Geometric Algorithms for De Novo Modeling

Sasakthi Abeysinghe[1] · Stephen Schuh[1]
Austin Abrams[1] · Tao Ju[1]
Matthew Baker[2] · Wah Chiu[2]

1. Washington University in St. Louis
2. Baylor College of Medicine

www.cs.wustl.edu/~taoju/research/modeling2010_tao.ppt

Geometric Problems

- Shape analysis of density maps
  - Tubular vs. plate-like regions
- Topology analysis of SSEs
  - Based on cryo-EM and primary sequence
- Flexible model fitting
  - Guided by SSEs

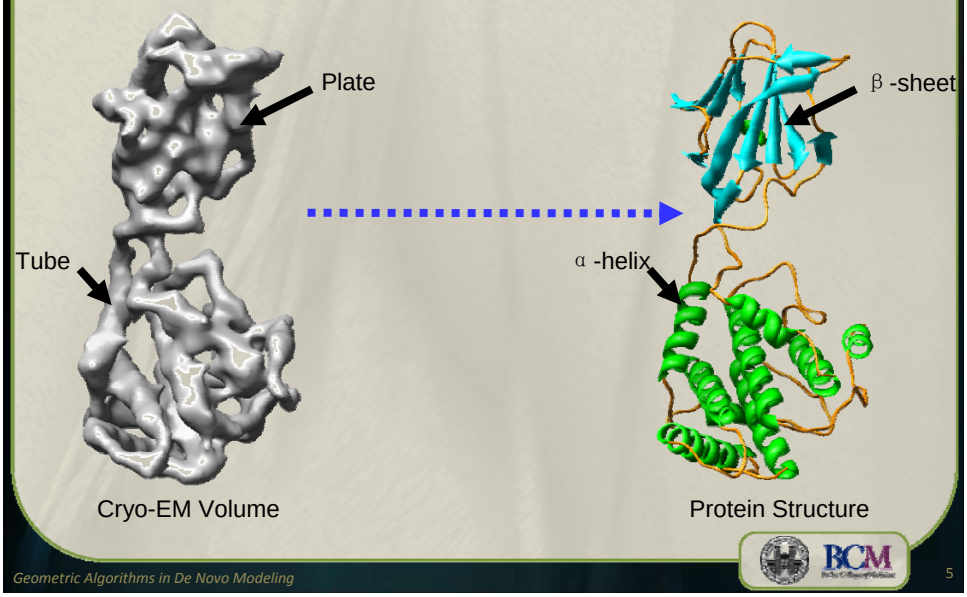
Geometric Algorithms

- Shape analysis of density maps
  - Skeletonization
- Topology analysis of SSEs
  - Graph matching
- Flexible model fitting
  - Shape registration

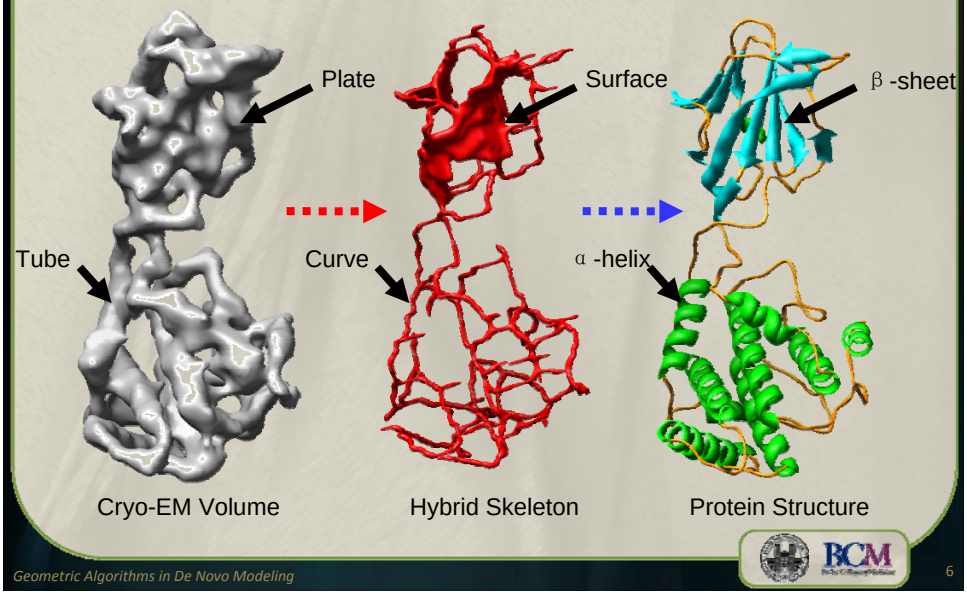
Geometric Algorithms

- Shape analysis of density maps
  - Skeletonization
- Topology analysis of SSEs
  - Graph matching
- Flexible model fitting
  - Shape registration

