

Automated Assignment of Backbone NMR Data using Artificial Intelligence

John Emmons

Drake University

john.emmons@drake.edu

April 19th, 2013



Overview



Overview

- Background



Overview

- Background
 - Nuclear Magnetic Resonance (NMR)



Overview

- Background
 - Nuclear Magnetic Resonance (NMR)
 - Research project summer 2012



Overview

- Background
 - Nuclear Magnetic Resonance (NMR)
 - Research project summer 2012
- Manual data assignment protocol



Overview

- Background
 - Nuclear Magnetic Resonance (NMR)
 - Research project summer 2012
- Manual data assignment protocol
 - Motivation for automation of assignment



Overview

- Background
 - Nuclear Magnetic Resonance (NMR)
 - Research project summer 2012
- Manual data assignment protocol
 - Motivation for automation of assignment
- Automated assignment algorithm



Overview

- Background
 - Nuclear Magnetic Resonance (NMR)
 - Research project summer 2012
- Manual data assignment protocol
 - Motivation for automation of assignment
- Automated assignment algorithm
- Results and future work



Background–Nuclear Magnetic Resonance Spectroscopy



800 MHz NMR Device



Background–Nuclear Magnetic Resonance Spectroscopy



Background–Nuclear Magnetic Resonance Spectroscopy

- Physical understanding



Background–Nuclear Magnetic Resonance Spectroscopy

- Physical understanding
 - NMR is the phenomenon in which atomic nuclei absorb and re-emit EM radiation based on their local chemical environment



Background–Nuclear Magnetic Resonance Spectroscopy

- Physical understanding

- NMR is the phenomenon in which atomic nuclei absorb and re-emit EM radiation based on their local chemical environment
- NMR spectroscopy involves two steps



Background–Nuclear Magnetic Resonance Spectroscopy

- Physical understanding
 - NMR is the phenomenon in which atomic nuclei absorb and re-emit EM radiation based on their local chemical environment
 - NMR spectroscopy involves two steps
 - ① Alignment of magnetic nuclear spins with a constant magnetic field



Background–Nuclear Magnetic Resonance Spectroscopy

- Physical understanding

- NMR is the phenomenon in which atomic nuclei absorb and re-emit EM radiation based on their local chemical environment
- NMR spectroscopy involves two steps
 - ① Alignment of magnetic nuclear spins with a constant magnetic field
 - ② Perturbation of this alignment with EM radiation, typically a radio frequency (RF) pulse



Background–Nuclear Magnetic Resonance Spectroscopy

- Physical understanding
 - NMR is the phenomenon in which atomic nuclei absorb and re-emit EM radiation based on their local chemical environment
 - NMR spectroscopy involves two steps
 - ① Alignment of magnetic nuclear spins with a constant magnetic field
 - ② Perturbation of this alignment with EM radiation, typically a radio frequency (RF) pulse
 - Information is gained by observing the changes in the perturbing EM signal, chemical shift



Background–Nuclear Magnetic Resonance Spectroscopy

- Physical understanding
 - NMR is the phenomenon in which atomic nuclei absorb and re-emit EM radiation based on their local chemical environment
 - NMR spectroscopy involves two steps
 - ① Alignment of magnetic nuclear spins with a constant magnetic field
 - ② Perturbation of this alignment with EM radiation, typically a radio frequency (RF) pulse
 - Information is gained by observing the changes in the perturbing EM signal, chemical shift
- The location of nuclei with respect to other atoms can be found with chemical shift



Background–Nuclear Magnetic Resonance Spectroscopy

- Physical understanding
 - NMR is the phenomenon in which atomic nuclei absorb and re-emit EM radiation based on their local chemical environment
 - NMR spectroscopy involves two steps
 - ① Alignment of magnetic nuclear spins with a constant magnetic field
 - ② Perturbation of this alignment with EM radiation, typically a radio frequency (RF) pulse
 - Information is gained by observing the changes in the perturbing EM signal, chemical shift
- The location of nuclei with respect to other atoms can be found with chemical shift
 - Especially useful in biology for determining molecule structure



Background–Nuclear Magnetic Resonance Spectroscopy



Background–Nuclear Magnetic Resonance Spectroscopy

- Applications in biology



Background–Nuclear Magnetic Resonance Spectroscopy

- Applications in biology
 - Investigating protein structure and dynamics in aqueous solutions



Background–Nuclear Magnetic Resonance Spectroscopy

- Applications in biology
 - Investigating protein structure and dynamics in aqueous solutions
 - Studying protein-drug interaction



Background–Nuclear Magnetic Resonance Spectroscopy

- Applications in biology
 - Investigating protein structure and dynamics in aqueous solutions
 - Studying protein-drug interaction
 - Protein hydration detection



Background–Nuclear Magnetic Resonance Spectroscopy

- Applications in biology
 - Investigating protein structure and dynamics in aqueous solutions
 - Studying protein-drug interaction
 - Protein hydration detection
- NMR provides useful data, but analysis can be nontrivial



Background–Nuclear Magnetic Resonance Spectroscopy

- Applications in biology
 - Investigating protein structure and dynamics in aqueous solutions
 - Studying protein-drug interaction
 - Protein hydration detection
- NMR provides useful data, but analysis can be nontrivial
 - Requires examination of multiple NMR spectra



Background–Nuclear Magnetic Resonance Spectroscopy

- Applications in biology
 - Investigating protein structure and dynamics in aqueous solutions
 - Studying protein-drug interaction
 - Protein hydration detection
- NMR provides useful data, but analysis can be nontrivial
 - Requires examination of multiple NMR spectra
 - Human error can easily slow down data analysis



Background–Summer 2012 Research



Background–Summer 2012 Research

- Calmodulin (CaM)



Background–Summer 2012 Research

- Calmodulin (CaM)
 - Calcium-binding protein expressed in all eukaryotic cells



Background–Summer 2012 Research

- Calmodulin (CaM)
 - Calcium-binding protein expressed in all eukaryotic cells
 - Multifunctional, i.e., used in many bodily processes



Background–Summer 2012 Research

- Calmodulin (CaM)
 - Calcium-binding protein expressed in all eukaryotic cells
 - Multifunctional, i.e., used in many bodily processes
- CaM in Drosophila muscle



Background–Summer 2012 Research

- Calmodulin (CaM)
 - Calcium-binding protein expressed in all eukaryotic cells
 - Multifunctional, i.e., used in many bodily processes
- CaM in Drosophila muscle
 - In muscle, CaM binds to ryanodine receptor



Background–Summer 2012 Research

- Calmodulin (CaM)
 - Calcium-binding protein expressed in all eukaryotic cells
 - Multifunctional, i.e., used in many bodily processes
- CaM in Drosophila muscle
 - In muscle, CaM binds to ryanodine receptor
 - Regulates Ca^+ ion flow in and out of cells



Background–Summer 2012 Research

- Calmodulin (CaM)
 - Calcium-binding protein expressed in all eukaryotic cells
 - Multifunctional, i.e., used in many bodily processes
- CaM in Drosophila muscle
 - In muscle, CaM binds to ryanodine receptor
 - Regulates Ca^+ ion flow in and out of cells
 - Serves as a biofeedback mechanism to prevent damage to muscle from excess Ca^+



Background–Summer 2012 Research

- Calmodulin (CaM)
 - Calcium-binding protein expressed in all eukaryotic cells
 - Multifunctional, i.e., used in many bodily processes
- CaM in Drosophila muscle
 - In muscle, CaM binds to ryanodine receptor
 - Regulates Ca^+ ion flow in and out of cells
 - Serves as a biofeedback mechanism to prevent damage to muscle from excess Ca^+
- Summer 2012 Research



Background–Summer 2012 Research

- Calmodulin (CaM)
 - Calcium-binding protein expressed in all eukaryotic cells
 - Multifunctional, i.e., used in many bodily processes
- CaM in Drosophila muscle
 - In muscle, CaM binds to ryanodine receptor
 - Regulates Ca^+ ion flow in and out of cells
 - Serves as a biofeedback mechanism to prevent damage to muscle from excess Ca^+
- Summer 2012 Research
 - Structural studies of a CaM mutant with defective regulation of muscle contraction

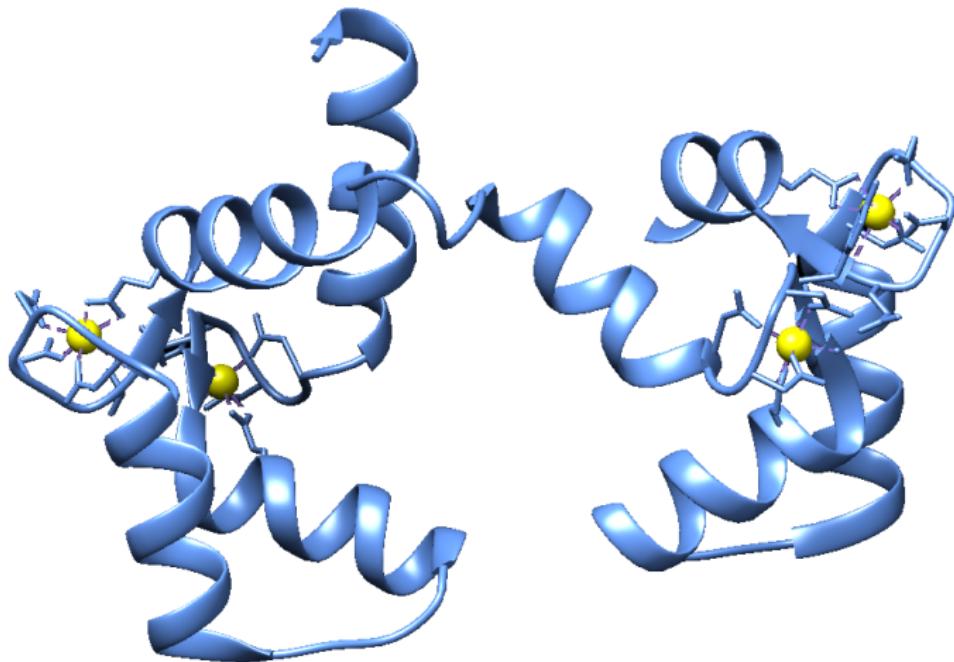


Background–Summer 2012 Research

- Calmodulin (CaM)
 - Calcium-binding protein expressed in all eukaryotic cells
 - Multifunctional, i.e., used in many bodily processes
- CaM in Drosophila muscle
 - In muscle, CaM binds to ryanodine receptor
 - Regulates Ca^+ ion flow in and out of cells
 - Serves as a biofeedback mechanism to prevent damage to muscle from excess Ca^+
- Summer 2012 Research
 - Structural studies of a CaM mutant with defective regulation of muscle contraction
 - Studied the structure of CaM mutant in a low Ca^+ environment



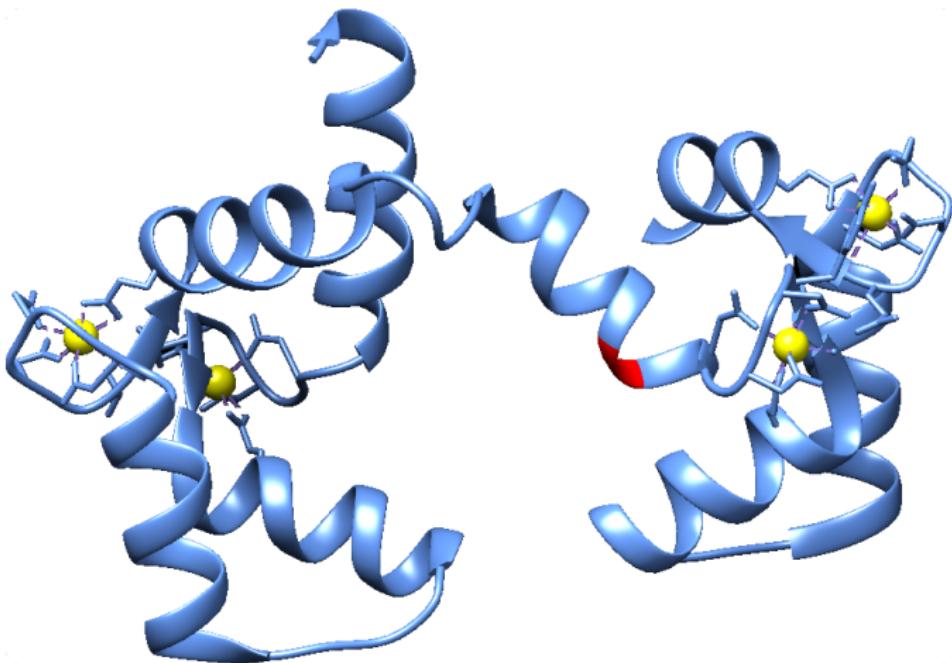
Background–Summer 2012 Research



Wild Type CaM Protein Structure



Background–Summer 2012 Research



V91G CaM Mutant Protein Structure



Background–Summer 2012 Research



Background–Summer 2012 Research



- Ability of CaM to bind calcium and regulate calcium levels in muscle is altered



Background–Summer 2012 Research



- Ability of CaM to bind calcium and regulate calcium levels in muscle is altered
- Excess calcium enters muscles causing hyper contractions



Background–Summer 2012 Research



- Ability of CaM to bind calcium and regulate calcium levels in muscle is altered
- Excess calcium enters muscles causing hyper contractions
- Preliminary data suggested the mutation causes conformational changes in CaM in low calcium environments



Background–Summer 2012 Research



Background–Summer 2012 Research

- Summer research goals



Background–Summer 2012 Research

- Summer research goals
 - Obtain 2D and 3D NMR spectra of CaM



Background–Summer 2012 Research

- Summer research goals
 - Obtain 2D and 3D NMR spectra of CaM
 - Analyze data and infer conformational changes



Background–Summer 2012 Research

- Summer research goals
 - Obtain 2D and 3D NMR spectra of CaM
 - Analyze data and infer conformational changes
- Procedure at a glance



Background–Summer 2012 Research

- Summer research goals
 - Obtain 2D and 3D NMR spectra of CaM
 - Analyze data and infer conformational changes
- Procedure at a glance
 - Expressed CaM in host bacteria, E. Coli



Background–Summer 2012 Research

- Summer research goals
 - Obtain 2D and 3D NMR spectra of CaM
 - Analyze data and infer conformational changes
- Procedure at a glance
 - Expressed CaM in host bacteria, E. Coli
 - Purified CaM sample using hydrophobic interaction



Background–Summer 2012 Research

- Summer research goals
 - Obtain 2D and 3D NMR spectra of CaM
 - Analyze data and infer conformational changes
- Procedure at a glance
 - Expressed CaM in host bacteria, E. Coli
 - Purified CaM sample using hydrophobic interaction
 - Obtained NMR data device from the University of Iowa

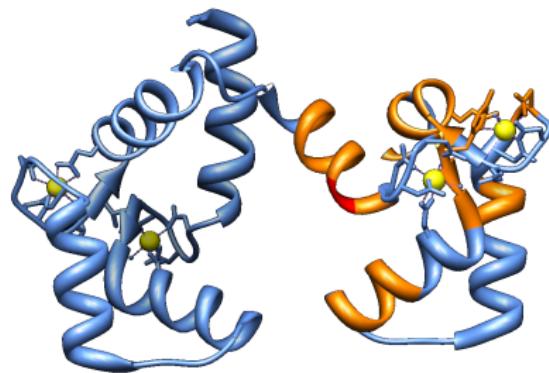


Background–Summer 2012 Research

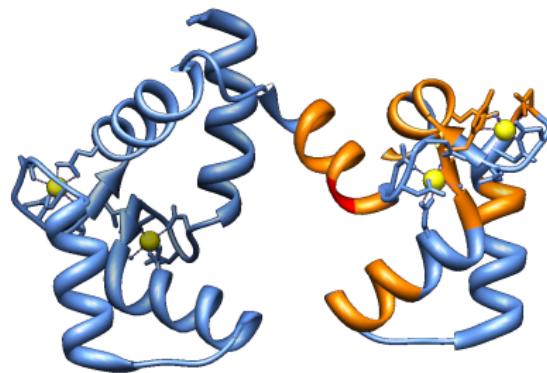
- Summer research goals
 - Obtain 2D and 3D NMR spectra of CaM
 - Analyze data and infer conformational changes
- Procedure at a glance
 - Expressed CaM in host bacteria, E. Coli
 - Purified CaM sample using hydrophobic interaction
 - Obtained NMR data device from the University of Iowa
 - Manually analyzed data and compared chemical shifts from V91G mutant with previously assigned wild type data



Background–Summer 2012 Research



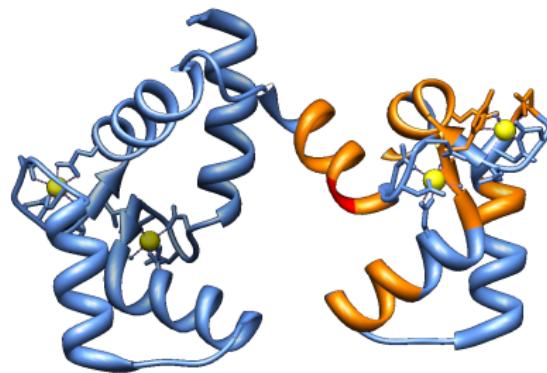
Background–Summer 2012 Research



- Summer research results



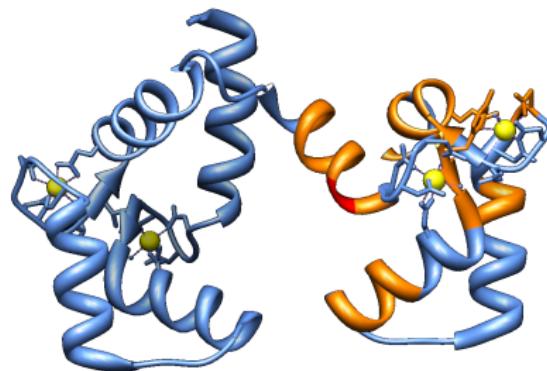
Background–Summer 2012 Research



- Summer research results
 - Collected dataset was incomplete (missing data in C-lobe)



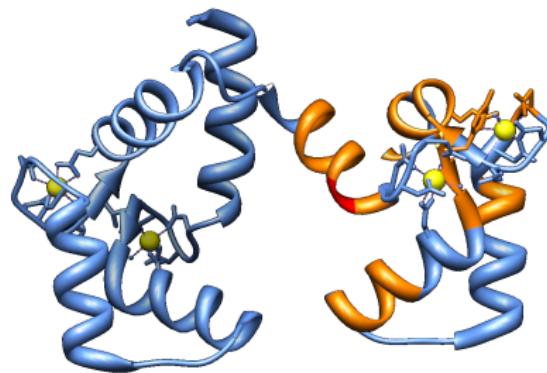
Background–Summer 2012 Research



- Summer research results
 - Collected dataset was incomplete (missing data in C-lobe)
 - Missing data could indicate V91G mutation destabilizes local helical structure



Background–Summer 2012 Research



- Summer research results
 - Collected dataset was incomplete (missing data in C-lobe)
 - Missing data could indicate V91G mutation destabilizes local helical structure
 - Further analysis is necessary



Background–Summer 2012 Research



Background–Summer 2012 Research

- Data analysis challenges



Background–Summer 2012 Research

- Data analysis challenges
 - Similar chemical environments in two lobes



Background–Summer 2012 Research

- Data analysis challenges
 - Similar chemical environments in two lobes
 - Missing data peaks



Background–Summer 2012 Research

- Data analysis challenges
 - Similar chemical environments in two lobes
 - Missing data peaks
 - Noisy data



Background–Summer 2012 Research

- Data analysis challenges
 - Similar chemical environments in two lobes
 - Missing data peaks
 - Noisy data
- Analysis of the V91G CaM mutant took on the order of weeks



Background–Summer 2012 Research

- Data analysis challenges
 - Similar chemical environments in two lobes
 - Missing data peaks
 - Noisy data
- Analysis of the V91G CaM mutant took on the order of weeks
 - NMR data analysis has traditionally been done manually



Background–Summer 2012 Research

- Data analysis challenges
 - Similar chemical environments in two lobes
 - Missing data peaks
 - Noisy data
- Analysis of the V91G CaM mutant took on the order of weeks
 - NMR data analysis has traditionally been done manually
 - Data analysis for larger proteins would be even more difficult



Background–Summer 2012 Research

- Data analysis challenges
 - Similar chemical environments in two lobes
 - Missing data peaks
 - Noisy data
- Analysis of the V91G CaM mutant took on the order of weeks
 - NMR data analysis has traditionally been done manually
 - Data analysis for larger proteins would be even more difficult
- Difficult analysis procedure prompted research for automated assignment program



Motivation



Motivation

- Scientific advancement



Motivation

- Scientific advancement
 - Genetic research has lead to the discovery of many new genes with associated protein of unknown structure and function



Motivation

- Scientific advancement
 - Genetic research has lead to the discovery of many new genes with associated protein of unknown structure and function
 - Diseases may be better understood with analysis by NMR



Motivation

- Scientific advancement

- Genetic research has lead to the discovery of many new genes with associated protein of unknown structure and function
- Diseases may be better understood with analysis by NMR
- High throughput data analysis is key to continued advancement



Motivation

- Scientific advancement
 - Genetic research has lead to the discovery of many new genes with associated protein of unknown structure and function
 - Diseases may be better understood with analysis by NMR
 - High throughput data analysis is key to continued advancement
- Artificial Intelligence



Motivation

- Scientific advancement
 - Genetic research has lead to the discovery of many new genes with associated protein of unknown structure and function
 - Diseases may be better understood with analysis by NMR
 - High throughput data analysis is key to continued advancement
- Artificial Intelligence
 - Data analysis could be greatly improved with simple AI strategies

