Using Circular Dichroism-Based Neural Networks to Predict the Secondary Structure of Proteins in Solution

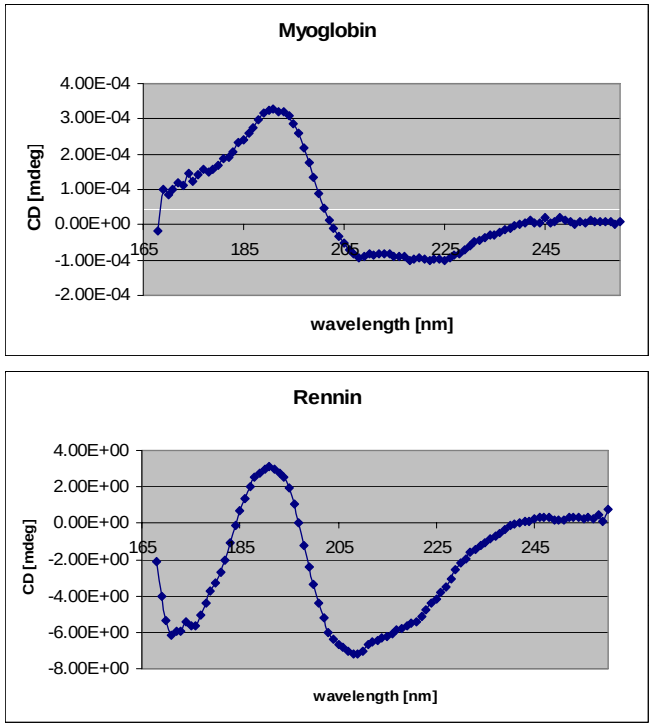
Beamline:
U11

Technique:
Circular Dichroism Spectroscopy

Researchers:
Imke Bodendiek, Mathew Engel, M. Allaire, Hui Zhong, L. M. Miller (BNL)

Publication:
In progress

Motivation: The functionality of proteins is intimately related to the secondary structure. A change of a protein's structure - as it occurs in denaturing processes - usually comes along with a loss of its biological function or activity. Proteins, especially enzymes, play a major role in metabolic processes and malfunctions / misfolding are often associated with disease. To gain a deeper understanding of the biochemical processes and to find an effective cure against illnesses, it is necessary to get information about the secondary structure of proteins. The most accurate results can be derived from crystallography experiments. But since crystallization of proteins is in often difficult and time-consuming, researchers sometimes need a quick method to get an idea about the secondary structure of their proteins. Circular dichroism spectroscopy has proven to be a valuable tool in this respect. Moreover, CD spectroscopy offers the advantage of studying proteins in their natural environment, namely a solution state. To extract the structural information from CD spectra, the help of algorithms is needed. One possibility for analyzing the data is to use a neural network (NN). A currently well-used NN, which is publicly available on an open-access web server, is K2d (Andrade et al., 1993). However, its capability of predicting several different kinds of secondary structure elements is limited due to the relatively small number of reference proteins used to train the NN and due to the small wavelength range of the data. The goal of this project is to create an open-access NN, which is capable of predicting more different secondary structure elements with a higher accuracy than K2d.



(Top) Circular Dichroism spectra of a mostly α-helical (myoglobin) and (bottom) a mostly β-sheet (rennin) protein.

Results: For the training of a neural network, it is mandatory to have very pure proteins with a solved crystal structure. Currently, our reference protein set contains 30 proteins that match these criteria. Compared to this, the actual number of proteins in the training set of K2d is relatively small (18 proteins) and the spectral range of the data is limited because the spectra were taken on a conventional CD instrument without a synchrotron source (240 to 200 nm), which restricts the prediction to 3 secondary structure elements (α-helix, β-sheet, remainder). Here, the CD data were taken on beamline U11 at the NSLS. The spectra encompass a wavelength range from 280 to 168 nm, which permits the prediction of five different secondary structure elements (α-helix, 310-helix, β-sheet, turns and remainder). Test runs of proteins with a known crystal structure show that the predictions of the new NN are more accurate than K2d's predictions. Furthermore, it is capable of predicting more different structural elements. In the near future, the NN is expected to become accessible to the public.