





CU Anschutz Medical School, January 2020

Model building (cryo-EM)

Dorothee Liebschner Lawrence Berkeley Laboratory

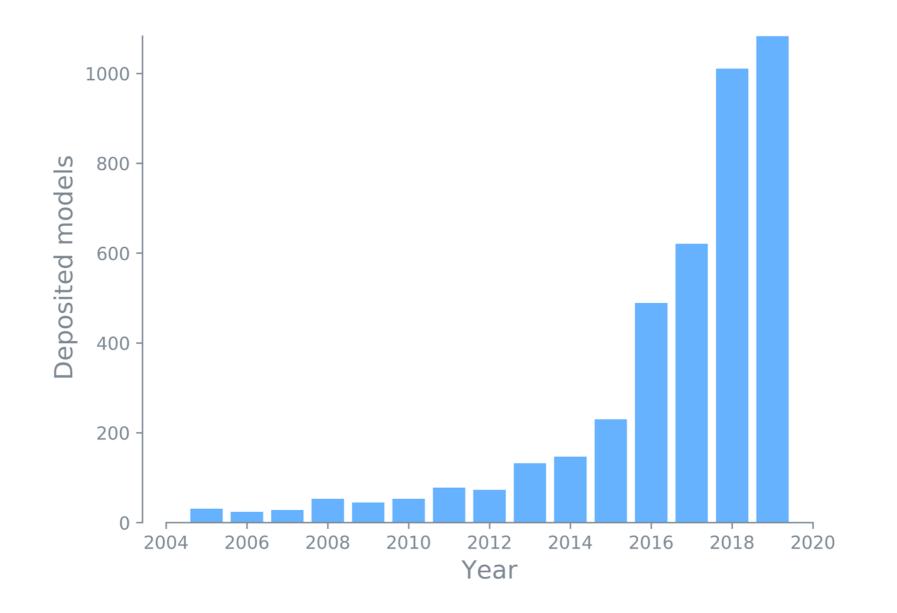






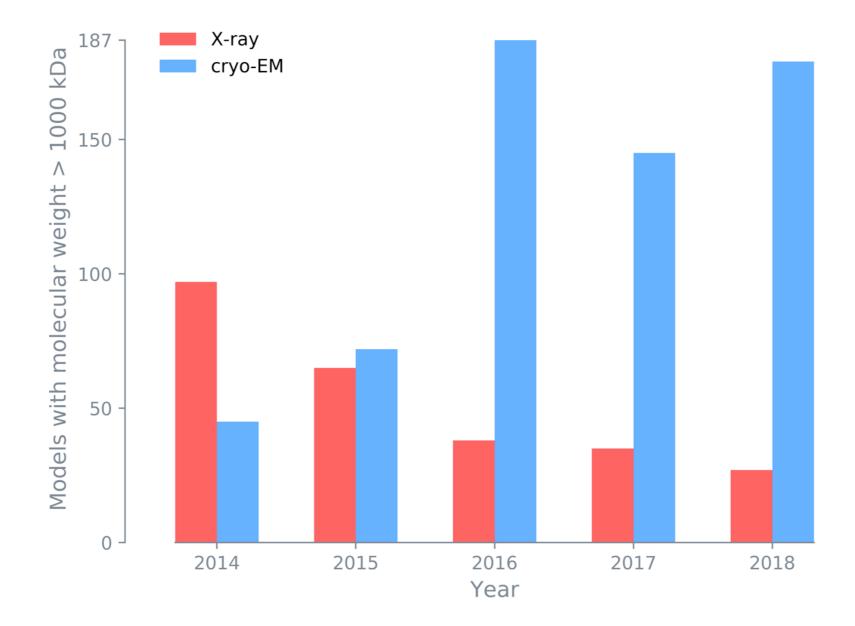


Cryo-EM models in the PDB

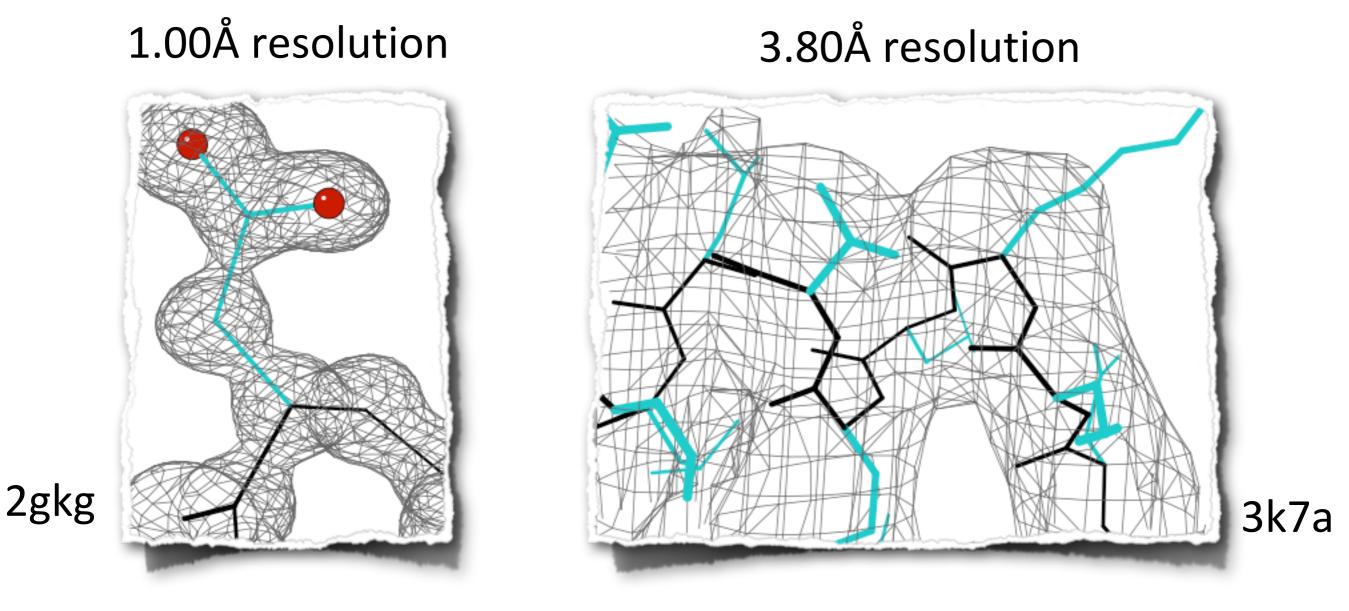


- Rapid growth since 2014
- More than 1000 models per year
- ~3% percent of entire PDB

Cryo-EM is used to determine large structures.



EM maps have low resolution



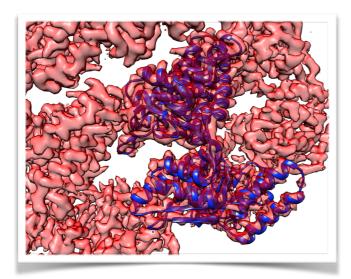
Challenges:

- How to interpret "featureless" maps?
- How to optimize models with sparse data?

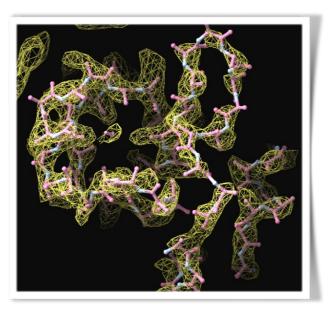
Model building tools in Phenix

Goal:

Obtain a molecular (atomic) model that fits the cryo-EM map.