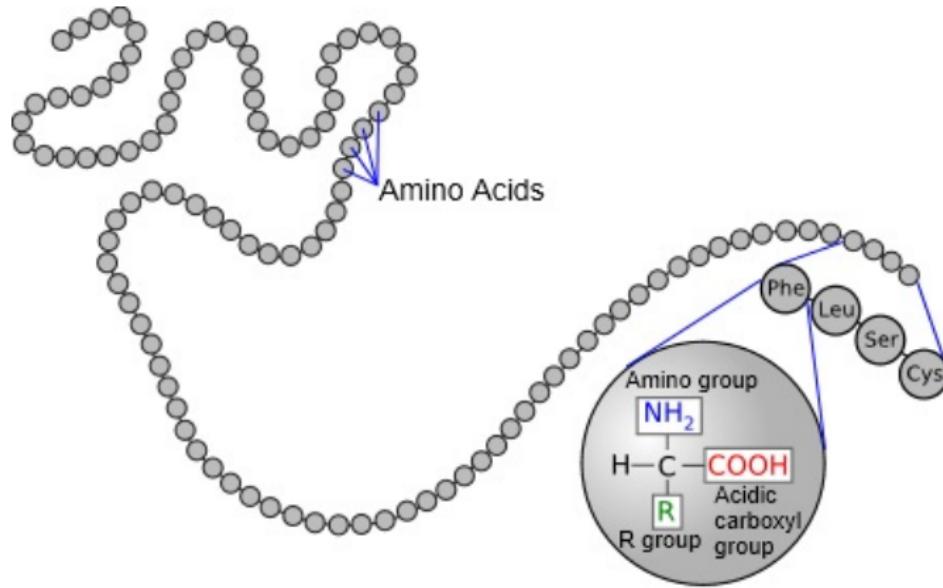


Motivation

- Scientific advancement
 - Genetic research has lead to the discovery of many new genes with associated protein of unknown structure and function
 - Diseases may be better understood with analysis by NMR
 - High throughput data analysis is key to continued advancement
- Artificial Intelligence
 - Data analysis could be greatly improved with simple AI strategies
 - Connected well with research from the summer and undergraduate coursework



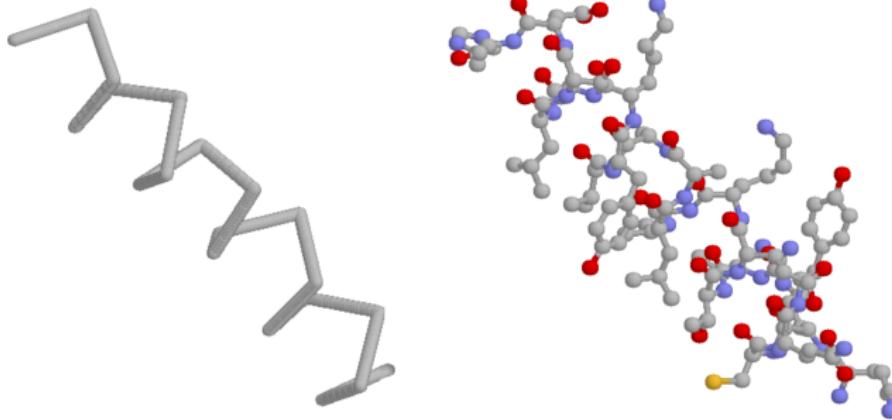
Data Collection and Structural Inferences



Protein chain amino acids/residues



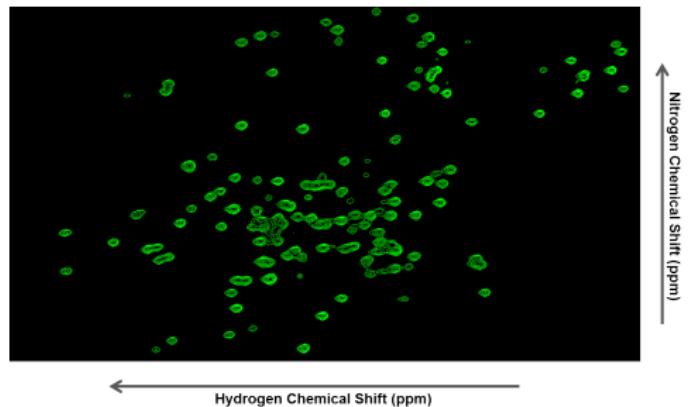
Data Collection and Structural Inferences



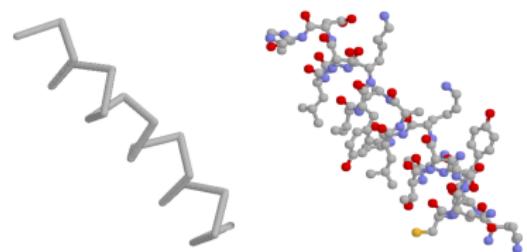
Left: protein backbone Right: protein backbone + side chains



Data Collection and Structural Inferences



HSQC NMR Spectrum



Data Collection and Structural Inferences



Data Collection and Structural Inferences

- Obtaining structural information from NMR



Data Collection and Structural Inferences

- Obtaining structural information from NMR
 - NMR provides information about each residue with a numeric value, chemical shift



Data Collection and Structural Inferences

- Obtaining structural information from NMR
 - NMR provides information about each residue with a numeric value, chemical shift
 - Chemical shifts depend on the local chemical environment of the residue



Data Collection and Structural Inferences

- Obtaining structural information from NMR
 - NMR provides information about each residue with a numeric value, chemical shift
 - Chemical shifts depend on the local chemical environment of the residue
 - By comparing chemical shifts with known data, structural information can be inferred



Data Collection and Structural Inferences

- Obtaining structural information from NMR
 - NMR provides information about each residue with a numeric value, chemical shift
 - Chemical shifts depend on the local chemical environment of the residue
 - By comparing chemical shifts with known data, structural information can be inferred
- Data analysis requires the assignment of chemical shifts



Data Collection and Structural Inferences

- Obtaining structural information from NMR
 - NMR provides information about each residue with a numeric value, chemical shift
 - Chemical shifts depend on the local chemical environment of the residue
 - By comparing chemical shifts with known data, structural information can be inferred
- Data analysis requires the assignment of chemical shifts
 - Assignment means making a correspondence between residues in the protein sequence and chemical shift from spectra

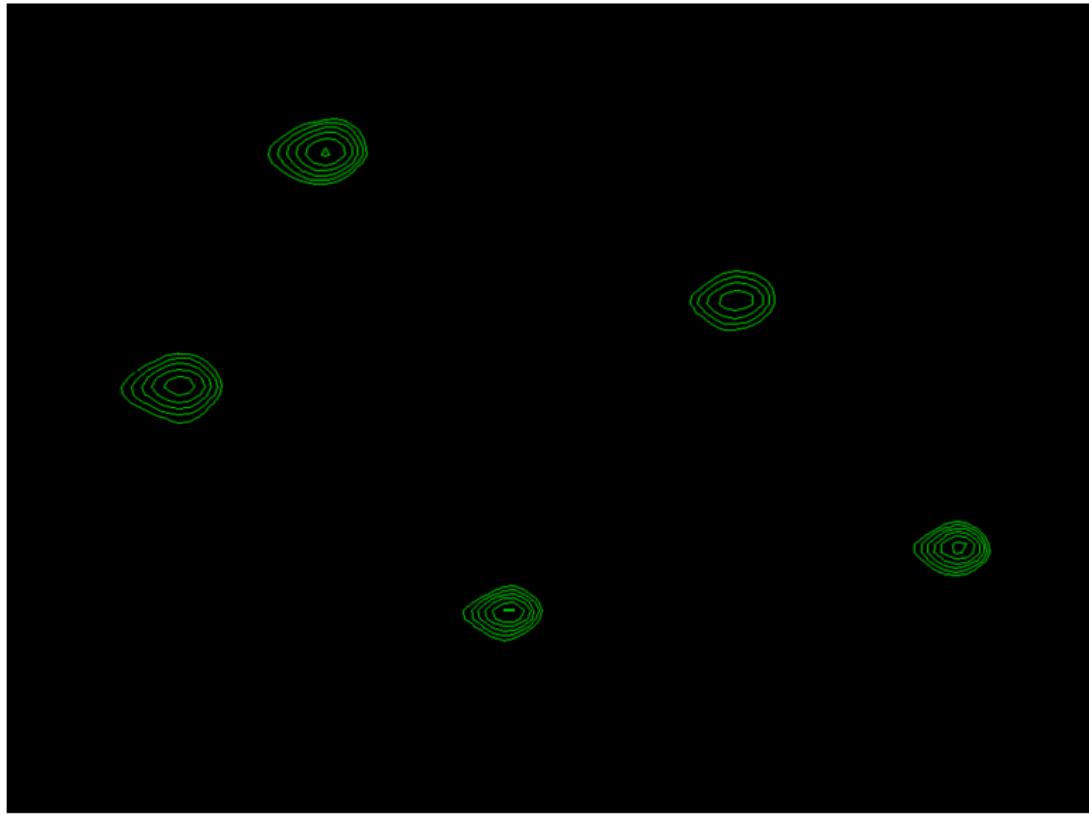


Data Collection and Structural Inferences

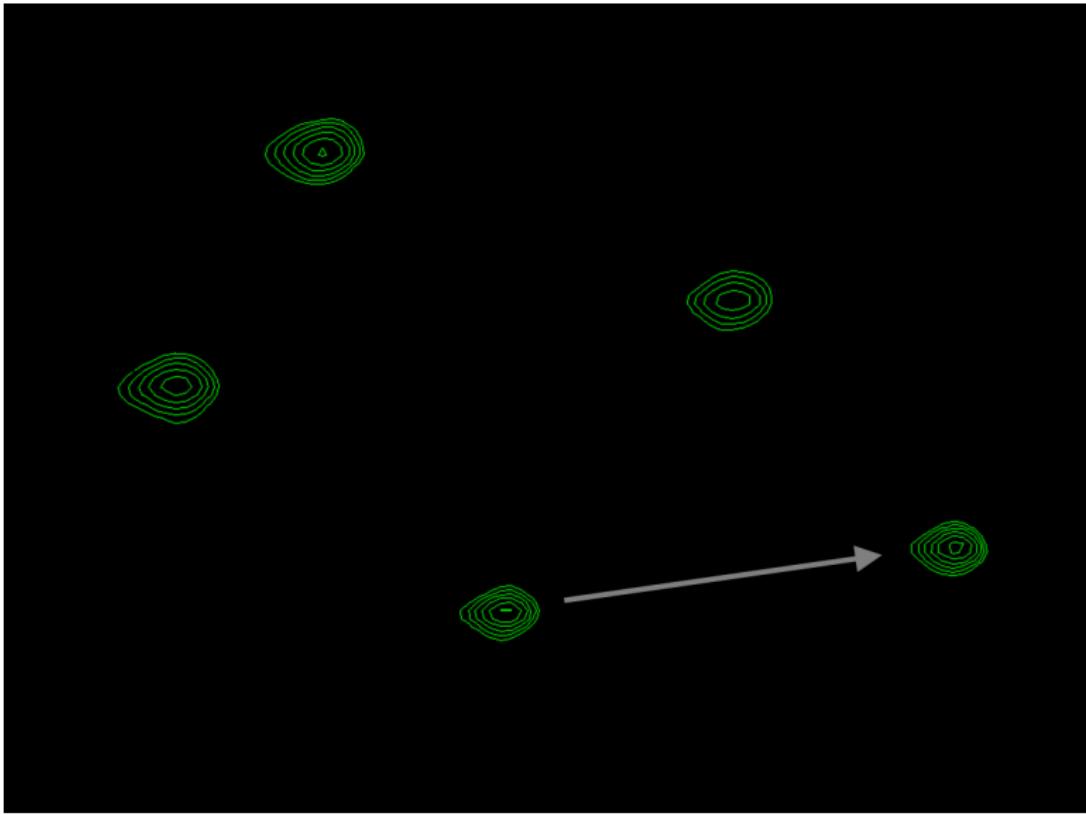
- Obtaining structural information from NMR
 - NMR provides information about each residue with a numeric value, chemical shift
 - Chemical shifts depend on the local chemical environment of the residue
 - By comparing chemical shifts with known data, structural information can be inferred
- Data analysis requires the assignment of chemical shifts
 - Assignment means making a correspondence between residues in the protein sequence and chemical shift from spectra
 - Typically done manually; requires substantial time and expertise with NMR



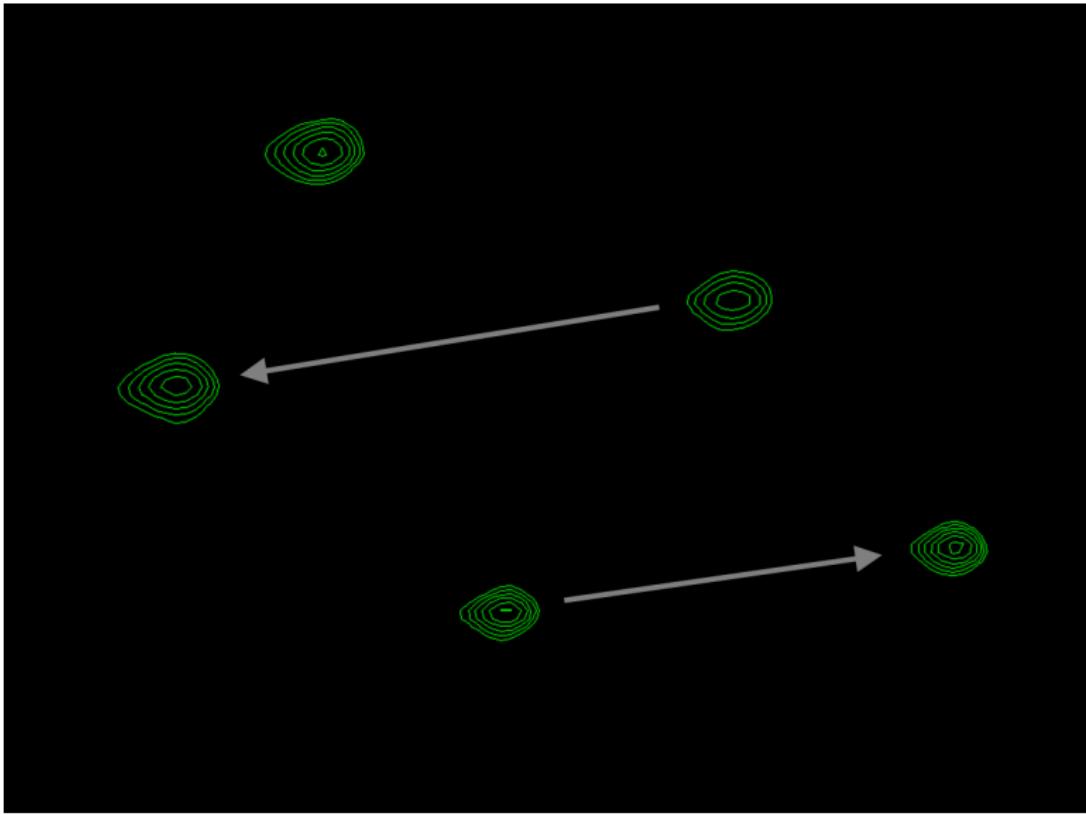
NMR Assignment Goal



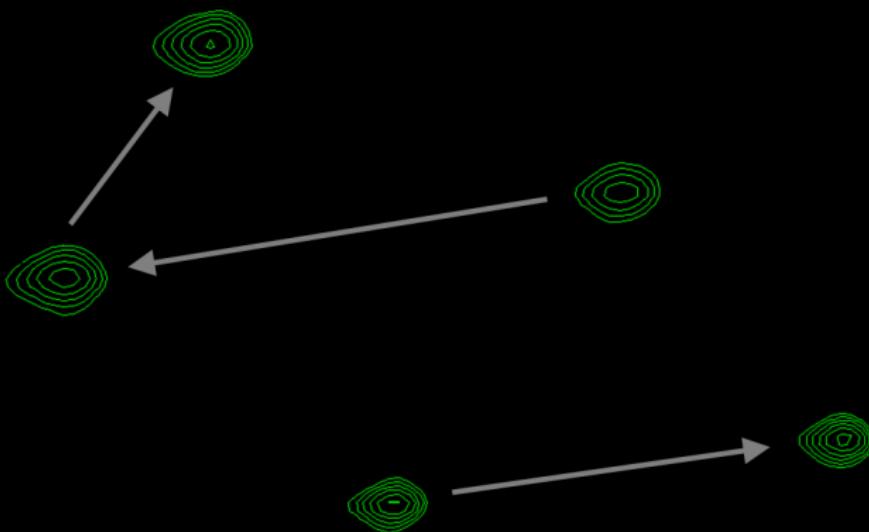
NMR Assignment Goal



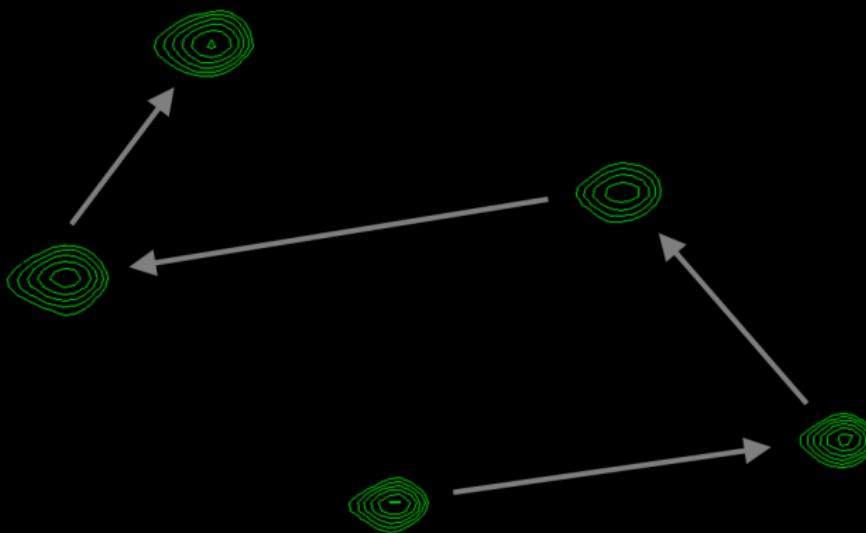
NMR Assignment Goal



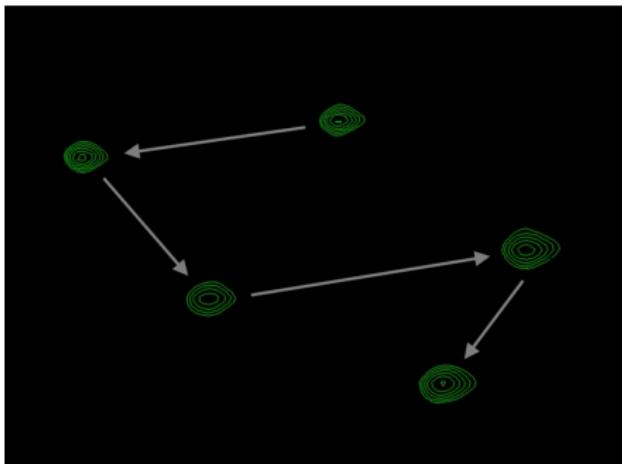
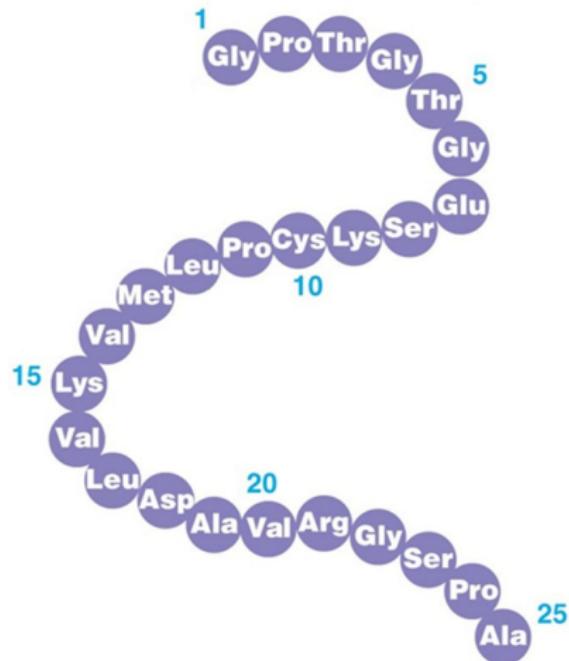
NMR Assignment Goal



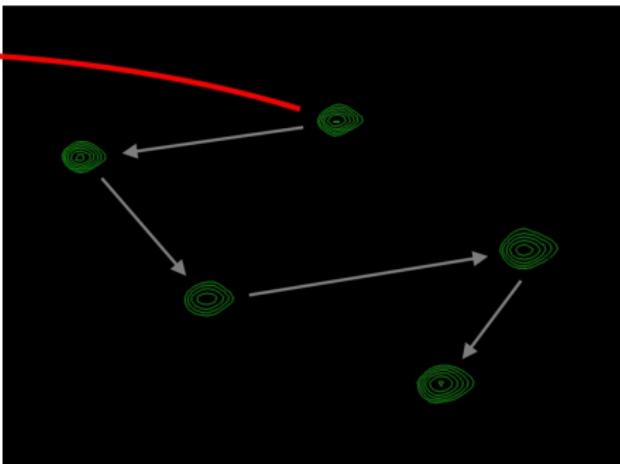
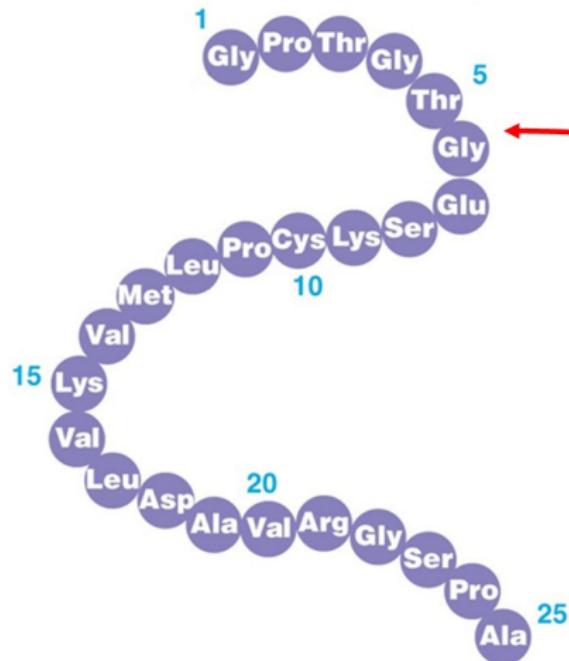
NMR Assignment Goal



