CHEM 110L: EXP. 07

Introductory Biochemistry Laboratory

Prelab: Circular Dichroism Characterization of Proteins

Heather MacGregor

Department of Chemistry and Biochemistry
UC Santa Barbara
Santa Barbara, CA, 93106

The scientific goal of this experiment is to determine the percentage of different secondary structure elements in a folded based on the electronic vibrational circular dichroism spectrum of the dissolved protein. The educational goals of this experiment are i) to expose you to spectropolrimetric techniques, ii) to allow you to review of protein secondary structures, and ii) the application of linear algebra to multi-wavelength data.

1 | INTRODUCTION

Circular Dichroism (CD) measures the differential absorbance of left- and right-handed CPL by optically active, or chiral, compounds. This phenomenon is known as circular dichroism (CD):

$$\Delta\varepsilon(\vec{\nu}) = \varepsilon^{L}(\vec{\nu}) - \varepsilon^{R}(\vec{\nu}) \tag{1}$$

Similarly to absorbance in UV-Vis spectrophotometry, the observed differential absorbance is dependent on the concentration of the sample in solution and the path length of the cuvette. The quantity $\Delta\varepsilon$ is known as the molar circular dichroism and characterizes the chiral molecule.

Although the information obtained from CD techniques is limited in comparison to X-ray crystallography or NMR, it is a well-established tool for the investigation of secondary structural elements of proteins. CD has a number of key advantages, including its ability to explore proteins in a wide range of solution conditions and temperatures, quick data collection, and consumption of relatively small amounts of sample.

Protein secondary structural information can be derived from CD signals in the far ultraviolet (UV) wavelength region between 190 and 240 nm that arise from numerous $\pi \rightarrow \pi*$ (222 nm) and $n \rightarrow \pi*$ (both parallel and perpendicular orientations; 208 and 190 nm) electronic transitions of the peptide backbone, which appear around. The intensities and signs of the CD bands produced from these overlapping transitions are strongly dependent on the secondary structure of the protein (Figure 5). Inspection of the spectrum between 200–230 nm often reveals the dominant secondary structure in the protein. If the protein has a mixture of secondary structure elements, their approximate percentages can be determined from the CD spectrum as well.

The longer wavelength region (240–300 nm) is most responsive to asymmetric environments around the standard protein chromophores and the chirality of disulfide bonds. This region can be used to assess if the protein is folded (aromatic residues in chiral environment) or not (aromatic residues surrounded by water). [5]

2 | MATERIALS AND METHODS

The sample of unknown protein #6 was dissolved in a 10 mM Potassium phosphate buffer, pH 7.6 at 0.03 mg/mL. UV-Vis spectra were taken from 340–190 nm using a Shimadzu UV-2600 spectrophotometer (Shimadzu) after performing a baseline correction.

Next, CD spectra were measured from 280–190 nm using a JASCO-1500 CD Spectropolarimeter (JASCO). All spectra were acquired with a data pitch of 1 nm, a DIT value of 2s, a bandwith of 1 nm, and a scanning speed of 50 nm/min. After

aquiring the baseline, a total of 10 spectra of the sample were collected and averaged to yield the final spectrum. All spectra described were acquired with a quartz cuvette with path length 0.1 cm.

The spectra was baseline-corrected and smoothed using the Spectragryph (v1.2; [7]) software. Secondary structure content was calculated using online resources DichroWeb [3] and BESTSEL [9];[8].

3 | RESULTS

The protein tested here was Protein #6 ($M_r \approx 60$ kDa, which has 526 residues and a molar extinction coefficient of $\varepsilon \approx 77{,}000$ $M^{-1} \cdot cm^{-1}$. [6]

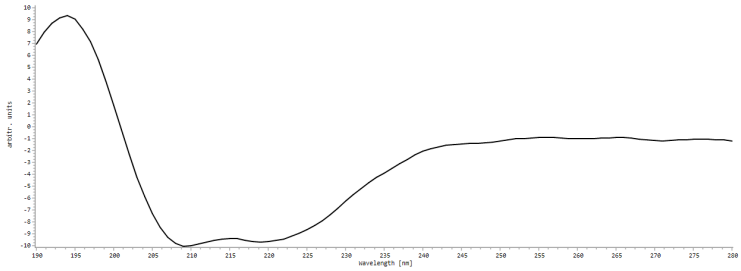


FIGURE 1 Post-corrected CD spectrum of Unknown Protein #6.

