

ACDP – a Java application for data processing and analysis of protein circular dichroism spectra

A. Hofmann

Structural Chemistry Program, Eskitis Institute for Cell & Molecular Therapies, Griffith University, Brisbane, Qld 4111, Australia

ACDP is a Java application for processing protein circular dichroism (CD) spectra either individually or within a series. Data processing includes spectrum subtraction ('baseline correction'), conversion of the raw CD into units of mean residue ellipticity, wavelength monitoring and graphical inspection. Three different algorithms for secondary structure deconvolution have been implemented, and spectra can be analysed without the need for data reformatting using a neural network approach, variable selection or linear combination of prototype spectra. The application is written entirely in Java and is thus portable to a wide variety of platforms, requiring only the Java Runtime Environment.

1. Introduction

Circular dichroism (CD) spectroscopy is a versatile tool for structural biochemistry, owing to its ability to monitor the secondary structure of proteins and peptides. The chirality of the C α atom in all amino acids but glycine induces asymmetry into the vicinal peptide bond which absorbs light in the far-UV region (250 – 190 nm). The secondary structure of a polypeptide thus induces an 'overall chirality' which gives rise to the CD phenomenon of a protein in the far-UV region.

While CD instruments record the raw CD in units of milli-degrees, protein spectra are commonly reported as molar ellipticity θ , or mean residue ellipticity θ_{res} which is normalised with respect to the number of chromophores (i.e. peptide bonds). The normalisation makes CD spectra of peptides of different length comparable.

The ability of CD spectroscopy to monitor the secondary structure of peptides and proteins makes it a versatile and useful tool for structural biology. The deconvolution of secondary structure contents of a protein is a frequently used application to assess the protein fold in solution or membrane-bound. The thermal denaturation of peptides and proteins can conveniently be monitored by CD spectroscopy and the be used to assess the fold stability. Since the binding of ligands increases the fold stability, this type of assay presents a convenient way of assessing ligand binding in a label-free manner. CD spectroscopy may thus be applied in a wide variety of fields ranging from structural biology and biotechnology to drug discovery.

Spectroscopic assays in the laboratory often comprise elaborate experimentation that includes acquisition of spectra for the samples, as well as 'baseline spectra' for control. In subsequent analyses, sample spectra have to be processed by subtracting the appropriate control spectra, and in the case of CD spectroscopy, data have to be converted from raw CD to mean residue ellipticity. While one spectrum at a time can in most cases be processed online using the acquisition software provided by the manufacturer of the instrument, there is no generally available software that enables processing of many spectra in batch mode.

Additionally, in many cases, secondary structure deconvolution requires time-consuming reformatting of the data. Since the 1980s, many different deconvolution algorithms have been generated and implemented in software applications. However, most of these are written either in Fortran or C, and require some 'advanced knowledge' to operate.

An important step towards unifying the scattered knowledge base for CD spectroscopy has been the online web service Dichroweb (<http://www.cryst.bbk.ac.uk/cdweb/html/>) which not only provides links to CD tutorials and deconvolution algorithms, but also offers the possibility to analyse CD

spectra by secondary structure deconvolution, albeit only for registered users. The reformatting of CD data into specific formats is still a requirement when using this web interface. Currently, Dicroprot (Deleage & Geourjon, 1993), CDPPro (Sreerama & Woody, 2000) and CDtool (Lees *et al.*, 2004) are the only available stand-alone programs that integrate different deconvolution methods into one single package. Dicroprot is only executable under the Windows operating system. In order to provide an intuitive, user-friendly and widely applicable way of processing and analysing protein CD spectra, we have generated the Java application ACDP.

2. Results and discussion

2.1 The application

Based on previous efforts in our laboratory to utilise Java for structural and biophysical tasks (Hofmann & Wlodawer, 2002), we have designed the stand-alone Java application ACDP (Automated CD Data Processing). All features of the program are available by a graphical user interface (GUI), and the results are also written into ASCII files for convenient import into other software.

