Analysis and Modeling of a Cryo-EM Density Map

Matthew L. Baker, Ph.D.
National Center for Macromolecular Imaging
Baylor College of Medicine
Resolution in Cryo-EM: GroEL

EMDB 5143: 18Å
EMDB 1042: 10Å
EMDB 1200: 8Å
EMDB 5001: 4Å
Tools

Density Manipulations
• **Filtering**: Mask or de-noise the density map
• **Segmentation**: Identifying and isolating single subunits or domains in the density
• **Feature Recognition**: Identify and characterize density features

Modeling Tools
• **Fitting**: Localizing a known structure within the density map
• **Constrained Modeling**: Generating computational models in the context of density
• **Feature Recognition**: Identify density features
• **De novo modeling**: Model building without a template
• **Optimization**: Refinement of a model in the density map
Analyzing Cryo-EM Density Maps

From Cryo-EM Density Maps to Models

Building an IP3R1 Subunit

Fitting

Poorly resolved
de novo homology
Fitting Known Structures

• Known structures/models (probes) for one or more of the components can be fit to a cryoEM density map (target)

• Interactive and automated programs

• Assess fit to map (Methods include correlation, atom inclusion/exclusion and clashes avoidance

• Sequential or simultaneous fit of multiple models

• Rigid body vs flexible fit

• Map resolvability key in determining quality of fit

Fitting Software

• Rigid body

  ➤ Foldhunter, Gorgon, EMFIT, UCSF Chimera, CoAn, Situs, UROX

• Flexible

  ➤ Gorgon, MDFF, EMFF, Phenix, FlexEM, MDFit, MVP-Fit, Direx, Norma
IP3R1: Fitting a Known Structure

- X-ray structure of the N-terminal domain
  - 4 overlapping structures for residues 7-587
  - Includes both ligand-bound and un-bound structures
- Corresponding N-terminal domain of related RyR1 also available and can “fill-in” missing portions of IP3R1 model
- Fit to entire density map with Chimera, Foldhunter and Situs
  - ~2Å RMSD between different models and fits
IP3R1: Flexible Fitting

- Models refined to density map using real-space refinement tools in Phenix and FlexEM
- 2.4Å RSMD from original rigid-body fit structure
Density Constrained Modeling

- Sequence-based searches reveal two regions of sequence homology using Phyre and RaptorX
  - Armadillo repeat domain (ARM2, \(\sim 1050-1500\))
  - Transmembrane domain (TMD, \(\sim 2200-2600\))
- Homology models fit with Chimera, Foldhunter and Situs
- Models refined against density map with Rosetta and Phenix
Feature Recognition

• Localization of individual secondary structure elements within a density map
• Interactive and automated programs
• Can provide a simple topological model

Secondary Structure Detection Software

• Helixhunter, SSEHunter, StrandTwister, HelixTracer, sheetminer, sheettracer, Gorgon, Pathwalker
Detecting Secondary Structure Elements

- SSEHunter: guided identification of alpha helices and beta sheets at intermediate resolutions
- Scoring based on correlation, skeletonization and local geometry
- >95% helix (2+ turns) detection accuracy
- >99% detection of 3+ stranded sheets

SSEHunter Methodology
82 “good” helices identified in the density of map with SSEHunter

Two sheet domains

SSEHunter helices matched helices in fitted structures

>90 helices per IP3R1 monomer predicted in sequence
From Map to Model
Building SSE Connections

• Density skeletonization - compact geometric representation of a volume
• Feature preserving
  - Sheets are represented as flat surfaces
  - Helices and loops are represented as curves
• Topology preserving
  - Maintains density connectivity
  - minimizes branches and breaks

Step 3: de novo Modeling

...IDDTEKTVGSKTVGVNSAAMTQGTVVAVLAIICNVEKTEAVTEDPAKMEFIQEEKMKEIVAVRASVGAYFCGEGSDLQHVLAKAGLVA...