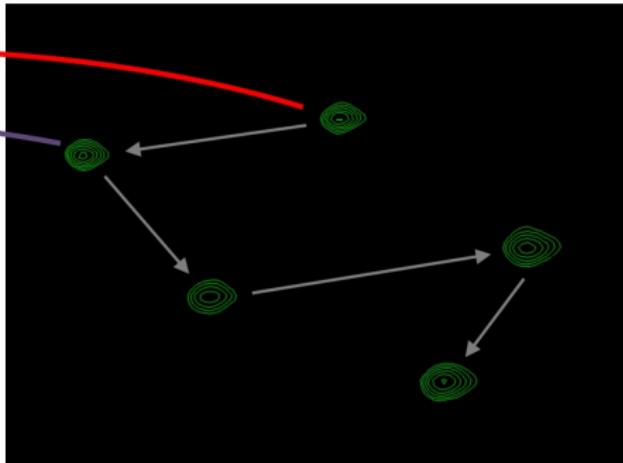
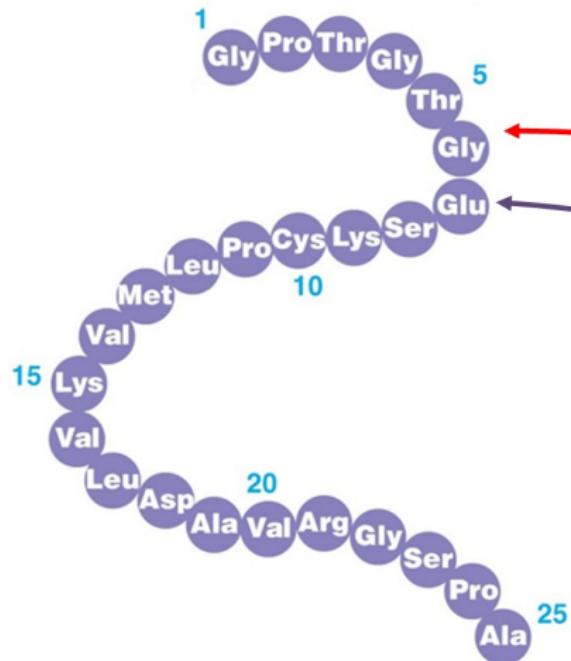
NMR Assignment Goal



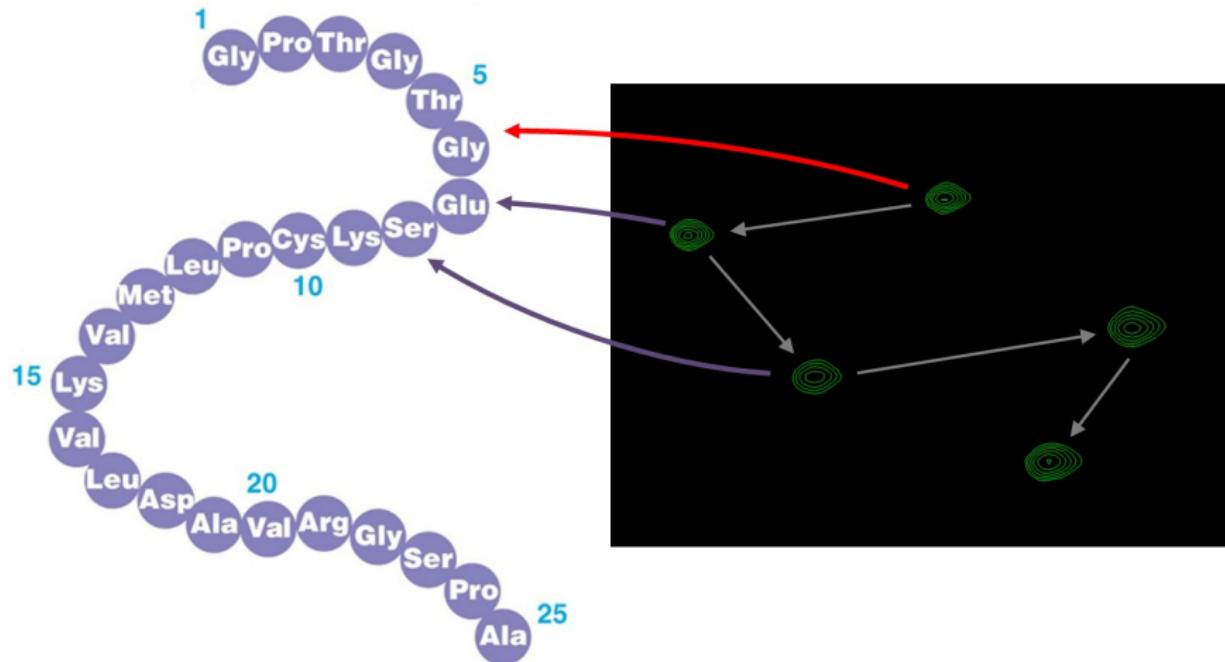
NMR Assignment Goal



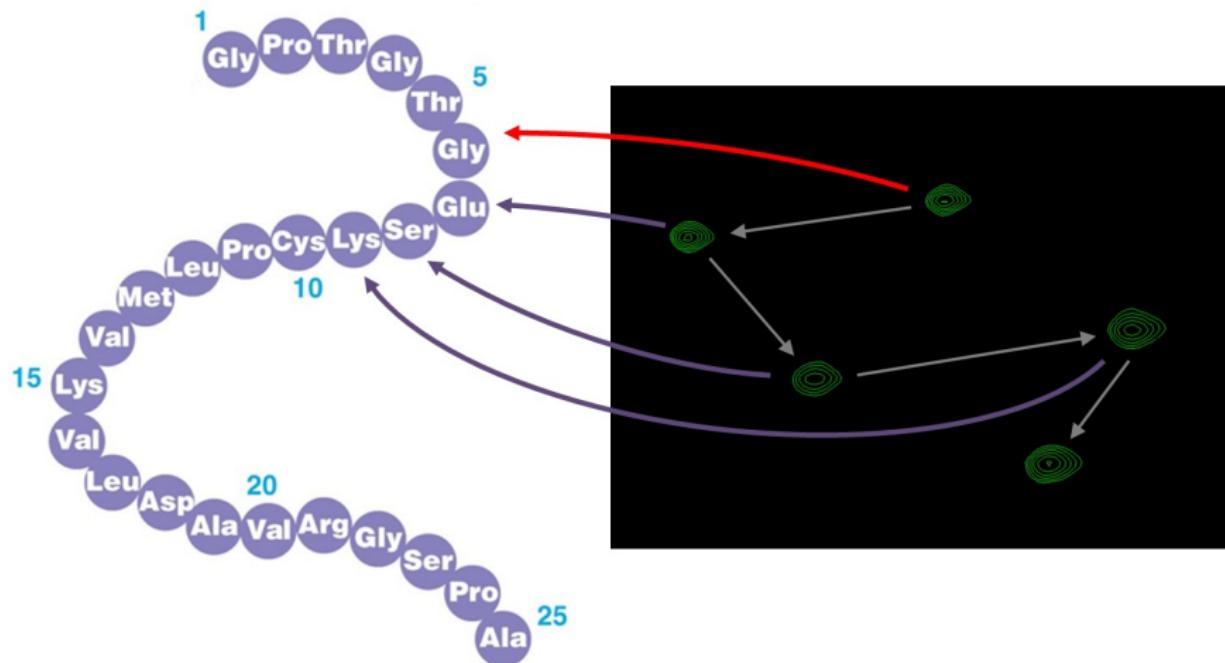
NMR Assignment Goal



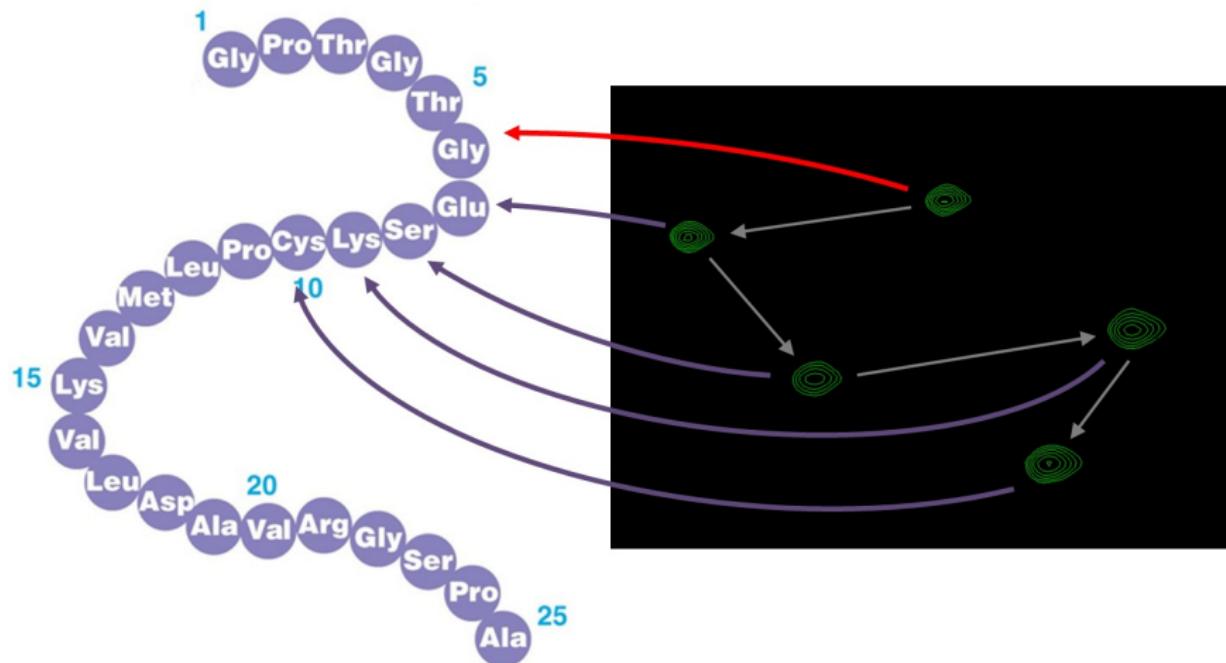
NMR Assignment Goal



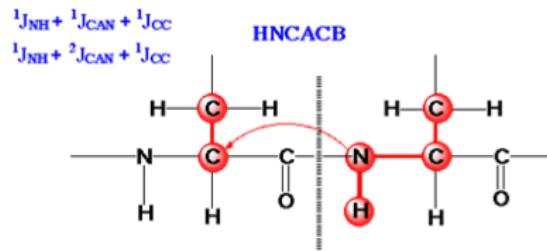
NMR Assignment Goal



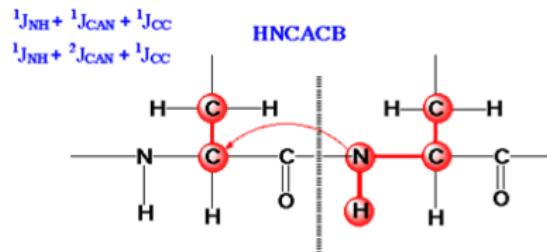
NMR Assignment Goal



Assignment Protocol



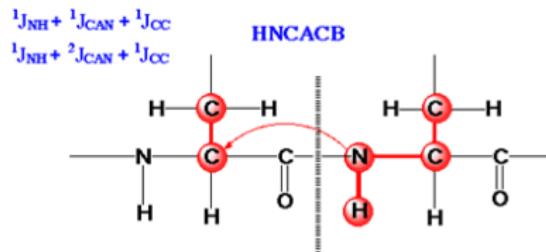
Assignment Protocol



- Some NMR experiments describe the sequence of chemical shifts



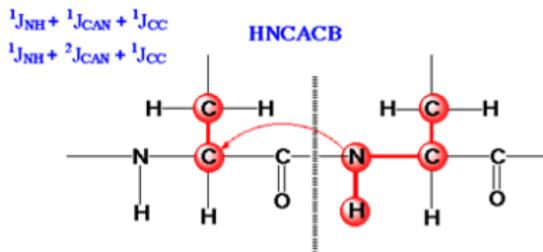
Assignment Protocol



- Some NMR experiments describe the sequence of chemical shifts
 - 3D NMR experiments yield C_α and C_β chemical shifts which help link residues to chemical shifts



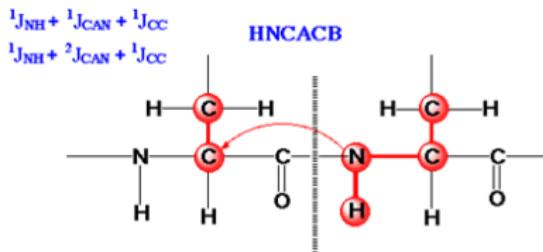
Assignment Protocol



- Some NMR experiments describe the sequence of chemical shifts
 - 3D NMR experiments yield C_α and C_β chemical shifts which help link residues to chemical shifts
 - HNCACB gives C_α and C_β chemical shifts corresponding to a given residue and the residue 'in front' of it



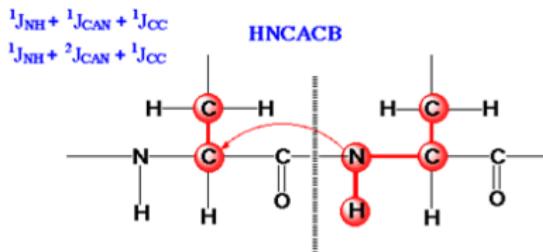
Assignment Protocol



- Some NMR experiments describe the sequence of chemical shifts
 - 3D NMR experiments yield C_α and C_β chemical shifts which help link residues to chemical shifts
 - HNCACB gives C_α and C_β chemical shifts corresponding to a given residue and the residue 'in front' of it
 - CBCA(CO)HN NMR experiment produces C_α and C_β chemical shifts corresponding to only the 'in front' residue



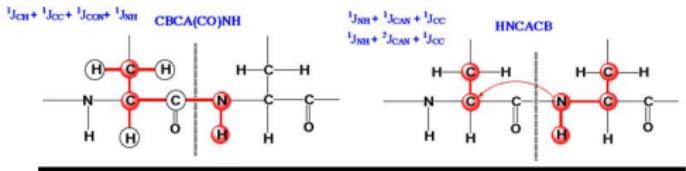
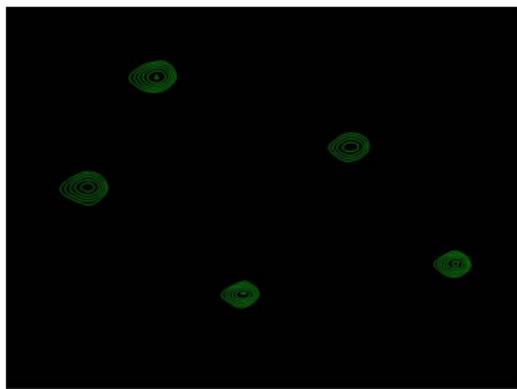
Assignment Protocol



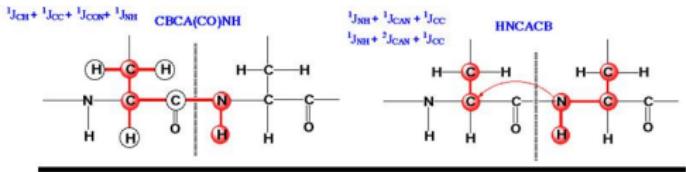
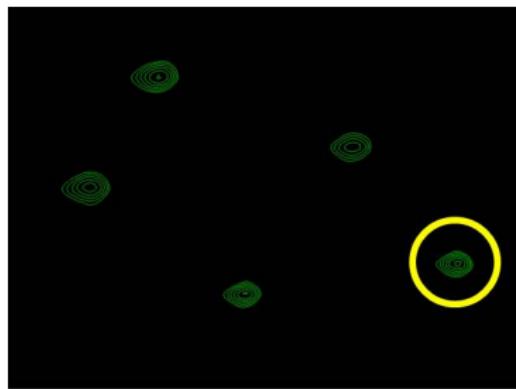
- Some NMR experiments describe the sequence of chemical shifts
 - 3D NMR experiments yield C_α and C_β chemical shifts which help link residues to chemical shifts
 - HNCACB gives C_α and C_β chemical shifts corresponding to a given residue and the residue 'in front' of it
 - CBCA(CO)HN NMR experiment produces C_α and C_β chemical shifts corresponding to only the 'in front' residue
- Analyzing multiple NMR experiments simultaneously allows for unambiguous assignment of residues



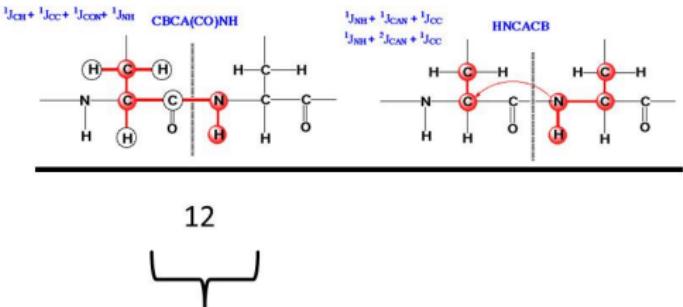
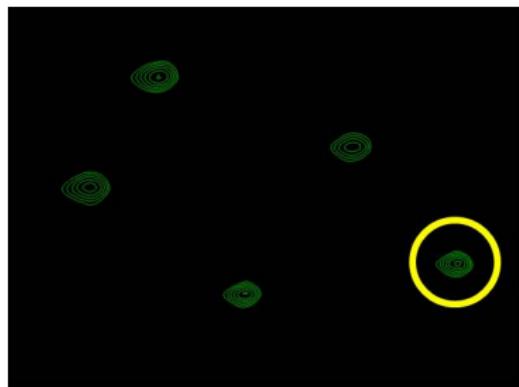
Assignment Protocol



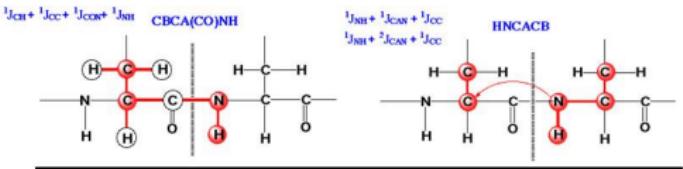
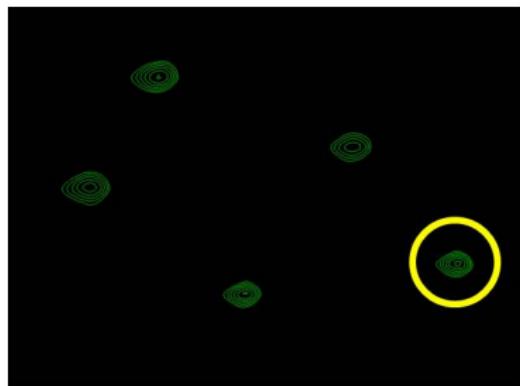
Assignment Protocol



Assignment Protocol



Assignment Protocol



12



CBCA(CO)NH yields the i-1 residue chemical shifts

5

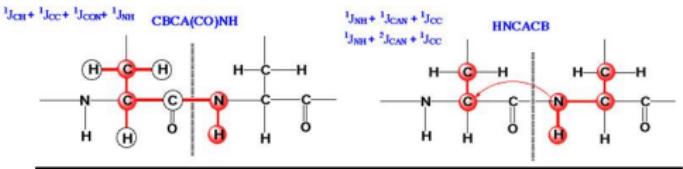
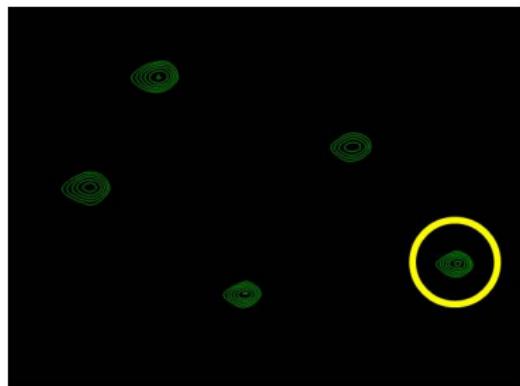
12



HNCACB yield the i and i-1 residue chemical shifts



Assignment Protocol



12 (i-1)

5

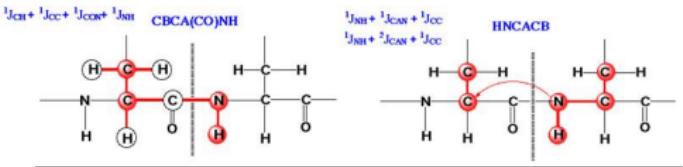
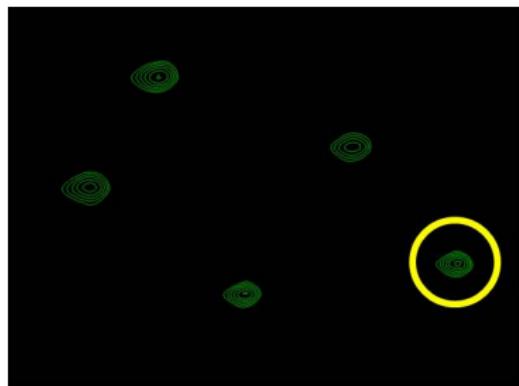
CBCA(CO)NH yields the i-1 residue chemical shifts

12

HNCACB yield the i and i-1 residue chemical shifts



Assignment Protocol

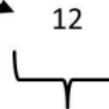


12 (i-1)



CBCA(CO)NH yields the i-1 residue chemical shifts

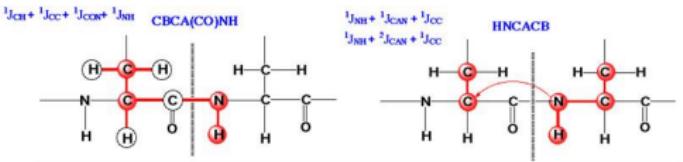
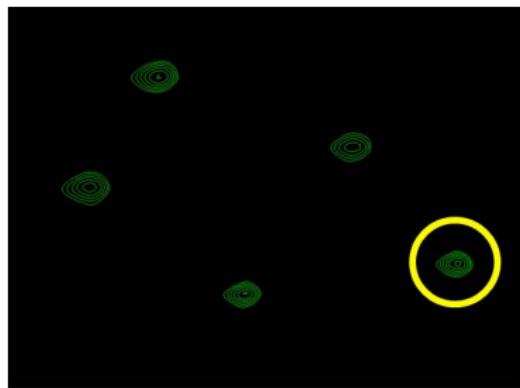
5