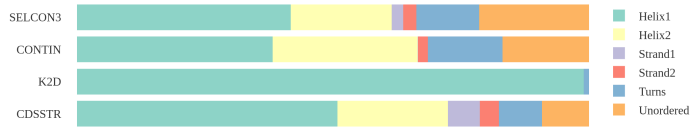


FIGURE 2 Percent composition of each secondary structure element as predicted by SELCON3, CONTIN, K2D, and CDSSTR.

TABLE 1 Secondary structure content of Unknown Protein #6 as predicted by a variety of methods.

Method	Reference Set	RMSD	Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
CDSSTR	4	0.001	0.52	0.16	0.09	0.05	0.1	0.09	1.01
	7	0.003	0.53	0.25	0.04	0.03	0.06	0.09	1
	SP175	0.003	0.57	0.23	0.03	0.02	0.1	0.06	1.01
	SMP180	0.003	0.43	0.23	0.09	0.05	0.08	0.13	1.01
K2D	———	0.095	0.99	———	———	———	0.01	———	1
CONTIN	4	0.058	0.4	0.334	0	0.054	0.203	0.009	1
		0.058	0.392	0.308	0.001	0.036	0.219	0.045	1.001
	7	0.058	0.426	0.323	0	0.029	0.146	0.076	1
		0.058	0.424	0.327	0	0.031	0.156	0.061	0.999
	SP175	0.103	0.37	0.224	0	0	0.092	0.314	1
		0.103	0.405	0.213	0	0.006	0.096	0.281	1.001
	SMP180	0.159	0.328	0.266	0	0	0.132	0.274	1
		0.159	0.312	0.273	0	0	0.122	0.292	0.999
SELCON3	4	0.338	0.452	0.214	0.024	0.024	0.112	0.208	1.034
	7	0.249	0.401	0.194	0.02	0.033	0.163	0.214	1.025
	SP175	0.304	0.443	0.204	0.026	0.024	0.105	0.244	1.045
	SMP180	———	———	———	———	———	———	———	———

4 | DISCUSSION

Additonal Question: Outline the implementation of the computer algorithms that you used to determine the secondary structure of your protein and highlight its key advantages or disadvantages over other existing algorithms.

Generally, CD spectra in the far-UV region are processed by an algorithm; a variety of such algorithms have been developed with different bases. Here, the web server DichroWeb, which implements many different methods, and BESTSEL (β-structure selection), which claims to better distinguish between parallel and antiparallel β-sheets than previously-developed methods.

For the spectral range sampled in this experiment, DichroWeb allowed analysis to be run using the CDSSTR, K2D, CONTIN, and SELCON3 methods. CONTIN is a ridge regression-based method that fits the CD of the unknown protein by comparing it to a linear combination of database spectra. K2D is a neural network, which is an artificial intelligence trained on experimentally verified CD data. Unfortunately, it does not estimate turns. CDSSTR is a variable selection-based method that creates a large database of standard spectra from proteins with known spectra and secondary structures, then utilizes single-value decomposition to find the structure of the unknown protein. Overall, variable selection-based methods claim to be the best available methods for prediction of the structure of globular proteins, but they does not always give fractional contents that add up to 100% and can be very computationally costly. SELCON3 is a self-consistent method that includes the spectrum of the unknown protein in the basis set, makes an initial guess, and solves the resulting matrix equation until it is self-consistent. [4]

As seen in Table 3, SELCON3 performs poorly in comparison to K2D and CONTIN, while BESTSEL and CDSSTR produce the most accurate results.

Overall, this procedure was able to produce a reasonably accurate estimate of the secondary structure content of our unknown protein. However, potential improvements to the experiment could be made by further improving the parameters of CD spectral acquisition (here, our parameters were reasonable, but do not obtain the "best" quality spectra), such as increasing the number of spectra collected and averaged over. A possible source of ambiguities in the experimental spectra here was the malfunctioning cuvette used in the lab, but this appears to have not posed any issues after subtraction of the background spectrum.

