

Geometric Algorithms for De Novo Modeling

Sasakthi Abeysinghe¹ · Stephen Schuh¹
Austin Abrams¹ · Tao Ju¹
Matthew Baker² · Wah Chiu²

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

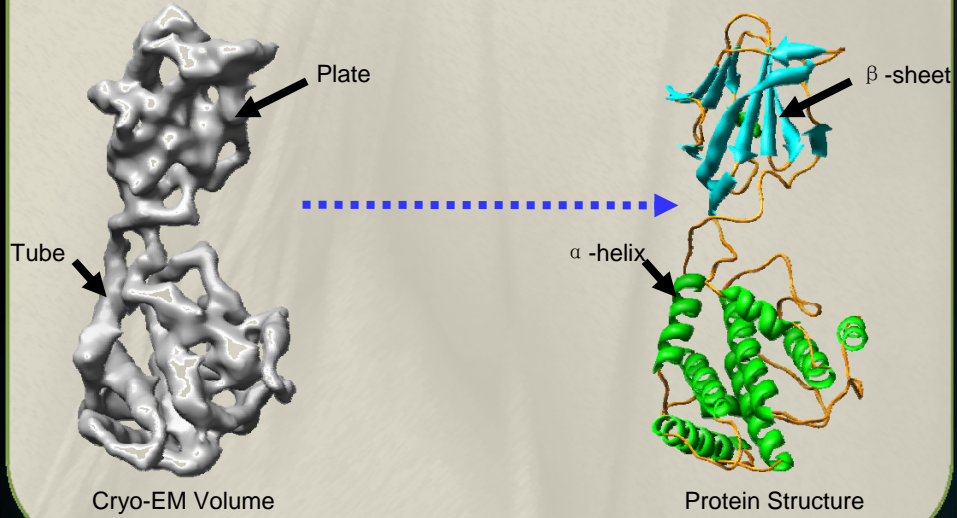
Geometric Algorithms

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

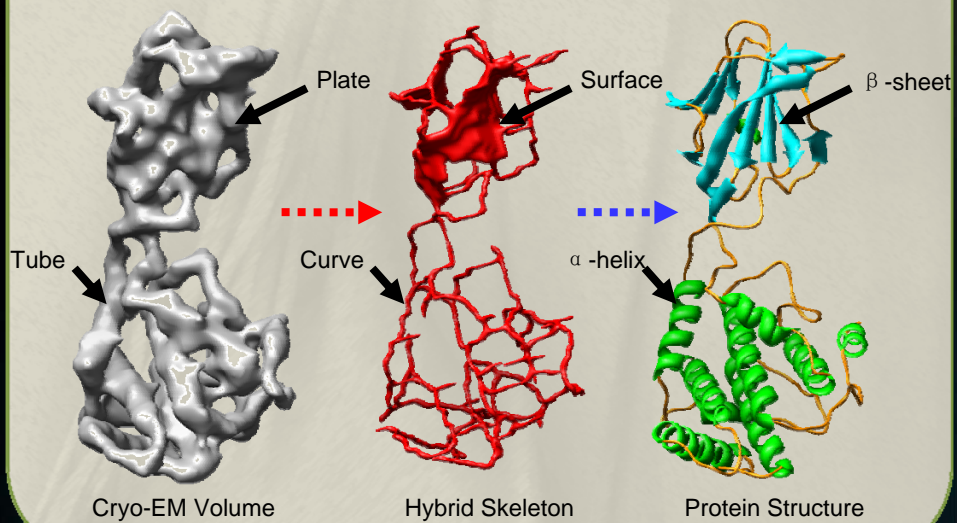
Geometric Algorithms

- 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

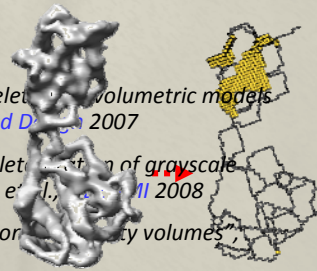


Computing Hybrid Skeletons

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

