

# Lecture 6

## JIGSAW and NMR

In this lecture, we mainly talk about the JIGSAW algorithm [1] for automated NMR peak assignment.

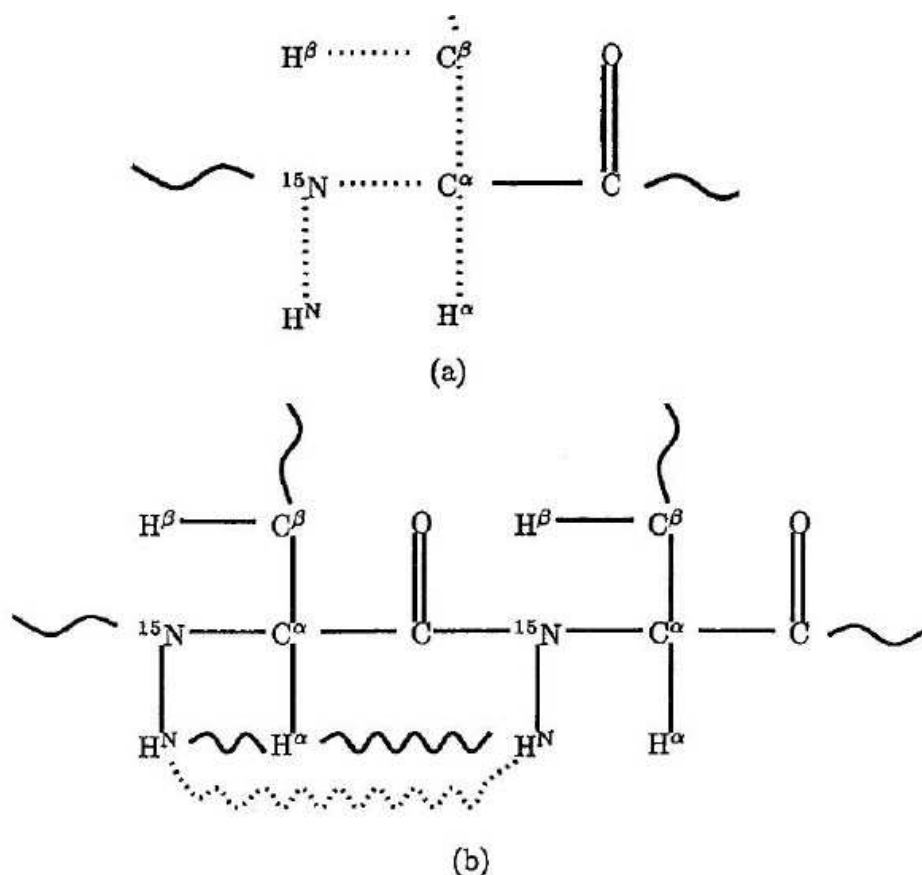
### 1 Overview of JIGSAW

The JIGSAW algorithm is a novel high-throughput and automated approach for secondary structure determination and resonance assignment from few, cheap NMR experiments. It applies a graph theoretic framework to represent the atom interactions from the NMR spectra. The contributions of JIGSAW lie in the following parts: First, based on the observation that most protein structures consist of some regular secondary structures, such as  $\alpha$ -helices,  $\beta$ -sheets, etc., and thus result in certain patterns of NMR-measurable distances, JIGSAW shows that the spectrum characteristics of secondary structures can be efficiently identified, and useful for the resonance assignment. Second, JIGSAW implies that a good resonance assignment can be obtained by using the fact that the proton chemical shifts of different amino acid types statistically display distinct patterns in the NMR spectra. Third, JIGSAW can identify  $\beta$ -sheets, which are really tertiary structures.

### 2 NMR Spectra Used in JIGSAW

JIGSAW deals with the spectra from the following four NMR experiments: NOESY, HSQC, HNHA and TOCSY. The NOESY spectra measure the through-space atom interactions, while the other three spectra measure the through-bond atom interactions. We briefly review these four NMR experiments. Figure 1 shows the atom interactions that can be captured by these spectra.