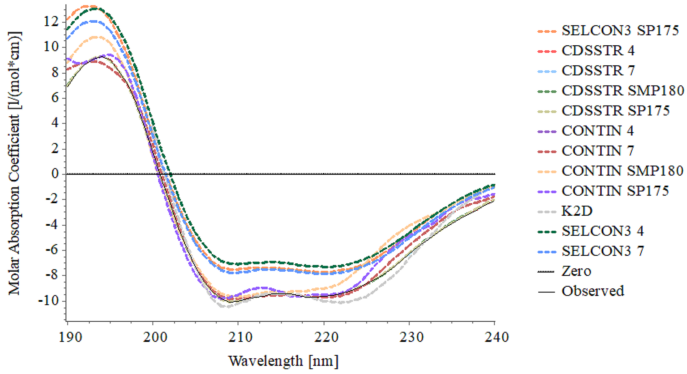


FIGURE 3 CD spectra predicted by each algorithm to best fit the experimental spectrum.

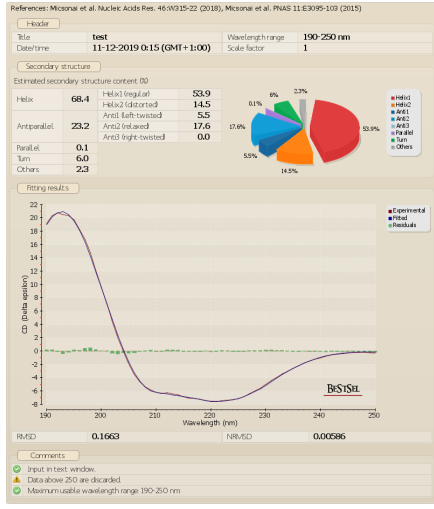


FIGURE 4 CD spectra predicted by BESTSEL.

5 | CONCLUSION

Overall, this experiment successfully produced spectra that could be input into a variety of computer algorithms that provide generally accurate predictions of the secondary structure content of the unknown protein. Further analyses may be made by investigating other regions of the spectrum to probe the environment around aromatic residues or metal ions bound to the protein of interest.

Appendix I: Rationale

5.1 | Sample Preparation

You will be given about 10 mg of dry crystalline unknown protein. You need to decide how to prepare the appropriate solutions, and decide (maybe via a trial-and-error approach) the appropriate data collection settings. You can use a variety of cuvettes that vary in path length and total volume. [6]

As a general rule, the absorbance of the sample should be ~1 at the wavelength of interest (λ_{220}) in a cuvette with the path length that will be used for data collection. A general concentration to start with would be **0.1–0.5 mg/mL protein** in a **0.1 cm cell**. [6] If this concentration produces too weak of a CD signal ($\lesssim$5–10 mdeg), the concentration of the sample can be doubled. However, since CD measures the difference between absorbance of circularly polarized light, using arbitrarily high concentrations prevents light from passing through the sample and no meaningful measurements can be made. [10]

The quartz cuvettes in the lab come in multiple sizes (path-lengths): 0.2 mm (dismountable), 1 mm (slightly defective), 5 mm, and 1 cm. [6]

An ideal buffer will produce little to no absorbance in the wavelength region of interest. Buffers, detergents or other additives that contain aromatic groups, thiol reducing agents (ME or DTT), imidazole, metal ions, and chloride salts should be avoided. [2],[1] A popular buffer for CD is **5-50 mM potassium phosphate buffer**. Others are listed along with their lower wavelength limits in Table 3.