HNCACB yield the i and i-1 residue chemical shifts



Assignment Protocol



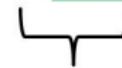
12 (i-1)



CBCA(CO)NH yields the i-1 residue chemical shifts

5

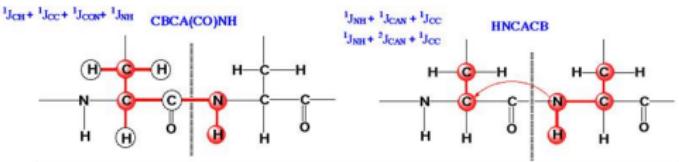
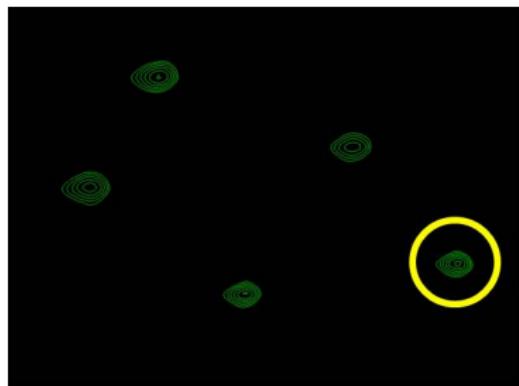
12 (i-1)



HNCACB yield the i and i-1 residue chemical shifts



Assignment Protocol



12 (i-1)



CBCA(CO)NH yields the i-1 residue chemical shifts

5 (i)

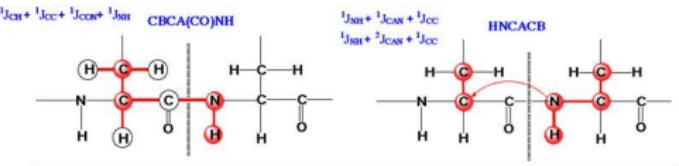
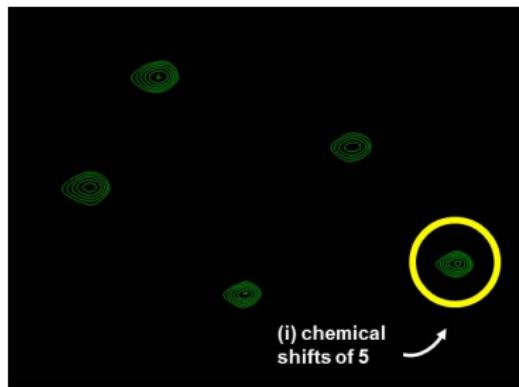
12 (i-1)



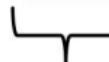
HNCACB yield the i and i-1 residue chemical shifts



Assignment Protocol



12 (i-1)



5 (i)

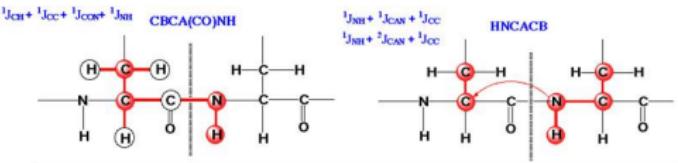
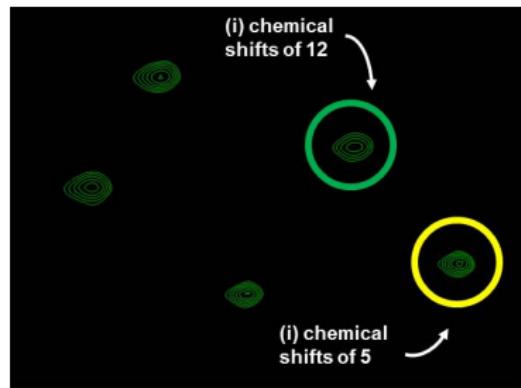
12 (i-1)

CBCA(CO)NH yields the i-1 residue chemical shifts

HNCACB yield the i and i-1 residue chemical shifts



Assignment Protocol



12 (i-1)

5 (i)

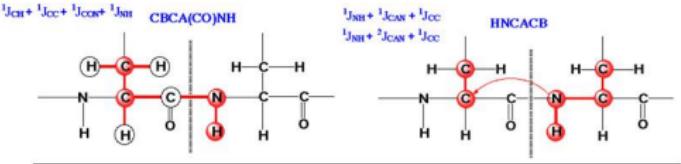
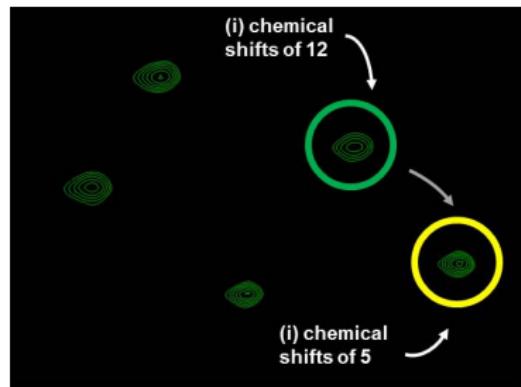
12 (i-1)

CBCA(CO)NH yields the i-1 residue chemical shifts

HNCACB yield the i and i-1 residue chemical shifts



Assignment Protocol



12 (i-1)

5 (i)

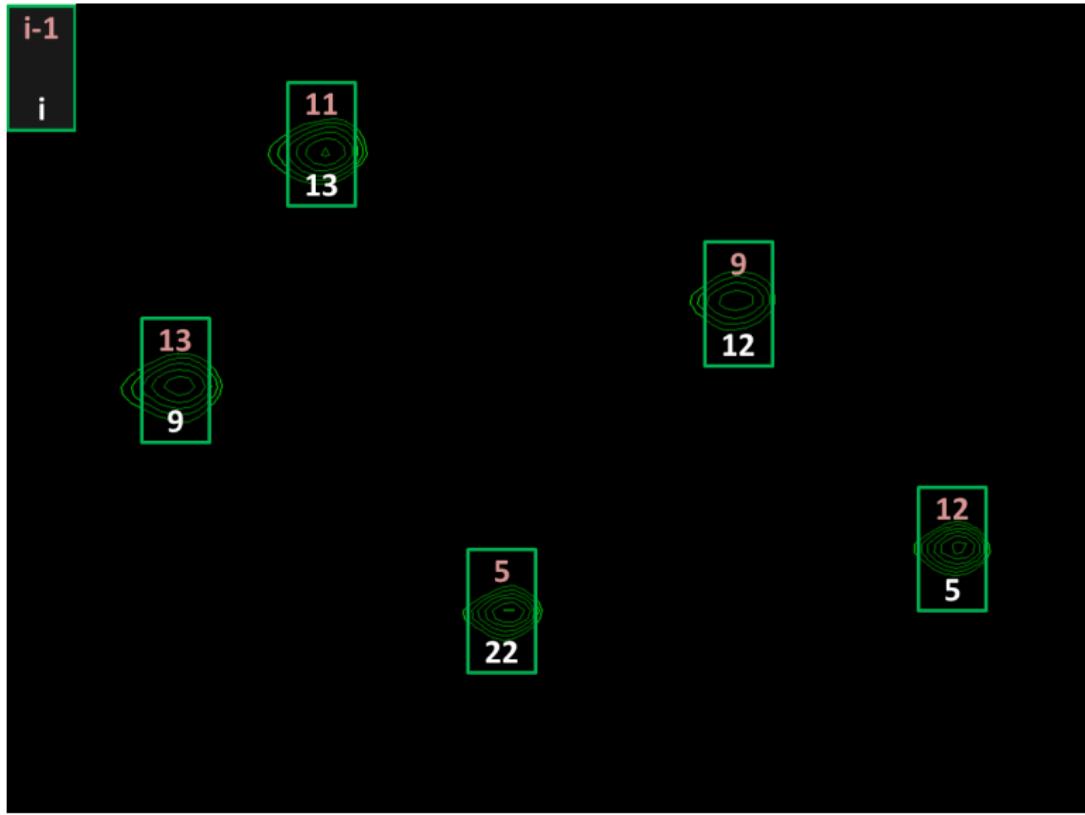
12 (i-1)

CBCA(CO)NH yields the i-1 residue chemical shifts

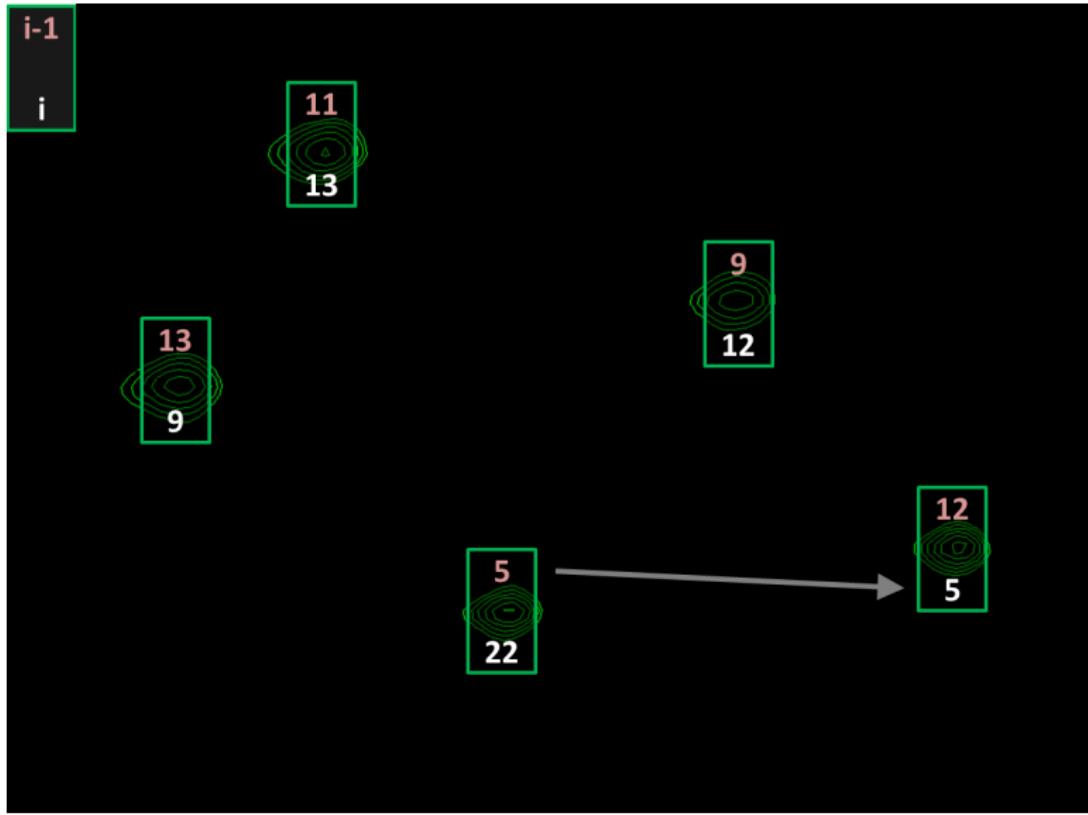
HNCACB yield the i and i-1 residue chemical shifts



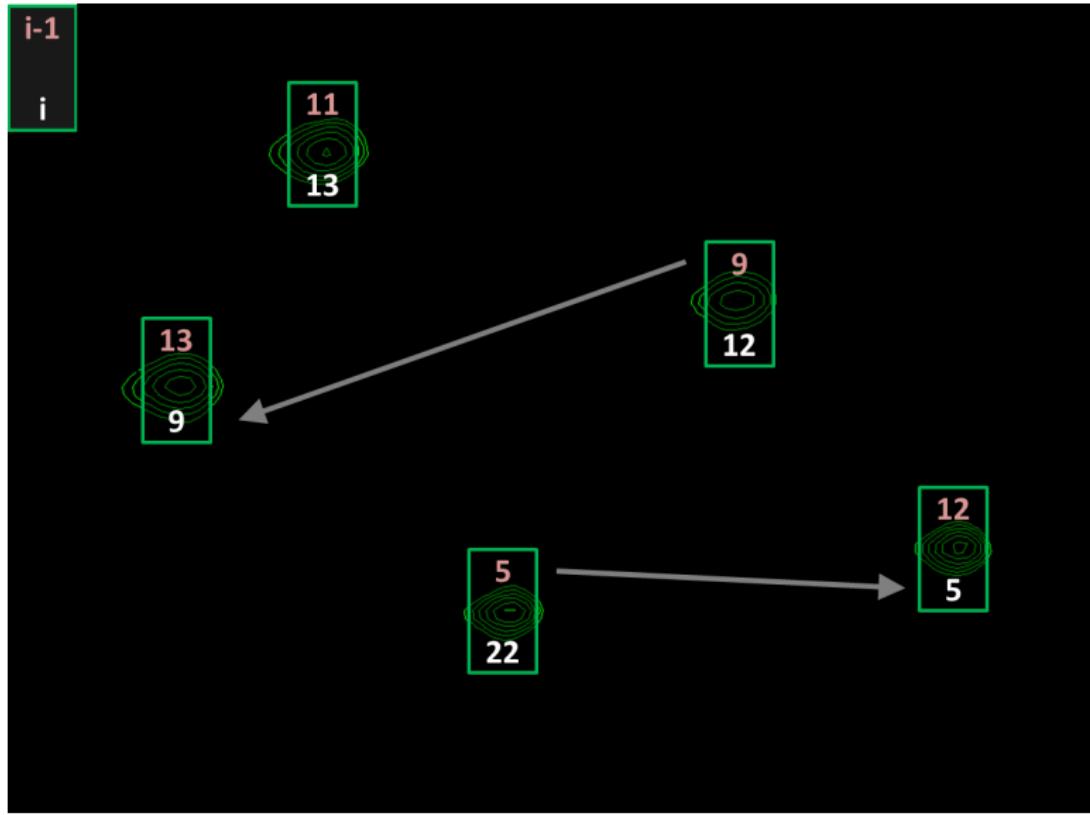
Assignment Protocol



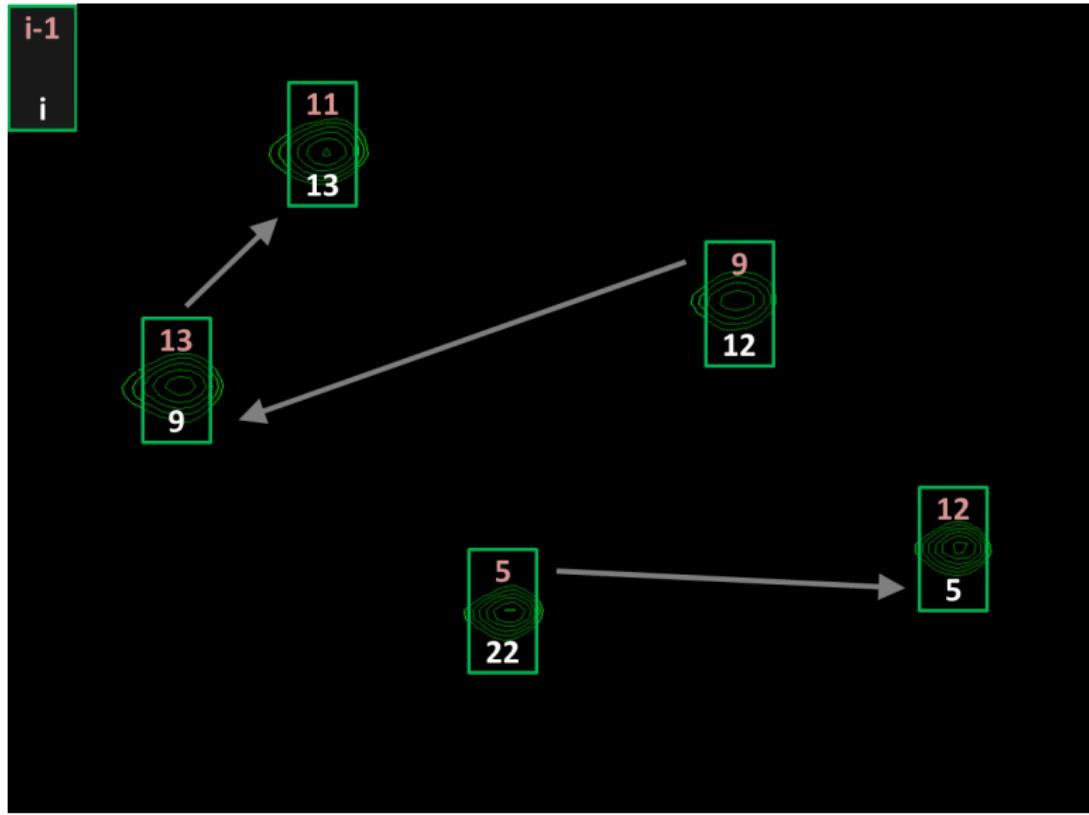
Assignment Protocol



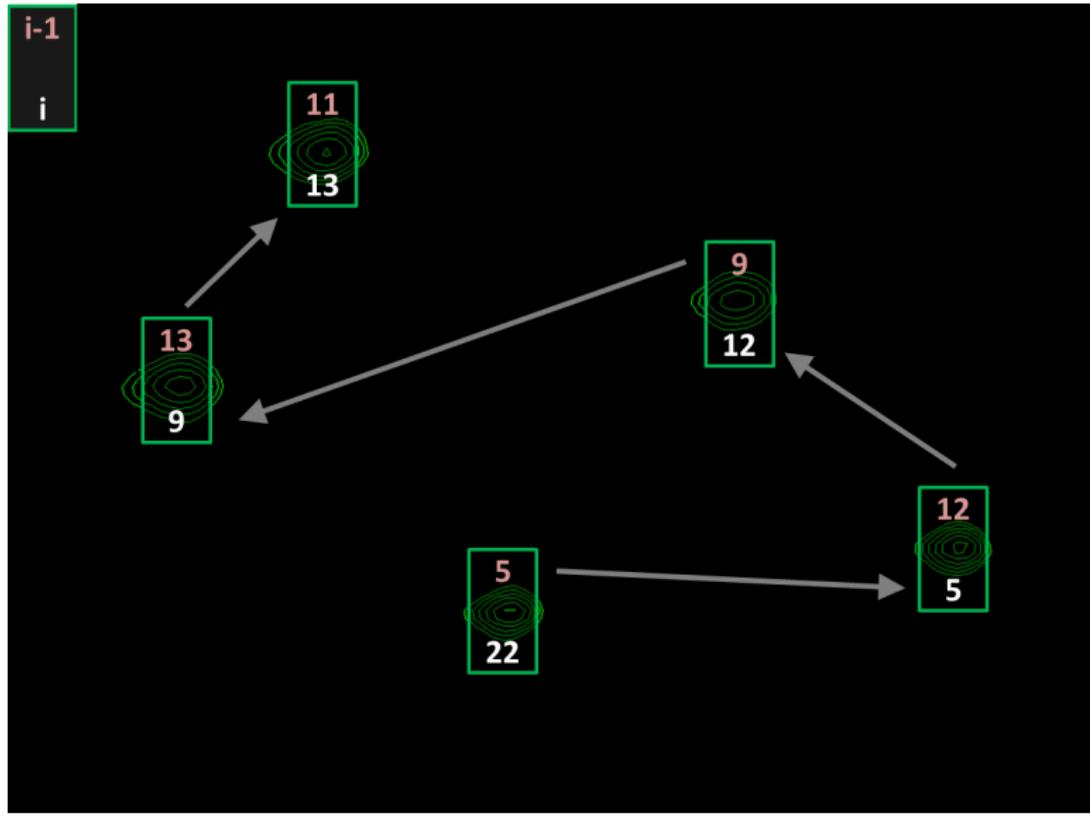
Assignment Protocol



Assignment Protocol



Assignment Protocol



Assignment Protocol



Assignment Protocol

- Characteristics of chemical shifts



Assignment Protocol

- Characteristics of chemical shifts
 - Most residues have very similar chemical shift values



Assignment Protocol

- Characteristics of chemical shifts
 - Most residues have very similar chemical shift values
 - Example, isoleucine and leucine have almost identical chemical shifts



Assignment Protocol

- Characteristics of chemical shifts
 - Most residues have very similar chemical shift values
 - Example, isoleucine and leucine have almost identical chemical shifts
 - Some residues/amino acids have very unique values for their chemical shifts



Assignment Protocol

- Characteristics of chemical shifts
 - Most residues have very similar chemical shift values
 - Example, isoleucine and leucine have almost identical chemical shifts
 - Some residues/amino acids have very unique values for their chemical shifts
 - Example, alanine can be identified also immediately by its chemical shifts



Assignment Protocol

- Characteristics of chemical shifts
 - Most residues have very similar chemical shift values
 - Example, isoleucine and leucine have almost identical chemical shifts
 - Some residues/amino acids have very unique values for their chemical shifts
 - Example, alanine can be identified also immediately by its chemical shifts
 - These residues are important because they can be easily identified without other information



Assignment Protocol

- Characteristics of chemical shifts
 - Most residues have very similar chemical shift values
 - Example, isoleucine and leucine have almost identical chemical shifts
 - Some residues/amino acids have very unique values for their chemical shifts
 - Example, alanine can be identified also immediately by its chemical shifts
 - These residues are important because they can be easily identified without other information
- Matching chemical shifts to residues in the protein sequence



Assignment Protocol

- Characteristics of chemical shifts
 - Most residues have very similar chemical shift values
 - Example, isoleucine and leucine have almost identical chemical shifts
 - Some residues/amino acids have very unique values for their chemical shifts
 - Example, alanine can be identified also immediately by its chemical shifts
 - These residues are important because they can be easily identified without other information
- Matching chemical shifts to residues in the protein sequence
 - Protein sequence is known before the experiment