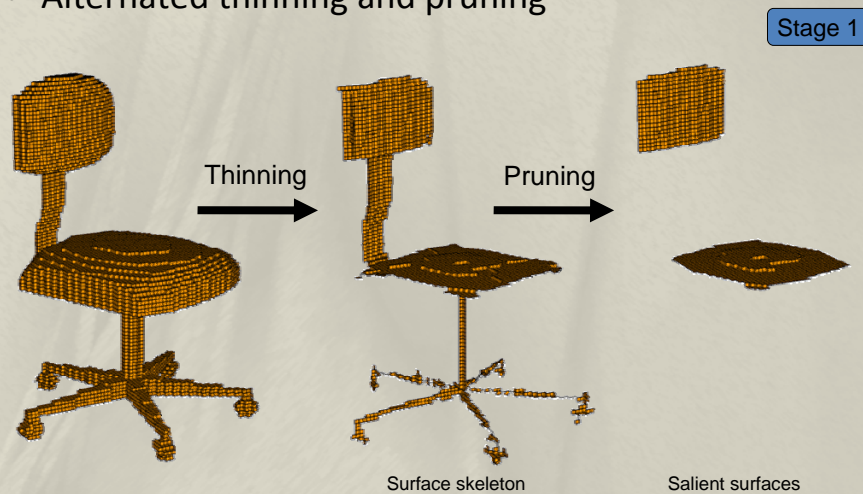
- Three Algorithms

- For binary volumes: "Computing a family of skeletons from volumetric models for shape description", Ju et al., *Computer-Aided Design* 2007
- For grayscale volumes: "Segmentation-free skeletonization of grayscale volumes for shape understanding", Abeyasinghe et al., *Computer-Aided Design* 2008
- Interactive sketching: "Interactive skeletonization of binary volumes", Abeyasinghe et al., *The visual computer* 2009



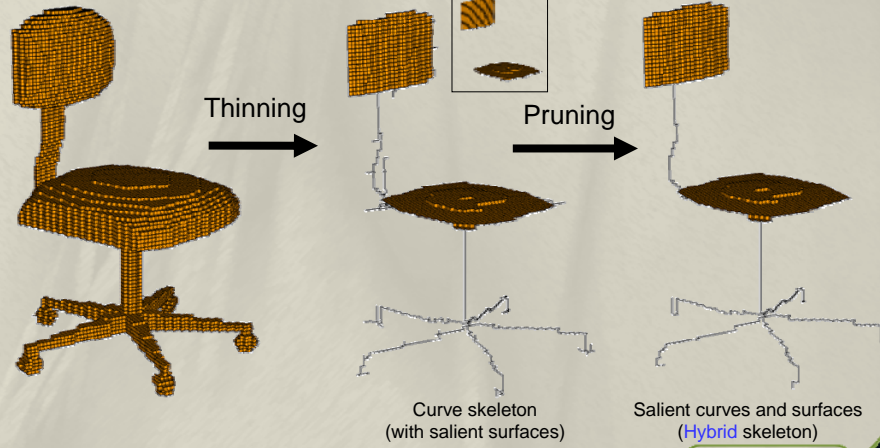
Algorithm 1: skeletons from Binary Volumes

