

Computation and Analysis of Protein Circular Dichroism Spectra

Narasimha Sreerama and Robert W. Woody

*Department of Biochemistry and Molecular Biology,
Colorado State University,
Fort Collins, CO 80523, USA.*

Introduction

Circular dichroism (CD) is the most widely used form of chiroptical spectroscopy, spectroscopic techniques that utilize the differential interaction of molecules with left- and right-circularly polarized light. In the absence of a magnetic field, the molecule must be chiral to give rise to a difference in the interaction with the two types of circularly polarized light. The phenomenon of CD consists of the differential absorption of left- and right-circularly polarized light by a chiral molecule. Circular dichroism due to electronic transitions is generally referred to as CD (also as ECD), and that due to vibrational transitions as VCD. There are other types of chiroptical spectroscopy, which include circularly polarized luminescence, circular intensity differential scattering, and Raman optical activity.

In this chapter we discuss the origins of electronic CD in proteins, theoretical methods for computing protein CD, and empirical analysis of CD for estimating structural composition of proteins. A large number of scientists have contributed to our understanding of protein CD. Given the wide scope of this chapter and space limitations, we have been forced to be highly selective in our references, citing representative papers and review articles.

Basic Definitions

In this section we summarize the definition and units for CD. We will also discuss the origin of CD in proteins and the terminology utilized in reporting and discussing protein CD spectra.

CD and Rotational Strength

The phenomenon of CD involves the absorption of light and it can be considered as a special type of absorption spectroscopy. CD is the difference in the absorption of left- and right-circularly polarized light and is defined as:

$$\Delta\epsilon(\lambda) = \epsilon_L(\lambda) - \epsilon_R(\lambda) \quad (1)$$

where ϵ_L and ϵ_R , respectively, are the extinction coefficients for the left- and right-circularly polarized components at wavelength λ . The units for CD, when defined as $\Delta\epsilon$, are $M^{-1}\cdot\text{cm}^{-1}$, where M is the molar concentration.

Molar ellipticity is another measure for CD, and is defined as:

$$[\theta] = 100 \theta / C l \quad (2)$$

where C and l are the molar concentration and pathlength (cm) of the sample. Molar ellipticity is reported either as $\text{deg}\cdot\text{cm}^2\cdot\text{dmole}^{-1}$ or as $\text{deg}\cdot M^{-1}\cdot\text{m}^{-1}$, and these two units are equivalent. Molar ellipticity and molar CD (as it is called when expressed in $\Delta\epsilon$) are interconvertible by a factor

$$[\theta] = 3298 \Delta\epsilon \quad (3)$$

The integrated intensity of a CD band gives a measure of the strength of CD, called the rotational strength. Rotational strength has the units of $\text{erg}\cdot\text{cm}^3$, and is experimentally defined as

$$\begin{aligned} R &= (hc/32 \pi^3 N_A) \int (\Delta\epsilon/\lambda) d\lambda \\ &= 2.295 \times 10^{-39} \int (\Delta\epsilon/\lambda) d\lambda \end{aligned} \quad (4)$$

where h is Planck's constant, c is the velocity of light, and N_A is Avogadro's number.

Rotational strength is defined theoretically, following Rosenfeld's treatment,¹ as the imaginary part of the scalar product of the electric ($\boldsymbol{\mu}$) and magnetic (\mathbf{m}) dipole transition moments of an electronic transition

$$R = \text{Im}\{\boldsymbol{\mu} \cdot \mathbf{m}\} \quad (5)$$

This definition suggests the more frequently used units for rotational strength of Debye–Bohr magnetons (DBM; 1 DBM = $0.9274 \times 10^{-38} \text{erg}\cdot\text{cm}^3$).

Using Eq. 5 and quantum mechanical wave functions for the ground and excited states one can calculate the rotational strength for a given transition as

$$R_{0a} = \text{Im} \{ \langle 0 | \boldsymbol{\mu} | a \rangle \cdot \langle a | \mathbf{m} | 0 \rangle \} \quad (6)$$

where $\boldsymbol{\mu}$ is the electric dipole transition moment operator, a measure of the linear displacement of charge upon excitation; and \mathbf{m} is the magnetic dipole transition moment operator, a measure of the circular displacement of electron density upon excitation. The superposition of $\boldsymbol{\mu}$ and \mathbf{m} results in a helical displacement of charge, which interacts differently with left- and right-circularly polarized light.

The expression for rotational strength as given in Eq. 6 is origin-dependent because of the origin-dependence of the magnetic dipole transition moment operator. An alternate origin-independent formulation of rotational strength, known as the dipole-velocity formulation,^{2,3} uses the gradient operator, ∇ , and is normally used in the theoretical computation of CD.

$$R_{oa} = - (eh / 2\pi m v_{oa}) \text{Im} \{ \langle o | \nabla | a \rangle \cdot \langle a | \mathbf{m} | o \rangle \} \quad (7)$$

where e and m are, respectively, charge and mass of an electron, and v_{oa} is the frequency of the transition $o \rightarrow a$.

Calculation of the CD spectrum at a given wavelength, λ , is done assuming Gaussian bands for all transitions and using the relation between molar CD, $\Delta\epsilon_k$, and rotational strength, R_k , for a given transition k with a half-bandwidth (one-half of the width of the CD band at $1/e$ of its maximum) of Δ_k .

$$\Delta\epsilon_k = 2.278 R_k \lambda_k / \Delta_k \quad (8)$$

Protein CD

The CD spectra of proteins are generally divided into three wavelength ranges, based on the energy of the electronic transitions that dominate in the given range (Figure 1). These are: (1) the far UV (below 250 nm), where the peptide contributions dominate, (2) the near UV (250–300 nm), where aromatic side-chains contribute and, (3) the near UV–visible region (300–700 nm), where extrinsic chromophores contribute. Applications of CD to protein structure and folding have been developed based on these characteristic origins of protein CD spectra. The protein CD spectrum in the far UV is almost always reported on a residue basis, which corresponds to the concentration expressed in molar concentration of protein times the number of residues. The near-UV CD is generally reported on the basis of

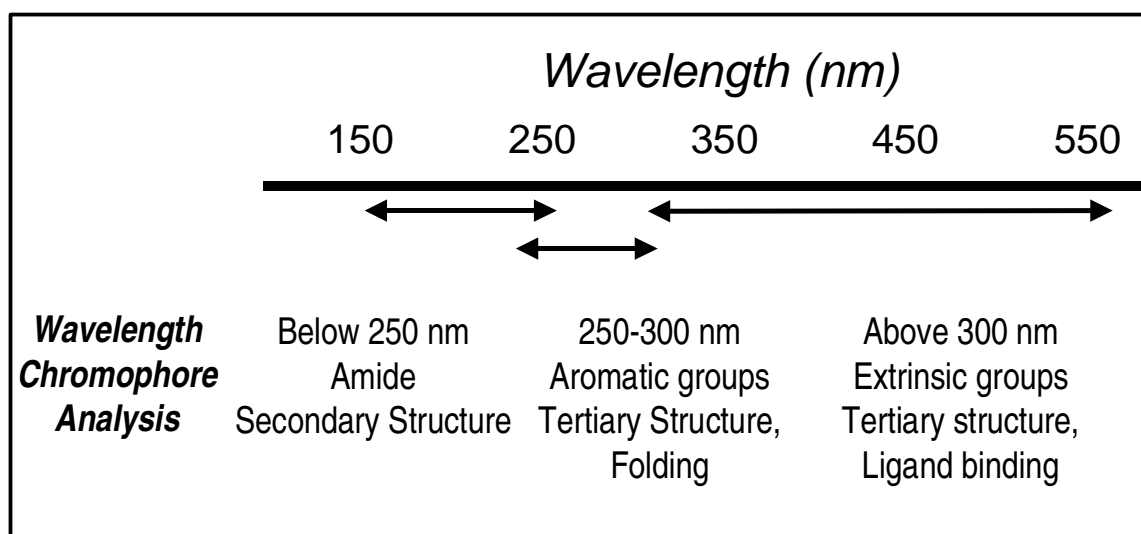


Figure 1. CD spectral regions and contributing chromophores in proteins.

residue molar concentration, but sometimes the protein concentration or aromatic residue concentration is used. Extrinsic chromophore CD is reported on the basis of protein concentration.

Protein CD has also been classified into "intrinsic" and "extrinsic", based on the contributing chromophores. Intrinsic CD of proteins is due to chromophores that are part of the protein (peptides, side-chain groups, and disulfides), and extrinsic CD is due to chromophores that are not part of the protein chain (heme, flavin, ligands, etc.). The early literature on protein CD considered the near-UV CD due to aromatic side chains as extrinsic CD, but we consider amino acid side chains as part of the protein and their CD contributions as "intrinsic" to the protein. Intrinsic protein CD is different from the "intrinsic CD" of a chromophore, which arises from the inherent chirality of the chromophore.⁴ Chromophores that contribute to protein CD are generally achiral. The peptide and the aromatic side-chain groups have a plane of symmetry, and do not exhibit intrinsic optical activity. Many extrinsic chromophores are also achiral in the isolated state and, when free, do not give CD bands. The "intrinsic CD" contributions of protein chromophores are generally assumed to be non-existent, except for the disulfide group. The disulfide chromophore is non-planar ($|\chi_{ssl}| \sim 90^\circ$), making it inherently dissymmetric and giving rise to intrinsic CD.

The electronic transitions in an achiral molecule have either the electric (e.g., $n\pi^*$ transitions) or the magnetic (e.g., $\pi\pi^*$ transitions) dipole transition moment equal to zero, or the two kinds of transition moments are perpendicular to each other (e.g., $n\sigma^*$ transitions), which results in zero rotational strength. The interaction between protein chromophores in the chiral field of the protein introduces perturbations leading to optical activity. Moffitt's seminal contributions² laid the foundations for the theoretical treatment of protein CD. The mechanisms that generate rotational strength have been worked out by Born, Kuhn, Kirkwood, Moffitt, Condon, Eyring, Tinoco, Schellman, and coworkers. These can be grouped into three basic types of interactions between transitions on different chromophores: (1) the coupled-oscillator mechanism⁵ (also known as the $\mu-\mu$ mechanism) describing the Coulomb interactions between two electrically allowed transitions, such as the $\pi\pi^*$ transitions of the peptide group, and resulting in composite transitions with non-vanishing electric and magnetic transition moments; (2) the $\mu-m$ mechanism⁶ wherein

the electrically allowed transition on one chromophore mixes with the magnetically allowed transition on another; and (3) the one-electron mechanism⁷ (also known as static-field mixing) where the electrically and magnetically allowed transitions on the same chromophore mix in the electrostatic field of the rest of the molecule. Exciton coupling is a special case of the coupled-oscillator mechanism in which degenerate transitions on identical chromophores mix. Theoretical calculation of CD of proteins is performed by considering each chromophore to be independent of others, thus neglecting overlap and electron exchange between chromophores, and constructing a composite wave function for the protein molecule from the wave functions of the constituent chromophores.

Computation of Protein CD

In this section we discuss the computation of protein CD using theoretical methods. The basic approach used to compute the CD of complex systems, such as proteins and nucleic acids, is the so-called "divide and conquer" approach.⁸ As one can glean from the previous section, the protein is treated as a collection of independent and mostly achiral chromophores, which are perturbed in the chiral environment of the protein. The characterization of the chromophores and their interactions forms the basis of theoretical calculation of CD. Methods based both on classical physics and on quantum mechanics have been developed to compute CD. The classical approach is based on the interaction of electromagnetic radiation with matter and the changes in polarizability, while the characteristics of electronic transitions form the basis of the quantum mechanical approach.

