Rydberg orbital), six electrons in six orbitals for *p*-cresol and

ten electrons in ten active orbitals (nine valence π orbitals and one Rydberg orbital[251]) for 3-MI, respectively. The calculations of the Franck-Condon factors require the equilibrium geometries of the ground and the excited states, the vibrational mode vectors and the related frequencies; these were all taken from the *ab initio* calculations.

Experimental CD spectra of 40 proteins (Table 4.1) were assembled from the literature. The units of the experimental spectra were all standardized to be mean residue ellipticity to facilitate comparison of calculated and experimental spectra. The crystal structures of these proteins and the solution NMR structures of 13 proteins are available from the PDB.[62] The selection of a certain PDB entry was made to match as closely as possible factors such as species, bound ligands and other experimental conditions under which the CD spectra had been measured. All chains of multimeric proteins are used in our calculations. The calculated spectra of NMR structures are obtained by averaging the spectra of individual structures in the PDB file.

Table 4.1 The 40 proteins studied: name, PDB codes, length (*i.e.* number of residues in the X-ray structure) and the number of aromatic residues and disulfide bonds. The four proteins whose mutants are studied are indicated by asterisk. Phe, Tyr, Trp and S-S represents phenylalanine, tyrosine, tryptophan and disulfide bond, respectively.

Protein	**PDB entry**		**Length**	**No. aromatic amino acids**			
	X-ray	NMR		Phe	Tyr	Trp	S-S
Acetylcholinesterase[252]	2ACE		537	34	17	14	3
Adenylate kinase[253]	2ECK		428	10	14	0	0
α-Lactalbumin[254]	1A4V		123	3	5	3	4
α-Toxin[255]	1QM6		740	30	50	20	0
Apolipophorin III[256]	1AEP	1LS4	161	2	0	2	0
Barnase[94]*	1A2P	1FW7	330	12	21	9	0
β-2 microglobulin (human)[257]	1LDS	1JNJ	100	5	6	2	1
β-Lactamase[258]	1BTL		263	5	4	4	1
P.69 pertactin[259]	1DAB		539	10	5	7	0
Bovine pancreatic trypsin inhibitor[93]*	5PTI	1PIT	58	4	4	0	3
Calmodulin[225]	4CLN		148	9	1	0	0
Cardiotoxin[260]	4OM4	1CRF	300	10	15	0	20
Chymotrypsinogen A[261]	2CGA		490	6	4	8	10
Dehydroquinase II[262]	2BT4		1884	36	48	12	0
Dihydrofolate reductase[263]	4P3Q		159	6	4	5	0
Glucose oxidase[264]	1CF3		583	18	27	10	1
Hen egg white lysozyme[227]	1HF4	1E8L	258	6	6	12	8
Human carbonic anhydrase II[92]*	2CBA		260	12	8	7	0
Human serum albumin[265]	1AO6		1170	62	35	2	34

Insulin[266]	5ENA		51	3	4	0	3
Interleukin 4 (cytokine)[267]	2B8U	1CYL	129	6	2	1	3
Interleukin 6[268]	1ALU	2IL6	186	7	3	1	2
Monellin[269]	1IV9	1MNL	192	12	14	2	0
Myoglobin (whale)[270]	1UFP		154	6	3	2	0
Neocarzinostatin[271]	1NOA		113	5	1	2	2
Odorant binding protein[272]	1A3Y		298	17	10	2	1
Papain[273]	3LFY		424	8	38	10	6
Pectate lyase C[274]	2PEC		353	8	17	7	2
Phosphatidylethanol-amine-binding protein[275]	1A44		185	3	10	5	0
Phospholipase A2 (Ca^{2+})[276]	1PSJ		124	6	7	3	7
Relaxin[277]	6RLX		104	2	2	3	6
Rhodanese[278]	1DP2		293	15	11	8	0
Ribonuclease T1[279]	1RN1	1YGW	104	4	9	1	2
RNase A[95]*	1AFU	2AAS	248	6	12	0	8
Staphylococcal nuclease[280]	1STN	2KQ3	149	3	7	1	0
Extracellular domain of human tissue factor[281]	2HFT		218	9	11	4	2
Sticholysin II[282]	1O72		350	12	24	10	0
Subtilisin BPN'[235]	1ST2		275	3	10	3	0
Thioredoxin[283]	2TRX	1XOA	216	8	4	4	2
Tryptophan synthase α-subunit[234]	1WQ5		536	24	14	0	0

CD spectra of mutants of four proteins were also calculated. Mutations from tyrosine to phenylalanine were modelled by deleting the hydroxyl group without any subsequent energy minimization. For mutations involving tryptophan, the residue is first substituted by the corresponding amino acid by the mutation prediction function on the WHAT IF server[284] followed by a steric clashes refinement.[285] Side chain replacement is one sub-step in the algorithm of building protein homology model employed by WHAT IF server. We mutate a single side chain from known crystal structure. Thus, the procedure of predicting the backbone conformation will not be covered here. For the mutated position, 40 rotamers will be extracted from a non-redundant database based on the adjacent backbone conformation. A scoring function is employed to determine which rotamer will be put in the model. The scoring function considers six terms including (1) the ratio of population of χ_1 within 75° of three most populated χ_1 value (60°, 180°, -60°) of this rotamer to the population counterpart of all rotamer; (2) RMSE of α carbon of the rotamer backbone to the corresponding α carbon of the model; (3) function of the backbone dihedral angle (φ, ϕ) differences of the model and the database hit; (4) quality of packing; (5) number of hydrogen bonds; (6) van der Waals' clashes of the side chain. Two atoms are considered to be clashes if the distance between them is shorter than the sum of their van der Waals' radii minus 0.5 Å. The rotamer with the highest score is selected. After placing in the model, the new side chain torsion angle is free to change

up to 5^{o} to get rid of van der Waals' clashes; or left completely free if there are still clashes. The mutated structure is then optimized by a 100-step energy minimization in vacuum.

Comparing the calculated CD of the native protein structure and a set of artificial conformations generated by computer (decoys) to the experiment can evaluate the ability of our vibrational parameters in identifying the native structure from the non-native conformations of protein. The 3DRobot server[286] was used to construct these decoys for 34 of the 40 proteins in our protein dataset. The other six proteins were not considered, because 3DRobot was unable to generate sufficient decoys in these cases. 3DRobot is a computer algorithm developed to generate decoys from the user-defined initial model instead of using a pre-calculated decoy set for specific proteins, which gives us the possibility to explore most of the proteins considered in this chapter. This algorithm also overcome some serious flaws in some existing decoy sets, such as bias to specific compactness pattern or secondary structure. These flaws would make the native structure to be easily distinguished from the decoys. 3DRobot employs a four-step procedure to generate a set of decoys for the target protein. Firstly, the native model is threaded through a non-redundant protein library whose pair-wise sequence identity is less than 70 % among 27822 protein structures. Up to a hundred structural scaffolds, which rank top in the

alignment to the native structure, are selected from this step. The pair-wise RMSD of the selected scaffolds is also restricted to above a threshold to ensure the diversity of the scaffolds. Then, the full length native model is split into unaligned and structurally aligned regions and is assembled by replica-exchanged Monte Carlo simulation with the selected scaffolds. The unaligned regions are modeled on a lattice system (grid = 0.87 Å) while the aligned region is modeled off-lattice. Decoys are stored in *n* bins (whose interval is 1 Å) according to RMSD to the native model. For example, 12 bins will be filled if the largest RMSD of decoys is set to 12 Å. An ideal case will be an equally distribution of the required number of decoys into all bins when generating them. A variety of the scaffolds are used to maximum the diversity of the decoys in a specific bin. The generated structures from last step are reduced models in which the residues are represented by α carbon atom and the center of mass of each side chain. The final step is to extend the reduced model into a full atom structure. A look-up table involving four neighboring α carbon atoms is used to quickly construct the backbone atoms. The all-atom structure is then refined by energy minimization. 1000 decoys were generated for each protein with an all-atom structural RMSD cut-off set to 12 Å. For multimeric proteins where the chains have identical sequences, only a single chain was used to generate decoys. For cases where the chains have distinct sequences, *e.g.*, relaxin (6RLX) and insulin (5ENA), chain

identifiers in the .pdb file were unified so that 3DRobot generated appropriate decoys.

4.3 Results and Discussion

4.3.1 *ab initio* calculation

Toluene

Hameka and Jensen[250] reported calculated energies, geometries and vibrational frequencies of the ground and first excited states of toluene using the CASSCF method with four electrons in an active space of four orbitals. Our calculation employed a larger active space: six electrons in seven orbitals, which comprise three occupied π orbitals, three virtual π orbitals and one Rydberg orbital. The latter is essential for the calculation to converge. The transition to the first singlet excited state of toluene is localized in the π orbitals of the ring. In Appendix Table A.1, the geometries of the ground and first excited states are compared with those determined by electron diffraction.[245] Although the barriers to internal rotation are very small, both states adopt staggered configurations, which is consistent with the predominant conformation of the β-carbon of phenylalanine. For the ground state, our calculations agree well with the experimental and previously calculated results. For the excited state, our calculations predict the C-C bonds of the ring to be extended by 0.04 Å compared to the ground state

geometry. This is qualitatively similar to the calculations with the smaller active space, but there were some quantitative differences.

The assignment and the calculated frequencies of the selected vibrational normal modes are in good agreement with the experimentally identified IR frequencies of toluene[242] (Table 4.2 and Appendix Table A.2). There are cases where multiple internal coordinates contribute to the same mode while some vibrational motions distribute in different fundamentals in the excited state (Appendix Table A.2).

Table 4.2 Calculated vibrational frequencies of selected normal modes of the ground (S_0), the first (^1L_b) and the second (^1L_a) excited state of toluene, *p*-cresol and 3-methylindole compared to their experimental counterparts.

	assignment	**S_0 frequency/cm^{-1}**		**^1L_b, ^1L_a frequency/cm^{-1}**		
				^1L_b		**^1L_a**
		Calc.	Expt.	Calc.	Expt.	Calc.
toluene						
	C-C bend	838	788	787	736	
	C-H bend	1107	1030	978	935	
p-cresol						
	C-C bend	890	844	845		
	C-H bend	1280	1174	1242		
3-methylindole						
	C-C stretch	806	758	775	740	784
	benzene C-H bend	1049	1009	957	919	1047

Experimentally, the vibrational structure in the absorption and the CD spectra in the near-UV of phenylalanine and derivatives shows two progressions both with a spacing of 930 cm^{-1}.[230] One arises from excitations from the ground state to the lowest vibrational level of the first electronic excited state (denoted 0-0), while the other corresponds to the 0+520 cm^{-1} vibrational state of the excited state. The progression of the transition to the 0+520 cm^{-1} vibrational level has bands of the opposite sign to the 0-0 progression in the CD spectra of phenylalanine and derivatives.

There is no severe torsion between the equilibrium geometries of the ground and the first excited state of toluene. The overlap of the vibrational wave functions between the ground and the excited states is described well by one-dimensional Franck-Condon factors in the double-harmonic approximation. Two vibrational normal modes (Figure 4.1) are Franck-Condon active. The normalized Franck-Condon integrals and the transition energies are given in Table 4.3. The CASSCF calculation gives a 0-0 transition energy of 37890 cm^{-1} (264 nm) to the first excited state. This is close to the experimental peak position of 268 nm observed in the near-UV CD spectra of phenylalanine and its derivatives.[230] In our parameters for CD calculations, we place the origin of toluene at 268 nm. The spacing in each progression used the excited state frequency of C-H bend mode. The frequency of C-C bend mode was taken to be 520 cm^{-1} (instead of the calculated 787 cm^{-1}) to

match the experimental spectra of phenylalanine.[230] The discrepancy comes from the influence of the amine and the carboxylic acid groups in phenylalanine, which elongate the C1-C7 bond. At higher energies, the vibrational structure of the spectra becomes more complicated and overlapped, due to combination modes and anharmonicity. Calculation of the CD spectra considering only the predominant modes represents a step-wise approach to increasing the complexity of the computational methodology.

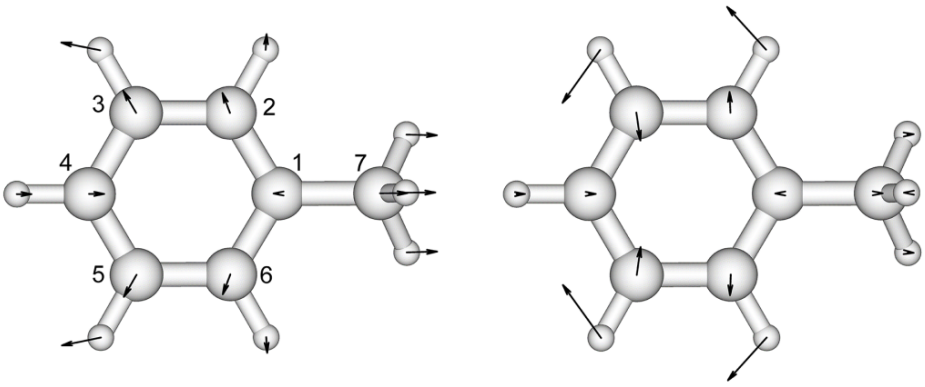


Figure 4.1 The vibrational modes which are Franck-Condon active in the excitation process of toluene. The direction and the magnitude are illustrated by the arrows located on the atoms.

Table 4.3 Transition energies and Franck-Condon integrals of transitions to different vibrational levels ('*vib*' level) of the ^1L_b states of toluene, *p*-cresol and 3-MI and the ^1L_a state of 3-MI. The vibrational level is denoted as M*v*N, where N is the assignment of the vibrational mode and M is the vibrational level of that mode. The C-C bend, C-H bend and C-C stretch mode are denoted as B_{CC}, B_{CH} and S_{CC}, respectively.

toluene			***p*-cresol**		
'*vib*' level	energy / cm^{-1}	FC integral	'*vib*' level	energy / cm^{-1}	FC integral
0	37310	0.59	0	34965	0.52
$1vB_{CC}$	37830	-0.43	$1vB_{CC}$	35810	0.58
$1vB_{CH}$	38288	0.49	$1vB_{CH}$	36221	0.17
$1vB_{CC}/1vB_{CH}$	38808	-0.35	$2vB_{CC}$	36655	0.45
$2vB_{CH}$	39266	0.26	$1vB_{CH}/1vB_{CC}$	37066	0.18
$1vB_{CC}/2vB_{CH}$	39786	-0.18	$3vB_{CC}$	37500	0.28
			$1vB_{CH}/2vB_{CC}$	37911	0.14
			$4vB_{CC}$	38345	0.14

3-MI ^1L_b			**3-MI ^1L_a**		
'*vib*' level	energy / cm^{-1}	FC integral	'*vib*' level	energy / cm^{-1}	FC integral
0	34424	0.62	0	35088	0.57
$1vS_{CC}$	35199	0.47	$1vS_{CC}$	35872	0.53
$1vB_{CH}$	35381	0.44	$1vB_{CH}$	36135	0.32
$2vS_{CC}$	35974	0.23	$2vS_{CC}$	36656	0.34
$1vB_{CH}/1vS_{CC}$	36156	0.33	$1vS_{CC}/1vB_{CH}$	36919	0.30
$3vS_{CC}$	36749	0.09	$3vS_{CC}$	37440	0.18
$1vB_{CH}/2vS_{CC}$	36931	0.17	$1vS_{CC}/2vB_{CH}$	37966	0.19
			$4vS_{CC}$	38224	0.08
			$1vS_{CC}/3vB_{CH}$	39013	0.10

p-Cresol

p-Cresol is used instead of phenol to represent tyrosine in this study to take into account the influence of the methyl group on the geometries and vibrational frequencies of the ring. Its electronic ground and first singlet excited state have been studied experimentally[238-240,244]and theoretically.[246,248-250] Our calculated CASSCF geometries of these states agree well with the X-ray determined geometry[287] and calculations using other levels of theory (Figure 4.2 and Appendix Table A.3).[244,248] The hydroxyl group is in the plane of the benzene ring while the methyl group is in a staggered configuration with one hydrogen perpendicular to the benzene ring. This is consistent with calculations at the B3LYP/cc-pVDZ level.[246] In the excited state, the bond lengths of the C-C bond of the ring are extended about 0.04 Å compared to the ground state, while the C-O bond length barely changes. In Table 4.2 (and Appendix Table A.4) the predominant normal modes that contribute to the vibrational features in the CD spectra are assigned and the calculated frequencies are compared to reported IR and Raman measurements[239] of *p*-cresol.

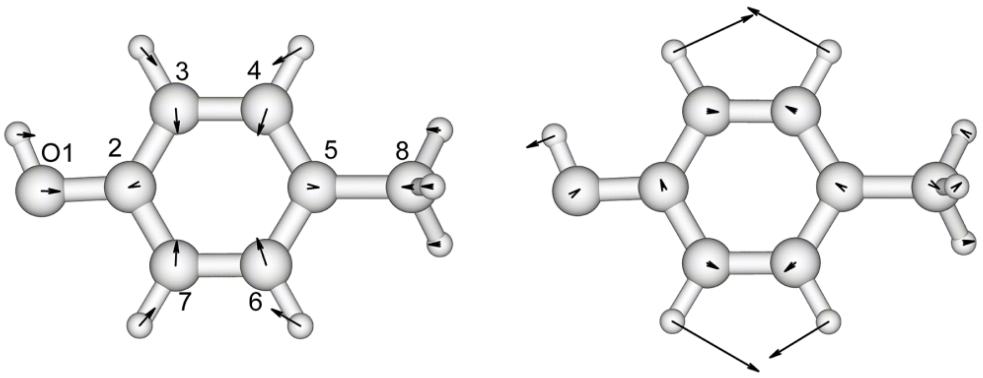


Figure 4.2 The vibrational modes which are Franck-Condon active in the excitation process of *p*-cresol. The direction and the magnitude are illustrated by the arrows located on the atoms.

The assignments and the normalized Franck-Condon integrals of *p*-cresol are shown in Table 4.3. The CASSCF calculation gave an overestimated 0-0 transition energy of 37908 cm^{-1}. We located the 0-0 transition at 34965 cm^{-1} (286 nm) based on the CD spectrum of *L*-tyrosine ethyl ester in EPA (ethyl ether-isopentane–ethanol) at low temperature.[88] The calculated excited state frequencies 890 cm^{-1} and 1280 cm^{-1} (Table 4.2) are used as the vibrational level interval in the excited state. This has been based on the transitions resolved and assigned in the resonance Raman spectrum[248] of *p*-cresol and in the absorption and CD spectra of tyrosine derivatives.[288] Two progressions, 0-0 and 0 + 1250 cm^{-1}, were identified. The interval between transitions is 800 cm^{-1} for both progressions. Horwitz *et al.*[88] suggested that

a 0+420 cm^{-1} transition is required to fill the gap between the first two vibrational transitions of the 0-0 progression. This is not taken into account in our CD calculations.

3-Methylindole

Indole, as a model compound representing tryptophan, has been studied extensively by experimental[240,289,290] and theoretical[291-296] methods. The geometries and vibrational frequencies of the ground and two excited states, ^1L_b and ^1L_a, of 3-MI have been compared to those of indole, to investigate the influence of the methyl group on the pyrrole ring.[240,247,297,298] Our CASSCF calculations of these three states of 3-MI are compared with the experimental ground state geometry[299] and the calculated ^1L_b and ^1L_a excited state geometries at various levels of theory (Figure 4.3 and Appendix Table A.5). The calculated geometries of the ^1L_b and ^1L_a states are in good agreement with those computed for indole[298] and for 3-MI[251] using CASSCF and an atomic natural orbital (ANO-s) type basis set. The C4-C5 and C8-C9 bonds show the greatest change in the ^1L_b state, while the N1-C2, C2-C3 and C6-C7 bonds change most in the ^1L_a state.

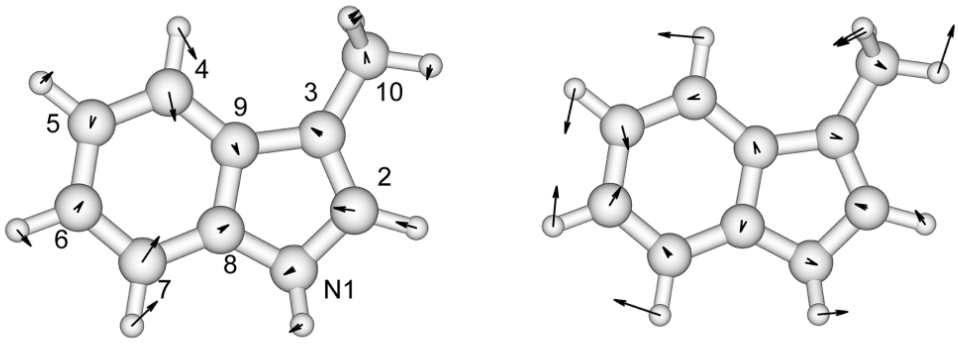


Figure 4.3 C-C stretch mode (left panel) and benzene C-H bend mode (right panel) of 3-MI which are Franck-Condon active in the excitation process. The direction and the magnitude of each normal mode are illustrated by the arrows located on the atoms.

We have compared the calculated vibrational frequencies of the ground and the two excited states with the experimental data measured with IR[300] and one-colour resonant two-photon ionization (1C-R2PI) spectroscopy (Appendix Table A.6).[301] The frequencies are in good agreement with the experiments for both the ground and the ^1L_b excited state. The contribution of each vibrational normal mode to the absorption spectrum of the ^1L_b transition has previously been studied[298] using TDDFT and different functionals with the 6-311+G** basis set. The Franck-Condon, the Herzberg-Teller and Duschinsky effects were considered. The predominant two modes identified by those calculations, which contribute to the absorption spectrum,

are with frequencies 806 cm^{-1} and 1049 cm^{-1}; we also found this using ezSpectrum.[196] The consistency of the description of the ^1L_b state by several independent sources provides confidence in the ezSpectrum calculations and their application to the less studied ^1L_a state.

The experimental absorption spectra and vibronic transitions of indole, 3-MI, *N*-acetyl-*L*-tryptophanamide and several proteins have been reported.[297] Different solvents were used to perturb the spectra in order to help to assign bands, which is otherwise difficult due to the overlap of the ^1L_b and ^1L_a bands. Transitions to the ^1L_a state are more sensitive to the environment than those to the ^1L_b state.[302] The 0-0 ^1L_a transition energy may shift by approximately 5 to 10 nm due to interactions with the environment. However, it is difficult to distinguish all the bands from various electronic transitions. For the purposes of our protein CD calculations, the origins of the ^1L_b and ^1L_a bands are placed at 290.5 and 285.0 nm, respectively, in accordance with the experimental measurements of tryptophan derivatives[238] and account for the energy shift induced by the surrounding residues or solvent. The Franck-Condon integrals for both states are shown in Table 4.3. The two progressions of the ^1L_b transition are 0-0 and 0+775 cm^{-1} with a spacing of 957 cm^{-1}, while the progressions of the ^1L_a transition are 0-0 and 0+1047 cm^{-1} with a spacing of 784 cm^{-1}.

4.3.2 Calculated CD spectra

We assess the influence of the vibrational structure of the aromatic electronic transitions on the calculated CD spectra. Two parameter sets are considered corresponding to inclusion or neglect of vibrational structure, respectively: '*vib*' and '*non-vib*'. The calculated spectra are compared with the experimental spectra for each protein, as shown in Figure 4.4 for adenylate kinase.[253] Spectra of other proteins are given in Appendix Figures C.1 to C.33. We use RMSE (Figure 4.5) and MRE to assess the accuracy (Appendix Table C.1). The Spearman rank correlation between the calculated and the experimental intensities of all 40 proteins for each integer wavelength provides another useful statistic.

The RMSE between the two parameter sets are summarised in Figure 4.5 and compared in detail in Appendix Table C.1. The wavelength range over which the RMSE is calculated depends on the associated published spectra. The '*vib*' calculation gives a smaller RMSE for all the proteins except staphylococcal nuclease (SNase). For the '*vib*' parameters, all 40 proteins have an RMSE less than 175 deg cm^2 $dmol^{-1}$, in contrast to a broader distribution of RMSE for the '*non-vib*' calculations. The analogous distributions of MRE are quantitatively similar to the RMSE. Half of the proteins show an improvement in RMSE of over 40% with the '*vib*' parameters. The RMSE is reduced by over 150 deg cm^2 $dmol^{-1}$ in the case of

five proteins: chymotrypsinogen A (Appendix Figure C.10), human carbonic anhydrase II (HCAII) (Appendix Figure C.15), papain (Appendix Figure C.22), phospholipase A2 (Appendix Figure C.25) and rhodanese (Appendix Figure C.27).

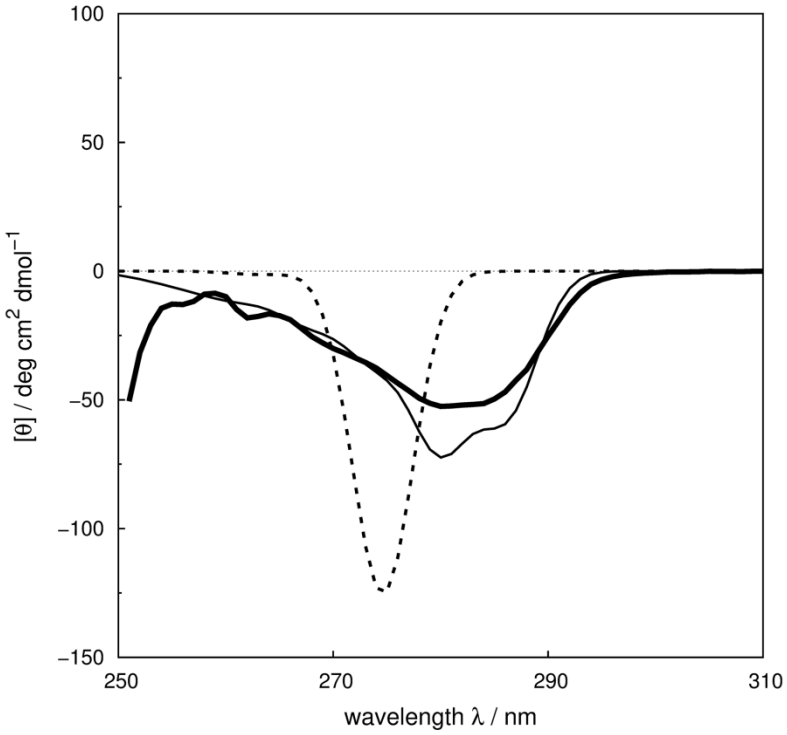


Figure 4.4 Experimental[253] near-UV CD spectrum (bold solid line) and calculated spectra of adenylate kinase (PDB code: 2ECK) with '*non-vib*' (dotted line) or '*vib*' (thin solid line) parameter sets. The RMSE of the spectrum calculated with the '*vib*' parameters is 11 deg cm^2 $dmol^{-1}$ and with the '*non-vib*' parameters is 34 deg cm^2 $dmol^{-1}$.

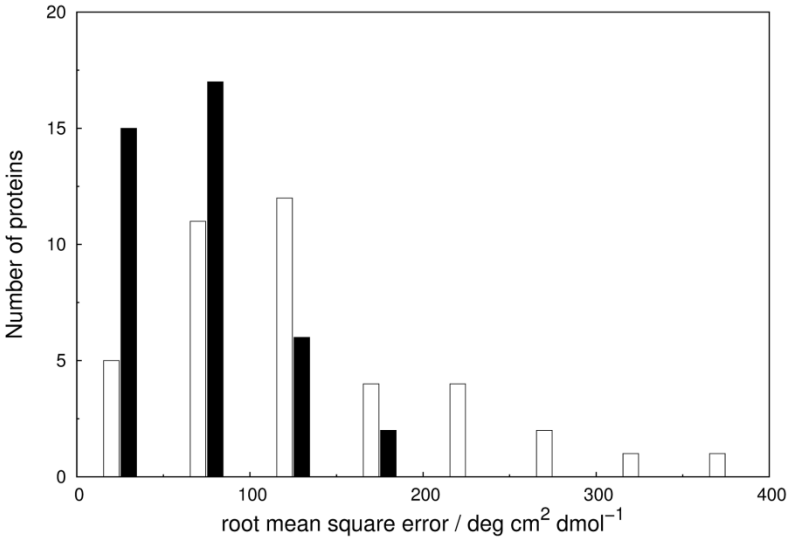


Figure 4.5 RMSE of the intensities between the experimental and calculated near-UV CD spectra, computed with '*non-vib*' (open bar) and '*vib*' (solid bar) parameters for the 40 proteins.

The Spearman rank correlation coefficients between the calculated and the experimental mean residue ellipticities are compared in Figure 4.6 for the '*non-vib*' and '*vib*' parameters. The wavelengths covered are from 260 to 300 nm. There are limited experimental spectra beyond this range; the correlation coefficients between 260 and 264 nm are calculated without myoglobin and bovine pancreatic trypsin inhibitor (BPTI), as experimental data were not available from the literature. The current wavelength range encompasses the excitation energies of the origins and the transitions with

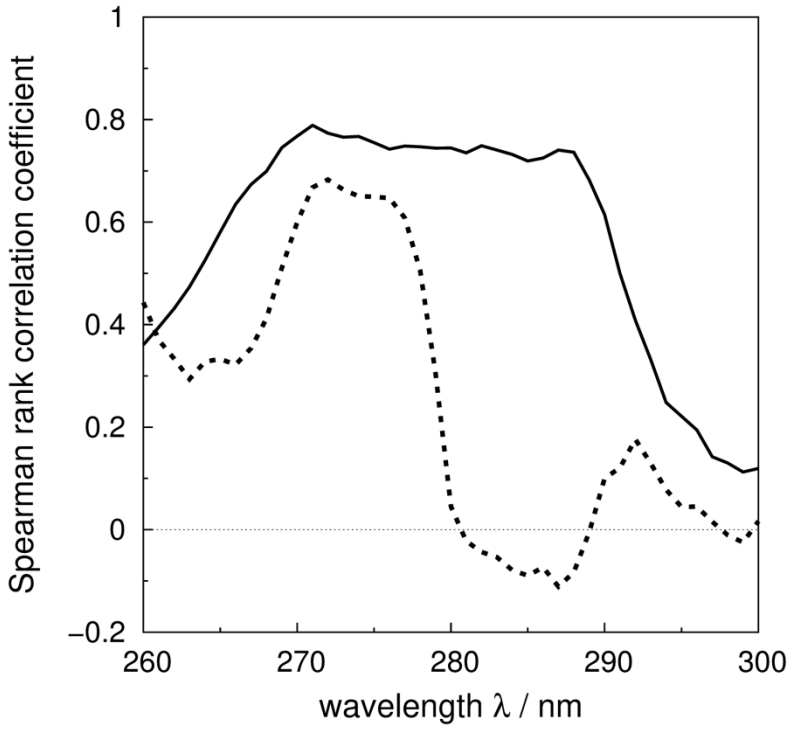


Figure 4.6 Spearman rank correlation coefficients between the calculated and experimental intensity for the '*vib'* (solid line) and the '*non-vib'* (dotted line) parameter sets, based on 40 proteins.

the largest Franck-Condon integrals of all three aromatic amino acids. The '*vib'* parameter set gives greater correlation across nearly the whole wavelength range. The correlation coefficients increase by over 0.7 between 281 and 288 nm. The lack of correlation above 290 nm may be due to neglect of the influence of the environment on the excitation energy of tryptophan.

An energy shift was calculated with a model considering the electrostatic interaction between the chromophore and the surrounding atoms.[142] However, inclusion of this effect did not noticeably improve the accuracy of the calculations (data not shown).

Illustrative spectra

We use four proteins from the set of 40 with various numbers of tryptophan residues and disulfide bonds to illustrate the effectiveness of '*vib*' parameter set in the CD calculations. *Drosophila melanogaster* calmodulin with calcium bound (Figure 4.7a) has no tryptophans or disulfide bonds.[225] The negative bands at 262 and 268 nm are from the nine phenylalanine residues in the protein, while its single tyrosine leads to a less intense broad band at longer wavelength. Compared to the '*non-vib*' parameters, which predicted a signal at 274 nm from the tyrosine, the calculation considering the vibrational structure locates the bands for the tyrosine and the phenylalanines at the same wavelengths as the experimental spectrum, with an average error in the intensity of less than 10 deg cm^2 $dmol^{-1}$ from 255 to 300 nm. *E. coli* adenylate kinase (Figure 4.4) and *E. coli* tryptophan synthase α-subunit (Appendix Figure C.33) are also proteins without tryptophans or disulfide bonds. The experimental spectrum of adenylate kinase,[253] which shows little vibrational fine structure, is reproduced quantitatively with the '*vib'* parameters.