Rigid model docking

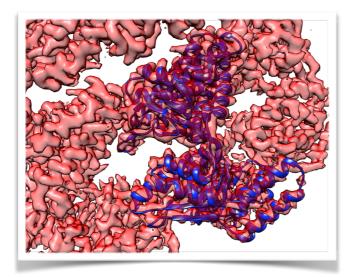


Automated model building

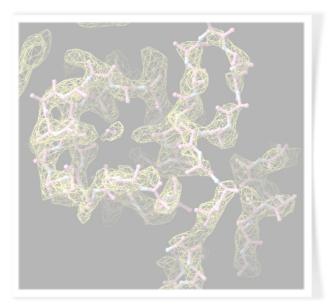
Model building tools in *Phenix*

1. Docking

Place an existing model into a map (move model as rigid body)

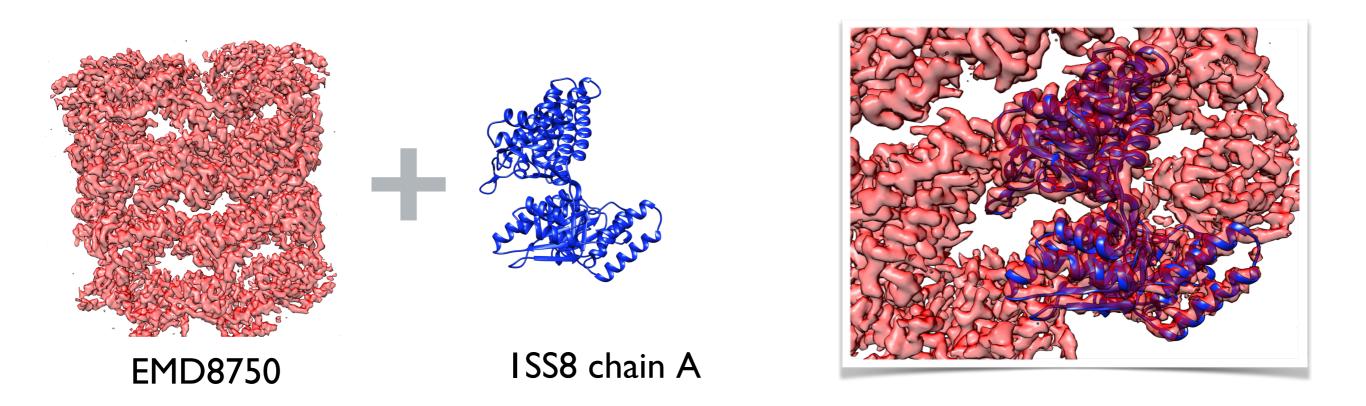


Rigid model docking



Automated model building

Rigid model docking

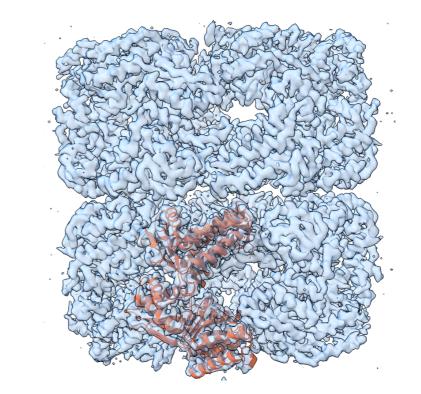


- Systematic cross correlation search of rotations and translations
- Performed in reciprocal space using FFT (very fast)
- Rigid body optimization of position

Rigid model docking

Dock in map (GUI) phenix.dock_in_map

•••			Dock in map (Project: groel_dock_refine)							
X	? {	े 😢 –	□ (3)							
Preferences	Help R	tun Abort	Ask for help							
Configure	DockInMap_1			4 Þ 🗙						
Dock a model or models in a cryo-EM map Inputs: Map file (mrc/ccp4/mtz map coefficients) Resolution limit (typically half-dataset FSC resolution) Search model file(s) (PDB, mmCIF) and/or search map files Optional fixed model file (PDB, mmCIF) Optional sequence file (any standard format or plain text) Optional symmetry file (phenix ncs_spec file, BIOMTR records) Dock_model_in_map will try to place the model in the map using a Fourier convolution search centered at the center of the model. The region occupied by the fixed model (if any) will be excluded. Job title :										
Input										
Map file :			/Users/dcliebschner/Documents/groel_dock_refine/emd_8750.mag	Browse						
Search model :			/Users/dcliebschner/Documents/groel_dock_refine/1ss8_A.pdb	Browse Q - +						
Sequence	file (optional) :			Browse						
Search m	odel copies (opt	ional) :								
Search map file (optional) :		l):		Browse 🤍 – +						
Search map copies (optional) :		nal) :								
Symmetry file (for multiple copies) :		e copies) :		Browse						
🗹 Use s	ymmetry									
Fixed model (optional) :				Browse						
🗹 Skip f	fixed model in p	db_out								
Options										
Resolutio	n :	4.0	Scattering table : electron ᅌ							
Number o	of processors :	4	Search options All parameters							
O Idle			Project: groel_c	lock_refine						



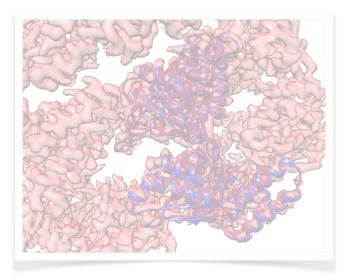


Map, model, resolution

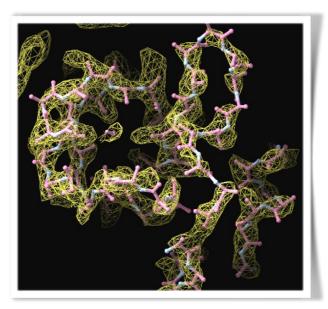
Model building tools in *Phenix*

2. Automated model building

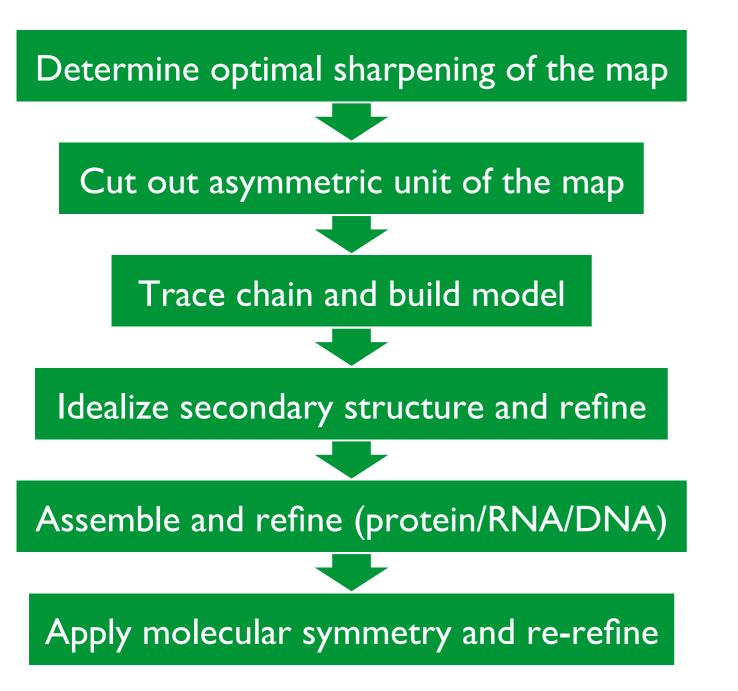
Build a model into a map from scratch.