- Alternated thinning and pruning

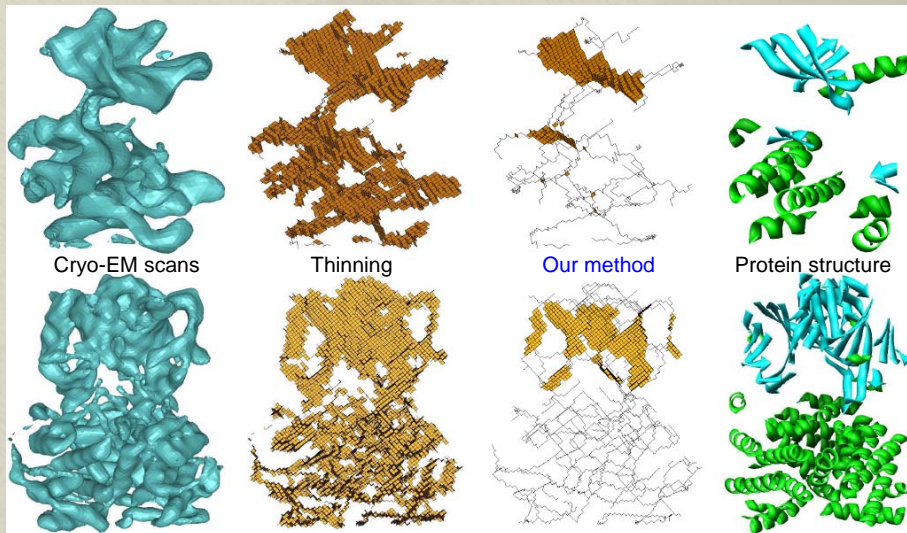


Algorithm 1: skeletons from Binary Volumes

- Alternated thinning and pruning

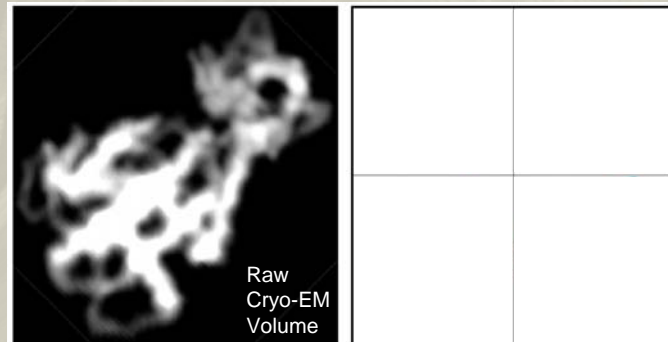


Results



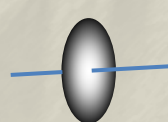
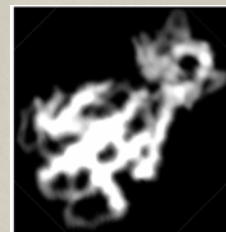
Algorithm 2: Skeletons from Grayscale Volumes

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

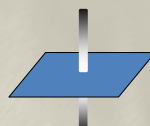


Algorithm 2: Skeletons from Grayscale Volumes

- 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



Skeleton curve



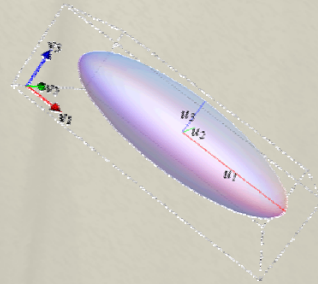
Skeleton surface

Algorithm 2: Skeletons from Grayscale Volumes

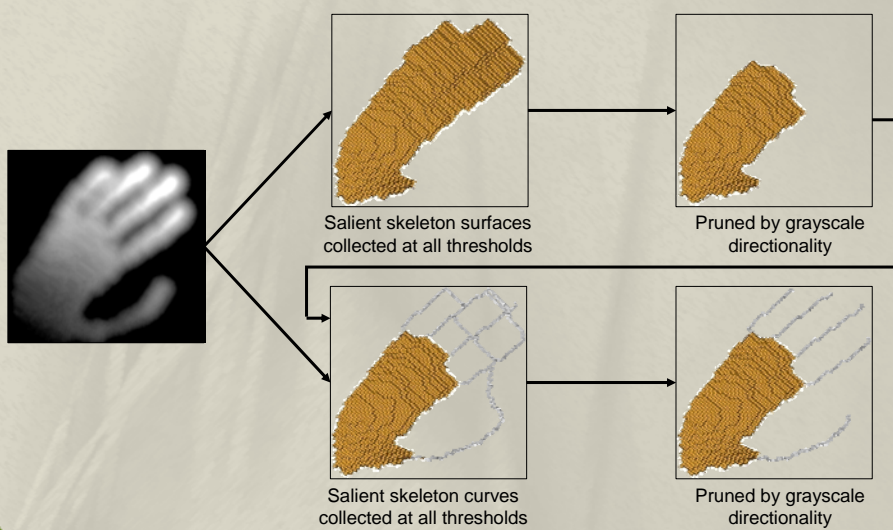
- Combine skeletons at various thresholds and prune using grayscale directionality
 - Directionality analysis: structure tensor
 - Eigenvectors and eigenvalues give principle directions and amount of intensity variation

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

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



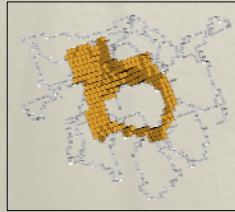
Algorithm 2: Skeletons from Grayscale Volumes



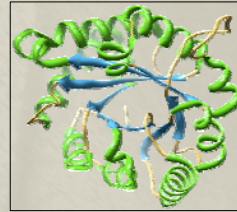
Results



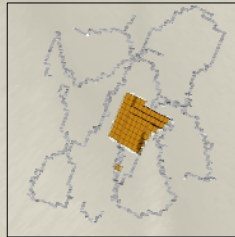
CryoEM scans



Hybrid skeletons



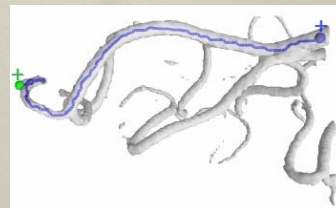
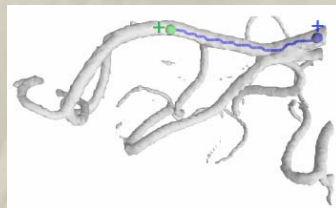
Protein structure