Skeleton-based Shape Analysis

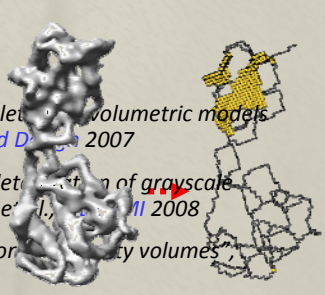


Skeleton-based Shape Analysis



Computing Hybrid Skeletons

- Hybrid Skeletons
  - Made up of curves and surfaces
  - Located at tubular and plate-like parts of the shape
  - Robust to noise

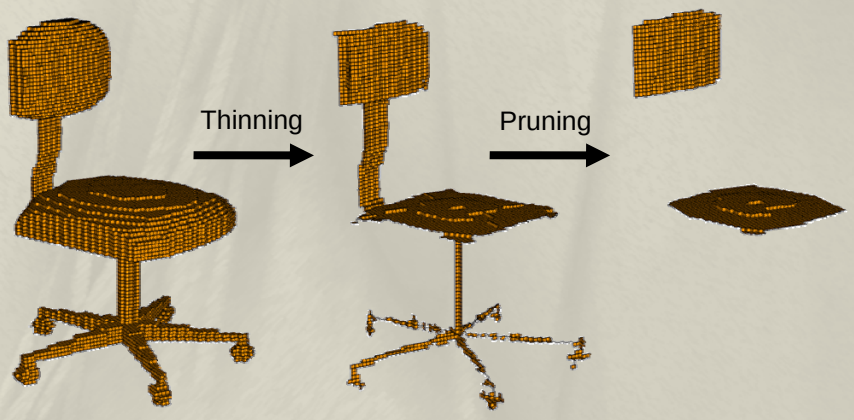


- Three Algorithms
  - For binary volumes: *"Computing a family of skeleton volumetric models for shape description", Ju et al., Computer-Aided D... 2007*
  - For grayscale volumes: *"Segmentation-free skeleton... of grayscale volumes for shape understanding", Abeysinghe e... ll 2008*
  - Interactive sketching: *"Interactive skeletonization... ty volumes", Abeysinghe et al., The visual computer 2009*

