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

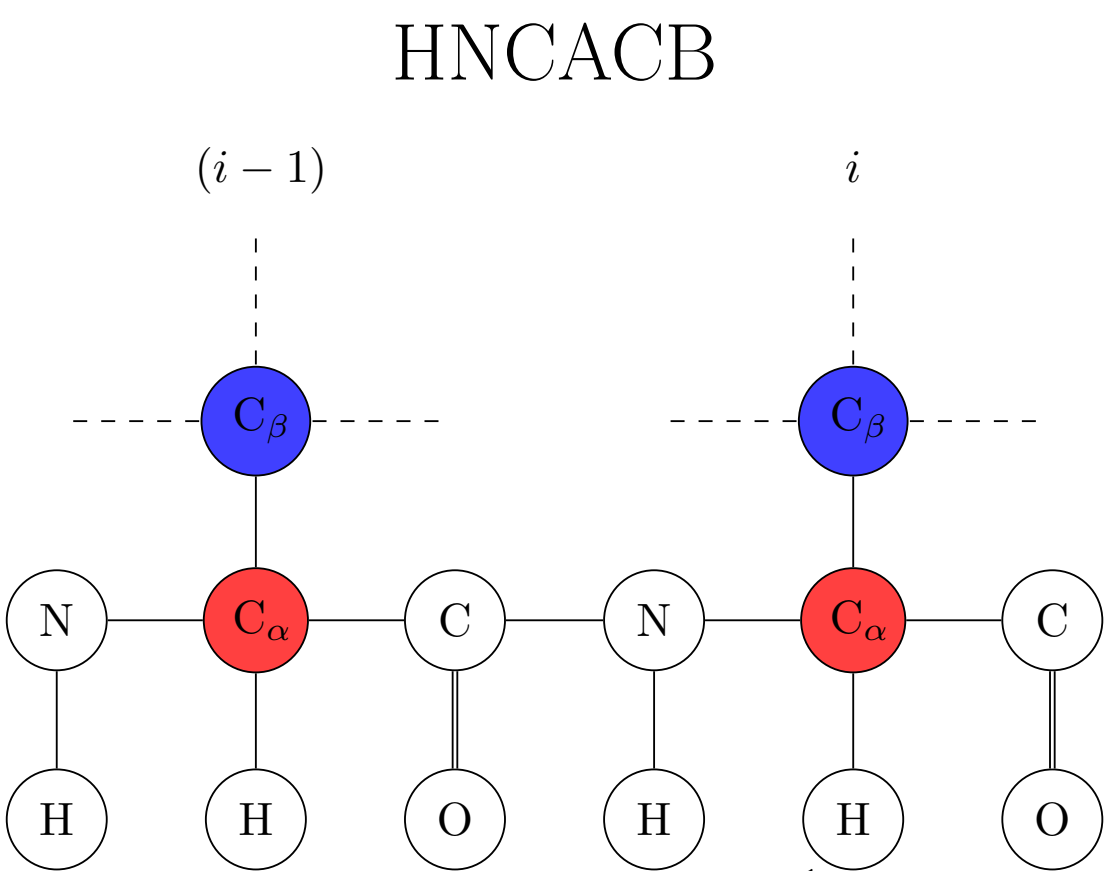
Motivation

- 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



Data Collection Time Line

- Protein production
  - At least 5 days [1]
- NMR Experiments
  - 1 to 2 days per spectrum involved [1]
- Assignment can begin

