Chapter 5

Conclusions

In Chapter 2, the application of NMR to protein structure determination was described, and the central role of chemical shift assignment in this process was made clear. The advantages of higher-dimensionality spectra for the purposes of assignment were shown, and an outline of the different types of NMR experiment was given. It was also noted that the assignment process is very time consuming, which has motivated researchers to look into ways of automating it.

A classification system for already-existing programs for the automated assignment of NMR spectra was presented, and a number of them were described in some detail. In all cases, the information used as a starting point by these programs was a manually-edited peak list. The rationale for a new type of program to find spin systems was described, using a knowledge of the expected *patterns* of peaks for different amino acids to find spin systems.

In Chapter 3, the implementation of such a program (*patt_recog*) was described. It was designed to work on patterns of peaks in multiple spectra simultaneously. The operation comprised three essential steps:

- convolutional filtering to emphasise spectrum peaks;
- search for peak patterns;
- heuristic filtering of results.

Section 3.2 in the same chapter summarised the results obtained during the use of this program. In Section 3.3, it was shown that pattern search within the original spectra allowed patterns to be found even in cases where there was chemical shift displacement between spectra, or where peaks were weak or nonexistent. The results presented in Section 3.5 showed that the program is capable of assigning a significant proportion of the chemical shifts in the spectra presented to it. Section 3.6, demonstrated that existing patterns could be used in the assignment of completely new spectra without substantial adjustment.

The sequential assignment program (*chain*) was described in detail in Chapter 4. This program uses *patt_recog* results from both backbone spectra and spin system assignments as constraints during the sequential assignment process. It's operation had two phases: first a chaining algorithm was used to form candidate chains of backbone results. The sequence of the protein was used to constrain this algorithm. The resulting list of chains, which constituted hypothesised partial sequential assignments, were subjected to further processing to delete similar chains, and penalise chains containing undesirable features.

Two high-scoring fragments of significant length were found (Section 4.2), which was actually a promising result if the poor quality of the data is taken to account, and it is likely that it could be further improved if a larger and more consistent set of spectra were available.