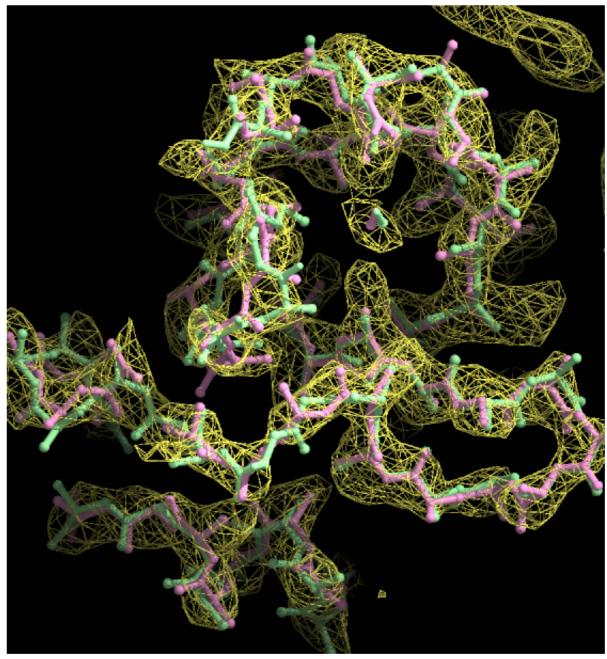
Rigid model docking



Automated model building



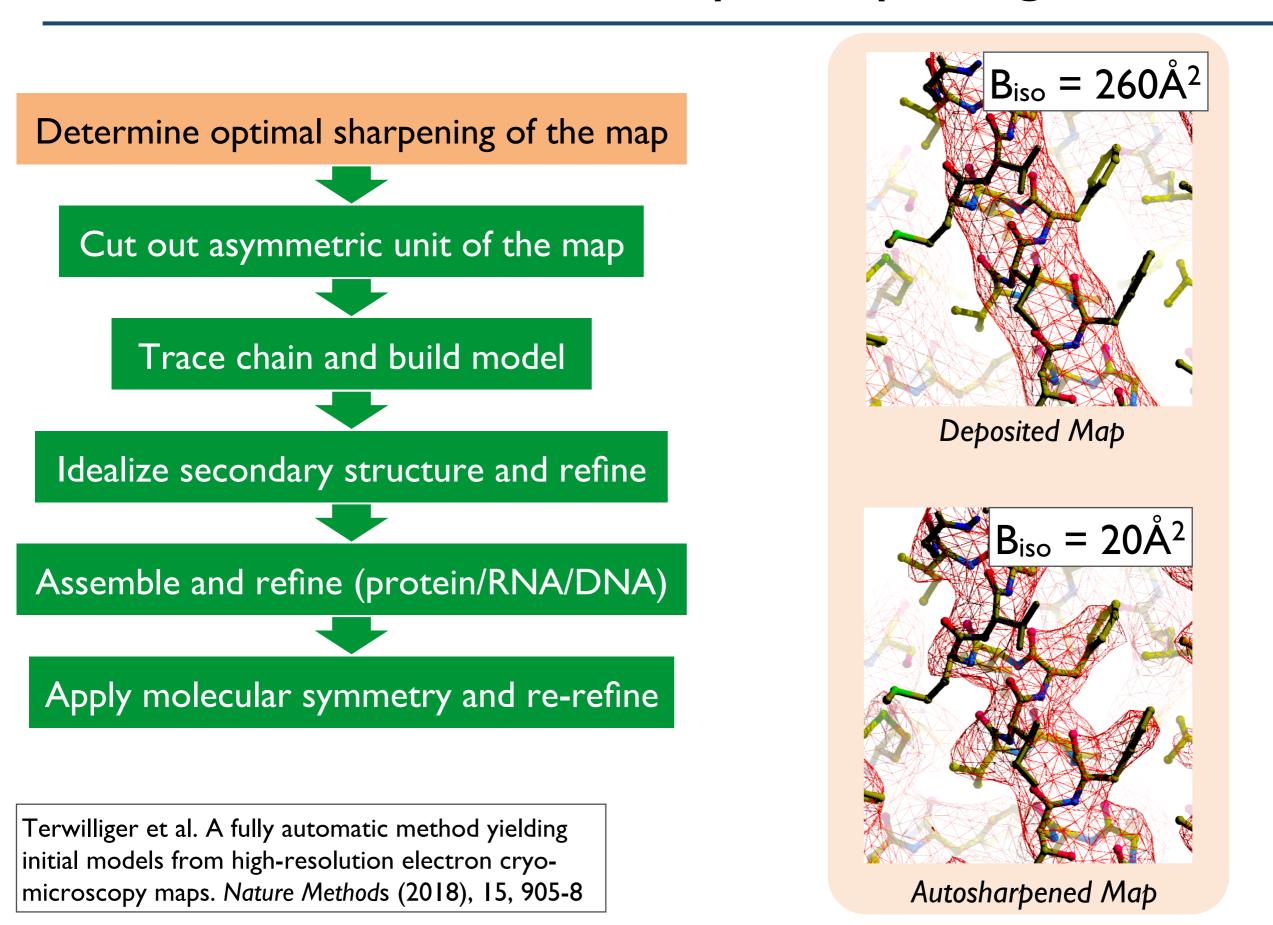
Terwilliger et al. A fully automatic method yielding initial models from high-resolution electron cryomicroscopy maps. *Nature Methods* (2018), 15, 905-8



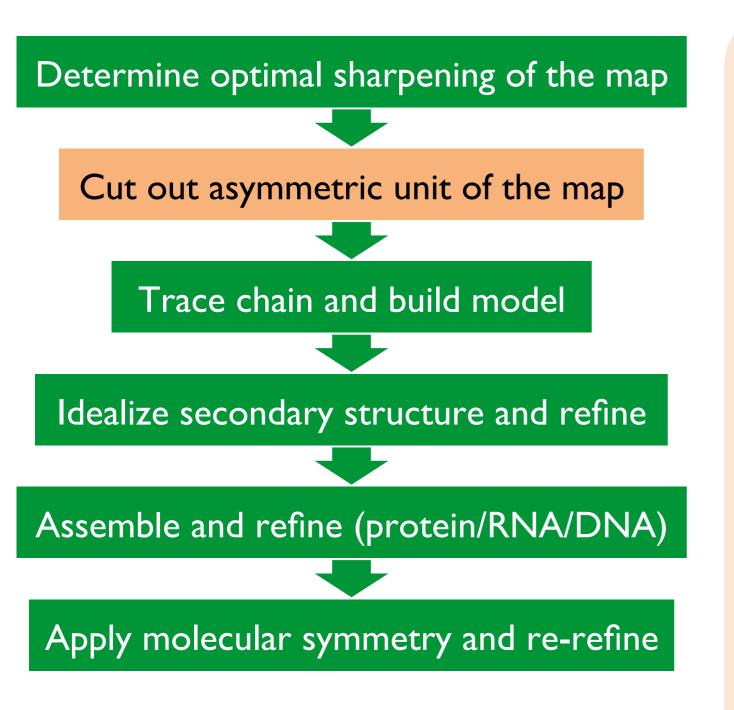
Cryo-EM map from the yeast mitochondrial ribosome (chain I of large subunit, 3.2Å, Amunts et al., 2014)

Autobuilt model (pink) Deposited model (green)

Automated Map Sharpening

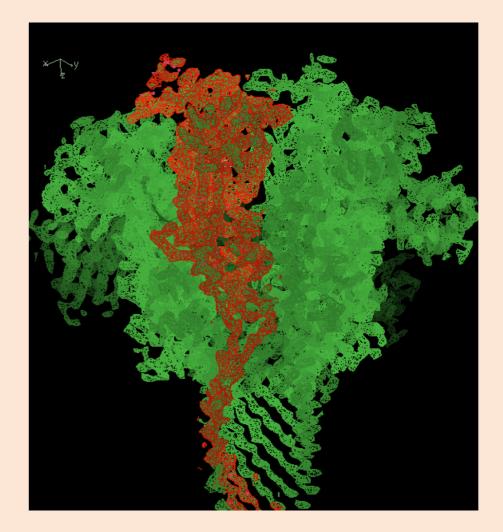


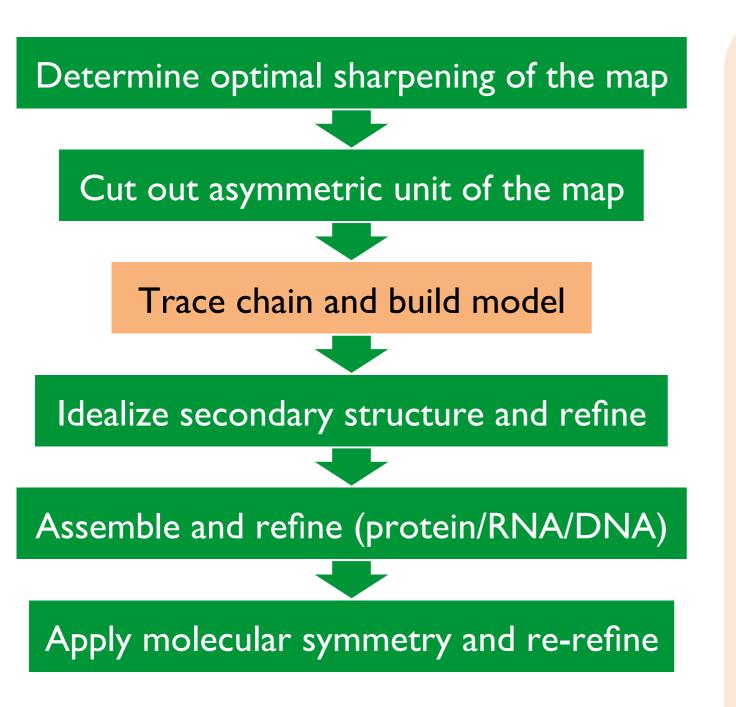
Automated Segmentation



Terwilliger et al. A fully automatic method yielding initial models from high-resolution electron cryomicroscopy maps. *Nature Methods* (2018), 15, 905-8

