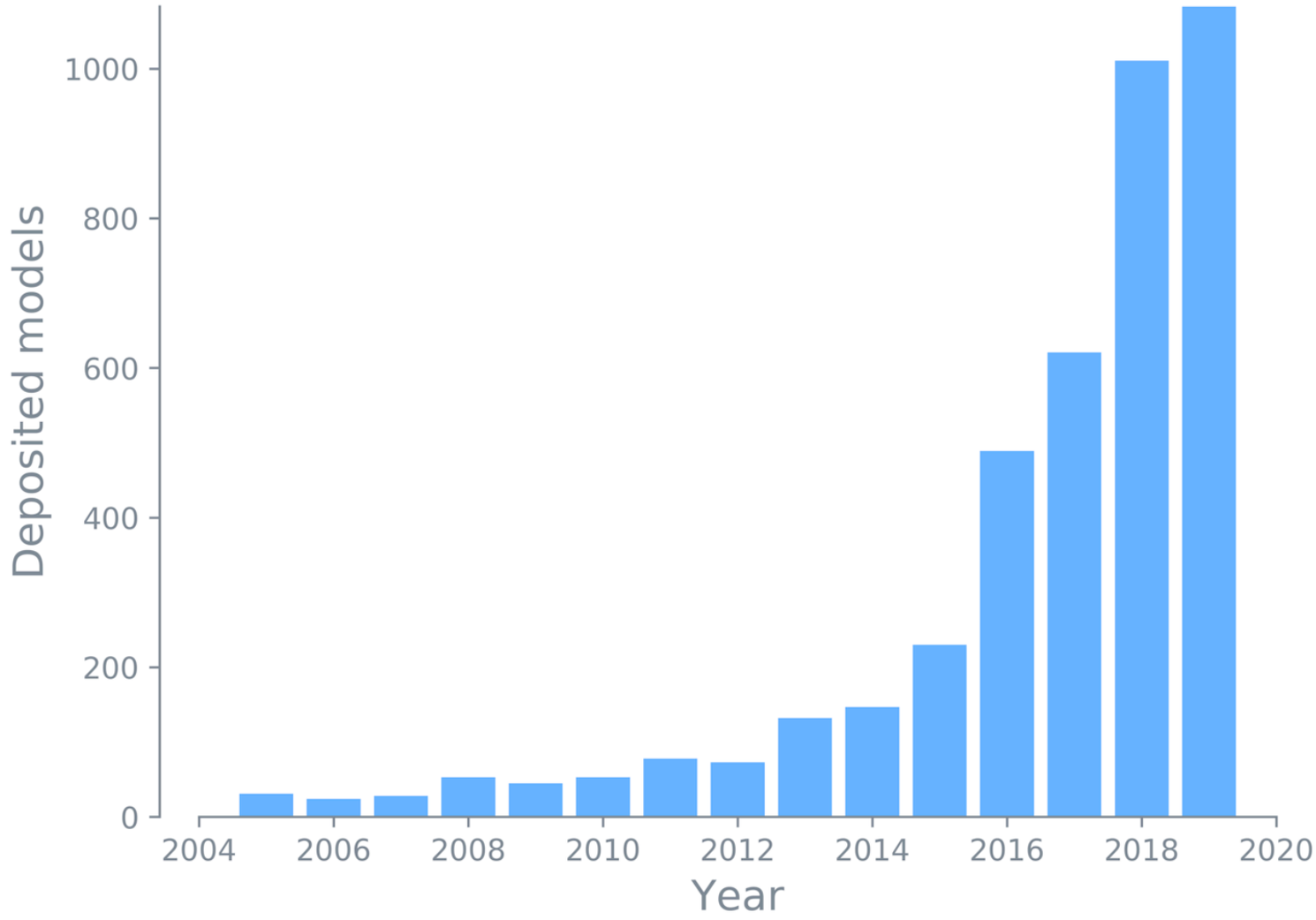
CU Anschutz Medical School, January 2020

Model building (cryo-EM)