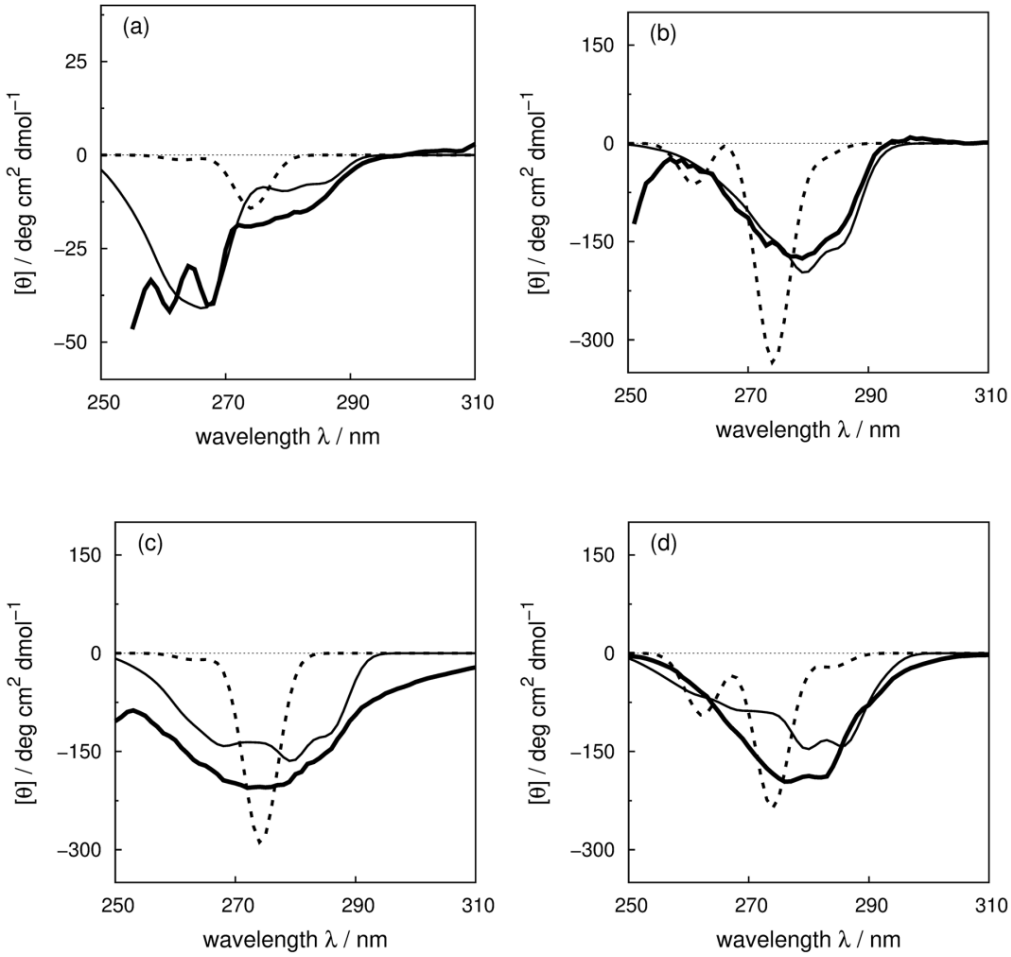


Figure 4.7 Experimental near-UV CD spectra (bold solid line) and calculated spectra with '*non-vib*' (dotted line) or '*vib*' (thin solid line) parameter sets: (a) calmodulin[225] (PDB code: 4CLN); (b) subtilisin BPN'[235] (PDB code: 1ST2); (c) insulin[266] (PDB code: 5ENA); (d) odorant binding protein[272] (PDB code: 1A3Y).

Tryptophan synthase α-subunit has a distinctive near-UV CD spectrum with both positive and negative bands,[234] while our calculation predicts an intense negative broad band with shifted peak positions. This protein has 20 pairs of aromatic rings in close proximity (less than 8 Å) which lead to strong couplings.

Bacillus amyloliquefaciens subtilisin BPN' (Figure 4.7b) consists of 275 residues including three phenylalanines, ten tyrosines, three tryptophans and no disulfide bonds.[235] The broad negative band at 278 nm in the experimental spectrum has no distinguishable fine structure and is calculated quantitatively at 279 nm with the '*vib'* parameters, but less accurately at 274 nm with the '*non-vib'* parameters. Both tyrosine and tryptophan contribute to this band. The well-reproduced rising edge of the experimental spectrum above 286 nm is solely from tryptophan, since there are no disulfide bonds. This level of accuracy reflects the effectiveness of the new parameters in describing tryptophans. Dehydroquinase II (Appendix Figure C.11) and α-lactalbumin (Appendix Figure C.2) are further similar examples. Some proteins, including barnase (Appendix Figure C.4), HCAII (Appendix Figure C.15), rhodanese (Appendix Figure C.27) and phosphatidylethanolamine-binding protein (Appendix Figure C.24), have distinct peaks or shoulders of moderate intensity above 290 nm. These

features must arise from tryptophans, since there are no other chromophores that have CD signals in this region.

Disulfide bonds forming between the side chains of two cysteines are another contributor to the near-UV protein CD spectra, which is currently omitted in our calculations. The CD bands arising from disulfide bonds are much broader than those of the aromatic amino acids appearing between 240 and 350 nm.[266] Human insulin[266] (Figure 4.7c) and porcine odorant binding protein[272] (Figure 4.7d) illustrate the contribution of the disulfide moieties. Human insulin has three disulfide bonds and no tryptophans. Tyrosine is the only aromatic amino acid that contributes to the spectrum above 270 nm in the calculation. Insulin exists mostly as a dimer when there are no stabilizing moieties, such as Zn^{2+}, phenol or glycerol, which is the case in our calculation; otherwise insulin is in an inactive hexamer form. The experimental spectrum of insulin is an intense broad negative band without clear fine structure. The two calculated extrema at 280 and 286 nm quantitatively match the rising edge at the same wavelengths which arises from a combination of individual bands of the four tyrosines. However, the broad band above 290 nm is not generated, due to the absence of the transitions of the disulfide bond in the calculation. The experimental and calculated spectra of BPTI[93] (Appendix Figure C.8) and bovine RNase A[95] (Appendix Figure C.29) are very similar to those of insulin. The overall

shapes of the experimental spectra of these two proteins are reproduced well, especially for the region where only tyrosine contributes.

The assessment of the calculations becomes more complicated when all three aromatic chromophores and disulfide bonds are present. Figure 4.7d displays the experimental and calculated CD spectra of porcine odorant binding protein.[272] There are two negative peaks with comparable intensities at 276 and 282 nm and a shoulder at 290 nm in the experimental spectrum. Our calculation reproduced peaks at 276 nm and 282 nm, which arise from both tyrosines and tryptophans, although approximately 15 deg cm^2 $dmol^{-1}$ less intense and red-shifted by about 4 nm compared to the experiment. The shoulder at 290 nm is predicted by the calculation with a 17 deg cm^2 $dmol^{-1}$ discrepancy at the correct wavelength. Other examples in the dataset where the vibrational fine structure in the experimental spectra is well reproduced with the '*vib*' parameters include papain (Appendix Figure C.22), thioredoxin (Appendix Figure C.32), phospholipase A2 (Appendix Figure C.25), ribonuclease T1 (Appendix Figure C.28) and pectate lyase C (Appendix Figure C.23). However, the calculated intensities are not well estimated for some of these proteins. One reason may be the omission of the disulfide bond in our calculation. The band shapes and intensities vary widely depending on conformation[303-305] and vicinal interactions.[306] However, there must be other explanations for the band shift and the intensity difference for

proteins without disulfide bonds; such cases include dihydrofolate reductase (Appendix Figure C.12), rhodanese (Appendix Figure C.27) and phosphatidylethanolamine-binding protein (Appendix Figure C.24).

Solvent is a universal factor that affects both the excitation of the aromatic chromophores and the structure of the protein. The influence of solvent on the ground and the excited state orbitals is evident in the difference in the absorption and CD spectra measured in gas phase and in various solutions.[297] Another consequence is broadened spectra, caused by interactions of varying degrees between the chromophore and surrounding solvent molecules. There are many environments in proteins for aromatic residues, spanning a range from buried to solvent exposed. These are currently omitted in our calculations; as noted earlier, exploratory calculations, to model these effects did not yield more accurate calculated CD spectra and this will be addressed in future work.

4.3.3 CD spectra of mutants

Comparison of the CD spectra of a wild-type protein and its mutants can reveal the contributions of individual aromatic residues to the spectrum. This kind of data can be difficult to extract directly from experimental spectra of wild-type proteins due to the complexity caused by a combination of multiple

contributors. Tryptophan and tyrosine are often mutated to phenylalanine, while phenylalanine is usually mutated to leucine as the similar volume of the side chain minimizes the perturbation of the overall structure of the protein and no other ambiguous features should appear in the CD spectra of the mutants. The availability of CD spectra of mutants allows us to assess the effectiveness of our new parameters.

Barnase

Vuilleumier *et al.*[94] have reported CD spectra of three tyrosine and three tryptophan mutants of barnase, which is a homo-trimer. The monomer consists of 110 residues including four phenylalanines, seven tyrosines and three tryptophans. The three chains in the crystal structure (PDB: 1A2P) adopt essentially identical core conformations.[307] However, the residues on surface show some flexibility, which is consistent with the NMR structure (PDB: 1FW7).[308] Comparable calculated spectra (Appendix Figure C.4) confirm the similarity of the crystal and NMR structures. The ^{17}Tyr residue adopts distinct conformations in the three chains of the crystal structure. The multimeric structure is used in the calculations of both wild-type and mutants. The predominant changes in the experimental spectra arising from the tyrosine mutations appear between 260 and 290 nm, the region in which tyrosine chromophore absorbs. The trend in intensities is predicted well by the calculation (Appendix Figures C.34a and C.34b) across the mutants,

although the intensities are underestimated for all calculated spectra compared to their experimental counterparts. The intensity changes are reproduced quantitatively for all three tyrosine mutants above 270 nm in this protein (Figure 4.8b). Both $^{1}L_{b}$ states of tyrosine and phenylalanine contribute to the region below 270 nm where the calculated spectra did not predict the changes of Y78F and Y97F observed in the experimental spectra.

Our theoretical model is sensitive to subtle ring flips. There is an X-ray structure of the Y78F mutant of barnase (PDB: 1BAO) whose calculated spectrum is compared in Figure 4.8b with the result calculated using a structure generated by deleting the OH group from the wild-type structure. Both structures give a positive band. The X-ray structure gives a difference spectrum slightly closer to the experimental one. Table 4.4 compares the differences in key residue-residue interactions (distances and TDM orientations) of two Y78F mutants (the X-ray determined structure and the one derived from deleting the OH group directly) for residues less than 8 Å apart, which might be considered to be strongly interacting. The nearest aromatic residue to ^{78}Tyr is ^{35}Trp at 8.8 Å, which means the intensity changes between the wild-type and mutated barnase are predominantly from the tyrosine. Whilst the various pairwise distances are very similar, the relative orientations of ^{13}Tyr-^{17}Tyr, ^{56}Phe-^{103}Tyr and ^{103}Tyr-^{106}Phe appear to be better

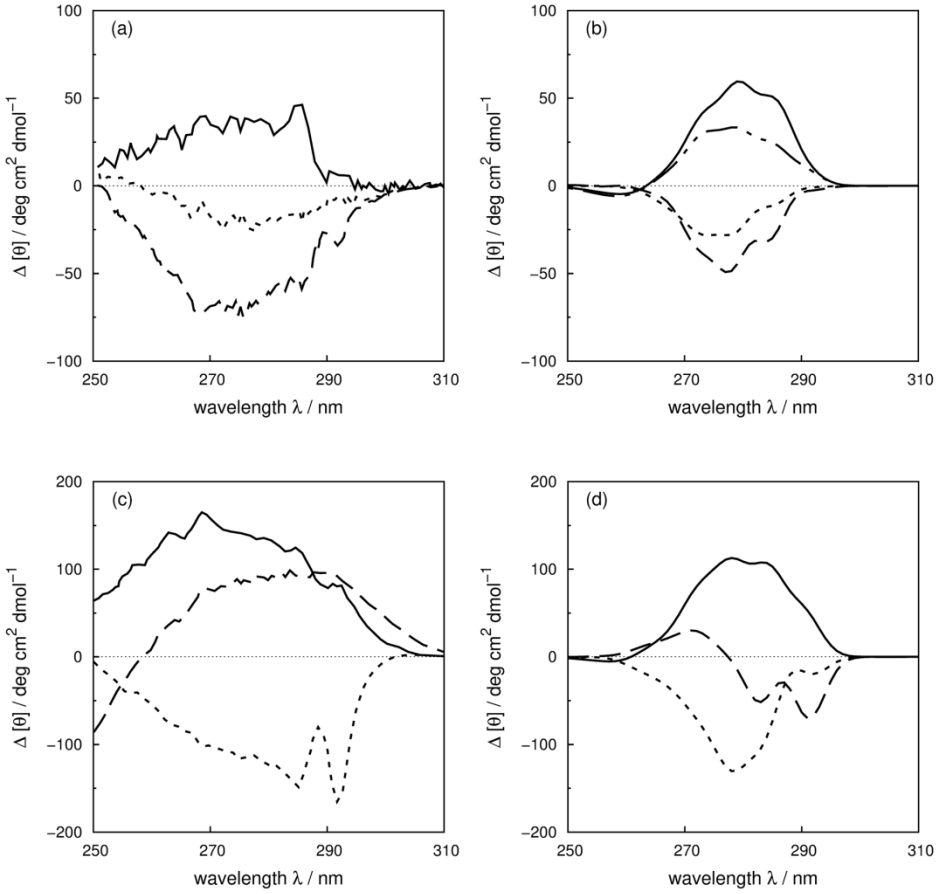


Figure 4.8 Difference spectra (wild-type minus the mutants) of barnase (experimental: (a), (c); calculated: (b), (d)). (a) and (b): ΔY78F (solid line), ΔY90F (dotted line), ΔY97F (dashed line); (c) and (d): ΔW35F (solid line), ΔW71F (dotted line), ΔW94F (dashed line). The dotted-dash line in Figure 4.8(b) is the difference spectrum calculated with X-ray structure of Y78F barnase (PDB: 1BAO).

Table 4.4 Comparison of distances and angles between TDMs of residue pairs with strong interactions. The distance is measured between the centroids of the aromatic side chains. The structures used in the calculations were generated by directly deleting the hydroxyl group from ^{78}Tyr of wild-type barnase (1A2P-Y78F) or taken from the structure of the X-ray structure mutant (1BAO).

Residues	Distance/Å		Angle between TDMs/degree			
	1A2P-Y78F	1BAO	1A2P-Y78F		1BAO	
			^1L_b	^1L_a	^1L_b	^1L_a
^{13}Tyr – ^{17}Tyr	4.2	4.4	115		83	
^{56}Phe – ^{71}Trp	7.5	7.7	45	53	43	61
^{56}Phe – ^{103}Tyr	5.6	5.3	44		151	
^{56}Phe – ^{106}Phe	6.2	6.2	68		74	
^{71}Trp – ^{97}Tyr	6.6	6.5	108	86	114	82
^{90}Tyr – ^{94}Trp	7.0	7.0	86	173	87	168
^{97}Tyr – ^{106}Phe	7.4	7.4	118		113	
^{103}Tyr – ^{106}Phe	7.3	7.1	111		80	

represented by the X-ray structure of the mutant than the mutant structure modelled from the wild-type.

For the tryptophan mutants of barnase, the trend of intensity changes is predicted well by the calculation (Appendix Figures C.35a and C.35b), although the calculated contribution of ^{94}Trp agrees less well with the experiment above 270 nm. ^{35}Trp and ^{71}Trp are more buried, with solvent accessible areas of 18 and 8 Å^2, compared to 80 Å^2 for ^{94}Trp. This explains

the more obvious vibrational features of ^{35}Trp and ^{71}Trp in the difference spectra (Figure 4.8c). The contributions of ^{35}Trp and ^{71}Trp are quantitatively reproduced around 285 nm, the origin of the $^{1}L_a$ state of tryptophan in our parameters. The $^{1}L_b$ transition of ^{71}Trp, which is the intense transition for this residue,[94] is underestimated by our calculation. The calculated spectrum of W94F gave a band with the opposite sign compared to the experimental spectrum above 270 nm. This may be related to the larger solvent accessible area; also ^{94}Trp is not in a well-defined secondary structure and the surrounding local structure may be sensitive to the mutation.

Bovine pancreatic trypsin inhibitor

Sreerama *et al.*[93] have systematically studied the contributions of four phenylalanines and four tyrosines in bovine BPTI in the far- and the near-UV region, by replacing them with leucine. The CD spectra of the BPTI mutants show distinct changes in intensities and peak positions in the far-UV, which may come from the contributions of the mutated residues or secondary structure changes associated with the mutations. There are indeed some conformational differences among the X-ray structures of BPTI mutants: F22A (PDB: 1BTI), F45A (PDB: 1FAN) and Y23A (PDB: 1BPT). The root mean square distances are 0.92, 0.59 and 0.25 Å between the wild-type and the mutated structures, respectively. These three aromatic residues are buried; the solvent accessible surface areas are 22, 16 and 15 Å^2, respectively.[93]

^{22}Phe and ^{23}Tyr are in a β-strand, while ^{45}Phe is in a β-bridge. Calculations with the crystal structures of the mutants are compared with experimental spectra of wild-type and the mutants of F22L, F45L and Y23L in Figure 4.9, which shows the trend of the intensity changes is reproduced well.

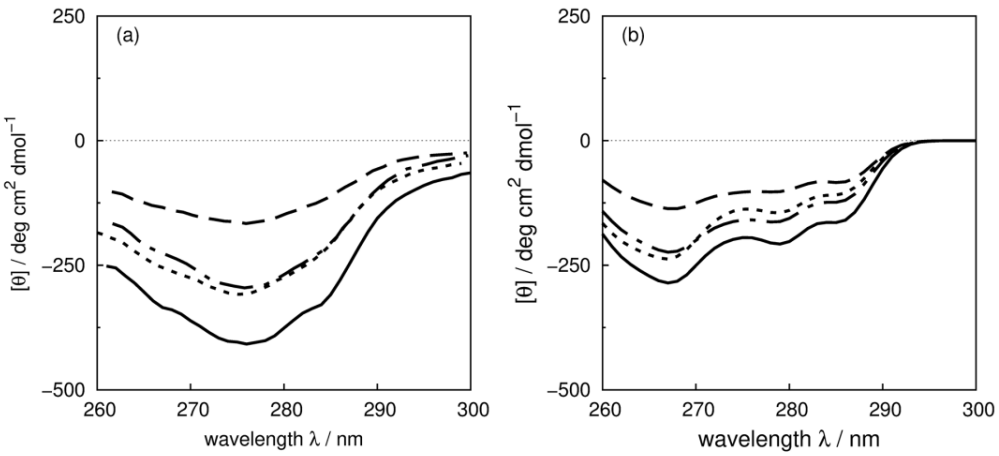


Figure 4.9 (a) Experimental CD spectra of wild-type BPTI (solid line) and mutants F22L (dashed line), F45L (dot-dashed line) and Y23L (dotted line); (b) calculated CD spectra from X-ray structures: wild-type (PDB: 5PTI, solid line), F22A (PDB: 1BTI, dashed line), F45A (PDB: 1FAN, dot-dashed line) and Y23A (PDB: 1BPT, dotted line).

Ribonuclease A

Woody and Woody[95] have studied the contributions of tyrosines in the CD spectrum of RNase A by replacing tyrosine with phenylalanine. There is no obvious secondary structures change, as shown in the far-UV spectra of the wild-type protein and the mutants.[95] Figure 4.10 compares the experimental and calculated changes arising from the mutations. The contributions of ^{25}Tyr, ^{76}Tyr and ^{115}Tyr are reproduced quantitatively. The calculated shift of 6 nm of the intense peaks of ^{25}Tyr and ^{115}Tyr agrees well with the 5 nm shift observed experimentally. The contribution of ^{73}Tyr is

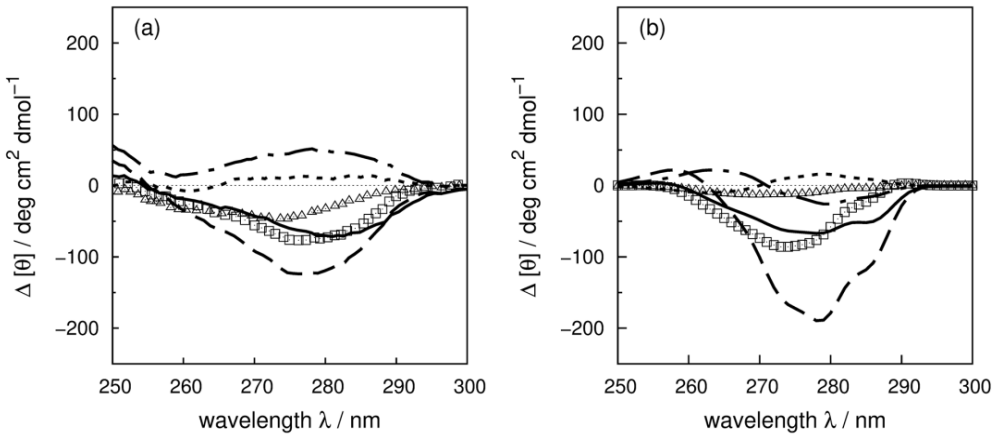


Figure 4.10 Difference spectra (wild-type minus the mutants) of RNase A ((a): experiment; (b): calculation). ΔY25F (solid line), ΔY73F (dashed line), ΔY76F (dotted line), ΔY92F (triangle), ΔY97F (dot-dashed line) and ΔY115F (square).

overestimated, while that of ^{92}Tyr is underestimated but has the right sign. There are two pairs of interactions, ^{73}Tyr-^{115}Tyr and ^{25}Tyr-^{46}Tyr, with respective separations of 5.2 and 5.7 Å, in RNase A, which explains the intense contributions of ^{25}Tyr, ^{73}Tyr and ^{115}Tyr. The calculated contribution due to ^{97}Tyr has the wrong sign between 270 and 290 nm. The hydroxyl group of ^{97}Tyr forms a hydrogen bond with the carbonyl oxygen of ^{41}Lys which is in the binding pocket. The activity of this protein is affected by replacing ^{97}Tyr with phenylalanine.[95] This may indicate a conformational change of local structure of the binding site, which may be missing from our modeled structure of the mutant and may explain the discrepancy between the calculated and experimental spectra of mutant Y97F.[309]

Human carbonic anhydrase II

The contributions of tryptophan to the CD spectrum of HCAII have been studied by Freskgård *et al.*[92] through site-directed mutagenesis. The mutants are generated from a pseudo-wild-type HCAII, in which the ^{206}Cys has been mutated to ^{206}Ser to eliminate the potential disulfide bond forming between this residue and the cysteine used to replace tryptophan. All tryptophans in HCAII have at least one other aromatic residue within 8 Å which leads to complex spectra.

Figure 4.11 shows the experimental and the calculated difference spectra for the wild-type HCAII and the tryptophan mutants (experimental and calculated spectra of the wild-type and the mutants are shown in Appendix Figure C.36). ^{97}Trp and ^{245}Trp have the largest contributions to the near-UV CD spectrum which is reproduced well by the calculation. The extrema at 287 nm in the experimental spectra are slightly overestimated by the calculation. The extrema at 296 nm, which indicate a significant shift in the transition energy due to the environment, are not reproduced.

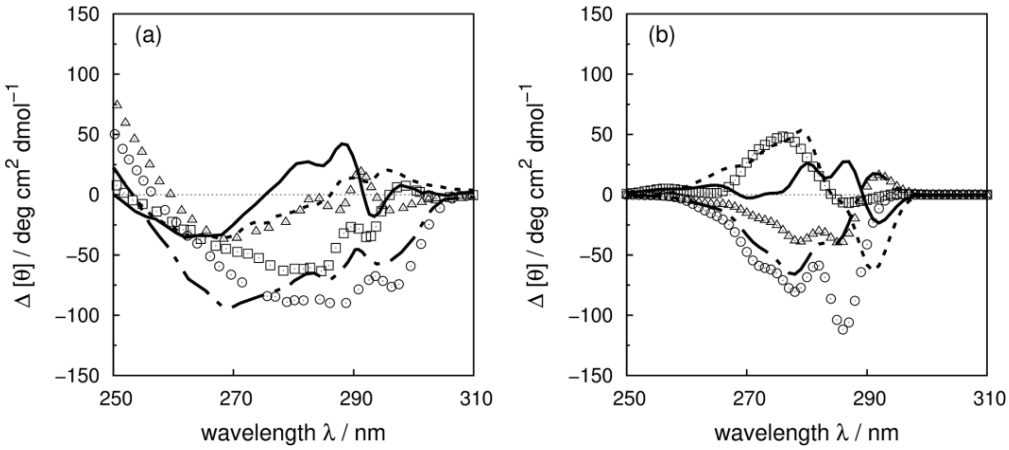


Figure 4.11 Difference spectra (wild-type minus the mutants) of HCAII ((a): experiment; (b): calculation); ΔW5F (solid line), ΔW16F (square), ΔW97C (circle), ΔW123C (triangle), ΔW209F (dotted line) and ΔW245C (dot-dashed line).

The 0-0 transitions are at 34424 and 35088 cm^{-1} (290.5 and 285.0 nm) for the ^1L_b and ^1L_a states of 3-MI, respectively. The contribution of ^{97}Trp, which is reproduced well, shows clear vibrational fine structure due to the buried environment. However, the shifts of ^{97}Trp and ^{245}Trp spectra due to the various local environments are not reproduced by the current calculation. ^{5}Trp and ^{16}Trp at the N-terminus are separated from most of the aromatic residues, except ^{7}Tyr and ^{20}Phe. The contribution of ^{5}Trp is reproduced well both for the intensity and the peak position. However, the calculation of ^{16}Trp gives the opposite sign to the experimental spectrum. This may be caused by local conformational changes of ^{16}Trp mutants due to the loss of the hydrogen bonds between the indole ring and the carbonyl oxygen on ^{7}Tyr and ^{11}Asn. W209F is the other mutant where the calculation does not reproduce the experiment well. ^{209}Trp is part of the active site where Zn^{2+} is bound to HCAII. Mutation of ^{209}Trp decreases the activity by 66% and may indicate a less tight binding mode of Zn^{2+} and a weaker interaction between ^{209}Trp and the ion.[310] This may cause a conformational change in the protein.

4.3.4 Influence of conformational dynamics

Compared to the crystal structure determined by the X-ray diffraction, protein structures obtained by solution NMR contain information about the

conformational dynamics. Calculations with an ensemble of NMR structures can account for some of the conformational flexibility in solution. The NMR structures for 13 of the proteins in our dataset (Table 4.1) are available in the PDB.[62] The influence of the structures used in the calculation (X-ray *vs* NMR) to the calculated spectra is similar for both '*non-vib*' and '*vib*' parameter sets. Combination of the NMR structures with the '*vib*' parameters better reproduced the experimental spectra than with the 'non-*vib*' parameters. Comparison between calculations with X-ray and NMR structures of apolipophorin III[256] (Figure 4.12a) and SNase[280] (Figure 4.12b) illustrates the influence of taking multiple conformations into account on the calculated intensity. Calculations considering a family of conformations also better predict the intensities of BPTI (Appendix Figure C.8) and thioredoxin (Appendix Figure C.32). Compared to the calculations with a crystal structure, the average intensity calculated over multiple frames can be stronger or weaker. For BPTI and thioredoxin, the spectra calculated from individual frames vary only in intensity, without significant changes in the sign of bands or the overall shape compared to calculations on crystal structures. This reflects the minor differences in conformations between NMR and X-ray structures. Calculations on NMR ensembles of other proteins show greater variation in band shapes due to distinct conformations.

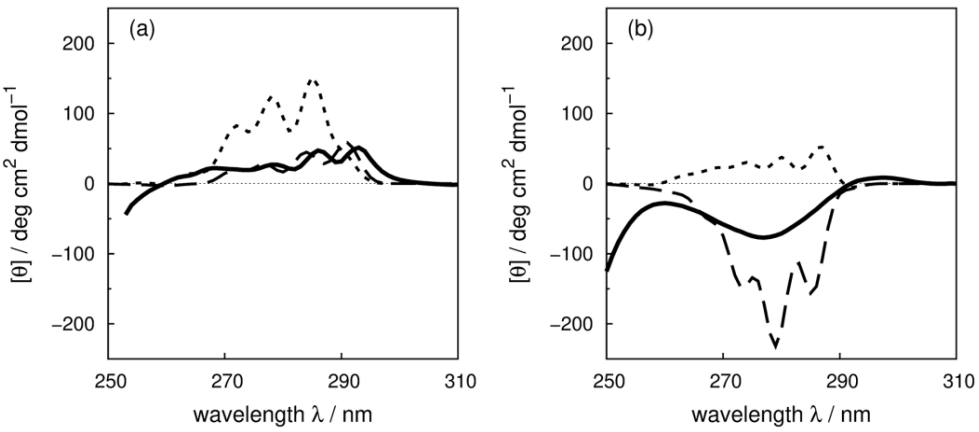


Figure 4.12 Comparison between the experimental spectrum (solid line) and calculated spectrum with '*vib*' parameters from the X-ray structure (dotted line) and the NMR structures (dashed line) of (a): apolipophorin III[256] and (b): SNase[280].

The spectrum of apolipophorin III[256] (Figure 4.12a) calculated from the NMR structures compared to the spectrum from the X-ray structure is less intense and red-shifted by approximately 6 nm and is much closer to the experimental spectrum. For SNase[280] (Figure 4.12b), spectra calculated using individual frames from the NMR structures all show a broad negative band at 279 nm with an intensity ranging from -310 to -50 deg cm^2 dmol^{-1}; the opposite sign is obtained using the X-ray structure. The X-ray structure of SNase is five residues shorter than the NMR structure at the N terminus

due to a lack of electron density in the X-ray structure determination;[311] ^{141}Ser is the last residue at the C-terminus in the X-ray structure, whereas it is ^{140}Trp in the NMR structure (denoted as SNase140). However, there are no missing aromatic amino acids in the truncated sequence. A study of SNase folding has highlighted the role of ^{141}Ser in maintaining the native structure and fluorescence spectra of SNase truncated at ^{140}Trp and ^{141}Ser show variation in the conformation of ^{140}Trp.[312] Calculations based on NMR structures do not always reproduce the experimental spectrum better than the X-ray structure. Such cases include lysozyme (Appendix Figure C.14), interleukin 6 (Appendix Figure C.18), monellin (Appendix Figure C.19) and ribonuclease T1 (Appendix Figure C.28). The influence of conformational dynamics could also be explored using MD simulations. Some individual cases have been studied[152,221,313] and we are planning a more comprehensive investigation.

4.3.5 CD calculation of native proteins and their decoys

In the final part of this chapter, we investigate to what extent a comparison between the computed and experimental near-UV CD spectra can be used to identify a native structure from a large number of plausible but incorrect model structures (decoys). Decoys are computer-generated conformations of proteins, which are widely used for training protein force

field and studying protein folding. Good parameters in the CD calculations should be able to predict the features of the experimental spectrum better with the native conformation than the decoys. We employed the online server 3DRobot to generate a set of 1000 decoys for each protein in our dataset. 3DRobot server exported two scores, the evenness score and the normalized pair-wise RMSD (npwRMSD), ranging between zero and one to characterize the quality of the decoy sets (Table 4.5).

The evenness score evaluates the distribution of RMSD of decoys to the initial model. The requested RMSD space 0 to 12 Å is divided into 12 bins with an interval 1 Å. 3DRobot employs eqn. 4.1 to calculated this score.

$$\text{Evenness} = 1 - \sqrt{\frac{\sum_{i=1}^{n_b}(n_i - \bar{n})^2}{\bar{n}^2 n_b (n_b - 1)}} \qquad (4.1)$$

where n_b is number of bins, n_i is the number of decoys in the i^{th} bin and $\bar{n}$ is the average number of decoys in n_b bins. An evenness score of unity corresponds to an even RMSD distribution of the decoys into n_b bins. An evenness score of zero represents the extreme case that all decoys fall within a single bin. A low evenness score indicates gaps in the RMSD space. Decoy sets of 20 proteins have an evenness score greater than 0.9, *i.e.*, the distribution of decoys is appropriate. The lowest evenness score is for

rhodanese (1DP2, 0.551) with around 700 decoys accumulating in the 2 and 3 Å bins. The decoy sets of three proteins, relaxin (6RLX), tryptophan synthase α-subunit (1WQ5) and HCAII (2CBA), with evenness scores less than 0.8 are either lacking decoys in the near-native state (RMSD < 3 Å) or in the distant-native end (RMSD above 7 Å).

The npwRMSD score reflects the diversity of the decoys. A good decoy set is necessarily non-redundant so that a maximum conformational space can be represented by a limited number of decoys. Since the decoys with a low RMSD to the native structure are expected to have high structural similarity, a normalization function is used when calculating the pair-wise RMSD of decoys. The normalization function is represented by $RMSD_n(r_i, r_j, L)$ in eqn. 4.2 for calculating npwRMSD.

$$npwRMSD = \frac{2}{N(N-1)} \sum_{i,j}^{N} \frac{RMSD_{i,j}}{RMSD_n(r_i, r_j, L)} \qquad (4.2)$$

where N is the total number of decoys defined by the user, the $RMSD_{i,j}$ is the pair-wise RMSD of decoys in the set. $RMSD_n(r_i, r_j, L)$ is calculated using a set of random non-redundant proteins (4928 PDB structures) with low sequence identity (less than 30%) as a reference pool to eliminate the dependence of pair-wise RMSD of decoys on their structural similarity with the native structure. The variable r_i and r_j represent the RMSD of the i^{th}

and j^{th} decoys to the native conformation while L is the target protein length. To calculate this normalization function, a protein with length L in the reference pool are selected and the pair-wise RMSD of the rest proteins in the pool is calculated for proteins whose RMSD = r_i and r_j to the selected one. Only proteins with length ≥ L are used to calculate the pair-wise RMSD of the reference protein pool. All proteins with length L in the reference pool are considered individually and then averaged. The decoy sets of all the proteins have an npwRMSD over 0.85 corresponding to good diversity.

Table 4.5 The 34 proteins studied: name, PDB codes, evenness score, npwRMSD, Spearman rank correlation of the native structure (experimental *vs* calculated intensities, in integer intervals between 250 nm or the lowest available signals and 300 nm) and its location (percentile) in the distribution generated by the decoy sets.