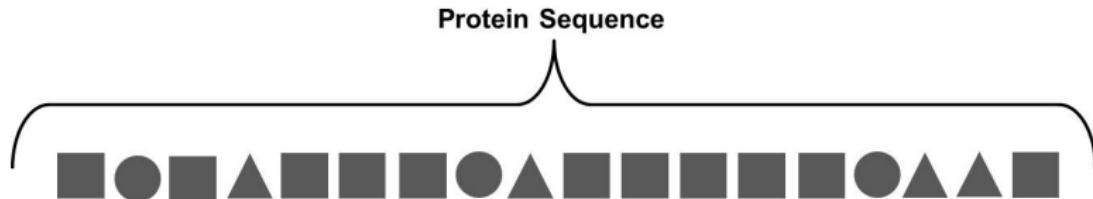


Assignment Protocol

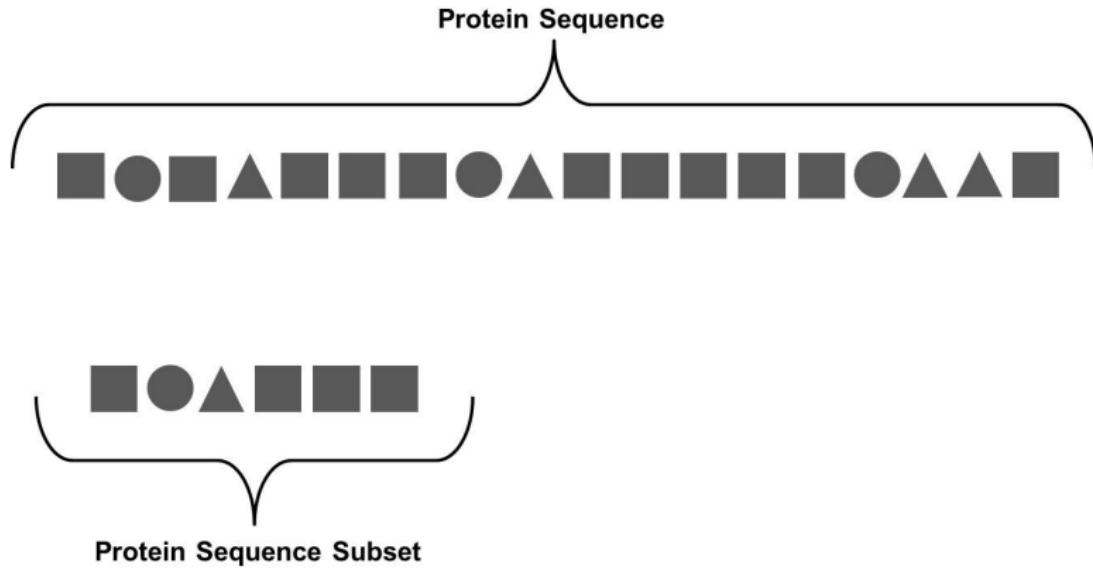
- Characteristics of chemical shifts
 - Most residues have very similar chemical shift values
 - Example, isoleucine and leucine have almost identical chemical shifts
 - Some residues/amino acids have very unique values for their chemical shifts
 - Example, alanine can be identified also immediately by its chemical shifts
 - These residues are important because they can be easily identified without other information
- Matching chemical shifts to residues in the protein sequence
 - Protein sequence is known before the experiment
 - The relative location of the easily identifiable residues is used to place sequentially-linked chemical in the protein sequence



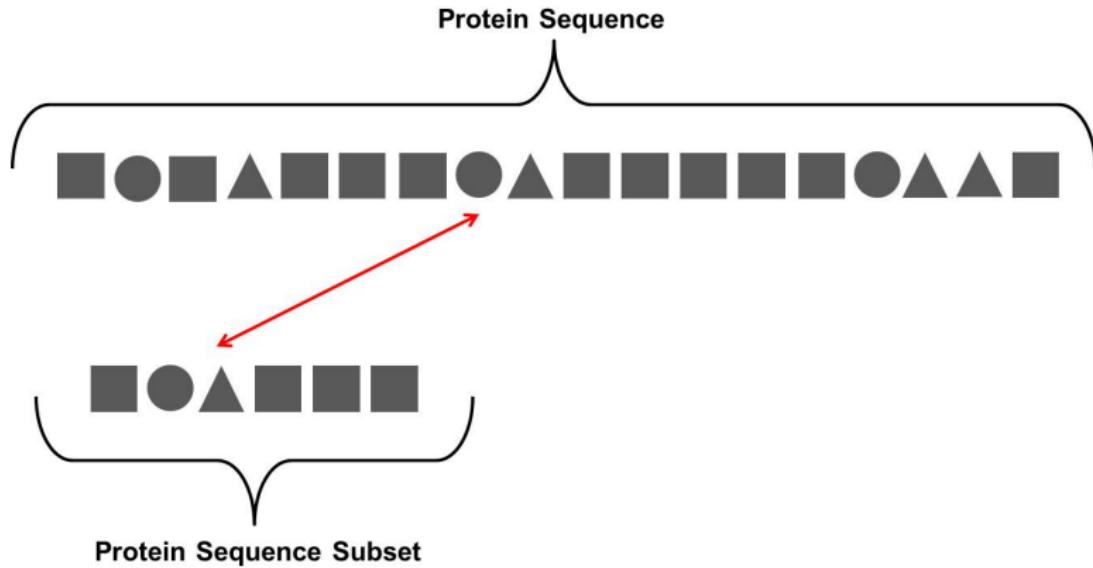
Assignment Protocol



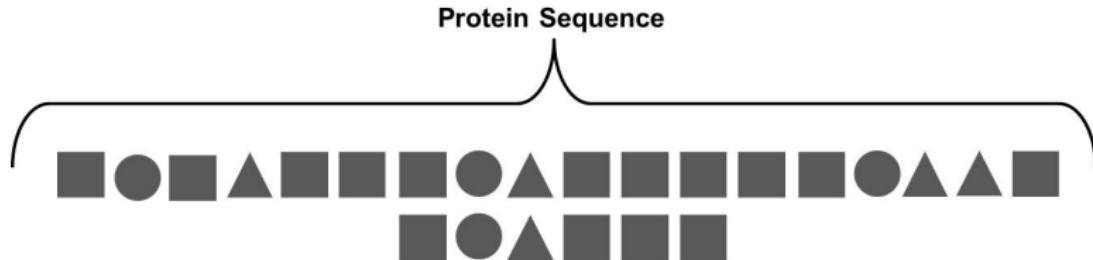
Assignment Protocol



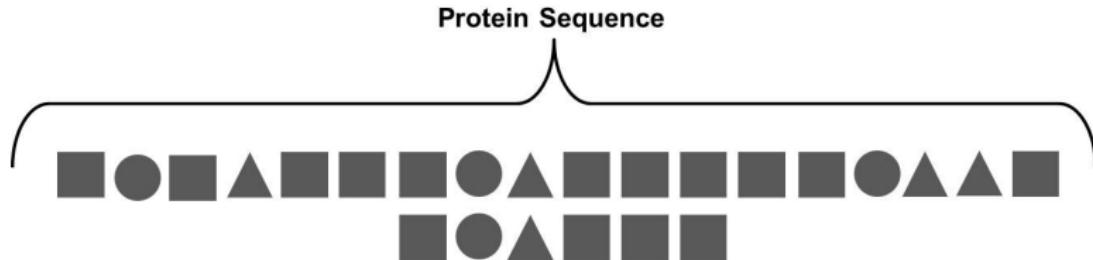
Assignment Protocol



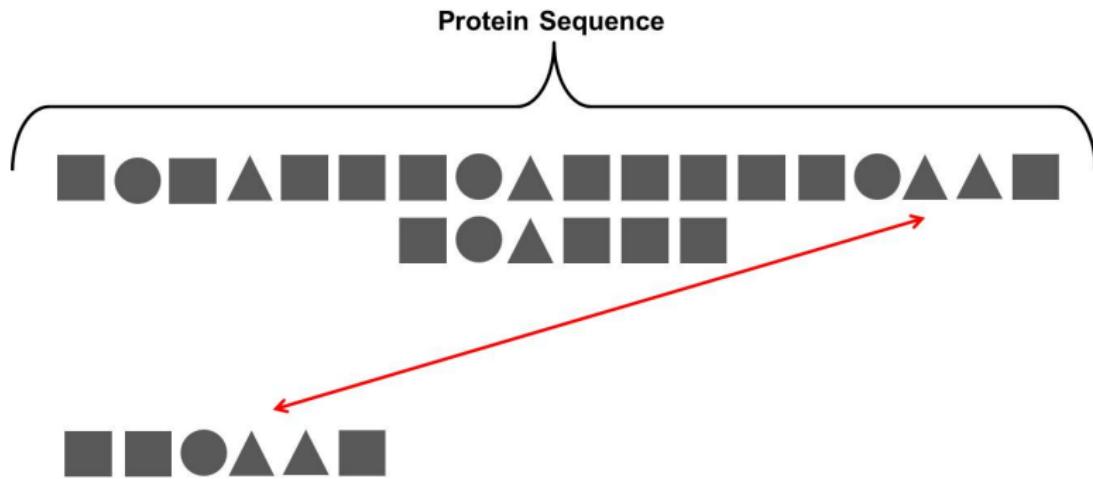
Assignment Protocol



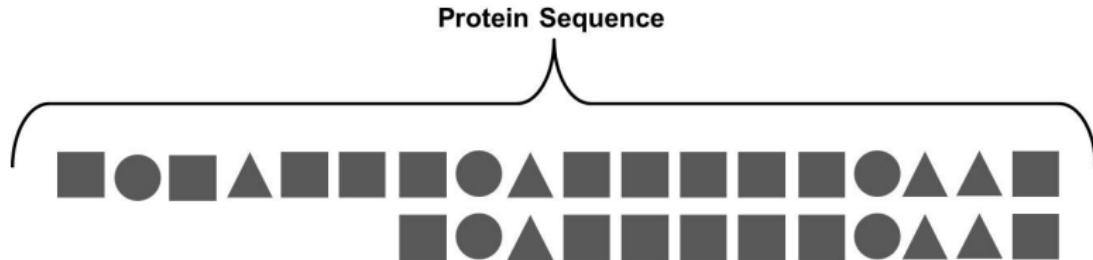
Assignment Protocol



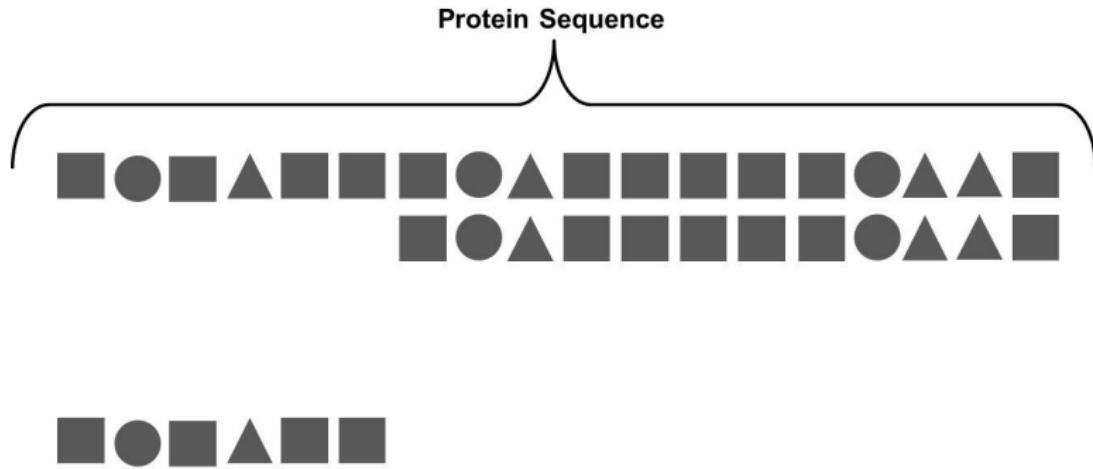
Assignment Protocol



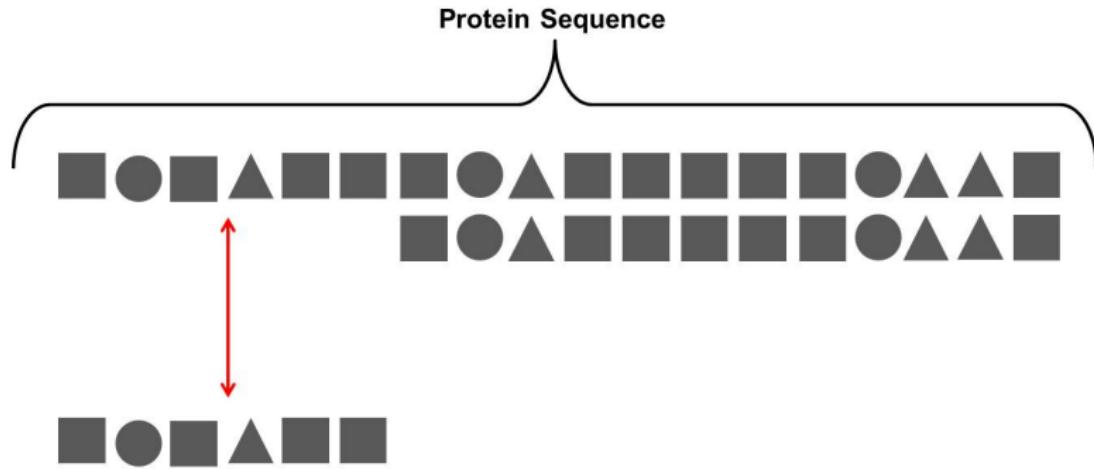
Assignment Protocol



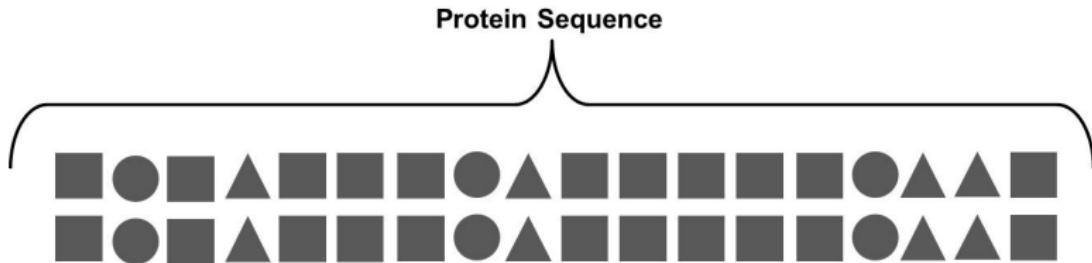
Assignment Protocol



Assignment Protocol



Assignment Protocol



Assignment Algorithm

Algorithm Outline



Assignment Algorithm

Algorithm Outline

- ① Mark highly unique subsets of protein sequences



Assignment Algorithm

Algorithm Outline

- ① Mark highly unique subsets of protein sequences
- ② Assign protein sequence subsets to sequentially-linked chemical shifts



Assignment Algorithm

Algorithm Outline

- ① Mark highly unique subsets of protein sequences
- ② Assign protein sequence subsets to sequentially-linked chemical shifts
- ③ Use iterative greedy search algorithm to complete assignment of all remaining residues



Assignment Algorithm

Algorithm Outline

- ① Mark highly unique subsets of protein sequences
 - ② Assign protein sequence subsets to sequentially-linked chemical shifts
 - ③ Use iterative greedy search algorithm to complete assignment of all remaining residues
-
- Steps (1) and (2) are designed to reduce the complexity of the dataset



Assignment Algorithm

Algorithm Outline

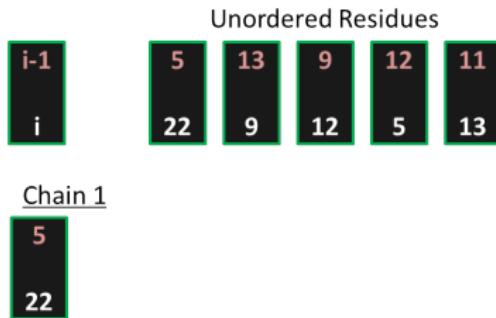
- ① Mark highly unique subsets of protein sequences
 - ② Assign protein sequence subsets to sequentially-linked chemical shifts
 - ③ Use iterative greedy search algorithm to complete assignment of all remaining residues
-
- Steps (1) and (2) are designed to reduce the complexity of the dataset
 - Step (3) is computationally expensive but will assign all remaining residues