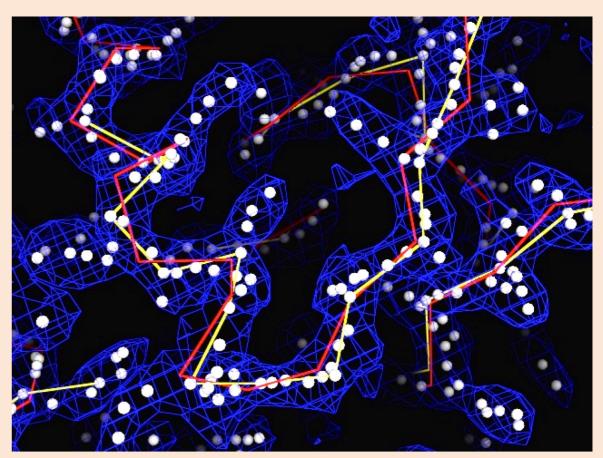
- Use symmetry of the map
- Identify regions representing the asymmetric unit of the map
- Choose symmetry copies that make compact molecule

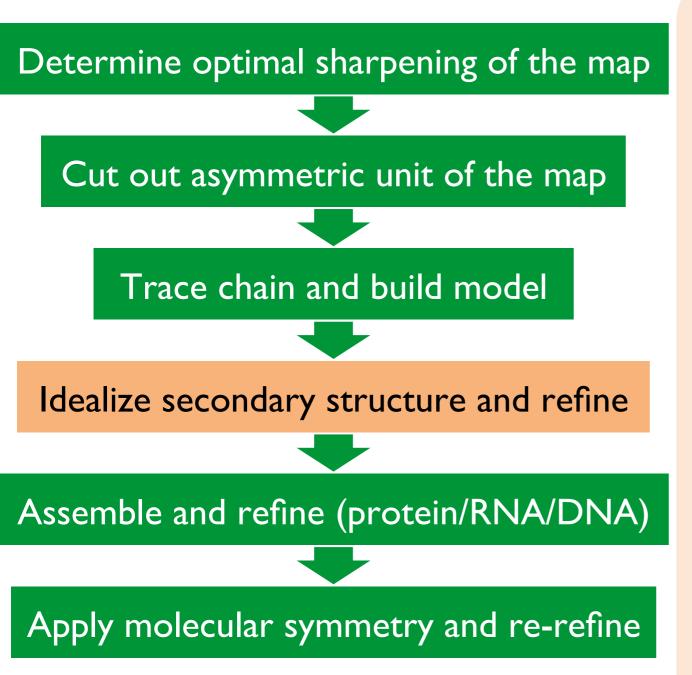




Terwilliger et al. A fully automatic method yielding initial models from high-resolution electron cryomicroscopy maps. *Nature Methods* (2018), 15, 905-8

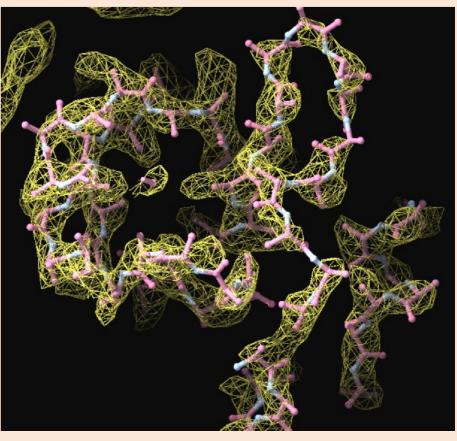
- Variable map thresholding
- Trace protein main chain
- Identify direction of main chain by fit to density



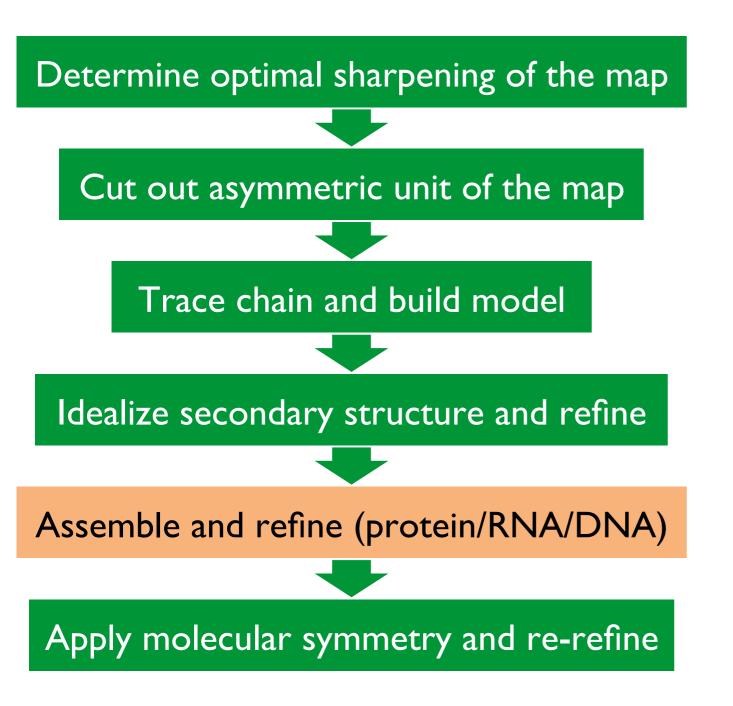


Terwilliger et al. A fully automatic method yielding initial models from high-resolution electron cryomicroscopy maps. *Nature Methods* (2018), 15, 905-8

- Refine and rebuild model (simulated annealing, rebuilding and combination of best parts of each model)
- Replace segments with idealized structure
- Identify hydrogen-bonding (β-sheets, α-helices) and use them as restraints in real-space refinement

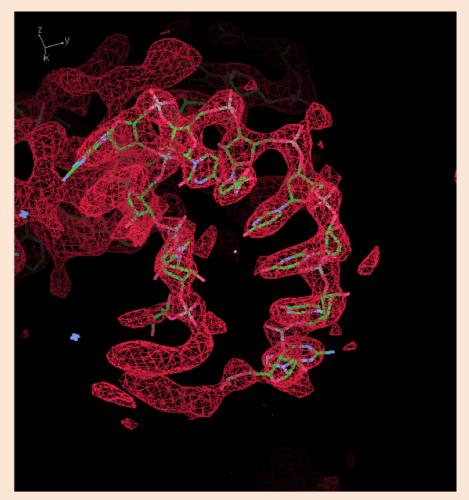


Chain I, yeast mitochondrial ribosome large subunit, 3.2 Å, 3j6b

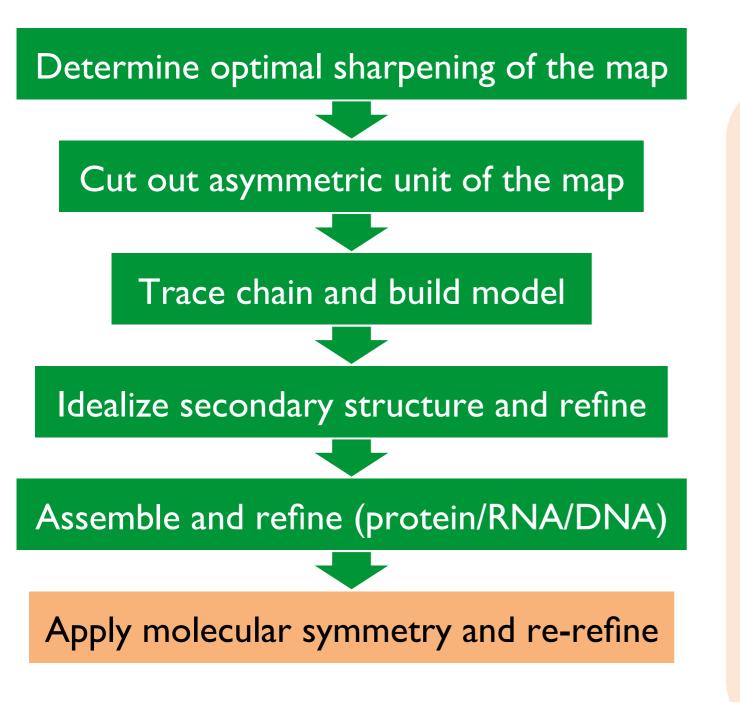


Terwilliger et al. A fully automatic method yielding initial models from high-resolution electron cryomicroscopy maps. *Nature Methods* (2018), 15, 905-8