Protein	**PDB entry**	**even-ness**	**npw-RMSD**	**Spearman rank correlation**	**rank in decoys / %**
Adenylate kinase	2ECK	0.879	0.990	0.89	1
Calmodulin	4CLN	0.911	0.931	0.93	1
Papain	3LFY	0.902	0.982	0.91	1
Ribonuclease T1	1RN1	0.909	0.971	0.75	1
Sticholysin II	1O72	0.815	0.991	0.81	5
Apolipophorin III	1AEP	0.943	0.990	0.71	6
HCAII	2CBA	0.719	0.980	0.77	6
Phospholipase A2 (Ca^{2+})	1PSJ	0.917	0.924	0.87	6
RNase A	1AFU	0.907	1.007	0.96	6
β–Lactamase	1BTL	0.835	0.987	0.93	8

Insulin	5ENA	0.825	0.874	0.93	8
Hen egg white lysozyme	1HF4	0.900	0.971	0.76	9
Subtilisin BPN’	1ST2	0.818	0.979	0.89	9
Neocarzinostatin	1NOA	0.933	0.977	0.81	10
Phosphatidylethan-olamine-binding protein	1A44	0.864	0.992	0.78	10
Dihydrofolate reductase	4P3Q	0.914	0.988	0.63	13
Relaxin	6RLX	0.733	0.864	0.82	15
Monellin	1IV9	0.923	0.933	0.59	21
Extracellular domain of human tissue factor	2HFT	0.838	0.975	0.66	25
BPTI	5PTI	0.916	0.895	0.66	29
Thioredoxin	2TRX	0.932	0.993	0.67	29
Odorant binding protein	1A3Y	0.911	0.979	0.76	34
α-Lactalbumin	1A4V	0.900	0.953	0.2	37
Rhodanese	1DP2	0.551	0.998	-0.1	41
β-2 microglobulin (human)	1LDS	0.912	0.959	0.14	45
Barnase	1A2P	0.911	0.953	0.07	46
Interleukin 6	1ALU	0.874	0.955	-0.38	48
Chymotrypsinogen A	2CGA	0.797	0.946	-0.08	53
Dehydroquinase II	2BT4	0.918	0.980	0.01	54
Cardiotoxin	4OM4	0.911	0.949	-0.47	71
Myoglobin (whale)	1UFP	0.895	0.978	-0.8	73
SNase	1STN	0.909	0.996	-0.39	81
Interleukin 4 (cytokine)	2B8U	0.911	0.974	-0.43	90
Tryptophan synthase α-subunit	1WQ5	0.784	0.918	-0.74	100

We computed CD spectra of the native conformation and 1000 decoys of each protein in our dataset with the vibrational parameters. All 1001 calculations were compared with the experimental spectrum of the native protein individually to assess the similarity between the prediction and the observation. To assess the similarity of two spectra, we use the Spearman rank correlation coefficient between the computed and experimental intensities at each integer wavelengths (in nm) across the near-UV range (Table 4.5). This measure gives a useful indication of the overall shape and qualitative features of the spectra. For example, the Spearman correlation coefficient approximates one will indicate similar shapes of two spectra; levels of similar intensity are not necessary though. The rank of the Spearman correlation of the native conformation in the pool of the results of decoys tells the effectiveness of our vibrational parameters in distinguishing conformations from various sources in the CD calculations. Higher rank means more potential in identifying the native structure. For 17 out of the 34 proteins, the Spearman rank correlation between the computed and experimental spectra of the native structure is in the top 15% of all the evaluated models. The native structure of ribonuclease T1, papain and adenylate kinase are in the top 1%. The rank of the native structure is related to its actual value of the Spearman rank correlation and the distribution patterns of the Spearman rank correlation of the decoys which are different

for various proteins. Histograms are shown in Figure 4.13 to represent distinctive behaviour of the Spearman rank correlation distribution of decoys.

4.4 Conclusions

In this chapter, the vibrational structure of electronic transitions of aromatic amino acids is incorporated into the calculations of the near-UV CD spectra of proteins by extending the matrix method approach to include new parameterizations of chromophores based on the Franck-Condon integrals of toluene, *p*-cresol and 3-MI. Qualitative and quantitative improvements in the computed near-UV CD spectra of proteins have been achieved through the first systematic inclusion of vibrational structure. Calculations on site-directed mutants give good predictions of the contribution of individual aromatic residues when the conformation of the protein is not severely affected by the mutation. Further improvements are obtained for some proteins by computing with the NMR determined structure which emphasizes the important role of conformation diversity. Calculations with our new vibrational parameters predict half of the cases (34 proteins in total) that the native structures of those proteins are in the top 15 % Spearman rank

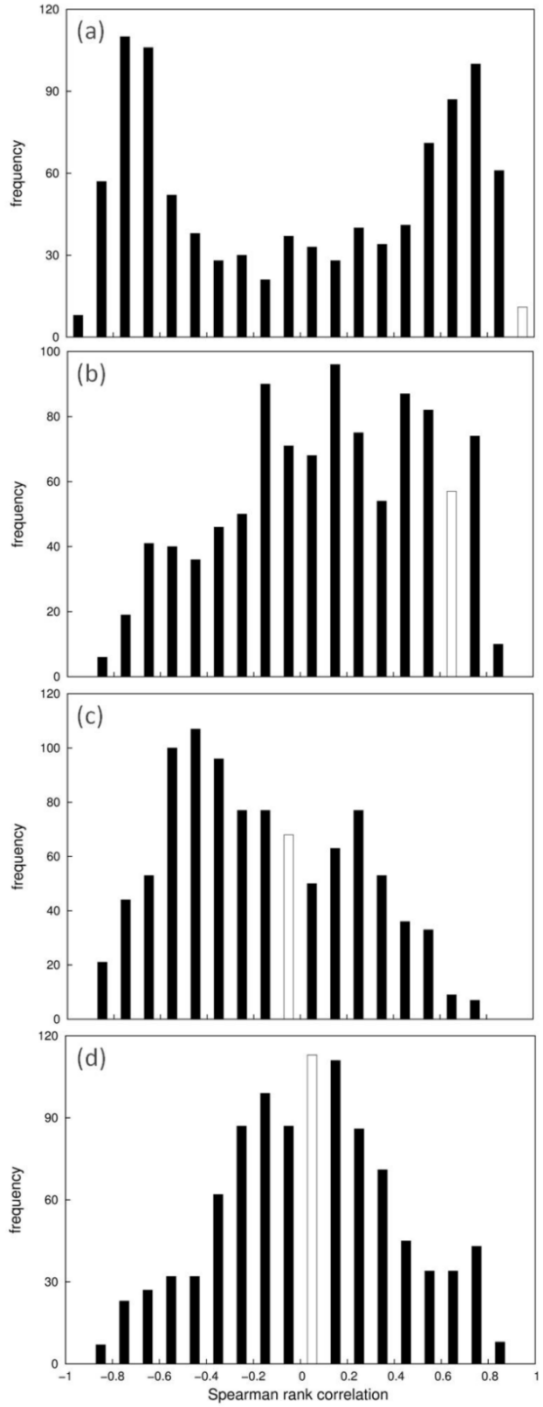


Figure 4.13 Histogram representing the proteins whose Spearman rank correlation of the native structure in decoys set falls in the (a) upper 10%; (b) upper 25%; (c) upper 50%; (d) bottom 50%. The open bar shows where the native structure appears in the Spearman rank correlation distribution of the native and the decoy structures.

correlation among 1000 computer-generated misfolded structures, which shows a potentially useful ability to distinguish between native and incorrect protein structures. We will be developing the work further, for example, by generating parameters for disulfide bonds and by further consideration of the influence of solvent and conformational dynamics.

Chapter 5

Conclusions

Proteins are an important class of biomolecules, studied in different scientific disciplines on a diverse range of aspects, such as manipulating structures by engineering methods, serving as drug targets, studying their biological functions, etc. A solid basis is knowledge of the structures. CD spectroscopy is a useful technique in revealing information about conformational changes of proteins, though not at atomic level, because of its ease of measurement. Thus, CD spectroscopy has been much studied both experimentally and theoretically, providing new insight into fundamental aspects of protein behaviour. However, the difficulty in interpretation of the signal remains, as the CD spectrum is a summation of signals arising from individual chromophores and also an averaging of a distribution of conformers. Theoretical calculation is very promising in providing a linkage between the spectroscopic information and atomic level structures. This thesis has aimed to shed light on the importance of considering more detailed

aspects of electronic structure of chromophores in the theoretical model, the matrix method.

The parameterization of the matrix method calculation is based on the *ab initio* calculations of chromophores. Description of the electronic ground and excited of chromophores accurately is essential for reproducing experimental features from CD calculation. In this thesis we have expended the parameters describing the electronic features of chromophores. We have calculated the ground and excited state geometries and frequencies of NMA, toluene, *p*-cresol and 3-MI. These results have been used to derive the Franck-Condon overlap integrals and to incorporate the vibronic coupling in the matrix method. Our high-level *ab initio* calculations were consistent with the previous experimental and theoretical results for these molecules. Franck-Condon integrals calculations led to five vibronic transitions in the $\pi_{nb}\pi^*$ excited state of NMA. Six and eight vibrational levels were taken into account in the ^1L_b excited state of toluene and *p*-cresol, respectively. Two excited states, ^1L_b and ^1L_a, of 3-MI appear in the near-UV region. We extended the parameter of 3-MI by seven vibronic states for the ^1L_b state and nine for the ^1L_a state. Considering the vibrational features of the chromophores provided some further insight into the physical origins of the nature of protein CD spectra.

We compared our CD calculations with experiment to assess the influence of incorporation of vibrational structure in both far- and near-UV. A set of 49 proteins were calculated in the far-UV. The negative peaks between 205 and 225 nm, from $\pi_{nb}\pi^*$ and $n\pi^*$ transitions, are better reproduced with vibronic parameters in α-helical proteins. CD spectra of β-II proteins with less regular conformations are difficult to reproduce. When taking the vibrational structures into account, the Spearman correlation coefficients between the experimental and the calculated intensities show improvement over most wavelengths for β-II proteins. No obvious improvement is observed for mixed-α,β and β-I proteins. In the near-UV, we calculated CD spectra of 40 proteins with their X-ray structures to assess the effectiveness of the vibrational parameter sets of aromatic amino acids. Comparison between the calculation and the experiment shows that qualitative and quantitative improvements have been achieved. Further improvements are obtained for nine out of 13 proteins calculated with NMR structures compared to results calculated from X-ray structures. This shows the importance of the conformation diversity of proteins in the CD calculation. Our calculations on four proteins show good predictions of the contribution of individual aromatic amino acids when studying the difference spectrum of wild-type protein minus the site-directed mutants when the conformation of the protein is not severely affected by the mutation. We further shows the potentially

ability of our calculations to distinguish the native protein structure from a large number of incorrect misfolded decoy structures when comparing to the experimental spectra. This new feature has been implemented in the DichroCalc web interface and is freely available to the community.

The success of considering the vibrational structures of chromophores of interest in the CD calculation confirms the effectiveness of our model and encourages us to further develop it. Some possible aspects are listed in the following section. The better the calculation can aid the interpretation of the experimental data, the more that CD spectroscopy will be used in protein studies. Also, the technical aspects of CD will be pushed forward continuously.

Future work

The parameterization of the matrix method in our CD calculations relies on the accuracy of the *ab initio* calculations of the chromophores. The calculation may be further improved by using a more accurate model describing the excitation process. Solvent effects are crucial in *ab initio* calculations. For example, a sequential Monte Carlo/QM calculation of NMA surrounded by 150 water molecules showed a 1755 cm^{-1} blue shifted $n\pi^*$ transition and a 1180 cm^{-1} red shifted $\pi\pi^*$ transition.[314] In this thesis, the

electronic ground and excited states are calculated in gas phase for all the small molecules representing the chromophore. It would be worthwhile to consider solvent effects further. The calculated vibrational frequencies need to be scaled to match with the experimental value depending on the theory used. The accurate description of the vibrational normal modes will affect the calculated FC integrals, and consequently influence the extended parameter sets in CD calculations. The frequencies will also affect the interval between the vibrational levels.

The CD signal can be sensitive to the environment surrounding the chromophore. The calculations of the mutants in chapter 4 showed our method can describe this feature very nicely. However, certain interactions are omitted in our current model. The solvent is also important when the water molecules comprise the environment of the protein. This means the properties of the chromophores will be different depending on its position, buried or exposed to solvent, in the protein. Also, the groups adjacent to the chromophore within a certain distance (usually 8 Å) will show strong electrostatic interactions which will affect the electronic structures of the chromophore. We developed a simple model considering the influence of the diagonal elements, namely the transition energies, arising from the surrounding atoms by computing the Coulomb interactions with point charges. However, no obvious improvement was achieved. A better model

should be considered to evaluate this effect. Hydrogen bonding is another interaction that should be considered. Different secondary structures of proteins are connected by hydrogen bonds between peptide bonds in different pattern. The aromatic side chains of tyrosine and tryptophan can also form hydrogen bonds with other residues or water molecules. It is quite common to have $\pi\pi$ stacking structures (parallel or T-shape) between two aromatic rings in proteins. All the above mentioned interactions will shift the CD band on a case by case basis. Considering these effects would help better assign the CD signals and further improve our understanding of the influence of protein structure.

The successful of calculation with NMR ensembles of structures encourages us to develop the work further by considering conformational dynamics. Protein structures are flexible in solution. Some less well-defined motifs may be better represented by an ensemble of structures instead of by a fixed X-ray structure, for example the β-II proteins. MD simulation may provide a better description of protein conformations in solution. Generation of a trajectory with implicit water models may also be used to take into account the influence of solvent effect on the electronic structure. We could perform MD simulations under different conditions such as different pH, various temperatures or bound/unbound states of protein with ligands. Currently, the CD spectroscopy is used to investigate if protein conformations

indeed change with varying conditions. CD calculation with the trajectory generated in the experimental condition may provide insight into the mechanisms of protein conformational change.

Generation of parameters of other chromophores is also of interest. The most relevant chromophore to this thesis, which has not considered, would be the disulfide bond forming between side chains of two cysteines. The disulfide bond is important for the structure and function of many proteins.[315, 316] Thus, interpretation the CD changes in the near-UV arising from formation or reduction of disulfide structures could provide useful information in understanding protein structure-function relationships. The transition of interest is from the lone pairs on sulfur to the σ* orbital the disulfide bond and the character of this transition depends on the dihedral angle of the C-S-S-C structure. This has been studied empirically and implemented in the calculation of optical properties.[317,318] Systematic *ab initio* calculations of the disulfide conformers with various dihedral angles would provide a more accurate description of this chromophore and is expected to further improve the CD calculation in the near-UV.

In our current CD calculation, we considered that the electrons are populated at the lowest vibrational level in the electronic ground state, namely no hot band is included. However, temperature effects may be

important in spectroscopic calculation. More signals are expected to be generated when the transitions start from higher vibrational levels of the electronic ground state to the excited state, which could further fill in the gaps in the current calculated CD envelope.

In this thesis, we considered the influence of the Franck-Condon effect in CD calculation. There are, no doubt, other effects that have correlation with the current model of CD calculation of proteins. For example, electronically forbidden transition can ‘borrow’ intensity from an allowed one which is known as Herzberg-Teller effect and such bands usually exhibit more fine structure than the electronically allowed transitions. A normal vibration is needed to connect these transitions.

There are other macromolecules that are of interest and can be studied by CD spectroscopy, such as DNA or polymers. Further developments on different systems not only contribute to our understanding of their own spectroscopic properties but also may provide supplementary information for other systems. Also, transitions at the other energy ranges, such as the visible and the VUV, are not included in this thesis; charge transfer parameters are available in DichroCalc, though. The interaction between these higher energy transitions or other chromophores and the currently studied chromophores is also relevant to CD calculation.

This thesis shows some successful results in improving the calculation of protein CD calculation by incorporation of vibrational structure of the chromophores and also indicates other possible aspects to further study CD spectroscopy by theoretical methods. A detailed CD calculation model would facilitate the interpretation of CD spectra and encourage the usage of CD in a wider range of studies.

References

1. Arago, F. J. D., Memoire sur une modification remarquable qu'eprouvent les rayons lumineux dans leur passage a travers certains corps diaphanes, et sur quelques autres nouveaux phenomenes 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.

2. Biot, J. B., Memoire sur un nouveau genre d'oscillation que les molecules de lumiere eprouvent 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.

3. Biot, J. B., Phenomenes de polarisation successive, observes dans des fluides homogenes. *Bull. Soc. Phil.* **1815**, 190.

4. Pasteur, L., Memoire sur la relation qui peut exister entre la forme cristalline et la composition chimique, et sur la cause de la polarisation 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.

5. Le Bel, J. A., 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.

6. van't Hoff, J. H., *La Chimie dans L'Espace*. Bazendijk: Rotterdam, 1875.

7. Morell, E. S.; Chico, L.; Brey, L., Twisting Dirac Fermions: Circular Dichroism in Bilayer Graphene. *2D Mater.* **2017,** *4*, 035015.

8. Sato, N.; Tatsumi, Y.; Saito, R., Circular Dichroism of Single-Wall Carbon Nanotubes. *Phys. Rev. B* **2017,** *95*, 155436.

9. Sprecher, C. A.; Baase, W. A.; Johnson, W. C. J., Conformation and Circular Dichroism of DNA. *Biopolymers* **1979,** *18*, 1009.

10. Baase, W. A.; Johnson, W. C. J., Circular Dichroism and DNA Secondary Structure. *Nucl. Acids Res.* **1979,** *6*, 797.

11. Girod, J. C.; Johnson, W. C. J.; Huntington, S. K.; Maestre, M. F., Conformation of Deoxyribonucleic Acid in Alcohol Solutions. *Biochemistry* **1973,** *12*, 5092.

12. Pohl, F. M.; Jovin, T. M., Salt-Induced Co-Operative Conformational

Change of a Synthetic DNA: Equilibrium and Kinetic Studies with Poly (dG-dC). *J. Mol. Biol.* **1972,** *67*, 375.

13. Johnson, K. H.; Gray, D. M.; Sutherland, J. C., Vacuum UV CD Spectra of Homopolymer Duplexes and Triplexes Containing A.T or A.U Base Pairs. *Nucl. Acids Res.* **1991,** *19*, 2275.

14. Howard, F. B.; Miles, H. T.; Ross, P., The Poly(dT)•2Poly(dA) Triple Helix. *Biochemistry* **1995,** *34*, 7135.

15. Gondeau, C.; Maurizot, J. C.; Durand, M., Spectroscopic Investigation of an Intramolecular DNA Triplex Containing Both G.G:C and T.A:T Triads and its Complex with Netropsin. *J. Biomol. Struct. Dyn.* **1998,** *15*, 1133.

16. Gondeau, C.; Maurizot, J. C.; Durand, M., Circular Dichroism and UV Melting Studies on Formation of an Intramolecular Triplex Containing Parallel T*A:T and G*G:C Triplets: Netropsin Complexation with the Triplex. *Nucl. Acids Res.* **1998,** *21*, 4996.

17. Clark, C. L.; Cecil, P. K.; Singh, D.; Gray, D. M., CD, Absorption and Thermodynamic Analysis of Repeating Dinucleotide DNA, RNA and Hybrid Duplexes [d/r(AC)]12.[d/r(GT/U)]12 and the Influence of Phosphorothioate Substitution. *Nucl. Acids Res.* **1997,** *25*, 4098.

18. Hashem, G. M.; Pham, L.; Vaughan, M. R.; Gray, D. M., Hybrid Oligomer Duplexes Formed with Phosphorothioate DNAs: CD Spectra and Melting Temperatures of S-DNA•RNA Hybrids are Sequence-Dependent but Consistent with Similar Heteronomous Conformations. *Biochemistry* **1998,** *37*, 61.

19. Verma, S.; Eckstein, F., Modified Oligonucleotides: Synthesis and Strategy for Users. *Annu. Rev. Biochem.* **1998,** *67*, 99.

20. Behe, M.; Felsenfeld, G., Effects of Methylation on a Synthetic Polynucleotide: the B-Z Transition in Poly(dG-m5dC).Poly(dG-m5dC). *Proc. Natl. Acad. Sci. (USA)* **1982,** *78*, 1619.

21. Lerman, L. S., Structural Considerations in the Interaction of DNA and Acridines. *J. Mol. Biol.* **1961,** *3*, 18.

22. Fornasiero, D.; Kurucsev, J., Circular Dichroism Spectra and the Interaction between Acridine Dyes and Deoxyribonucleic Acid. *J. Phys. Chem.* **1981,** *85*, 613.

23. Schipper, P. E.; Norden, B.; Tjerneld, F., Determination of Binding Geometry of DNA-Adduct Systems through Induced Circular Dichroism. *Chem. Phys. Lett.* **1980,** *70*, 17.

24. Norden, B.; Tjerneld, F., Structure of Methylene Blue–DNA Complexes Studied by Linear and Circular Dichroism Spectroscopy. *Biopolymers* **1982,** *21*, 1713.

25. Tuite, E.; Norden, B., Sequence-Specific Interactions of Methylene Blue with Polynucleotides and DNA: A Spectroscopic Study. *J. Am. Chem. Soc.* **1994,** *116*, 7548.

26. Kliger, D. S.; Lewis, J. W., *Fast Time-Resolved Circular Dichroism Measurements*. Wiley-VCH: United States, 2000.

27. Lewis, J. W.; Tilton, R. F.; Einterz, C. M.; Milder, S. J.; Kuntz, I. D.; Kliger, D. S., New Technique for Measuring Circular Dichroism Changes on a Nanosecond Time Scale. Application to (Carbonmonoxy)myoglobin and (Carbonmonoxy)hemoglobin. *J. Phys. Chem.* **1985,** *89*, 289.

28. Wen, Y.-X.; Chen, E.; Lewis, J. W.; Kliger, D. S., Nanosecond Time-Resolved Circular Dichroism Measurements using an Upconverted Ti:Sapphire Laser. *Rev. Sci. Inst.* **1996,** *67*, 3010.

29. Milder, S. J.; Bjorling, S. C.; Kuntz, I. D.; Kliger, D. S., Time-Resolved Circular Dichroism and Absorption Studies of the Photolysis Reaction of (Carbonmonoxy) myoglobin. *Biophys. J.* **1988,** *53*, 659.

30. Bjorling, S. C.; Goldbeck, R. A.; Paquette, S. J.; Milder, S. J.; Kliger, D. S., Allosteric Intermediates in Hemoglobin. 1. Nanosecond Time-Resolved Circular Dichroism Spectroscopy. *Biochemistry* **1996,** *35*, 8619.

31. Elove, G. A.; Chaffotte, A. F.; Roder, H.; Goldberg, M. E., Early Steps in Cytochrome C Folding Probed by Time-Resolved Circular Dichroism and Fluorescence Spectroscopy. *Biochemistry* **1992,** *31*, 6876.

32. Milder, S. J.; Weiss, P. S.; Kliger, D. S., Time-Resolved Absorption, Circular Dichroism, and Emission of tRNA. Evidence that the Photo-Cross-Linking of 4-Thiouridine in tRNA Occurs from the Triplet State. *Biochemistry* **1989,** *28*, 2258.

33. Dooley, D. M.; Dawson, J. H., Bioinorganic Applications of Magnetic Circular Dichroism Spectroscopy. Copper, Rare-earth Ions, Cobalt and Non-heme Iron Systems. *Coord. Chem. Rev.* **1984,** *60*, 1.

34. Johnson, M. K.; Robinson, A. E.; Thomson, A. J., *Low-Temperature Magnetic Circular Dichroism of Iron-Sulfur Proteins*. Wiley: New York, 1982; p 367.

35. Dawson, J. H.; Dooley, D. M., *Magnetic Circular Dichroism Spectroscopy of Iron Porphyrins and Heme Proteins*. VCH publishers: New York, 1989; p 1.

36. Cheesman, M. R.; Greenwood, C.; Thomson, A. J., *Magnetic Circular Dichroism of Hemoproteins*. in Advances in Inorganic Chemistry, Academic Press: San Diego, 1991; Vol. 36.

37. Solomon, E. I.; Pavel, E. G.; Loeb, K. E.; Campochiaro, C., Magnetic Circular Dichroism Spectroscopy as a Probe of the Geometric and Electronic

Structure of Non-heme Ferrous Enzymes. *Coord. Chem. Rev.* **1995,** *144*, 369.

38. Holzwarth, G.; Hsu, E. C.; Mosher, H. S.; Faulkner, T. R.; Mscowitz, A., Infrared Circular Dichroism of Carbon-Hydrogen and Carbon-Deuterium Stretching Modes. Observations. *J. Am. Chem. Soc.* **1974,** *96*, 251.

39. Nafie, L. A.; Cheng, J. C.; Stephens, P. J., Vibrational Circular Dichroism. *J. Am. Chem. Soc.* **1976,** *98*, 2715.

40. Bour, P.; Keiderling, T. A., Ab Initio Simulations of the Vibrational Circular Dichroism of Coupled Peptides. *J. Am. Chem. Soc.* **1993,** *115*, 9602.

41. Lal, B. B.; Diem, M.; Polavarapu, P. L.; Oboodi, M.; Freedman, T. B.; Nafie, L. A., Vibrational Circular Dichroism in Amino Acids and Peptides. 6. Localized Molecular Orbital Calculations of the Carbon-Hydrogen Stretching Vibrational Circular Dichroism in Deuterated Isotopomers of Alanine. *J. Am. Chem. Soc.* **1982,** *104*, 3336.

42. Freedman, T. B.; Chernovitz, A. C.; Zuk, W. M.; Paterlini, M. G.; Nafie, L. A., Vibrational Circular Dichroism in the Methine Bending Modes of Amino Acids and Dipeptides. *J. Am. Chem. Soc.* **1988,** *110*, 6970.

43. Han, W. G.; Jalkanen, K. J.; Elstner, M.; Suhai, S., Theoretical Study of Aqueous N-Acetyl-l-Alanine N'-Methylamide: Structures and Raman, VCD, and ROA Spectra. *J. Phys. Chem. B* **1998,** *102*, 2587.

44. Zhang, Y.; Rouxel, J. R.; Autschbach, J.; Govind, N.; Mukamel, S., X-ray Circular Dichroism Signals: A Unique Probe of Local Molecular Chirality. *Chem. Sci.* **2017,** *8*, 5969.

45. Štěpánek, P.; Coriani, S.; Sundholm, D.; Ovchinnikov, V. A.; Vaara, J., Relation between Molecular Electronic Structure and Nuclear Spin-Induced Circular Dichroism. *Sci. Rep.* **2017,** *7*, 46617.

46. Condon, E., A Theory of Intensity Distribution in Band Systems. *Phys. Rev.* **1926,** *28*, 1182.

47. Franck, J.; Dymond, E. G., Elementary Processes of Photochemical Reactions. *Trans. Faraday Soc.* **1926,** *21*, 536.

48. Neugebauer, J.; Baerends, E. J.; Nooijen, M.; Autschbach, J., Importance of Vibronic Effects on the Circular Dichroism Spectrum of Dimethyloxirane. *J. Chem. Phys.* **2005,** *122*, 234305.

49. Seidler, T.; Andrzejak, M.; Pawlikowski, M. T., The Circular Dichroism (CD) Studies of Bis-1,8-Naphthalimide. The Theoretical Analysis in terms of Density Functional and Vibronic Dimer Theories. *Chem. Phys. Lett.* **2010,** *496*, 74.

50. Bulheller, B. M.; Pantos, G. D.; Sanders, J. K. M.; Hirst, J. D., Electronic Structure and Circular Dichroism Spectroscopy of

Naphthalenediimide Nanotubes. *Phys. Chem. Chem. Phys.* **2009,** *11*, 6060.

51. Mikulás, T., *A Documentary History of Biochemistry,1770-1940*. Fairleigh Dickinson University Press: Rutherford, 1992.

52. Hofmeister, F., *Ergeb. Physiol. Biol. Chem. Exp. Pharmacol.* **1902,** *1*, 759.

53. Fischer, E., *Ber. dtsch. Chem. Ges.* **1902,** *35*, 1095.

54. Whitford, D., *Proteins: Structure and Function*. Wiley: 2005.

55. Pauling, L.; Corey, R. B.; Branson, H. R., The Structure of Proteins; Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *Proc. Natl. Acad. Sci. (USA)* **1951,** *37*, 205.

56. Kendrew, J. C.; Dickerson, R. E.; Strandberg, B. E.; Hart, R. G.; Davies, D. R.; Phillips, D. C.; Shore, V. C., Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å Resolution. *Nature* **1960,** *185*, 422.

57. Barlow, D. J.; Thornton, J. M., Helix Geometry in Proteins. *J. Mol. Biol.* **1988,** *201*, 601.

58. Pauling, L.; Corey, R. B., Configurations of Polypeptide Chains with Favored Orientations around Single Bonds.Two New Pleated Sheets. *Proc. Natl. Acad. Sci. (USA)* **1951,** *37*, 729.

59. Davydov, A. S., *Theory of Molecular Excitations*. Springer: Berlin, 1971.

60. Schellman, J. A., Symmetry Rules for Optical Rotation. *Accts. Chem. Res.* **1968,** *1*, 144.

61. Condon, E. U.; Altar, W.; Eyring, H., One-Electron Rotatory Power. *J. Chem. Phys.* **1937,** *5*, 753.

62. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000,** *28*, 235.

63. Klechner, I. R.; Foster, M. P., An Introduction to NMR-Based Approaches for Measuring Protein Dynamics. *Biochim. Biophys. Acta*. **2011,** *1814*, 942.

64. Rodger, A.; Norden, B., *Circular Dichroism and Linear Dichroism*. Oxford University Press: 1997.

65. Narvaez-Pita, X.; Ortega-Zuniga, C.; Acevedo-Morantes, C. Y.; Pastrana, B.; Olivero-Verbel, J.; Maldonado-Rojas, W.; Ramirez-Vick, J. E.; Melendez, E., Water Soluble Molybdenocene Complexes: Synthesis, Cytotoxic Activity and Binding Studies to Ubiquitin by Fluorescence Spectroscopy, Circular Dichroism and Molecular Modeling. *J. Inorg. Biochem.* **2014,** *132*, 77.

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

67. Sutherland, J. C.; Desmond, E. J.; Takacs, P. Z., Versatile Spectrometer for Experiments using Synchrotron Radiation at Wavelengths Greater than 100 nm. *Nuc. Inst. Meth.* **1980,** *172*, 195.

68. Safar, J.; Roller, P. P.; Ruben, G. C.; Gajdusek, D. C.; Gibbs, C. J. J., Secondary Structure of Proteins Associated in Thin Films. *Biopolymers* **1993,** *33*, 1461.

69. Bowman, R. L.; Kellerman, M.; Johnson, W. C. J., Optical Properties of Cyclic Dimers of Amino Acids: An Experimental and Theoretical Study. *Biopolymers* **1983,** *22*, 1045.

70. Moffitt, W., Optical Rotatory Dispersion of Helical Polymers. *J. Chem. Phys.* **1956,** *25*, 467.

71. Woody, R. W.; Tinoco, I., Optical Rotation of Oriented Helices. III. Calculation of Rotatory Dispersion and Circular Dichroism of Alpha- and 3_{10}-Helix. *J. Chem. Phys.* **1967,** *46*, 4927.

72. Schellman, J. A.; Oriel, P., Origin of the Cotton Effect of Helical Polypeptides. *J. Chem. Phys.* **1962,** *37*, 2114.

73. Greenfield, N. J.; Fasman, G. D., Computed Circular Dichroism Spectra for the Evaluation of Protein Conformation. *Biochemistry* **1969,** *8*, 4108.

74. Sreerama, N.; Woody, R. W., Structural composition of beta(I)- and beta(II)-proteins. *Protein Sci.* **2003,** *12*, 384.

75. Micsonai, A.; Wien, F.; Kernya, L.; Lee, Y. H.; Goto, Y.; Refregiers, M.; Kardos, J., Accurate Secondary Structure Prediction and Fold Recognition for Circular Dichroism Spectroscopy. *Proc. Natl. Acad. Sci. (USA)* **2015,** *112*, e3095.

76. Brahms, S.; Brahms, J.; Spach, G.; Brack, A., Identification of Beta, Beta-Turns and Unordered Conformations in Polypeptide Chains by Vacuum Ultraviolet Circular Dichroism. *Proc. Natl. Acad. Sci. (USA)* **1977,** *74*, 3208.

77. Chang, C. T.; Wu, C. S.; Yang, J. T., Circular Dichroic Analysis of Protein Conformation: Inclusion of the Beta-Turns. *Anal. Biochem.* **1978,** *91*, 13.

78. Compton, L. A.; Johnson, W. C. J., Analysis of Protein Circular Dichroism Spectra for Secondary Structure using a Simple Matrix Multiplication. *Anal. Biochem.* **1986,** *155*, 155.

79. Provencher, S. W.; Glockner, J., Estimation of Globular Protein Secondary Structure from Circular Dichroism. *Biochemistry* **1981,** *20*, 33.

80. Sreerama, N.; Woody, R. W., A Self-Consistent Method for the Analysis of Protein Secondary Structure from Circular Dichroism. *Anal. Biochem.* **1993,** *209*, 32.

81. Perczel, A.; Park, K.; Fasman, G. D., Analysis of the Circular Dichroism Spectrum of Proteins using the Convex Constraint Algorithm: a Practical Guide. *Anal. Biochem.* **1992,** *203*, 83.

82. Whitmore, L.; Wallace, B. A., Protein Secondary Structure Analyses from Circular Dichroism Spectroscopy: Methods and Reference Databases. *Biopolymers* **2008,** *89*, 392.

83. Balcerski, J. S.; Pysh, E. S.; Bonora, G. D.; Toniolo, C., Vacuum Ultraviolet Circular Dichroism of Beta-Forming Alkyl Oligopeptides. *J. Am. Chem. Soc.* **1976,** *98*, 3470.

84. Nardi, F.; Kemmink, J.; Sattler, M.; Wade, R. C., The Cisproline(i-1)-Aromatic(i) Interaction: Folding of the Ala-cisPro-Tyr Peptide Characterized by NMR and Theoretical Approaches. *J. Biomol. NMR* **2000,** *17*, 63.

85. Kim, P. S.; Baldwin, R. L., Intermediates in the Folding Reactions of Small Proteins. *Annu. Rev. Biochem.* **1990,** *59*, 631.

86. Kuwajima, K., *Circular Dichroism and the Conformational Analysis of Biomolecules*. Prenum Press: New York, 1996.

87. Burley, S. K.; Petsko, G. A., Aromatic-Aromatic Interaction-a Mechanism of Protein-Structure Stabilization. *Science* **1985,** *229*, 23.

88. Horwitz, J.; Strickland, E. H.; Billups, C., Analysis of Vibratinal Structure in Near-Ultraviolet Circular Dichroism and Absorption Spectra of Tyrosine Derivatives and Ribonuclease A at 77 Degrees K. *J. Am. Chem. Soc.* **1970,** *92*, 2119.