Algorithm 1: skeletons from Binary Volumes

- Alternated thinning and pruning

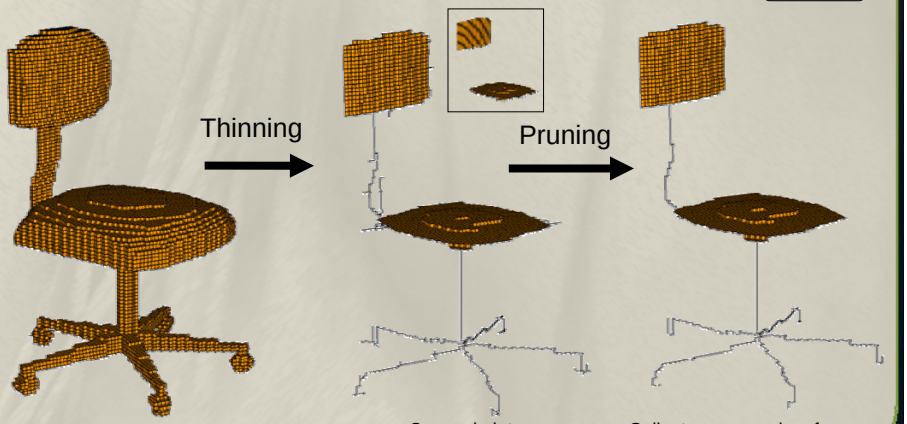
Stage 1



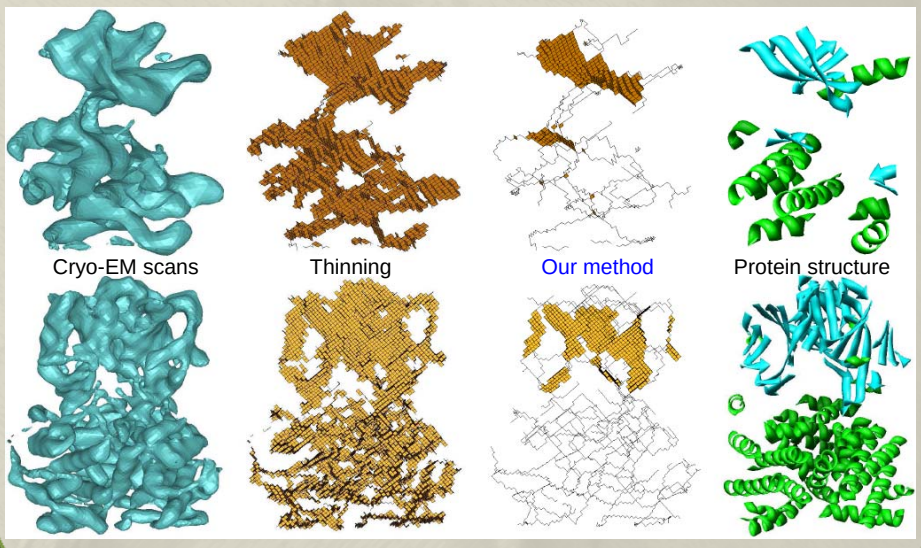
Algorithm 1: skeletons from Binary Volumes

- Alternated thinning and pruning

Stage 2

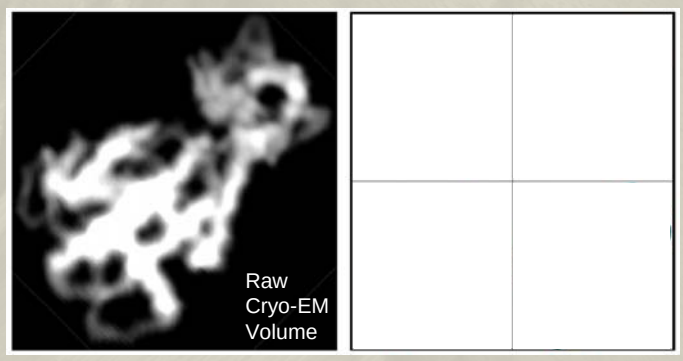


Results



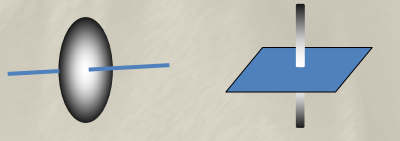
Algorithm 2: Skeletons from Grayscale Volumes

- Binary algorithms that work on thresholded volumes are sensitive to the choice of thresholds



Algorithm 2: Skeletons from Grayscale Volumes

- Observations
  - Skeletons lie close to local maxima of grayscale intensity
  - Intensity around the skeleton varies most in directions orthogonal to the skeleton
    - Curve: intensity varies most on a plane
    - Surface: intensity varies most in a line

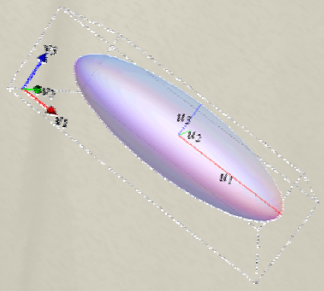


Algorithm 2: Skeletons from Grayscale Volumes

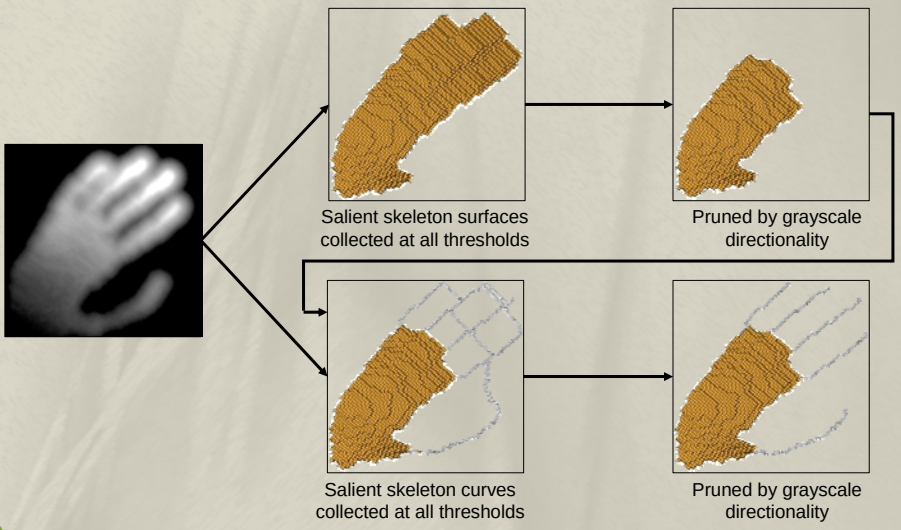
- Combine skeletons at various thresholds and prune using grayscale directionality
  - Directionality analysis: structure tensor
    - Eigenvectors and eigenvalues give principle directions and amount of intensity variation