Dorothee Liebschner

Lawrence Berkeley Laboratory

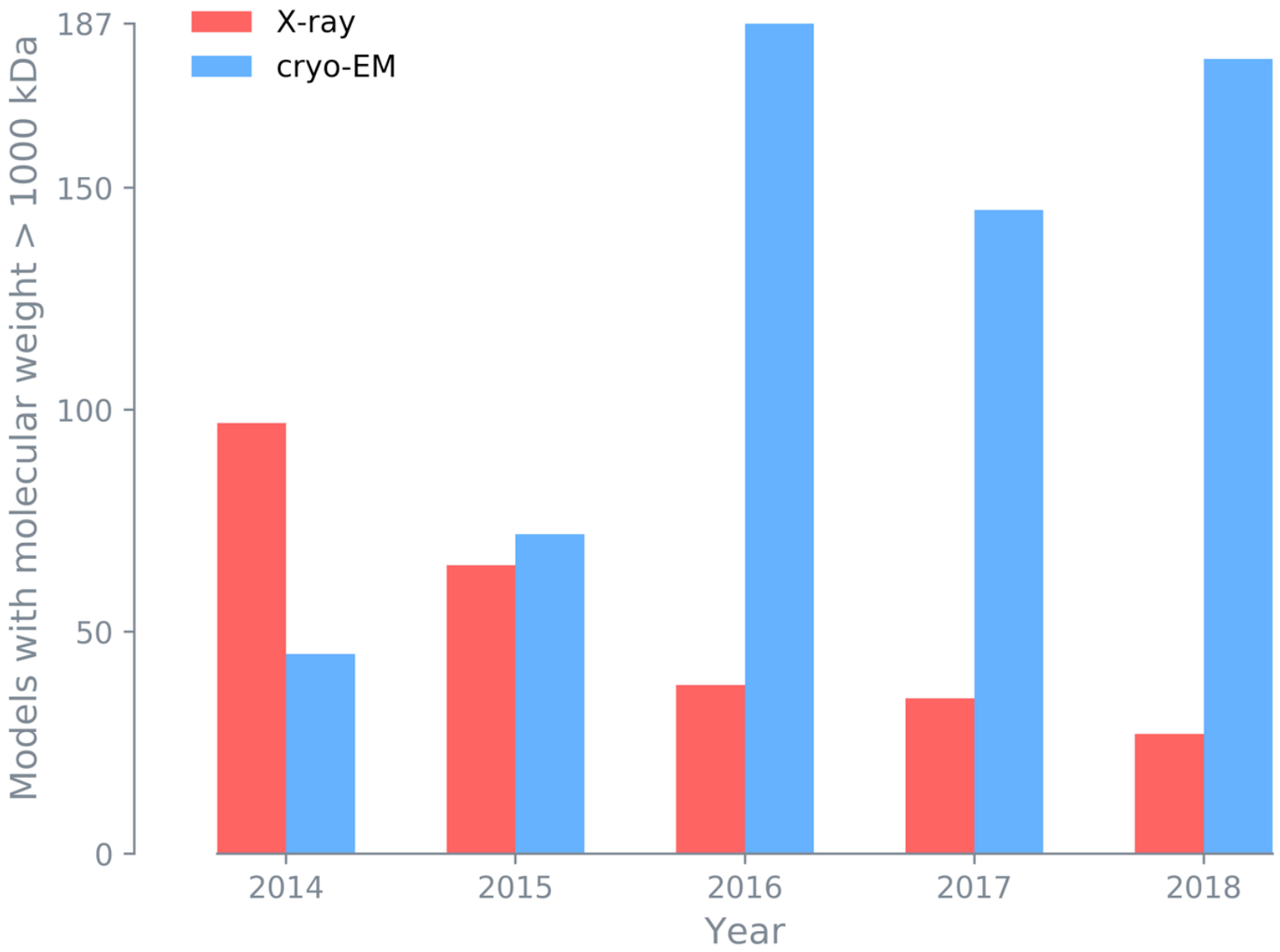
Cryo-EM models in the PDB



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

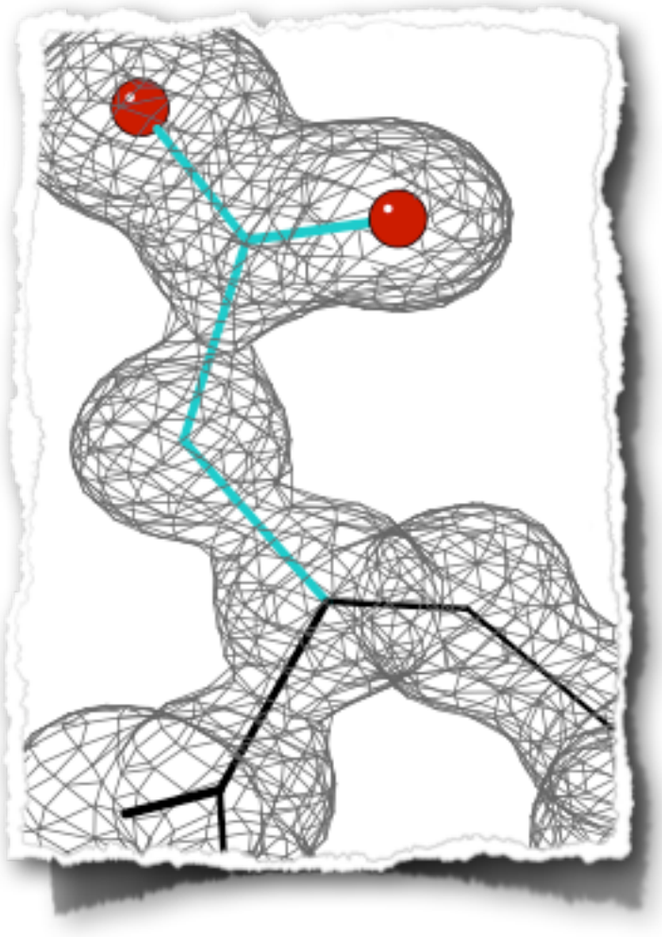
Cryo-EM models in the PDB

Cryo-EM is used to determine large structures.