89. Strickland, E. H., Interactions Contributing to the Tyrosyl Circular Dichroism Bands of Ribonuclease S and A. *Biochemistry* **1972,** *11*, 3465.

90. Hatatanaka, A.; Kawai, G.; Yamasaki, K.; Ito, Y.; Kajiura, H.; Ha, J. M.; Miyazawa, T.; Yokoyama, S.; Nishimura, S., Sping-Labeling Proton NMR-Study on Aromatic Amino-Acid Residues in the Guanine-Nucleotide Binding Site of Human c-Ha-fas(1-171) Protein. *Biochemistry* **1989,** *28*, 9550.

91. Kopecny, D.; Tylichova, M.; Snegaroff, J.; Popelkova, H.; Sebela, M., Carboxylate and Aromatic Active-Site Residues are Determinants of High-Affinity Binding of Omega-Aminoaldehydes to Plant Aminoaldehyde Dehydrogenases. *FEBS J.* **2011,** *278*, 3130.

92. Freskgard, P. O.; Martensson, L. G.; Jonasson, P.; Jonsson, B. H.; Carlsson, U., Assignment of the Contribution of the Tryptophan Residues to the Circular Dichroism Spectrum of Human Carbonic Anhydrase II. *Biochemistry* **1994,** *33*, 14281.

93. Sreerama, N.; Manning, M. C.; Powers, M. E.; Zhang, J. X.; Goldenberg, D. P.; Woody, R. W., Tyrosine, Phenylalanine, and Disulfide Contributions to the Circular Dichroism of Proteins: Circular Dichroism Spectra of Wild-Type and Mutant Bovine Pancreatic Trypsin Inhibitor. *Biochemistry* **1999,** *38*, 10814.

94. Vuilleumier, S.; Sancho, J.; Loewenthal, R.; Fersht, A. R., Circular Dichroism Studies of Barnase and its Mutants-Characterization of the Contribution of Aromatic Side Chains. *Biochemistry* **1993,** *32*, 10303.

95. Woody, A. Y. M.; Woody, R. W., Individual Tyrosine Side-Chain Contributions to Circular Dichroism of Ribonuclease. *Biopolymers* **2003,** *72*, 500.

96. Greenfield, N. J., Enzyme Ligand Complexes: Spectroscopic Studies. *CRC Crit. Rev. Biochem.* **1975,** *3*, 71.

97. Honig, B.; Ebrey, T. G., The Structure and Spectra of the Chromophore of the Visual Pigments. *Ann. Rev. Biophys. Bioeng.* **1974,** *3*, 151.

98. Blauer, G.; Wagniere, G., Conformation of Bilirubin and Biliverdin in Their Complexes with Serum Albumin. *J. Am. Chem. Soc.* **1975,** *97*, 1949.

99. Wicken, J. S.; Woody, R. W., Circular Dichroism of Liver Alcohol Dehydrogenase Complexes with Auramine O. *Biochemistry* **1973,** *12*, 3459.

100. Chang, J. F.; Phillips, K.; Lundback, T.; Gstaiger, M.; Ladbury, J. E.; Luisi, B., Oct-1 POU and Octamer DNA Co-operate to Recognise the Bob-1 Transcription Co-activator via Induced Folding. *J. Mol. Biol.* **1999,** *288*, 941.

101. Clarke, D. T.; Jones, G. R., Extended Circular Dichroism Measurements using Synchrotron Radiation Show That the Assembly of Clathrin Coats Requires No Change in Secondary Structure. *Biochemistry* **1999,** *38*, 10457.

102. Clarke, D. T.; Doig, A. J.; Stapeley, B. J.; Jones, G. R., The α-Helix Folds on the Millisecond Time Scale. *Proc. Natl. Acad. Sci. (USA)* **1999,** *96*, 7232.

103. Rosenfeld, L., Quantenmechanische Theorie der naturlichen optischen Aktivitat von Flussigkeiten und Gasen [Quantum Mechanical Theory of the Optical Activity of Liquids and Gases]. *Z. Phys.* **1928,** *52*, 161.

104. Cantor, C. R.; Schimmel, P. R., *Biophysical Chemistry*. W. H. Freeman: San Francisco, 1980.

105. Salvadori, P.; Rosini, C.; Di Bari, L., *Conformation and Chiroptical Properties of Dienes and Polyenes*. John Wiley: Chichester, 1997.

106. Rosini, C.; Scamuzzi, S.; Focati, M. P.; Salvadori, P., A General, Multitechnique Approach to the Stereochemical Characterization of 1,2-

Diarylethane-1,2-diols. *J. Org. Chem.* **1995,** *60*, 8289.

107. DeVoe, H., Optical Properties of Molecular Aggregates. I. Classical Model of Electronic Absorption and Refraction. *J. Chem. Phys.* **1964,** *41*, 393.

108. DeVoe, H., Optical Properties of Molecular Aggregates. 2. Classical Theory of Refraction Absorption and Optical Activity of Solutions and Crystals. *J. Chem. Phys.* **1965,** *43*, 3199.

109. Applequist, J., Full Polarizability Treatment of the Pi-Pi-Star Absorption and Circular Dichroic Spectra of Alpha-Helical Polypeptides. *J. Chem. Phys.* **1979,** *71*, 4332.

110. Applequist, J.; Sundberg, K. R.; Olson, M. L.; Weiss, L. C., Normal Mode Treatment of Optical Properties of A Classical Coupled Dipole Oscillator System with Lorentzian Band Shapes. *J. Chem. Phys.* **1979,** *70*, 1240.

111. Bode, K. A.; Applequist, J., 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.

112. Tinoco, I. J., Theoretical Aspects of Optical Activity. *Adv. Chem. Phys.* **1962,** *4*, 113.

113. Grimme, S.; Bannwarth, C., Ultra-fast Computation of Electronic Spectra for Large Systems by Tight-Binding Based Simplified Tamm-Dancoff Approximation (sTDA-xTB). *J. Chem. Phys.* **2016,** *145*, 054103.

114. Seibert, J.; Bannwarth, C.; Grimme, S., Biomolecular Structure Information from High-Speed Quantum Mechanical Electronic Spectra Calculation. *J. Am. Chem. Soc.* **2017,** *139*, 11682.

115. Hallberg, A.; Isaksson, R.; Martin, A. R.; Sandstrom, J., Chromatographic Resolution, Circular Dichroism Spectra, and Absolute Configurations of Dimers of 5H-Indolo[1,7-ab][1]benzazepine and Coumarin with C2 Symmetry. *J. Am. Chem. Soc.* **1989,** *111*, 4387.

116. Fleischhauer, J.; Koslowski, A.; Schiffer, J.; Wollmer, A.; Enders, D.; Zamponi, A.; Raabe, G., CD-Spectroscopic Confirmation of the Absolute Configuration of a Conformationally Flexible Aldehyde Ester Containing a Quarternary Stereogenic Center. *Z. Naturforsch.* **1998,** *53a*, 891.

117. Moffitt, W., The Optical Rotatory Dispersion of Simple Polypeptides. II. *Proc. Natl. Acad. Sci. (USA)* **1956,** *42*, 736.

118. Madison, V.; Schellman, J. A., Optical Activity of Polypeptides and Proteins. *Biopolymers* **1972,** *11*, 1041.

119. Besley, N. A.; Brienne, M. J.; Hirst, J. D., Electronic Structure of a Rigid Cyclic Diamide. *J. Phys. Chem. B* **2000,** *104*, 12371.

120. Fleischhauer, J.; Grotzinger, J.; Kramer, B.; Kruger, P.; Wollmer, A.; Woody, R. W.; Zobel, E., Calculation of the Circular Dichroism Spectrum of Cyclo(L-Tyr-L-Tyr) Based on a Molecular Dynamics Simulation. *Biophys. Chem.* **1994,** *49*, 141.

121. Li, Z.; Mukamel, S., First-Principles Simulation of Amide and Aromatic Side Chain Ultraviolet Spectroscopy of a Cyclic Dipeptide. *J. Phys. Chem. A* **2007,** *111*, 11579.

122. Manning, M. C.; Illangasekare, M.; Woody, R. W., Circular Dichroism Studies of Distorted Alpha-Helices, Twisted Beta-Sheets, and Beta Turns. *Biophys. Chem.* **1988,** *31*, 77.

123. Kabsch, W.; Sander, C., Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983,** *22*, 2577.

124. Wilmot, C. M.; Thornton, J. M., Beta-Turns and Their Distortions: A Proposed New Nomenclature. *Protein Eng.* **1990,** *3*, 479.

125. Smith, J. A.; Pease, L. G., Reverse Turns in Peptides and Proteins. *CRC Crit. Rev. Biochem.* **1980,** *8*, 315.

126. Manning, M. C.; Woody, R. W., Theoretical CD Studies of Polypeptide Helices: Examination of Important Electronic and Geometric Factors. *Biopolymers* **1991,** *31*, 569.

127. Sugita, Y.; Nagai, M.; Yoneyama, Y., Circular Dichroism of Hemoglobin in Relation to the Structure Surrounding the Heme. *J. Biol. Chem.* **1971,** *246*, 383.

128. Hsu, M.-C.; Woody, R. W., The Origin of the Heme Cotton Effects in Myoglobin and Hemoglobin. *J. Am. Chem. Soc.* **1971,** *93*, 3515.

129. Woody, R. W., *Protein-Ligand Interactions*. Gruyter: Berlin, 1975.

130. Pescitelli, G.; Sreerama, N.; Salvadori, P.; Nakanishi, K.; Berova, N.; Woody, R. W., Inherent Chirality Dominates the Visible/Near-Ultraviolet CD Spectrum of Rhodopsin. *J. Am. Chem. Soc.* **2008,** *130*, 6170.

131. Hirst, J. D.; Hirst, D. M.; Brooks, C. L., Ab Initio Calculations of the Excited States of Formamide. *J. Phys. Chem.* **1996,** *100*, 13487.

132. Hirst, J. D.; Persson, B. J., Ab Initio Calculations of the Vibrational and Electronic Spectra of Diketopiperazine. *J. Phys. Chem. A* **1998,** *102*, 7519.

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

134. Besley, N. A.; Hirst, J. D., Ab Initio Study of the Effect of Solvation on the Electronic Spectra of Formamide and N-methylacetamide. *J. Phys. Chem.*

A **1998,** *102*, 10791.

135. Watson, T. M.; Hirst, J. D., Calculating Vibrational Frequencies of Amides: From Formamide to Concanavalin A. *Phys. Chem. Chem. Phys.* **2004,** *6*, 998.

136. Oakley, M. T.; Bulheller, B. M.; Hirst, J. D., First-Principles Calculations of Protein Circular Dichroism in the Far-Ultraviolet and beyond. *Chirality* **2006,** *18*, 340.

137. Oakley, M. T.; Hirst, J. D., Charge-Transfer Transitions in Protein Circular Dichroism Calculations. *J. Am. Chem. Soc.* **2006,** *128*, 12414.

138. Li, Z.; Robinson, D.; Hirst, J. D., Vibronic Structure in the Far-UV Electronic Circular Dichroism Spectra of Proteins. *Faraday Discuss.* **2015,** *177*, 329.

139. Li, Z.; Hirst, J. D., Quantitative First Principles Calculations of Protein Circular Dichroism in the Near-Ultraviolet. *Chem. Sci.* **2017,** *8*, 4318.

140. Besley, N. A.; Hirst, J. D., Hydrogen Bonding in Protein Circular Dichroism Calculations. *J. Mol. Struct.: THEOCHEM* **2000,** *506*, 161.

141. Bhattacharjee, S.; Toth, G.; Lovas, S.; Hirst, J. D., Influence of Tyrosine on the Electronic Circular Dichroism of Helical Peptides. *J. Phys. Chem. B* **2003,** *107*, 8682.

142. Jiang, J.; Abramavicius, D.; Bulheller, B. M.; Hirst, J. D.; Mukamel, S., Ultraviolet Spectroscopy of Protein Backbone Transitions in Aqueous Solution: Combined QM and MM Simulations. *J. Phys. Chem. B* **2010,** *114*, 8270.

143. Bulheller, B. M.; Hirst, J. D., DichroCalc-Circular and Linear Dichroism Online. *Bioinformatics* **2009,** *25*, 539.

144. Besley, N. A.; Hirst, J. D., Theoretical Studies toward Quantitative Protein Circular Dichroism Calculations. *J. Am. Chem. Soc.* **1999,** *121*, 9636.

145. Rogers, D. M.; Hirst, J. D., First-Principles Calculations of Protein Circular Dichroism in the Near Ultraviolet. *Biochemistry* **2004,** *43*, 11092.

146. Bulheller, B. M.; Miles, A. J.; Wallace, B. A.; Hirst, J. D., Charge-Transfer Transitions in the Vacuum-Ultraviolet of Protein Circular Dichroism Spectra. *J. Phys. Chem. B* **2008,** *112*, 1866.

147. Louis-Jeune, C.; Andrade-Navarro, M. A.; Perez-Iratxeta, C., Prediction of Protein Secondary Structure from Circular Dichroism using Theoretically Derived Spectra. *Proteins: Struct., Funct., Bioinf.* **2012,** *80*, 374.

148. Hamley, I. W.; Nutt, D. R.; Brown, G. D.; Miravet, J. F.; Escuder, B.; Rodriguez-Llansola, F., Influence of the Solvent on the Self-Assembly of a Modified Amyloid Beta Peptide Fragment. II. NMR and Computer Simulation

Investigation. *J. Phys. Chem. B* **2010,** *114*, 940.

149. Nguyen, T. H.; Fabio, A.; Simone, S.; Giulia, R.; Emiliano, I.; Paolo, C.; Giovanni, N., Structural Determinants of Cisplatin and Transplatin Binding to the Met-Rich Motif of Ctr1: A Computational Spectroscopy Approach. *J. Chem. Theory Comput.* **2012,** *8*, 2912.

150. Dixon, A. S.; Pendley, S. S.; Bruno, B. J.; Woessner, D. W.; Shimpi, A. A.; Cheatham, T. E., III; Lim, C. S., Disruption of Bcr-Abl Coiled Coil Oligomerization by Design. *J. Biol. Chem.* **2011,** *286*, 27751.

151. Majava, V.; Polverini, E.; Mazzini, A.; Nanekar, R.; Knoll, W.; Peters, J.; Natali, F.; Baumgaertel, P.; Kursula, I.; Kursula, P., Structural and Functional Characterization of Human Peripheral Nervous System Myelin Protein P2. *Plos One* **2010,** *5*, e10300.

152. Meissner, R. H.; Schneider, J.; Schiffels, P.; Ciacchi, L. C., Computational Prediction of Circular Dichroism Spectra and Quantification of Helicity Loss upon Peptide Adsorption on Silica. *Langmuir* **2014,** *30*, 3487.

153. McCann, D. M.; Stephens, P. J., Determination of Absolute Configuration using Density Functional Theory Calculations of Optical Rotation and Electronic Circular Dichroism: Chiral Alkenes. *J. Org. Chem.* **2006,** *71*, 6074.

154. Mori, T.; Inoue, Y.; Grimme, S., Quantum Chemical Study on the Circular Dichroism Spectra and Specific Rotation of Donor-Acceptor Cyclophanes. *J. Phys. Chem. A* **2007,** *111*, 7995.

155. Bak, K. L.; Hansen, A. E.; Ruud, K.; Helgaker, T.; Olsen, J.; Jorgensen, P., Ab Initio Calculation of Electronic Circular Dichroism for Trans-Cyclooctene using London Atomic Orbitals. *Theor. Chim. Acta* **1995,** *90*, 441.

156. Pedersen, T. B.; Koch, H.; Ruud, K., Coupled Cluster Response Calculation of Natural Chiroptical Spectra. *J. Chem. Phys.* **1999,** *110*, 2883.

157. Gaigeot, M. P.; Besley, N. A.; Hirst, J. D., Modeling the Infrared and Circular Dichroism Spectroscopy of a Bridged Cyclic Diamide. *J. Phys. Chem. B* **2011,** *115*, 5526.

158. Bode, K. A.; Applequist, J., 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.

159. Bayley, P. M.; Nielsen, E. B.; Schellman, J. A., Rotatory Properties of Molecules Containing 2 Peptide Groups-Theory. *J. Phys. Chem.* **1969,** *73*, 228.

160. Woody, R. W., Improved Calculation of NPi' Rotational Strength in Polypeptides. *J. Chem. Phys.* **1968,** *49*, 4797.

161. Woody, R. W.; Sreerama, N., 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.

162. Ottiger, P.; Köppel, H.; Leutwyler, S., Excitonic Splittings in Molecular Dimers: Why Static Ab Initio Calculations Cannot Match Them. *Chem. Sci.* **2015,** *6*, 6059.

163. Goldmann, E.; Asher, S. A.; Mukamel, S., Electronic Excitations of Polyalanine; Test of the Independent Chromophore Approximation. *Phys. Chem. Chem. Phys.* **2001,** *3*, 2893.

164. Volkhard, M.; Oliver, K., *Charge and Energy Transfer Dynamics in Molecular Systems*. Wiley-VCH: Weinheim, 2011; p 581.

165. Yang, M., Influence of Energy Transfer on the Intensity Pattern of Vibronic Excitation Studied by Reduced Density-Matrix Theory. *J. Mol. Spectrosc.* **2006,** *239*, 108.

166. Bulheller, B. M.; Rodger, A.; Hirst, J. D., Circular and Linear Dichroism of Proteins. *Phys. Chem. Chem. Phys.* **2007,** *9*, 2020.

167. Zerner, M. C., *Semiempirical Molecular Orbital Methods*. in Reviews in Computational Chemistry, Wiley-VCH: 1991; Vol. 2.

168. Born, M.; Oppenheimer, R., Zur Quantentheorie der Molekeln (On the Quantum Theory of Molecules). *Ann. Phys.* **1927,** *389*, 457.

169. Sherrill, C. D.; Schaefer, H. F., *The Configuration Interaction Method: Advances in Highly Correlated Approaches*. in Advances in Quantum Chemistry, Academic Press: 1999; Vol. 34.

170. Cremer, D., From Configuration Interaction to Coupled Cluster Theory: The Quadratic Configuration Interaction Approach. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013,** *3*, 482.

171. van Lenthe, J. H.; Pulay, P., A Space-Saving Modification of Davidson's Eigenvector Algorithm. *J. Comput. Chem.* **1990,** *11,* 1164.

172. Roos, B. O.; Siegbahn, P. E. M., *Methods of Electronic Structure Theory*. Schaefer H. F., III, Ed., Plenum: New York, 1977.

173. Saebo, S.; Pulay, P., Local Treatment of Electron Correlation. *Annu. Rev. Phys. Chem.* **1993,** *44*, 213.

174. Shepard, R., The Multiconfiguration Self-Consistent Field Method. *Adv. Chem. Phys.* **1987,** *69*, 63.

175. Szalay, P. G.; Mueller, T.; Gidofalvi, G.; Lischka, H.; Shepard, R., Multiconfiguration Self-Consistent Field and Multireference Configuration Interaction Methods and Applications. *Chem. Rev.* **2012,** *112*, 108.

176. Roos, B. O., *Complete Active Space Self-Consistent Field Method and its Applications in Electronic Structure Calculations*. in Ab Initio Methods In Quantum Chemistry Part II, John Wiley & Sons, Inc.: New York, 1987; Vol. 69.

177. Roos, B. O.; Andersson, K., Multiconfigurational Perturbation-Theory with Level Shift-the Cr_2 Potential Revisited. *Chem. Phys. Lett.* **1995,** *245*, 215.

178. Werner, H. J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schuetz, M., Molpro: A General-Purpose Quantum Chemistry Program Package. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012,** *2*, 242.

179. Werner, H. J.; Knowles, P. J., A Second Order MCSCF Method with Optimum Convergence. *J. Chem. Phys.* **1985,** *82*, 5053.

180. Knowles, P. J.; Werner, H. J., An Efficient Second Order MCSCF Method for Long Configuration Expansions. *Chem. Phys. Lett.* **1985,** *115*, 259.

181. Werner, H. J.; Meyer, W., A Quadratically Convergent MCSCF Method for the Simultaneous Optimization of Several States. *J. Chem. Phys.* **1981,** *74*, 5794.

182. Knowles, P. J.; Handy, N. C., A New Determinant-Based Full Configuration Interaction Method. *Chem. Phys. Lett.* **1984,** *111*, 315.

183. Parr, R. G.; Yang, W., *Density-Functional Theory of Atoms and Molecules*. Oxford Univ. Press: New York, 1989.

184. Hohenberg, P.; Kohn, W., Inhomogeneous Electron Gas. *Phys. Rev. B* **1964,** *136*, B864.

185. Bauernschmitt , R.; Ahlrichs , R., Treatment of Electronic Excitations within the Adiabatic Approximation of Time Dependent Density Functional Theory. *Chem. Phys. Lett.* **1996,** *256*, 454.

186. Runge, E.; Gross, E. K. U., Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984,** *52*, 997.

187. Tanaka, T.; Kodama, T. S.; Morita, H. E.; Ohno, T., Electronic and Vibrational Circular Dichroism of Aromatic Amino Acids by Density Functional Theory. *Chirality* **2006,** *18*, 652.

188. Hehre, W. J.; Ditchfie, R.; Pople, J. A., Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972,** *56*, 2257.

189. Hill, J. G., Gaussian Basis Sets for Molecular Applications. *Int. J. Quantum Chem.* **2013,** *113*, 21.

190. Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M.

S.; Defrees, D. J.; Pople, J. A., Self-Consistent Molecular Orbital Methods. XXIII. A Polarization-Type Basis Set for Second-Row Elements. *J. Chem. Phys.* **1982,** *77*, 3654.

191. Serrano-Andrés, L.; Fülscher, M. P., Theoretical Study of the Electronic Spectroscopy of Peptides. III. Charge-Transfer Transitions in Polypeptides. *J. Am. Chem. Soc.* **1998,** *120*, 10912.

192. Walden, S. E.; Wheeler, R. A., Structural and Vibrational Analysis of Indole by Density Functional and Hybrid Hartree-Fock/Density Functional Methods. *J. Chem. Soc., Perkin Trans. 2* **1996**, 2653.

193. Slater, L. S.; Callis, P. R., Molecular Orbital Theory of the (1)L(A) and (1)L(B) States of Indole. 2. an Ab Initio Study. *J. Phys. Chem.* **1995,** *99*, 8572.

194. Liu, D.; Fang, W. H.; Fu, X. Y., Ab Initio Molecular Orbital Study of the Mechanism of Photodissociation of Formamide. *Chem. Phys. Lett.* **2000,** *318*, 291.

195. Kroemer, R. T.; Liedl, K. R.; Dickinson, J. A.; Robertson, E. G.; Simons, J. P.; Borst, D. R.; Pratt, D. W., Conformationally Induced Changes in the Electronic Structures of Some Flexible Benzenes. A Molecular Orbital Model. *J. Am. Chem. Soc.* **1998,** *120*, 12573.

196. Mozhayskiy, V. A.; Krylov, A. I., *ezSpectrum*. http://iopenshell.usc.edu/downloads.

197. Lopes, J. L. S.; Orcia, D.; Araujo, A. P. U.; DeMarco, R.; Wallace, B. A., Folding Factors and Partners for the Intrinsically Disordered Protein Micro-Exon Gene 14 (MEG-14). *Biophys. J.* **2013,** *104*, 2512.

198. Vanhove, M.; Lejeune, A.; Pain, R. H., Beta-Lactamases as Models for Protein-Folding Studies. *Cell Mol. Life Sci.* **1998,** *54*, 372.

199. Miles, A. J.; Fedosova, N. U.; Hoffmann, S. V.; Wallace, B. A.; Esmann, M., Stabilisation of Na,K-ATPase Structure by the Cardiotonic Steroid Ouabain. *Biochem. Biophys. Res. Commun.* **2013,** *435*, 300.

200. Hirst, J. D.; Hirst, D. M.; Brooks, C. L., Multireference Configuration Interaction Calculations of Electronic States of N-methylformamide, Acetamide, and N-methylacetamide. *J. Phys. Chem. A* **1997,** *101*, 4821.

201. Besley, N. A.; Oakley, M. T.; Cowan, A. J.; Hirst, J. D., A Sequential Molecular Mechanics/Quantum Mechanics Study of the Electronic Spectre of Amides. *J. Am. Chem. Soc.* **2004,** *126*, 13502.

202. Dang, Z. J.; Hirst, J. D., Short Hydrogen Bonds, Circular Dichroism, and Over-Estimates of Peptide Helicity. *Angew. Chem. Int. Ed.* **2001,** *40*, 3619.

203. Sanematu, T.; Mizuno, Y., Vibronic Exciton Approach to NV1

Absorption Bands of Polypeptides. 1. Alpha-Helix. *J. Phys. Soc. Jpn.* **1976,** *40*, 1733.

204. Sanematu, T., Vibronic Exciton Approach to NV1 Absorption Bands of Polypeptides. 2. Beta-Form. *J. Phys. Soc. Jpn.* **1977,** *43*, 600.

205. Mayne, L. C.; Hudson, B., Resonance Raman Spectroscopy of N-Methylacetamide Overtones and Combinations of the C-N Stretch (Amide-II') and Effect of Solvation on the C=O Stretch (Amide-I) Intensity. *J. Phys. Chem.* **1991,** *95*, 2962.

206. Weiss, M. S.; Metzner, H. J.; Hilgenfeld, R., Two Non-Proline Cis Peptide Bonds may be Important for Factor XIII Function. *FEBS Lett.* **1998,** *423*, 291.

207. Mirkin, N. G.; Krimm, S., Ab Initio Vibrational Analysis of Hydrogen Bonded Trans-N-Methylacetamide and Cis-N-Methylacetamide. *J. Am. Chem. Soc.* **1991,** *113*, 9742.

208. Whitmore, L.; Woollett, B.; Miles, A. J.; Klose, D. P.; Janes, R. W.; Wallace, B. A., PCDDB: the Protein Circular Dichroism Data Bank, A Repository for Circular Dichroism Spectral and Metadata. *Nucleic Acids. Res.* **2011,** *39*, D480.

209. Robin, M. B., *Higher Excited States of Polyatomic Molecules*. New York, 1985; Vol. 3.

210. Hagler, A. T.; Leiserowitz, L.; Tuval, M., Experimental and Theoretical Studies of Barrier to Rotation about N-C-Alpha and C-Alpha-C' Bonds (Phi and Psi) in Amides and Peptides. *J. Am. Chem. Soc.* **1976,** *98*, 4600.

211. Itoh, K.; Shimanou, T., Far-Infrared Spectra of N-Methylacetamide and Related Compounds and Hydrogen Bond Force Constants. *Biopolymers* **1967,** *5*, 921.

212. Fillaux, F.; Deloze, C., Structure of Matrix Isolated N-Methylacetamide. *Chem. Phys. Lett.* **1976,** *39*, 547.

213. Li, Y.; Garrell, R. L.; Houk, K. N., Mechanism of Cis-Trans Isomerizations of Amide and Peptide Excited States. *J. Am. Chem. Soc.* **1991,** *113*, 5895.

214. Chen, X. G.; Schweitzer-Stenner, R.; Krimm, S.; Mirkin, N. G.; Asher, S. A., N-Methylacetamide and its Hydrogen Bonded Water Molecules are Vibrationally Coupled. *J. Am. Chem. Soc.* **1994,** *116*, 11141.

215. Chen, X. G.; Asher, S. A.; Schweitzer-Stenner, R.; Mirkin, N. G.; Krimm, S., UV Raman Determination of the Pi-Pi* Excited State Geometry of N-Methylacetamide Vibrational Enhancement Pattern. *J. Am. Chem. Soc.* **1995,** *117*, 2884.

216. Katz, J. L.; Post, B., The Crystal Structure and Polymorphism of N-

Methylacetamide. *Acta Crystallogr.* **1960,** *13*, 624.

217. Kitano, M.; Fukuyama, T.; Kuchitsu, K., Molecular Structure of N-Methylacetamide as Studied by Gas Electron Diffraction. *Bull. Chem. Soc. Jpn.* **1973,** *46*, 384.

218. Mayne, L. C.; Ziegler, L. D.; Hudson, B., Ultraviolet Resonance Raman Studies of N-Methylacetamide. *J. Phys. Chem.* **1985,** *89*, 3395.

219. Pople, J. A.; Scott, A. P.; Wong, M. W.; Radom, L., Scaling Factors for Obtaining Fundamental Vibrational Frequencies and Zero-Point Energies form HF/6-31G* and MP2/6-31G* Harmonic Frequencies. *Isr. J. Chem.* **1993,** *33*, 345.

220. Kaya, K.; Nagakura, S., Electronic Absorption Spectra of Hydrogen Bonded Amides. *Theor. Chim. Acta* **1967,** *7*, 124.

221. Hirst, J. D.; Bhattacharjee, S.; Onufriev, A. V., Theoretical Studies of Time-Resolved Spectroscopy of Protein Folding. *Faraday Discuss.* **2003,** *122*, 253.

222. Kelly, S. M.; Jess, T. J.; Price, N. C., How to Study Proteins by Circular Dichroism. *Biochim. Biophys. Acta, Proteins Proteomics* **2005,** *1751*, 119.

223. Li, J.; Du, X.; Hashim, S.; Shy, A.; Xu, B., Aromatic–Aromatic Interactions Enable α-Helix to β-Sheet Transition of Peptides to Form Supramolecular Hydrogels. *J. Am. Chem. Soc.* **2017,** *139*, 71.

224. Ptitsyn, O. B., Molten Globule and Protein Folding. *Adv. Protein Chem.* **1995,** *47*, 83.

225. Findlay, W. A.; Martin, S. R.; Beckingham, K.; Bayley, P. M., Recovery of Native Structure by Calcium Binding Site Mutants of Calmodulin upon Binding of sk-MLCK Target Peptides. *Biochemistry* **1995,** *34*, 2087.

226. Pflumm, M. N.; Beychok, S., Optical Activity of Cystine-Containing Proteins. 2. Circular Dichroism Spectra of Pancreatic Ribonuclease A, Ribonuclease S, and Ribonuclease S-Protein. *J. Biol. Chem.* **1969,** *244*, 3973.

227. Buck, M.; Radford, S. E.; Dobson, C. M., A Partially Folded State of Hen Egg White Lysozyme in Trifluoroethanol-Structural Characterization and Implications for Protein Folding. *Biochemistry* **1993,** *32*, 669.

228. Nielsen, L.; Khurana, R.; Coats, A.; Frokjaer, S.; Brange, J.; Vyas, S.; Uversky, V. N.; Fink, A. L., Effect of Environmental Factors on the Kinetics of Insulin Fibril Formation: Elucidation of the Molecular Mechanism. *Biochemistry* **2001,** *40*, 6036.

229. Mendonça, L.; Hache, F.; Changenet-Barret, P.; Plaza, P.; Chosrowjan, H.; Taniguchi, S.; Imamoto, Y., Ultrafast Carbonyl Motion of the Photoactive Yellow Protein Chromophore Probed by Femtosecond Circular Dichroism. *J. Am. Chem. Soc.* **2013,** *135*, 14637.

230. Horwitz, J.; Strickland, E. H.; Billups, C., Analysis of Vibrational Structure in Near-Ultraviolet Circular Dichroism and Absorption Spectra of Phenylalanine and its Derivatives. *J. Am. Chem. Soc.* **1969,** *91*, 184.

231. Strickland, E. H.; Wilchek, M.; Billups, C.; Horwitz, J., Effects of Hydrogen-Bonding and Temperature upon Near Ultraviolet Circular Dichroism and Absorption Spectra of Tyrosine and O-Methyl Tyrosine Derivatives. *J. Biol. Chem.* **1972,** *247*, 572.

232. Strickland, E. H.; Horwitz, J.; Billups, C., Fine Structure in the Near UV Circular Dichroism and Absorption Spectra of Tryptophan Derivatives and Chymotrypsinogen A at 77 Kelvin. *Biochemistry* **1969,** *8*, 3205.

233. Strickland, E. H.; Billups, C., Oscillator Strengths of 1La and 1Lb Absorption Bands of Tryptophan and Several Other Indoles. *Biopolymers* **1973,** *12*, 1989.

234. Yutani, K.; Ogasahara, K.; Suzuki, M.; Sugino, Y., Effect of a Single Amino Acid Substitution on the Near Ultraviolet CD Spectra of Tryptophan Synthase Alpha-Subunit. *J. Biochem.* **1980,** *87*, 117.

235. Eder, J.; Rheinnecker, M.; Fersht, A. R., Folding of Subtilisin BPN': Characterization of a Folding Intermediate. *Biochemistry* **1993,** *32*, 18.

236. Barth, A.; Martin, S. R.; Bayley, P. M., Resolution of Trp Near UV CD Spectra of Calmodulin-Domain Peptide Complexes into the L-1(a) and L-1(b) Component Spectra. *Biopolymers* **1998,** *45*, 493.

237. McCann, J.; Rauk, A.; Shustov, G. V.; Wieser, H.; Yang, D. Y., Electronic and Vibrational Circular Dichroism of Model β-Lactams: 3-Methyl- and 4-Methylazetidin-2-One. *Appl. Spectrosc.* **1996,** *50*, 630.

238. Andrejeva, A.; Gardner, A. M.; Tuttle, W. D.; Wright, T. G., Consistent Assignment of the Vibrations of Symmetric and Asymmetric para-Disubstituted Benzene Molecules. *J. Mol. Spectrosc.* **2016,** *321*, 28.

239. Asher, S. A.; Ludwig, M.; Johnson, C. R., UV Resonance Raman Excitation Profiles of the Aromatic Amino Acids. *J. Am. Chem. Soc.* **1986,** *108*, 3186.

240. Carney, J. R.; Zwier, T. S., Infrared and Ultraviolet Spectroscopy of Water-Containing Clusters of Indole, 1-Methylindole, and 3-Methylindole. *J. Phys. Chem. A* **1999,** *103*, 9943.

241. Myszkiewicz, G.; Meerts, W. L.; Ratzer, C.; Schmitt, M., The Structure of 4-Methylphenol and its Water Cluster Revealed by Rotationally Resolved UV Spectroscopy using a Genetic Algorithm Approach. *J. Chem. Phys.* **2005,** *123*, 044304.