Classical Approaches

The framework for the classical method was developed by DeVoe.^{9,10} Here the electronic transitions, either individually or collectively, are considered as oscillators. The dipole moment induced by the electromagnetic field in an oscillator is described by the complex polarizability tensor. The real part of the polarizability tensor, which is important in wavelength regions away from the absorption band, is related to the index of refraction, and the complex part is proportional to the extinction coefficient; the real and imaginary parts of a complex function of a real variable, such as the complex polarizability, are related to one another by Kramers-Kronig relations.¹¹

The induced dipole moment in a given oscillator i has two sources: the electromagnetic field of the incident light, and the field produced by the induced dipole moments of all other oscillators. The expression for the induced dipole moment for oscillator i is given as,

$$\boldsymbol{\mu}_i^{\text{ind}} = \boldsymbol{\alpha}_i \mathbf{E}_i^{\text{ext}} - \sum_{j \neq i} \boldsymbol{\alpha}_i \mathbf{T}_{ij} \boldsymbol{\mu}_j^{\text{ind}} \quad (9)$$

where $\boldsymbol{\alpha}_i$ is the complex polarizability, $\mathbf{E}_i^{\text{ext}}$ is the external field, \mathbf{T}_{ij} is the dipole interaction tensor for oscillators i and j , $\boldsymbol{\mu}_j^{\text{ind}}$ is the induced dipole moment for oscillator j . The set of equations for N oscillators can be written as a matrix equation

$$\boldsymbol{\mu}_i^{\text{ind}} = \mathbf{G}^{-1} \mathbf{E}_i^{\text{ext}} = \mathbf{A} \mathbf{E}_i^{\text{ext}} \quad (10)$$

where \mathbf{G} is a $3N \times 3N$ matrix comprised of N^2 3×3 submatrices, $\boldsymbol{\alpha}_i^{-1}$ as diagonal elements and \mathbf{T}_{ij} as off-diagonal elements. The inverse of matrix \mathbf{G} is matrix \mathbf{A} .

The induced dipole moments $\boldsymbol{\mu}^{\text{ind}}$ are analogous to the electric dipole transition moments given in Eq. 7, and the moments of induced moments with respect to the origin, $\mathbf{r} \times \boldsymbol{\mu}^{\text{ind}}$, are analogous to the magnetic dipole transition moments. The expressions for absorption and CD for electrically allowed transitions are then derived as^{9,10}

$$\boldsymbol{\epsilon} = 8 \pi^2 N_A / 6909 \lambda \sum_i \sum_j \text{Im}(A_{ij}) \mathbf{R}_{ij} \mathbf{e}_i \bullet \mathbf{e}_j \quad (11)$$

$$\Delta\boldsymbol{\epsilon} = 16 \pi^3 N_A / 6909 \lambda^2 \sum_i \sum_j \text{Im}(A_{ij}) \mathbf{R}_{ij} \bullet \mathbf{e}_i \times \mathbf{e}_j \quad (12)$$

where \mathbf{e}_i and \mathbf{e}_j are the unit vectors along $\boldsymbol{\mu}_i^{\text{ind}}$ and $\boldsymbol{\mu}_j^{\text{ind}}$, and \mathbf{R}_{ij} is the vector from i to j . In this approach, the extinction coefficient and CD at a given wavelength are calculated directly as a summation of contributions from all oscillators in a given molecule. These expressions describe the coupled oscillator interactions between the electrically allowed transitions, which are normally considered in classical polarizability theory applications for organic molecules and biopolymers.

Extensive applications of classical polarizability theory for proteins have been performed by Applequist and coworkers.¹²⁻¹⁶ They have developed the atom dipole interaction model in which each atom has an isotropic polarizability, independent of wavelength, and the dipole moments induced on the atoms are incorporated in the calculation of matrix **A**. Small organic molecule polarizability data at the NaD line were used to obtain polarizability parameters for C, O, N, H, and the halogens. In addition to the isotropic atomic polarizabilities, the anisotropic polarizability of the first electrically allowed peptide transition ($\pi\pi^*$ at 190 nm) was included in the dipole interaction model for protein CD calculations. Polarizability anisotropy data for simple amides were used for obtaining polarizability parameters. The results for proteins and polypeptides in specific conformations obtained with the dipole-interaction model are satisfactory in the $\pi\pi^*$ region. The major drawback of the dipole-interaction model is the neglect of the electrically forbidden amide $n\pi^*$ transition (at 220 nm), which is very important in proteins.

Matrix Method

Direct calculation of rotational strengths of proteins using quantum chemical methods and Eq. 7 is, in principle, possible. However, the size and complexity of a protein makes such a computation next to impossible. The matrix method^{17,18} performs this task by combining the interactions between the chromophores in the protein, and can be considered as a quantum chemical method depending on the description of chromophores. The matrix method has its origins in the coupled-oscillator and the exciton models, wherein a secular determinant is constructed based on the energy and interactions between the transitions on chromophores. The secular matrix for a two-chromophore system with one transition on each chromophore is written as,

$$\begin{pmatrix} E_1 & V_{12} \\ V_{12} & E_2 \end{pmatrix} \quad (13)$$

where E_1 and E_2 are the energies of the transitions and V_{12} is the interaction energy between the two transition moments. The secular matrix is diagonalized, yielding two excited states that are linear combinations of individual transitions.

An analogous matrix can be constructed for a protein, comprised of a set of chromophores with localized transitions, with the energies of the transitions forming the diagonal elements and the interactions between different

$$\begin{pmatrix}
 E_{11} & V_{11,12} & V_{11,13} & V_{11,21} & V_{11,22} & V_{11,23} & \bullet & \bullet & \bullet & V_{11,N1} & V_{11,N2} & V_{11,N3} \\
 \bullet & E_{12} & V_{12,13} & V_{12,21} & V_{12,22} & V_{12,23} & & & & V_{12,N1} & V_{12,N2} & V_{12,N3} \\
 \bullet & & E_{13} & V_{13,21} & V_{13,22} & V_{13,23} & & & & V_{13,N1} & V_{13,N2} & V_{13,N3} \\
 \bullet & & & E_{21} & V_{21,22} & V_{21,23} & & & & V_{21,N1} & V_{21,N2} & V_{21,N3} \\
 \bullet & & & & E_{22} & V_{22,23} & & & & V_{22,N1} & V_{22,N2} & V_{22,N3} \\
 \bullet & & & & & E_{23} & & & & V_{23,N1} & V_{23,N2} & V_{23,N3} \\
 \bullet & & & & & & \bullet & & & \bullet & \bullet & \bullet \\
 \bullet & & & & & & & \bullet & & \bullet & \bullet & \bullet \\
 \bullet & & & & & & & & \bullet & \bullet & \bullet & \bullet \\
 \bullet & & & & & & & & & E_{N1} & V_{N1,N2} & V_{N1,N3} \\
 \bullet & & & & & & & & & & E_{N2} & V_{N2,N3} \\
 \bullet & & & & & & & & & & & E_{N3}
 \end{pmatrix} \quad (14)$$

localized transitions on chromophores forming the off-diagonal elements. For example, one can construct a $3N \times 3N$ interaction energy matrix, \mathbf{H} , for a protein with N residues with three transitions on each amide chromophore. Here E_{ij} represents the energy of the transition j on chromophore i , and $V_{ij,kl}$ represents the interaction between transition j on chromophore i and the transition l on chromophore k . A transition localized on a given chromophore is represented by an appropriate charge distribution, for example, a set of point charges describing the dipole transition moment. Electrically allowed $\pi\pi^*$ transitions have a non-zero electric dipole transition moment and are represented by dipolar charge distributions. Electrically forbidden $n\pi^*$ transitions are represented by a set of quadrupolar charges centered on relevant atoms. The point charges are determined either from the quantum chemical wave function or from experiment, or a combination of both. The Coulomb interaction between localized transitions j and l on chromophores i and k , respectively, are calculated using the monopole-monopole approximation as,

$$V_{ij,kl} = \sum_m \sum_n q_{ijm} q_{kln} / r_{ijm,kl} \quad (15)$$

where indices m and n correspond to the point charges for transitions j and l , respectively, and the distance between point charges is given by r . The monopole-monopole approximation is a better approach to compute the Coulomb interaction between two charge distributions because the dipole-dipole approximation breaks down at short distances, e.g., when chromophores are at or near van der Waals distances.

In the example given above, the first $n\pi^*$ ($j = 1$) and the first two $\pi\pi^*$ ($j = 2$ and 3) transitions on the peptide group have been considered. The three types of interactions between different transitions described by the $\mu-\mu$, $\mu-m$, and one-electron mechanisms are distributed among the various off-diagonal elements of this matrix. For the terms with $i \neq k$, $j = 2$ or 3 , and $l = 2$ or 3 , both interacting transitions are electrically allowed, and the terms correspond to the $\mu-\mu$ mechanism. Terms with $i \neq k$, $j = 1$, and $l = 2$ or 3 describe the interaction between an $n\pi^*$ and a $\pi\pi^*$ transition on different chromophores and correspond to the $\mu-m$ mechanism. When $i = k$, both transitions lie on the same chromophore representing the mixing of excited states in the static field of the molecule, corresponding to the one-

electron mechanism. An additional term, with $i \neq k$, $j = 1$, and $l = 1$, describes the interaction between quadrupolar charge distributions of two $n\pi^*$ transitions on two different chromophores.

Diagonalization of matrix \mathbf{H} gives the eigenvalues and eigenvectors that describe the excited states of the composite molecule; these can be called group-states.

$$\mathbf{H} = \mathbf{U} \mathbf{H}_d \mathbf{U}^T \quad (16)$$

Eigenvalues, the diagonal elements of matrix \mathbf{H}_d , give the energies of the transitions; eigenvectors, elements of the unitary matrix \mathbf{U} ($\mathbf{U} \mathbf{U}^T = \mathbf{I}$, where matrix \mathbf{U}^T is the transpose of matrix \mathbf{U} and \mathbf{I} is the identity matrix), describe the mixing of localized transitions. By combining eigenvectors with the properties of localized transitions, $\boldsymbol{\mu}_k^0$ and \mathbf{m}_k^0 , one can determine the transition moments for each group-state for the composite system, and calculate the rotational strength.

$$\mathbf{m}_i = \sum_k \mathbf{U}_{ki} \mathbf{m}_k^0 \text{ and } \boldsymbol{\mu}_i = \sum_k \mathbf{U}_{ki} \boldsymbol{\mu}_k^0 \quad (17)$$

Tinoco's first-order perturbation method,¹⁹ which predates the matrix method, considers doubly excited states (excitation of electrons on two chromophores) in addition to the singly excited states. While this method gives expressions for $\mu-\mu$, $\mu-m$, and one-electron terms generating rotational strength, it requires the diagonalization of the Hamiltonian matrix for degenerate or near-degenerate transitions before the perturbative treatment. On the other hand, the matrix method is equivalent to an all-order perturbation approach and is much easier to implement. Nowadays, most protein CD calculations are performed using the matrix method.

Chromophores and Transitions

The chromophores that are important in protein CD calculations are the amide, the aromatic side chains, disulfides, and extrinsic groups. Of these, the amide chromophore, which forms the backbone of the polypeptide chain, is the most abundant and dominates the far-UV CD; the contributions from aromatic side chains and extrinsic groups, respectively, dominate the near-UV and visible regions (Figure 1). Only a few of the amino acid side chains have transitions at wavelengths longer than 180 nm²⁰ (aromatic side chains, disulfides, histidine, arginine, glutamine, asparagine, glutamic acid and aspartic acid), and among those only the aromatic transitions and disulfides have large enough electric or magnetic dipole transition moments to make appreciable spectral contributions. Most protein CD calculations, however, include only the amide transitions. Only a limited number of CD calculations have incorporated the aromatic side-chain and disulfide transitions, and those that include extrinsic chromophores (for example, the heme group in myoglobin and retinal in rhodopsins) are even rarer.

The transition parameters are generally obtained from molecular orbital calculations on structures modeling the chromophores, combined with experimentally available transition dipole moments and wavelengths.²¹ The electrically allowed $\pi\pi^*$ transitions, experimental data for which can be obtained, are represented by an electric dipole transition moment. The magnetically allowed $n\pi^*$ transitions, which are difficult to fully characterize experimentally, are represented by a magnetic dipole transition moment and an electric quadrupole transition moment,²² obtained from quantum chemical calculations.