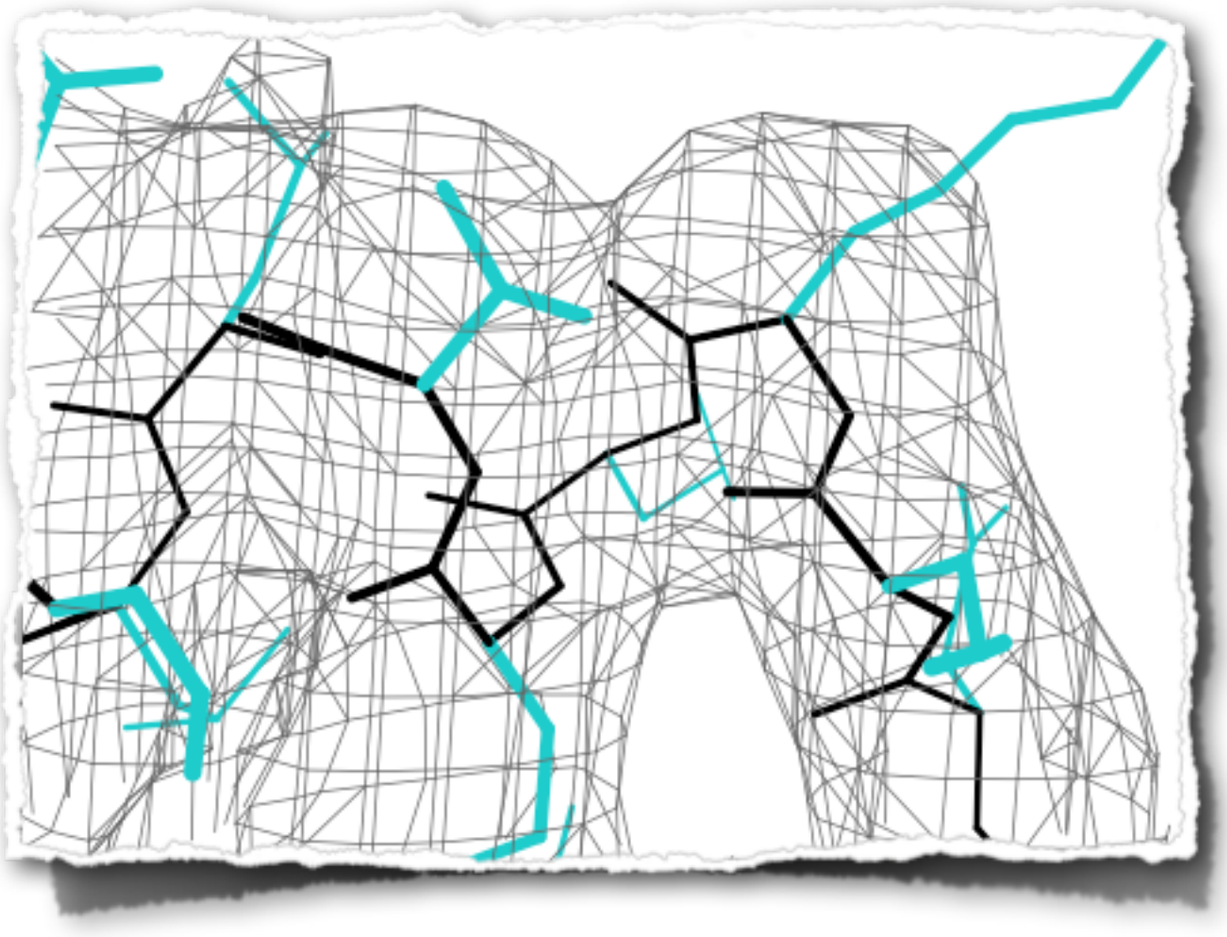
EM maps have low resolution

1.00Å resolution



2gkg

3.80Å resolution



3k7a

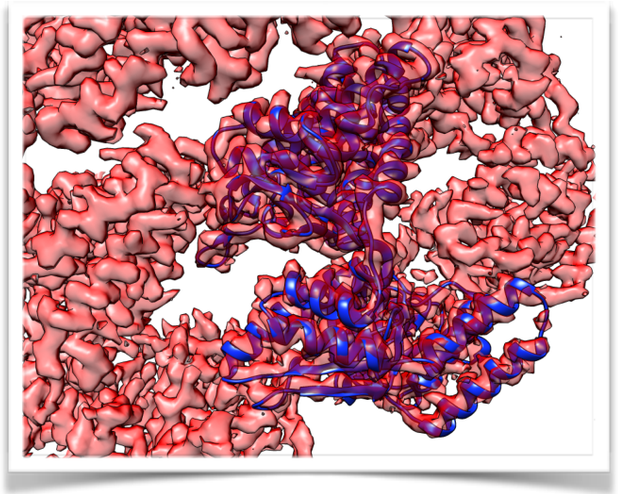
Challenges:

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

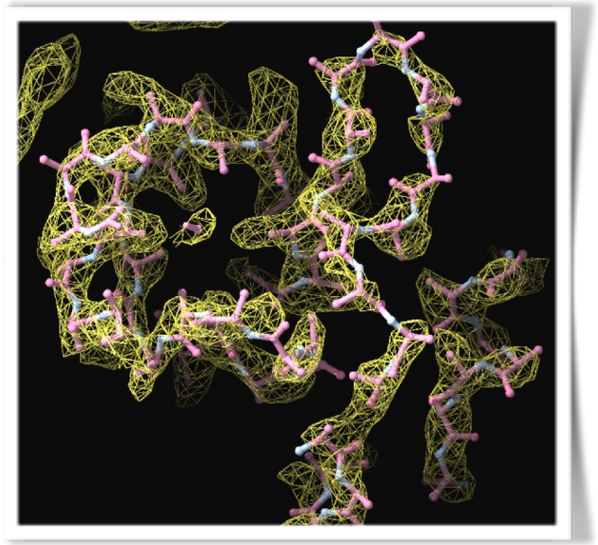
Model building tools in *Phenix*

Goal:

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



Rigid model docking

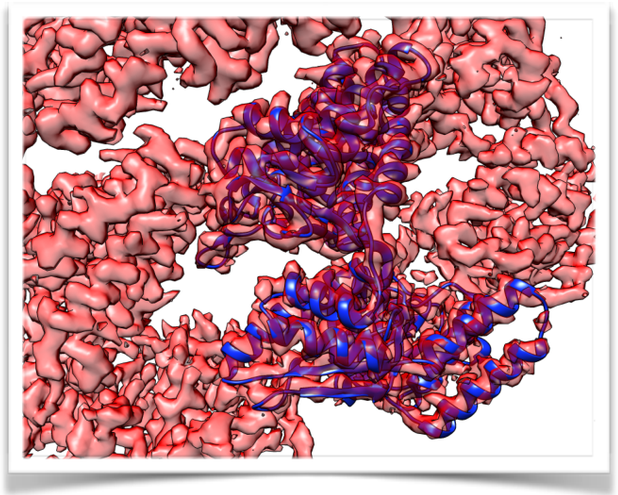


Automated model building

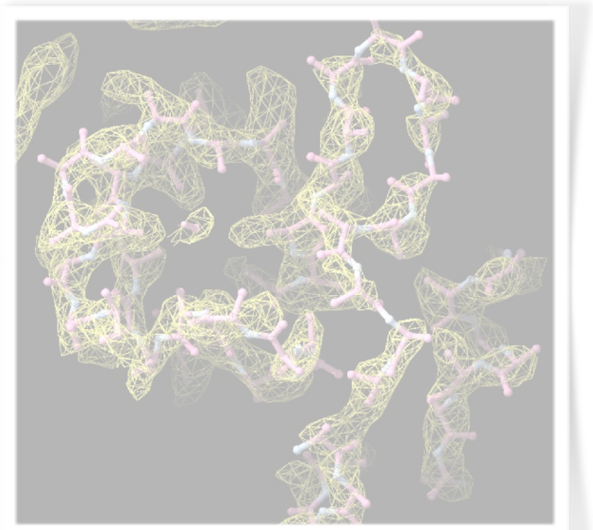
Model building tools in *Phenix*

1. Docking

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

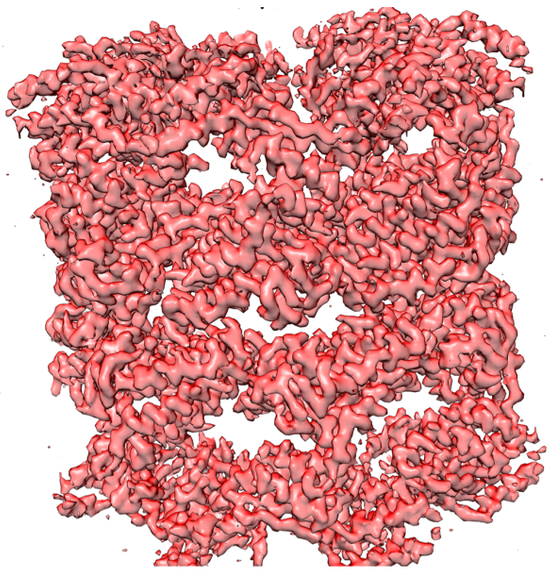


Rigid model docking



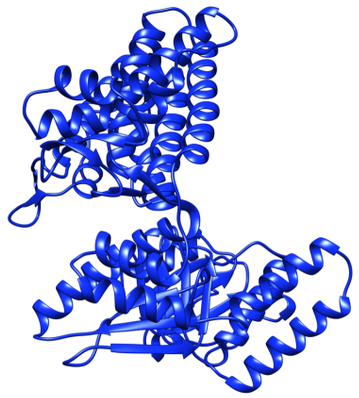
Automated model building

Rigid model docking

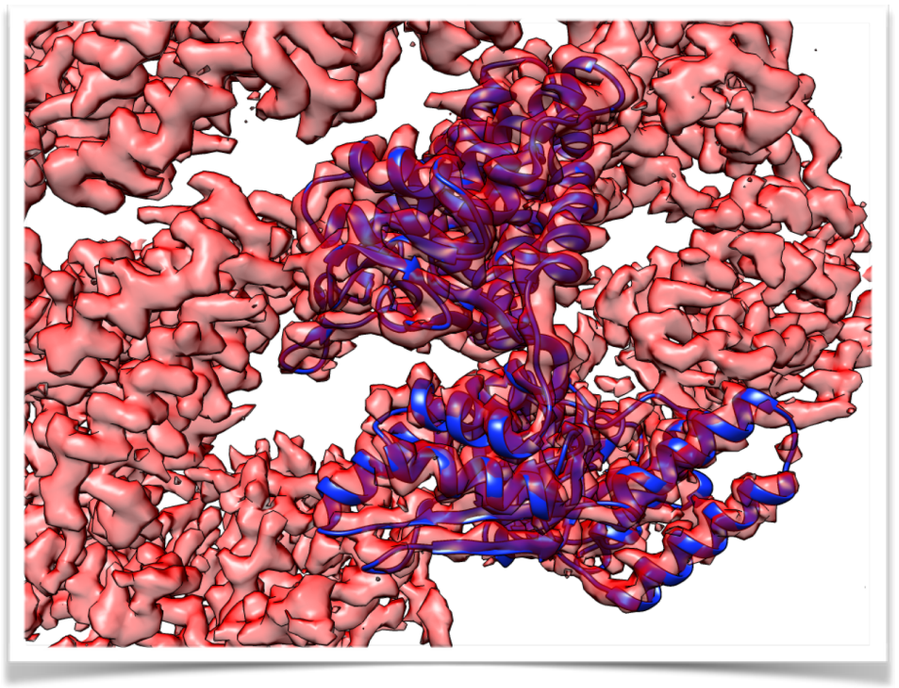


EMD8750

+



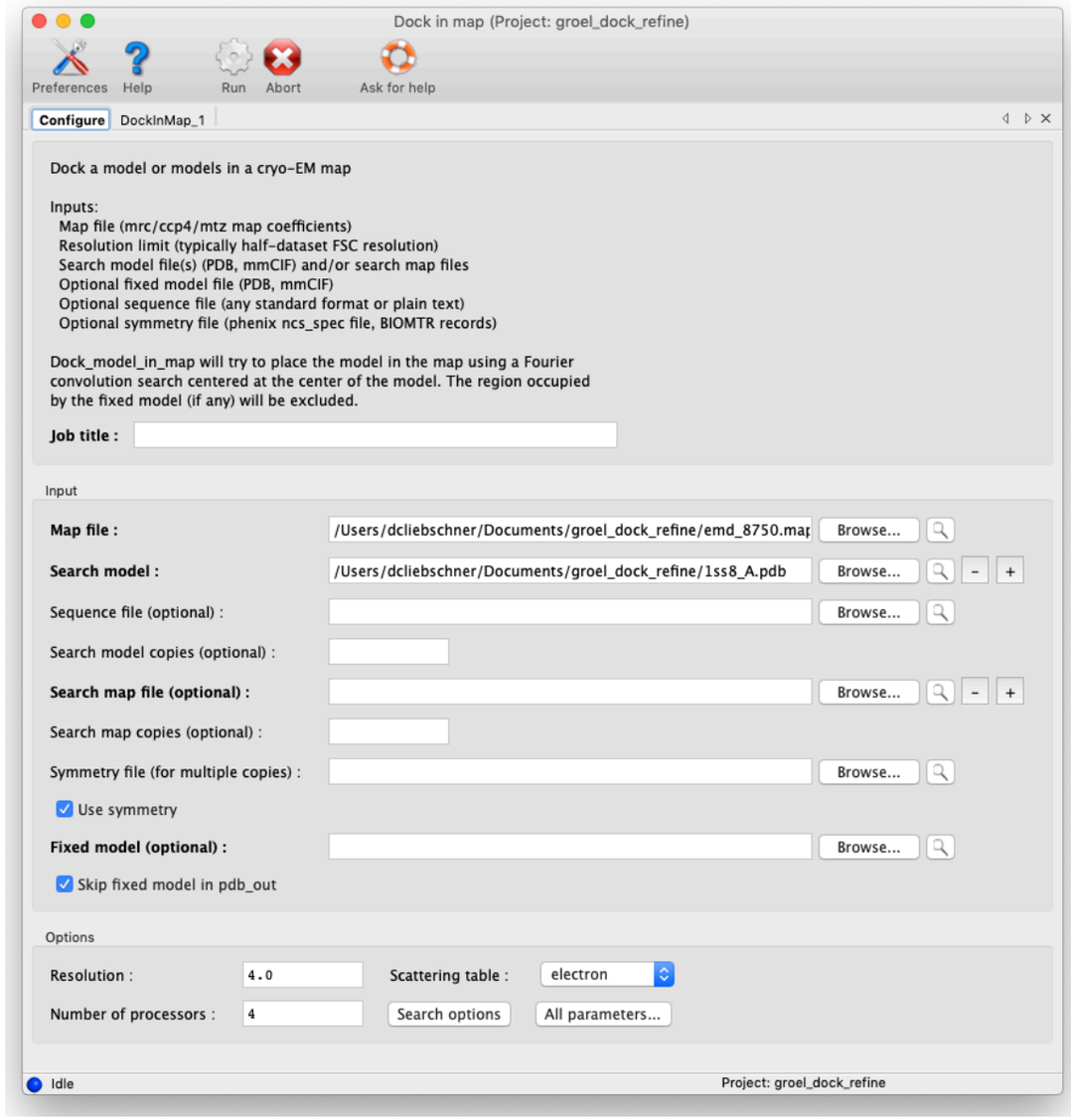
1SS8 chain A

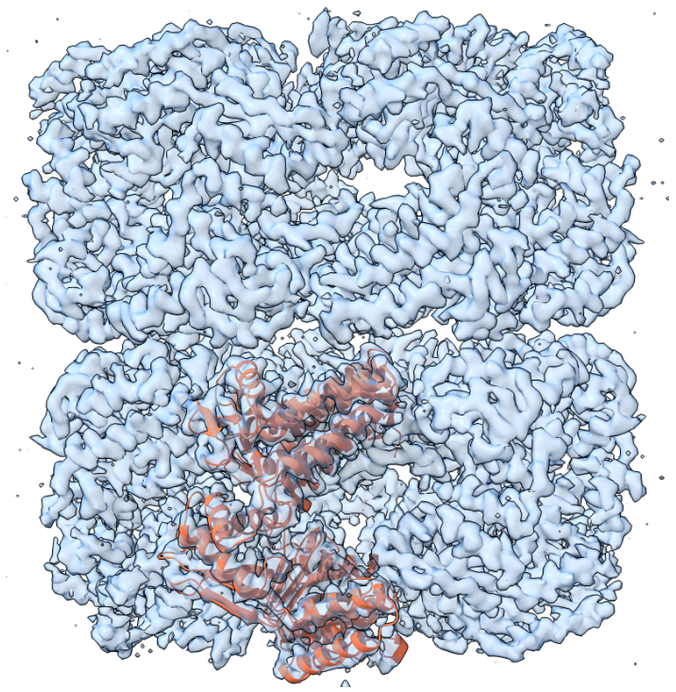


- Systematic cross correlation search of rotations and translations
- Performed in reciprocal space using FFT (very fast)
- Rigid body optimization of position