242. Lau, C. L.; Snyder, R. G., Valence Force Field for Alkyl Benzenes Toluene, p-Xylene, m-Xylene Mesitylene, and Some of Their Deuterated

Analogues. *Spectrochim. Acta Part A* **1971,** *27*, 2073.

243. Gardner, A. M.; Green, A. M.; Tame-Reyes, V. M.; Reid, K. L.; Davies, H. A.; Parkes, V. H. K.; Wright, T. G., The 700-1500 cm^{-1} Region of the S_1 ($\tilde{A}^1B_2$) State of Toluene Studied with Resonance-Enhanced Multiphoton Ionization (REMPI), Zero-Kinetic-Energy (ZEKE) Spectroscopy, and Time-Resolved Slow-Electron Velocity-Map Imaging (tr-SEVI) Spectroscopy. *J. Chem. Phys.* **2014,** *140*, 114308.

244. Siamwiza, M. N.; Lord, R. C.; Chen, M. C.; Takamatsu, T.; Harada, I.; Matsuura, H.; Shimanouchi, T., Interpretation of Doublet at 850 and 830 cm^{-1} in Raman SPectra of Tyrosyl Residues in Proteins and Certain Model Compounds. *Biochemistry* **1975,** *14*, 4870.

245. Seip, R.; Schultz, G.; Hargittai, I.; Szabo, Z. G., Electron Diffraction Study of Toluene Molecular Geometry. *Z. Naturforsch. A: Phys. Sci.* **1977,** *32*, 1178.

246. Balachandran, V.; Murugan, M.; Nataraj, A.; Karnan, M.; Ilango, G., Comparative Vibrational Spectroscopic Studies, HOMO-LUMO, NBO Analyses and Thermodynamic Functions of p-Cresol and 2-Methyl-p-Cresol based on DFT Calculations. *Spectrochim. Acta, Part A* **2014,** *132*, 538.

247. Dieng, S. D.; Schelvis, J. P. M., Analysis of Measured and Calculated Raman Spectra of lndole, 3-Methylindole, and Tryptophan on the Basis of Observed and Predicted Isotope Shifts. *J. Phys. Chem. A* **2010,** *114*, 10897.

248. Gaff, J. F.; Franzen, S., Excited-State Geometry Method for Calculation of the Absolute Resonance Raman Cross Sections of the Aromatic Amino Acids. *J. Phys. Chem. A* **2009,** *113*, 5414.

249. Hellweg, A.; Haettig, C., On the Internal Rotations in p-Cresol in its Ground and First Electronically Excited States. *J. Chem. Phys.* **2007,** *127*, 024307.

250. Hameka, H. F.; Jensen, J. O., Calculations of the Energies, Geometries and Vibrational Frequencies of the Ground and First Excited Singlet States of Toluene and p-Cresol. *J. Mol. Struct.: THEOCHEM* **1995,** *331*, 203.

251. Somers, K. R. F.; Ceulemans, A., Absorption and Fluorescence of 3-Methylindole: A Theoretical Study, Including H_2O Interactions. *J. Phys. Chem. A* **2004,** *108*, 7577.

252. Jean, L.; Lee, C. F.; Shaw, M.; Vaux, D. J., Structural Elements Regulating Amyloidogenesis: A Cholinesterase Model System. *Plos One* **2008,** *3*, e1834.

253. Monnot, M.; Gilles, A. M.; Saintgirons, I.; Michelson, S.; Barzu, O.; Fermandjian, S., Circular Dichroism Investigation of Escherichia Coli Adenylate Kinase. *J. Biol. Chem* **1987,** *262*, 2502.

254. Permyakov, S. E.; Pershikova, I. V.; Zhadan, A. P.; Goers, J.; Bakunts, A. G.; Uversky, V. N.; Berliner, L. J.; Permyakov, E. A., Conversion of Human alpha-Lactalbumin to An Apo-Like State in the Complexes with Basic Poly-Amino Acids: Toward Understanding of the Molecular Mechanism of Antitumor Action of HAMLET. *J. Proteome Res.* **2005,** *4*, 564.

255. Naylor, C. E.; Jepson, M.; Crane, D. T.; Titball, R. W.; Miller, J.; Basak, A. K.; Bolgiano, B., Characterisation of the Calcium-Binding C-Terminal Domain of ClostridiumPerfringens Alpha-Toxin. *J. Mol. Biol.* **1999,** *294*, 757.

256. Weers, P. M. M.; Kay, C. M.; Oikawa, K.; Wientzek, M.; Vanderhorst, D. J.; Ryan, R. O., Factors Affecting the Stability and Conformation of Locusta Migratoria Apolipophorin III. *Biochemistry* **1994,** *33*, 3617.

257. Brown, E. M.; Groves, M. L., Effect of Temperature on the Circular Dichroism Spectra of Beta-2-Microglobulins. *FEBS Lett.* **1985,** *184*, 36.

258. Vanhove, M.; Raquet, X.; Frere, J. M., Investigation of the Folding Pathway of the Tem-1 Beta-Lactamase. *Proteins: Struct. Funct. Genet.* **1995,** *22*, 110.

259. Bolgiano, B.; Crane, D. T.; Xing, D.; Williams, L.; Jones, C.; Corbel, M. J., Physico-Chemical Analysis of Bordetella Pertussis Antigens. *Biologicals* **1999,** *27*, 155.

260. Galat, A.; Yang, C. C.; Blout, E. R., Circular Dichroism Study of the Unfolding Refolding of A Cardiotoxin from Taiwan Cobra (Naja-Naja-Atra) Venom. *Biochemistry* **1985,** *24*, 5678.

261. Khan, F.; Khan, R. H.; Muzammil, S., Alcohol-Induced versus Anion-Induced States of Alpha-Chymotrypsinogen A at Low pH. *BBA-Protein Struct. M.* **2000,** *1481*, 229.

262. Krell, T.; Horsburgh, M. J.; Cooper, A.; Kelly, S. M.; Coggins, J. R., Localization of the Active Site of Type II Dehydroquinases-Identification of a Common Arginine-Containing Motif in the Two Classes of Dehydroquinases. *J. Biol. Chem.* **1996,** *271*, 24492.

263. Svensson, A. K. E.; O'Neill, J. C.; Matthews, C. R., The Coordination of the Isomerization of A Conserved Non-Prolyl Cis Peptide Bond with the Rate-Limiting Steps in the Folding of Dihydrofolate Reductase. *J. Mol. Biol.* **2003,** *326*, 569.

264. Ranjbar, B.; Gill, P., Circular Dichroism Techniques: Biomolecular and Nanostructural Analyses- A Review. *Chem. Biol. Drug Des.* **2009,** *74*, 101.

265. Sjoholm, I.; Ljungste, I., Studies on Tryptophan and Drug Binding Properties of Human Serum Albumin Fragments by Affinity Chromatography and Circular Dichroism Measurements. *J. Biol. Chem.* **1973,** *248*, 8434.

266. Hinds, K.; Koh, J. J.; Joss, L.; Liu, F.; Baudys, M.; Kim, S. W.,

Synthesis and Characterization of Poly(Ethylene Glycol)-Insulin Conjugates. *Bioconjugate Chem.* **2000,** *11*, 195.

267. Windsor, W. T.; Syto, R.; Le, H. V.; Trotta, P. P., Analysis of the Conformation and Stability of Escherichia Coli Derived Recombinant Human Interleukin 4 by Circular Dichroism. *Biochemistry* **1991,** *30*, 1259.

268. Matthews, J. M.; Hammacher, A.; Howlett, G. J.; Simpson, R. J., Physicochemical Characterization of An Antagonistic Human Interleukin 6 Dimer. *Biochemistry* **1998,** *37*, 10671.

269. Patra, A. K.; Udgaonkar, J. B., Characterization of the Folding and Unfolding Reactions of Single-Chain Monellin: Evidence for Multiple Intermediates and Competing Pathways. *Biochemistry* **2007,** *46*, 11727.

270. Sirangelo, I.; Bismuto, E.; Tavassi, S.; Irace, G., Near-Ultraviolet Circular Dichroic Activity of Apomyoglobin: Resolution of the Individual Tryptophanyl Contributions by Site-Directed Mutagenesis. *Eur. Biophys. J. Biophy. Lett.* **1998,** *27*, 27.

271. Chi, H.-W.; Chien, Y.-C.; Liu, C.-Y.; Tseng, C.-J.; Lee, Y.-J.; Chan, J.-L.; Chu, Y.-R.; Chin, D.-H., Role of Steric Effects in Protein-Directed Enediyne Cycloaromatization of Neocarzinostatin. *Chem. Eur. J.* **2011,** *17*, 1493.

272. Parisi, M.; Mazzini, A.; Sorbi, R. T.; Ramoni, R.; Grolli, S.; Favilla, R., Unfolding and Refolding of Porcine Odorant Binding Protein in Guanidinium Hydrochloride: Equilibrium Studies at Neutral pH. *BBA-Proteins Proteom.* **2003,** *1652*, 115.

273. Edwin, F.; Jagannadham, M. V., Sequential Unfolding of Papain in Molten Globule State. *Biochem. Biophys. Res. Commun.* **1998,** *252*, 654.

274. Kamen, D. E.; Woody, R. W., A Partially Folded Intermediate Conformation is Induced in Pectate Lyase C by the Addition of 8-Anilino-1-Naphthalenesulfonate (ANS). *Prot. Sci.* **2001,** *10*, 2123.

275. Vallee, B.; Teyssier, C.; Maget-Dana, R.; Ramstein, J.; Bureaud, N.; Schoentgen, F., Stability and Physicochemical Properties of the Bovine Brain Phosphatidylethanolamine-Binding Protein. *Eur. J. Biochem.* **1999,** *266*, 40.

276. Ikeda, K.; Samejima, Y., Bindings of Ca^{2+} Ion and Monodispersed n-Alkylphosphorylcholines to Phospholipase A_2-II of A. Halys Blomhoffii. *J. Biochem. (Tokyo, Jpn.)* **1981,** *89*, 1175.

277. Shire, S. J.; Holladay, L. A.; Rinderknecht, E., Self-Association of Human and Porcine Relaxin as Assessed by Analytical Ultracentrifugation and Circular Dichroism. *Biochemistry* **1991,** *30*, 7703.

278. Cannella, C.; Costa, M.; Pensa, B.; Ricci, G.; Pecci, L.; Cavallini, D., Multiple Forms of Bovine Liver Rhodanese. *Eur. J. Biochem.* **1981,** *119*, 491.

279. Shastry, M. C. R.; Eftink, M. R., Reversible Thermal Unfolding of

Ribonuclease T-1 in Reverse Micelles. *Biochemistry* **1996,** *35*, 4094.

280. Filfil, R.; Chalikian, T. V., Volumetric and Spectroscopic Characterizations of the Native and Acid-Induced Denatured States of Staphylococcal Nuclease. *J. Mol. Biol.* **2000,** *299*, 827.

281. Andersson, D.; Carlsson, U.; Freskgard, P. O., Contribution of Tryptophan Residues to the CD Spectrum of the Extracellular Domain of Human Tissue Factor: Application in Folding Studies and Prediction of Secondary Structure. *Eur. J. Biochem.* **2001,** *268*, 1118.

282. Mancheno, J. M.; De Los Rios, V.; Del Pozo, A. M.; Lanio, M. E.; Onaderra, M.; Gavilanes, J. G., Partially Folded States of the Cytolytic Protein Sticholysin II. *BBA-Protein Struct. M.* **2001,** *1545*, 122.

283. Georgescu, R. E.; Braswell, E. H.; Zhu, D.; Tasayco, M. L., Energetics of Assembling An Artificial Heterodimer with An Alpha/Beta Motif: Cleaved versus Uncleaved Escherichia Coli Thioredoxin. *Biochemistry* **1999,** *38*, 13355.

284. Vriend, G., What If-A Molecular Modeling and Drug Design Program. *J. Mol. Graphics* **1990,** *8*, 52.

285. Ramachandran, S.; Kota, P.; Ding, F.; Dokholyan, N. V., Automated Minimization of Steric Clashes in Protein Structures. *Proteins: Struct. Funct. Bioinf.* **2011,** *79*, 261.

286. Deng, H.; Jia, Y.; Zhang, Y., 3DRobot: Automated Generation of Diverse and Well-packed Protein Structure Decoys. *Boinformatics* **2016,** *32*, 378.

287. Bois, C., Structure of p-Cresol at Low Temperature. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **1970,** *26*, 2086.

288. Roos, B. O.; Fulscher, M. P.; Malmqvist, P.-A.; Merchan, M.; Serrano-Andres, L., *Theoretical Studies of Electronic Spectra of Organic Molecules*. Dordrecht: Netherlands, 1995.

289. Carney, J. R.; Hagemeister, F. C.; Zwier, T. S., Hydrogen Bonding Topologies of Indole-(water)(n) Clusters from Resonant Ion-Dip Infrared Spectroscopy. *J. Chem. Phys.* **1998,** *108*, 3379.

290. Nibu, Y.; Abe, H.; Mikami, N.; Ito, M., Electronic Spectra of Hydrogen Bonded Indoles in a Supersonic Free Jet. *J. Phys. Chem.* **1983,** *87*, 3898.

291. Walden, S. E.; Wheeler, R. A., Structural and Vibrational Analysis of Indole by Density Functional and Hybrid Hartree-Fock/Density Functional Methods. *J. Chem. Soc., Perkin Trans. 2* **1996,** *0,* 2653.

292. Brand, C.; Kuepper, J.; Pratt, D. W.; Meerts, W. L.; Kruegler, D.; Tatchen, J.; Schmitt, M., Vibronic Coupling in Indole: I. Theoretical Description of the (1)L(a)-(1)L(b) Interaction and the Electronic Spectrum.

Phys. Chem. Chem. Phys. **2010,** *12*, 4968.

293. Chabalowski, C. F.; Garmer, D. R.; Jensen, J. O.; Krauss, M., Reaction Field Calculation of the Spectral Shifts of Indole. *J. Phys. Chem.* **1993,** *97*, 4608.

294. Callis, P. R.; Vivian, J. T.; Slater, L. S., Ab Initio Calculations of Vibronic Spectra for Indole. *Chem. Phys. Lett.* **1995,** *244*, 53.

295. Kuepper, J.; Pratt, D. W.; Meerts, W. L.; Brand, C.; Tatchen, J.; Schmitt, M., Vibronic Coupling in Indole: II. Investigation of the (1)L(a)-(1)L(b) Interaction using Rotationally Resolved Electronic Spectroscopy. *Phys. Chem. Chem. Phys.* **2010,** *12*, 4980.

296. Ten, G. N.; Yakovleva, A. A.; Berezin, M. K.; Baranov, V. I., Calculation and Interpretation of the Absorption and Fluorescence Spectra of Indole in the Isolated State and Aqueous Solution. *Opt. Spectrosc.* **2013,** *114*, 590.

297. Strickland, E. H.; Horwitz, J.; Billups, C., Near-Ultraviolet Absorption Bands of Tryptophan-Studies using Indole and 3-Methylindole as Models. *Biochemistry* **1970,** *9*, 4914.

298. Yang, P.; Pang, M.; Shen, W.; Li, M.; He, R., Vibronic Analysis of the L-1(b) <-> S-0 Transitions of Indole and 3-Methylindole: The Influence of Anharmonic, Duschinsky, and Herzberg-Teller Contributions. *J. Mol. Spectrosc.* **2015,** *313*, 40.

299. Chadwick, D. J., *Comprehensive Heterocyclic Chemistry: The Structure, Reactions, Synthesis, and Uses of Heterocyclic Compounds*. Pergammon: Oxford, 1984.

300. Bunte, S. W.; Jensen, G. M.; McNesby, K. L.; Goodin, D. B.; Chabalowski, C. F.; Nieminen, R. M.; Suhai, S.; Jalkanen, K. J., Theoretical Determination of the Vibrational Absorption and Raman Spectra of 3-Methylindole and 3-Methylindole Radicals. *Chem. Phys.* **2001,** *265*, 13.

301. Lin, J. L.; Zhang, S.; Tzeng, W. B., Mass Analyzed Threshold Ionization Spectroscopy of 5-Methylindole and 3-Methylindole Cations and the Methyl Substitution Effect. *J. Chem. Phys.* **2004,** *120*, 5057.

302. Platt, J. R., Experimental Determination of Even-Odd Charater of Excited Electronic States of Molecules with A Center of Symmetry. *J. Chem. Phys.* **1951,** *19*, 1418.

303. Strickland, E. H., Aromatic Contributions to Circular Dichroism Spectra of Proteins. *CRC Crit. Rev. Biochem.* **1974,** *2*, 113.

304. Woody, R. W., Application of Bergson Model to Optical Properties of Chiral Disulfides. *Tetrahedron* **1973,** *29*, 1273.

305. Webb, J.; Strickland, R. W.; Richardson, F. S., Optical Activity of Chiral Disulfide Chromophores. *J. Am. Chem. Soc.* **1973,** *95*, 4775.

306. Casey, J. P.; Martin, R. B., Disulfide Stereochemistry. Conformations and Chiroptical Properties of L-Cystine Derivatives. *J. Am. Chem. Soc.* **1972,** *94*, 6141.

307. Martin, C.; Richard, V.; Salem, M.; Hartley, R.; Mauguen, Y., Refinement and Structural Analysis of Barnase at 1.5 Angstrom Resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999,** *55*, 386.

308. Bycroft, M.; Ludvigsen, S.; Fersht, A. R.; Poulsen, F. M., Determination of the 3-Dimensional Solution Structure of Barnase using Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **1991,** *30*, 8697.

309. Eberhardt, E. S.; Wittmayer, P. K.; Templer, B. M.; Raines, R. T., Contribution of A Tyrosine Side Chain to Ribonuclease A Catalysis and Stability. *Protein Sci.* **1996,** *5*, 1697.

310. Martensson, L. G.; Jonasson, P.; Freskgard, P. O.; Svensson, M.; Carlsson, U.; Jonsson, B. H., Contribution of Individual Tryptophan Residues to the Fluorescence Spectrum of Native and Denatured Forms of Human Carbonic Anhydrase II. *Biochemistry* **1995,** *34*, 1011.

311. Hynes, T. R.; Fox, R. O., The Crystal-Structure of Staphylococcal Nuclease Refined at 1.7 A Resolution. *Proteins: Struct., Funct., Bioinf.* **1991,** *10*, 92.

312. Wang, M.; Feng, Y.; Yao, H.; Wang, J., Importance of the C-Terminal Loop L137-S141 for the Folding and Folding Stability of Staphylococcal nuclease. *Biochemistry* **2010,** *49*, 4318.

313. Evans, C. L.; Long, J. E.; Gallagher, T. R. A.; Hirst, J. D.; Searle, M. S., Conformation and Dynamics of the Three-Helix Bundle UBA Domain of p62 from Experiment and Simulation. *Proteins* **2008,** *71*, 227.

314. Rocha, W. R.; De Almeida, K. J.; Coutinho, K.; Canuto, S., The Electronic Spectrum of N-Methylacetamide in Aqueous Solution: A Sequential Monte Carlo/Quantum Mechanical Study. *Chem. Phys. Lett.* **2001,** *345*, 171.

315. Takahashi, N.; Koseki, T.; Doi, E.; Hirose, M., Role of an Intrachain Disulfide Bond in the Conformation and Stability of Ovalbumin. *J. Biochem.* **1991,** *109*, 846.

316. McAuley, A.; Jacob, J.; Kolvenbach, C. G.; Westland, K.; Lee, H. J.; Brych, S. R.; Rehder, D.; Kleemann, G. R.; Brems, D. N.; Matsumura, M., Contributions of a Disulfide Bond to the Structure, Stability, and Dimerization of Human IgG1 Antibody C_H3 Domain. *Protein Sci.* **2008,** *17*, 95.

317. Woody, R. W., Application of Bergson Model to Optical Properties of Chiral Disulfides. *Tetrahedron* **1973,** *29*, 1273.

318. Kurapkat, G.; Kruger, P.; Wollmer, A.; Fleischhauer, J.; Kramer, B.; Zobel, E.; Koslowski, A.; Botterweck, H.; Woody, R. W., Calculations of the

CD Spectrum of Bovine Pancreatic Ribonuclease. *Biopolymers* **1997,** *41*, 267.

319. Hickman, C. G.; Gascooke, J. R.; Lawrance, W. D., The S-1-S-0(B-1(2)-(1)A(1)) Transition of Jet-Cooled Toluene: Excitation and Dispersed Fluorescence Spectra, Fluorescence Lifetimes, and Intramolecular Vibrational Energy Redistribution. *J. Chem. Phys.* **1996,** *104*, 4887.

320. Pitzer, K. S.; Scott, D. W., The Thermodynamics and Molecular Structure of Benzene and its Methyl Derivatives. *J. Am. Chem. Soc.* **1943,** *65*, 803.

321. Varsanyi, G., *Assignments for Vibrational Spectra of Seven Hundred Benzene Derivatives*. Wiley: New York, 1974.

322. Arp, Z.; Autrey, D.; Laane, J.; Overman, S. A.; Thomas, G. J., Structural Studies of Viruses by Raman Spectroscopy Part LXXI-Tyrosine Raman Signatures of the Filamentous Virus Ff are Diagnostic of Non-Hydrogen-Bonded Phenoxyls: Demonstration by Raman and Infrared Spectroscopy of p-Cresol Vapor. *Biochemistry* **2001,** *40*, 2522.

323. Jakobsen, R. J., Vibrational Spectra of p-Cresol. *Spectrochim. Acta* **1965,** *21*, 433.

324. Combs, A.; McCann, K.; Autrey, D.; Laane, J.; Overman, S. A.; Thomas, G. J., Raman Signature of the Non-Hydrogen-Bonded Tryptophan Side Chain in Proteins: Experimental and Ab Initio Spectra of 3-Methylindole in the Gas Phase. *J. Mol. Struct.* **2005,** *735*, 271.

325. Jensen, G. M.; Goodin, D. B.; Bunte, S. W., Density Functional and MP2 Calculations of Spin Densities of Oxidized 3-Methylindole: Models for Tryptophan Radicals. *J. Phys. Chem.* **1996,** *100*, 954.

Appendix A

Calculated Geometries and Vibrational Frequencies Assignment for Toluene, p-Cresol and 3-Methylindole

Toluene

Our calculated geometry (Table A.1) and vibrational frequencies (Table A.2) of the ground and the first excited states of toluene are in good agreement with the experimental[242,245,319,320] and previous calculated results.[250] The assignment of the vibrational modes is differs from the previous study.[319,321] The C-H bending mode (1163 cm^{-1}) was assigned to the band 342 cm^{-1} in the experiment by Varsanyi[321] and to the band 1080 cm^{-1} by Pitzer and Scott.[320] There are some discrepancies in Varsanyi's and Pitzer's assignments. Our assignments match better with Varsanyi's. The vibrational motion at band 1328 cm^{-1} in the ground state appears at both bands 1893 cm^{-1} and 1262 cm^{-1} in the excited state, which refer to the C-C and C-H streches, respectively. The vibrational modes assigned to the bands at 1561 cm^{-1} and 1566 cm^{-1} in the excited state appear at 1569 cm^{-1} and 1648 cm^{-1} in the ground state.

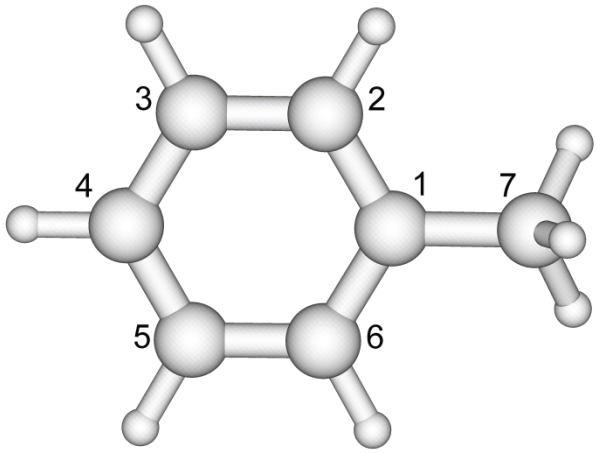


Figure A.1 Structure of toluene and numbering convention of the non-hydrogen atoms.

Table A.1 Calculated geometries of the ground (S_0) and the first (^1L_b) excited state of toluene compared to their experimental and calculated counterparts.

bonds	**S_0 geometry / Å**			**^1L_b geometry / Å**	
	Cal.	**Exp.**[245]	**Lit. Cal.**[250]	**Cal.**	**Lit. Cal.**[250]
C1-C2	1.401	1.399	1.396	1.439	1.472
C2-C3	1.396	1.399	1.398	1.433	1.474
C3-C4	1.396	1.399	1.378	1.432	1.428
C4-C5	1.396	1.399	1.398	1.432	1.359
C5-C6	1.396	1.399	1.394	1.433	1.430
C1-C6	1.401	1.399	1.405	1.439	1.478
C1-C7	1.513	1.511	1.511	1.504	1.497
C2-H	1.076	1.098	1.074	1.074	1.072
C3-H	1.076	1.098	1.074	1.073	1.071
C4-H	1.075	1.098	1.073	1.073	1.074
C5-H	1.076	1.098	1.073	1.073	1.074
C6-H	1.076	1.098	1.075	1.074	1.072
C7-H	1.085	1.12	1.083	1.085	1.084
C7-H	1.085	1.12	1.085	1.085	1.085
C7-H	1.087	1.12	1.085	1.089	1.090
angles	**S_0 geometry / degree**			**^1L_b geometry / degree**	
C1-C7-H	111.3	-	111.3	111.3	111.3

Table A.2 Calculated vibrational frequencies of the ground (S_0) and the first (^1L_b) excited state of toluene compared to their experimental and calculated counterparts.

symmetry	assignment	S_0 frequency / cm^{-1}				^1L_b frequency / cm^{-1}		
		Calc.	Expt.		Lit. Calc.[a,321]	Calc.	Expt.[319]	Lit. Calc.[a,321]
			Expt.[242]	Expt.[320]				
In Plane								
a_1	C-C stretch	1755	1606	1605	1601	1695	-	1636
	C-C stretch	1648	1497	1494	1506	1561/1566	-	-
	C-H bend (CH_3)	1637	1497	-	1513	1632	-	1504
	C-H bend (CH_3)	1569	-	-	-	1561/1566	-	-
	C-CH_3 stretch	1312	1211	-	1214	1286	-	1178
	C-H bend	1281	1182	1175	1197	1250	1021	1203
	C-H bend	1107	-	1030	-	978	935	-
	C-C bend	1069	1005	1003	1030	1035	966	977
	C-C bend	838	788	785	779	787	736	555
	C-C bend	555	522	521	522	503	457/462	459
b_2	C-C stretch	1724	1587	1586	1587	1667	-	1520
	C-H bend (CH_3)	1642	1497	-	1513	1625	-	1504
	C-C stretch	1586	1445	1445	1455	1519	-	1451
	C-H bend	1469	-	-	-	1435	-	-
	C-H bend	1328	-	1312	-	1893/1262	-	-

	C-H bend	1225	1157	1155	1153	1139	-	1126
	C-H bend	1163	1084	1080	1090	1139	-	1058
	C-CH_3 bend	1074	-	-	-	986	-	-
	C-C bend	668	622	623	637	588	532	721
	C-CH_3 wag	362	348	342	354	349	332	296
	Out of plane							
	C-H bend (CH_3)	1149	1041	-	1078	1110	-	1012
	C-C bend	1029	978	978	1032	711	-	1005
	C-H bend	931	898	895	904	647	-	519
b_1	C-H bend	768	732	728	733	562	-	458
	C-C bend	728	698	695	698	495	423	312
	C-C bend	492	467	464	466	353	320	264
	C-CH_3 wag	221	220	216	225	158	157	85
	C-C bend	998	967	964	966	653	687	790
a_2	C-H bend	881	841	843	836	583	-	471
	C-C bend	436	407	407	409	283	228	161
	Other modes							
	C-CH_3 rotation	18	-	-	-	39	-	-

a. The calculated vibrational frequencies are the scaled values reported in the literature.

p-cresol

The calculated geometry (Table A.3) and vibrational frequencies (Table A.4) of the ground and the first excited states of *p*-cresol are compared with the experimental[287,322] and previous reported calculated results.[246,250] Both geometries and vibrational frequencies are in good agreement with experimental and theoretical results in the literature. DFT calculations with various size of basis set give similar results.[246,322] We compared our calculations with B3LYP/6-311++G** calculations (Table A.4). Four vibrational modes, which were assigned to 842 cm^{-1}, 1374 cm^{-1}, 1470 cm^{-1} and 1568 cm^{-1} in the ground state, were assigned to more than one band in the excited state.

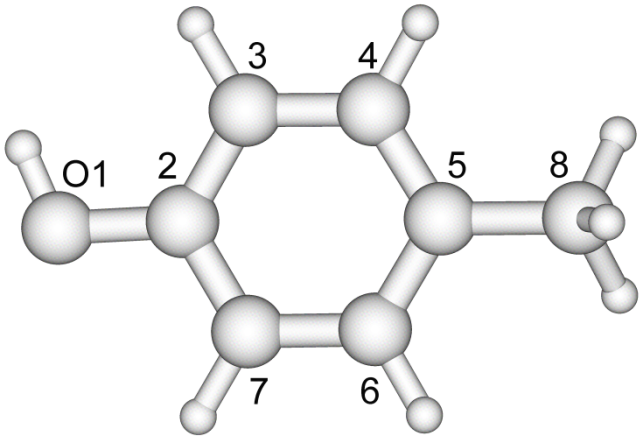


Figure A.2 Structure of *p*-cresol and numbering convention of the non-hydrogen atoms.

Table A.3 Calculated geometries of the ground (S_0) and the first (^1L_b) excited state of *p*-cresol compared to their experimental and calculated counterparts.

bonds	**S_0 geometry / Å**				**^1L_b geometry / Å**	
	Cal.	**Exp.**[287]	**Cal. Lit.**		**Cal.**	**Cal. Lit.**
			CASSCF[250]	**B3LYP**[246]		**CASSCF**[250]
O1-C2	1.361	1.387	1.380	1.368	1.354	1.377
C2-C3	1.392	1.397	1.371	1.399	1.429	1.358
C3-C4	1.398	1.401	1.402	1.396	1.431	1.426
C4-C5	1.397	1.391	1.392	1.403	1.439	1.476
C5-C6	1.403	1.406	1.408	1.403	1.438	1.470
C6-C7	1.393	1.397	1.387	1.396	1.434	1.473
C2-C7	1.396	1.388	1.398	1.402	1.425	1.427
C5-C8	1.513	1.520	1.511	1.511	1.504	1.497
O1-H	0.947	0.9	0.950	0.968	0.947	0.950
C3-H	1.077	1.0	1.075	1.092	0.947	1.075
C4-H	1.076	0.9	1.074	1.094	1.073	1.071
C6-H	1.076	1.0	1.074	1.094	1.073	1.071
C7-H	1.074	1.0	1.071	-	1.072	1.069
C8-H	1.085	0.9	1.085	1.103	1.085	1.084
C8-H	1.085	1.0	1.083	1.103	1.085	1.085
C8-H	1.087	1.1	1.085	1.100	1.087	1.090
angles	**S_0 geometry / degree**				**^1L_b geometry / degree**	
H-O1-C2	110.4	120.0	114.6	108.6	110.6	114.7
C5-C8-H	111.2	-	111.3	111.6	111.2	111.3
C7-C2-O1	117.6	118.7	116.3	117.8	116.8	114.0

Table A.4 Calculated vibrational frequencies of the ground (S_0) and the first (^1L_b) excited state of *p*-cresol compared to their experimental and calculated counterparts.

symmetry	assignment	S_0 frequency / cm^{-1}				^1L_b frequency / cm^{-1}	
		Calc.	Expt.[323]	Calc. Lit.		Calc.	Calc. Lit.[c,250]
				CASSCF[c,250]	B3LYP[322]		
	In Plane						
a_1	C-C stretch	1778	1621	1624	1656	1720	1649
	C-C stretch	1677	1520	1537	1545	1590	1466
	CH_3 bend	1637	1385[b]	1436	1415	1632	1430
	C-O stretch	1387	1255	1245	1274	1379	1182
	C-CH_3 stretch	1310	1215[a]	1194	1229	1279	1153
	C-H bend	1280	1178	1128	1195	1242	1105
	C-H bend	1099	1015	1008	1031	1014	1069
	C-C bend	890	840[b]	823	851	847	769
	C-H bend	790	739	727	745	751	576
	C-H bend	490	459	459	467	444	407
b_2	C-C stretch	1738	-	1599	1632	1680	1525
	C-C stretch	1568	1431	1456	1456	1486/1566/1895	1450
	C-H bend	1470	1298[a]	1294	1332	1434/1486	1270
	C-H bend	1374	1337	1366	1359	1296/1895	1369
	C-H bend	1194	1114	1077	1129	1156	-
	C-C bend	695	642	657	657	618	696

	C-O wag	453	420	430	426	433	414
	C-CH_3 wag	324	333	313	304	320	303
	Out of Plane						
b_1	C-C bend	967	918[a]	936	919	634	454
	C-C bend	842	811	805	804	412/564	116
	C-C bend	721	698	698	693	488	576
	C-C bend	535	503	510	513	564	244
	C-H bend	347	294	294	331	255	335
	CH wag, O-H wag	155	-	157	144	114	74
a_2	C-H bend	863	-	824	830	564	430
	C-C bend	442	-	416	419	273	225
	Other modes						
	CH_3 bend	1571	1472[a]	1513	1487	1566	1505
	O-H wag	1261	1187[a]	1214	1185	1895	1310
	CH_3 rock	1085	1105	1073	1060	1018	975
	O-H wag	259	-	-	296	247	-
	C-CH_3 rotation	25	-	55[d]	18	63	-

a. Data are from Raman vapor.[322] b. IR vapor data from different reference.[322] c. The calculated vibrational frequencies are scaled values as reported in the literature. d. This value has not been scaled.