One or two $\pi\pi^*$ and one or two $n\pi^*$ amide transitions have been used in protein and polypeptide CD calculations, and the parameters were mostly obtained from semi-empirical MO calculations and available

experimental data on secondary amides. The amide group has four π electrons, three π orbitals (π_+ , π_0 and π^*) and two lone pairs in the non-bonding orbitals (n and n') of the valence shell. Two $\pi\pi^*$ transitions, $\pi_0 \rightarrow \pi^*$ (denoted NV_1 , at 190 nm, electric dipole transition moment $|\mu| \approx 3.1$ Debyes, D) and $\pi_+ \rightarrow \pi^*$ (NV_2 , 139 nm, $|\mu| \approx 1.8$ D), and one $n\pi^*$ ($n \rightarrow \pi^*$, 220 nm) transition in the amide chromophore have been identified. Semi-empirical calculations predict the second $n\pi^*$ transition ($n' \rightarrow \pi^*$) near the NV_1 transition, but it has not been identified experimentally and ab initio calculations place it at much higher energies.²³ The experimental transition moment directions of NV_1 and NV_2 , characterized by single-crystal polarized reflection studies of N-acetylglycine,²⁴ are -55° and 10° or 61° , respectively. The $n \rightarrow \pi^*$ transition has a large magnetic dipole transition moment along the carbonyl bond. The wavelength of the $n \rightarrow \pi^*$ transition in amides is dependent on solvent (230 nm in apolar solvents to 210 nm in strongly polar solvents), while those of $\pi \rightarrow \pi^*$ transitions are much less sensitive to solvent. The parameters for the tertiary amide, formed by the imino acid proline, are slightly different from those for secondary amide; the wavelength of the first $\pi\pi^*$ transition is at ~ 200 nm, which is ~ 10 nm longer than that in secondary amide. The polarizations of the $\pi\pi^*$ transitions in tertiary amides are not known, but ab initio calculations indicate that they are similar to those for secondary amides.²³

The transitions of Phe and Tyr side chains that contribute to near- and far-UV protein CD are well understood, as the chromophores are derivatives of benzene; the Phe side chain is an alkyl-substituted benzene, and the Tyr side chain is an alkylated phenol. These benzenoid chromophores have four $\pi\pi^*$ transitions: L_b , L_a , B_b and B_a , in Platt's notation.²⁵ Of these, the L transitions are forbidden and the B transitions are allowed. Furthermore, the B transitions are degenerate in benzene. The L transitions, however, are coupled to B transitions vibronically and borrow most of their intensity from the B transitions.²⁶ In the Phe side chain, the weakly perturbing alkyl substitution does not change the situation; here, the B transitions are very nearly degenerate and the L transitions derive most of their intensity through vibronic coupling with the B transitions. Phe side-chain transitions are: L_b , 260 nm, $|\mu| \approx 0.4$ D; L_a , 210 nm, $|\mu| \approx 2.7$ D; B_b and B_a , 185 nm, $|\mu| \approx 6$ D. In the Tyr side chain, stronger perturbation by the phenolic oxygen shifts the L transitions to longer wavelengths, makes the L transitions electrically allowed, and makes the B transitions nondegenerate. In addition, the vibronic coupling between the B and L transitions is weaker than that in the Phe side chain due to larger energy separation. Tyr side-chain transitions are: L_b , 275 nm, $|\mu| \approx 1.2$ D; L_a , 230 nm, $|\mu| \approx 2.8$ D; B_b and B_a , 190 nm, $|\mu| \approx 6$ D.

The Trp side-chain transitions can be approximated by those of indole, a perturbed naphthalene, which in Platt's model are also: L_b , L_a , B_b and B_a . These $\pi\pi^*$ transitions occur in the near- and far-UV region, and are more complex than those of the Tyr side chain due to vibronic fine structure and overlapping transitions. The first three transitions have been characterized experimentally.²⁷ The L_b and L_a transitions are between 270 and 280 nm and the L_b band has well-resolved vibronic components; the L_a transition is quite sensitive to the environment; the B_b transition occurs near 225 nm. There are two to three strong transitions between 180-210 nm, as determined by the absorption spectra of indole derivatives. Parameters for six transitions of the Trp side chain were obtained by combining experimental results with Pariser-Parr-Pople calculations.

Two $n \rightarrow \sigma^*$ transitions have been characterized in the disulfide group, which arise from excitation of lone pair orbitals on the sulfurs to the σ^* orbital of the S-S bond. These are electrically forbidden transitions but have strong

magnetic dipole transition moments. The transition energies and transition dipole moments depend on the C_{β} -S-S- C_{β} dihedral angle, and the excitation energies were determined empirically by fitting experimental transition energies for model disulfides with different C_{β} -S-S- C_{β} angles to a cubic polynomial.²⁸ By extending the Bergson model accounting for the C_{β} -S-S- C_{β} angle dependence of the absorption bands, analytical expressions for transition monopoles as a function of the C_{β} -S-S- C_{β} angle have been derived.²⁹

Protein CD spectra

The polypeptide chain in a protein is primarily made up of secondary amides. Proline, which accounts for ~5% of residues in proteins, forms tertiary amide. Polypeptides form different secondary structures based upon the arrangement of amide groups dictated by the backbone conformation. The interaction between transition moments leads to CD, and the geometrical relation between amide groups in different secondary structures determines the characteristic secondary structure CD spectra. Proteins vary in the extent of their secondary structure, and the secondary structure composition and spatial arrangement of amides determines the far-UV CD spectrum. The correspondence between the secondary structure and CD spectrum is exploited in developing applications of CD.

α -helices and β -sheets are the two most important secondary structures in proteins and they are stabilized, respectively, by intra- and inter-chain hydrogen bonds. They are characterized by a set of dihedral angles ϕ and ψ (α : -57° , -47° ; β : -120° , $+120^\circ$) that repeat along the polypeptide chain so that the successive amide groups are oriented identically with respect to the overall direction of chain propagation. Amide groups in α -helices form an approximately cylindrical surface with intra-chain hydrogen bonds parallel to the helix axis, and those in β -sheets span a planar surface. Two types of β -sheets exist, having polypeptide strands either parallel or antiparallel to each other. In general, α -helices are longer and more rigid than β -sheets because of the nature of the hydrogen bonds that stabilize them. Proteins are often characterized by the relative contents of α and β structures as α -rich, β -rich, and $\alpha\beta$ proteins. Another important secondary structure is the β -turn, generally formed by three residues and stabilized by a hydrogen bond between the first and the third amide group, which effectively reverses the direction of chain propagation. Turns are characterized by sets of differing dihedral angles for successive residues³⁰ and need not have a stabilizing hydrogen bond. Another secondary structure that has repeating backbone dihedral angles (ϕ , ψ : -78° , 150°) and is present in proteins, although to limited extent, is the poly(Pro)II type structure (P_2). This structure is the preferred conformation of proline-rich polypeptides, such as collagen, and is formed due to the limited conformational mobility of the proline side chain. Short stretches of P_2 structures involving Pro and non-Pro residues do occur in proteins, and they form an appreciable fraction of the residues that do not form α - or β - or turn structures.³¹ Amino acid residues that do not form any of these secondary structures exist in proteins, and these are called "unordered" for lack of ordered conformation. In many instances, the P_2 conformation is not explicitly identified and is implicitly treated as part of the unordered structure.

The three regular secondary structures, α , β , and P_2 , that have repeating ϕ and ψ angles have characteristic CD spectra (Figure 2). The α -helical CD spectrum³² is characterized by two negative bands at 222 and 208 nm, and a positive band at 192 nm, which are normally used in CD analysis. Additional bands, a positive shoulder at 175 nm, a

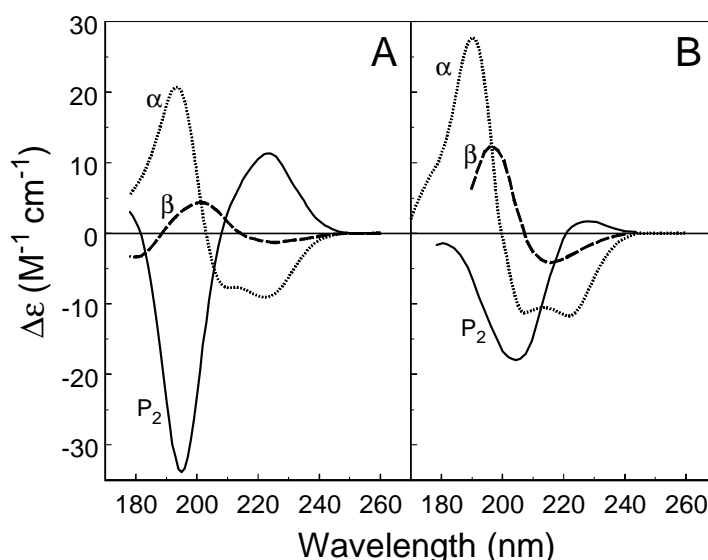


Figure 2. (A) CD spectra of α -helix, β -sheet, and P_2 structure deconvoluted from a reference-protein set of 37 proteins. (B) CD spectra of model polypeptides in α -helical,³² β -sheet,³⁶ and poly(Pro)II helix³⁹ conformations.

negative band at 160 nm and a positive band centered below 140 nm, have been observed in the vacuum UV,³³ but the latter two are generally inaccessible for solution studies. In proteins, α -helices show variations in their structure and the average geometry is different from the canonical form. Calculated CD spectra of α -helical fragments are consistent with the model α -helical CD. That the intensity of α -helical CD increases with the chain length has been confirmed by theoretical calculations. An empirical expression

$$V_r = V_\infty (r-k)/r \quad (18)$$

where V is the CD amplitude at a specific wavelength for a helix of r residues, V_∞ is the CD of an infinite helix at the same wavelength, and k is an empirical parameter, describes the chain-length dependence of α -helical CD.³⁴ Physically, the parameter k corresponds to the number of residues "missing" due to end effects. Recent experimental data suggests that the value of k is not a constant but decreases with the helix length for short helices.³⁵ However, for longer helices it can be considered a constant (~ 4).

The CD spectrum of the typical β -sheet³⁶ has a negative band near 215 nm and a positive band near 198 nm (Figure 2); additional short-wavelength bands near 175 (negative) and 168 nm (positive) are also observed.³⁷ The β -sheet CD is difficult to characterize due to the variations in the geometry of β -structure in polypeptides and proteins, and the limited solubility of polypeptides in the β -structure. Among the variations in the β -sheet geometry, twisting of the β -sheet has been examined, and it was found that twisting led to increased amplitudes of the CD bands. Weakly twisted β -sheets have similar amplitudes for the two bands of opposite sign at 215 and 198 nm, whereas strongly twisted β -sheets have a stronger 198 nm band.³⁸ Theoretical studies of chain-length dependence of β -sheet CD predict an increase in the intensity of both bands with increasing length of the β -sheet.

The P_2 conformation is a left-handed helix with three residues/turn, the geometry of which is derived from the poly(Pro)II helix formed by *trans* peptides; the poly(Pro)I-helix has *cis* peptides. The CD spectrum of the poly(Pro)II-

helix is characterized by a weak positive band near 226 nm and a strong negative band near 206 nm (Figure 2).³⁹ This spectrum is qualitatively similar to those of unordered (called random coil in older literature) and unfolded proteins, which suggested the existence of P_2 conformation in them.⁴⁰ Subsequent analyses of protein crystal structures confirmed the existence of short stretches of P_2 structure in globular proteins.^{31,41,42}

The variable nature of turns in proteins, with the characterization of at least eight different kinds of turns,³⁰ makes its CD characterization difficult. Model cyclic and linear peptide studies combined with theoretical calculations have been used in identifying salient features for the major types of turns in proteins.⁴³ Type I turns have CD spectra qualitatively resembling that of the α -helix. Type II turns have CD spectra similar to that of the β -sheet, but with red-shifted extrema. Type II' turns are predicted and observed to have an α -helix-like CD spectrum.