Raw CD (in 10^{-3} °) data obtained in ASCII format from a variety of different CD acquisition software programs can be processed by ACDP using the appropriate filter (AVIV, ASCII, JASCO, JASCO-J715, JASCO-J810). If specified, ACDP carries out baseline correction and calculates spectra in units of mean residue ellipticity ($^{\circ}\text{cm}^2\text{dmol}^{-1}$). An ASCII file for each corrected spectrum is written out. Two wavelengths can be selected to be monitored for each spectrum, and their mean residue ellipticities, as well as the ratio of these are tabulated in the results file. All results are also tabulated in a table for direct copy-paste into spreadsheet programs.

CD spectra are displayed in the 'Graph' pane for visual inspection and manual analysis. The displayed spectra can be interrogated for wavelength and CD/ellipticity using either a line cursor or direct key entry in the 'Spectrum' tool panel. The tool panel also allows for smoothing using either mean or median filtering.

Secondary structure deconvolution can be conveniently performed within the program without the need for manual data reformatting. Three different deconvolution algorithms have been implemented.

2.2 Neural network analysis

Two neural network applications for CD spectroscopy have been reported so far: CDNN (Bohm *et al.*, 1992) and K2D (Andrade *et al.*, 1993; Merelo *et al.*, 1994) both treat secondary structure as a superposition of secondary structure information, aromatic side chain and cysteine absorption, random noise, and errors from protein concentration. While CDNN uses a single hidden layer approach with a linear transfer function to convey information between layers, K2D is a Kohonen neural network with a two-dimensional output layer. Using the K2D program, we have implemented the *K2D* Java class into ACDP.

2.3 Variable selection

The Java class *VarSelec* is based on the program *Varslc1.for* (Variable Selection) by Manavalan, and modified by Toumadje & Johnson. *VarSelec* computes the secondary structure of a protein from its CD by using the method of Hennessey & Johnson (Hennessey & Johnson, 1981), and the singular value decomposition (SVD) algorithm (Compton & Johnson, 1986) for all possible combinations in the basis set (Manavalan & Johnson, 1987).

2.4 Linear combination of Fasman prototype spectra

The Java class *LinCombFasman* computes the best fit of the experimental CD spectrum by linear combination of three or five prototype Fasman spectra (Greenfield & Fasman, 1969; Perczel *et al.*, 1992). The algorithm implemented here uses a brute force approach and the following constraints: (i) the sum of all fractional contents must equal 1, and (ii) the individual fractional content must be

greater than 0.

References

- Andrade, M., Chacon, P., Merelo, J. & Moran, F. (1993). *Protein Eng.* **6**, 383-390.
- Bohm, G., Muhr, R. & Jaenicke, R. (1992). *Protein Eng.* **5**, 191-195.
- Compton, L. & Johnson, C. (1986). *Anal. Biochem.* **155**, 155-167.
- Deleage, G. & Geourjon, C. (1993). *Comput. Appl. Biosci.* **2**, 197-199.
- Greenfield, N. & Fasman, G. (1969). *Biochemistry* **8**, 4108-4116.
- Hennessey, J. & Johnson, C. (1981). *Biochemistry* **20**, 1085-1094.
- Hofmann, A. & Wlodawer, A. (2002). *Bioinformatics* **18**, 209-210.
- Lees, J., Smith, B., Wien, F., Miles, A. & Wallace, B. (2004). *Anal. Biochem.* **332**, 285-289.
- Manavalan, P. & Johnson, W. (1987). *Anal. Biochem.* **167**, 76-85.
- Merelo, J., Andrade, M., Prieto, A. & Moran, F. (1994). *Neurocomputing* **6**, 1-12.
- Perczel, A., Park, K. & Fasman, G. (1992). *Anal. Biochem.* **203**, 83-93.
- Sreerama, N. & Woody, R. (2000). *Anal. Biochem.* **287**, 252-260.

Figure Legends

Figure 1

Screen shots of the *Input Files* panel (A) and the *Graph* panel (B). After processing, spectra can be automatically loaded into the *Graph* panel for inspection.

Figure 2

Screen shots of the *Results* panel (A) and the *Secondary Structure Prediction* panel (B).

The spectra specified in Figure 1A have been processed and the CD signal/mean residue ellipticity at 208 nm and 222 nm are tabulated in the *Results* panel. After highlighting two columns, an xy-plot can be produced by a simple right-click on the mouse.