reconstruction

segmented subunit

SSEs

skeleton

initial backbone model
IP3R1 de novo Model

- Fitted models and SSE correspondence serve as anchor points
- Complete topological model for ~85% of the IP3R1 monomer
  - Density for the loops at the three splice variants are missing
  - Small loops missing in N-terminal domain
  - Missing connection between ARM2 and ARM3 ~(100aa)
- Model consists of both full-atom domains (Fitted and homology models) and C-alpha only domains (de novo)
Model Building

4.2Å resolution GroEL

4.5Å resolution ε15


Gorgon

Interactive molecular modeling toolkit for intermediate resolution density maps focused on providing a simple and efficient framework for de novo modeling

- Annual workshops and trainings
- On-line videos and tutorials with sample data
- Cross platform (Windows, Linux, OS X 10.5+)

http://gorgon.wustl.edu

Model Building in Gorgon

- Density map and model visualization
- Density skeletonization (binary, grey-scale and interactive)
- SSE identification and building using SSEHunter
- SSE correspondence searches with helices and sheets
- Semi-automated atom placement
- Rigid body and flexible fitting
Interactive, semi-automated model building with density constraints
- Interactive, sketching of loops
- Auto-build of SSE
- Manual editing with local fitting

Sequence to structure correspondence using graph matching
- Based on helix position and distance/connectivity
- Gallery of correspondences
- Partial assignments

Modeling Tools
Features in Gorgon v2.2

- Improved SSE correspondence search with β sheets
- Complete SSEHunter and SSEBuilder integration
- Rigid-body and flexible fitting routines
- Session support to save and load work-in-progress
- Improved user-interface

DE NOVO MODEL BUILDING
TUTORIAL: GORGON
Model Building With Gorgon

- **The data:** Rotavirus VP6
  - 3.80 Å resolution
  - 1.23 Å/pixel
  - Monomer segmented with Chimera
  - EMDB ID: 1461
  - X-ray structure: 1QHD
Gorgon Basics: Layout

- **Menu Bar**
- **Visualization Window**
- **Options Window**
- **Volume/Surface Editor**
Gorgon Basics: Opening a map

- File>Open>Volume “…Data-Sets/Gorgon/vp6-96o.mrc”
- Adjust transparency and color in options menu:
  - click on grey box next to “show model colored”
- Adjust isosurface in volume/surface editor options
## Gorgon Basics: Controls

<table>
<thead>
<tr>
<th></th>
<th>Left button</th>
<th>Middle button</th>
<th>Right button</th>
<th>Wheel</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Click</strong></td>
<td>Selection</td>
<td>-</td>
<td>Focus (atom only)</td>
<td>Zoom</td>
</tr>
<tr>
<td><strong>Ctrl+click</strong></td>
<td>Toggle selection</td>
<td>-</td>
<td>Focus (atom only)</td>
<td>Change isosurface</td>
</tr>
<tr>
<td><strong>Click+drag</strong></td>
<td>Rotate</td>
<td>-</td>
<td>Translate</td>
<td>-</td>
</tr>
</tbody>
</table>

Use the Apple key instead of Ctrl on a Mac
Gorgon: SSEHunter

• Build a skeleton
  – Actions> Volume> Skeletonization
  – Binary skeleton; select a threshold where the separation of stands and loops can be first seen (~0.40)
  – Sheets are in yellow, loops are in red
Gorgon: SSEHunter

- Calculate SSEHunter scores
  - Actions> Secondary Structure Elements> Identify SSEs
  - Threshold ~0.40, resolution 8.00*
  - Click on Find Scored Pseudoatoms OR load from file “skeleton-vp6-b0.40.mrc”

* The resolution of the map is ~4Å; it is not necessary to adjust this parameter in general when working with high resolution data
Gorgon: SSEHunter

