
CDNN USER MANUAL

by E.Bucci

Supported OS:

This program will work only with Microsoft Windows 95/98 or Windows NT. Windows 2000 not yet tested.

To install the program:

- 1- Create a directory (e.g. c:\cdnn);
- 2- Copy the installation files in this directory;
- 3- Double click on instdsk1.exe and then on instdsk2.exe.

To use the program:

- 1- Open the CD file you want to analyse with the Jasco program "Standard analysis";
 - 2- Transform the CD spectrum in a molar ellipticity spectrum (menu Mathematics - Optical constant);
 - 3- Divide the obtained spectrum for the number of protein residues (menu Mathematics - Arithmetics w. constant);
 - 4- Execute the "dump" routine (menu Spectra - Dump) with a 1 nm sampling interval and a 185-260 nm spectral window;
 - 5- Press the copy button on the results box;
 - 6- Paste the data in a Excel worksheet;
 - 7- Clear all the rows containing non-numerical characters and clear also the third column of data;
 - 8- Use the Excel function to rearrange the data from 185 to 260 nm;
 - 9- Save the obtained data file as a text file (not in the Unicode format!);
 - 10- Close Excel and open the new data file by Word;
 - 11- Change the tab characters (they should be 71) into 4 spaces (use the Word search&change function and specify "^t" as text to search for and " " as text to insert);
 - 12- Save the file as simple text;
 - 13- Execute the CDNN program and load the data file using molar ellipticity as measurement unit;
 - 14- In the Options-Preferences menu select the option "Net using 13 basespectra";
 - 15- Press the "Deconvolute" button;
 - 16- Write down the results for the 185-260 nm interval and repeat step 14,15 and 16 setting 23 and 33 basespectra respectively;
 - 17- The best deconvolution out of the three obtained is the one where the total secondary structure content is more similar to 100%.
-

Deconvolution reliability:

The program output is an ordered table showing you the secondary structure content prediction as a function of the wavelength range considered for the fitting. Most reliable is, of course, the lowest-wavelength prediction. However, the other predictions should not differ too far. Moreover, the secondary structure total content should be as close as possible to 100 %. If this value deviates more than 5-10%, then either the data are incorrect (for example, units were converted wrongly), or the network structure / base spectra are not suitable for solving your problem (i.e. your spectrum is very different from any other in the program database).

Concerning the prediction error, even if there is no way to know it *a-priori*, the authors report the average error in estimating the secondary structure of a protein contained in the program database. At the current state of the trained networks (CDNN version 2.0.3.188), the average error (%) for the prediction of one of these protein structures is:

	180 - 260	185 - 260	190 - 260	195 - 260	200 - 260	205 - 260	210 - 260
NNET_13	4,32 %	4,37 %	4,35 %	4,51 %	4,63 %	4,84 %	4,91 %
NNET_23	4,87 %	5,06 %	5,03 %	5,16 %	5,38 %	5,47 %	5,44 %
NNET_33	3,98 %	5,92 %	5,57 %	6,20 %	6,45 %	6,39 %	6,60 %

These values are to be referred as the minimum error made by the program on average and WITHOUT FURTHER TRAINING.