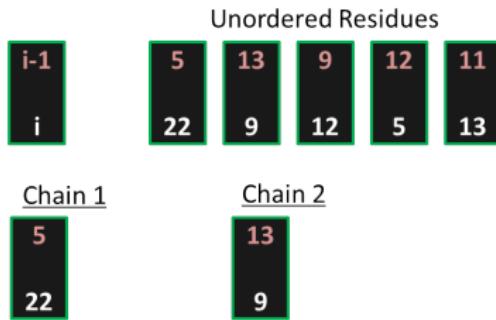
Assignment Algorithm Illustrated



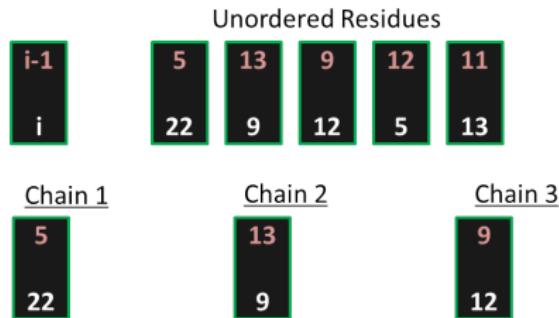
Assignment Algorithm Illustrated



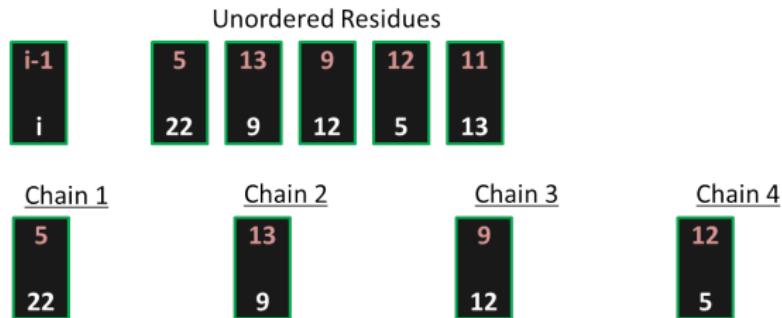
Assignment Algorithm Illustrated



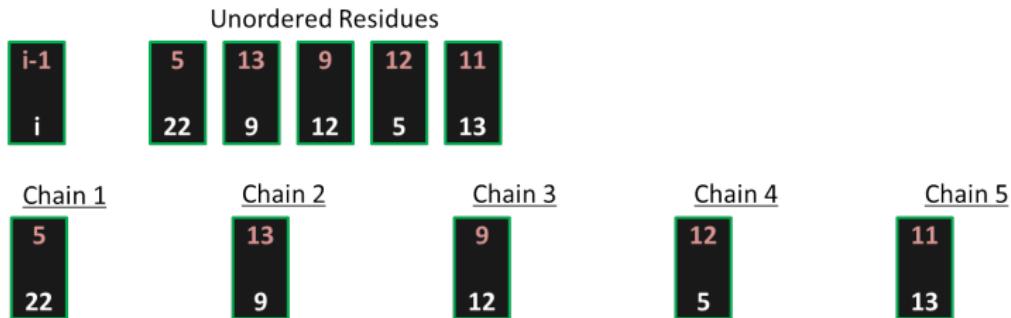
Assignment Algorithm Illustrated



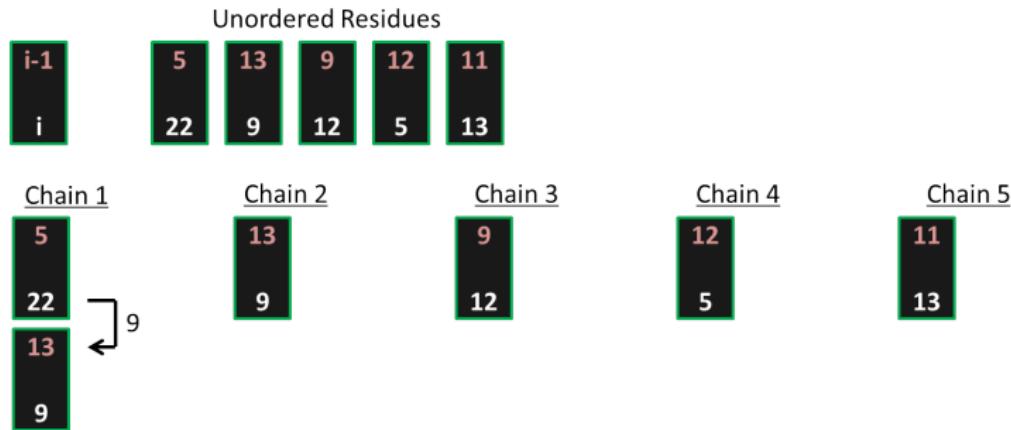
Assignment Algorithm Illustrated



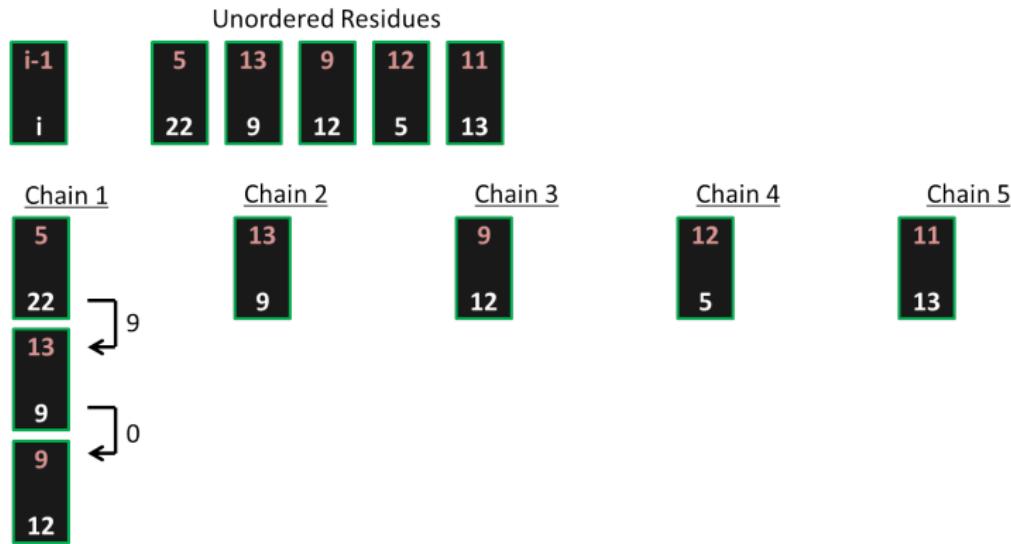
Assignment Algorithm Illustrated



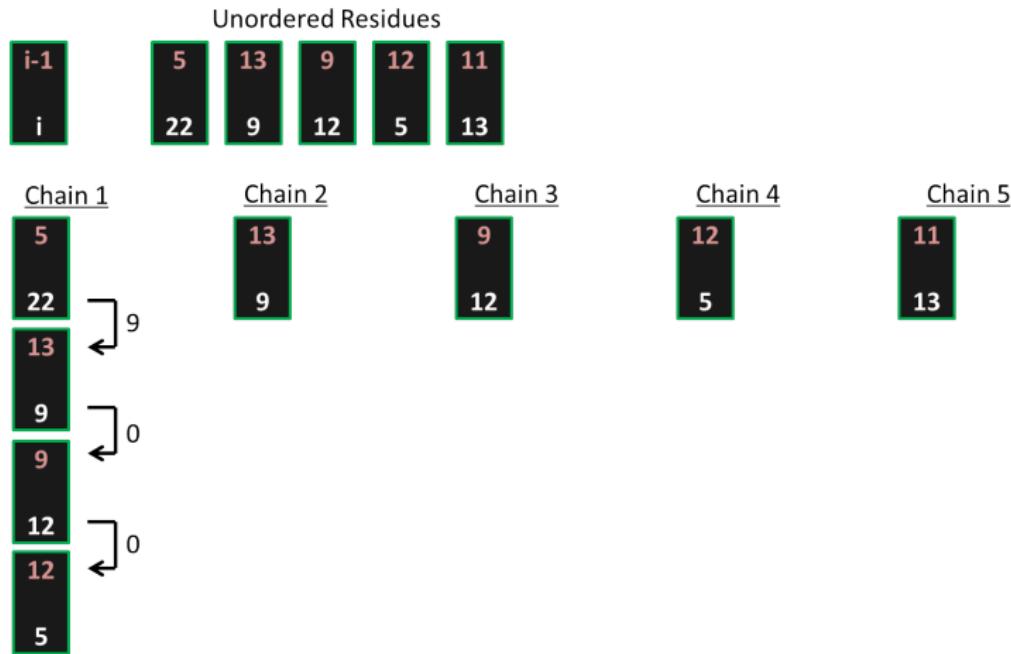
Assignment Algorithm Illustrated



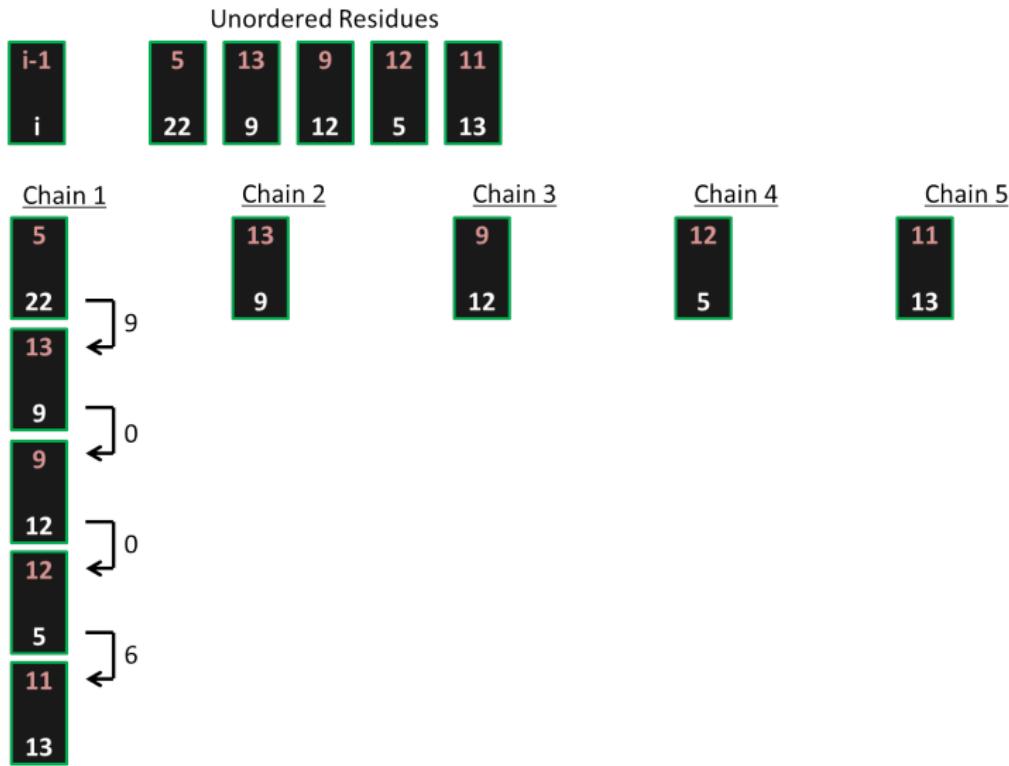
Assignment Algorithm Illustrated



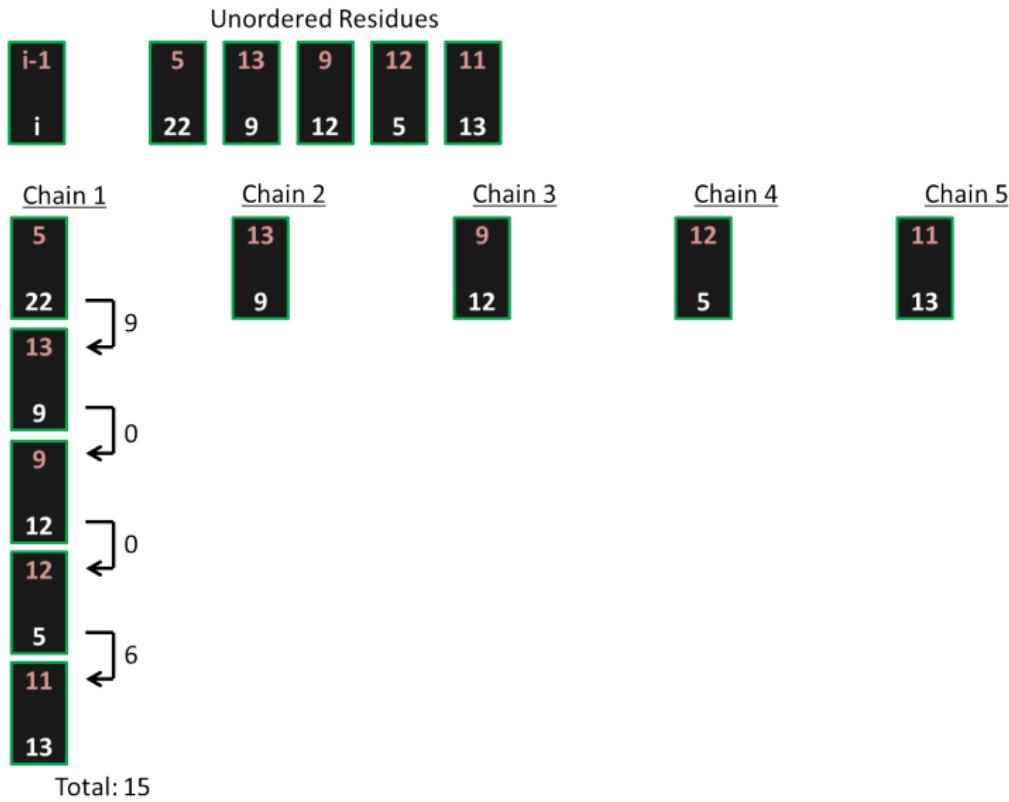
Assignment Algorithm Illustrated



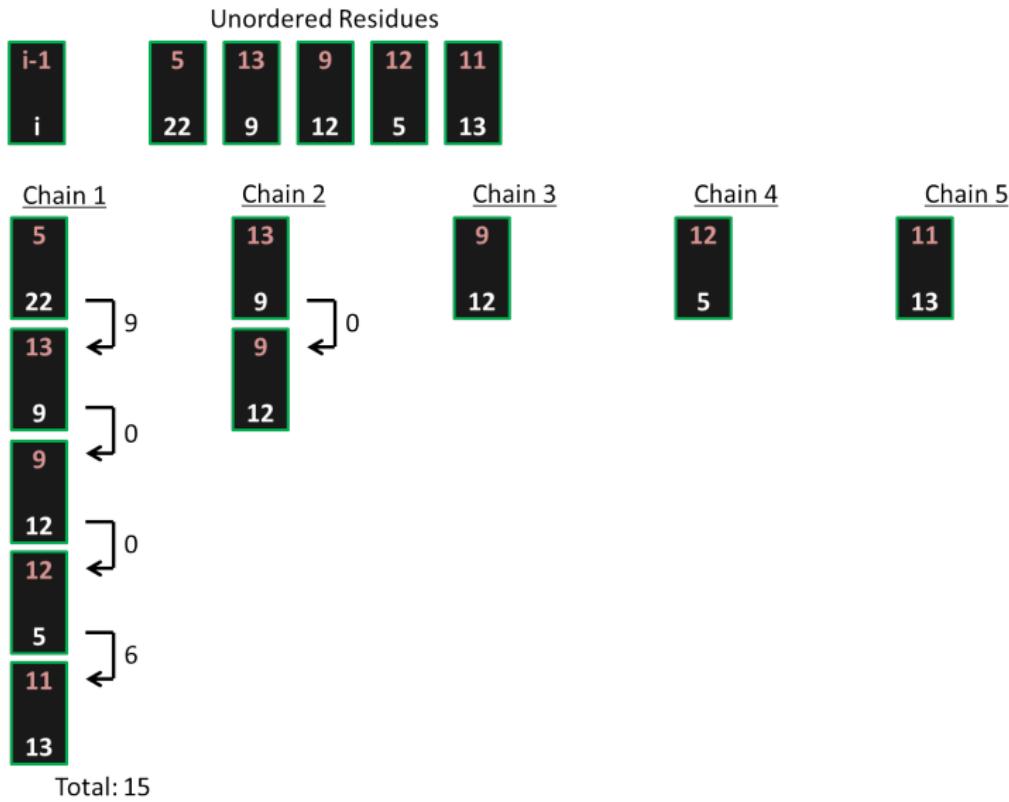
Assignment Algorithm Illustrated



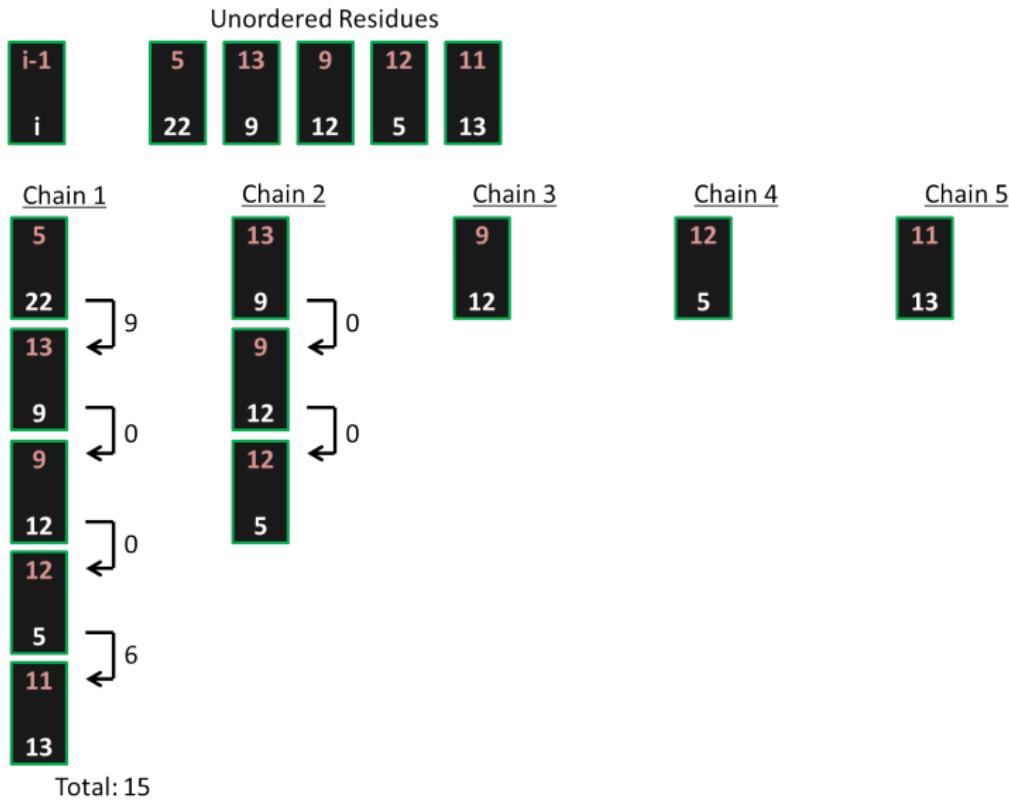
Assignment Algorithm Illustrated



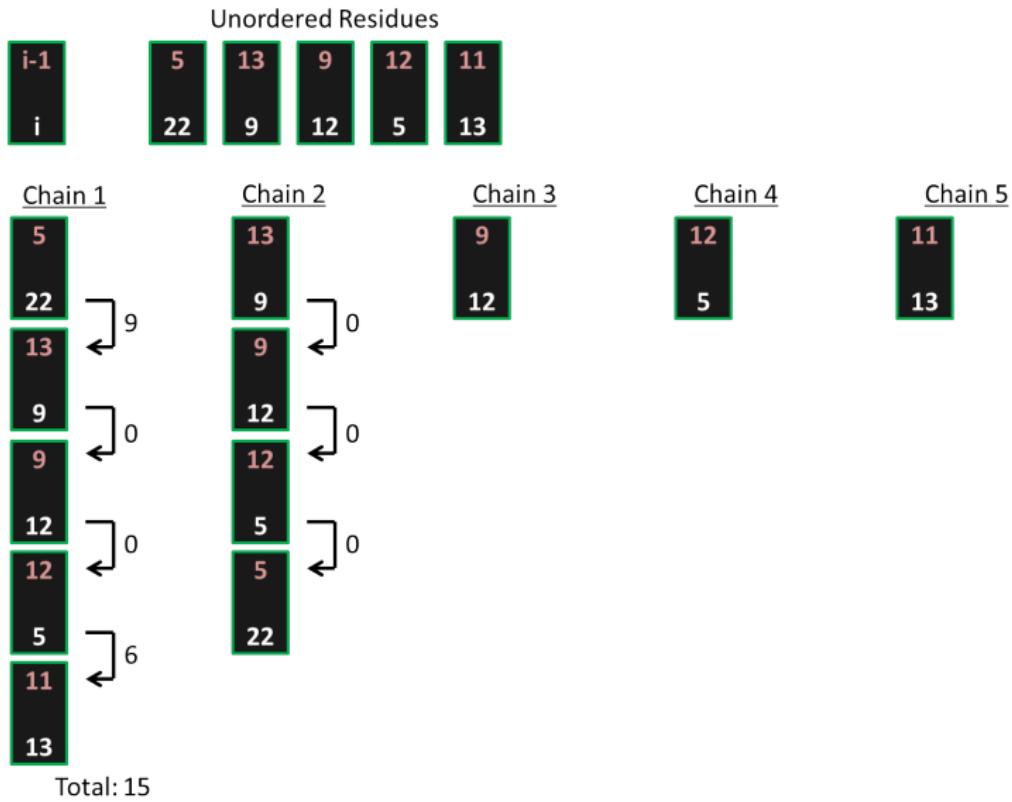
Assignment Algorithm Illustrated



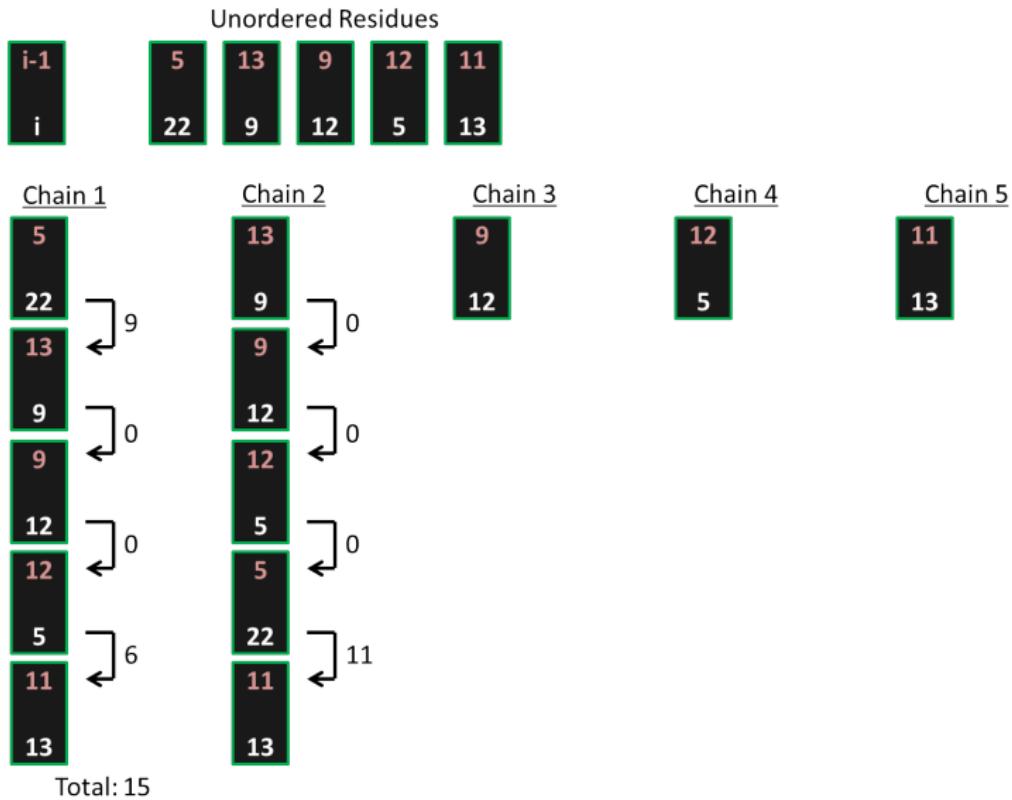
Assignment Algorithm Illustrated



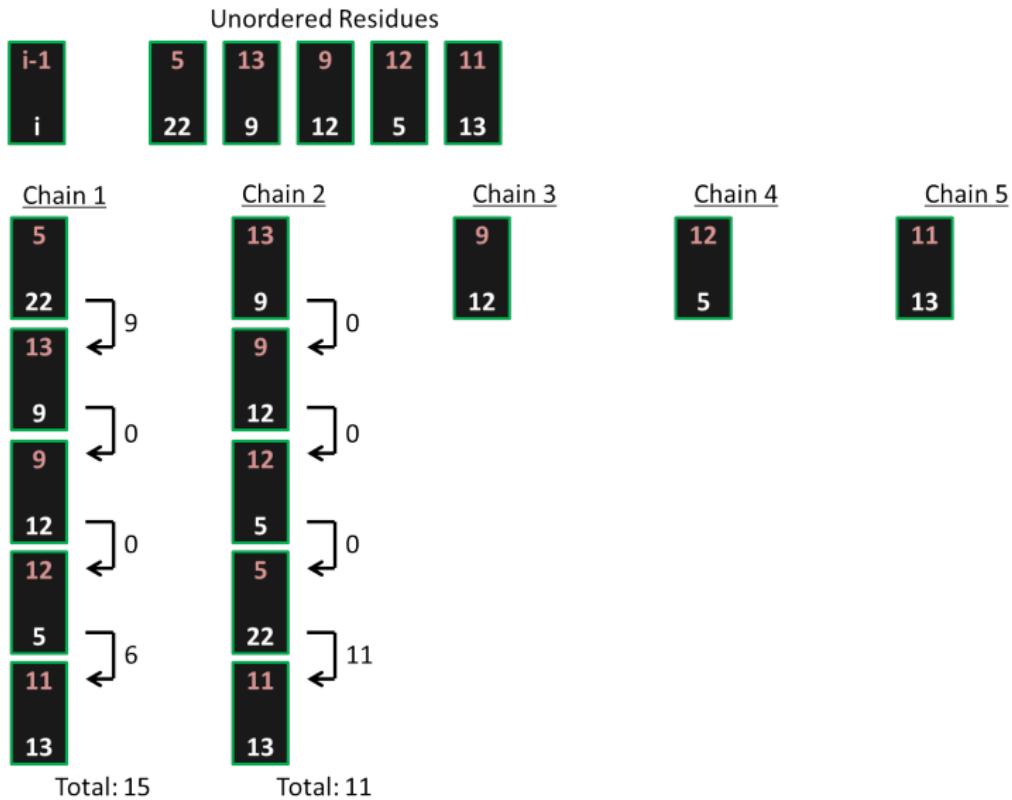
Assignment Algorithm Illustrated



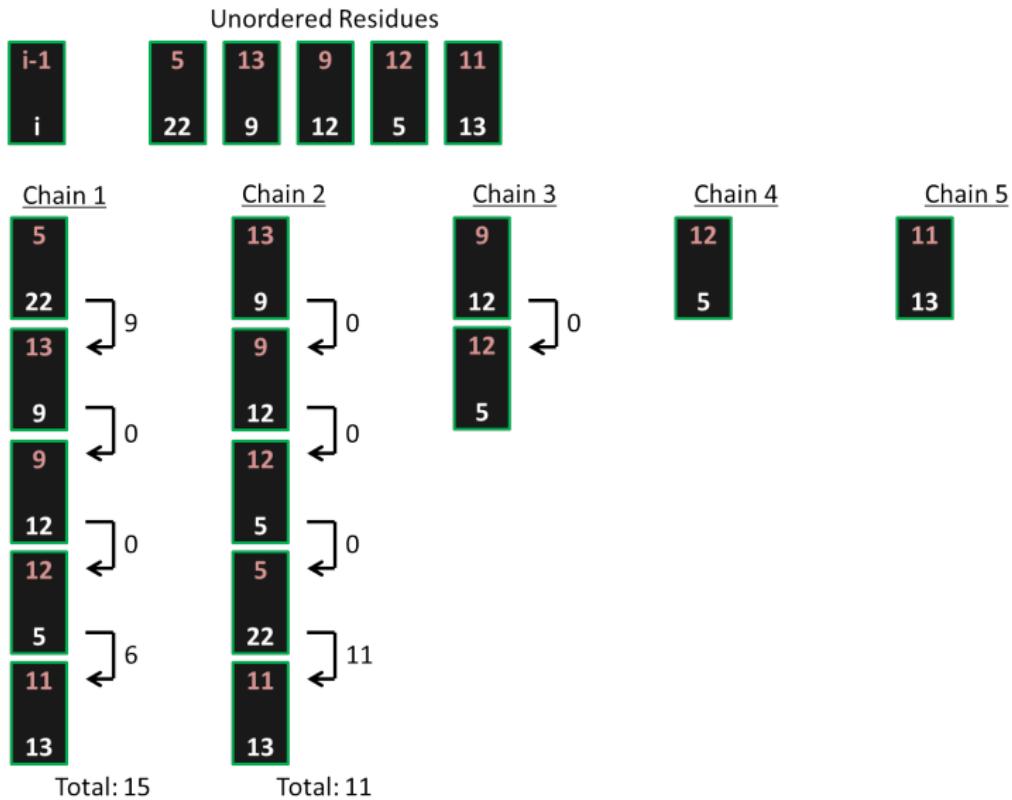
Assignment Algorithm Illustrated



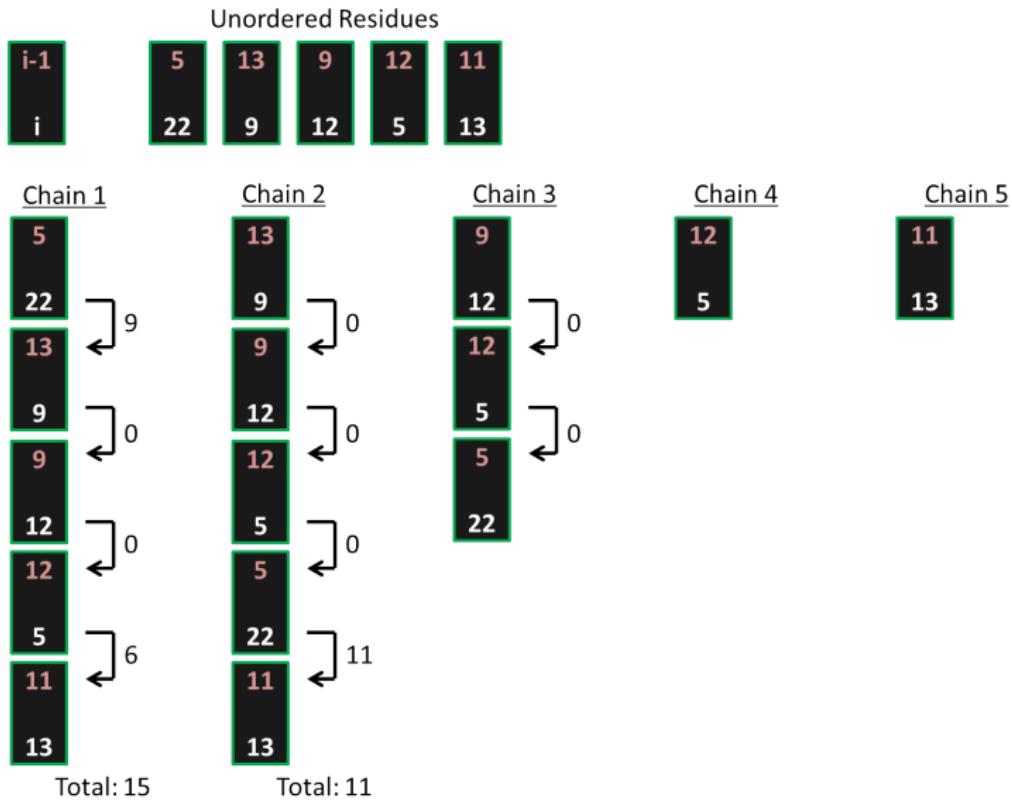
Assignment Algorithm Illustrated



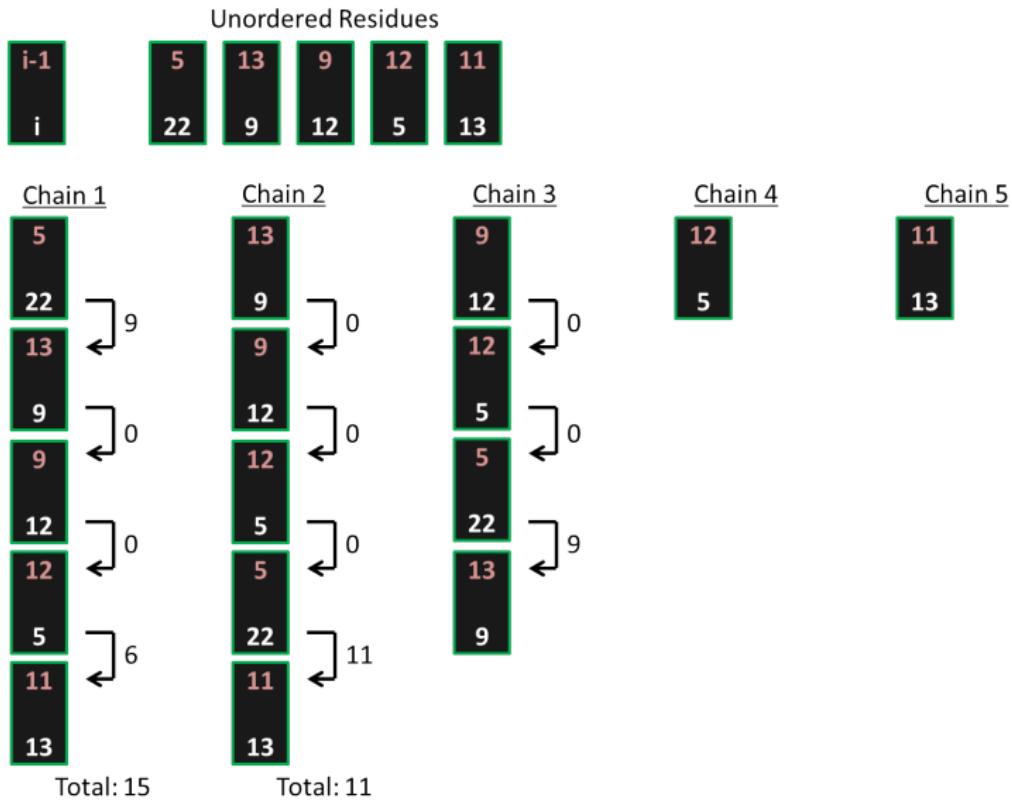
Assignment Algorithm Illustrated



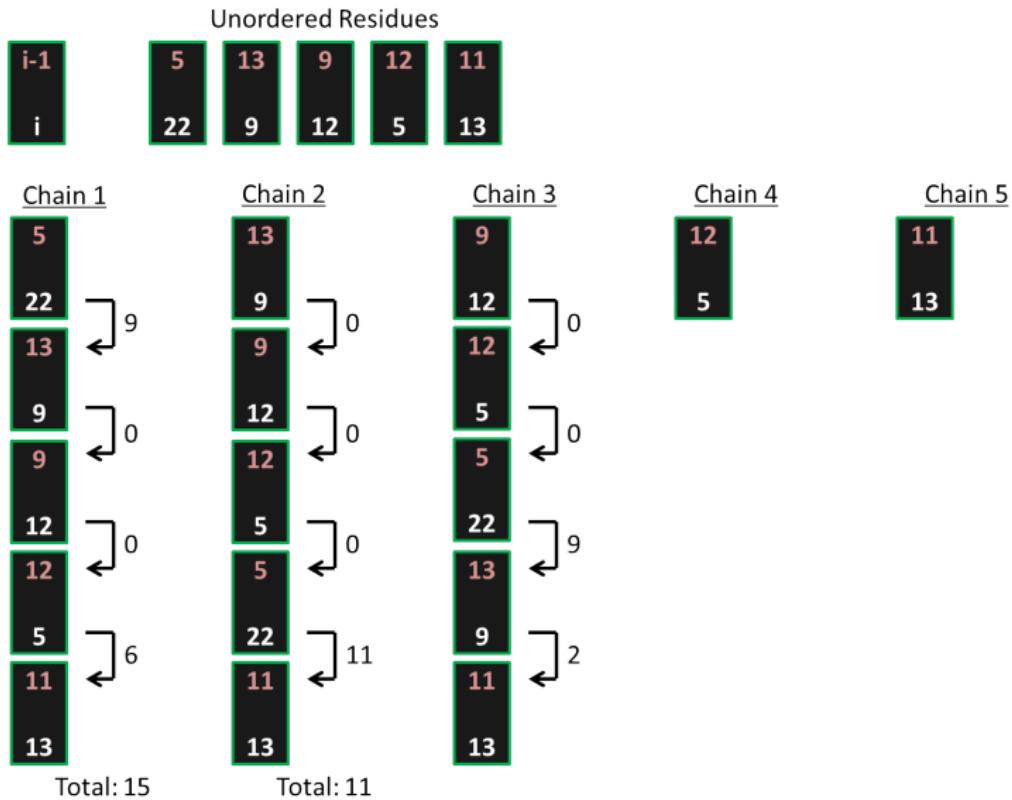
Assignment Algorithm Illustrated



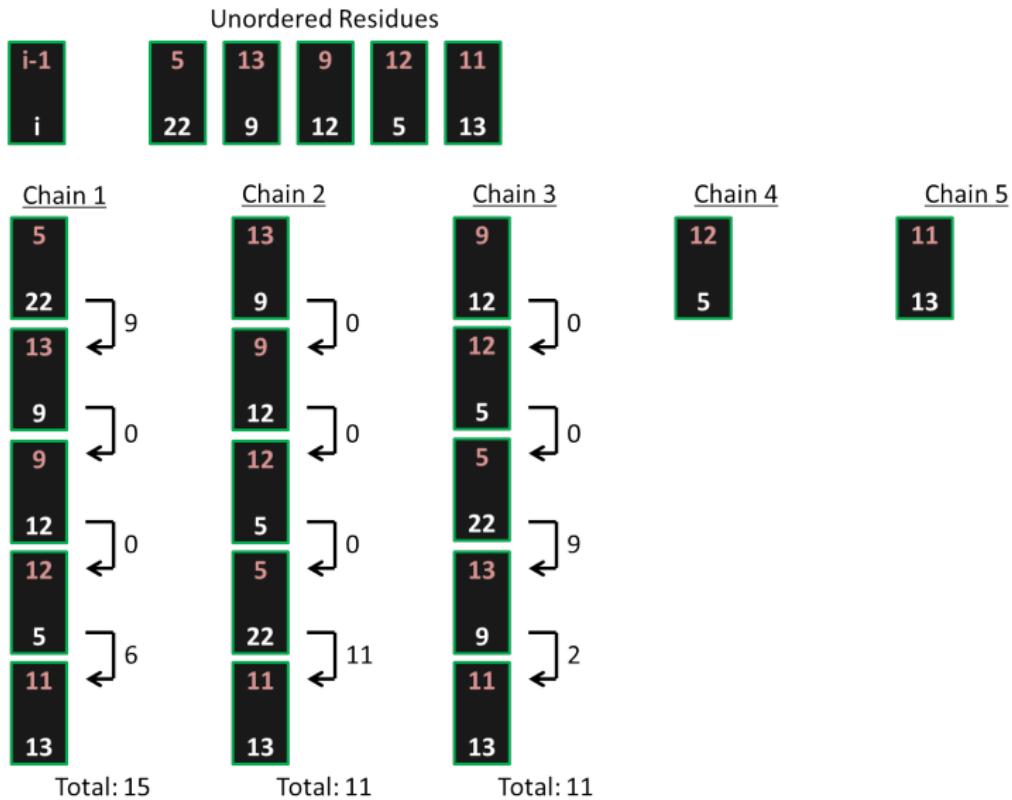
Assignment Algorithm Illustrated



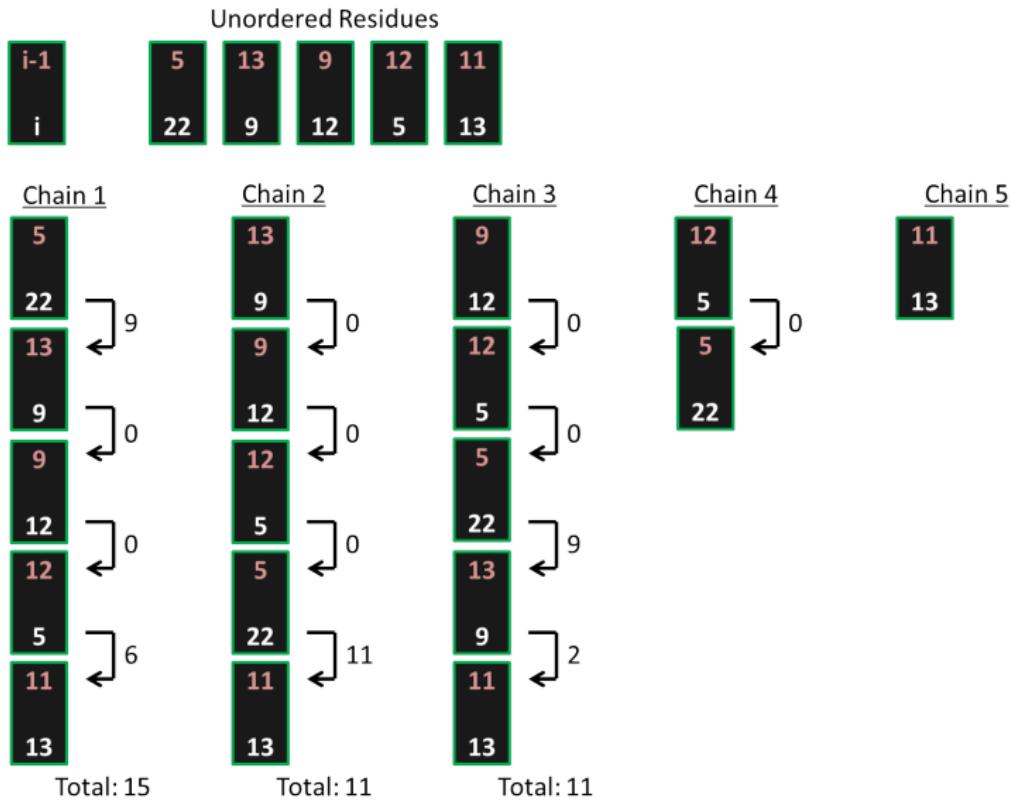
Assignment Algorithm Illustrated



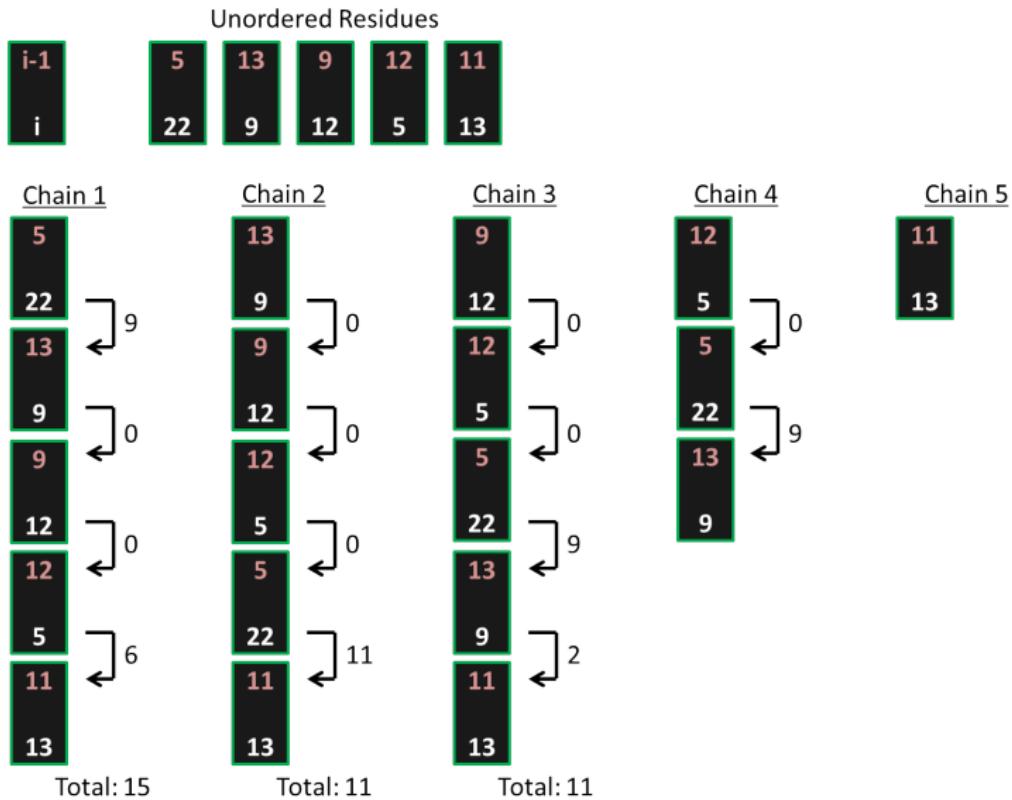
Assignment Algorithm Illustrated



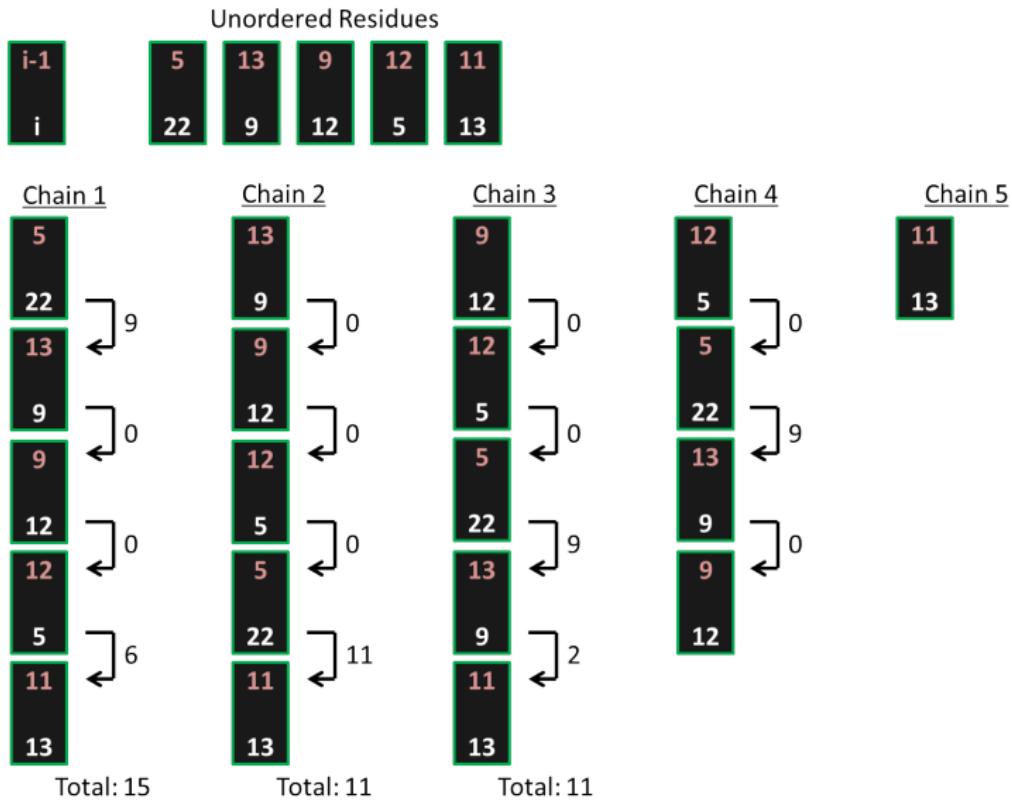
Assignment Algorithm Illustrated



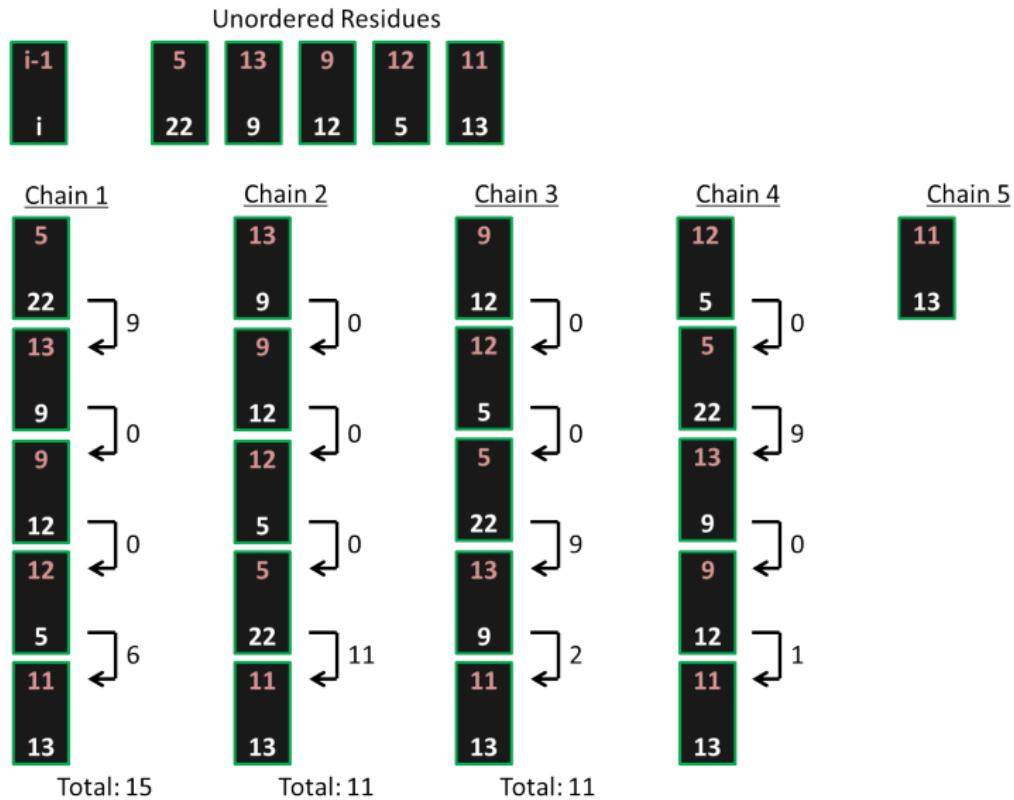
Assignment Algorithm Illustrated



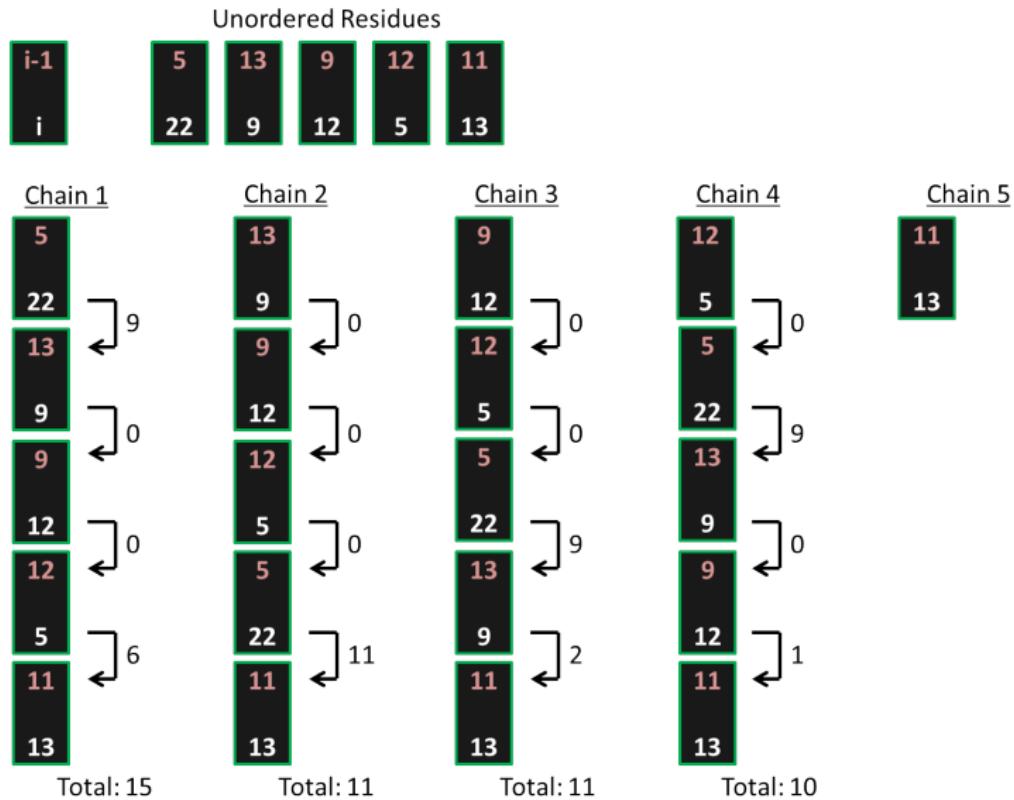