Rigid model docking

Dock in map (GUI)
phenix.dock_in_map





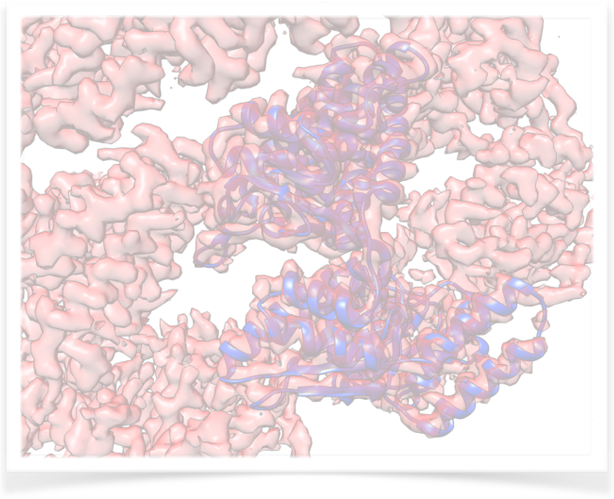
Model docked in map

Map, model, resolution

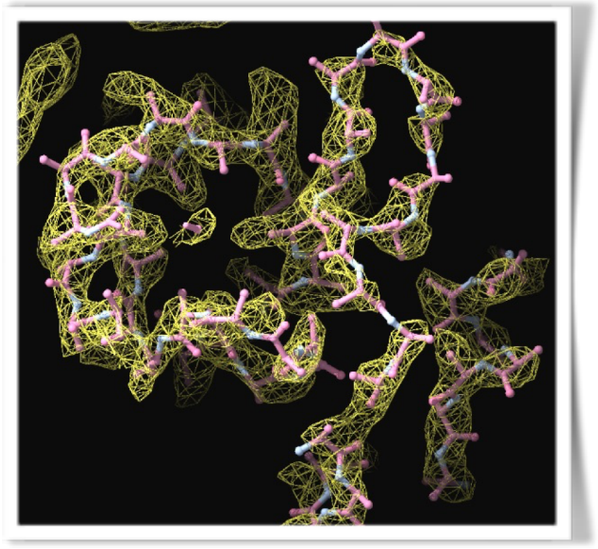
Model building tools in *Phenix*

2. Automated model building

Build a model into a map from scratch.



Rigid model docking



Automated model building

Automated Model Building Procedure

Determine optimal sharpening of the map

↓

Cut out asymmetric unit of the map

↓

Trace chain and build model

↓

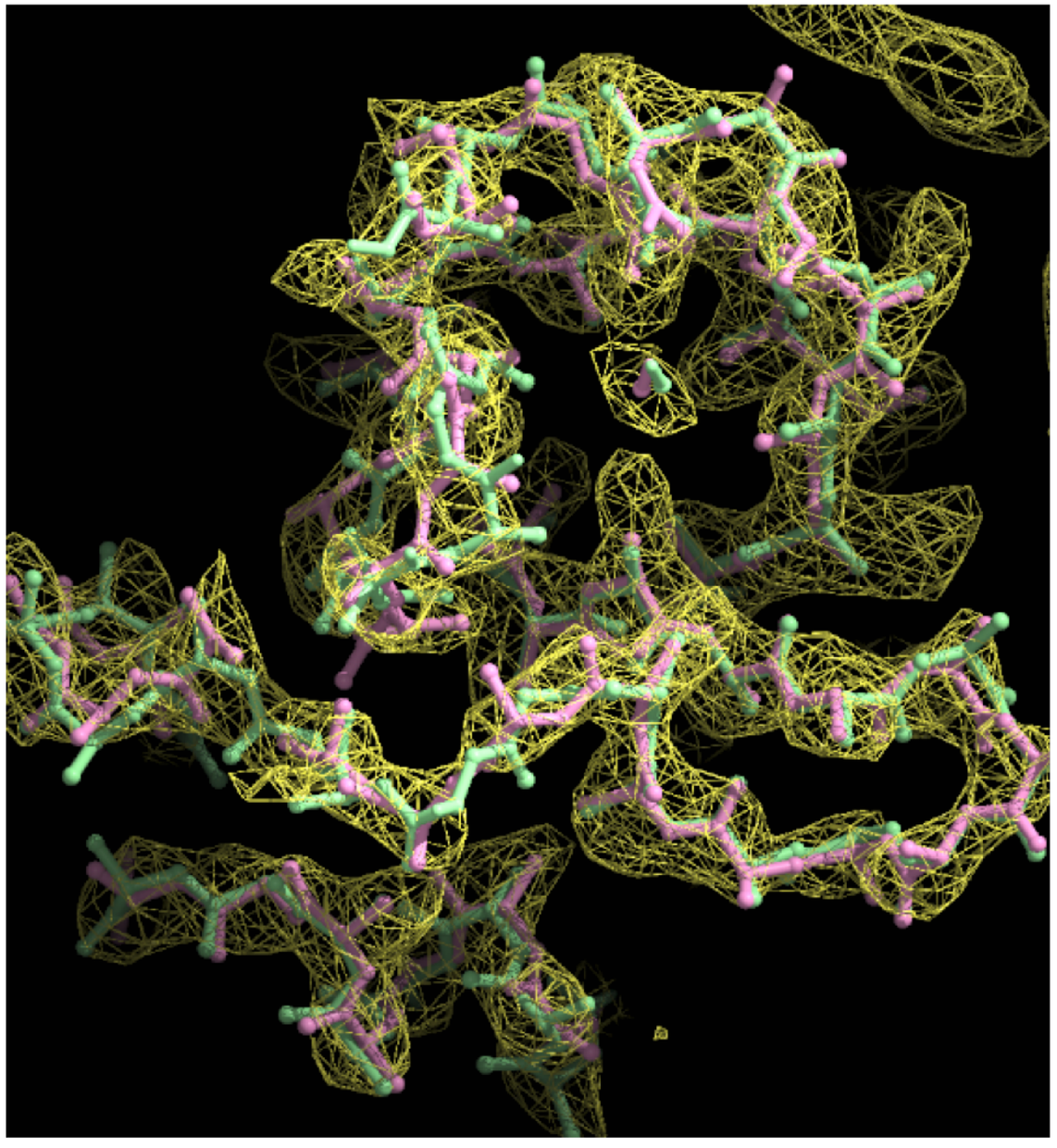
Idealize secondary structure and refine

↓

Assemble and refine (protein/RNA/DNA)