- Build SSE models
  - Ctrl+click on red Ca-atoms at beginning and end of helix
  - Click “Add Helix”
  - Select Helix in visualization window and press Ctrl+f to refine fit to density
  - Ctrl+click on blue Ca-atoms in a plane
  - Click “Add Sheet”
  - Iterate until all visible SSEs are annotated

*** When finished, close Ca file (the ssehunter score) in the file menu and save helices/sheets as VRML ***
Gorgon: SSE Correspondence

- Generating an SSE correspondence
  - Actions > Secondary Structure Elements > Find SSE correspondence
- Files
  - Cryo-EM skeleton: “skeleton-vp6-b0.40.mrc”
  - Sequence: “vp6.pdb”
  - 3D Helix locations: “helices-vp6.vrml”
  - 3D Sheet locations: “sheets-vp6.vrml”
- Click “OK”

*** Close all SSEs (VRML files) from the previous step before loading the SSEs in the example ***
Gorgon: SSE Correspondence

- Select a correspondence
  - Evaluate correspondence by comparing lengths, percentiles and overall topology
  - Click on helices in visualization window to view correspondence
  - Constrain “good” matches by clicking on “Constrained” box
  - Click “OK” to re-run correspondence routine with desired selection
Gorgon: Model Building

- Atom placement
  - Actions > C-alpha atoms > Semi-automated atom placement
  - Top panel is sequence viewer with SSE predictions
  - Current location in sequence is shown in grey bar and below in the zoom view
  - Bottom panel (atom panel) has 4 tabs for atom placement
Gorgon: Model Building

- Add a helix
  - Select “helix editor” in the atom panel
  - Ctrl+Click on a helix in the visualization window
  - Click on the corresponding helix segment in sequence viewer; sequence will be highlighted in black
  - Adjust length/position in the helix editor if desired
  - Click “Accept” to build a helix; atoms will appear in visualization window
  - Click on an assigned atom and locate it in the sequence view
  - If helix is reversed, flip helix by clicking on “flip” (at least 1 atom in the helix must be selected)
  - Repeat until all helices are assigned
Gorgon: Model Building

- Build a loop
  - Select “atomic editor” in the atom panel
  - Set C-alpha distance to 3.5Å
  - Click on a starting/ending atom in the vis window
  - Select direction in Atomic editor panel (next atom increments, previous atom decrements)
  - Selected residue is highlighted, residue to be placed is in green
  - Cycle through the “Use choice” positions to find best placement, current position for the atom to be placed is in cyan
  - Click on “Accept”
  - Repeat until next assigned atom or terminus is reached
Gorgon: Model Building

- Adding a loop
  - Select “loop editor” in the atom panel
  - Select residues between two assigned residues in the sequence window
  - Click “start loop placement”
  - Select start point by Ctrl+click on the desired start point
  - Move loop through density with alt+move
  - Click “End loop placement” when finished
Gorgon: Model Building

- Adjust atom positions
  - Select residue (click)
  - Adjust position by Ctrl+click +drag or use Position editor in Atom panel
    - Blue bonds are too short (<3.5Å)
    - Red bonds are too long (>4.2Å)
  - Relative sidechain size can be shown by selecting “mock sidechains”
  - Repeat until all atoms are adjusted

*** When finished, save model in the File menu as C-alpha atoms or use export to PDB***
DE NOVO MODEL BUILDING
TUTORIAL: PATHWALKING
Deriving a Model With Limited Constraints

Goal: Find a path or sets of paths that trace the complete path of a protein through a density map at near-atomic resolutions such that:
- No SSEs required
- No explicit sequence information required
- No structural template required
- Automated
- Optimized against biophysical constraints
Repurposing the Traveling Salesman Problem

- TSP calculates optimal route between cities by minimizing distance travelled
- Each city can only be visited once
- Several exact and approximate TSP solvers for thousands of nodes
Connecting the “Dots”
Protein Structure Determination with TSP