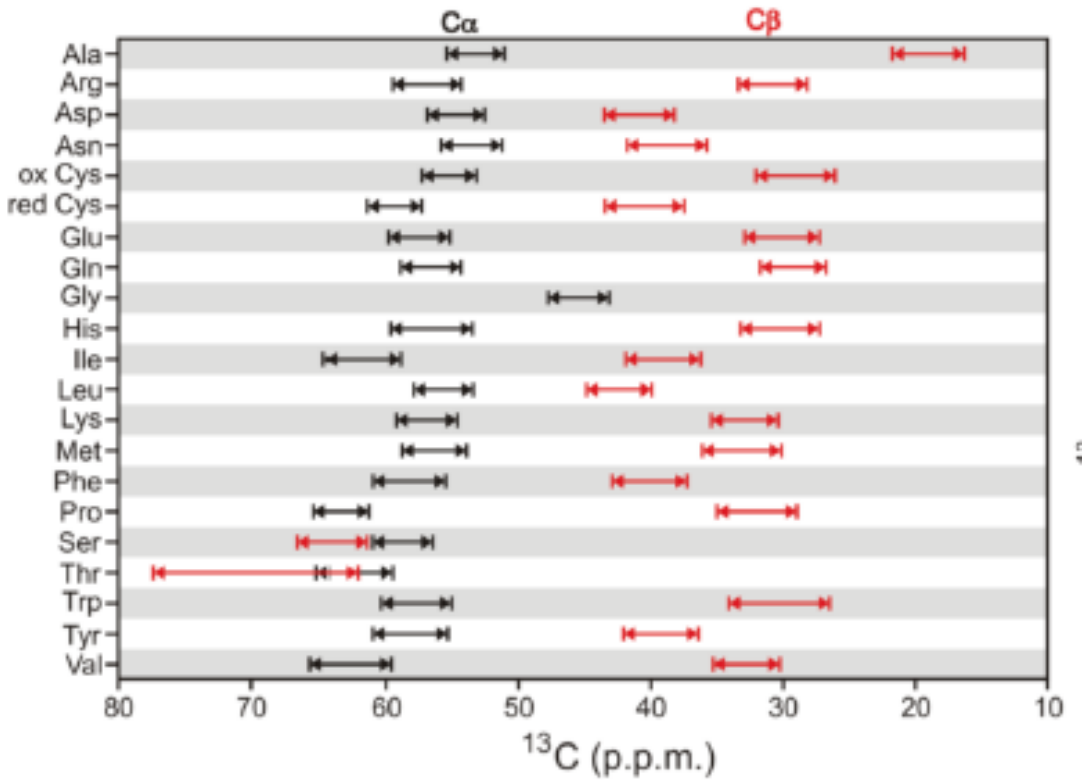
****** graphic *********

Manual Methods

Initialization

- Input
  - Expected amino acid sequence
    - Covered to expectation chemical shift values
    - Stored as the protein chain
  - NMR 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


Cα
Cβ
Ala
Arg
Asp
Asn
ox Cys
red Cys
Glu
Gln
Gly
His
Ile
Leu
Lys
Met
Phe
Pro
Ser
Thr
Trp
Tyr
Val
80
70
60
50
40
30
20
10
13C (p.p.m.)

[2]

Starting the assignment

- Include Diagram of first step in the assignment process

Cost Calculation

- Accuracy matching the protein chain residue
- Accuracy matching the tile above current tile
- Cost of all tiles place before current tile
  - Obtained from parent node

Generating child nodes

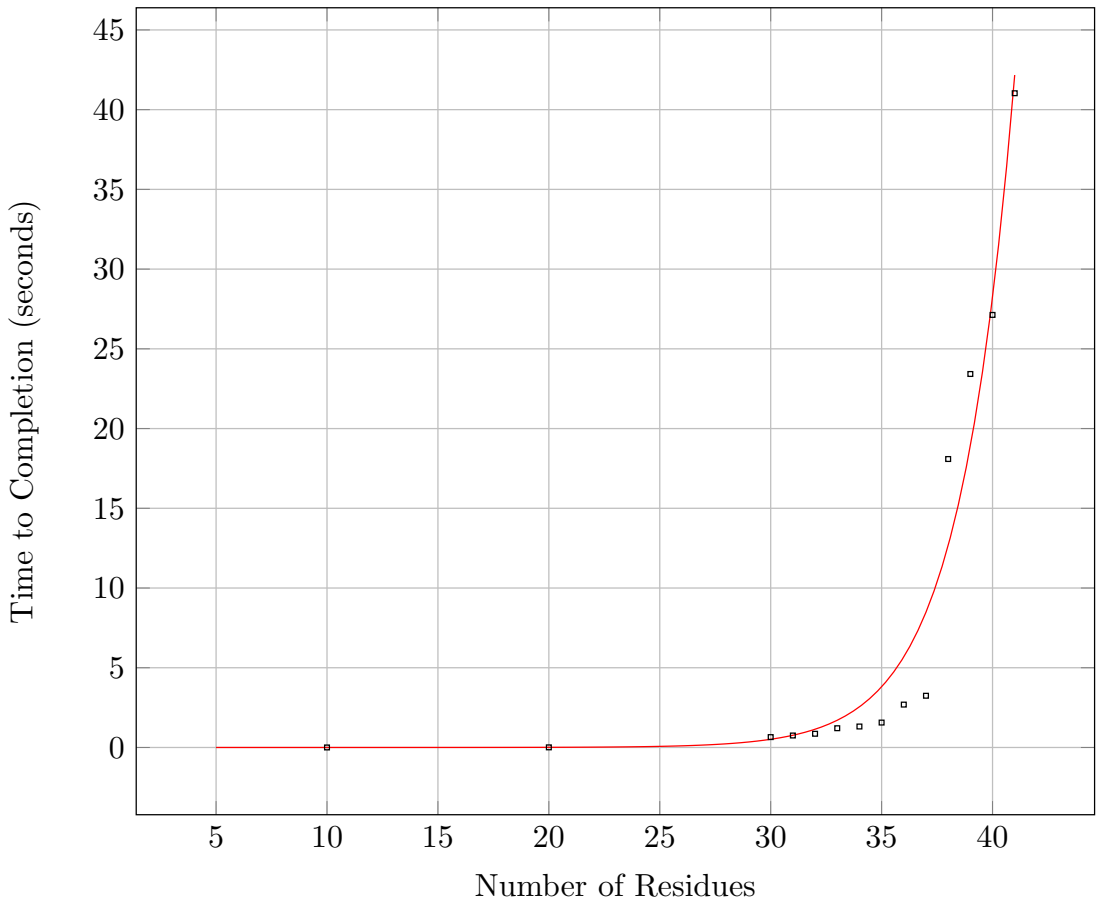
- The node with the lowest cost is selected

****** Include graphic *********

Goal State

- Child nodes are generated until goal state is reached
- Goal state
  - All tiles are placed

Time of Assignment



Assignment Issues

- Missing data decreases accuracy
  - increases assignment time
- ****** add more or remove slide ********

Future Goals

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

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.