↓

Apply molecular symmetry and re-refine



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)

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

Automated Map Sharpening

Determine optimal sharpening of the map

Cut out asymmetric unit of the map

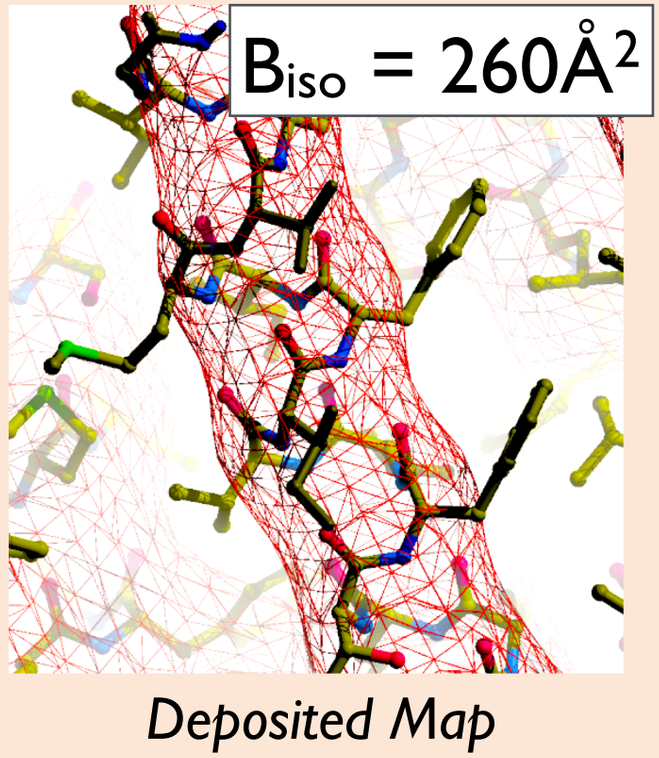
Trace chain and build model

Idealize secondary structure and refine

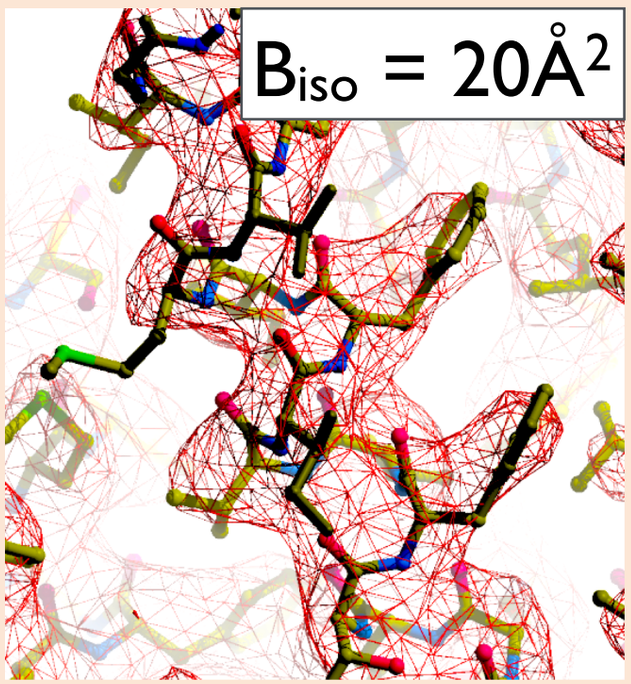
Assemble and refine (protein/RNA/DNA)

Apply molecular symmetry and re-refine

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



Deposited Map



Autosharpened Map

Automated Segmentation

Determine optimal sharpening of the map

Cut out asymmetric unit of the map

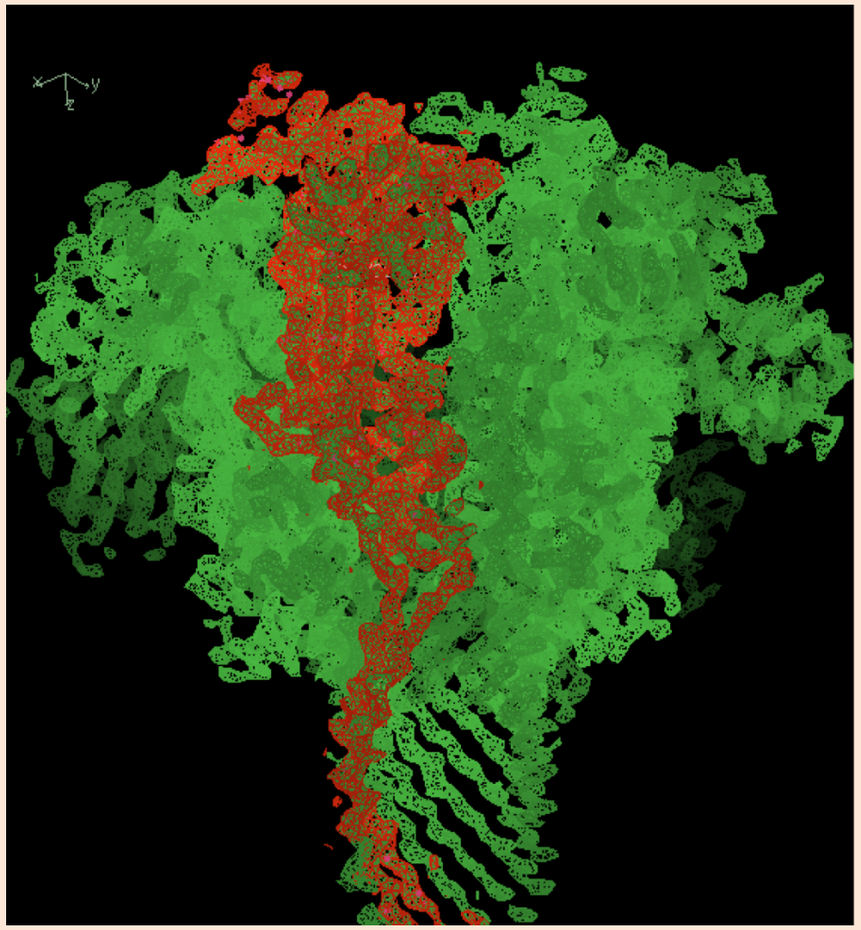
Trace chain and build model

Idealize secondary structure and refine

Assemble and refine (protein/RNA/DNA)