TABLE 2 Properties of buffers. [4]

Buffer	Lower Wavelength Limit (nm)[A]
10 mM Potassium Phosphate, 100 mM potassium fluoride	185
10 mM Potassium Phosphate, 100 mM $(NH_4)_2SO_4$	185
10 mM,Potassium Phosphate, 50 mM Na_2SO_4	185
10 mM Potassium Phosphate, 100 mM KCl	195
20 mM Sodium Phosphate, 100 mM NaCl	195
Dulbecco's Phosphate buffered saline (PBS)[B]	200
2 mM Hepes, 50 mM NaCl, 2 mM EDTA, 1 mM Dithiothreitol.	200
50 mM Tris, 150 mM NaCl, 1 mM Dithiothreitol, 0.1 mM EDTA.	201

[A]The lower limit values are typical for solutions containing ~0.1 mg/ml protein in 0.1 cm cells. Below the lower wavelength cutoffs the dynode voltages rapidly increase, the signal to noise is poor and the ellipticity is not a linear function of the path length of the cell.
[B]9.33 mM Potassium Phosphate, 136 mM NaCl, 2.7 mM KCl, 0.6 mM $MgCl_2$, 0.9 mM $CaCl_2$

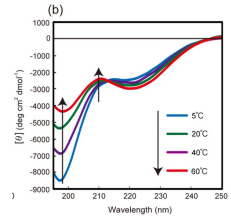
5.2 | Spectral acquisition parameters

- The user can select the number of channels to view during the collection period. Generally, one should monitor **HT (High Tension)** in **Channel 2** with **CD (Circular Dichroism)** signal in **Channel 1**. As a general rule, the $\frac{S}{N}$ ratio worsens as the HT voltage rises, and data points collected while the HT voltage $\geq$ 700 V are unreliable and should be discounted.
- Routine analyses of protein secondary structure are performed in the **far UV region (280–190 nm)**.
- The **data pitch** determines the number of data points taken during the scan. A data pitch of 1 nm is appropriate for routine analyses of protein secondary structure. [6]
- The **digital integration time (DIT)** is the amount of time that the data is integrated over, or the length of time the detector collects photons. Thus, a longer DIT value will generally result in a better $\frac{S}{N}$. A 2 second DIT with a 1 nm **bandwidth** is reported to be sufficient for most samples. If the signal is too low, the bandwidth can be increased to 2 nm.
- The **scanning speed** determines how quickly the instrument acquires data across the region of interest. The DIT must be taken into account when setting the scanning speed, as the response wavelength λ_R = DIT $\times$ Response (s). To prevent distortion in the measured spectrum, the $\lambda_R \leq \frac{1}{10}$FWHM. A good initial scanning speed is 50 nm/min, but the $\frac{S}{N}$ may be improved with slower speeds.
- Spectral scans collected together may be averaged, or 'accumulated.' As $\frac{S}{N} \sim \sqrt{\text{Response} \times \text{Number of Accumulations}}$, noise can be further reduced by increasing the number of **accumulations** once the previous scanning parameters have

been optimized.

- CD spectra is temperature-dependent. Thus, the temperature must be kept constant at a temperature about 10 degrees less than RT.

FIGURE 5 Temperature-dependence of CD spectra.



5.3 | Data Presentation and Analysis

You should outline a general plan on how to analyze your data as part of the prelab. More details about data analysis strategies will be available after completion of the experiment. [6]

TABLE 3 Sources of Circular Dichroism Analysis Software. [4]

	Website	Software	Operating systems
(1)	CDPro	SELCON3,CONTIN, CONTINLL,CDSSTR	Windows 95, 98, XP LINUS, UNIX
	Advantages: Data conversion program included. Superior fits of data on globular proteins. Source code available and can be compiled for use with LINUX or UNIX machines.		
(2)	Circular Dichroism at UMDNJ	LINCOMB, MS-DOS, MLR, SELCON, SELCON2, CONTIN, VARSLC, K2D, CCA	Windows 95, 98, XP
	Advantages: Data conversion programs are supplied. Data collected at any range of wavelengths can be analyzed. Peptide references for least squares analyses are included.		
	Disadvantages: Programs are not user friendly. Data conversion and text editing necessary.		
(3)	CONTIN	CONTIN	LINUX Source Code
	Disadvantages: Programs must be compiled by user.		
(4)	CCA + the CD Spectrum Analyser System	CCA	Windows 95, 98
	Advantages: Windows operating system.		
(5)	DICROPROT	LINEAR, REGRESSI, ON,SELCON2, SELCON3, CONTIN, VARSLC	Windows 95, 98, XP
(6)	K2D	K2D, DOS	On-line, Windows 95, 98, XP
	Advantages: Simple to use.		
(7)	SOMCD	SOMCD	On-line
	Advantages: Update of K2D algorithm. Analyzes turns as well as α-helix and β-sheets and uses data from 240 to 190 nm as well as from 240 to 200 nm.		
	Advantages: Easy to use. Every method uses the same input format. Output files are shown on the screen immediately.		
	Disadvantages: The package does not have a constrained least squares analysis module where the fractional spectral components are constrained to be positive. Many of the programs do not run unless the CD data is in the range of 260 to 178 nm.		
(8)	Dicroweb	CONTINLL, SELCON3, CDSSTR, VARSLC, K2D	On-line
	Advantages: On-line analysis available. Accepts input directly from the output of many different CD machines.		