Moffitt's exciton picture² and the subsequent theoretical developments explain the basic interactions resulting in the CD patterns for various polypeptide secondary structures and proteins. In the α -helix the exciton splitting of the first peptide $\pi\pi^*$ transition gives rise to three components. A long-wavelength component polarized along the helix axis is responsible for the negative 208 nm band, and two degenerate components polarized perpendicular to the helix axis result in the positive 192 nm band. The negative band at 222 nm is due to the peptide $n\pi^*$ transition. In addition to the Moffitt exciton bands, theory predicts a "helix band", which is a consequence of helical geometry.⁴⁴ The helix band corresponds to two strong, slightly split bands of opposite sign, giving rise to a positive couplet centered at the perpendicularly polarized band position. (A couplet is a pair of closely spaced CD bands of opposite sign and is designated as positive or negative according to the sign of the long-wavelength component.) While direct experimental evidence is lacking for the helix band in the α -helix, its existence has been established indirectly using multiple spectroscopic techniques.

The exciton interactions in β -sheets result in the positive band near 195 nm and a negative band near 175 nm. The negative band near 215 nm is assigned to the $n\pi^*$ transition. The amplitude of the 215 nm band is largely determined by the mixing of $n\pi^*$ and $\pi\pi^*$ via the μ - m mechanism. Exciton coupling theory predicts a positive $\pi\pi^*$ couplet in both parallel and antiparallel β -sheet structures, but the effect is smaller than that in the α -helix, resulting in smaller amplitudes. The helix band is absent in β -sheets because the two directions perpendicular to the chain direction are not equivalent (nondegenerate).

The assignments of far-UV CD bands in the P_2 conformation are more complicated because of the non-conservative nature of its CD spectrum, i.e., the sum of the rotational strengths does not vanish. The long-wavelength band (~217 nm) is generally assigned to the $n\pi^*$ transition and the short-wavelength band (~200 nm) to the $\pi\pi^*$ transition. Strong mixing with the higher energy transitions and/or asymmetric solvent effects are believed to be responsible for the non-conservative CD spectrum.

Protein CD can be approximated as a linear combination of secondary structure spectra determined by their relative contents in the protein structure. The coupled-oscillator interactions between various $\pi\pi^*$ transitions and the mixing of $\pi\pi^*$ and $n\pi^*$ transitions describe the basic mechanisms. Madison and Schellman⁴⁵ calculated protein CD spectra using the matrix method¹⁷ and by combining the CD from different segments. The results for α -helical segments resembled those of ideal α -helices. For β -structures, agreement with ideal β -sheet CD was only qualitative,

which was attributed to variations of β -structures in proteins. The nonperiodic chain fragments showed the greatest disagreement between experiment and theory, and even poorer agreement was obtained for poly(Pro)II-helix.

Systematic comparisons^{16,46-50} of calculated CD with experiment have been carried out for a set of proteins with known crystal structures and different compositions of secondary structures, using both matrix and polarizability methods. These calculations were performed using a description of transitions, obtained either by high-level quantum chemical methods or by experimental data on transition moment directions.

The dipole interaction model calculations performed by Bode and Applequist¹⁶ used the empirical polarizabilities of the amide $\pi_0\pi^*$ transition and the aliphatic side chains, and were performed on protein fragments representing different structural components (helix, sheet, and unordered; here, the unordered structure includes turns). The CD spectra of the reassembled proteins constructed from the individual component spectra were comparable to the weighted average component spectra. Bandwidths of 4000 and 6000 cm^{-1} were employed in these calculations and, in general, the larger bandwidth (6000 cm^{-1} , which is ~ 22 nm at 190 nm) gave better results. The $n\pi^*$ transition was not included in these calculations, and the theoretical predictions for the 190 nm band compared qualitatively with experiment. The inclusion of solvent effects in the dipole interaction model indicated that the effects of the solvent on the calculated CD spectra are small for proteins. The dipole interaction model predicts a qualitatively correct CD spectrum for the poly(Pro)II-helix,¹³ but fails to reproduce the CD spectra of unordered regions.

Protein CD spectra were calculated by Hirst and coworkers^{46,47} using the matrix method¹⁷ and amide transition parameters derived from various quantum chemical methods, ranging from semi-empirical CNDO/S to solution-phase *ab initio* CASSCF/SCRF methods, with acetamide and N-methyl acetamide as models for the peptide chromophore. Their best results were obtained with transition monopoles that reproduce the *ab initio* electrostatic potential arising from the transition charge densities.⁴⁷ A bandwidth of 15.5 nm was used for all transitions. The Spearman rank correlation coefficients between the experimental and calculated values of $[\theta]_{190}$ for 29 proteins improved from 0.44 (CNDO/S parameters) to 0.66, and from 0.48 to 0.90 for $[\theta]_{220}$. The improvement appears to be mainly due to the use of transition charge densities fitted to the electrostatic potential rather than use of the CASSCF parameters per se, since the correlation coefficients obtained when the solution-phase transition monopole charges were placed near atomic centers were poorer or comparable to those obtained with the gas-phase parameters. The results for 15 proteins that were common between this protein set and that used in the dipole interaction model were comparable. The dipole interaction model showed better correlation at 190 nm while the matrix method with improved parameters showed better correlation at 220 nm.

Woody and Sreerama⁴⁹ also used the matrix method¹⁷ in their calculation of protein CD spectra, but their transition parameter set was derived from a combination of experimental data and theoretical parameters. For the two amide $\pi\pi^*$ transitions, they used experimental data for secondary amides, and for the $n\pi^*$ transition they used parameters derived from INDO/S wavefunctions. The transition parameters used for protein chromophores are described in the preceding section. Bandwidths used were: 10.5 nm for the $n\pi^*$ transition; 11.3 nm for the $\pi_0\pi^*$ transition; 7.2 nm for the $\pi_+\pi^*$ transition; and 12.8 to 7.1 nm for the various aromatic transitions, calculated from an empirical relationship between bandwidth and transition wavelength. The Spearman rank correlation coefficients obtained for a set of 23 proteins at 190 and 220 nm were 0.66 and 0.84, respectively. Inclusion of side-chain transitions

led to inferior correlation coefficients (0.46 and 0.71, respectively). The rank correlation coefficient, which calculates the correlation between the ranking of the proteins determined by the CD value at a given wavelength, is a poor measure. A more detailed comparison indicates larger disagreements between theory and experiment, which is pronounced in the 195-208 nm region where the mean absolute error in ellipticity is 2 – 3 $\Delta\epsilon$.

Theoretical CD spectra (from Woody & Sreerama⁴⁹) for eight proteins, two each from α -rich, $\alpha\beta$, β_I , and β_{II} classes,⁵¹ are compared with experiment in Figure 3. These represent typical results one obtains from CD calculations. The theoretical CD spectra of proteins with large α -helix content from all three sets of calculations generally agree with the experiment, as do those for ideal α -helix and β -sheet structures. The results for β -sheet-rich proteins are

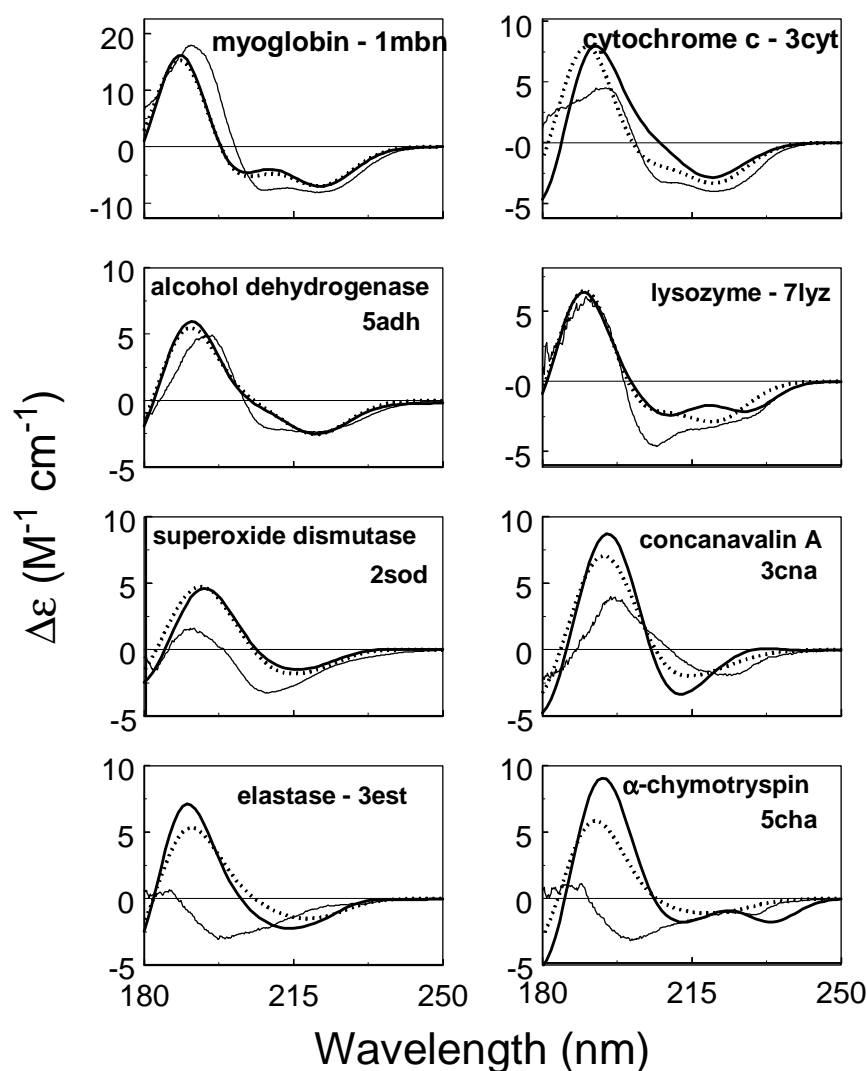


Figure 3. Comparison of theoretical and experimental CD spectra for selected proteins. The two theoretical CD spectra⁴⁹ were obtained by excluding (including only peptide groups, dotted line) or including aromatic chromophores (including all aromatic side chains and peptide groups, thick line) in the calculation. The experimental CD spectra⁷² are shown as thin lines. The PDB code of the crystal structure used in the CD calculation is given in the figure.

much poorer, predicting positive and larger than experimentally observed amplitudes for the 195 nm band. Agreement is especially poor for some β -proteins that have a negative band around 200 nm. The results for proteins with comparable α and β contents are intermediate. Overall, the agreement between theoretical and experimental protein CD spectra could be called semiquantitative, and the β -proteins are the major source of disagreement.