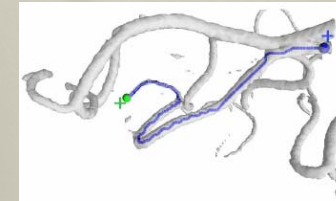
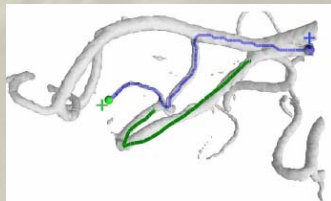
Algorithm 3: Interactive Skeleton Sketching

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

End-point
clicking:



Scribbling:



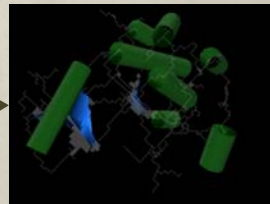
Geometric Algorithms

- 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

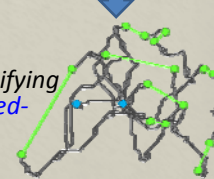
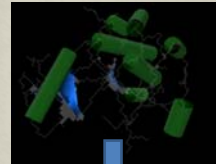
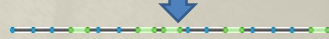
```
SSVWPEVMSREK1STGKLEK2QV3QV4QV5QV6QV7QV8QV9QV10QV11QV12QV13QV14QV15QV16QV17QV18QV19QV20QV21QV22QV23QV24QV25QV26QV27QV28QV29QV30QV31QV32QV33QV34QV35QV36QV37QV38QV39QV40QV41QV42QV43QV44QV45QV46QV47QV48QV49QV50QV51QV52QV53QV54QV55QV56QV57QV58QV59QV60QV61QV62QV63QV64QV65QV66QV67QV68QV69QV70QV71QV72QV73QV74QV75QV76QV77QV78QV79QV80QV81QV82QV83QV84QV85QV86QV87QV88QV89QV90QV91QV92QV93QV94QV95QV96QV97QV98QV99QV100QV101QV102QV103QV104QV105QV106QV107QV108QV109QV110QV111QV112QV113QV114QV115QV116QV117QV118QV119QV120QV121QV122QV123QV124QV125QV126QV127QV128QV129QV130QV131QV132QV133QV134QV135QV136QV137QV138QV139QV140QV141QV142QV143QV144QV145QV146QV147QV148QV149QV150QV151QV152QV153QV154QV155QV156QV157QV158QV159QV160QV161QV162QV163QV164QV165QV166QV167QV168QV169QV170QV171QV172QV173QV174QV175QV176QV177QV178QV179QV180QV181QV182QV183QV184QV185QV186QV187QV188QV189QV190QV191QV192QV193QV194QV195QV196QV197QV198QV199QV200QV201QV202QV203QV204QV205QV206QV207QV208QV209QV210QV211QV212QV213QV214QV215QV216QV217QV218QV219QV220QV221QV222QV223QV224QV225QV226QV227QV228QV229QV230QV231QV232QV233QV234QV235QV236QV237QV238QV239QV240QV241QV242QV243QV244QV245QV246QV247QV248QV249QV250QV251QV252QV253QV254QV255QV256QV257QV258QV259QV260QV261QV262QV263QV264QV265QV266QV267QV268QV269QV270QV271QV272QV273QV274QV275QV276QV277QV278QV279QV280QV281QV282QV283QV284QV285QV286QV287QV288QV289QV290QV291QV292QV293QV294QV295QV296QV297QV298QV299QV300QV301QV302QV303QV304QV305QV306QV307QV308QV309QV310QV311QV312QV313QV314QV315QV316QV317QV318QV319QV320QV321QV322QV323QV324QV325QV326QV327QV328QV329QV330QV331QV332QV333QV334QV335QV336QV337QV338QV339QV340QV341QV342QV343QV344QV345QV346QV347QV348QV349QV350QV351QV352QV353QV354QV355QV356QV357QV358QV359QV360QV361QV362QV363QV364QV365QV366QV367QV368QV369QV370QV371QV372QV373QV374QV375QV376QV377QV378QV379QV380QV381QV382QV383QV384QV385QV386QV387QV388QV389QV390QV391QV392QV393QV394QV395QV396QV397QV398QV399QV400QV401QV402QV403QV404QV405QV406QV407QV408QV409QV410QV411QV412QV413QV414QV415QV416QV417QV418QV419QV420QV421QV422QV423QV424QV425QV426QV427QV428QV429QV430QV431QV432QV433QV434QV435QV436QV437QV438QV439QV440QV441QV442QV443QV444QV445QV446QV447QV448QV449QV450QV451QV452QV453QV454QV455QV456QV457QV458QV459QV460QV461QV462QV463QV464QV465QV466QV467QV468QV469QV470QV471QV472QV473QV474QV475QV476QV477QV478QV479QV480QV481QV482QV483QV484QV485QV486QV487QV488QV489QV490QV491QV492QV493QV494QV495QV496QV497QV498QV499QV500QV501QV502QV503QV504QV505QV506QV507QV508QV509QV510QV511QV512QV513QV514QV515QV516QV517QV518QV519QV520QV521QV522QV523QV524QV525QV526QV527QV528QV529QV530QV531QV532QV533QV534QV535QV536QV537QV538QV539QV540QV541QV542QV543QV544QV545QV546QV547QV548QV549QV550QV551QV552QV553QV554QV555QV556QV557QV558QV559QV560QV561QV562QV563QV564QV565QV566QV567QV568QV569QV570QV571QV572QV573QV574QV575QV576QV577QV578QV579QV580QV581QV582QV583QV584QV585QV586QV587QV588QV589QV590QV591QV592QV593QV594QV595QV596QV597QV598QV599QV600QV601QV602QV603QV604QV605QV606QV607QV608QV609QV610QV611QV612QV613QV614QV615QV616QV617QV618QV619QV620QV621QV622QV623QV624QV625QV626QV627QV628QV629QV630QV631QV632QV633QV634QV635QV636QV637QV638QV639QV640QV641QV642QV643QV644QV645QV646QV647QV648QV649QV650QV651QV652QV653QV654QV655QV656QV657QV658QV659QV660QV661QV662QV663QV664QV665QV666QV667QV668QV669QV670QV671QV672QV673QV674QV675QV676QV677QV678QV679QV680QV681QV682QV683QV684QV685QV686QV687QV688QV689QV690QV691QV692QV693QV694QV695QV696QV697QV698QV699QV700QV701QV702QV703QV704QV705QV706QV707QV708QV709QV710QV711QV712QV713QV714QV715QV716QV717QV718QV719QV720QV721QV722QV723QV724QV725QV726QV727QV728QV729QV730QV731QV732QV733QV734QV735QV736QV737QV738QV739QV740QV741QV742QV743QV744QV745QV746QV747QV748QV749QV750QV751QV752QV753QV754QV755QV756QV757QV758QV759QV760QV761QV762QV763QV764QV765QV766QV767QV768QV769QV770QV771QV772QV773QV774QV775QV776QV777QV778QV779QV780QV781QV782QV783QV784QV785QV786QV787QV788QV789QV790QV791QV792QV793QV794QV795QV796QV797QV798QV799QV800QV801QV802QV803QV804QV805QV806QV807QV808QV809QV810QV811QV812QV813QV814QV815QV816QV817QV818QV819QV820QV821QV822QV823QV824QV825QV826QV827QV828QV829QV830QV831QV832QV833QV834QV835QV836QV837QV838QV839QV840QV841QV842QV843QV844QV845QV846QV847QV848QV849QV850QV851QV852QV853QV854QV855QV856QV857QV858QV859QV860QV861QV862QV863QV864QV865QV866QV867QV868QV869QV870QV871QV872QV873QV874QV875QV876QV877QV878QV879QV880QV881QV882QV883QV884QV885QV886QV887QV888QV889QV890QV891QV892QV893QV894QV895QV896QV897QV898QV899QV900QV901QV902QV903QV904QV905QV906QV907QV908QV909QV910QV911QV912QV913QV914QV915QV916QV917QV918QV919QV920QV921QV922QV923QV924QV925QV926QV927QV928QV929QV930QV931QV932QV933QV934QV935QV936QV937QV938QV939QV940QV941QV942QV943QV944QV945QV946QV947QV948QV949QV950QV951QV952QV953QV954QV955QV956QV957QV958QV959QV960QV961QV962QV963QV964QV965QV966QV967QV968QV969QV970QV971QV972QV973QV974QV975QV976QV977QV978QV979QV980QV981QV982QV983QV984QV985QV986QV987QV988QV989QV990QV991QV992QV993QV994QV995QV996QV997QV998QV999QV1000
```



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

```
SSVFPDDEWVSREK I TLLRELGG SFGM VYEGNARD I I KGE AETRVAVK I TVNESAS L RERIEFLNEASVM KGFTC ...
```



- 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)