$$T_p = \sum \theta_p T'_p$$

$$T'_p = \begin{bmatrix} I_x \\ I_y \\ I_z \end{bmatrix} \times \begin{bmatrix} I_x \\ I_y \\ I_z \end{bmatrix}^T = \begin{bmatrix} I_x^2 & I_x I_y & I_x I_z \\ I_x I_y & I_y^2 & I_y I_z \\ I_x I_z & I_y I_z & I_z^2 \end{bmatrix}$$

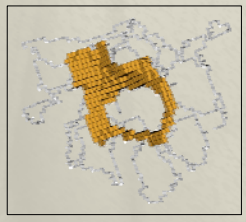


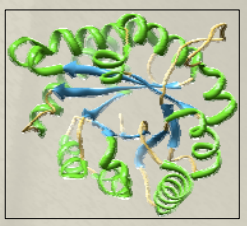
Algorithm 2: Skeletons from Grayscale Volumes



Results


CryoEM scans

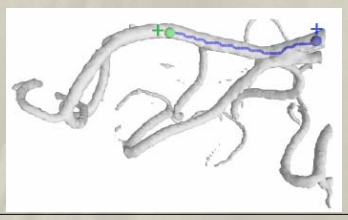

Hybrid skeletons

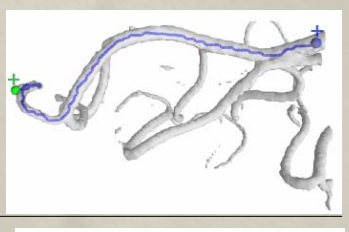

Protein structure

Algorithm 3: Interactive Skeleton Sketching

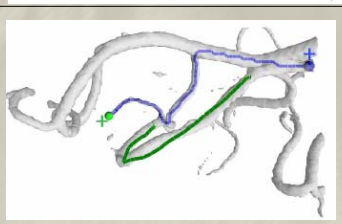
- Click-and-drag interaction in 3D
  - Suited for noisy and ambiguous density data

End-point clicking:





Scribbling:

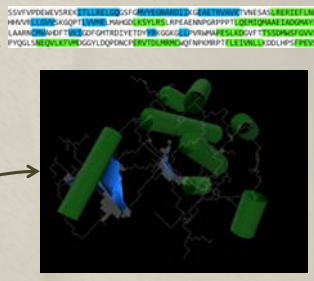


Geometric Algorithms

- Shape analysis of density maps
  - Skeletonization
- Topology analysis of SSEs
  - Graph matching
- Flexible model fitting
  - Shape registration

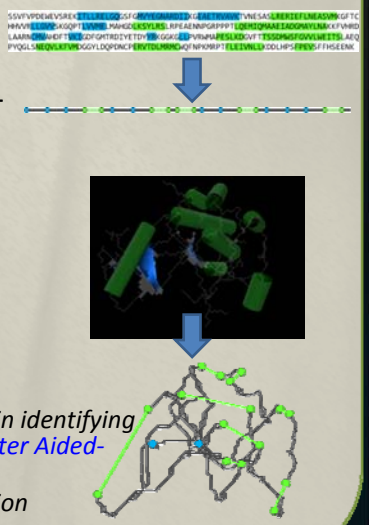
Objective

- Input:
  - Sequence with predicted helices and strands
  - Skeleton with observed helices and sheets
- Problem:
  - Find a connected path of helices and sheets that best matches the sequence



Method

- Graph matching
  - Represent sequence and cryo-EM as graphs
    - Encoding connectivity of SSEs
  - Compute the best-matching sub-graph
    - Match scored by SSE similarity