Apply molecular symmetry and re-refine

- 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 cryo-microscopy maps. *Nature Methods* (2018), 15, 905-8

Automated Model Building Procedure

Determine optimal sharpening of the map

↓

Cut out asymmetric unit of the map

↓

Trace chain and build model

↓

Idealize secondary structure and refine

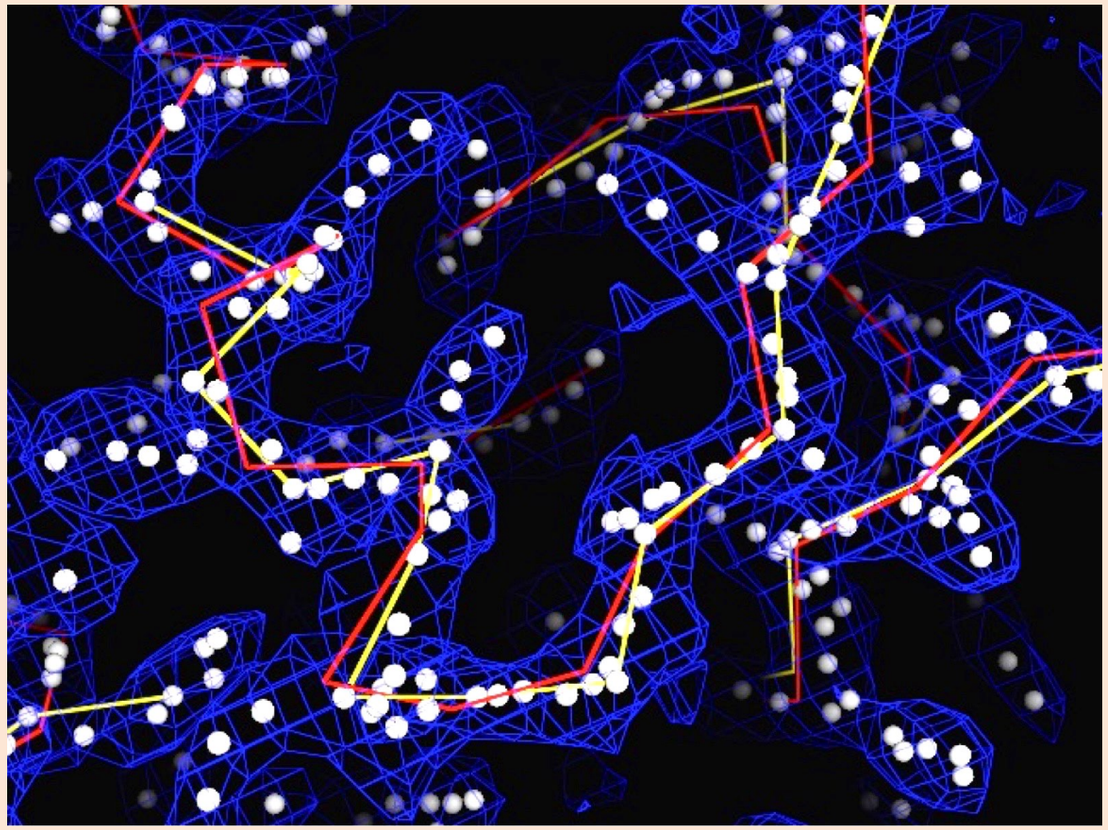
↓

Assemble and refine (protein/RNA/DNA)

↓

Apply molecular symmetry and re-refine

- 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 cryo-microscopy maps. *Nature Methods* (2018), 15, 905-8

Automated Model Building Procedure

Determine optimal sharpening of the map

Cut out asymmetric unit of the map

Trace chain and build model

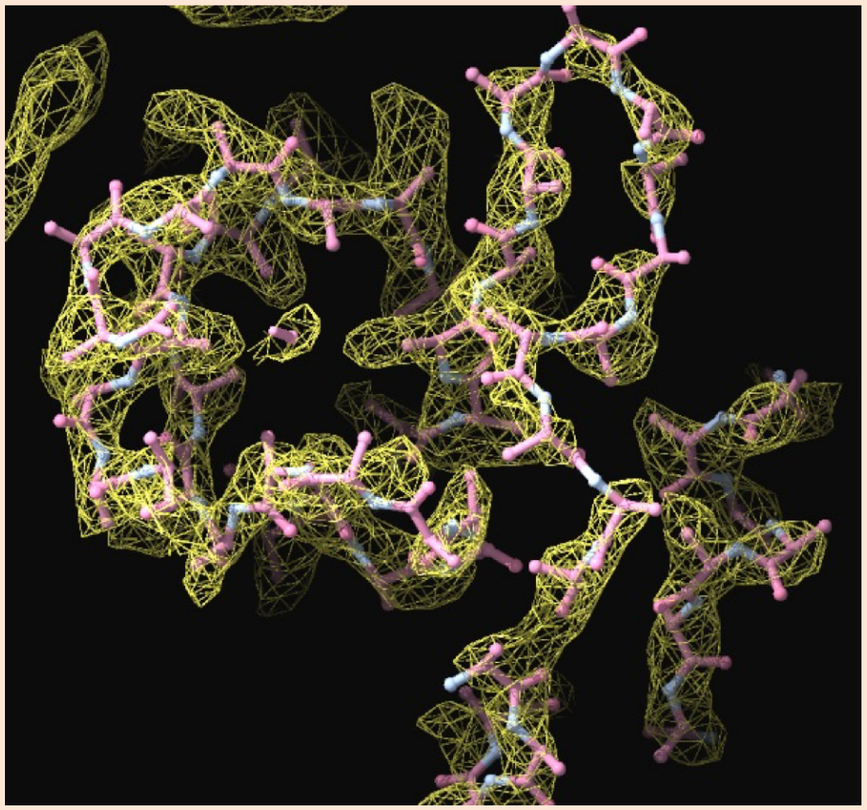
Idealize secondary structure and refine

Assemble and refine (protein/RNA/DNA)

Apply molecular symmetry and re-refine

Terwilliger et al. A fully automatic method yielding initial models from high-resolution electron cryo-microscopy 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