[A]The lower limit values are typical for solutions containing ~0.1 mg/ml protein in 0.1 cm cells. Below the lower wavelength cutoffs the dynode voltages rapidly increase, the signal to noise is poor and the ellipticity is not a linear function of the path length of the cell.
[B]9.33 mM Potassium Phosphate, 136 mM NaCl, 2.7 mM KCl, 0.6 mM $MgCl_2$, 0.9 mM $CaCl_2$

DICROPROT (DICHROism of PROTeins) [3] integrates most of the methods designed for the estimation of protein sequence secondary structure into a single package.

Appendix II: Protocol

UV-Vis spectrum of native protein

Initialize the Shimadzu UV-2600 spectrophotometer and set up spectral acquisition from 340–190 nm. Using the solution that you created, perform the baseline correction. Record the UV-Vis spectrum of your protein solution after blanking the instrument and decide if the optical properties of your protein solution are suitable for the next step.

CD spectrum of native protein

1. Start Spectra Manager. Select Measurement, right-click and select Start to initialize the instrument. A "Nitrogen Gas Replacement" notification pops up. Wait for purging to finish. Alternatively, if the instrument is pre-purged, select "Skip". The light source is automatically ignited.
2. Click on the "Parameters" icon that has a hand pointing to a table. Under "General" tab: Verify that Channel 1 is CD (Circular Dichroism) and Channel 2 is HT (High Tension)
   Select an appropriate start and end wavelengths:
   - For the analysis of secondary structures: 280–190 nm.
   - For the analysis of environment around aromatic residues: 320–240 nm.
   - For the analysis of environment around heme: 600–350 nm (use 1 cm cuvette for this)

   Data pitch 1 nm is appropriate for the routine analysis of secondary structure of proteins. For spectra that show sharp, or closely spaced peaks, one shall reduce data pitch to 0.5 nm.
   Select suitable scanning speed and Digital Integration Time (DIT).
   - For a fast scanning, you may choose 50 nm/min with D.I.T. of 2 sec.
   - To obtain better signal-to-noise, a slower scanning speed and longer D.I.T. is recommended.

   Under "Cell Unit" tab, Check "Control" option and enter the desired temperature.
   Under "Information", enter appropriate description.
   Under "Data", uncheck "Auto Save"; Make sure that "Send data To Spectra Analysis" is Checked.
   Hit OK to transfer parameters to the instrument.
3. Collect the blank spectrum:
   a. Choose which spectral range you will be using.
   b. Pipette the your buffer into the cuvette (300 μL in 1 mm cuvette, or 75 μL in 0.2 mm cuvette, or 2 mL in 1 cm cuvette). When using the dismountable cuvette, place the side with the oval cavity facing up, and slide the cover from one end to another while gradually pipetting the sample into the cavity.
   c. Place the cuvette with phosphate buffer into the sample compartment and carefully note the cuvette orientation. Close the lid of the compartment, it should "click".
   d. Hit the "Sample Measurement" button that has S with an arrow under it. The phosphate buffer background CD signal should be almost zero because water and phosphate ions are non-chiral; while DTT is chiral it's contribution to CD signal is small. Pay attention to the HT signal line, it should not exceed 800 V. Assess if the baseline is flat, or shows spurious signals (often due to a dirty or strained cuvette). Save this spectrum in case you decide to subtract it from your sample spectra later.
4. Collect the spectrum of your native protein.
   a. Pipette the solution of your protein into the cuvette (300 μL of 0.5 mg/mL in 1 mm cuvette, or 75 μL of 1 mg/mL in 0.2 mm cuvette, or 2 mL of 5 mg/mL in 1 cm cuvette).
   b. Place the cuvette with sample compartment and carefully note the cuvette orientation. Close the lid of the compartment, it should "click".
   c. Hit the "Sample Measurement" button that has S with an arrow under it. Verify that the Sample Info is correct and hit "OK".
   d. When the data acquisition has finished, analyze the spectrum. Record the positions of significant minima and maxima, and the corresponding CD signal value(s). Save this spectrum.

