A GRAPH-BASED AUTOMATED NMR BACKBONE RESONANCE SEQUENTIAL ASSIGNMENT

Xiang Wan and Guohui Lin
Department of Computing Science, University of Alberta
Edmonton, Alberta T6G 2E8, Canada

The success in backbone resonance sequential assignment is fundamental to protein three dimensional structure determination via NMR spectroscopy. Such a sequential assignment can roughly be partitioned into three separate steps, which are grouping resonance peaks in multiple spectra into spin systems, chaining the resultant spin systems into strings, and assigning strings of spin systems to non-overlapping consecutive amino acid residues in the target protein. Separately dealing with these three steps has been adopted in many existing assignment programs, and it works well on protein NMR data that is close to ideal quality, while only moderately or even poorly on most real protein datasets, where noises as well as data degeneracy occur frequently. We propose in this work to partition the sequential assignment not into physical steps, but only virtual steps, and use their outputs to cross validate each other. The novelty lies in the places where the ambiguities in the grouping step will be resolved in finding the highly confident strings in the chaining step, and the ambiguities in the chaining step will be resolved by examining the mappings of strings in the assignment step. In such a way, all ambiguities in the sequential assignment will be resolved globally and optimally. The resultant assignment program is called GASA, which was compared to several recent similar developments RIBRA, MARS, PACES and a random graph approach. The performance comparisons with these works demonstrated that GASA might be more promising for practical use.