3-methylindole

The calculated geometry (Table A.5) and vibrational frequencies (Table A.6) of 3-MI are compared with the experimental[299,300,324,325] and previous reported calculated results.[251,324] Our calculated geometries are in good agreement with earlier CASSCF/ANO-S calculations for all three states. The vibrational frequencies of the ground state are compared with the IR frequencies and B3LYP/TZ2P calculation. Our calculation shows good. However, the frequencies are overestimated for some high energy modes. The N-H wag mode and C-CH_3 bend mode, which were assigned to bands 1420 cm^{-1} and 1387 cm^{-1} in the IR spectrum[300] appear together in bands 1554 cm^{-1} and 1575 cm^{-1} in the ground state as similar motions. Most of the vibrational modes in the $1L_b$ state can be matched to corresponding ground state vibrations. However, the in-plane stretch mode 1477 cm^{-1} is assigned to two modes in the $1L_b$ state (1425 cm^{-1} contains the skeletal motion while 1520 cm^{-1} contains the C-H bend motion on the benzene ring). The assignments of the ^1L_a state differ from the other two states. There are cases where a normal mode was assigned to a single band in the ground state while its vibrational character appears in two frequencies in the ^1L_a state or two or more vibrational modes are combined in the second excited state.

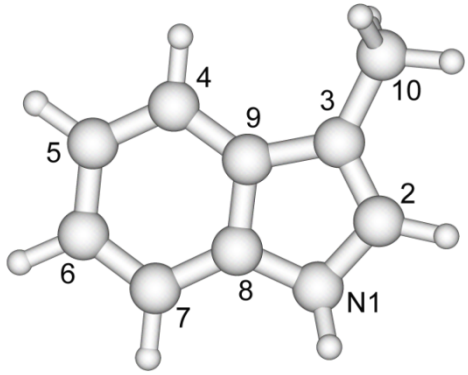


Figure A.3 Structure of 3-methylindole and numbering convention of the non-hydrogen atoms.

Table A.5 Calculated geometries of the ground (S_0), the first (^1L_b) and the second (^1L_a) excited state of 3-methylindole compared to their experimental and calculated counterparts.

bonds	S_0 geometry / Å			^1L_b geometry / Å		^1L_a geometry / Å	
	Cal.	Exp.[299]	Lit. Cal.[251]	Cal.	Lit. Cal.[251]	Cal.	Lit. Cal.[251]
N1-C2	1.394	1.380	1.386	1.404	1.399	1.322	1.303
C2-C3	1.368	1.358	1.368	1.385	1.384	1.442	1.441
C3-C9	1.444	1.436	1.449	1.426	1.432	1.441	1.417
C4-C9	1.411	1.398	1.409	1.421	1.418	1.421	1.417
C4-C5	1.382	1.381	1.390	1.438	1.444	1.426	1.400
C5-C6	1.418	1.396	1.415	1.445	1.441	1.380	1.400
C6-C7	1.390	1.372	1.390	1.431	1.434	1.448	1.414
C7-C8	1.397	1.395	1.403	1.410	1.411	1.395	1.373
C8-C9	1.413	1.410	1.407	1.463	1.460	1.404	1.413
N1-C8	1.379	1.370	1.375	1.377	1.370	1.417	1.407
C3-C10	1.501	-	1.489	1.502	1.474	1.491	1.476
N1-H	0.994			0.994		0.998	
C2-H	1.071			1.069		1.068	
C4-H	1.076			1.073		1.072	
C5-H	1.075			1.074		1.077	
C6-H	1.075			1.073		1.076	
C7-H	1.076			1.074		1.073	
C10-H	1.085			1.085		1.084	
C10-H	1.086			1.086		1.086	
C10-H	1.086			1.086		1.086	
angles	**S_0 geometry / degree**			**^1L_b geometry / degree**		**^1L_a geometry / degree**	
C9-C8-N1-H	156.1			156.7		180.0	

Table A.6 Vibrational frequencies from unscaled CASSCF calculations the ground (S_0), the first (^1L_b) and the second (^1L_a) excited state of 3-methylindole compared to their experimental counterparts and scaled calculations using DFT.

assignment	S_0 frequency / cm^{-1}			Cal. frequency / cm^{-1}	
	Cal.	IR[300]	Cal. Lit.[300]	$^1L_{b.}$	^1L_a
In plane					
C-CH_3 wag	235	231	221	235	236
C-C bend	495	462	464	481	474
C-C bend	561	532	531	525	542
C-C bend	599	565	562	576	589
C-C bend	742	708	708	727	733
C-C stretch	806	758	762	775	784
benzene C-C bend	940	876	878	901	733
benzene C-H bend	1049	1009	1018	957	1047
C-CH_3 bend	1078	1070	1075	1055	1032
C-CH_3 stretch	1160	1070	1075	1142	1180
N1-C2 stretch	1174	1080	1090	1172	1215/1243
C-C stretch	1215	1249	1255	1255	1287
C-C stretch	1235	1249	1255	1312	1089
pyrrole stretch	1330	1229	1226	1395	1392
pyrrole stretch	1369	1302	1302	1351	1544
benzene C-H bend	1393	1334	1346	1395	1569
benzene C-H bend	1413	1334	1346	1626	1287
benzene C-H bend	1477	1345	1351	1425/1520	1392/1453
N1-H wag, C-CH_3 bend	1554	1420/1387	1424/1403	1589	1525
N1-H wag, C-CH_3 bend	1575	1420/1387	1424/1403	1565	1578

C-H bend	1610	1455	1463	1534	1569
benzene C-C stretch	1623	1493	1503	1626	1569
C-CH_3 bend	1646	1488	1481	1639	1638
C-C stretch	1688	1557	1570	1777	1764
benzene C-C stretch	1727	1577	1592	1670	1596
benzene C-C stretch	1800	1617	1633	2075	1720
C-C stretch, C-H wag					1282
C-C stretch, CH_3 bend					1638
Out of plane					
C-CH_3 rotation	142	177	135	111	89
C-C bend	160	177	151	135	123
C-C bend	234	-	219	191	173
C-C bend	306	231	244	286	321/1387
C-C bend	351	347	346	440	321/1387
benzene C-C bend	449	426	424	321	314
benzene C-C bend	591	573	581	412	554
C-C bend	610	601	609	524	586
C2-H wag	750	780	784	551	624
benzene C-H bend	779	731	737	558	749
C-H bend	780	758	760	629	686
benzene C-H bend	887	-	842	608	963
benzene C-H bend	973	925	928	658	-
benzene C-H bend	1011	983	965	690	-
CH_3 bend	1151	-	1051	1144	1110
CH_3 bend	1630	-	-	1630	1616
C7-H wag					391
C4-H wag					454

Appendix B

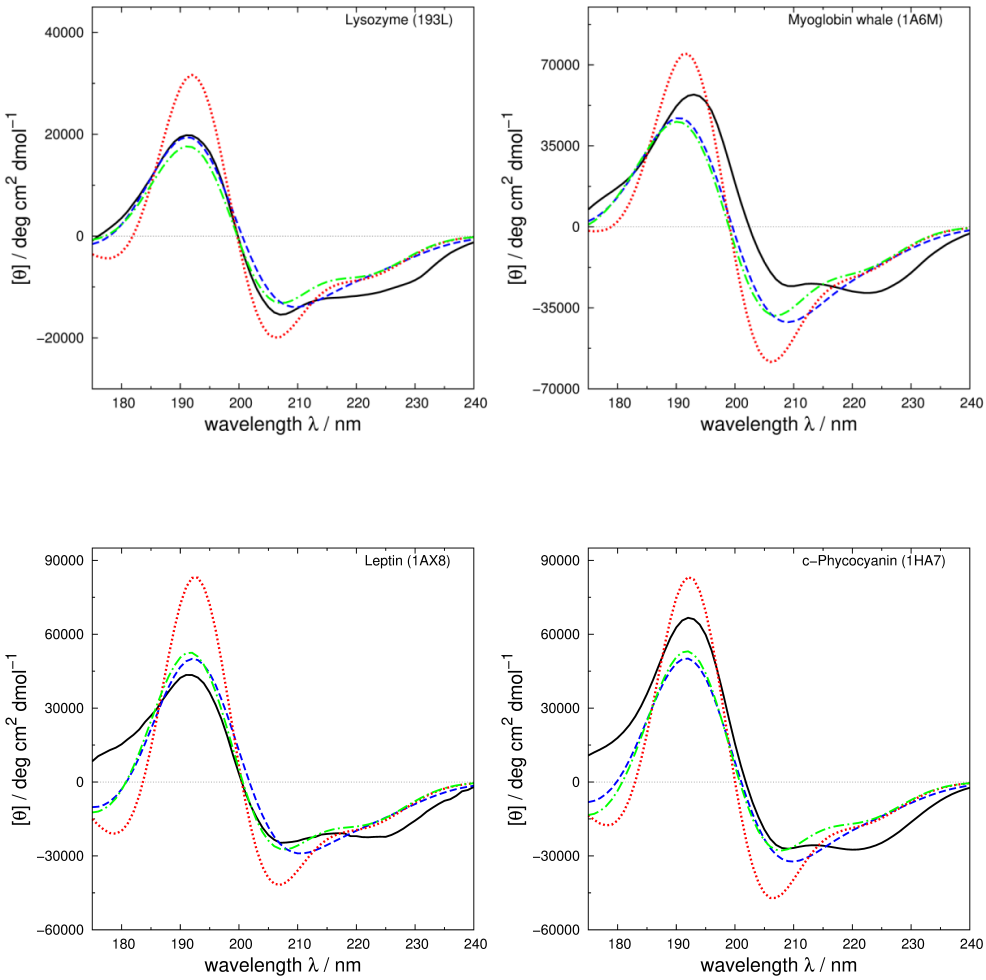
Experimental and Calculated Protein Far-UV CD Spectra

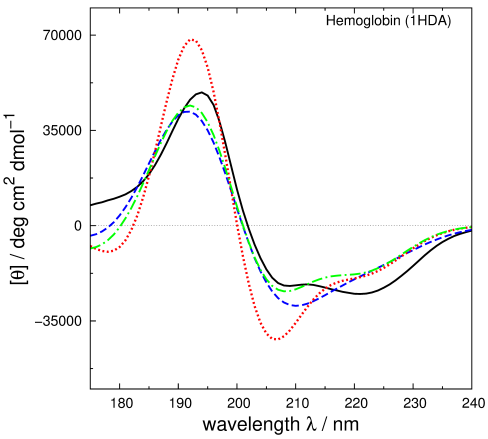
Experimental[208] and calculated CD spectra for protein set (Table 3.1) in the far-UV: black solid line, experiment; blue dashed line, non-vibronic calculation with a bandwidth, Δ = 12.5 nm; red dotted line, non-vibronic calculation with Δ = 10 nm; green dash-dotted line, vibronic calculation with Δ = 10 nm.

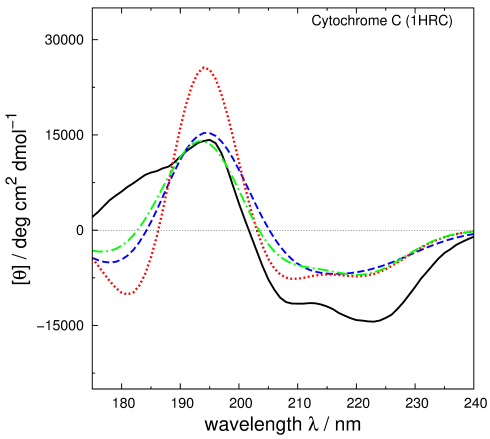
α-helix proteins

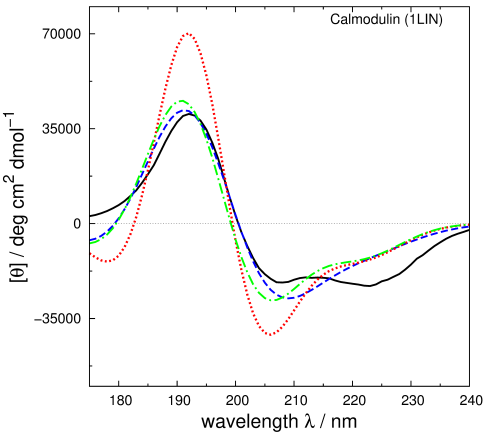
Calculated CD spectra of nine α-helical proteins are compared with their experimental spectra. The main features of α-helical proteins, which are a positive peak at 190 nm and two negative peaks at 208 and 222 nm, are reproduced by calculation taking the Franck-Condon effect into account. The two negative peaks are not distinguishable in the calculated spectra without the vibrational features in the parameter sets when the bandwidth is 12.5 nm. The intensities of the peaks are reproduced similar by both calculations with different bandwidth, namely Δ=10 nm for vibronic calculation and Δ=12.5 nm for non-vibronic calculation. A smaller bandwidth (Δ=10 nm) in

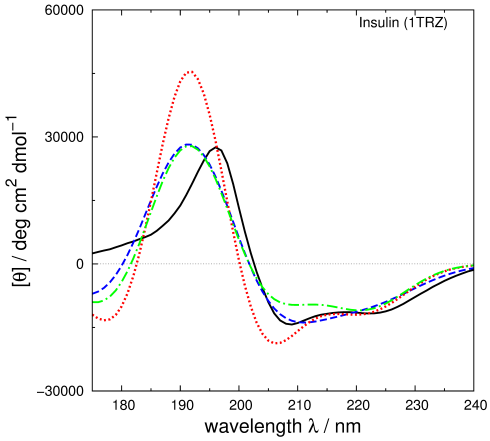
non-vibronic calculations would reproduce the two negative peaks. However, the intensities of CD signal are overestimated significantly.

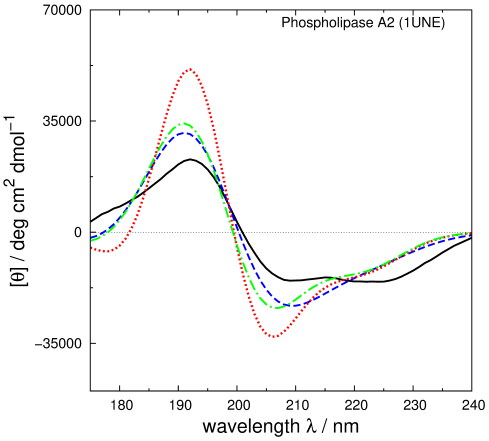



Hemoglobin (1HDA)
70000
35000
0
−35000
[θ] / deg cm2 dmol−1
180 190 200 210 220 230 240
wavelength λ / nm


Cytochrome C (1HRC)
30000
15000
0
−15000
[θ] / deg cm2 dmol−1
180 190 200 210 220 230 240
wavelength λ / nm

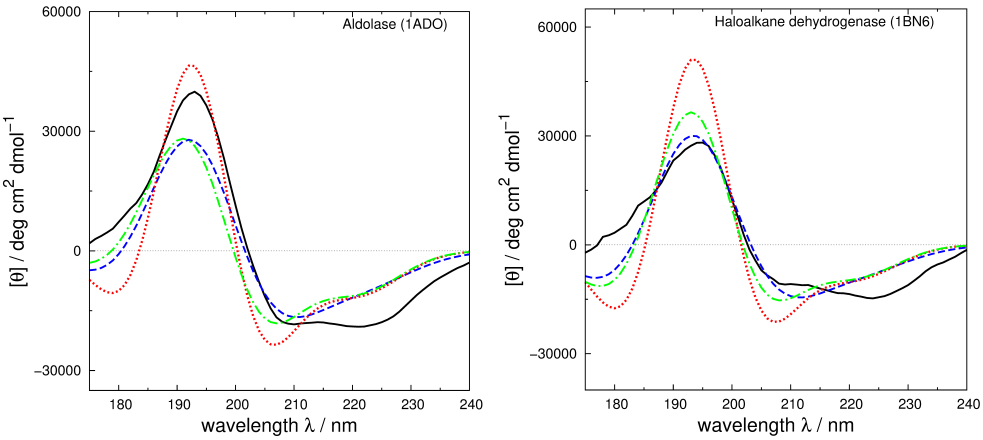

Calmodulin (1LIN)
70000
35000
0
−35000
[θ] / deg cm2 dmol−1
180 190 200 210 220 230 240
wavelength λ / nm

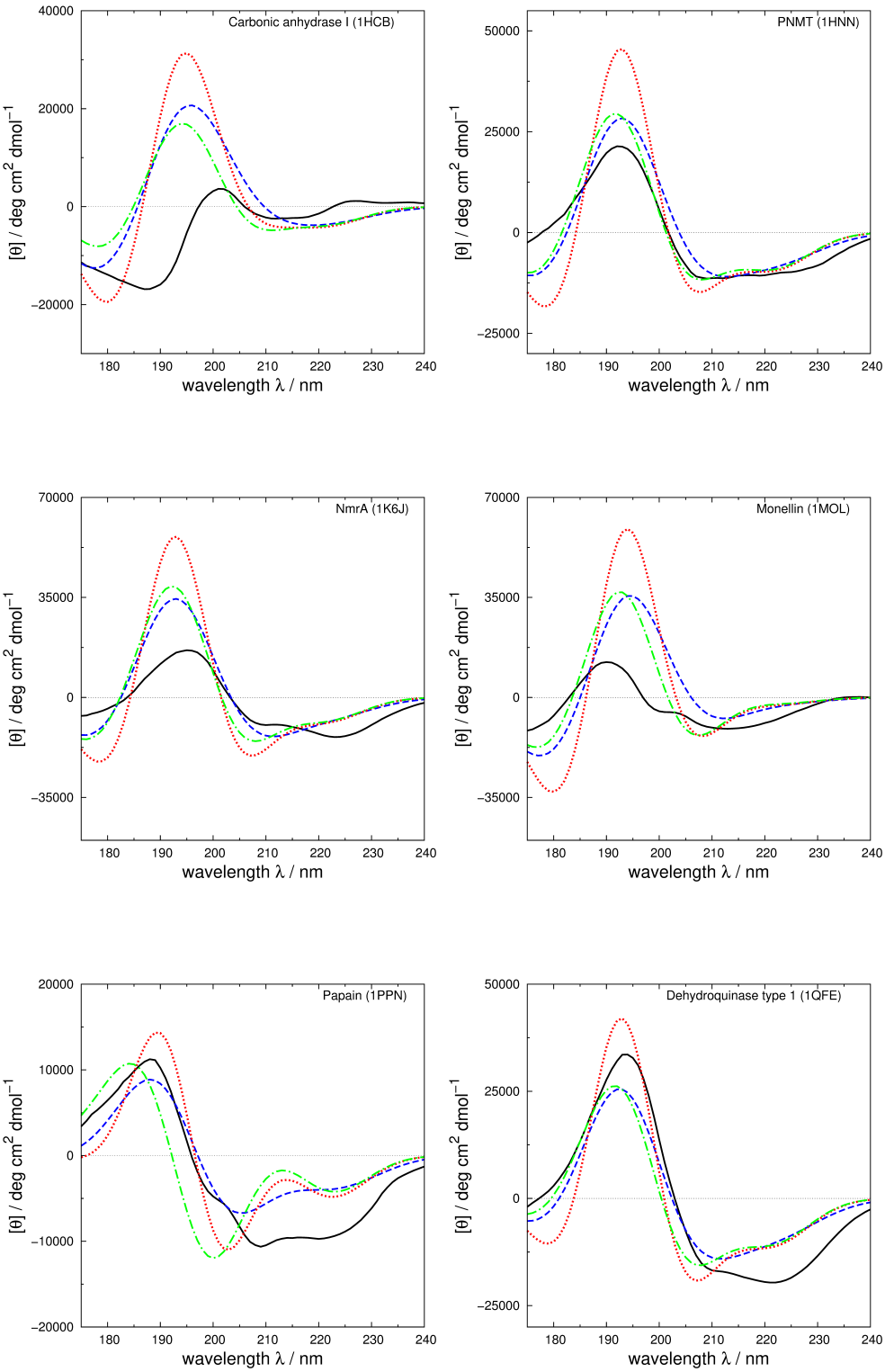

Insulin (1TRZ)
60000
30000
0
−30000
[θ] / deg cm2 dmol−1
180 190 200 210 220 230 240
wavelength λ / nm

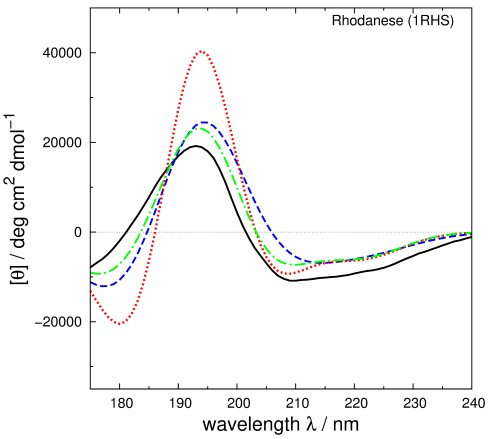

Phospholipase A2 (1UNE)
70000
35000
0
−35000
[θ] / deg cm2 dmol−1
180 190 200 210 220 230 240
wavelength λ / nm

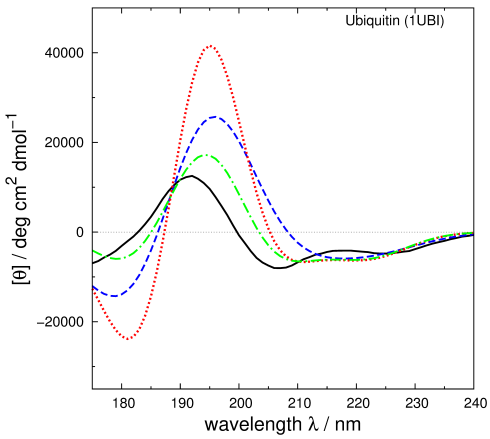
Mixed-α,β proteins

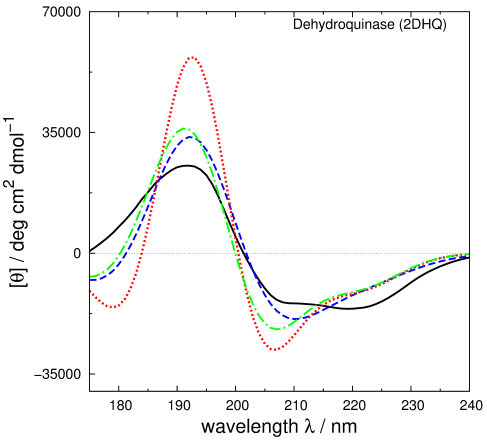
Calculated CD spectra of 14 mixed-α,β proteins are compared with their experimental spectra. The main features of mixed-α,β proteins are similar to the α-helical protein in some cases (e.g. aldolase, haloalkane delydrogenase, PNMT, etc.) depending on the content of the α-helical structure with fluctuated peak positions and intensities. The two negative peaks at wavelength larger than 200 nm, which were not generated by non-vibrational calculation, are reproduced by calculation with the Franck-Condon effect. The experimental spectra vary significantly among the mixed-α,β proteins. However, our calculations captured mainly the α-helical characters, which lead less well results for proteins like monellin, ubiquitin and ribonuclease A, etc.

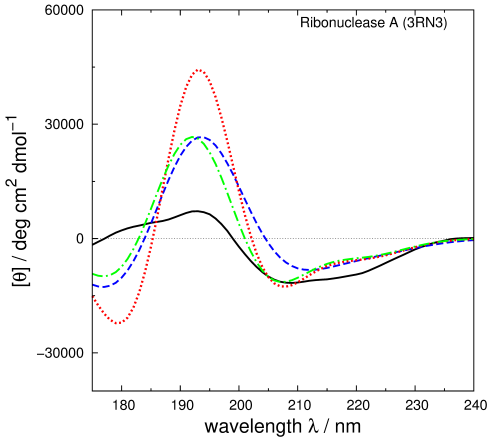


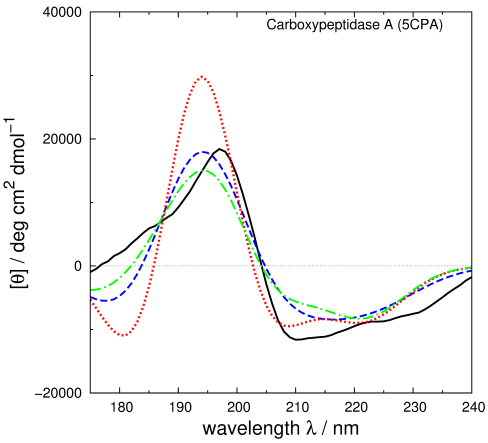

Carbonic anhydrase I (1HCB)
PNMT (1HNN)
NmrA (1K6J)
Monellin (1MOL)
Papain (1PPN)
Dehydroquinase type 1 (1QFE)
[θ] / deg cm2 dmol−1
wavelength λ / nm

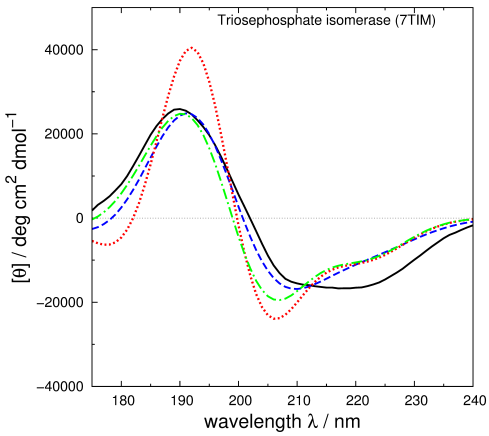

Rhodanese (1RHS)
[θ] / deg cm^2 dmol^{-1}
wavelength λ / nm


Ubiquitin (1UBI)
[θ] / deg cm^2 dmol^{-1}
wavelength λ / nm


Dehydroquinase (2DHQ)
[θ] / deg cm^2 dmol^{-1}
wavelength λ / nm

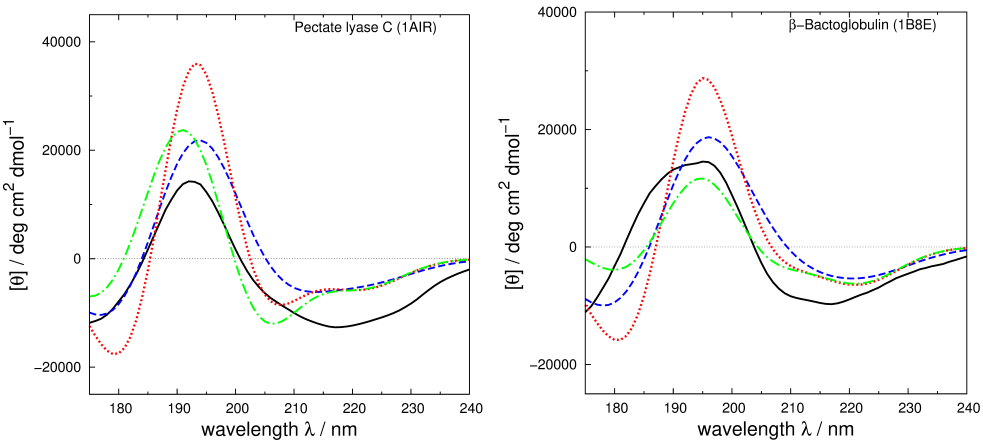

Ribonuclease A (3RN3)
[θ] / deg cm^2 dmol^{-1}
wavelength λ / nm

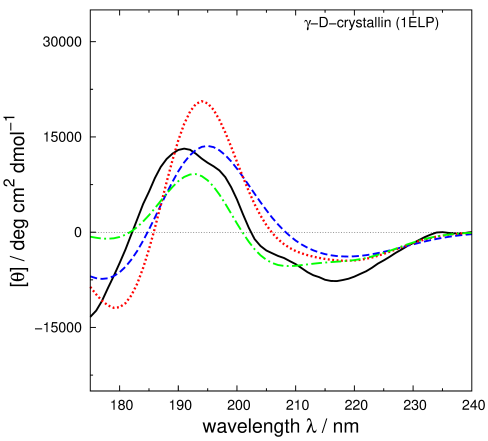

Carboxypeptidase A (5CPA)
[θ] / deg cm^2 dmol^{-1}
wavelength λ / nm

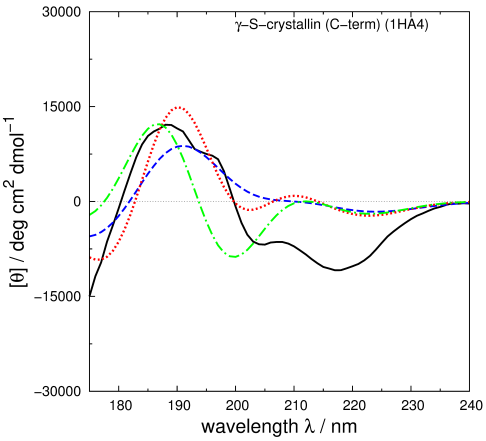

Triosephosphate isomerase (7TIM)
[θ] / deg cm^2 dmol^{-1}
wavelength λ / nm

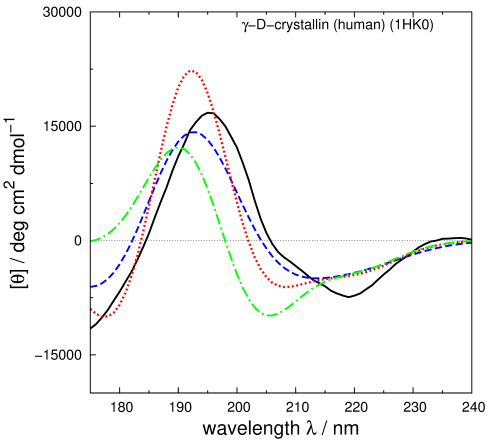
β-I proteins

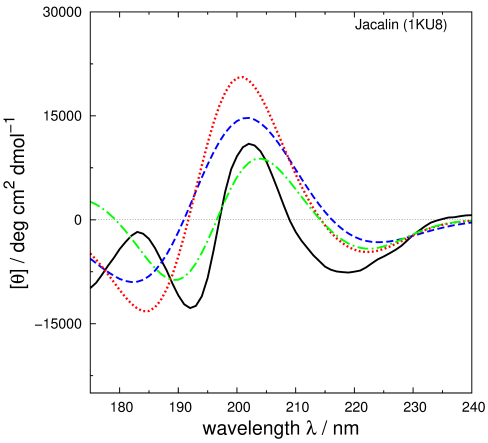
Calculated CD spectra of 14 β-I proteins are compared with their experimental spectra. The features of β-I proteins are a positive peak near 195 nm and a negative peak with comparable magnitude at around 216 nm. The vibronic calculations generated more detailed bands than non-vibronic between 200 and 240 nm. However, the quality of the calculations with both parameters is not in a general fashion, such as the peak at 196 nm in pectate lyase C is predicted red-shifted by non-vibronic calculation but blue-shifted by the vibronic one with similar amount. This positive peak was also, in some cases, overestimated by one parameter but underestimated by the other (e.g. β-bactoglobulin, γ-D-crystallin and pea lectin).