The β -rich proteins exhibit two types of CD spectra, one showing features that are reminiscent of model β -sheets, a positive $\pi\pi^*$ band (~ 195 nm) followed by a negative $n\pi^*$ band (~ 215 nm), and the other those of unordered polypeptides, a negative $\pi\pi^*$ band (~ 200 nm), leading to the classification as β_I - and β_{II} -proteins.⁵¹ The existence of P_2 structure in the unordered conformation has been established.³¹ An analysis of β -sheet and P_2 structure contents in β -rich proteins showed that β_{II} -proteins have larger ratios of P_2 to β -sheet contents (> 0.4) than β_I -proteins (< 0.4).⁵² The average P_2 structure content in globular proteins is small ($< 10\%$), and generally the P_2 structure appears in short stretches.^{31,41,42} In proteins with moderate to high α -content, CD contributions from P_2 structure is probably masked by the strong α -helical contribution. However, in β -rich proteins contributions from P_2 structure seems to be influencing the protein CD, and the inability of the matrix method to correctly calculate the P_2 CD appears to be largely

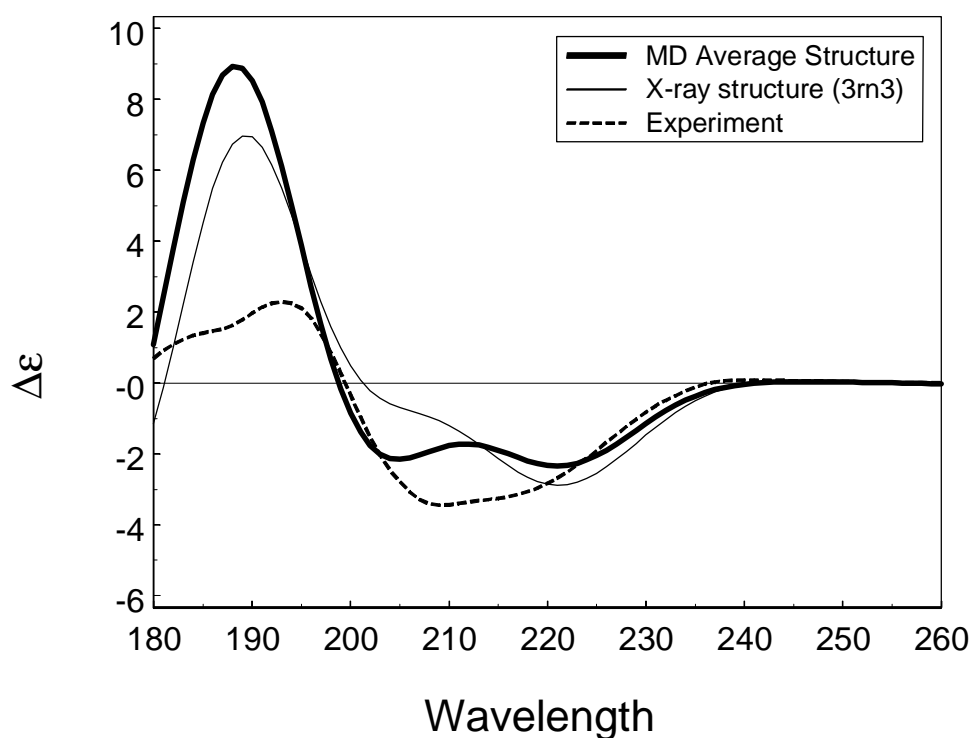


Figure 4. Comparison of theoretical (solid lines) and experimental (dotted line) CD spectra of ribonuclease A. One theoretical CD spectrum (thin line) corresponds to the static structure (PDB code of the crystal structure: 3rn3). The second theoretical spectrum (thick line) was the result of a combined molecular dynamics/CD study, and was obtained as an average of 75 CD spectra corresponding to 75 structures, at an interval of 4 ps, along a 300 ps molecular dynamics trajectory.⁵⁴

responsible for the greater disagreement between theory and experiment for β -rich proteins. A better description of transition parameters or improvements in the methodology that predict the non-conservative CD spectrum of P_2 structure is needed for a better understanding of protein CD spectra.

Most protein CD calculations use a static structure, generally determined by x-ray diffraction, while the experimental data correspond to the dynamic solution-state structure. Integration of molecular dynamics with CD calculations provides a means for introducing structural fluctuations in these calculations.^{48,53,54} Both equilibrium and folding/unfolding molecular dynamics trajectories can be used in such studies. The equilibrium dynamics of elastase and concanavalin A indicate that while the ensemble-averaged CD spectrum of concanavalin A, a β_1 -protein, was similar to that of the crystal structure, that of elastase, a β_{II} -protein, was quite different.⁴⁸ A shift in the population of dihedral angles towards the P_2 region in the Ramachandran plot that was observed in the elastase ensemble was deemed responsible for the changed CD spectrum, which reduced the disagreement with experiment.⁴⁸ The myoglobin unfolding trajectory was examined for changes in helicity, estimated from the $[\theta]_{220}$ values, which stayed near 0.3 while that calculated from the structure fluctuated between 0.2 and 0.4.⁴⁸ Equilibrium dynamics of ribonuclease A combined with CD calculations resulted in somewhat better agreement between theory and experiment than that obtained with the static crystal structure^{28,54} (Figure 4). These studies have demonstrated the feasibility of the combined molecular dynamics/CD approach for proteins, but the inability of the CD calculations to account for the non-conservative CD spectrum of the P_2 conformation still exists. Inclusion of other factors, such as the solvent effects, non-planarity of the peptide bond, and mixing with high-energy transitions should lead to further improvements in the calculation of protein CD.

Analysis of Protein CD

In this section we will discuss the application of CD spectroscopy to obtain insights on protein structure and conformation. The analysis generally starts with the protein CD spectrum and may be considered the inverse of the computation of protein CD, where one starts with the structure. However, computed CD spectra can also be used in deriving helpful information about the conformation of side chain and extrinsic chromophores.

The far-UV CD of a protein is generally reflective of the secondary structure content of the protein because it is dominated by contributions from the amide chromophore, which are determined by the backbone conformation. One of the most successful applications of CD, the structural characterization of proteins, is developed from this remarkable sensitivity of far-UV CD to secondary structure. Although the side-chain contributions to protein far-UV CD are generally submerged under amide contributions, some proteins have strong interactions between aromatic chromophores that can lead to significant far-UV CD bands. On the other hand, the near-UV CD is generally due to aromatic and disulfide contributions, providing insights into tertiary interactions. CD bands due to extrinsic chromophores, such as prosthetic groups, metal ions, inhibitors, and substrate analogs, are seen in the visible and near-UV regions and give information about protein-ligand interactions.

Secondary structure analysis

The approximation that a given protein CD spectrum (C_λ) can be expressed as a linear combination of secondary structure component spectra, B_{kl} , given as

$$C_\lambda = \sum f_k B_{k\lambda} \quad (19)$$

where f_k is the fraction of secondary structure k , forms the basis of analytical methods for protein CD analysis.⁵⁵⁻⁶⁰ With the knowledge of $B_{k\lambda}$ one can determine the fractions f_k corresponding to a given C_λ using Eq. 19. $B_{k\lambda}$ can be obtained from polypeptides in specific conformations, as was done in the earlier methods. Secondary structures in proteins deviate from ideal conformations, and in most current methods $B_{k\lambda}$ that give a better representation of component spectra in proteins are derived from a set of CD spectra of proteins with known secondary structures. Such a set of proteins is called a basis set or reference set.

For a reference set of N proteins with known CD spectra and secondary structures, one can write N linear equations.

$$[C_\lambda]_N = [\sum f_k B_{k\lambda}]_N \quad (20)$$

The common elements among them are the component spectra $B_{k\lambda}$, which can be obtained by solving the N equations simultaneously and minimizing the difference between the calculated and experimental CD spectra for the reference proteins by the least-squares method.^{55,56,61} However, such a simple procedure becomes unstable because the number of equations, N , far exceeds the number of unknowns, f_k . Ridge regression, adopted in the program CONTIN,⁶² fits the analyzed CD spectrum (C_λ^{obs}) as a linear combination of the reference CD spectra by minimizing the function:

$$\sum_\lambda (C_\lambda^{calc} - C_\lambda^{obs})^2 + \alpha^2 \sum_{j=1}^N (v_j - N^{-1})^2 \quad (21)$$

where α is the regularizer and v_j is the coefficient of the CD spectrum for the j^{th} reference protein used to construct the analyzed spectrum, C_λ^{calc} . The secondary structures corresponding to the analyzed CD spectrum are calculated using the coefficients v_j with the secondary structures of protein j :

$$f_k^{calc} = \sum v_j f_k^j \quad (22)$$

One can also formulate this problem as a matrix equation by constructing a CD spectral matrix \mathbf{C} of dimension $M \times N$, with M CD data points, and a structural matrix \mathbf{F} of dimension $K \times N$, with K secondary structures.

$$\mathbf{C} = \mathbf{B} \mathbf{F} \quad (23)$$

The inverse of matrix \mathbf{B} , which has a dimension $K \times M$, can be used with any C_λ for estimating f_k (represented as vectors \mathbf{C} and \mathbf{f} of dimensions M and K , respectively).⁶³

$$\mathbf{B}^{-1} \mathbf{C} = \mathbf{F} \quad \text{or} \quad \mathbf{B}^{-1} \mathbf{C} = \mathbf{f} \quad (24)$$

The matrix equation can be solved using singular value decomposition⁵⁷ or factor analysis⁶⁴ techniques, which are mathematically similar. In the singular value decomposition algorithm, one writes matrix \mathbf{C} as a product of three matrices, $\mathbf{U} \mathbf{S} \mathbf{V}^T$, where \mathbf{U} and \mathbf{V} are unitary matrices ($\mathbf{U}\mathbf{U}^T = \mathbf{I}$, where matrix \mathbf{U}^T is the transpose of matrix \mathbf{U} and \mathbf{I} is the identity matrix), and \mathbf{S} is a diagonal matrix. The diagonal elements of matrix \mathbf{S} are the singular values, and their magnitudes determine their importance in the analysis; smaller values generally correspond to experimental noise and are ignored in the analysis. Eq 24 is solved by determining \mathbf{B}^{-1} as a multiplicative inverse of the matrix \mathbf{B} , written as a matrix product: $\mathbf{F} \mathbf{V} \mathbf{S}^+ \mathbf{U}^T$, where \mathbf{S}^+ is the inverse of matrix \mathbf{S} .⁶⁵

Alternatively, one can also consider a given pair of CD spectrum and corresponding secondary structures as a pattern, $C_\lambda \leftrightarrow f_k$ and create N such patterns for an N -protein reference set. Pattern recognition techniques, such as

neural networks, are utilized to determine information flow from the spectral pattern to the structure pattern, which can then be used with any given spectral pattern.⁶⁶⁻⁶⁸

The protein secondary structures used are determined from their crystal structures, which are available in the Protein Data Bank as atomic coordinates. The secondary structure content is determined by an unbiased assignment of the secondary structures, using certain geometric rules that define a given structure. Of the many algorithms that are in use for assigning secondary structures to a set of coordinates, the DSSP⁶⁹ method is the most popular. Structures assigned by DSSP include α -, 3_{10} -, and π - helices, β -sheets, β -turns, and S-bends; α -, 3_{10} -, and π -helices are combined to obtain the α -helical fraction, and β -turns and S-bends are combined for the turns fraction. One can consider four or two residues, respectively, for each α -helical or β -strand segment as distorted, as an approximation for end-effects, and further split the α - and β -fractions into distorted and regular fractions.⁷⁰ Algorithms for assigning the P₂ structure are also available.^{31,71} The protein secondary structures that are included in CD analysis are: regular and distorted α -helices, regular and distorted β -strands, turns, P₂ structure, and unordered or remainder.

The construction of a good reference protein set that represents the spectral and structural variations present in proteins is very important for a successful protein CD analysis. The characteristics of protein CD spectra and their relation to secondary structure have been discussed in the previous section. The dominating effect of α -helical CD contributions in proteins is well documented. In comparison to α -helices, other secondary structures not only have weaker CD but also have greater structural variability. Presently, a 42-protein reference set that was created by combining different reference protein sets that were developed independently by three research groups is available for the analysis of protein CD spectra.⁷² Inclusion of a large number of proteins, belonging to different tertiary structure classes and with varying secondary structure contents, is expected to provide a good representation of the spectral and structural variability in proteins. A larger 48-protein reference set that includes six denatured proteins is also available⁷² and can be used in specific applications requiring the characterization of the unordered state.

Using all proteins in the reference set and a mathematical method based on the formulation of Eq. 19 one obtains a solution for a given CD spectrum C_λ , which gives the corresponding secondary structure fractions f_k . Ideally, the values of f_k thus obtained should be positive, the sum of f_k should be unity, and the reconstructed CD spectrum should match C_λ . These ideal conditions are rarely met. The additivity of individual secondary structure contributions involved in the linear approximation of Eq. 19 assumes the equivalence of ensemble-averaged solution and the time-averaged solid-state structures, and ignores the contributions from non-peptide chromophores and the effects of geometric variability of secondary structure and of tertiary structure. The nonideality of the situation often results in negative f_k and/or $\sum f_k$ not being unity, and one can use these criteria as selection rules for the validity of the analysis. They are used as constraints in some least-squares methods.^{61,62} By and large, the assumptions involved in protein CD analysis are valid, but a poor or inadequate representation of the analyzed spectrum in the reference set results in a failed analysis (nonideal f_k values) or erroneous results. The inadequacies of the assumptions and their effects on protein CD analysis are overcome by relaxing the physical constraints and/or introducing flexibility in the analysis.