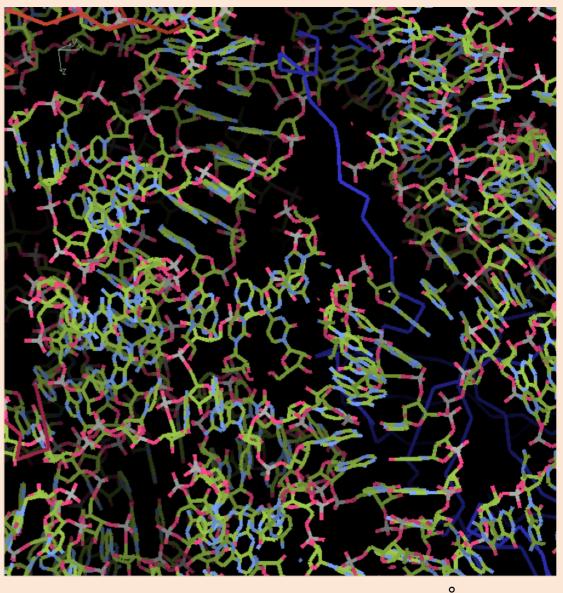
- Try building protein/RNA/DNA (whatever may be there)
- Choose segment type by map correlation



70S ribosome at 2.9 Å

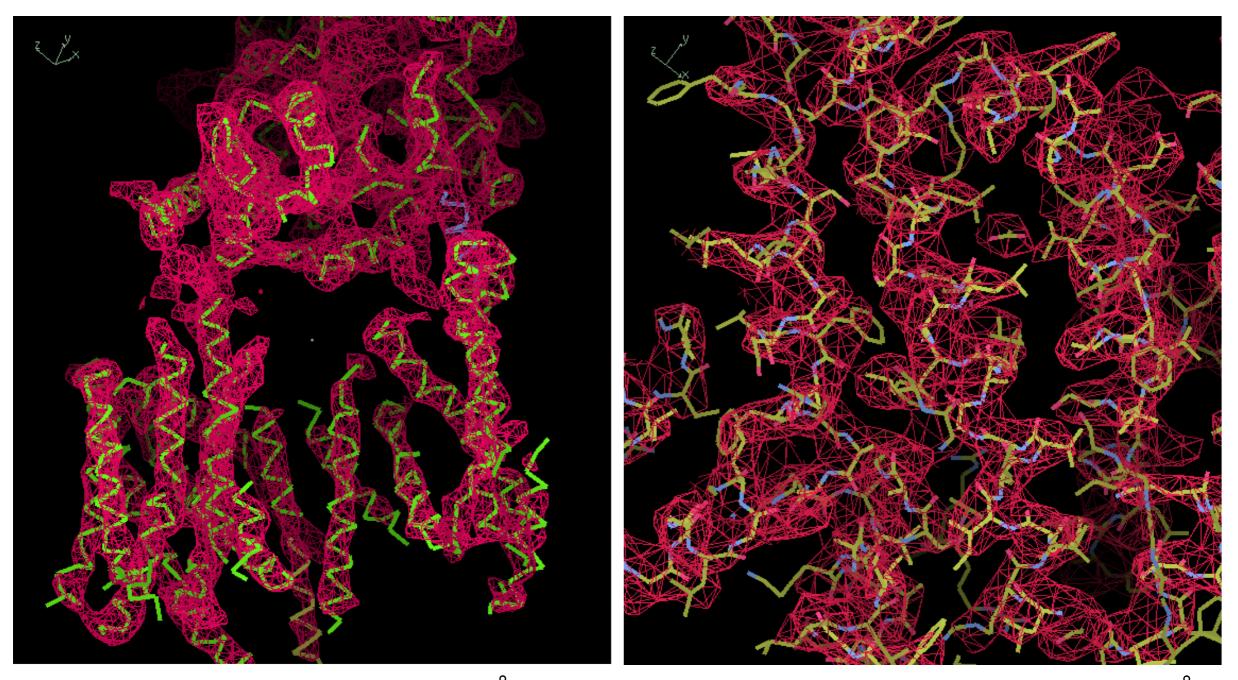


Terwilliger et al. A fully automatic method yielding initial models from high-resolution electron cryomicroscopy maps. *Nature Methods* (2018), 15, 905-8



30S Ribosome (1j5e, 2.9 Å)

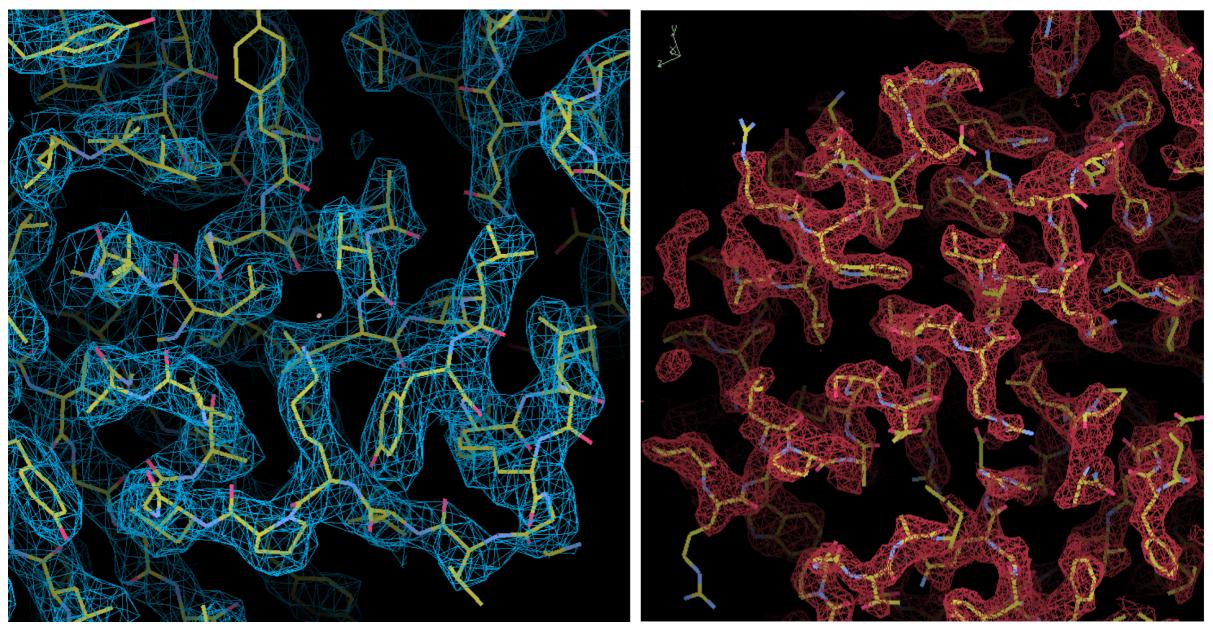
Building at Low Resolution



Gamma-secretase at 4.5 Å (autobuilt model; emd_2677)

Gamma-secretase structure at 3.4 Å (autobuilt model; emd_3061)