The secondary structure content of any spectrum can be deconvoluted by simply clicking *Predict* in the *Spectrum detail* panel shown in Figure 1B.

Figure 1

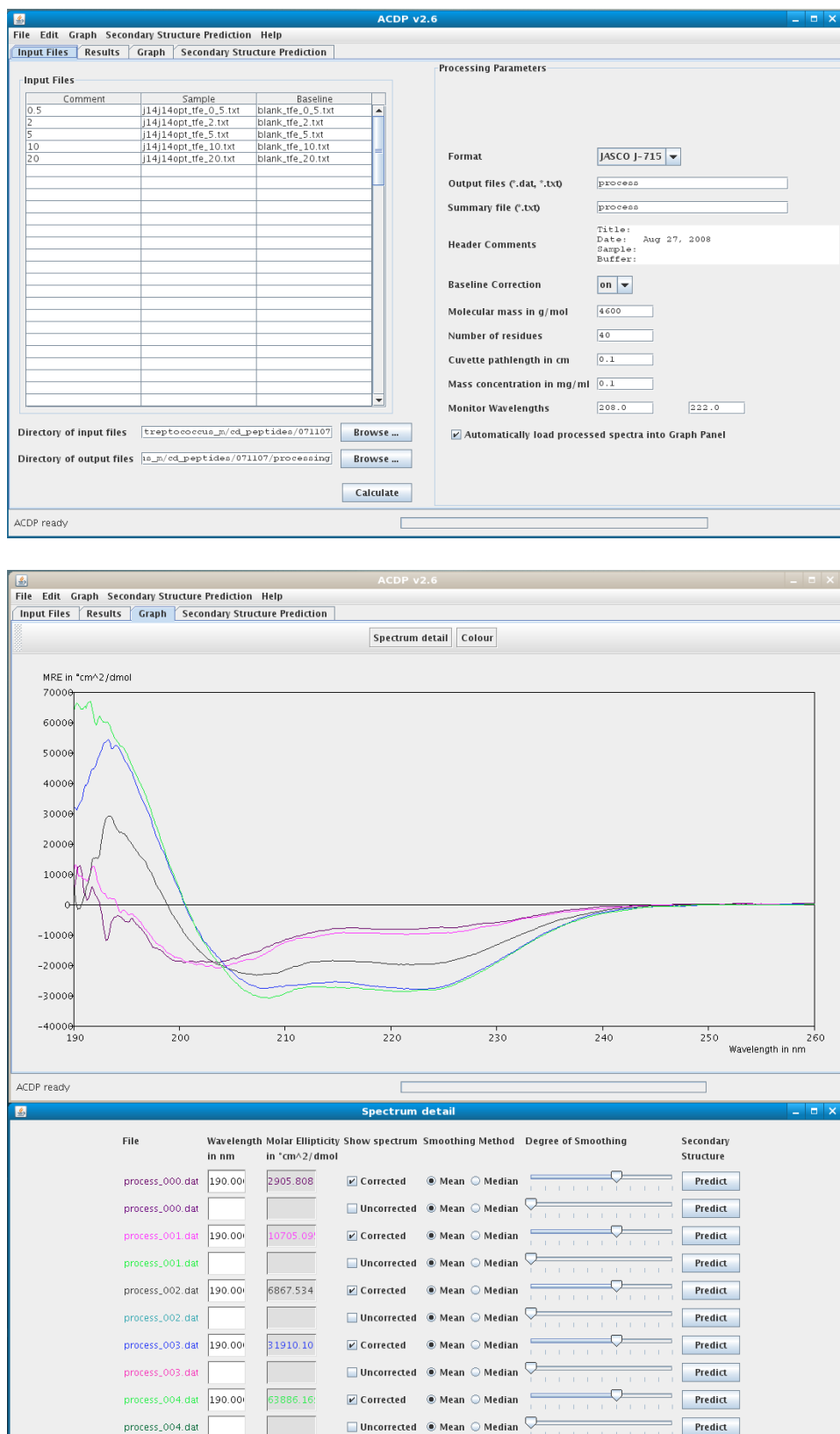


Figure 2