The principle of selecting a limited number of proteins from the reference set specifically for the analysis of a given CD spectrum, thus creating one or more subsets, is embodied in the variable selection method.⁷³ It effectively creates reference sets that include proteins that are important for the analysis and excludes proteins that adversely

affect the analysis. *A priori*, one does not know whether a given protein needs to be included or excluded in the analysis, and the selection is accomplished differently by different algorithms. In the basic variable selection method⁷³ one or more proteins are deleted from the reference set sequentially; removal of one protein results in N reference protein sets of size $N-1$; removal of two proteins gives $N \times (N-1)/2$ reference sets of size $N-2$; and so on. This results in a large number of solutions, with one solution from each reference set, and performing variable selection exhaustively becomes computationally prohibitive. Proteins in the reference set can also be removed based on the similarity of their CD spectra with the analyzed spectrum, with the least similar protein being removed first and a few most similar proteins always included in the analysis. This implementation is referred to as the locally linearized model⁷⁴ and is computationally much less intensive. One can also select a specific number of proteins, randomly chosen from the reference protein set, to create a minimal basis.⁷⁵ Since this approach generates a very large number of combinations of proteins the process is stopped once an acceptable number of valid solutions is obtained.⁷⁶ With the knowledge of the tertiary structure class (α -rich, β -rich, or $\alpha\beta$) of the analyzed protein, one can use tertiary class-specific reference proteins⁷⁷ for CD analysis. Proteins in the reference set can also be weighted variably in a least-squares fitting of the analyzed CD spectrum,⁶² as is done in the ridge regression technique implemented in the computer program CONTIN. Variable selection has been explicitly introduced⁷² in CONTIN via the locally linearized implementation (CONTIN/LL).

The inclusion of the analyzed protein in the reference set improves the quality of $B_{k\lambda}$ and the subsequent analysis, but the corresponding f_k values are unknown. The self-consistent method⁷⁸ resolves this problem by making a guess for the unknown structure, the secondary structure of the protein in the reference set that is spectrally most similar to the analyzed protein. Solution of the matrix equation gives a new set of f_k values that replace the initial guess and the process is iterated for convergence.

Implementation of variable selection invariably gives a large number of solutions. The validity of a given solution is generally measured by the relaxed selection rules, as applicable:⁵⁷⁻⁶⁰ *Sum rule*, $|\sum f_k - 1.0| \leq 0.05$; *Fraction rule*, $f_k \geq -0.025$; *Spectral rule*, error between the experimental and reconstructed spectra is $\leq 0.25 \Delta\epsilon$; *Helix rule*,⁷⁶ the difference between the helix fractions obtained with the all-protein reference set and the variably selected reference set is ≤ 0.25 . The first two selection rules are derived from the physical constraints of f_k , the third follows Eq. 19, and the fourth is a consequence of the dominance of α -helical CD in proteins. Generally, all valid solutions are averaged to obtain the secondary structure fractions f_k corresponding to a given CD spectrum.

Two approaches for introducing variable selection, locally linearized and minimal basis approaches, and the self-consistent method are combined with the singular value decomposition algorithm for solving matrix equation in two popular computer programs (Selcon3 and CDsstr) for protein CD analysis. The three computer programs Selcon3, CDsstr, and CONTIN/LL, provided with multiple reference protein sets with different wavelength ranges, have been made available with the CDPro software package (Internet: <http://lamar.colostate.edu/~sreeram/CDPro>). These programs and the reference sets are also available for use at the internet-based CD analysis site DICHROWEB (<http://www.cryst.bbk.ac.uk/cdweb/>). Computer programs that use neural network methods are also available for protein CD analysis (http://bioinformatik.biochemtech.uni-halle.de/cd_spec/), but these do not include variable selection and the resulting flexibility.

The success of an analytical method is generally measured by the accuracy of the results and reliability of the analysis. The accuracy of protein CD analysis is determined by the performance of the method for the members of a given reference set by cross-validation. CD estimates of the secondary structure fractions of each protein in the reference set are obtained by removing it from the reference set and analyzing its CD spectrum using the other members of the reference set. The performance of a method is assessed statistically using RMS deviations (δ) and correlation coefficients (r) between CD and x-ray estimates of secondary structures. Such a comparison of the performance of the three programs and three reference sets from CDPro⁷² is given in Table I, where δ and r for each

Table I. Performance of SELCON3, CDSSTR and CONTIN/LL programs, for analyzing protein CD spectra of 29 proteins in the wavelength range 190–240 nm.

Reference Proteins	Method	α_R		α_D		β_R		β_D		T		U			
		δ	r	δ	r	δ	r	δ	r	δ	r	δ	r		
29	SELCON3	0.052	0.949	0.053	0.689	0.102	0.547	0.036	0.709	0.075	0.302	0.118	0.268	0.078	0.773
	CDSSTR	0.059	0.938	0.052	0.785	0.083	0.655	0.030	0.790	0.074	0.337	0.097	0.491	0.070	0.817
	CONTIN/LL	0.058	0.936	0.055	0.679	0.102	0.486	0.035	0.719	0.074	0.323	0.103	0.317	0.075	0.784
37	SELCON3	0.047	0.960	0.050	0.715	0.094	0.638	0.036	0.704	0.063	0.538	0.116	0.142	0.073	0.795
	CDSSTR	0.059	0.939	0.047	0.811	0.087	0.648	0.030	0.801	0.066	0.452	0.098	0.413	0.069	0.819
	CONTIN/LL	0.054	0.944	0.052	0.706	0.093	0.624	0.033	0.753	0.066	0.447	0.095	0.328	0.069	0.813
43	SELCON3	0.051	0.953	0.048	0.747	0.086	0.659	0.034	0.746	0.073	0.382	0.110	0.181	0.072	0.802
	CDSSTR	0.064	0.929	0.042	0.792	0.081	0.704	0.028	0.843	0.067	0.462	0.089	0.444	0.065	0.833
	CONTIN/LL	0.053	0.942	0.048	0.756	0.084	0.674	0.031	0.781	0.076	0.373	0.096	0.262	0.069	0.817

^a The 29-proteins form the reference protein set in the wavelength range 178–260 nm and forms a subset of 37-protein and 43-protein reference sets, which have smaller wavelength ranges of 185–250 and 190–240 nm, respectively. The results presented in this table illustrate the improvements in the performance with increasing number of reference proteins. The secondary structures are: α_R , regular α -helix; α_D , distorted α -helix; β_R , regular β -strand; β_D , distorted β -strand; T, turns; and U, unordered.

secondary structure, separately and collectively, are given. These performance indices give an estimate of the accuracy of either a given method (e.g., δ and r , for CDSstr) or a given secondary structure (e.g., δ_T and r_T , for turns). On the whole, CDSstr performed better than the other two methods but the difference is mainly due to improved performance for the unordered fraction; results for α , β , and turn fractions from all three methods are comparable. Also, the performance of a given method improves with the size of the reference set due to better representation of spectral

variations. The perceived difficulty in the analysis of β -rich proteins with two types of characteristic CD spectra⁵¹ is presumably overcome by the use of a larger reference set that provides adequate representation of both types of spectra, and variable selection which allows the inclusion of similar proteins.⁵² The similarity of the performance of all three methods, despite the differences in the mathematical techniques or the means of implementation of variable selection, for α , β , and turn fractions gives a measure of the reliability of analysis. Improvements in both accuracy and reliability in the estimation of secondary structure from protein CD spectra are achieved by the use of multiple methods of analysis combined with a larger reference protein set.⁷²

Other structural information

Methods have also been developed to estimate the number of segments of secondary structures and the tertiary structure class from protein CD spectrum. A matrix-description of secondary structure segments,⁷⁹ with the number of α -helix, β -sheet, and coil segments forming the diagonal elements and the junctions between them forming the off-diagonal elements, has been used in conjunction with a neural network-based analysis of the protein CD spectra to estimate the number of α - and β -segments. The number of distorted residues in α -helices and β -sheets obtained from CD analysis has also been used to estimate the number of segments in a given protein,⁷⁰ considering four and two residues distorted in an α -helix and a β -sheet, respectively. The results from these two methods were comparable. The latter method is implemented in CDPro.

The tertiary structure class of a protein can be identified from its CD spectrum by representing proteins as vectors in a multidimensional space of CD data.⁸⁰ In such a hyperspace, proteins belonging to different tertiary structure classes form clusters separated by hyperplanes. The equations of the hyperplanes separating the different clusters are used to assign any new CD spectrum to a tertiary structure class (α -rich, β -rich, $\alpha\beta$, or denatured). A computer program, CLUSTER, which identifies the tertiary structure class and creates a reference set specific to the determined tertiary class, is available in CDPro. Such a tertiary class-specific reference set⁷⁷ provides limited but specific information for the analysis of protein CD spectra, e.g., an α -rich protein set for analyzing an α -rich protein CD, and improves the analysis.

Aromatic contributions

Aromatic contributions can be considered weak in comparison to those from backbone amides.⁸¹ However, changes in the environment of aromatic side chains in proteins caused by external perturbations, such as ligand binding, mutations, or experimental conditions, can lead to detectable changes in their CD signals. Such contributions⁸¹⁻⁸³ are generally monitored in the near-UV region, where backbone amides do not contribute. The analysis is, however, of a qualitative nature although in some proteins aromatic CD bands can be assigned to specific residues to extract information about their environment. Protein-folding studies often utilize the changes in aromatic CD as an indicator of tertiary interactions.

In some proteins, even in the far UV, aromatic side chains make significant contributions. This is especially true for proteins with low α -helix content, such as lectins, immunoglobulins and snake toxins. The positive CD bands in the 225–235 nm region reported for many proteins⁸⁴ are certainly due to Tyr or Trp side chains or to disulfides,

since the amide contributions in this region are generally negative, and these side-chain contributions can interfere with protein CD analysis for secondary structure estimation. Although the protein CD analysis methods do not explicitly consider aromatic contributions, they are implicit in the analysis as a part of spectral variations represented in the reference set. Both aromatic contributions and secondary structure variations are accommodated in the analysis by the flexibility of the analysis, variably selecting proteins specific for a given CD spectrum. In the extreme case, where the anomalous aromatic contribution is not represented in the reference set, even flexible methods either give poor results or fail. Such situations can be identified with the use of multiple methods of analysis that give dissimilar results, and may be corrected by employing other spectroscopic techniques such as VCD, IR, etc. Attempts to explicitly incorporate or extract aromatic CD contributions from a reference protein set⁸⁵ have been of limited value, due to lack of structural patterns and characteristic CD bands, and cancellation of contributions from different aromatic side chains.