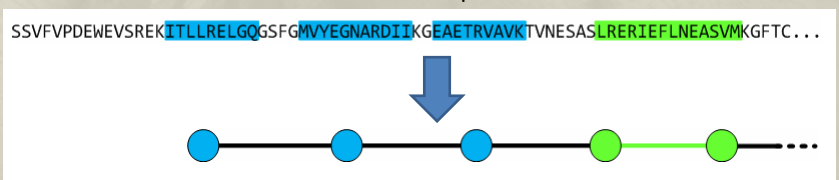
•Helix matching: *"Shape modeling and matching in identifying 3D protein structures.", Abeysinghe et al., Computer Aided-Design, 2008*

•Helix + Sheet matching: *Schuh et al., in preparation (functionality already in Gorgon V2.0)*

Sequence as Graph

- Nodes: 2 per helix (green), 1 per strand (blue)
- Edges: Connect adjacent nodes

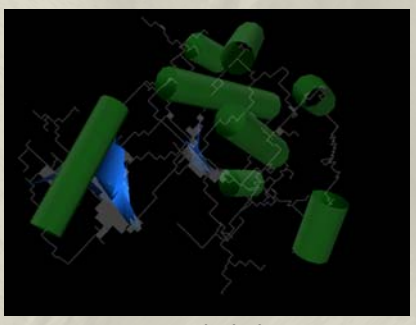
Annotated Sequence



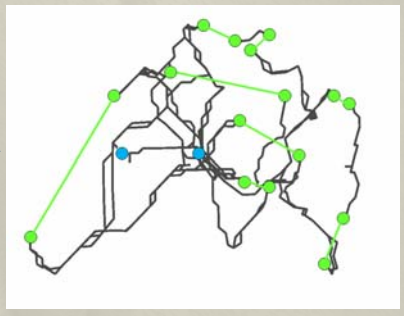
Graph

Skeleton as Graph

- Nodes: 2 per helix (green), 1 per sheet (blue)
- Edges: Paths along skeleton connecting nodes

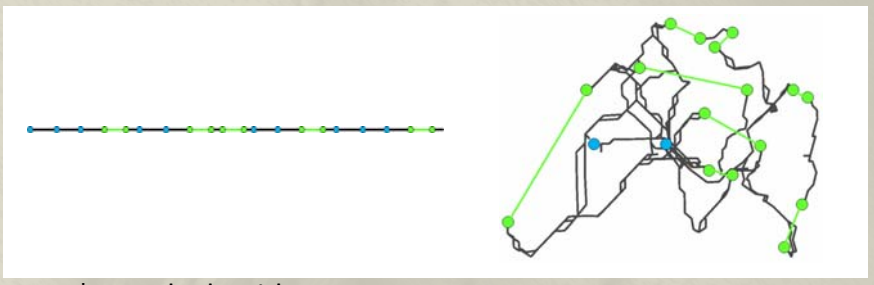


Annotated Skeleton



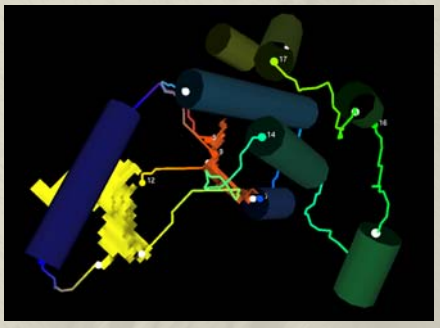
Graph

Graph Matching

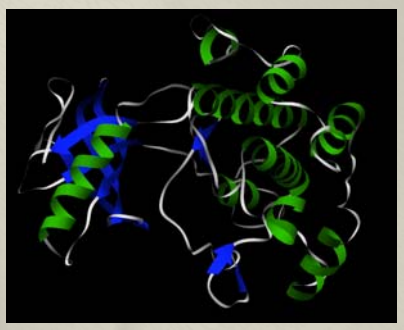


- A* search algorithm
- Sheet nodes (blue) in skeleton can be visited more than once
- Modify graph structure to handle noise
  - Missing helices and sheets
  - Missing loops in skeleton
  - Broken sheets
- Now available in Gorgon: http://cs.wustl.edu/~ssa1/gorgon

Results: 1IRK (simulated data)