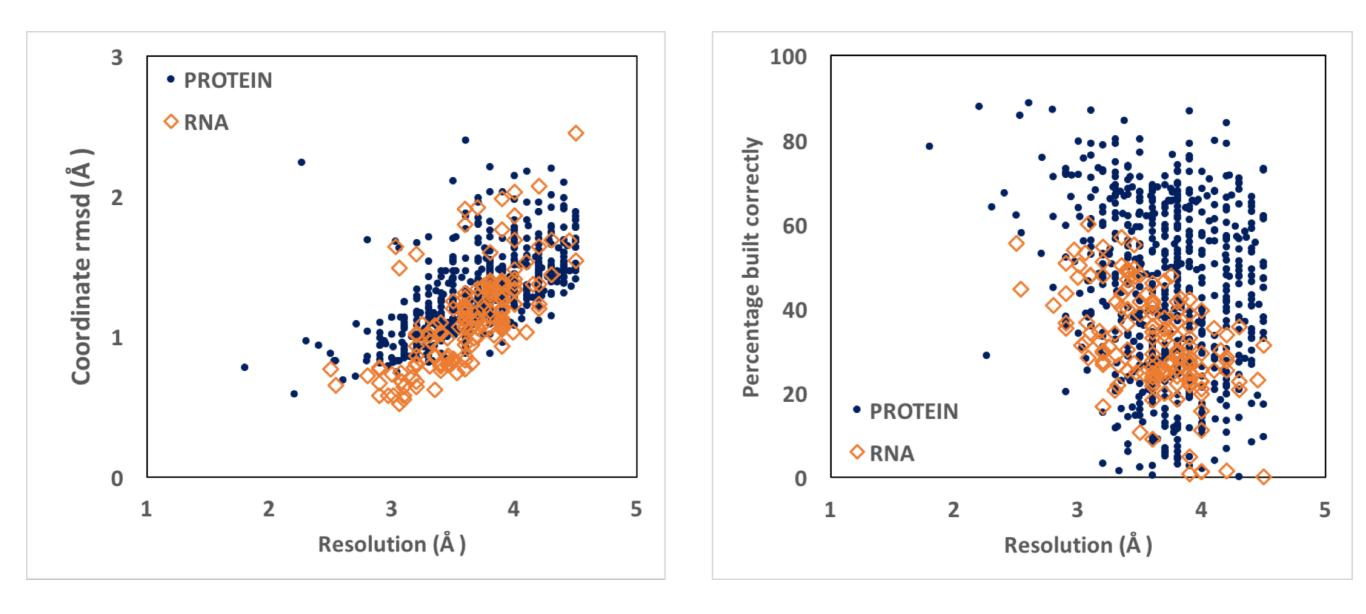
Building at Medium/High Resolution



Proteasome at 2.8 Å (autobuilt model; emd_6287)

Beta-galactosidase at 2.2 Å (autobuilt model; emd_2984)

Autobuilding Performance



Model Building Version 2

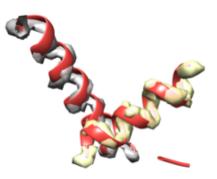
Trace chain the way a person does

Find secondary structure

Find clear regions of density

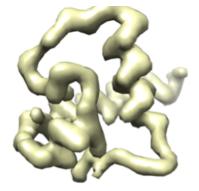
Adjust contour level until a region just connects to another



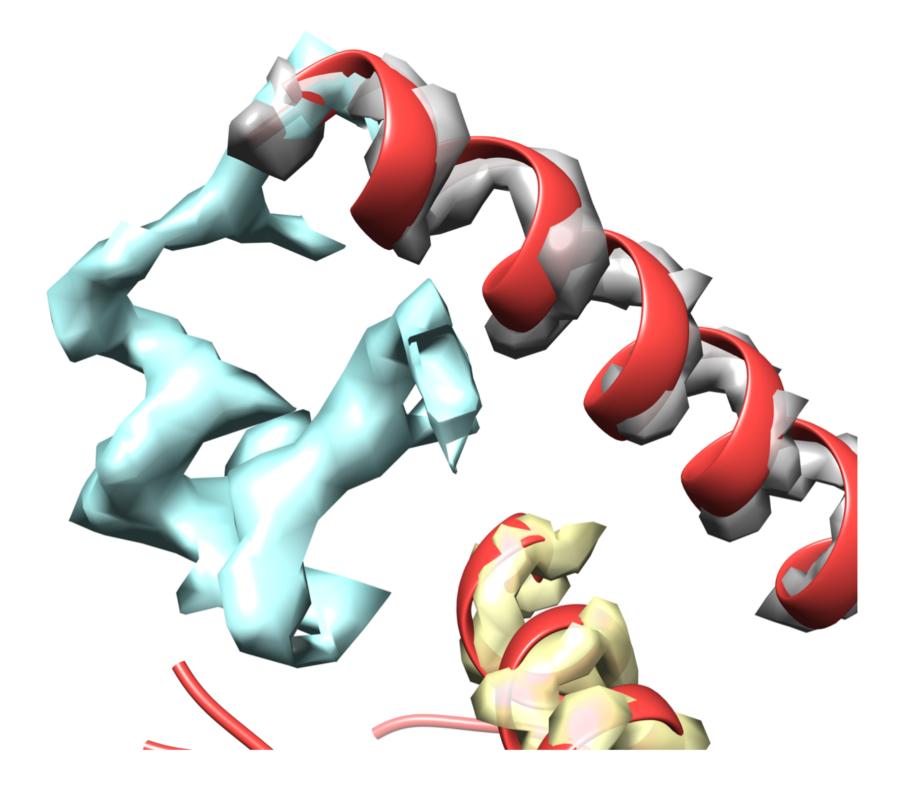




Iterate to build up a connected chain



Model Building Version 2



Finding C_{α} and C_{β} positions

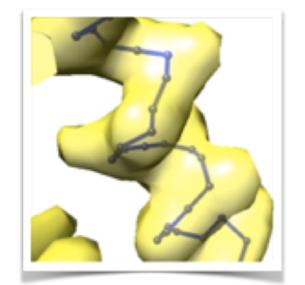
Trace chain path through high density

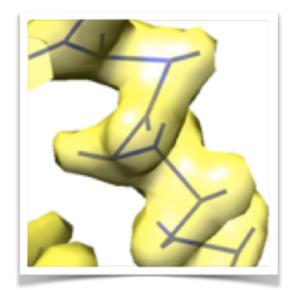
Find C_{β} positions from side-chain density

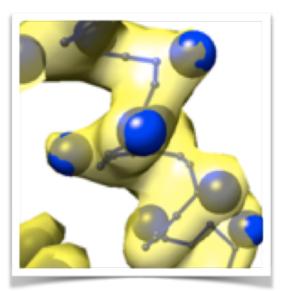
Choose C_{α} positions 3.8 Å apart and next to C_{β} positions

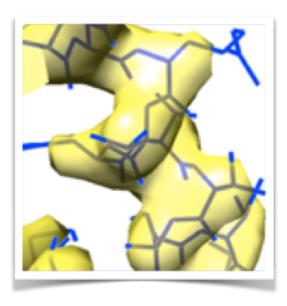
Construct all-atom model with Pulchra* and refine

*Rotkiewicz & Skolnick (2008). J. Comp. Chem. 29, 1460.