Re-pose de novo modeling as a 3D TSP problem

- C-alpha atoms are “cities”
- Protein backbone is not a minimal distance, rather optimal distance is 3.8Å*N
- Distance expressed as a deviation from 3.8Å

Building a Better Pathwalker

- Optimized pseudoatom generation
- Fully automated with “path checking"
- Density weighted paths
- Iterative SSE detection and pseudoatom placement
- Geometry filtering
- Sidechain filtering
- Multiple chains
Software

• EMAN2: cryo-EM image processing; pathwalker, a sequence free modeling tool
  – http://blake.bcm.tmc.edu/eman/eman2/

• UCSF Chimera: Visualization
  – http://www.cgl.ucsf.edu/chimera/

All software available for windows, OS X and linux
Model Building With Pathwalker

- The data: Beta-Galactosidase
  - 3.2 Å resolution
  - 0.6375 Å/pixel
  - Monomer segmented with Chimera
  - EMDB ID: 5995
  - X-ray structure: 3j7h
Pathwalker: Generating pseudoatoms

• In a terminal window run
  – e2proc3d.py sub-A.mrc map.mrc --
    process normalize.edgemean --
    process threshold.belowtozero
  – e2segment3d.py map.mrc --
    pdbout=pseudoatoms.pdb --
    process=segment.kmeans:ampweight
    =1:nseg=1022:verbose=1:minsegsep=
    1:pseudoatom=1:thr=10

• Open pseudoatoms.pdb file in
  Chimera
  – Show only atoms, not bonds or ribbons
Pathwalker: Calculating an initial path

• In a terminal window run
  – e2pathwalker.py pseudoatoms.pdb
    --mapfile=map.mrc --
    output=path0.pdb --solver=lkh --
    overwrite --dmin=1 --dmax=10 --
    mapthresh=12 —mapweight=200
  – Open path0.pdb file in chimera
  – Render as ball and stick
Pathwalker: Re-calculating the path

• In a terminal window run
  – `e2pathwalker.py pseudoatoms.pdb`  
    --mapfile=map.mrc  
    output=path1.pdb  
    --solver=lkh  
    overwrite  
    --dmin=1  
    --dmax=10  
    --mapthresh=12  
    --mapweight=200  
    --subunit=3
  – Open path1.pdb file in chimera
  – Render as ball and stick
Pathwalker: Fixing edges in the path

- In a terminal window run
  - `printf "870 1017\n827 829\n" > edge.txt`
  - `e2pathwalker.py pseudoatoms.pdb -- mapfile=map.mrc -- output=path2.pdb --solver=lkh -- overwrite --dmin=1 --dmax=10 -- mapthresh=12 --mapweight=200 --edgefile=edge.txt`
  - Open path2.pdb file in chimera
  - Render as ball and stick
Pathwalker: Setting the termini

• In a terminal window run
  – e2pathwalker.py pseudoatoms.pdb
    --mapfile=map.mrc --
    output=path3.pdb --solver=lkh --
    overwrite --dmin=1 --dmax=10 --
    mapthresh=12 --mapweight=200 --
    edgefile=edge.txt --start=91 --
    end=956
  – Open path3.pdb file in chimera
  – Render as ball and stick
Pathwalker: Finding helices

• In a terminal window run
  – /Applications/EMAN2/examples/e2pwhelixfit.py --mapin map.mrc --pdbin path3.pdb --output hlx.pdb --denthr 13 --mapwohelix map_nohlx.mrc --minlen 4 --lenthr 10
  – Open hlx.pdb file in chimera
Pathwalker: Finding sheets

• In a terminal window run
  – /Applications/EMAN2/examples/e2pwsheetfit.py --pdbin hlx.pdb --output sheet_0.pdb --nsht 30 --minlen 3
  – Open sheet_0.pdb file in chimera
## De Novo Modeling Utilities