- (1) **HSQC:** HSQC is used to capture the through-bond interaction of correlated  $H^N$  and  $^{15}N$  atoms. Since every residue has a unique  $H^N$ - $^{15}N$  pair on the protein backbone, and thus ideally has distinct frequency signals, the HSQC spectrum serves as the identification of each residue (unassigned), and reference interaction for all other spectra.
- (2) **HNHA:** HNHA is used to estimate the J coupling constant  $^3J_{HNHA}$ , which is related to the  $\phi$  torsion angle of a peptide plane. Thus, we can figure out the secondary structure type from a HNHA spectrum.
- (3) **TOCSY:** The TOCSY spectrum captures the through-bond interactions of protons on a residue's side chain. Since the chemical shifts of protons for different amino acids are characteristically different, TOCSY serves as a fingerprint of each amino acid type.



**Figure 1:** Through-bond and through-space atom interactions in a protein [1]. Atom nomenclature and interactions in a protein. (a) Through-bond interactions shown with dotted lines (HSQX:  $H^N-^{15}N$ ; HMHA:  $H^N-^{15}N-H^\alpha-H^\beta$ -...). (b) Through-space interactions in NOESY shown with wavy line ( $d_{\alpha N}$  solid and  $d_{NN}$  dashed).

- (4) **NOESY:** NOESY captures the through-space Nuclear Overhauser Effect (NOE) between an amide proton  $H^N$  and its neighboring  $^{15}N$ , which are within a distance less than 6 Å.

### 3 Graph Representation of Atom Interactions in NOESY Spectra

#### 3.1 Graph Representation

In JIGSAW, the atom interactions from the NOESY spectra are represented by a directed graph  $G = (V, E)$ , where  $V$  denotes the set of residues (unassigned, but sharing same atoms  $H^N$  and  $^{15}N$ ), and  $E$  denotes the set of possible NOESY proton interactions between two

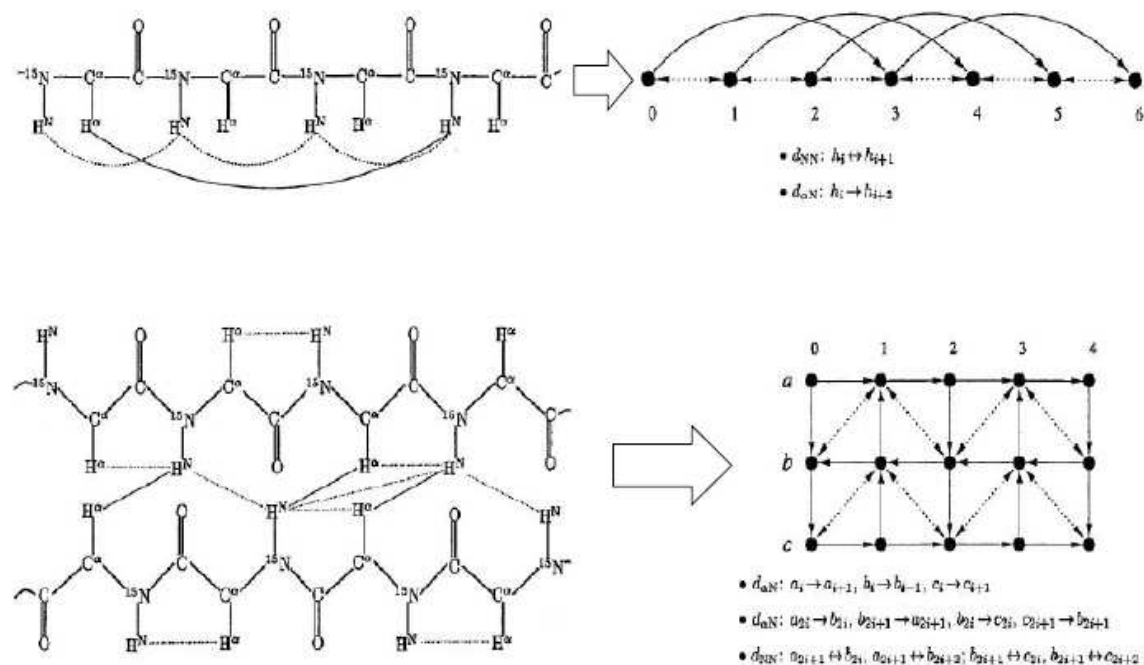


Figure 2: Graph constraints [1].