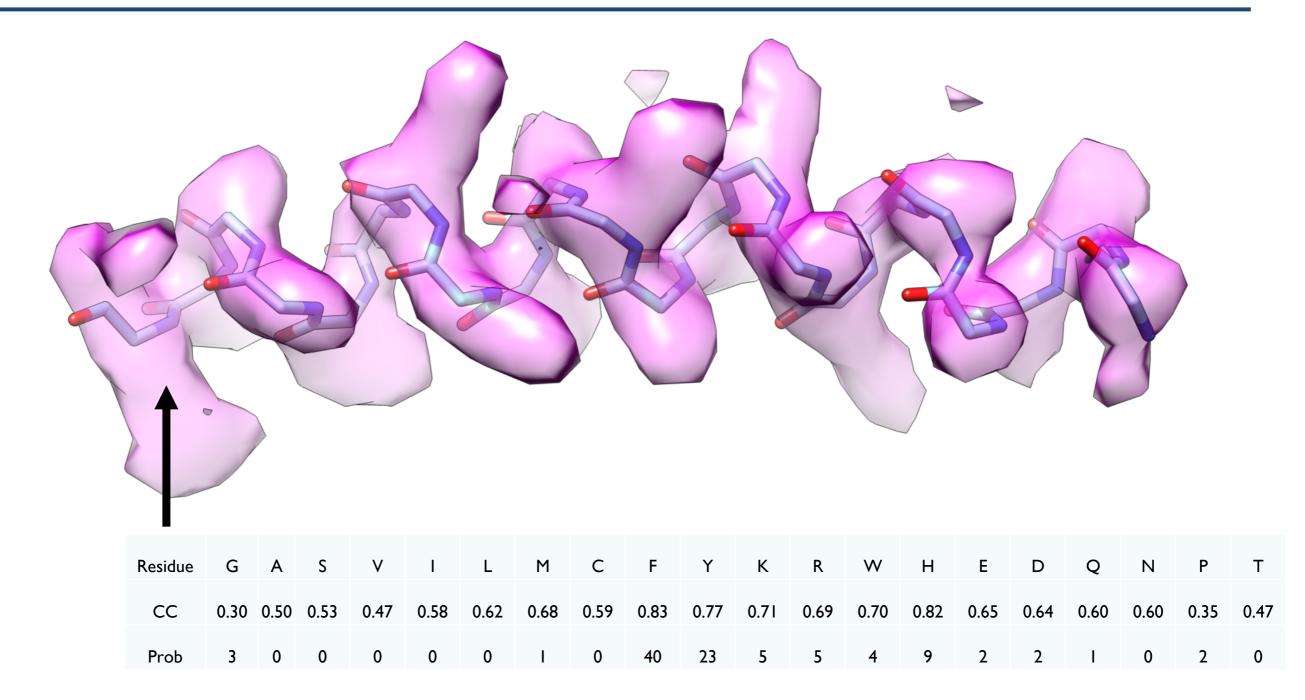






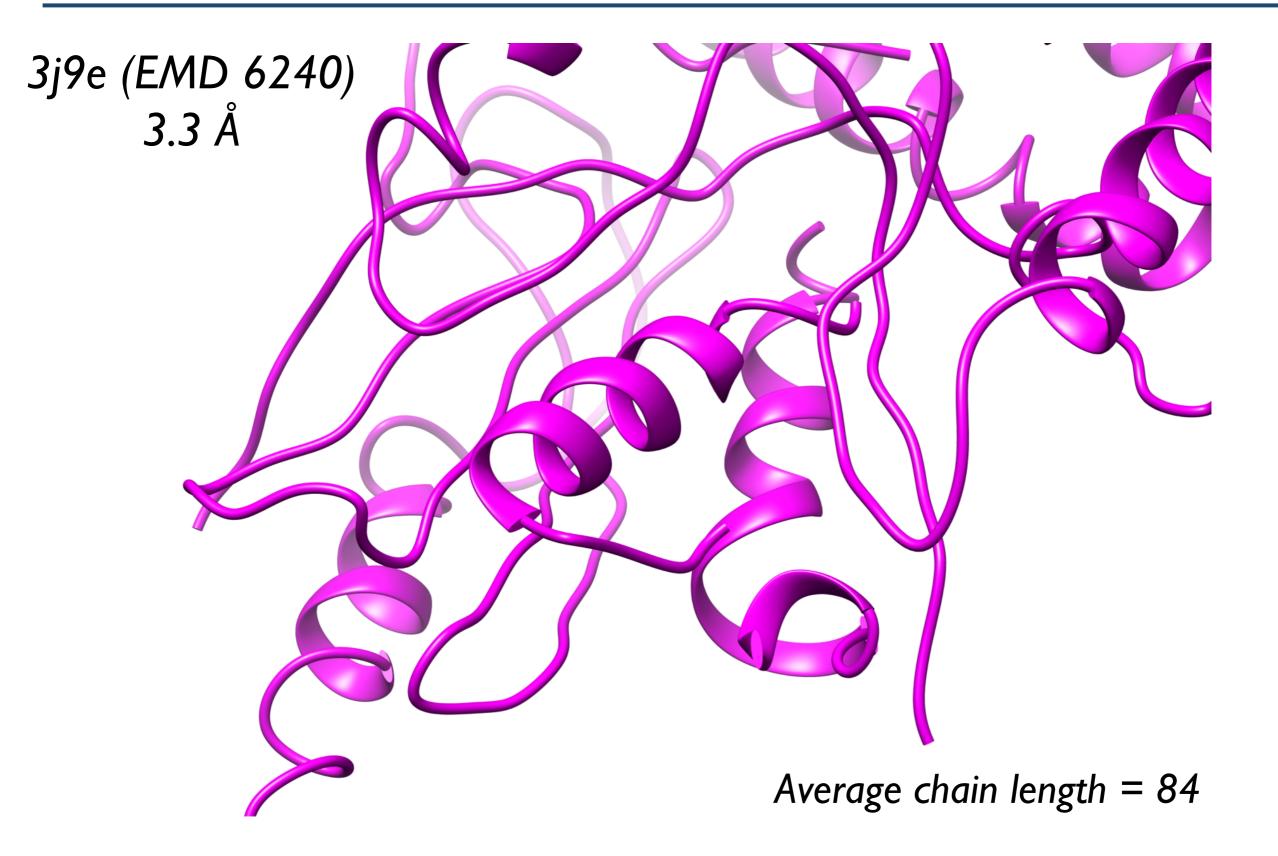


Sequence Assignment

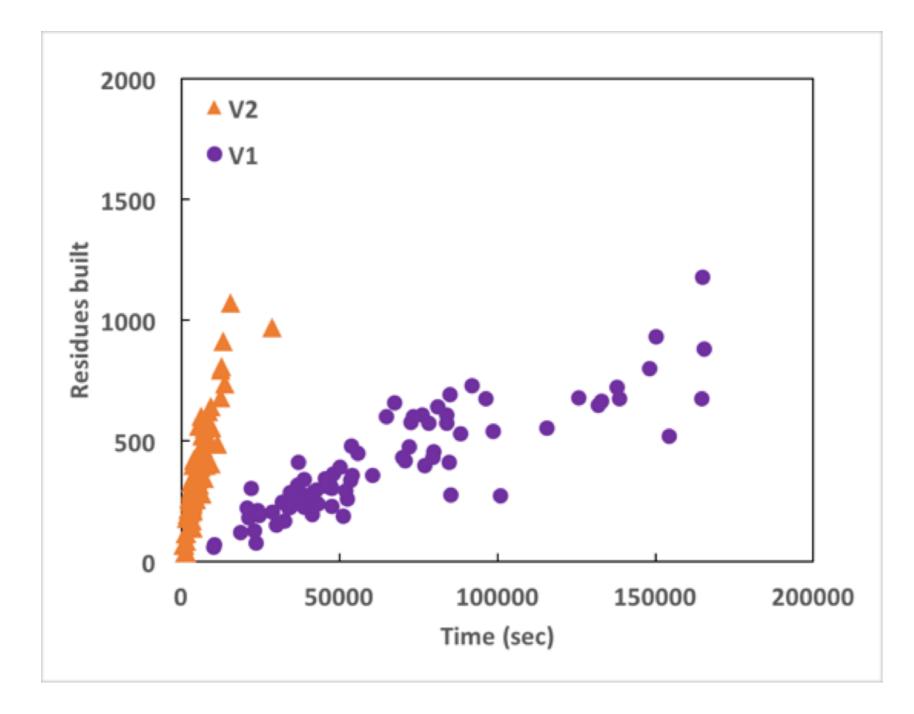


- Determine probability of side chain at each C α
- Align sequence to maximize total probability for the chain

Improved Connectivity

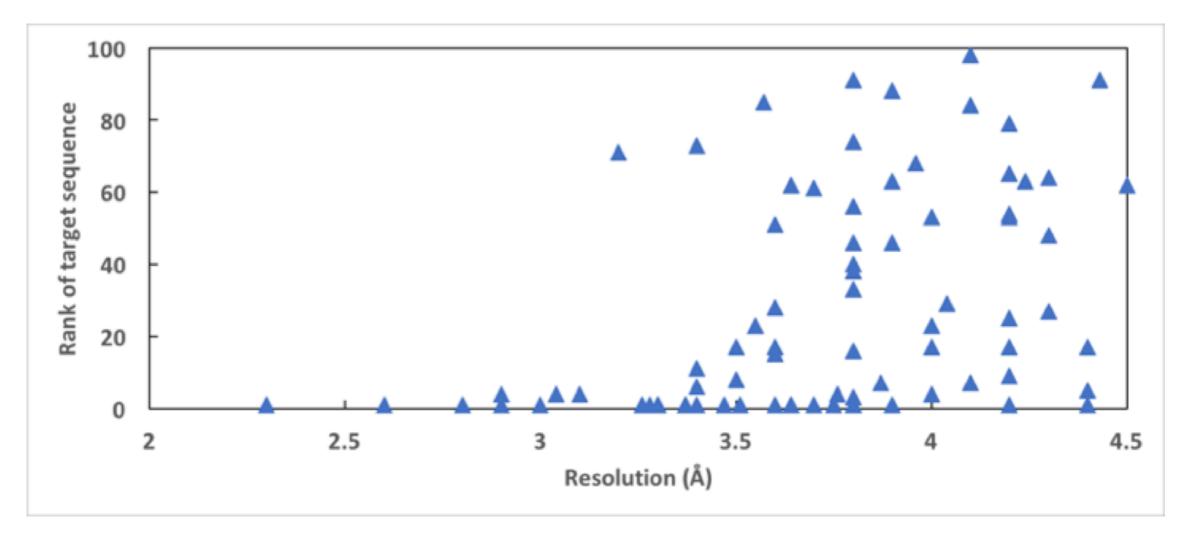


Improved Performance



What's The Molecule?