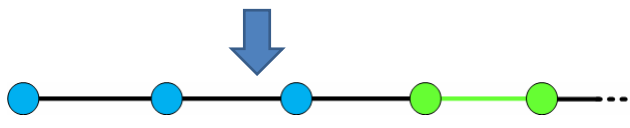


Sequence as Graph

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

Annotated Sequence

```
SSVFPDDEWVSREK I TLLRELGG SFGM VYEGNARD I I KGE AETRVAVK I TVNESAS L RERIEFLNEASVM KGFTC ...
```

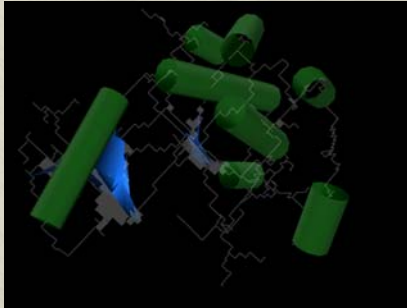


Graph

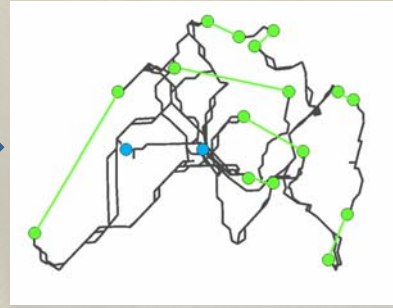


Skeleton as Graph

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

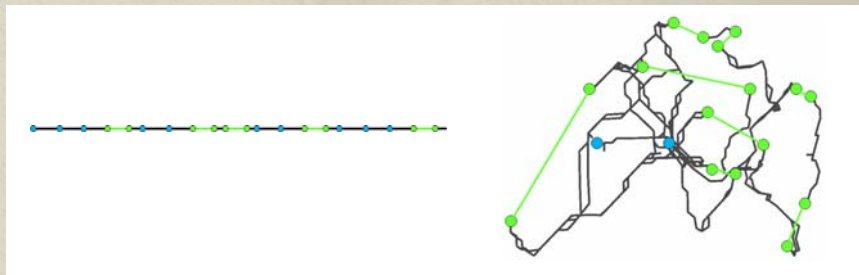


Annotated Skeleton



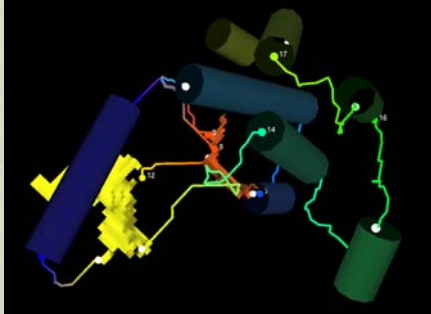
Graph

Graph Matching

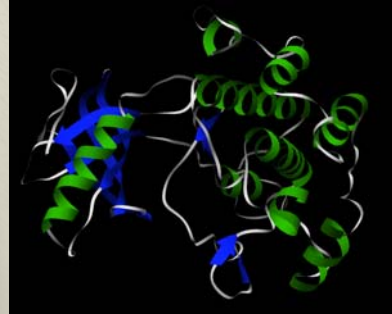


- A* search algorithm
- Sheet nodes (blue) in skeleton can be visited more than once
- Modify graph structure to handle noise
 - Missing helices and sheets
 - Missing loops in skeleton
 - Broken sheets
- Now available in Gorgon: <http://cs.wustl.edu/~ssa1/gorgon>

Results: 1IRK (simulated data)



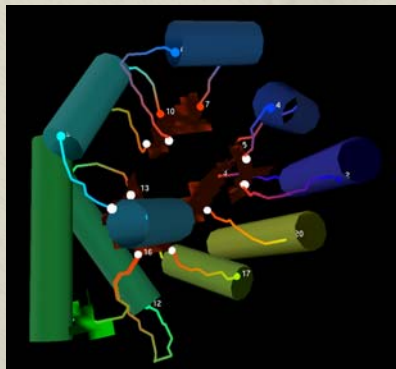
Lowest-cost match



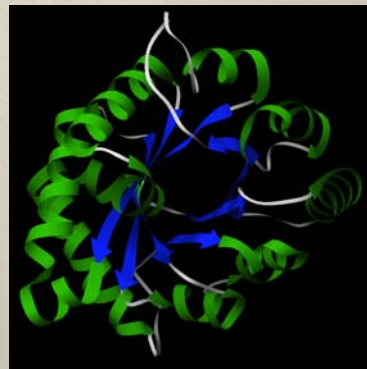
X-ray Structure

- 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



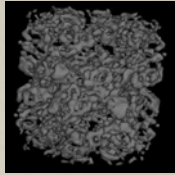
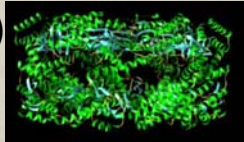
X-ray Structure