residues. Because of the noise from the experimental data, there are possible erroneous nodes and edges in a interaction graph. There are other characteristics associated with the nodes and edges in the interaction graph, such as secondary structure type label, match score, atom distance, etc. For more details, please refer to the definition of “NOESY interaction graph” in [1].

**Comments:** Why should we formulate the proton interactions into directed graphs instead of undirected ones? The reason is that, when two protons are connected to different heavy atoms, the NMR pulse sequence will cause the quantum coherence to transfer along different magnetization pathways.

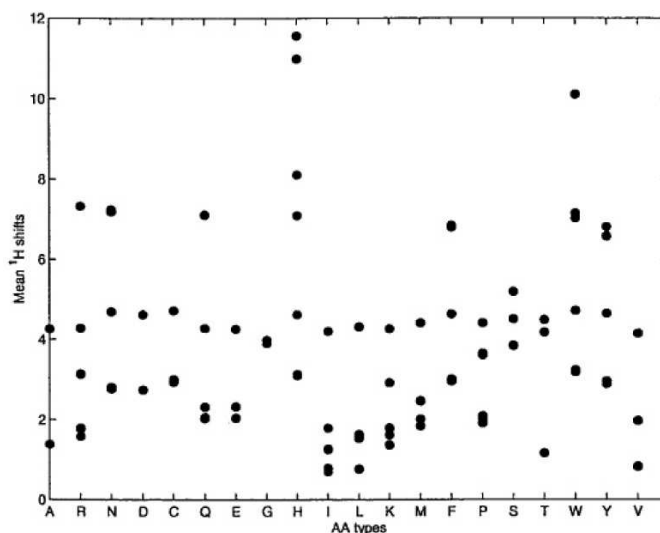
### 3.2 Graph Constraints for Identifying Secondary Structure

Ideally (if there is no noise), the secondary structures, such as  $\alpha$ -helix and  $\beta$ -sheet, will display certain graph constraints, as shown in Figure 2. Such graph constraints should correspond to some subgraphs in the interaction graph.

## 4 Assignment by Alignment of Side-Chain Fingerprints

The BioMagResBank (BMRB) has collected the statistics of observed chemical shifts of the 20 different amino acid types, and shown that different amino acid types display distinct

fingerprints. Figure 3 shows the mean chemical shifts of the protons for the 20 different amino acid types. JIGSAW applies these distinct fingerprints to predict amino acid type, given the chemical shifts from NMR spectra, and secondary structure types discovered.



**Figure 3:** The mean proton chemical shifts of the 20 different amino acid types in BMRB [1].

## 5 Secondary Structure Pattern Discovery

### 5.1 Algorithm Description

Since there are always some false nodes or edges in the interaction graph due to the experimental data noise, we need to develop algorithms to identify the subgraphs that are consistent with the secondary structure graph constraints. JIGSAW applies the following algorithm (Figure 4 shows the procedure of finding the  $\beta$ -sheet secondary structure from the interaction graph.):