- Use the highest side chain probabilities to determine a sequence (from the map)
- Search the sequence database to identify the molecule

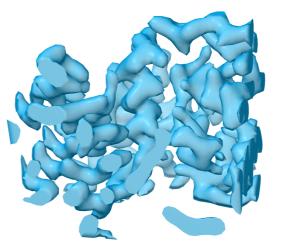


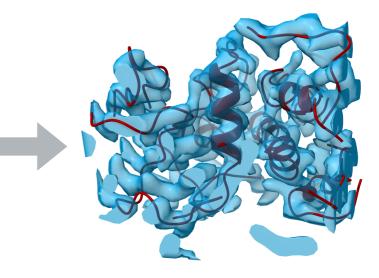
With Xiaorun Li, Chi-min Ho & Hong Zhou, UCLA

Automated model building

Map to Model (GUI) phenix.map_to_model

Map to Model (Project: rotavirus-model-building)	
💥 🥐 🛞 😫 📫	
Preferences Help Run Abort Ask for help	
Configure MapToModel_2	↓ ▷ ×
Automated map interpretation with map_to_model (version 2)	
Inputs: Map file (mrc/ccp4/mtz map coefficients) Resolution limit (typically half-dataset FSC resolution) Sequence file (any standard format or plain text)	
Normal use: Supply a map containing your already-sharpened map. Map_to_model will trace the chain, build and re	fine a model.
Job title :	
Input	
Map file: /Users/dcliebschner/Documents/rotavirus-model-building/rotaviru Browse	
High-resolution limit : 3.0	
Sequence file : /Users/dcliebschner/Documents/rotavirus-model-building/rotaviru Browse	۹.
Starting model (optional) : Browse	Q - +
Starting model (optional) .	<u> </u>
Output	
Output PDB file : map_to_model.pdb	
Output PDB me : mmp_co_mode1.pdb	
Run control	
Thoroughness : medium 📀 🗸 Build a new model	Number of processors : 4
multiprocessing type : multiprocessing 🗘 Queue run command :	Auto-sharpen map
Build cycles : 1	
Symmetry and strategy	
Asymmetric map E Find symmetry Symmetry type (optional) :	
Unique part only - Ignore map limitations Symmetry file (optional) :	Browse
Cryo-EM or X-ray	
Scattering table : electron ᅌ 😑 Is a crystal (optional)	
All parameters	
	Project: rotavirus-model-building
dle	Project. Totavirus-model-building







Map, Sequence, Resolution

Conclusions

Automated model building is possible, but can be improved

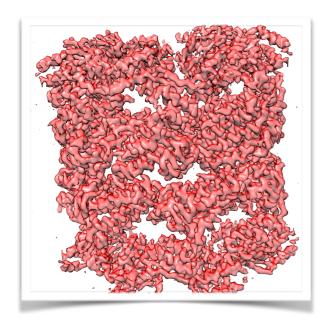
- Include information from secondary structure prediction, evolution etc.
- Combine structure-modeling tools (Rosetta) with Phenix model-building

Many challenges remain:

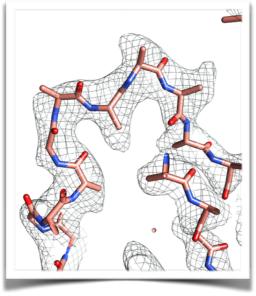
- Reliably accounting for uncertainty in magnification
- Local variation in resolution leads to uncertainties in interpretation

Title

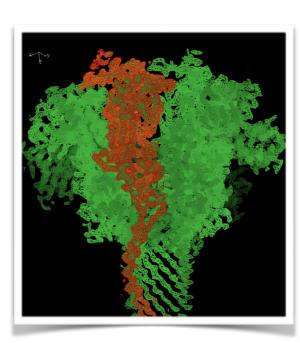
New Tools for Cryo-EM in Phenix



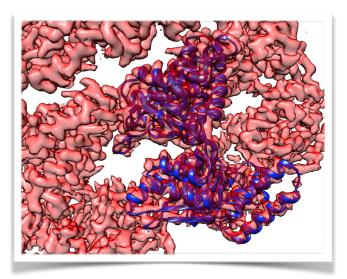
Symmetry from a map



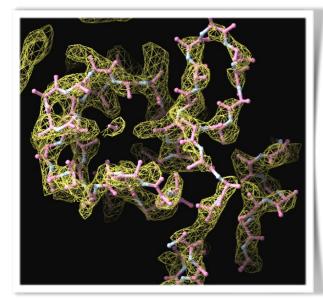
Automated map sharpening



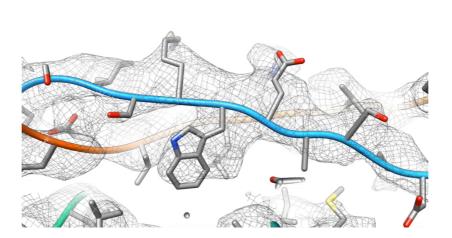
Map segmentation



Rigid model docking



Automated model building



Real space refinement

		Com	prehensive vali	dation (CryoE	M) (Project: rea	a-space-	refine-6crz)				
× ?	5) 💽) 🦿	2							
references Help	Ru	n Abo	rt Ask fo	or help							
Input/Output Valida	tionCryo	EM_7								4	1
Summary Model N	fodel vs. I	Data	Data							4	1
Files											
Model: /Users	/PDAdar	ns/Doc	uments/rea-sp	ace-refine-6ci	z/model.pdb						
Map: /Users	/PDAdar	ns/Doc	uments/rea-sp	ace-refine-6ci	z/map.ccp4						
- Open in (Coot							Exp	ort Table 1	٦	
tt⊂ open in v											
Clicking on a row w	Jring	, ap a p	and man more	detaned mon							
MolProbity					Ramachandra	n					
MolProbity score		1.72			Outliers (%)	0.00	(Goal: < 0.2%)			
Clash score		5.44			Allowed (%)	6.45					
Rotamer outliers (%)		0.00	(Goal: < 1%)		Favored (%)	93.55	(Goal: > 98%)				
CB outliers		0	(Goal: 0)								
CaBLAM					Peptide Plane						
Outliers (%)	3.88	(Goal	: <= 1%)		cis-proline (%))	0.00				
Disfavored (%)	8.96		: <= 5%)		twisted proline		0.00				
Cα outliers (%) 1.19		(Goal	: <= 0.5%)		cis-general (%		0.00				
					twisted genera	al (%)	0.00				
Geometry Restrain	ts										
Bond	Angle			Dihedral							
	_										

Model and map validation

Schedule

8:30 AM: Introduction to Phenix and overview of tools for cryo-EM
8:45 AM: Map tools (density modification, sharpening, map symmetry)
9:30 AM: Break and computer setup
10:00 AM Model building (docking, ab initio building)
10:30 AM: Atomic model refinement
11:30 AM: Validation (map, model, model to map fit)

12:00: Lunch

1:00 PM: Introduction to the GUI and setup
1:15 PM Map improvement and model building
(DM, sharpening, symmetry, segmentation + automated model building)
3:00 PM: Break

3:30 PM: Refinement and validation

4:30 PM: User questions, more select tutorials, discussion, etc

5:30 PM: Finish