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

Geometric Algorithms

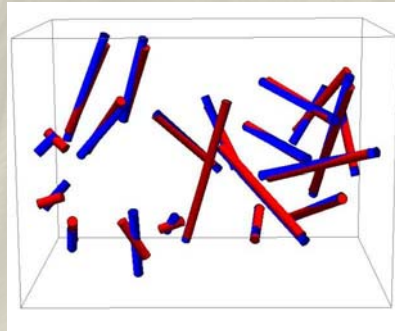
- Shape analysis of density maps
 - Skeletonization
- Topology analysis of SSEs
 - Graph matching
- Flexible model fitting
 - Shape registration

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

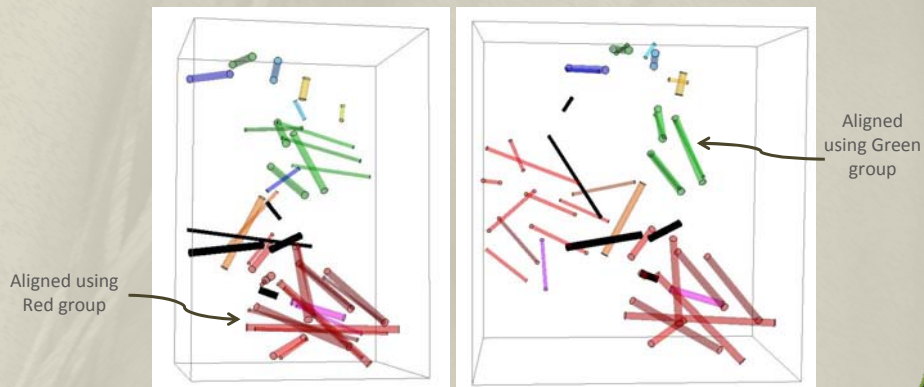
- In different conformations or in similar molecules, corresponding helices are often similar in shape



1OEL -A (red) and 1SS8 -A (blue)

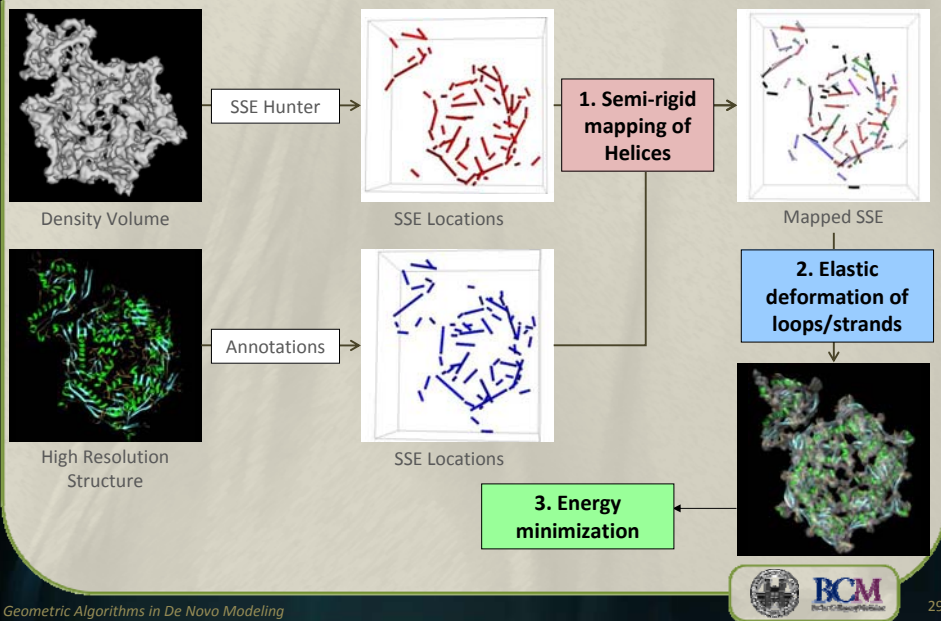
Observation #2

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



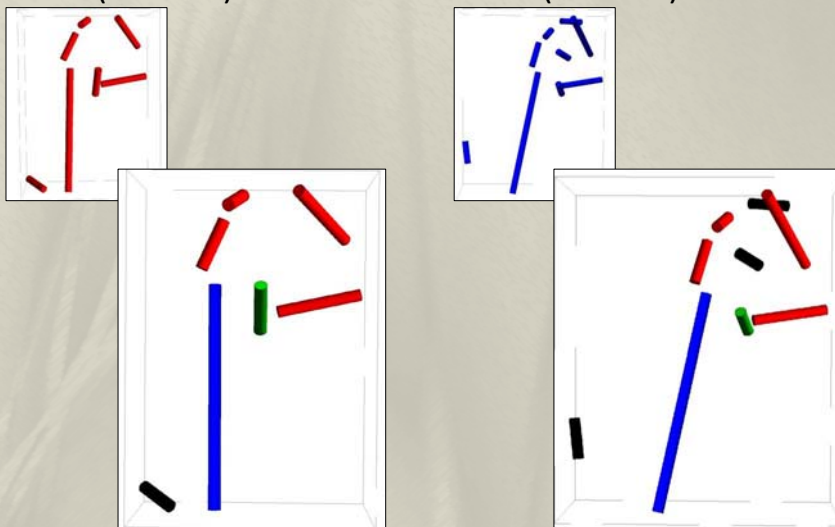
1OEL -A (thin cylinders) and 2C7C -A (thick cylinders)