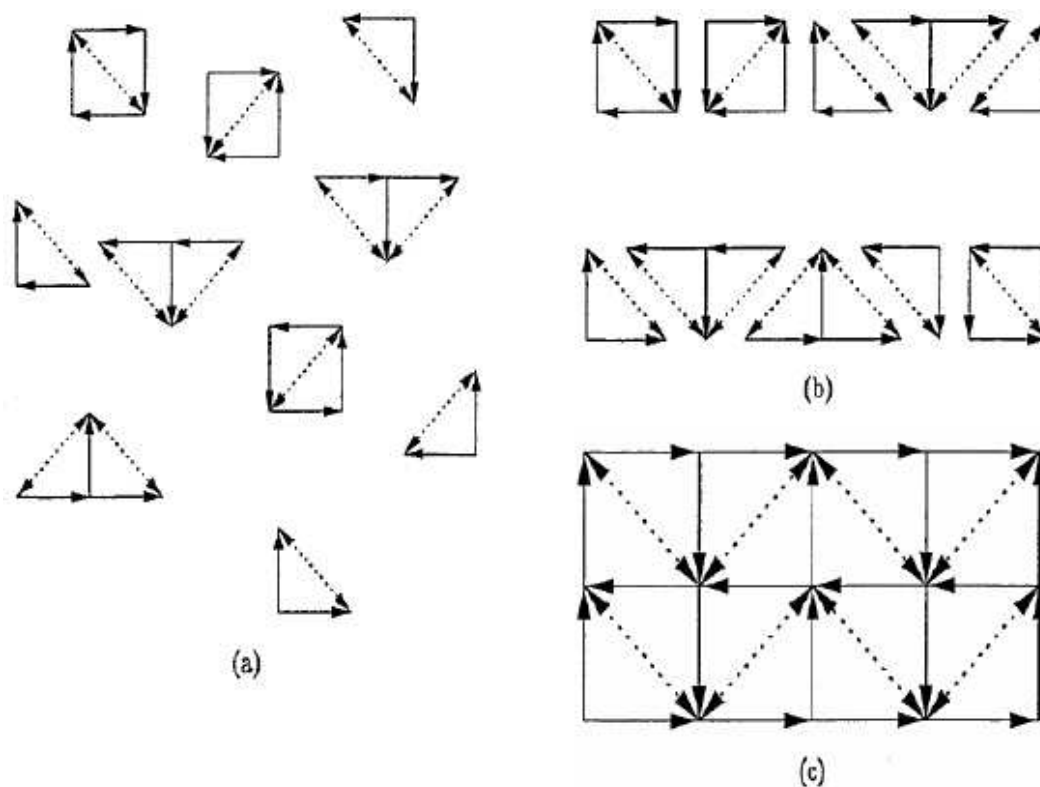
**Step 1:** Identify instances of fragment patterns.

**Step 2:** Form sequences of consistent fragments.

**Step 3:** Form sheets of consistent fragments.

**Step 4:** Identify the best secondary structure graphs from the set of collected possibilities.

Although the interaction graph from experimental NOESY spectra unavoidably includes many false edges, it seems that a self-consistent pattern of continuous false edges is unlikely to occur, based on a simple joint probability model. JIGSAW applies the insight that incorrect edges can be mutually inconsistent, while correct edges can consistently reinforce each other.



**Figure 4:** Algorithm for identifying  $\beta$ -sheet from the interaction graph [1]. JIGSAW algorithm overview: (a) identify graph fragment. (b) merge them sequentially, and (c) collect them into complete secondary graphs. Only correct fragments are shown here. Graphs from experimental data also generate a large number of incorrect fragments, but mutual inconsistencies prevent them from forming either long sequences or large secondary structure graphs.

## 5.2 Results

The following data were used for testing the performance of JIGSAW: the HSQC, HNHA, 80ms TOCSY and NOESY experimental data of huGrx, CBF- $\beta$  and vacGrx proteins. In other tests, the simulated J-coupling constant data are also used in the testing. The results show that JIGSAW is successful in structure discovery and NMR resonance assignment. For more details on the simulation results, please refer to the original paper.

## References

- [1] C. Bailey-Kellogg, A. Widge, J. J. Kelley III, M. J. Berardi, J. H. Bushweller, and B. R. Donald. The NOESY Jigsaw: Automated protein secondary structure and main-chain assignment from sparse, unassigned NMR data. *Jour. Comp. Biol.*, 3-4(7):537-558, 2000.