Aromatic side chains often form interacting pairs or clusters in proteins⁸⁶ and the coupled oscillator interactions between them is a source of aromatic CD in proteins, which explains the absence of correlation between the number of aromatic side chains and aromatic CD bands. Coupling of two aromatic groups depends on the distance separating them and the relative orientation of the aromatic rings and, normally, shorter distances lead to stronger CD signals.⁸⁷ In spite of the apparent randomness of the aromatic CD bands due to the nature of the coupling, analysis of aromatic contributions can be performed by a combination of experimental and computational tools. Experimentally, contributions from an aromatic residue can be inferred by replacing it by a non-aromatic amino acid or by an aromatic

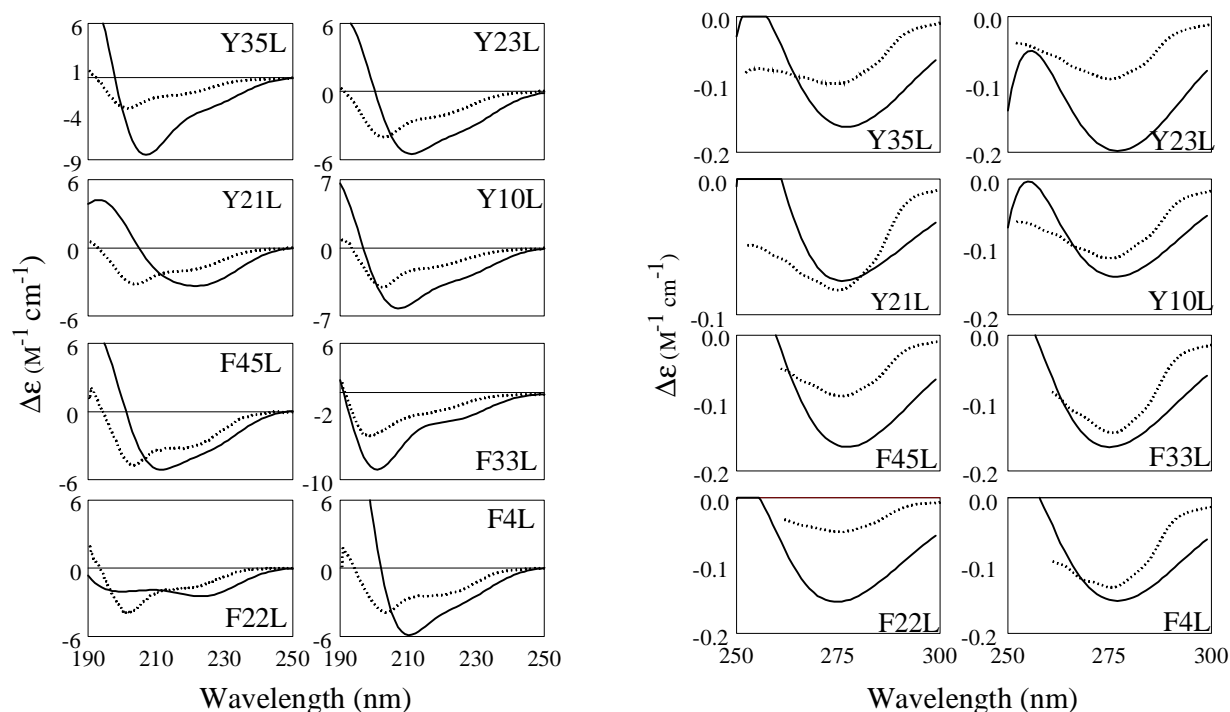


Figure 5. Comparison of calculated (including all aromatic side chains and peptide groups, solid curve) and experimental (dotted curve) CD spectra for the Y→L and F→L mutants of bovine pancreatic trypsin inhibitor.⁸⁸ (A) Far-UV CD spectra. (B) Near-UV CD spectra.

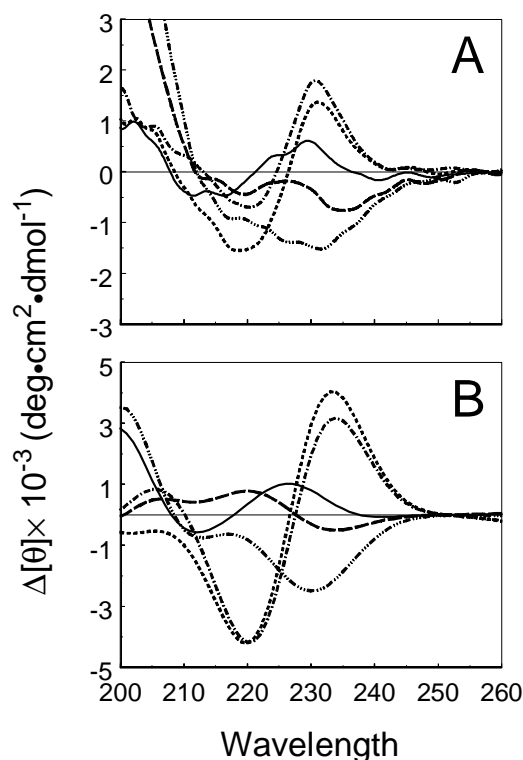


Figure 6. Comparison of experimental and theoretical difference CD spectra (wild-type – mutant) for dihydrofolate reductase. Panel A: Experimental difference spectra⁸⁹ for W22L (—), W30L (---), W47L (·····), W74L (— · — ·), W133V (— · — · — ·). Panel B: Theoretical difference spectra for W22A (—), W30A (---), W47A (·····), W74A (— · — ·), W133A (— · — · — ·). The mutant structures used in theoretical calculations were obtained by replacing the side chain of Trp by that of Ala in the DHFR crystal structure (PDB code: 4dfr, molecule B). Theoretical CD spectra were obtained using all aromatic side chains and peptide groups.

amino acid with a weaker chromophore, and examining the difference CD spectra (wild-type CD – mutant CD). Using computational tools one can perform a similar experiment in the computer to obtain aromatic contributions theoretically, assuming that the structure does not change with the mutation. Such analyses have been performed for bovine pancreatic trypsin inhibitor (BPTI),⁸⁸ bovine pancreatic ribonuclease,^{28,82} dihydrofolate reductase (DHFR),⁸⁹ human interleukin 1 β ,⁹⁰ barnase,⁹¹ carbonic anhydrase,⁹² and gene 5 protein.⁹³ In most cases, the experimental difference CD spectra and theoretically calculated difference CD spectra are in good agreement, implying structural similarity of the wild-type and the mutant proteins. However, the difficulties involved in modeling the structural changes due to the disruption of an aromatic cluster may lead to disagreement between theory and experiment.

Two examples of combined experimental and theoretical studies of mutants are given in Figures 5 and 6. The theoretical CD spectra of Y→L and F→L mutant BPTI agree reasonably well with experiment⁸⁸ (Figure 5). In the near UV, theory predicts a negative CD band for all mutants, in agreement with the experiment. In the far UV, however, experimental and theoretical CD spectra show some disagreements that can be correlated to the aromatic side-chain interactions in different aromatic clusters in the BPTI structure.⁸⁸ The theoretical difference CD spectra⁸⁷ (wild-type – mutant) of specific mutants of DHFR (Figure 6) are similar to those obtained experimentally.⁸⁹ In DHFR, two Trp side

chains, W74 and W47, are in close proximity and give rise to strong coupling. Such a coupling is absent in the W74F and W47F mutants, so the interaction between W74 and W47 in the wild-type protein shows up as a positive couplet (Figure 6) in both the theoretical and experimental difference CD spectra.⁸⁹ For the other three mutants, the theoretical difference CD spectra (Figure 6B) show features observed in the experimental spectra (Figure 6A). Molecular dynamics techniques can be used to model the local structural changes caused by a mutation, but large conformational changes may be difficult to reproduce.

Extrinsic chromophores

Protein CD bands at wavelengths longer than 300 nm, in the near-UV and visible regions, are attributed to extrinsic chromophores. These are not part of the polypeptide chain, but bind to proteins either covalently or noncovalently. Extrinsic chromophores by themselves, in the free state, may not show optical activity because they are either achiral or exist as enantiomeric mixtures, but their interaction with the chiral environment of the protein matrix upon binding generates optical activity. The protein matrix may prefer a particular conformation of the chromophore, which may make it chiral and lead to intrinsic optical activity, or it may choose one enantiomer, enhancing its CD signal. The electronic transitions on the extrinsic chromophore in the bound state, in addition to having intrinsic CD, can also generate optical activity by mixing with transitions on protein chromophores by coupled-oscillator interactions. The $n\pi^*$ and $\pi\pi^*$ transitions on the extrinsic chromophore can also mix in the electrostatic field of the protein. An analysis of extrinsic contributions thus requires the knowledge of both the extrinsic chromophore conformation and the protein structure. Such analyses often involve a combination of computational and experimental studies in obtaining information about the conformation of extrinsic chromophore and the binding pocket.

The origin of the Soret CD band in heme proteins has been extensively studied.⁹⁴⁻⁹⁸ The Soret band results from two degenerate, fully allowed $\pi\pi^*$ transitions of the porphyrin that appear in the 400 nm region. The coupled-oscillator interactions between the $\pi\pi^*$ transitions on the porphyrin and those in the protein aromatic side chains were identified as the source of positive Soret CD in sperm-whale and horse metmyoglobins, and the calculated CD spectra agreed with experiment in sign and approximately in magnitude.⁹⁴ Coupled-oscillator calculations in hemoglobins from *Chironomus thummi thummi*⁹⁵ and lamprey predicted negative Soret bands, in agreement with experiment.⁹⁶ However, the coupled-oscillator mechanism alone can not explain the observed CD differences between the two isomers of heme that differ by a 180° rotation about the α - γ axis. In the carbonmonoxy form of sperm-whale myoglobin the two isomers exist in a 9:1 ratio. The major form has a strong positive band while the minor form has a weak negative band.⁹⁷ A simple rotation about the in-plane α - γ axis, giving rise to heme isomerism, does not change the net rotational strength for a planar heme in the coupled-oscillator picture. The solid-state and solution structures correspond to the major form, and attempts to model the minor form using computer simulations have not been successful. Molecular dynamics studies of the major form give a picture of non-planar distortions of the heme and its interactions with the protein matrix, which are believed to be the source of Soret CD.⁹⁸ The intrinsic rotational strengths of the two Soret transitions, separated according to their polarization in the heme plane, show strong correlations with the ruffling deformation,⁹⁹ in which the opposite pyrrole rings are counter rotated, bringing the alternate *meso* carbons above and below the mean plane, and the average twist angle of the pyrrole rings, which

measures the counter rotation of the opposite pyrroles. Rotational strengths generated by the mixing of heme transitions with both aromatic and peptide transitions were comparable to the intrinsic Soret rotational strengths.

The extrinsic CD in visual rhodopsins¹⁰⁰ has two positive bands, α and β , from the chromophore 11-*cis* retinal linked to a lysine side chain via a protonated Schiff base (PSB). The extrinsic chromophore in bacterial rhodopsins¹⁰⁰ is all-*trans* retinal-PSB and it gives rise to a positive α and a negative β band. Theoretical calculations of retinal CD were performed¹⁰¹ using the available crystal structures of both visual and bacterial rhodopsins, and the calculated rotational strengths for the two bands agree reasonably well with the experiment, reproducing the pattern of signs for the α and β bands in all five retinal-binding proteins studied. However, the calculated intrinsic contributions from retinal-PSB, in both visual and bacterial rhodopsins, were generally small owing to small non-planar distortions along the polyene chain. The dominant contribution to the rotational strength was calculated from the coupling of retinal and aromatic transitions. The non-planar distortions in retinal-PSB are the source of its intrinsic CD, which may have been suppressed by crystal interactions.

References

1. L. Rosenfeld, *Z. Physik* **52**, 161 (1928).
2. W. Moffitt, *J. Chem. Phys.* **25**, 467 (1956).
3. A. Moscowitz, in: "Modern Quantum Chemistry - Istanbul Lectures" (O. Sinanoğlu, Ed.) Part III, p.31. Academic Press, New York (1965).
4. W. Moffitt and A. Moscowitz, *J. Chem. Phys.* **30**, 648 (1959).
5. J.G. Kirkwood, *J. Chem. Phys.* **5**, 479 (1937).
6. J.A. Schellman, *Acc. Chem. Res.* **1**, 144 (1968).
7. E.U. Condon, W. Altar, and H. Eyring, *J. Chem. Phys.* **5**, 753 (1937).
8. R.W. Woody, in: "Circular Dichroism and the Conformational Analysis of Biomolecules" (G.D. Fasman, Ed.), p.25. Plenum Press, New York (1996).
9. H. DeVoe, *J. Chem. Phys.* **41**, 393 (1964).
10. H. DeVoe, *J. Chem. Phys.* **43**, 3199 (1965).
11. A. Moscowitz, *Adv. Chem. Phys.* **4**, 67 (1962).
12. J. Applequist, *J. Chem. Phys.* **71**, 1983 (1979).
13. J. Applequist, *J. Chem. Phys.* **71**, 4332 (1979).
14. J. Applequist, *Biopolymers* **20**, 2311 (1981).
15. J. Applequist and K. Bode, *J. Phys. Chem. B* **103**, 1767 (1999).
16. K. Bode and J. Applequist, *J. Am. Chem. Soc.* **120**, 10938 (1998)
17. P.M. Bayley, E.B. Nielsen, and J.A. Schellman, *J. Phys. Chem.* **73**, 228 (1969).
18. P.M. Bayley, *Prog. Biophys. Mol. Biol.* **27**, 1 (1973).
19. I. Tinoco, Jr., *Adv. Chem. Phys.* **4**, 113 (1962).
20. C.R. Cantor and S.N. Timasheff, in: "The Proteins" Vol V, (H. Neurath, Ed.) p.145. Academic Press (1982).
21. M.C. Manning and R.W. Woody, *Biopolymers* **31**, 569 (1991).