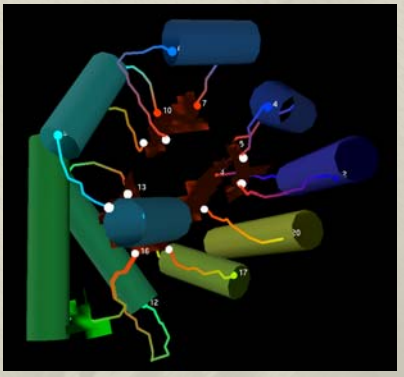
Lowest-cost match



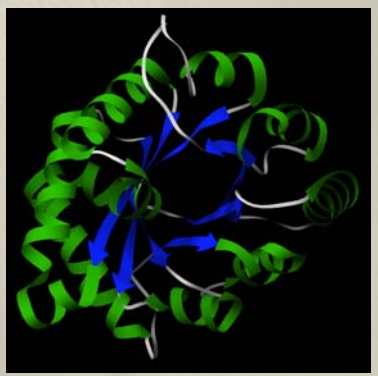
X-ray Structure

- Executed in 0.45 seconds
- 100% correct helix correspondence
- 60% correct strand correspondence (one missing sheet)

Results: 1TIM (simulated data)



Lowest-cost match, after user input



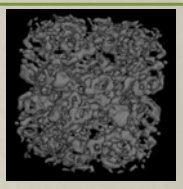
X-ray Structure

- Executed in 2.8 seconds
- 1 user-specified constraint
- 100% correct helix and strand correspondence

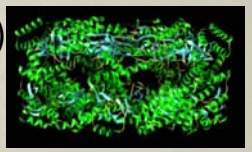
Geometric Algorithms

- Shape analysis of density maps
  - Skeletonization
- Topology analysis of SSEs
  - Graph matching
- Flexible model fitting
  - Shape registration

Objective



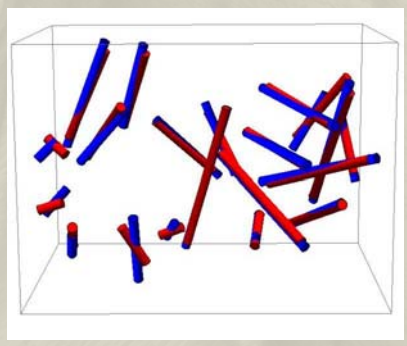
- Input:
  - Cryo EM density of a molecule
  - High-resolution structure of a similar molecule (can be at a different conformation)



- Problem:
  - Fit the high-resolution structure into the density while allowing for non-rigid deformations

Observation #1

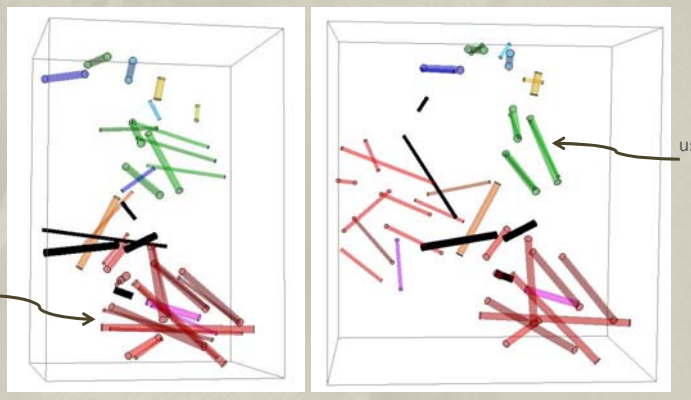
- In different conformations or in similar molecules, corresponding helices are often similar in shape



1OEL –A (red) and 1SS8 –A (blue)