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



Preliminary Results: Helix Mapping

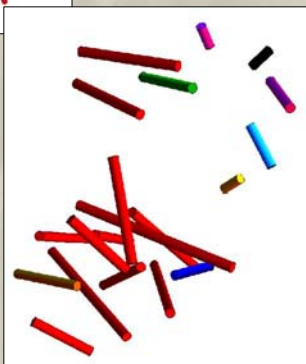
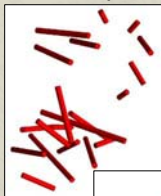
- 3E8K (Chain A) - 7 Helices
- 2GP1 (Chain A) - 9 Helices



Executed in 0.1s

Preliminary Results: Helix Mapping

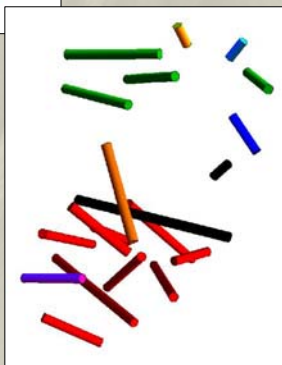
- 1OEL (Chain A) - 20 Helices
- 1SS8 (Chain A) - 20 Helices



Executed in 0.2s

Preliminary Results: Helix Mapping

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



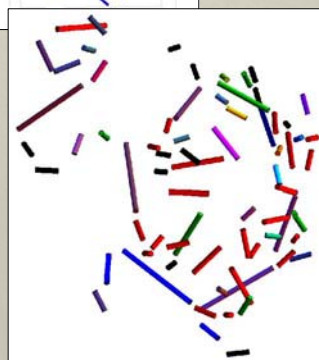
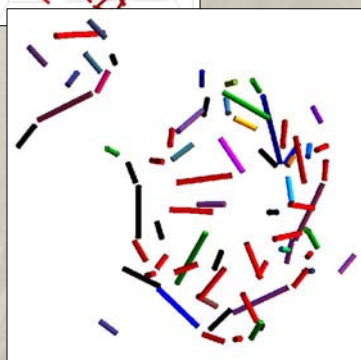
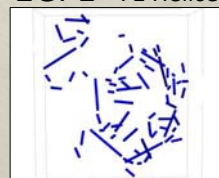
Executed in 0.3s

Preliminary Results: Helix Mapping

- 3E8K - 71 Helices



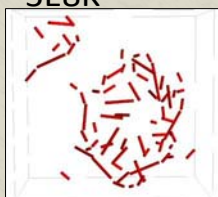
- 2GP1 - 71 Helices



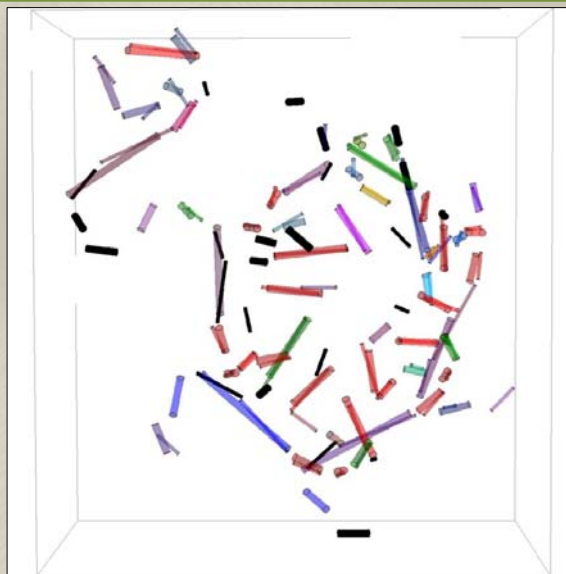
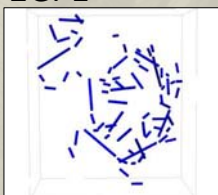
Executed in 84s

Preliminary Results: Helix Mapping

- 3E8K



- 2GP1



Summary

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