<table>
<thead>
<tr>
<th></th>
<th>de novo</th>
<th>Gorgon</th>
<th>Pathwalker</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Map requirements</strong></td>
<td>full map</td>
<td>full map (&lt;256³)</td>
<td>segmented map</td>
</tr>
<tr>
<td><strong>SSE</strong></td>
<td>required</td>
<td>required</td>
<td>optional</td>
</tr>
<tr>
<td><strong>SSE correspondence</strong></td>
<td>required</td>
<td>required</td>
<td>none</td>
</tr>
<tr>
<td><strong>Completion time</strong></td>
<td>long (weeks)</td>
<td>short (0.5-1 day)</td>
<td>short (0.5-1 day)</td>
</tr>
<tr>
<td><strong>Model accuracy</strong></td>
<td>+++</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td><strong>User interface</strong></td>
<td>varied software</td>
<td>graphical</td>
<td>EMAN2+Gorgon</td>
</tr>
<tr>
<td><strong>Resolution range</strong></td>
<td>3-5Å</td>
<td>3-7Å</td>
<td>3-6.5Å</td>
</tr>
<tr>
<td><strong>Multiple models</strong></td>
<td>no</td>
<td>partial</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Ease of use</strong></td>
<td>difficult</td>
<td>easy to moderate</td>
<td>easy</td>
</tr>
</tbody>
</table>
Gorgon vs. Pathwalking

- red = X-ray structure
- blue = Gorgon mode
- green = Pathwalker model
## What Tools Should I Use?

<table>
<thead>
<tr>
<th>Gorgon</th>
<th>Pathwalker</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Single protein subunit</td>
<td>• Single or multiple subunits</td>
</tr>
<tr>
<td>• Between 3 and ~60</td>
<td>• No secondary structure</td>
</tr>
<tr>
<td>helices of varying sizes</td>
<td>elements detected</td>
</tr>
<tr>
<td>• Well resolved density</td>
<td>• Between 3-7Å resolution</td>
</tr>
<tr>
<td>• No size limit</td>
<td>• Up to ~2000 amino acids</td>
</tr>
<tr>
<td>• Interactive</td>
<td>• Automated</td>
</tr>
</tbody>
</table>
All Atom Modeling

From C-alpha to All Atom

REM0 is an algorithm for constructing protein atomic structures from C-alpha traces by optimizing the backbone hydrogen-bonding networks. A downloadable package (2.1M) of REMO is available at REMO.tar.gz. More details can be found at README. A newer version of on-line protein structure refinements through molecular dynamic simulations can be found at FG-MD.

Cut and paste your C-alpha trace structure in PDB format here:

Or upload the structure file from your local computer:

Email: (Mandatory, where results will be sent to)

run REMO

or phenix.pulchra sheet_0.pdb
Coot
Model Refinement with Rosetta and Phenix

Comparison to the X-ray structure


grey= de novo model
purple= X-ray structure
Side Chain and Model Refinement

- Higher resolution features visible in TM, ILD, LNK, CTD and ARM3 domains
- Extend C-alpha domain models to full atom models using REMO
- Iterative real-space refinement in Phenix and manual optimization in COOT
  - optimize ramachandran plot, rotamer selection, clash minimization
- Model characterization and validation with Phenix and MolProbity
Acknowledgments

**IP3R**
- Mariah Baker (UTH)
- Guizhen Feng (UTH)
- Irina Serysheva (UTH)
- Zhou Wang (BCM)
- Steve Ludtke (BCM)
- Wah Chiu (BCM)

**Pathwalker**
- Mariah Baker (BCM, UTH)
- Ian Rees (BCM)
- Steve Ludtke (BCM)
- Muyuan Chen (BCM)
- Wah Chiu (BCM)

**Gorgon**
- Tao Ju (WUSTL)
- Sasakthi Abeysinghe, (WUSTL)
- Stephen Schuh (WUSTL)
- Derrick Burrows (WUSTL)
- Hang Dao (WUSTL)
- Mike Marsh (BCM)
- Ross Coleman (BCM)
- Tunay Durmaz (BCM)