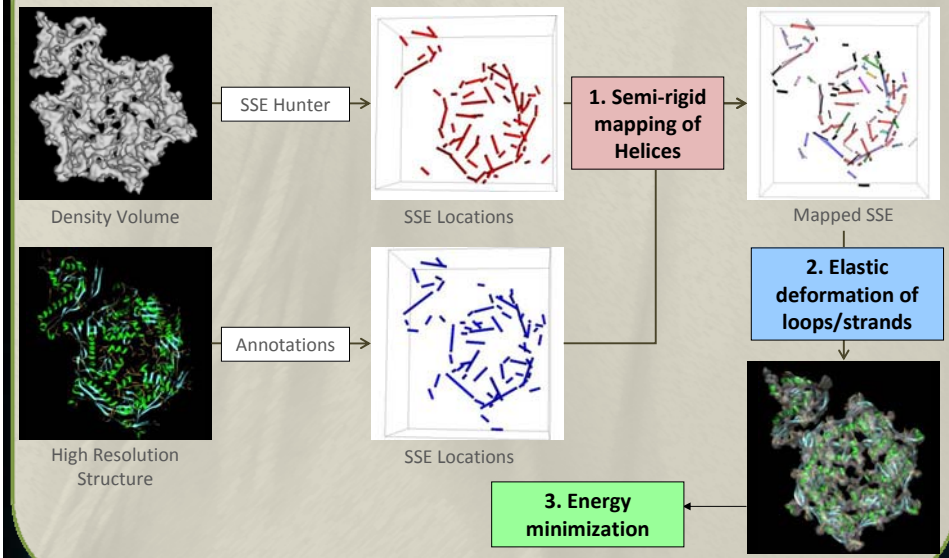
Observation #2

- Groups of helices often deform rigidly (but deformation may differ among groups)



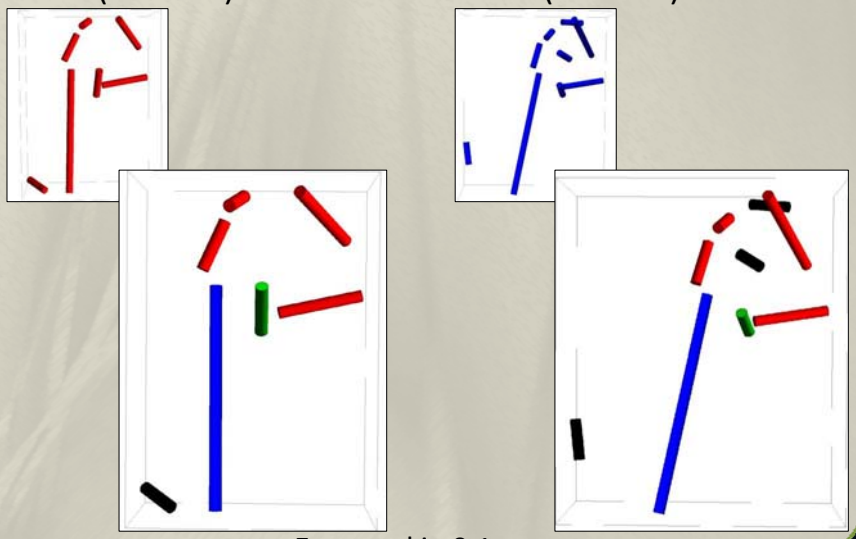
1OEL –A (thin cylinders) and 2C7C –A (thick cylinders)

Method: Coarse-to-fine, Helix-driven Fitting



Preliminary Results: Helix Mapping

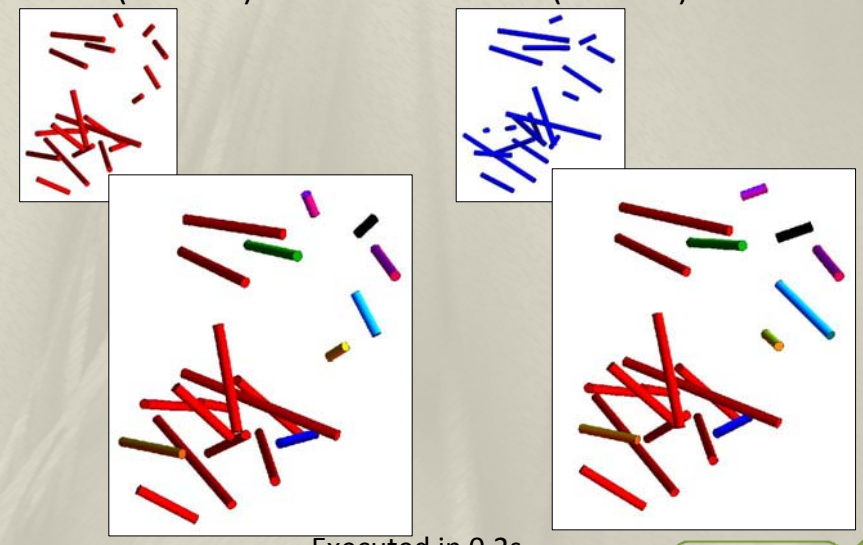
- 3E8K (Chain A) - 7 Helices
- 2GP1 (Chain A) - 9 Helices