Assignment Algorithm Illustrated



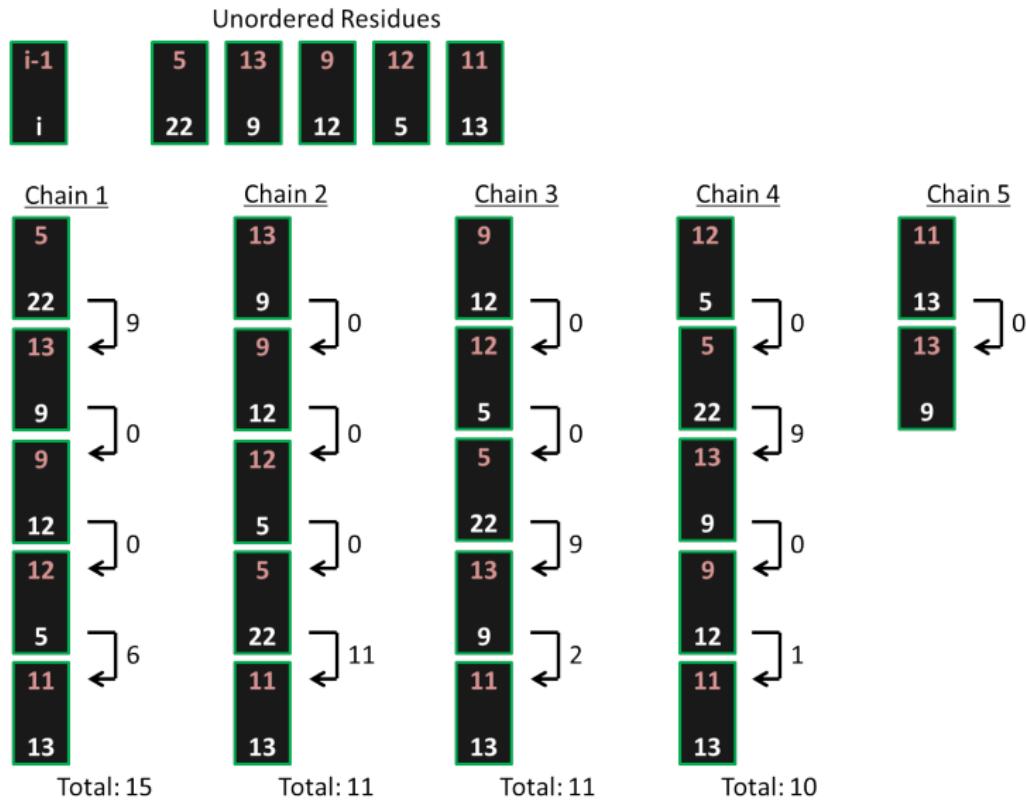
Assignment Algorithm Illustrated



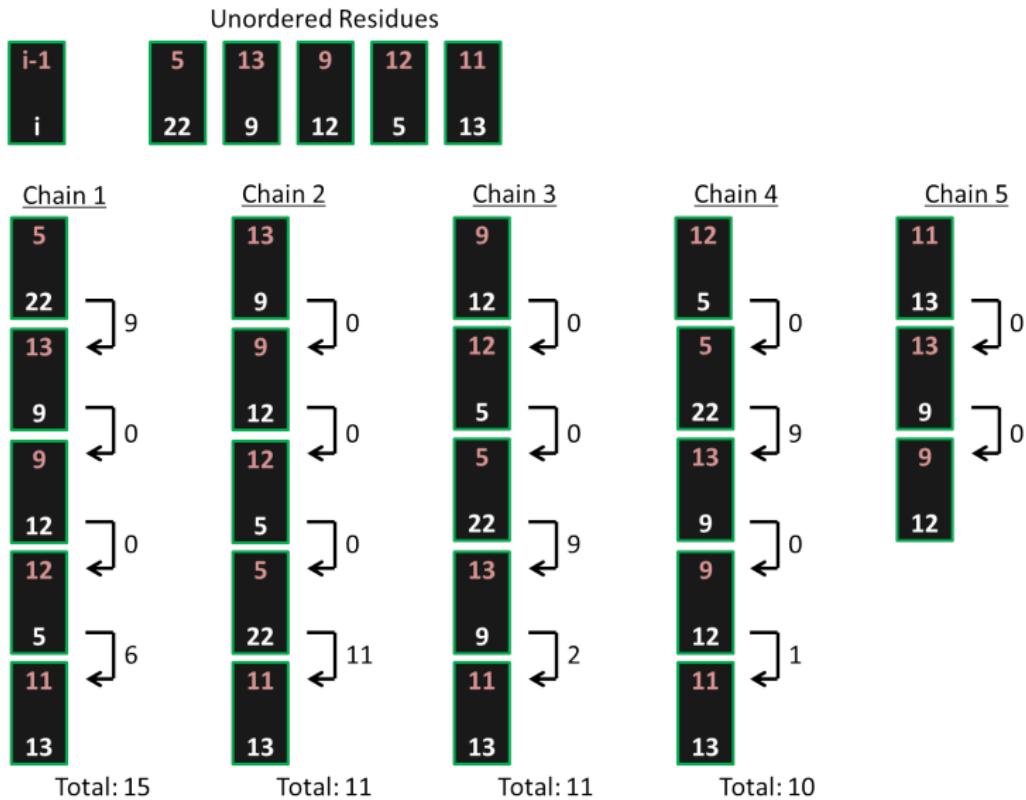
Assignment Algorithm Illustrated



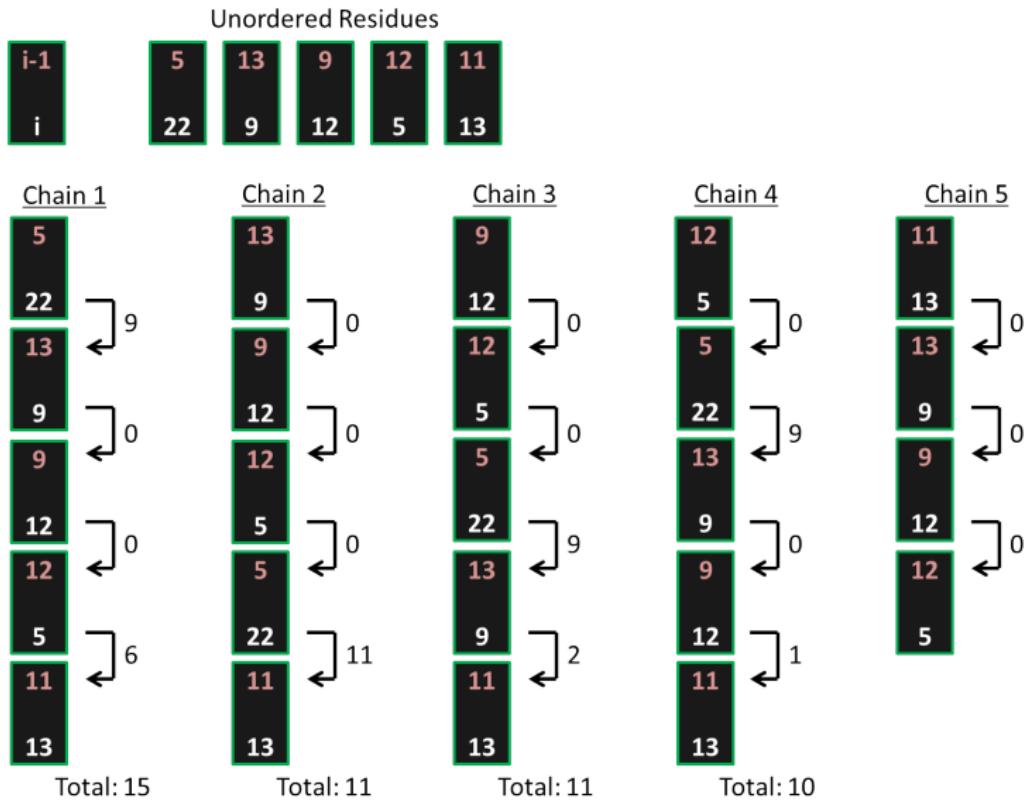
Assignment Algorithm Illustrated



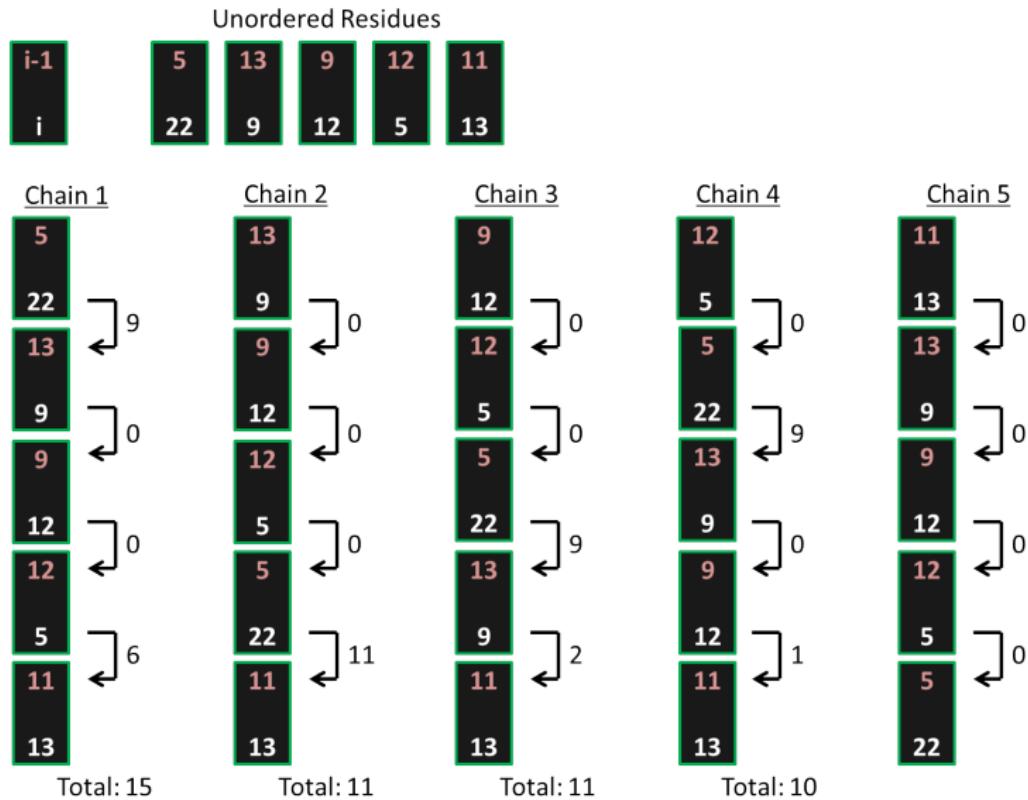
Assignment Algorithm Illustrated



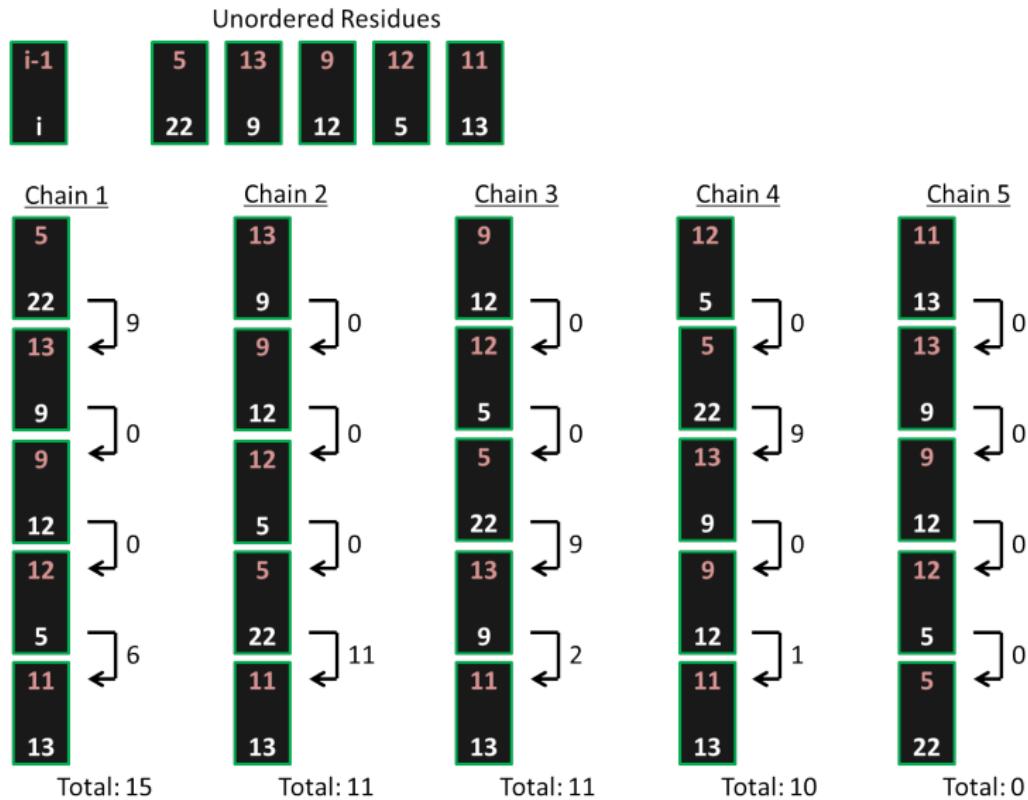
Assignment Algorithm Illustrated



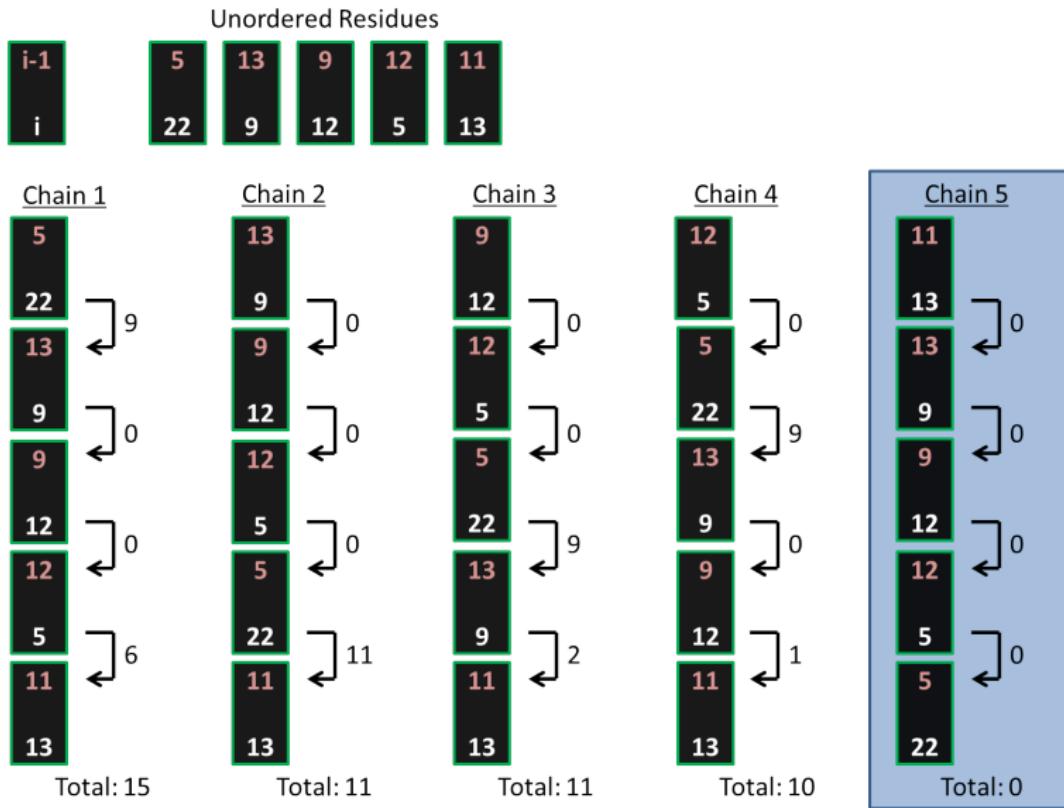
Assignment Algorithm Illustrated



Assignment Algorithm Illustrated



Assignment Algorithm Illustrated



Results and Conclusions



Results and Conclusions

- Successes



Results and Conclusions

- Successes
 - Assignment of small protein datasets was successful (<20 residues)



Results and Conclusions

- Successes

- Assignment of small protein datasets was successful (<20 residues)
- Longer datasets did not take exponentially longer to assign (polynomial time complexity)



Results and Conclusions

- Successes
 - Assignment of small protein datasets was successful (<20 residues)
 - Longer datasets did not take exponentially longer to assign (polynomial time complexity)
- Failures



Results and Conclusions

- Successes
 - Assignment of small protein datasets was successful (<20 residues)
 - Longer datasets did not take exponentially longer to assign (polynomial time complexity)
- Failures
 - Introduction of ambiguous data led to less than satisfactory results



Results and Conclusions

- Successes
 - Assignment of small protein datasets was successful (<20 residues)
 - Longer datasets did not take exponentially longer to assign (polynomial time complexity)
- Failures
 - Introduction of ambiguous data led to less than satisfactory results
 - Datasets less than 20 residues are trivial, even for manual assignment