ACKNOWLEDGMENTS

Acknowledgments should include contributions from anyone who does not meet the criteria for authorship (for example, to recognize contributions from people who provided technical help, collation of data, writing assistance, acquisition of funding, or a department chairperson who provided general support), as well as any funding or other support information.

References

[1] D. Barrick, *Sample Preparation for Far-UV CD Spectra*, Johns Hopkins University (2013), 1–2.

[2] CSB, *CD Sample Preparation*, Vanderbilt University (2014).

[3] D. G and G. C, *An interactive graphic program for calculating the secondary structure content of proteins from circular dichroism spectrum*, Comput Appl Biosci **2** (1993), 197–199.

[4] N.J. Greenfield, *Using circular dichroism spectra to estimate protein secondary structure*, Nat Protoc **1** (2006), 2876–2890.

[5] K. Kahn, *Biochemistry Laboratory Manual: Theory*, CHEM 110L **2** (2019), 128–131.

[6] K. Kahn, *Circular Dichroism Characterization of Proteins*, CHEM 110L (2019).

[7] F. Menges, *Spectragryph: Optical spectroscopy software*, Software-Entwicklung **1.2** (2016).

[8] A. Micsonai, F. wien, E. Bulyaki, J. Kun, E. Moussong, Y. Lee, Y. Goto, M. Refregiers, and J. Kardos, *ABeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra*, Nucleic Acids Res. **46** (2018), W315–322.

[9] A. Micsonai, F. Wien, L. Kernya, Y. Lee, Y. Goto, M. Refregiers, and J. Kardos, *Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy*, Proc Natl Acad Sci USA **112** (2015), E3095–3103.

[10] K. Stott and J. Maman, *Sample Preparation for CD Measurements*, University Cambridge Biophysics (2013), 1–2.

Supplementary Information

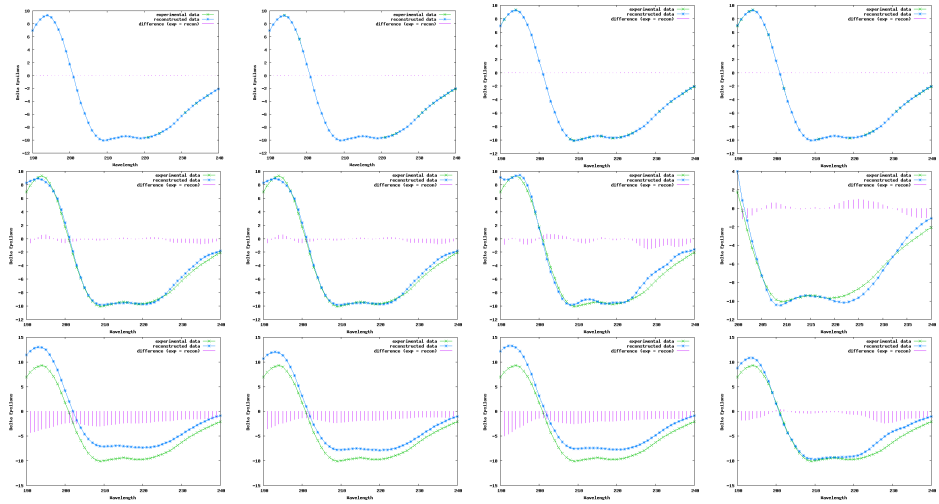


FIGURE 6 From left to right: CDSSTR 4, 7, SP175, SMP180. From left to right: CONTIN 4, 7, SP175. From left to right: SELCON3 4, 7, SP175, SMP180. Plots of observed and theoretical spectra generated by DICHROWEB using the K2D analysis program. No reference set was necessary.