22. R.W. Woody, *J. Chem. Phys.*, **49**, 4797 (1968).
23. L. Serrano-Andrés and M.P. Fülscher, *J. Am. Chem. Soc.* **118**, 12190 (1996).
24. L.B. Clark, *J. Am. Chem. Soc.* **117**, 7974 (1995).
25. J.R. Platt, *J. Chem. Phys.* **17**, 484 (1949).
26. J.N. Murrell and J.A. Pople, *Proc. Phys. Soc.* **A69**, 245 (1956).
27. B. Albinsson and B. Norden, *J. Phys. Chem.* **96**, 6204 (1992).
28. G. Kurapkat, P. Krüger, A. Wollmer, J. Fleischhauer, B. Kramer, E. Zobel, A. Koslowski, H. Botterweck, and R.W. Woody, *Biopolymers* **41**, 267 (1997).
29. R.W. Woody, *Tetrahedron* **29**, 1273 (1973).
30. C.M. Wilmot and J.M. Thornton, *Protein Eng.* **3**, 479 (1990).
31. N. Sreerama and R.W. Woody, *Biochemistry* **33**, 10022 (1994).
32. A. Toumadje, S.W. Alcorn, and W.C. Johnson, Jr., *Anal. Biochem.* **200**, 321 (1992).
33. W.C. Johnson, Jr. and I. Tinoco, Jr., *J. Am. Chem. Soc.* **94**, 4389 (1972).
34. P.J. Gans, P.C. Lyu, M.C. Manning, R.W. Woody, and N.R. Kallenbach, *Biopolymers* **31**, 1605 (1991).
35. D-H. Chin, R.W. Woody, C.A. Rohl, and R.L. Baldwin, *Proc. Natl. Acad. Sci.*, **99**, 15416 (2002).
36. S. Brahm, J. Brahm, G. Spach, and A. Brack, *Proc. Natl. Acad. Sci. USA* **74**, 3208 (1977).
37. B.A. Wallace and R.W. Janes, *Curr. Opin. Chem. Biol.* **5**, 567 (2001).
38. M.C. Manning, M. Illangasekare, and R.W. Woody, *Biophys. Chem.* **31**, 77 (1988).
39. G.D. Jenness, C. Sprecher, and W.C. Johnson, Jr., *Biopolymers* **15**, 513 (1976).
40. M.L. Tiffany and S. Krimm, *Biopolymers*. **6**, 1379 (1968).
41. A.A. Adzhubei and M. Sternberg, *J. Mol. Biol.* **229**, 472 (1993).
42. B.J. Stapley and T.P. Creamer, *Protein Sci.* **8**, 587 (1999).
43. A. Perczel and M. Hollosi, in: "Circular Dichroism and the Conformational Analysis of Biomolecules" (G.D. Fasman, Ed.), p.285. Plenum Press, New York (1996).
44. W. Moffit, D.D. Fitts, and J.G. Kirkwood, *Proc. Natl. Acad. Sci. USA* **43**, 723 (1957).
45. V. Madison and J.A. Schellman, *Biopolymers* **11**, 1041 (1972).
46. J.D. Hirst, *J. Chem. Phys.* **109**, 782 (1998).
47. N.A. Besley and J.D. Hirst, *J. Phys Chem. A* **102**, 10791 (1998).
48. J.D. Hirst, S. Bhattacharjee, and A.V. Onufriev, *Faraday Discuss.* **122**, 253 (2002).
49. R.W. Woody and N. Sreerama, *J. Chem. Phys.* **111**, 2844 (1999).
50. A. Koslowski, N. Sreerama, and R.W. Woody, in: "Circular Dichroism: Principles and Applications" 2nd ed. (Ed. N. Berova, K. Nakanishi, and R.W. Woody), p. 55, Wiley, New York. (2000).
51. J. Wu, J.T. Yang, and C-S.C. Wu, *Anal. Biochem.* **200**, 359 (1992).
52. N. Sreerama and R.W. Woody, *Protein Sci.* **12**, 384 (2003).
53. J. Fleischhauer, J. Grötzinger, B. Kramer, P. Krüger, A. Wollmer, R.W. Woody, and E. Zobel, *Biophys. Chem.* **49**, 141 (1994).
54. A. Koslowski, H. Botterweck, J. Fleischhauer, G. Kurapkat, A. Wollmer, and R.W. Woody, *Prog. Biophys. Mol. Biol.* **65**, suppl. 1, p. 43. (1996).

55. N.J. Greenfield and G.D. Fasman, *Biochemistry* **8**, 4108 (1969).
56. J.T. Yang, C-S. C. Wu, and H.M. Martinez, *Methods Enzymol.* **130**, 208 (1986)
57. W.C. Johnson, Jr., *Annu. Rev. Biophys. Biophys. Chem.* **17**, 145 (1988).
58. S.Y. Venyaminov and J.T. Yang, in: "Circular Dichroism and the Conformational Analysis of Biomolecules" (Ed. GD Fasman), p.69. Plenum, New York (1996).
59. N.J. Greenfield, *Anal. Biochem.* **235**, 1 (1996).
60. N. Sreerama, and R.W. Woody, in: "Circular Dichroism: Principles and Applications" 2nd ed. (N. Berova, K. Nakanishi, and R.W. Woody, Eds), p. 601. Wiley, New York. (2000).
61. Y.H. Chen, J.T. Yang, and H.M. Martinez, *Biochemistry* **11**, 4120 (1972).
62. S.W. Provencher and J. Glöckner, *Biochemistry* **20**, 33 (1981).
63. J.P. Hennessey, Jr. and W.C. Johnson, Jr., *Biochemistry* **20**, 1085 (1981).
64. P. Pancoska and T.A. Keiderling, *Biochemistry* **30**, 6885 (1991).
65. L.A. Compton and W.C. Johnson Jr., *Anal Biochem* **155**, 155 (1986).
66. G. Böhm, R. Muhr, and R. Jaenicke, *Protein. Eng.* **5**, 191 (1992).
67. M.A. Andrade, P. Chacán, J.J. Merolo, and F. Morán, *Protein Eng.* **6**, 383 (1993).
68. N. Sreerama and R.W. Woody, *J. Mol. Biol.* **242**, 497 (1994).
69. W. Kabsch and C. Sander, *Biopolymers* **22**, 2577 (1983).
70. N. Sreerama, S.Y. Venyaminov, and R.W. Woody, *Protein Sci.* **8**, 370 (1999).
71. S.M. King and W.C. Johnson, Jr., *Proteins: Struct. Func. Genet.* **35**, 313 (1999).
72. N. Sreerama and R.W. Woody, *Anal. Biochem.* **287**, 252 (2000).
73. P. Manavalan, W.C. Johnson, Jr., *Anal. Biochem.* **167**, 76 (1987).
74. I.H.M. van Stokkum, H.J.W. Spoelder, M. Bloemendal, R. van Grondelle, and F.C.A. Groen, *Anal. Biochem.* **191**, 110 (1990).
75. B. Dalmas and W. H. Bannister, *Anal. Biochem.* **225**, 39 (1995).
76. W.C. Johnson, Jr., *Proteins: Struct. Func. Genet.* **35**, 307 (1999).
77. N. Sreerama, S.Y. Venyaminov, and R.W. Woody, *Anal. Biochem.* **299**, 271 (2001)
78. N. Sreerama and R.W. Woody, *Anal. Biochem.* **209**, 32 (1993).
79. P. Pancoska, V. Janota, and T.A. Keiderling, *Anal. Biochem.* **267**, 72 (1999).
80. S.Y. Venyaminov and K.S. Vassilenko, *Anal. Biochem.* **222**, 176 (1994).
81. R.W. Woody and A.K. Dunker, in: "Circular Dichroism and the Conformational Analysis of Biomolecules" (G.D. Fasman, Ed.), p.109. Plenum Press, New York (1996).
82. E.H. Strickland, *CRC Crit. Rev. Biochem.* **2**, 113 (1974).
83. P.C. Kahn, *Meth. Enzymol.* **61**, 339 (1979).
84. R.W. Woody, *Eur. Biophys. J.* **23**, 253 (1994).
85. I.A. Bolotina and V.Y. Lugauskas, *Mol. Biol. (Eng. Transl. of Molekul. Biol.)* **19**, 1154 (1985).
86. A. Thomas, R. Meurisse, B. Charlotteaux, and R. Brasseur, *Proteins: Struct. Func. Genet.* **48**, 628 (2002).
87. I.B. Grishina and R.W. Woody, *Faraday Discuss.*, **99**, 245 (1994).

88. N. Sreerama, M.C. Manning, M.E. Powers, J. Zhang, D.M. Goldenberg, and R.W. Woody, *Biochemistry* **38**, 10814 (1998).
89. E. Ohmae, Y. Sasaki, and K. Gekko, *J. Biochem.* **130**, 439 (2001).
90. S. Craig, R.H. Pain, U. Schmeissner, R. Virden, and P.T. Wingfield, *Int. J. Pept. Protein Res.* **33**, 256 (1989).
91. S. Vuilleumier, J. Sancho, R. Loewenthal, and A.R. Fersht, *Biochemistry*, **32**, 10303 (1993).
92. P.-O. Freskgård, L.-G. Mårtensson, P. Jonasson, B.-H. Jonsson, and U. Carlsson, *Biochemistry* **33**, 14281 (1994).
93. T.M. Thompson, B.L. Mark, C.L. Gray, T.C. Terwilliger, N. Sreerama, R.W. Woody, and D.M. Gray, *Biochemistry* **37**, 7463 (1998).
94. M.-C. Hsu and R.W. Woody, *J. Am. Chem. Soc.* **93**, 3515 (1971).
95. J. Fleischhauer and A. Wollmer, *Zeit. Naturforsch.* **27b**, 530 (1972).
96. R.W. Woody, in: "Protein-Ligand Interactions" (Sund, H., and Blauer, G., Eds.), p.60, de Gruyter, Berlin (1975).
97. H.S. Aojula, M.T. Wilson, G.R. Moore, and D.J. Williamson, *Biochem. J.* **250**, 853 (1988).
98. C. Kiefl, N. Sreerama, R. Haddad, L. Sun, W. Jentzen, Y. Lu, Y. Qiu, J.A. Shelnutt, and R.W. Woody, *J. Am. Chem. Soc.* **124**, 3385 (2002).
99. J.A. Shelnutt, X.Z. Song, J.G. Ma, S.L. Jia, W. Jentzen, and C.J. Medforth, *Chem. Soc. Rev.* **27**, 31 (1998)
100. R.R. Birge, *Biochim. Biophys. Acta* **1016**, 293 (1990).
101. N. Sreerama, R.R. Birge, V. Buß, B.S. Hudson, D. Singh, and R.W. Woody, *Biophysical J.* **82**, 224A (2002).

Acknowledgements

The work of our research group described this review article was supported by an NIH grant, GM22994. We thank Dr. K. Gekko, Hiroshima University, Hiroshima, for providing the experimental CD data of wild-type and mutant DHFR.