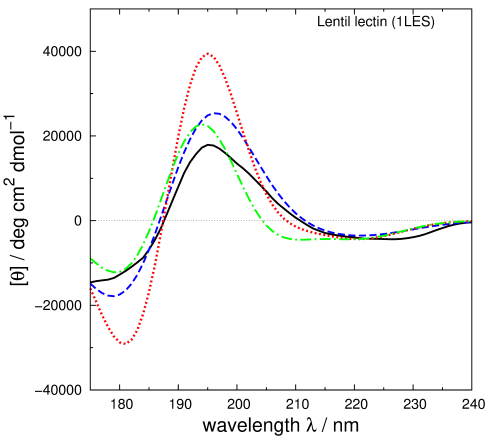


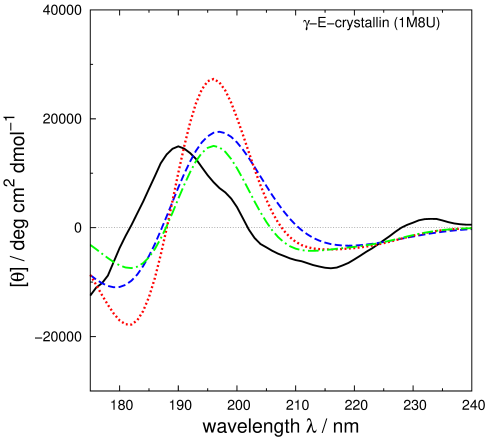

γ–D–crystallin (1ELP)
[θ] / deg cm^2 dmol^{-1}
wavelength λ / nm

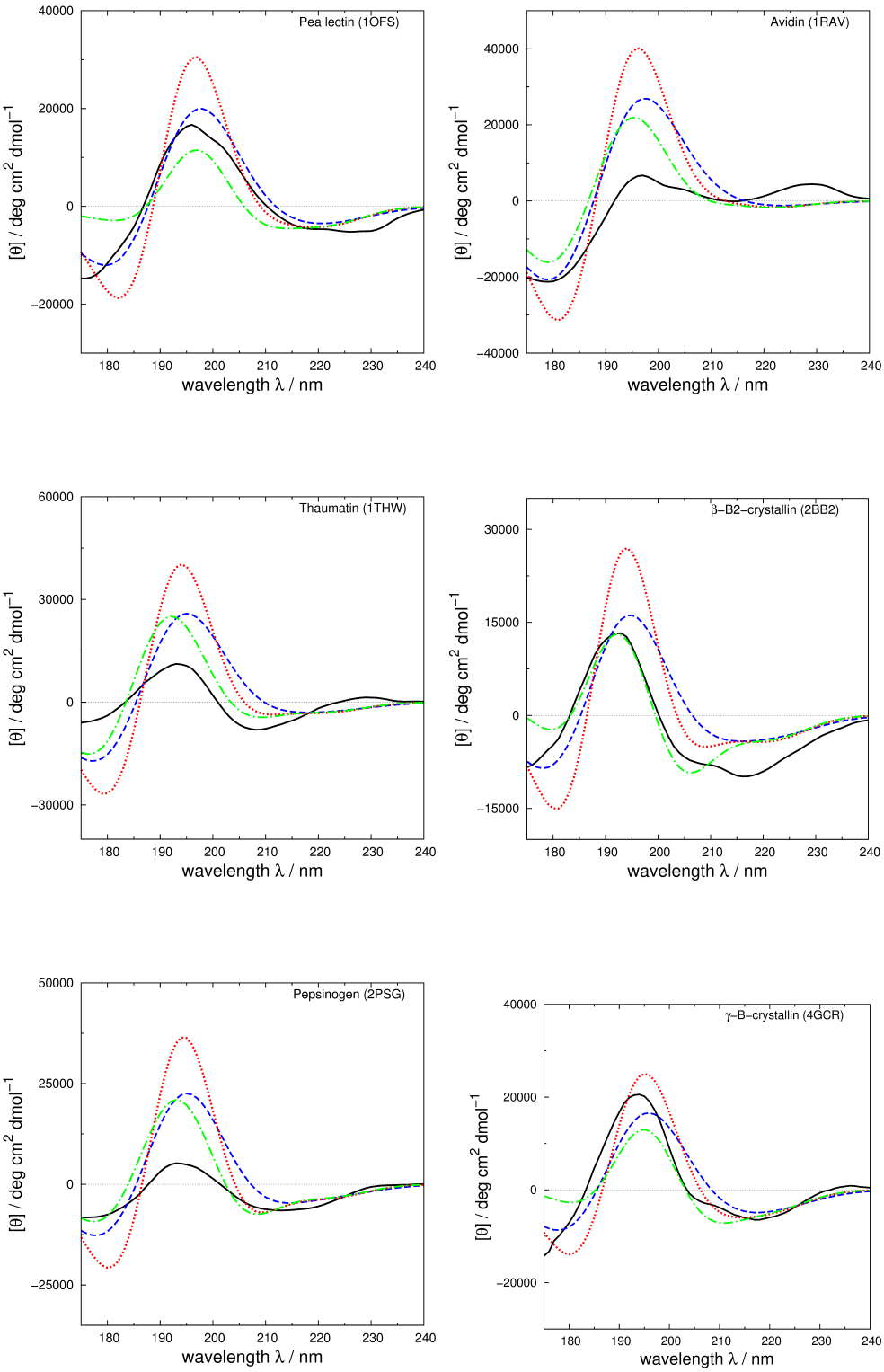

γ–S–crystallin (C–term) (1HA4)
[θ] / deg cm^2 dmol^{-1}
wavelength λ / nm


γ–D–crystallin (human) (1HK0)
[θ] / deg cm^2 dmol^{-1}
wavelength λ / nm


Jacalin (1KU8)
[θ] / deg cm^2 dmol^{-1}
wavelength λ / nm

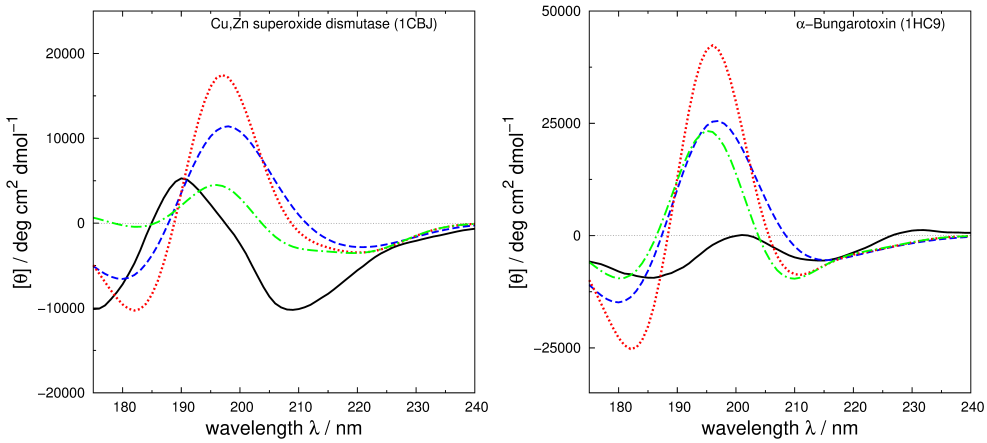

Lentil lectin (1LES)
[θ] / deg cm^2 dmol^{-1}
wavelength λ / nm

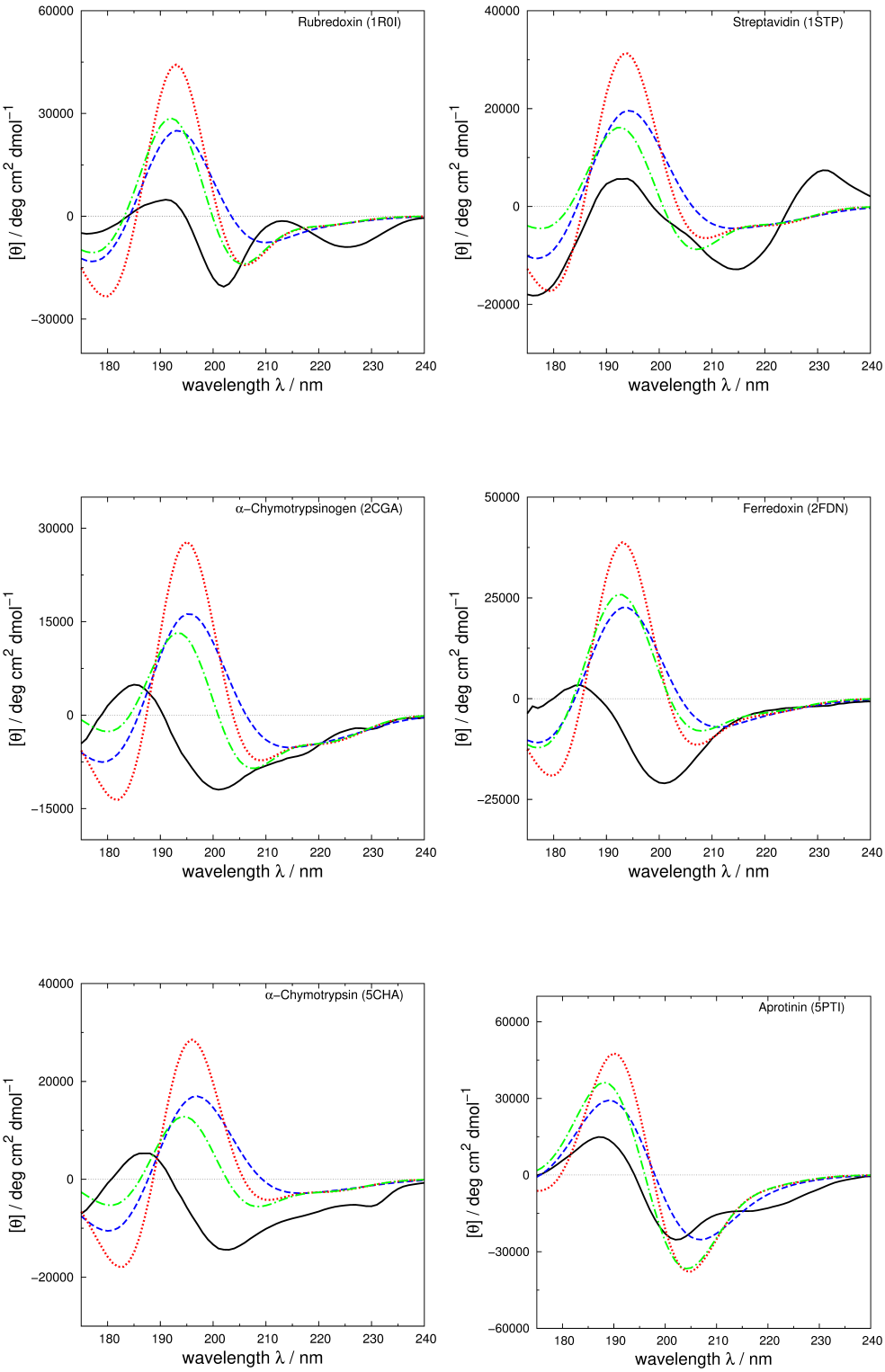

γ–E–crystallin (1M8U)
[θ] / deg cm^2 dmol^{-1}
wavelength λ / nm


Pea lectin (1OFS)
Avidin (1RAV)
Thaumatin (1THW)
β–B2–crystallin (2BB2)
Pepsinogen (2PSG)
γ–B–crystallin (4GCR)
[θ] / deg cm2 dmol–1
wavelength λ / nm

β-II proteins

Calculated CD spectra of eight β-II proteins are compared with their experimental spectra. The structures of β-II proteins are less regular with more conformational flexibility, whose features are a negative peak at about 200 nm and a positive band ranging from 185 to 190 nm. The spectra of these proteins were not reproduce well even with the vibronic structures under consideration. However, there are indeed noticeable improvements in the correlation of the calculated and experimental intensities as stated in chapter 3.




Rubredoxin (1R0I)
Streptavidin (1STP)
α–Chymotrypsinogen (2CGA)
Ferredoxin (2FDN)
α–Chymotrypsin (5CHA)
Aprotinin (5PTI)
[θ] / deg cm2 dmol–1
wavelength λ / nm

Appendix C

Experimental and Calculated Protein Near-UV CD Spectra

Table C.1 Comparison of the RMSE, MRE and Spearman rank correlation between the near-UV CD intensity calculated with different parameters and the experimental intensity.

Protein	**PDB entry**	**RMSE/cm^{-1}**		**MRE**		**Spearman correlation**	
		non-vib	*vib*	*non-vib*	*vib*	*non-vib*	*vib*
Acetylcholines-terase	2ACE	199	80	1.69	1.19	0.60	0.55
Adenylate kinase	2ECK	34	11	0.98	0.36	0.62	0.89
α-Lactalbumin	1A4V	195	144	3.00	1.39	0.27	0.20
α-Toxin	1QM6	197	92	5.36	6.98	-0.50	0.30
Apolipophorin III	1AEP	107	46	9.07	1.46	-0.59	0.71
Barnase	1A2P	173	121	1.30	1.07	-0.15	0.83
β-2 microglo-bulin (human)	1LDS	147	96	4.54	6.74	0.09	0.14
β-Lactamase	1BTL	137	76	0.94	0.59	0.28	0.93
P.69 pertactin	1DAB	45	23	9.68	2.50	-0.00	0.55
BPTI	5PTI	235	128	0.87	0.52	0.93	0.66
Calmodulin	4CLN	22	8	0.88	0.44	0.63	0.93
Cardiotoxin	4OM4	105	98	1.26	1.33	-0.71	-0.36
Chymotrypsin-ogen A	2CGA	352	169	3.06	9.54	-0.43	-0.03
Dehydroquinase II	2BT4	60	41	4.40	7.47	-0.74	0.12
Dihydrofolate reductase	4P3Q	62	40	14.28	6.40	0.67	0.63
Glucose oxidase	1CF3	68	65	0.79	0.67	0.75	0.65
Hen egg white lysozyme	1HF4	246	104	6.54	1.08	-0.04	0.90
HCAII	2CBA	307	75	2.40	0.63	0.17	0.77

Human serum albumin	1AO6	101	90	1.03	1.09	-0.60	0.44
Insulin	5ENA	114	56	0.84	0.49	0.97	0.94
Interleukin 4 (cytokine)	2B8U	64	46	4.51	1.50	-0.20	-0.43
Interleukin 6	1ALU	40	34	1.46	0.97	-0.66	-0.38
Monellin	1IV9	83	61	0.96	0.87	0.70	0.41
Myoglobin (whale)	1UFP	52	48	38.55	5.15	0.75	-0.80
Neocarzinostatin	1NOA	126	102	0.85	0.71	0.79	0.81
Odorant binding protein	1A3Y	74	40	0.68	0.49	0.70	0.91
Papain	3LFY	254	77	2.45	2.40	-0.06	0.89
Pectate lyase C	2PEC	111	26	5.74	1.20	-0.47	0.79
Phosphatidylet-hanolamine-binding protein	1A44	106	70	1.01	0.74	0.01	0.78
Phospholipase A2 (Ca^{2+})	1PSJ	279	102	52.18	4.44	0.78	0.87
Relaxin	6RLX	244	166	1.45	1.07	0.05	0.82
Rhodanese	1DP2	243	71	8.68	8.07	0.61	-0.10
Ribonuclease T1	1RN1	93	74	3.44	1.30	0.57	0.73
RNase A	1AFU	123	49	0.88	0.50	0.75	0.95
SNase	1STN	59	78	1.08	1.64	-0.14	-0.39
Extracellular domain of human tissue factor	2HFT	118	77	1.10	0.77	0.04	0.66
Sticholysin II	1O72	70	29	5.13	1.01	-0.18	0.83
Subtilisin BPN’	1ST2	78	28	0.77	0.55	0.71	0.89
Thioredoxin	2TRX	104	90	0.83	1.00	0.40	0.63
Tryptophan synthase α-subunit	1WQ5	31	27	5.89	10.68	-0.44	-0.37

Experimental and calculated spectra

Experimental and calculated CD spectra for protein set (Table 4.1) in the near-UV: bold solid line, experiment; dotted line, '*non-vib*' calculation with crystal structure; thin solid line, '*vib*' calculation with crystal structure; dotted-dash line, '*vib*' calculation with NMR structures. Bandwidth of calculations is 4 nm.

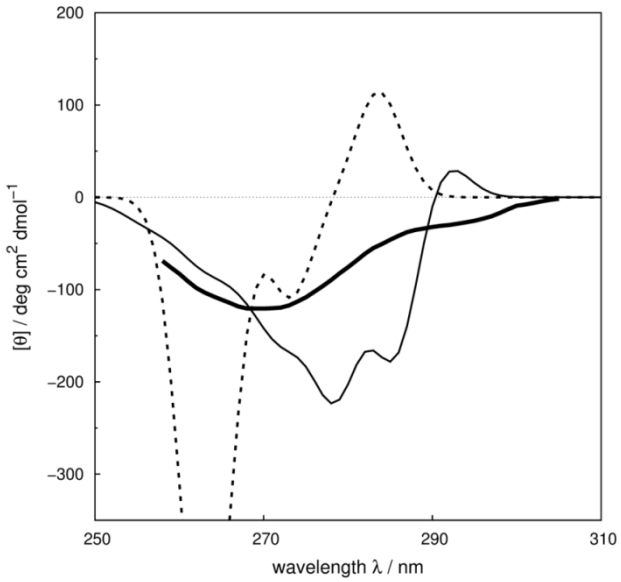


Figure C.1 Spectra of acetylcholinesterase (PDB code: 2ACE). There is noticeable qualitative and quantitative (RMSE drops from 199 to 80 cm^{-1}) improvement with the new parameters.

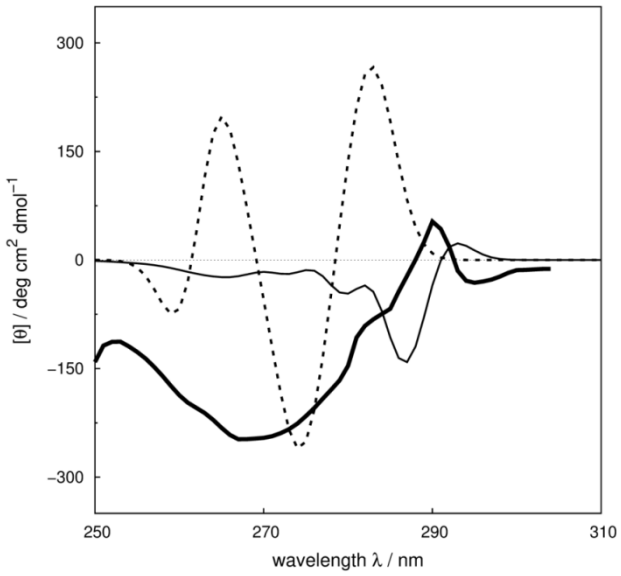


Figure C.2 Spectra of α-lactalbumin (PDB code: 1A4V). There is modest improvement (RMSE drops from 195 to 144 cm^{-1}), but quantitative agreement is still lacking.

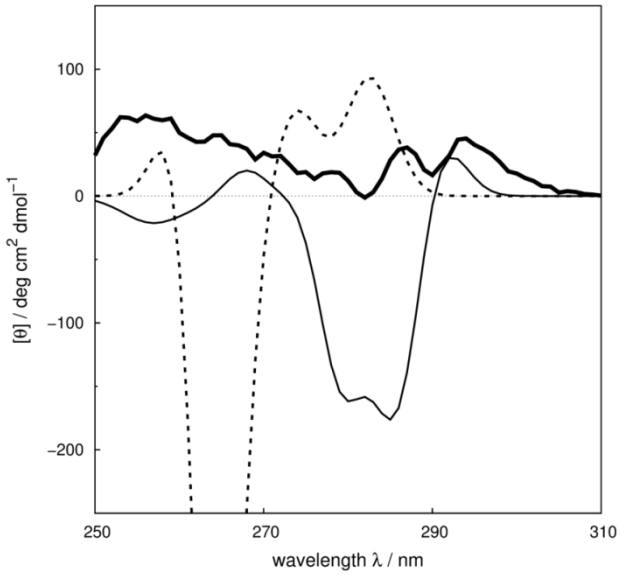


Figure C.3 Spectra of α-toxin (PDB code: 1QM6). There is modest improvement (RMSE drops from 197 to 92 cm^{-1}), but quantitative agreement is still lacking and the calculations wrongly predict an intense negative peak between 275 and 285 nm.

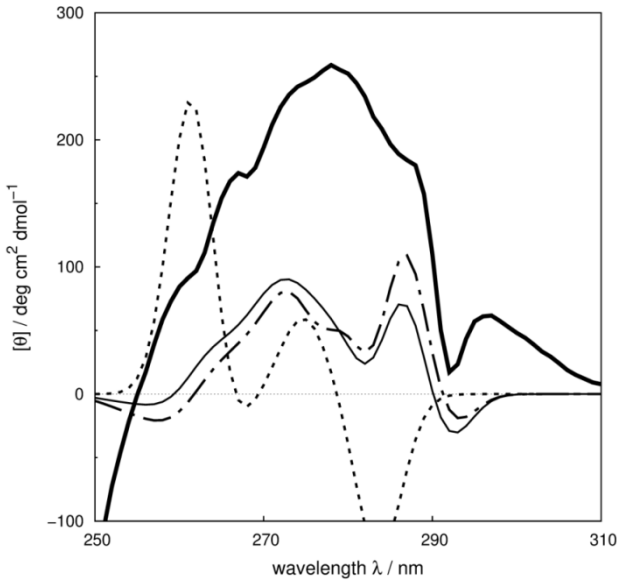


Figure C.4 Spectra of barnase (PDB code: 1A2P and 1FW7). The new parameters give a qualitatively better computed spectrum, with either the X-ray or NMR structures. The band structure agrees well with experiment, but the intensity is under-estimated.

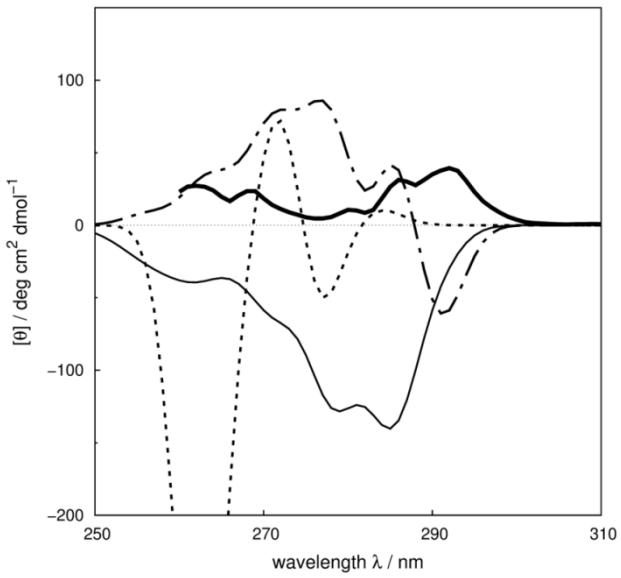


Figure C.5 Spectra of β-2 microglobulin (PDB code: 1LDS and 1JNJ). The calculated spectrum using the new parameters and the NMR structures gives a noticeable improvement with experiment, but full quantitative agreement is still lacking.

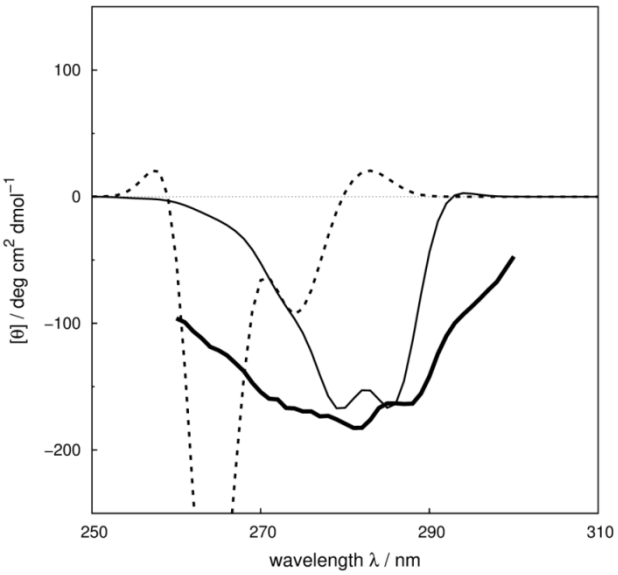


Figure C.6 Spectra of β-lactamase (PDB code: 1BTL). There is noticeable qualitative (MRE drops from 0.94 to 0.59) and quantitative (RMSE drops from 137 to 76 cm^{-1}) improvement with the new parameters.

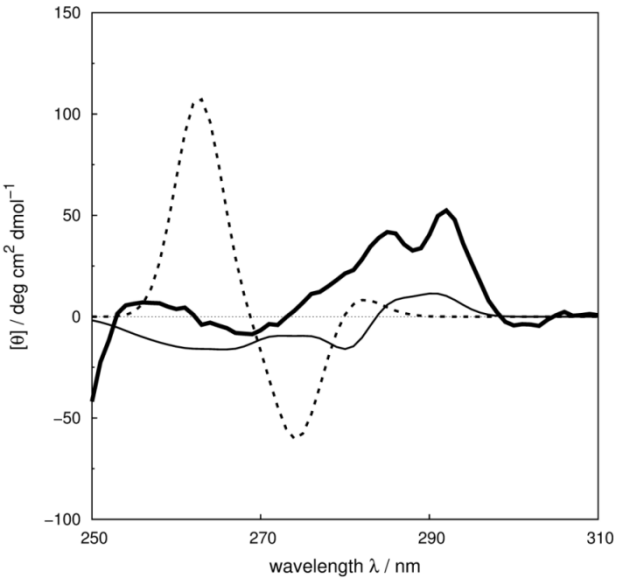


Figure C.7 Spectra of P. 69 pertactin (PDB code: 1DAB). The new parameters improve the calculation of what is quite a weak near-UV CD spectrum.

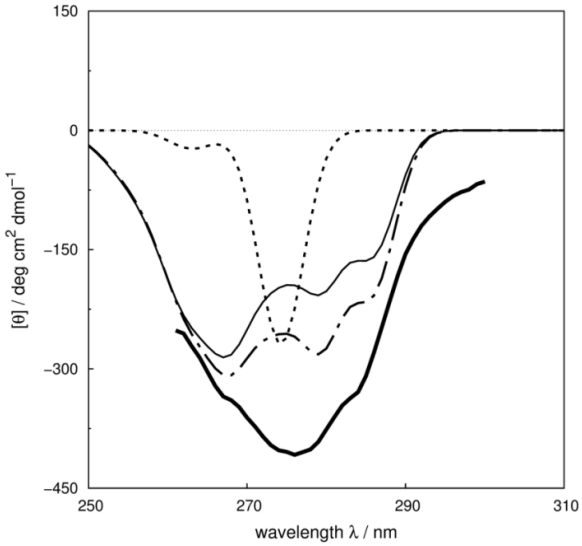


Figure C.8 Spectra of BPTI (PDB code: 5PTI and 1PIT). The new parameters improve the calculated spectrum (MRE drops from 0.87 to 0.52, and the RMSE drops from 235 to 128 cm^{-1}). The use of the NMR structures gives some further noticeable improvement. The band structure is reproduced well, although the intensity is under-estimated.

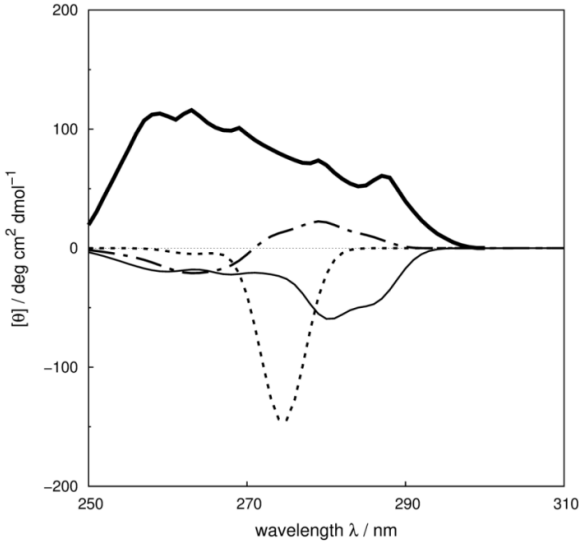


Figure C.9 Spectra of cardiotoxin (PDB code: 4OM4 and 1CRF). The calculation with the NMR structure predicts a positive band between 270 and 290 nm correctly which is predicted to be negative using the X-ray structure.

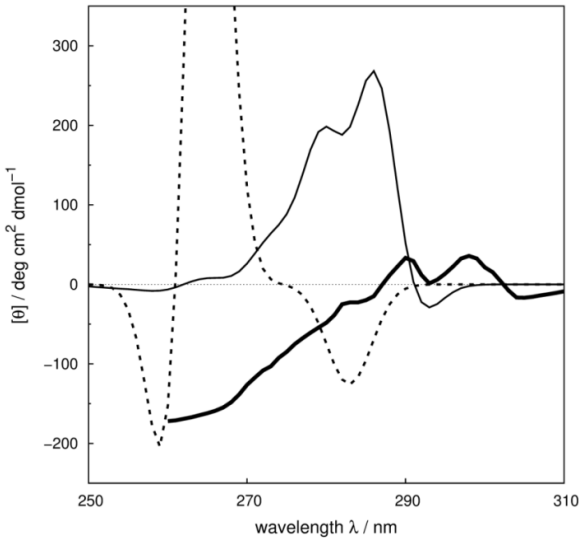


Figure C.10 Spectra of chymotrypsinogen A (PDB code: 2CGA). The calculated spectrum using the new parameter shows noticeable quantitative (RMSE drops from 352 to 169 cm^{-1}) improvement. The band structure agrees with the experiment, but the calculated spectrum is blue-shifted and the intensity is over-estimated.

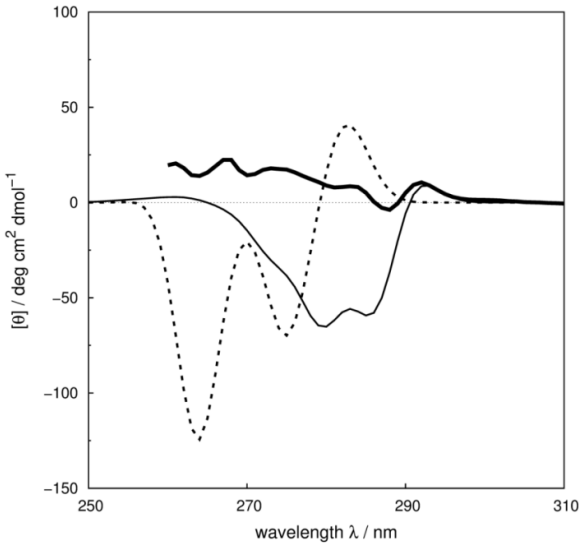


Figure C.11 Spectra of dehydroquinase II (PDB code: 2BT4). The calculation predicts negative bands between 265 and 290 nm while the experiment shows weak positive bands. However, the prediction of the band structure shows noticeable improvement (Spearman correlation increases from -0.74 to 0.12) with the new parameters.

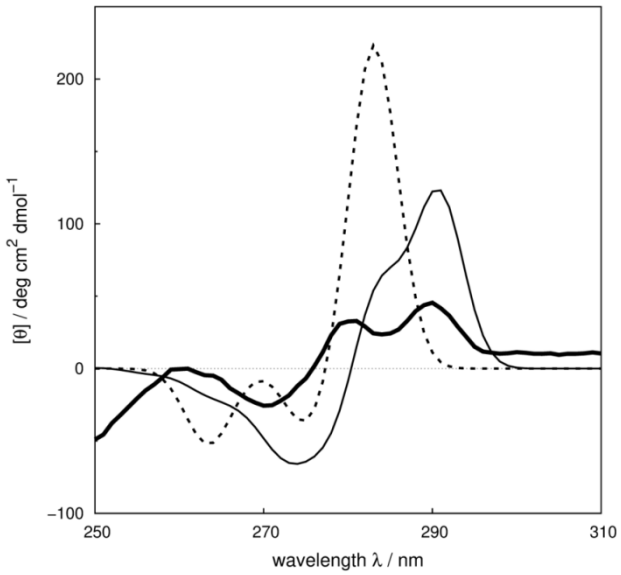


Figure C.12 Spectra of dihydrofolate reductase (PDB code: 4P3Q). The calculated spectrum with new parameters shows comparable statistical numbers with more detailed band structure.

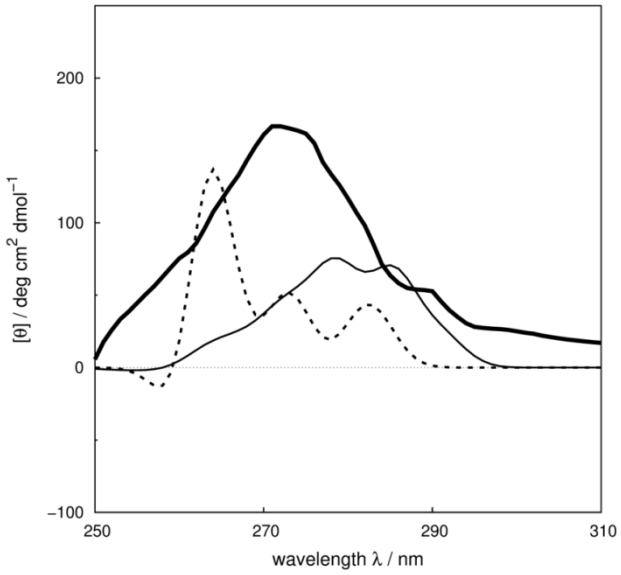


Figure C.13 Spectra of glucose oxidase (PDB code: 1CF3). The calculated spectra with both parameters show comparable statistical numbers. Calculation with the new parameters gives better prediction above 270 nm while the non-vibrational parameters show better results under 270 nm.

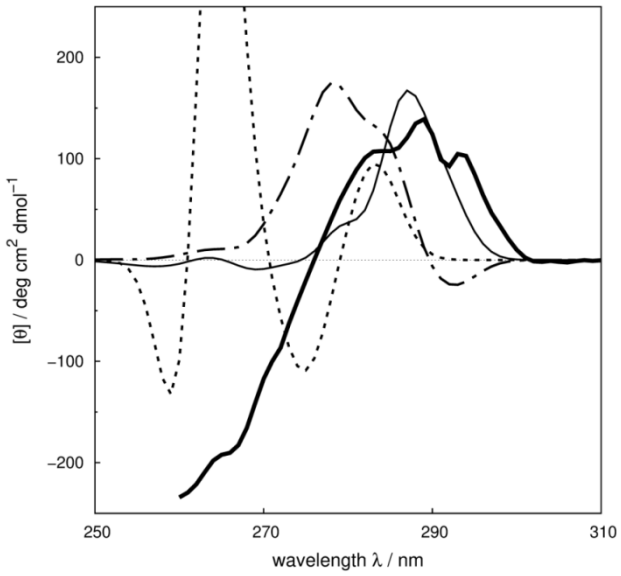


Figure C.14 Spectra of hen egg white lysozyme (PDB code: 1HF4 and 1E8L). The new parameters give a qualitatively better computed spectrum, with either the X-ray or NMR structures. The band structure agrees well with experiment with a comparable intensity.

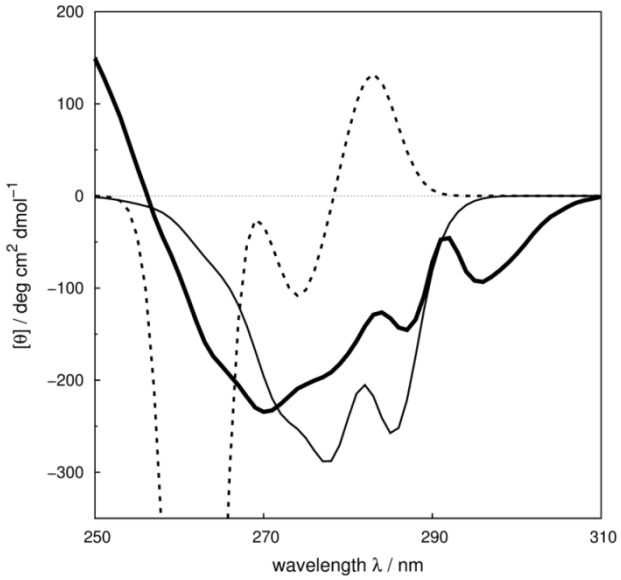


Figure C.15 Spectra of HCAII (PDB code: 2CBA). There is noticeable qualitative and quantitative (RMSE drops from 307 to 75 cm^{-1}) improvement with the new parameters.

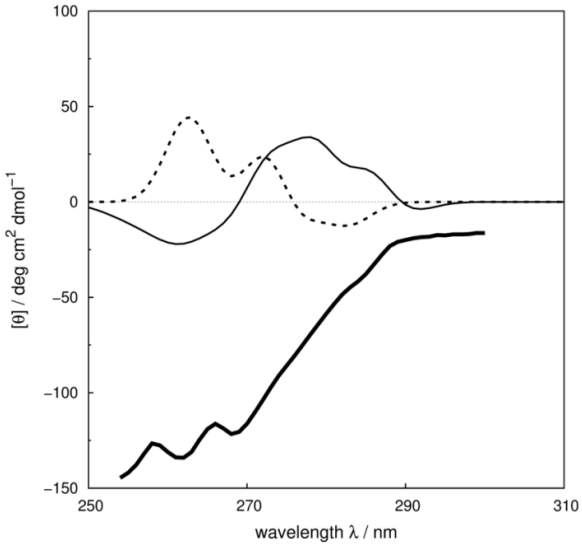


Figure C.16 Spectra of human serum albumin (PDB code: 1AO6). There is still lacking quantitative agreement for this protein. The calculated spectrum with new parameters shows better Spearman correlation due to the prediction of the negative band below 270 nm.

