

Accelerating Biomolecular Nuclear Magnetic Resonance Assignment with A*

Joel Venzke, Paxten Johnson, Rachel Davis, John Emmons,
Katherine Roth, David Mascharka, Leah Robison,
Timothy Urness and Adina Kilpatrick

Department of Mathematics and Computer Science
Drake University

joel.venzke@drake.edu

April 10, 2014

Overview

- 1 Introduction
 - Motivation
 - Nuclear Magnetic Resonance Spectroscopy
- 2 NMR Assignment Background
 - Data Collection and Manual Assignment
- 3 Automation Algorithm
 - Preprocessing
 - Assignment
 - Goal State
- 4 Conclusion
 - Results
 - Outlook

Motivation

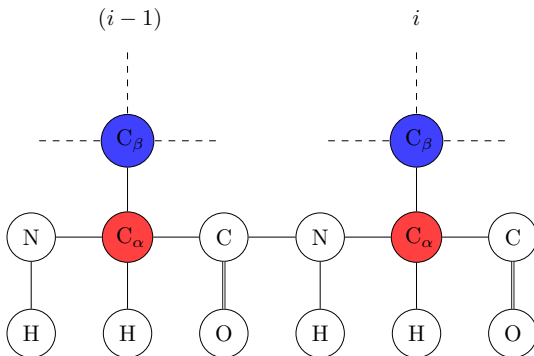
- Nuclear Magnetic Resonance Spectroscopy
 - Gain knowledge about protein structure
 - Study how mutations lead to diseases
- Problems
 - Generates large amounts of data
 - Data analysis is slow and error prone
- Goal
 - Automate the assignment process
 - Decrease human error
 - Increase productivity

Nuclear Magnetic Resonance (NMR)

- Used to obtain structural information
 - Chemical shift values
- HNCACB experiment
 - Generates C_α and C_β residue i and $i - 1$
- CBCA(CO) NH experiment
 - Generates C_α and C_β for residue i
 - Confirms residue data

Chemical Shift Values

HNCACB



Timeline

Protein
Production
at least 5 days

Data Assignment
20 days to 9 months

NMR
Experiments
1-2 days

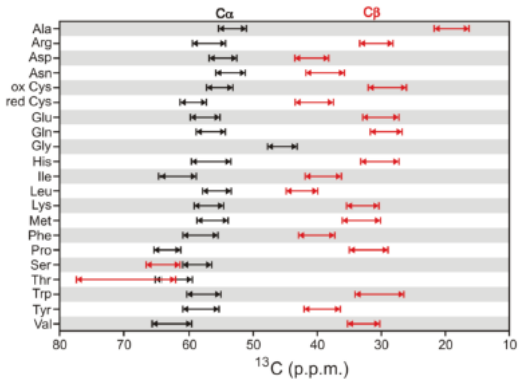
Manual Methods

- Most time consuming part
- Missing and ambiguous data forces chunks to be skipped
- Prone to human error

Initialization

- Expected amino acid sequence
 - Converted to expected chemical shift values
 - Stored as the reference protein chain
- NMR experiment's chemical shift data
 - C_α and C_β for residue i and $i - 1$
 - Stored in a tile
- Missing data
 - Place holder tile generation
- Grouping

Grouping



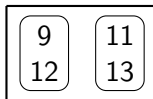
[2]

Starting the assignment

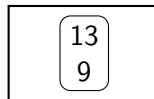
Tiles to assign:



Group 1



Group 2



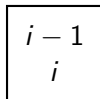
**Reference
Protein Chain**

Nodes

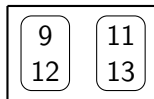
Chemical Shift	Group
13.5	1
9.5	2
11.4	1

Starting the assignment

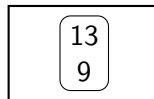
Tiles to assign:



Group 1



Group 2



Reference
Protein Chain

Nodes

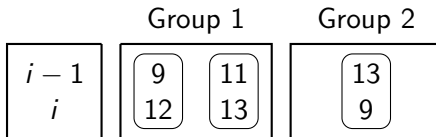
Chemical Shift	Group		
13.5	1	9 12	11 13
9.5	2		
11.4	1		

Cost Calculation

- Accuracy matching the protein chain residue
- Accuracy matching the tile above current tile
- Cost of placing all previous tiles

Generating child nodes

Tiles to assign:



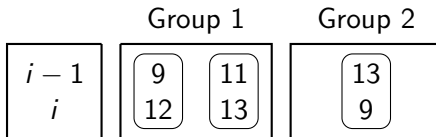
Reference
Protein Chain

Nodes

Chemical Shift	Group		
13.5	1	9 12	1.5
9.5	2		
11.4	1	11 13	0.5

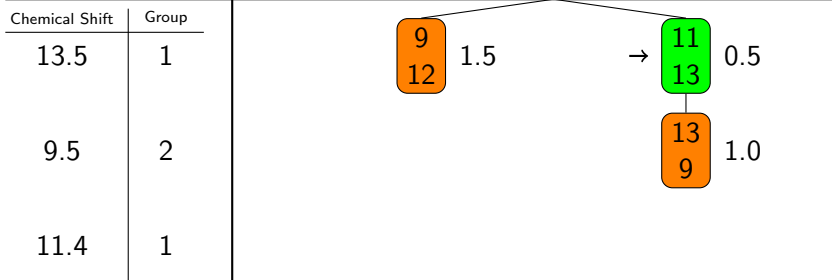
Generating child nodes

Tiles to assign:



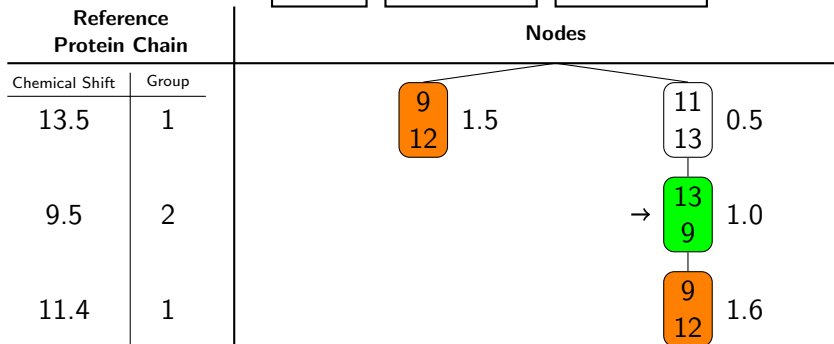
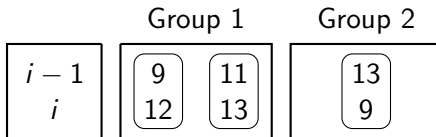
Reference
Protein Chain

Nodes



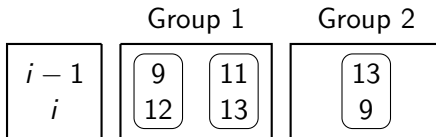
Goal State

Tiles to assign:



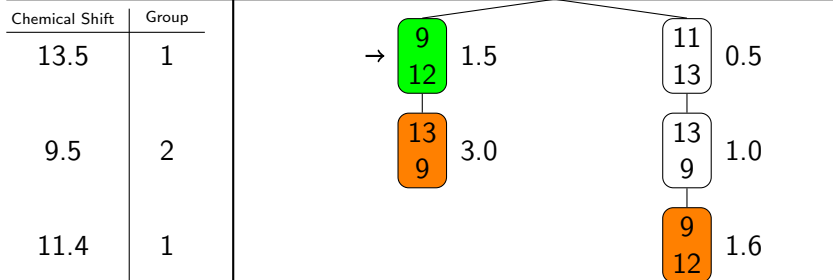
Goal State

Tiles to assign:



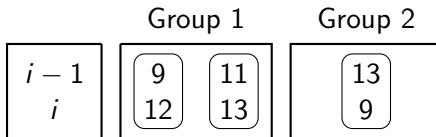
Reference
Protein Chain

Nodes



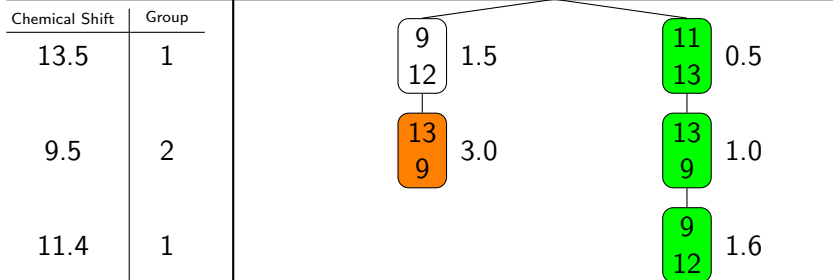
Solution State

Tiles to assign:

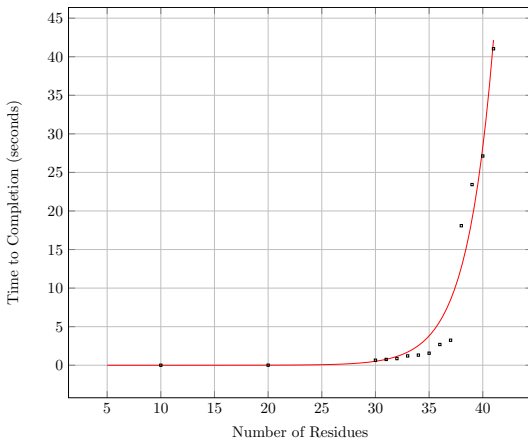


Reference
Protein Chain

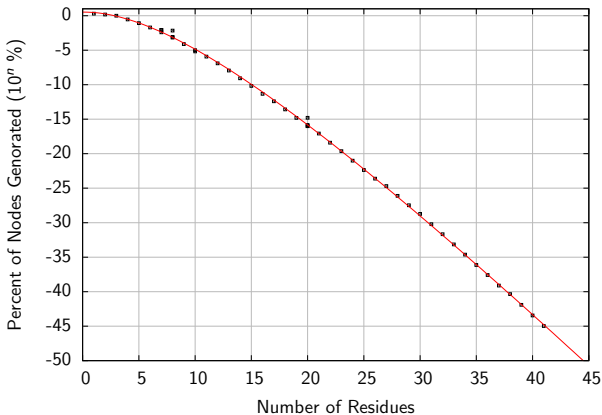
Nodes



Time of Assignment



Child Nodes Generated



Future Goals

- Parallelization
 - Decrease assignment time
 - Allow for larger data sets
- Machine learning
 - Optimize cost calculation
 - Increase accuracy of assignment

Acknowledgments

- Dr. Tim Urness (Mathematics and Computer Science)
- Dr. Adina Kilpatrick (Physics)
- Rachel Davis (research colleague)
- John Emmons (research colleague)
- Katherine Roth (research colleague)
- David Mascharka (research colleague)
- Leah Robison (research colleague)

Bibliography



Babak Alipanahi, Xin Gao, Emre Karakoc, Frank Balbach, Shuai Cheng Li, Guangyu Feng, Logan Donaldson and Ming Li, *Error tolerant NMR backbone resonance assignment and automated structure generation.*, Journal of bioinformatics and computational biology, **9** (2011), 15–41.



Sean Cahill and Mark Girvin.
Introduction to 3d triple resonance experiments.
2012.

Thank You