Results and Conclusions

- Successes

- Assignment of small protein datasets was successful (<20 residues)
- Longer datasets did not take exponentially longer to assign (polynomial time complexity)

- Failures

- Introduction of ambiguous data led to less than satisfactory results
- Datasets less than 20 residues are trivial, even for manual assignment
- Data had to pre-processed before the algorithm could be used



Results and Future Work



Results and Future Work

- Future work



Results and Future Work

- Future work
 - Use an optimal search strategy (e.g A*)



Results and Future Work

- Future work
 - Use an optimal search strategy (e.g A*)
 - Greedy search is not optimal



Results and Future Work

- Future work
 - Use an optimal search strategy (e.g A*)
 - Greedy search is not optimal
 - Optimal strategies will always generate best solution



Results and Future Work

- Future work
 - Use an optimal search strategy (e.g A*)
 - Greedy search is not optimal
 - Optimal strategies will always generate best solution
 - Improve heuristic so larger datasets can be assigned with greater a success rate



Results and Future Work

- Future work
 - Use an optimal search strategy (e.g A*)
 - Greedy search is not optimal
 - Optimal strategies will always generate best solution
 - Improve heuristic so larger datasets can be assigned with greater a success rate
 - Use multiple NMR experiments at the same time to improve accuracy with noisy or ambiguous data



Acknowledgements

- Drake Undergraduate Collaborative Science Initiative (DUSCI) summer research fellowship
- Dr. Adina Kilpatrick—Drake University (Physics)
- Dr. Tim Urness—Drake University (Mathematics and Computer Science)
- Steven Johnson—Drake University (research colleague)