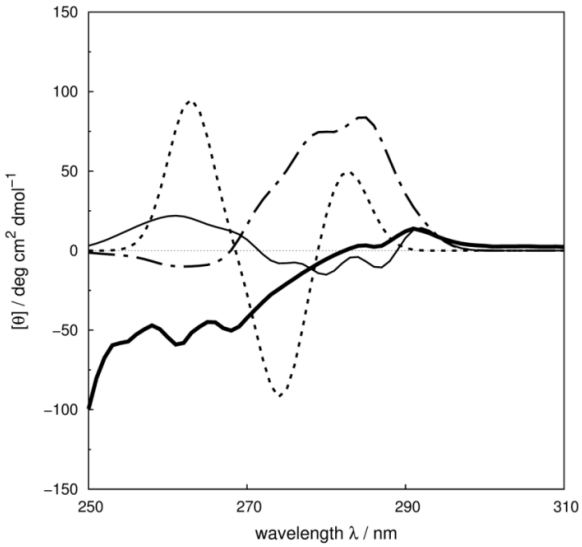


Figure C.17 Spectra of interleukin 4 (PDB code: 2B8U and 1CYL). There is modest improvement (RMSE drops from 64 to 46 cm^{-1}), but quantitative agreement is still lacking and the calculations wrongly predict an positive peak between 250 and 275 nm. The calculated spectrum with the NMR structures shows better prediction of the sign of the spectrum with over-estimated intensities.

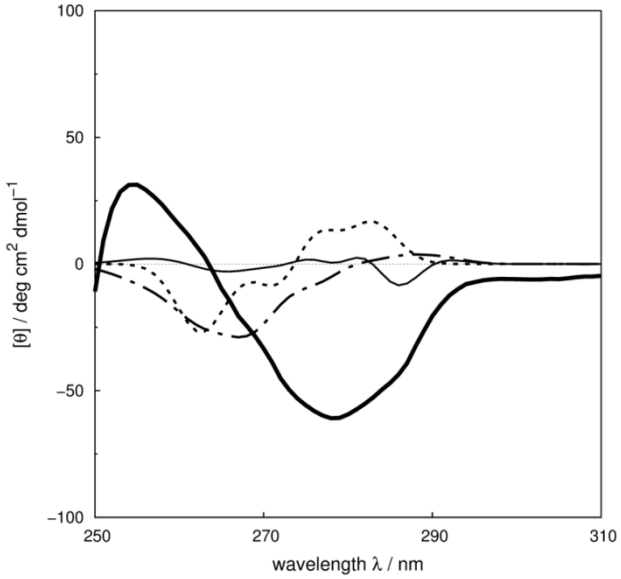


Figure C.18 Spectra of interleukin 6 (PDB code: 1ALU and 2IL6). The quantitative agreement is still lacking for this protein calculated with either X-ray or NMR structures.

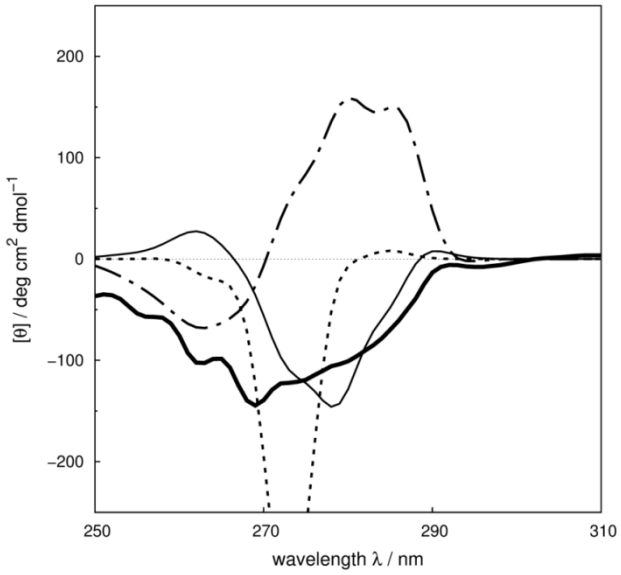


Figure C.19 Spectra of monellin (PDB code: 1IV9 and 1MNL). Calculation with the new parameters shows modest improvement between 265 and 290 nm, but quantitative agreement is still lacking. Calculation with the NMR structures show well reproduced band below 270 nm and wrongly predicted band over 270 nm.

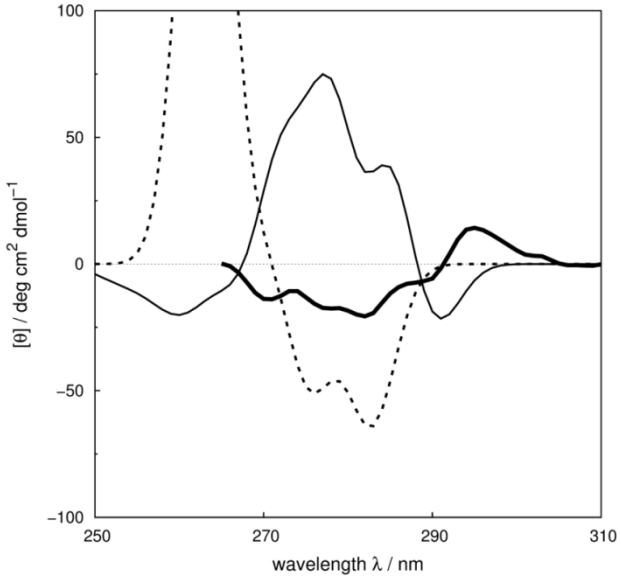


Figure C.20 Spectra of whale myoglobin (PDB code: 1UFP). The quantitative agreement is still lacking for this weak near-UV CD spectrum.

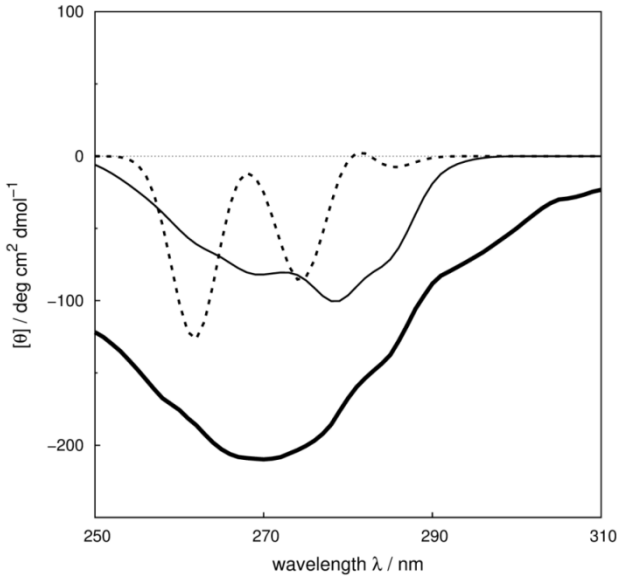


Figure C.21 Spectra of neocarzinostatin (PDB code: 1NOA). The new parameters improve the calculated spectrum with a better band structure.

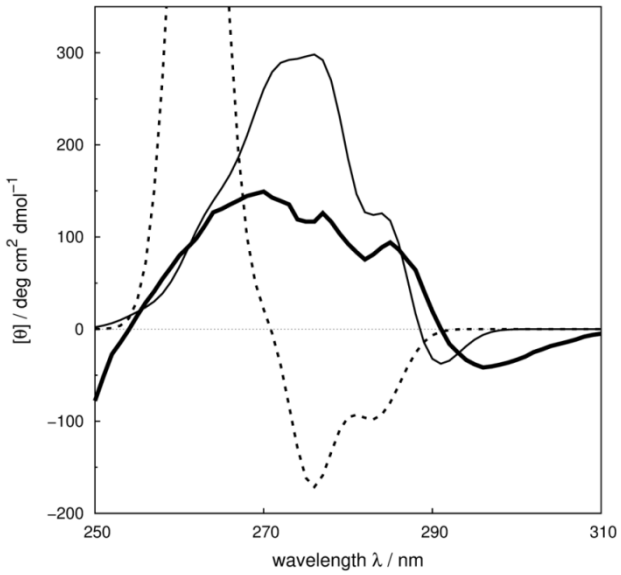


Figure C.22 Spectra of papain (PDB code: 3LFY). There is noticeable qualitative and quantitative (RMSE drops from 254 to 77 cm^{-1}) improvement with the new parameters. The band structure is well reproduced with the Spearman correlation 0.89.

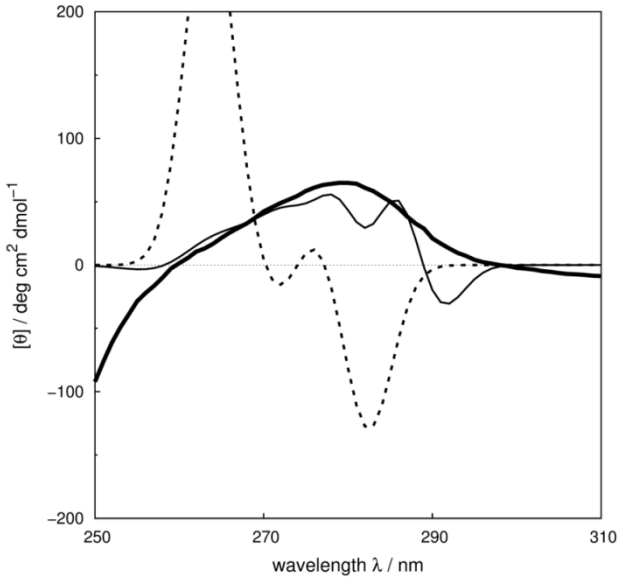


Figure C.23 Spectra of peptate lyase C (PDB code: 2PEC). There is noticeable qualitative (MRE drops from 5.74 to 1.20) and quantitative (RMSE drops from 111 to 26 cm^{-1}) improvement with the new parameters.

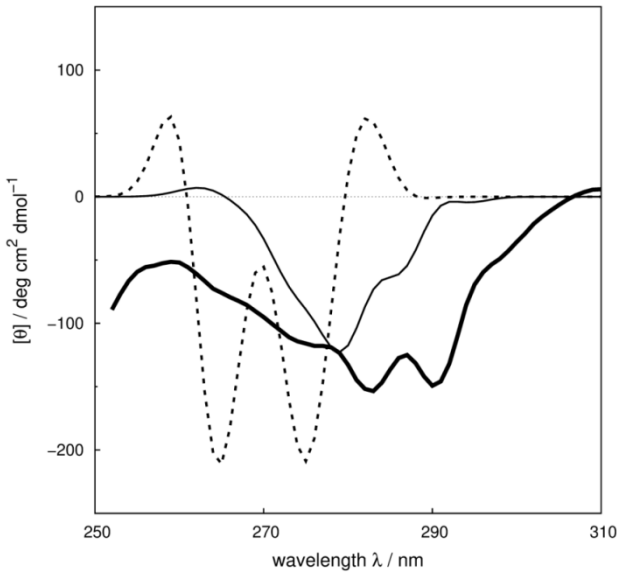


Figure C.24 Spectra of phosphatidylethanolamine-binding protein (PDB code: 1A44). There is modest quantitative improvement (RMSE drops from 106 to 70 cm^{-1}) with a better reproduced band structure (Spearman correlation increases from 0.01 to 0.78).

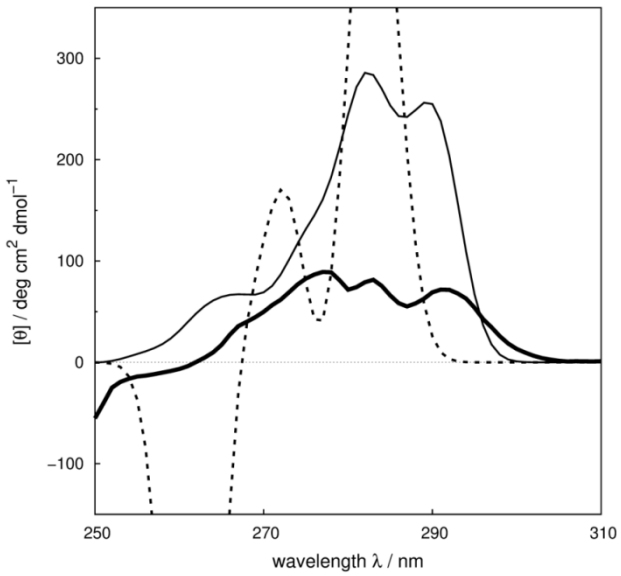


Figure C.25 Spectra of phospholipase A2 (Ca^{2+}) (PDB code: 1PSJ). The new parameters give a qualitatively and quantitative better computed spectrum. The band structure agrees well with experiment, but the intensity is over-estimated.

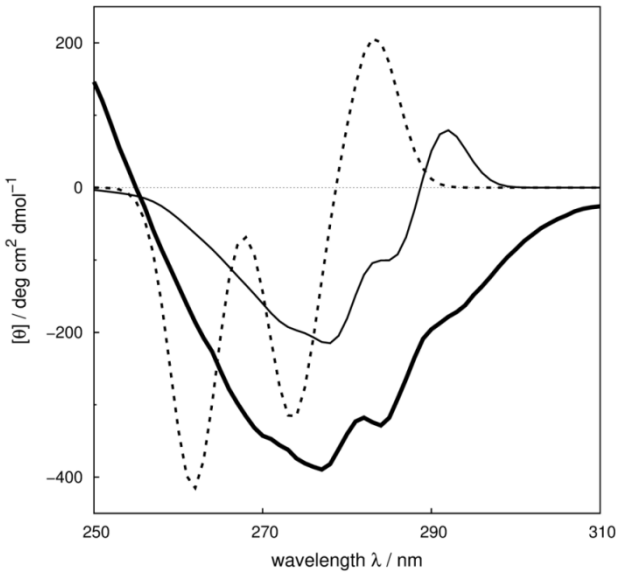


Figure C.26 Spectra of relaxin (PDB code: 6RLX). There is noticeable quantitative (RMSE drops from 244 to 166 cm^{-1}) improvement with the new parameters. The band structure is well reproduced (Spearman correlation increases from 0.05 to 0.82) with the new parameters.

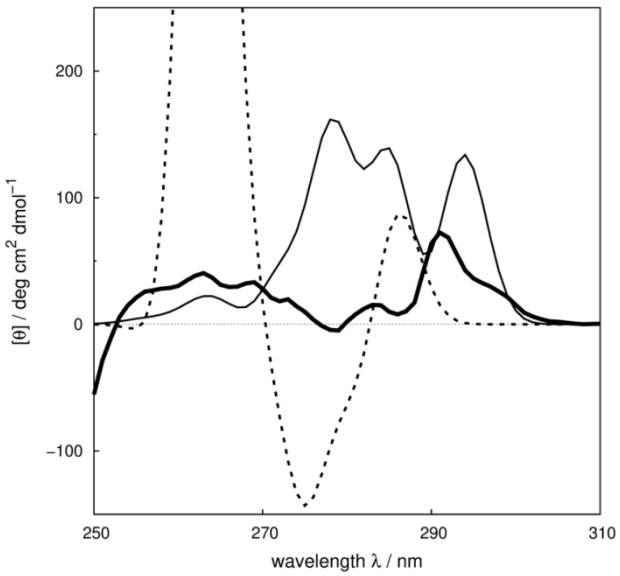


Figure C.27 Spectra of rhodanese (PDB code: 1DP2). There is modest improvement (RMSE drops from 243 to 71 cm^{-1}), but quantitative agreement is still lacking.

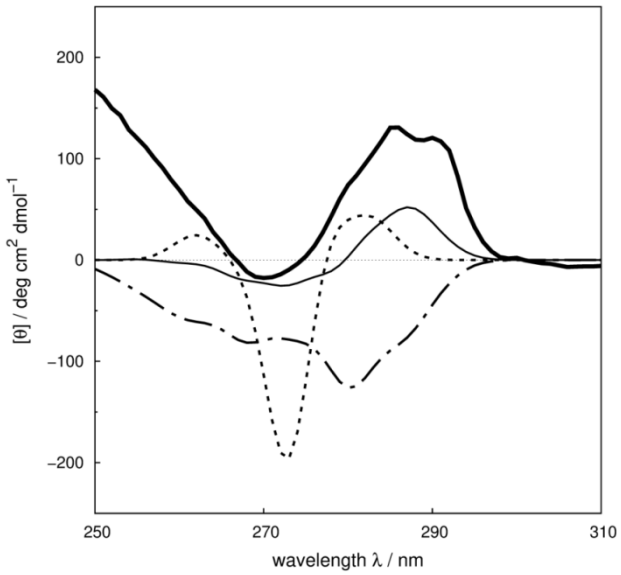


Figure C.28 Spectra of ribonuclease T1 (PDB code: 1RN1 and 1YGW). The calculated spectrum with the new parameters shows noticeable improvement between 270 and 300 nm using X-ray structure while the calculation with the NMR structures wrongly predict a negative band.

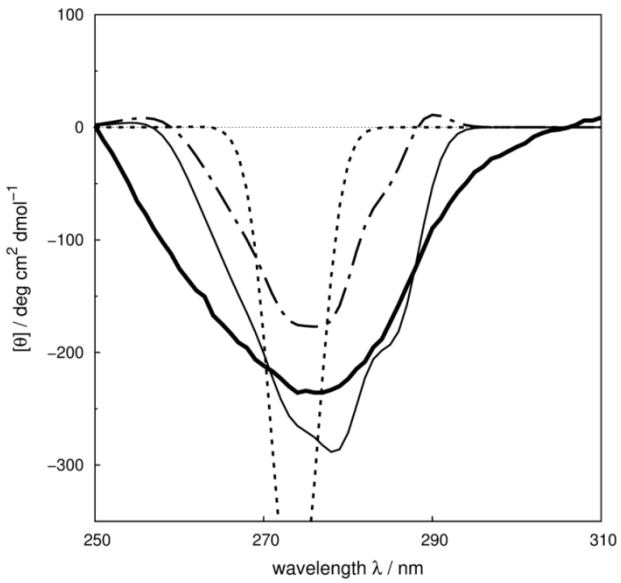


Figure C.29 Spectra of RNase A (PDB code: 1AFU and 2AAS). There is modest quantitative (RMSE drops from 123 to 49 cm^{-1}) improvement with the new parameters. Calculation with the NMR structures predicted a less intense spectrum.

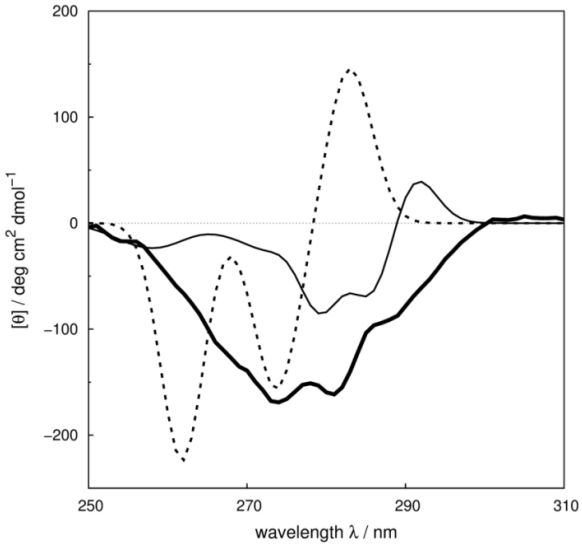


Figure C.30 Spectra of extracellular domain of human tissue factor (PDB code: 2HFT). There is noticeable quantitative (RMSE drops from 118 to 77 cm^{-1}) improvement with the new parameters. The band structure is well reproduced (Spearman correlation increases from 0.04 to 0.66) with the new parameters.

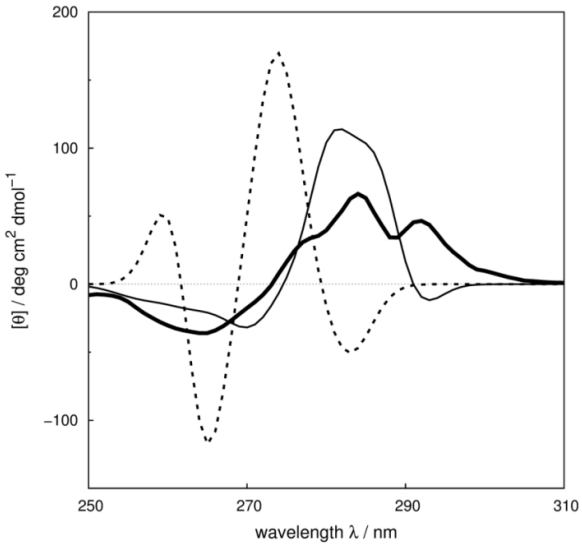


Figure C.31 Spectra of sticholysin II (PDB code: 1O72). There is noticeable qualitative (MRE drops from 5.13 to 1.01) and quantitative (RMSE drops from 70 to 29 cm^{-1}) improvement with the new parameters. The band structure is also better reproduced (Spearman correlation increases from -0.18 to 0.83).

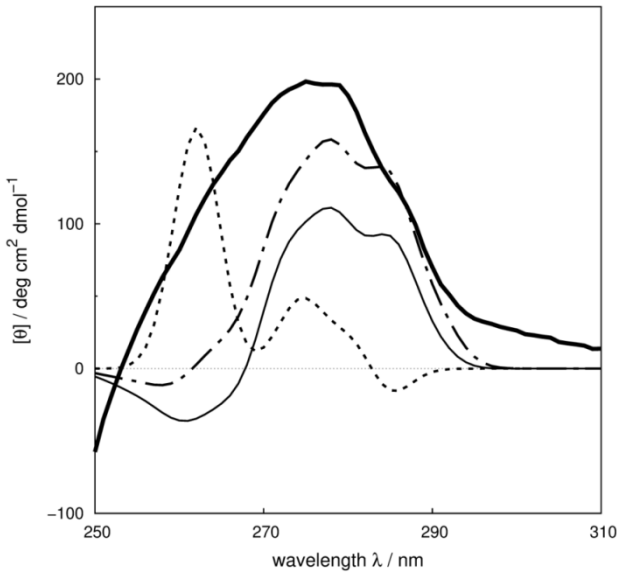


Figure C.32 Spectra of thioredoxin (PDB code: 2TRX and 1XOA). The new parameters improve the calculated spectrum (Spearman correlation increases from 0.40 to 0.63, and he RMSE drops from 104 to 90 cm^{-1}). The use of the NMR structures gives some further noticeable improvement.

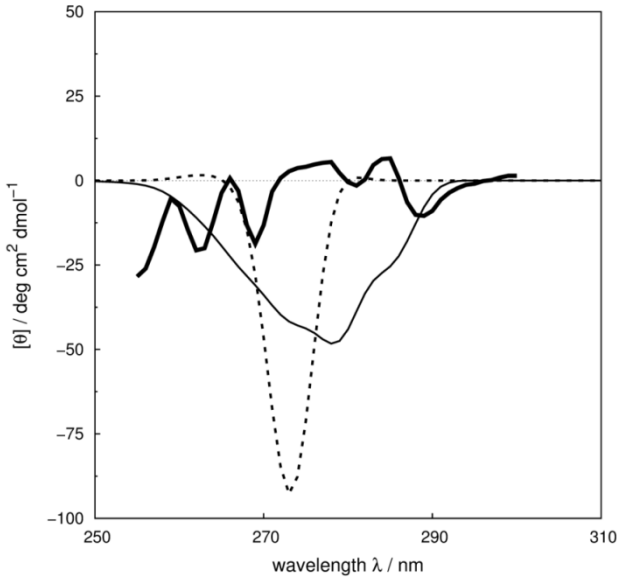


Figure C.33 Spectra of tryptophan synthase α-subunit (PDB code: 1WQ5). The quantitative agreement is still lacking for this protein. Both parameters predict negative bands while the experiment shows both negative and positive bands.

The X-ray structures we used in our calculation include 16 proteins that contain multiple chains with identical sequences. All chains are considered in the calculation; this is necessary if the chains (despite identical sequence) do not share similar conformations or there are interactions between aromatic residues from different chains. Chains in the same protein with the same sequence may have distinguishable conformational differences arising from a specific residue or motif which may lead to various optical features in their near-UV CD spectra of individual chains. Calculation with one chain instead of the whole protein may not be sufficient to reproduce the experimental intensities which may be affected by cancellation or enhancement from multiple chains at certain wavelength. Proteins containing multiple chains can be grouped into four cases (Table C.2). The first kind (Group 1, including adenylate kinase, human serum albumin, papain, RNase A, thioredoxin and ribonuclease T1) has low all-atom RMSD between chains. There are no interactions between aromatic residues on different chains for these proteins and the Spearman rank correlations of calculations using all chains or considering only single chain are comparable. Two proteins (Group 2), α-toxin and dehydroquinase II, have similar chain conformations with aromatic rings adjacent to each other from different chains. There are no aromatic interactions between chains for proteins in Group 3. However, ^{146}Tyr in chymotrypsinogen A, ^{62}Trp in hen egg white lysozyme and ^{17}Tyr in barnase are not superimposable among chains.

Another five proteins (Group 4) have both aromatic interactions between chains and residues or motifs that cannot superimpose among chains, such as ^{52}Phe in monellin, ^{10}Phe in odorant binding protein, ^{10}Trp and ^{62}Phe in sticholysin II, ^{72}Phe and ^{102}Tyr in tryptophan synthase α-subunit and ^{10}Phe in cardiotoxin. Comparison of the Spearman rank correlations (Table C.2) between calculations with each single chain or the whole protein and their experimental spectra confirm the necessity of calculation with multiple chains of proteins.

Table C.2 Spearman correlations of calculated spectra using vibrational parameters compared to their experiment spectra for all chains or individual chains of multimeric proteins.

Protein		No. of chain	Spearman rank correlation				all-atom RMSD[a]		aromatic interaction between chains[b]	Group
			All	Chain A	Chain B	Chain C				
Adenylate kinase	2ECK	2	0.89	0.89	0.89		0.05		0	1
α-Toxin	1QM6	2	0.3	0.32	0.32		0.11		Y65-Y65	2
Chymotrypsinogen A	2CGA	2	-0.03	-0.08	0.10		0.70		0	3
Hen egg white lysozyme	1HF4	2	0.9	0.76	0.91		0.52		0	3
Human serum albumin	1AO6	2	0.44	0.49	0.33		0.28		0	1
Monellin	1IV9	2	0.41	0.59	0.44		0.63		W3-Y163	4
Odorant binding protein	1A3Y	2	0.91	0.76	0.90		0.52		F66-Y92;F56-F10	4
Papain	3LFY	2	0.89	0.91	0.90		0.26		0	1
RNase A	1AFU	2	0.95	0.96	0.94		0.46		0	1
Sticholysin II	1O72	2	0.83	0.81	0.63		0.46		Y91-F62[b]	4
Thioredoxin	2TRX	2	0.63	0.67	0.62		0.66		0	1
Tryptophan synthase α-subunit	1WQ5	2	-0.37	-0.74	-0.57		0.77		F72-Y102;F72-F22	4
Barnase	1A2P	3	0.83	0.15	0.86	0.80	0.23	0.31	0	3
Ribonuclease T1	1RN1	3	0.73	0.75	0.65	0.70	0.41	0.41	0	1
Cardiotoxin	4OM4	5	-0.36	Chain A	Chain B	Chain C			F10(C)-F10(E)	4

				-0.47	-0.72	-0.29	0.63		0.39		
					Chain D	Chain E					
					0.34	-0.49	0.76		0.70		
Dehydroquinase II	2BT4	12	0.12	Chain A	Chain B	Chain C				F116(A)-Y321(B); F121(A)-F516(C); F140(A)-F740(D); F316(B)-Y521(C); F340(B)-F1340(G); F540(C)-F1940(J); F716(D)-Y921(E); F721(D)-F1116(F); F916(E)-Y1121(F); F940(E)-F2340(L); F1140(F)-F1540(H); F1316(G)-Y1521(H); Y1321(G)-F1716(I); F1516(H)-Y1721(I); F1740(I)-F2140(K); F1916(J)-Y2121(K); Y1921(J)-F2316(L); F2116(K)-Y2321(L)	2
				0.01	-0.07	-0.03		0.15	0.12		
				Chain D	Chain E	Chain F					
				-0.20	-0.10	-0.11	0.13	0.18	0.08		
				Chain G	Chain H	Chain I					
				-0.15	-0.05	-0.19	0.09	0.12	0.13		
				Chain J	Chain K	Chain L					
				-0.13	-0.18	0.01	0.10	0.08	0.08		

a. all-atom RMSD compares chains with the first chain in the PDB file. b. Aromatic groups within 8 Å are considered to be interacted.

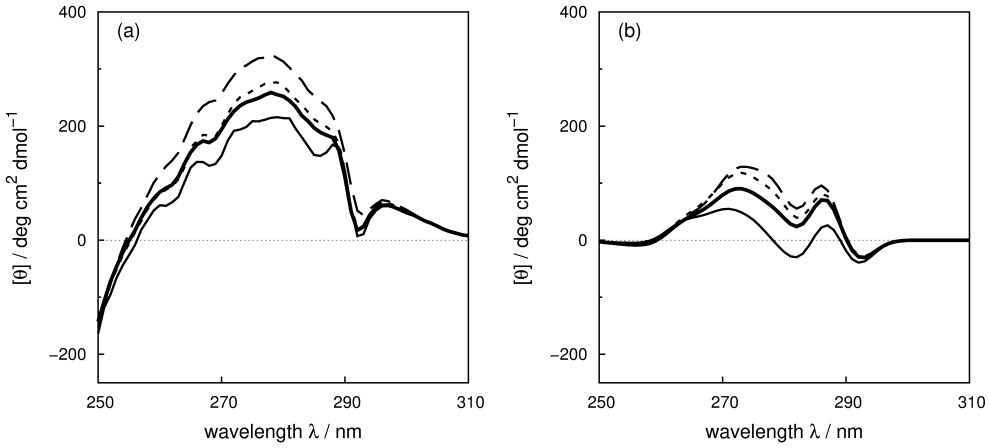


Figure C.34 CD spectra of wild-type barnase (bold solid line) and mutants Y78F (solid line), Y90F (dotted line) and Y97F (dashed line); (a) experimental spectra and (b) calculated spectra.

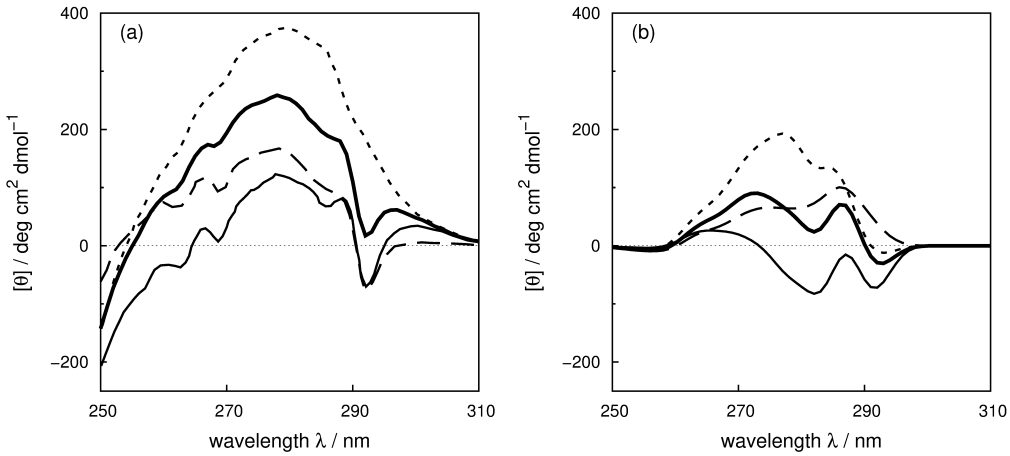


Figure C.35 CD spectra of wild-type barnase (bold solid line) and mutants W35F (solid line), W71F (dotted line) and W94F (dashed line); (a) experimental spectra and (b) calculated spectra.

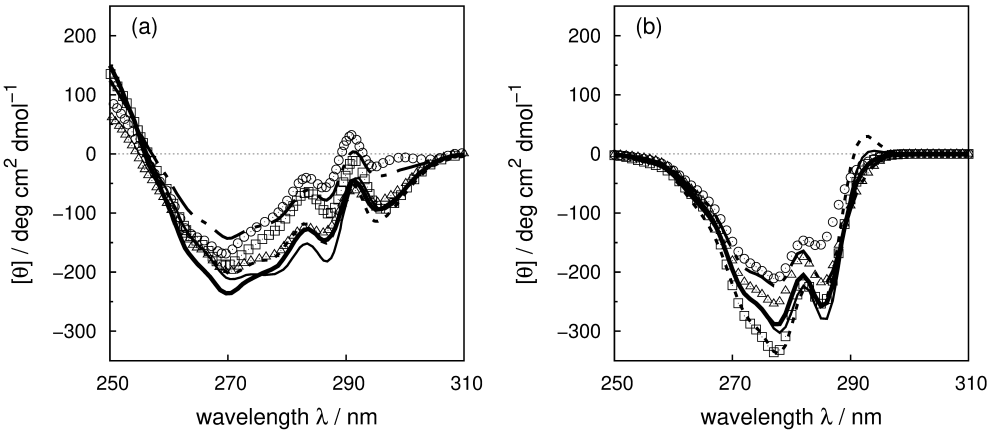


Figure C.36 CD spectra of wild-type HCAII (bold solid line) and mutants W5F (solid line), W16F (square), W97C (circle), W123C (triangle), W209F (dotted line) and W245C (dot-dashed line); (a) experiment and (b) calculation.