Executed in 0.1s

Preliminary Results: Helix Mapping

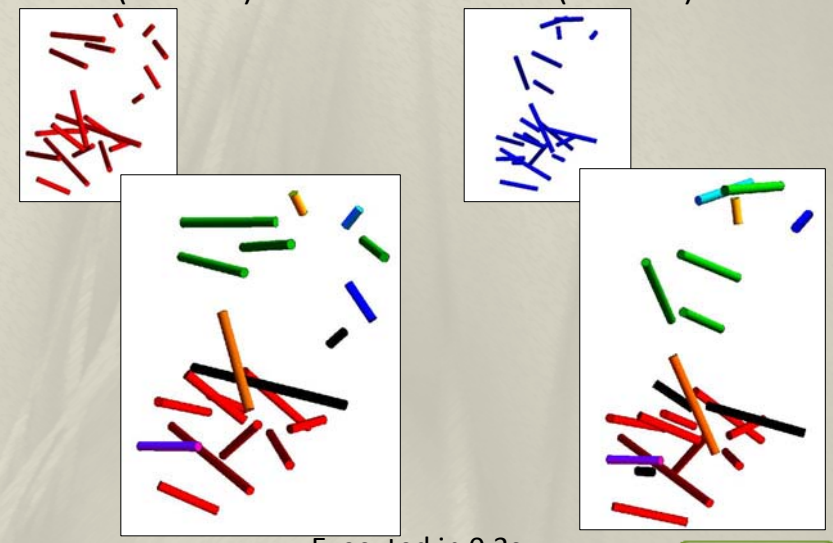
- 1OEL (Chain A) - 20 Helices
- 1SS8 (Chain A) - 20 Helices



Executed in 0.2s

Preliminary Results: Helix Mapping

- 1OEL (Chain A) - 20 Helices
- 2C7C (Chain A) - 21 Helices

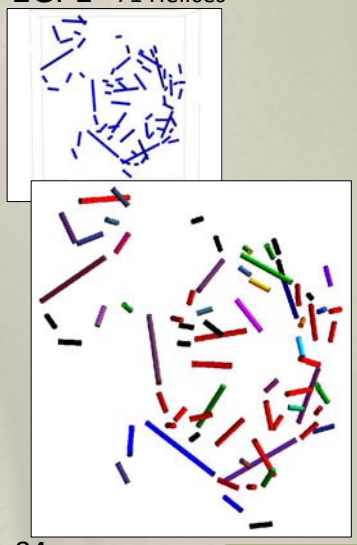


Executed in 0.3s

Preliminary Results: Helix Mapping

- 3E8K - 71 Helices

- 2GP1 - 71 Helices

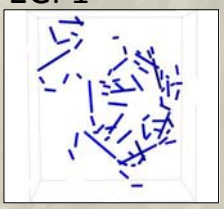


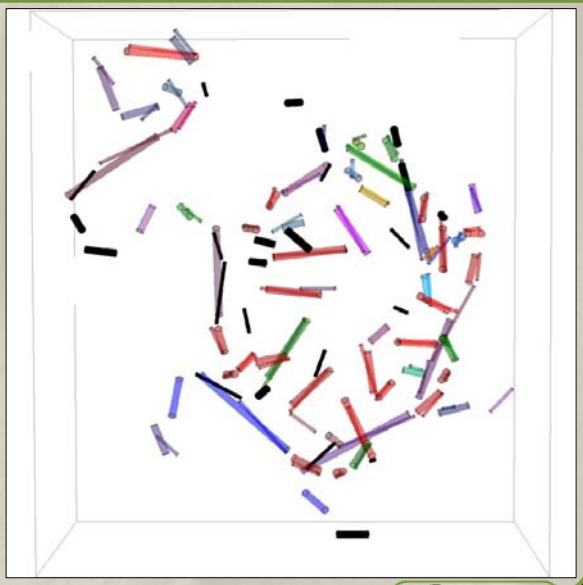
Executed in 84s

Preliminary Results: Helix Mapping

- 3E8K

- 2GP1





Summary

- Shape analysis of density maps
  - Skeletonization
- Topology analysis of SSEs
  - Graph matching
- Flexible model fitting
  - Shape registration