Automated Model Building Procedure

Determine optimal sharpening of the map

↓

Cut out asymmetric unit of the map

↓

Trace chain and build model

↓

Idealize secondary structure and refine

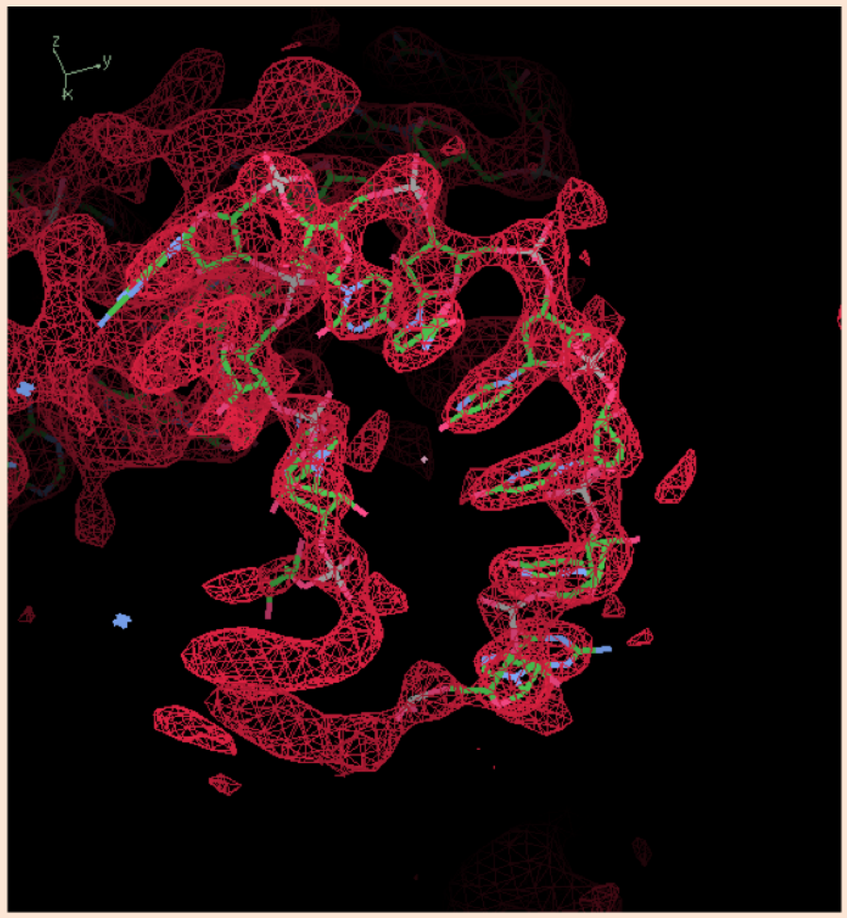
↓

Assemble and refine (protein/RNA/DNA)

↓

Apply molecular symmetry and re-refine

- 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 cryo-microscopy maps. *Nature Methods* (2018), 15, 905-8

Automated Model Building Procedure

Determine optimal sharpening of the map

↓

Cut out asymmetric unit of the map

↓

Trace chain and build model

↓

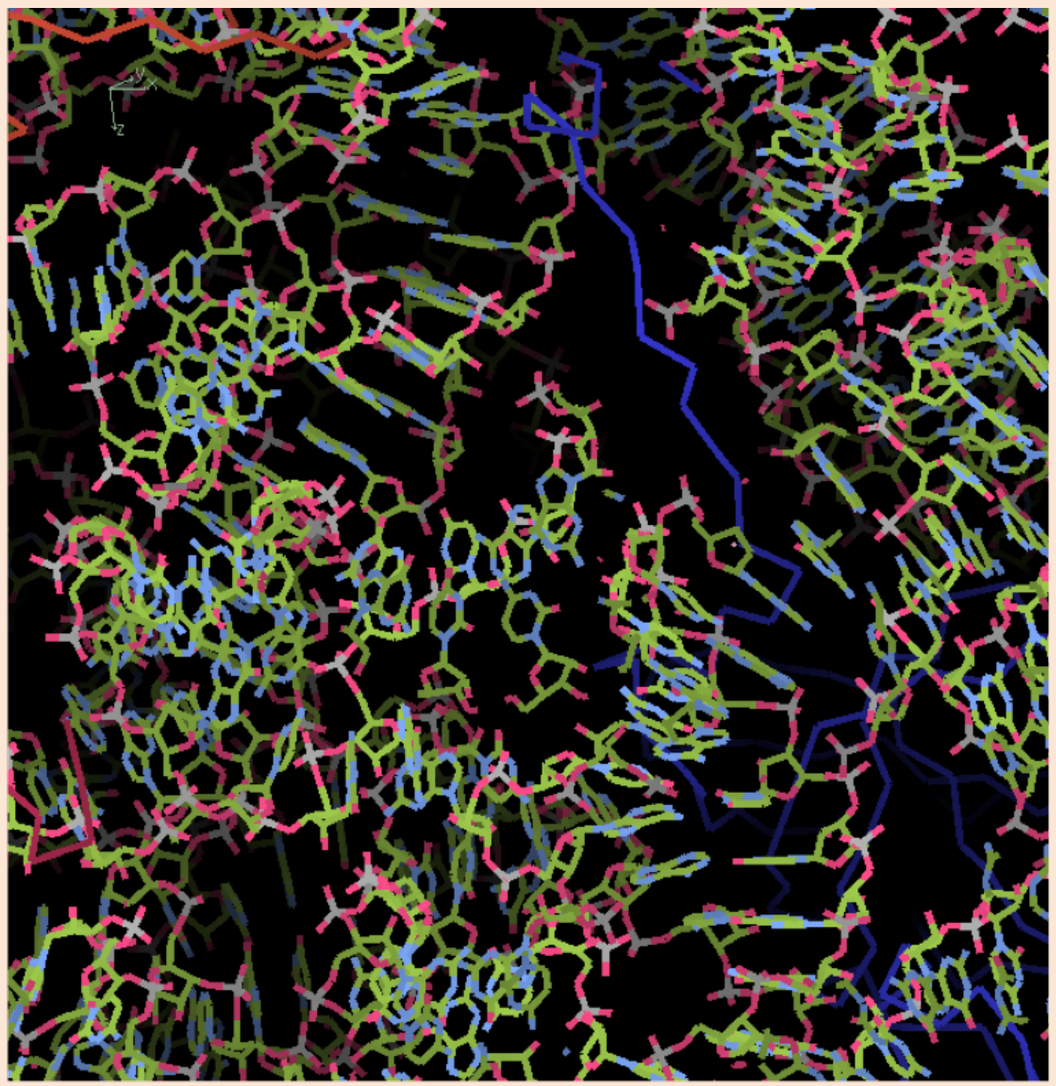
Idealize secondary structure and refine

↓

Assemble and refine (protein/RNA/DNA)

↓

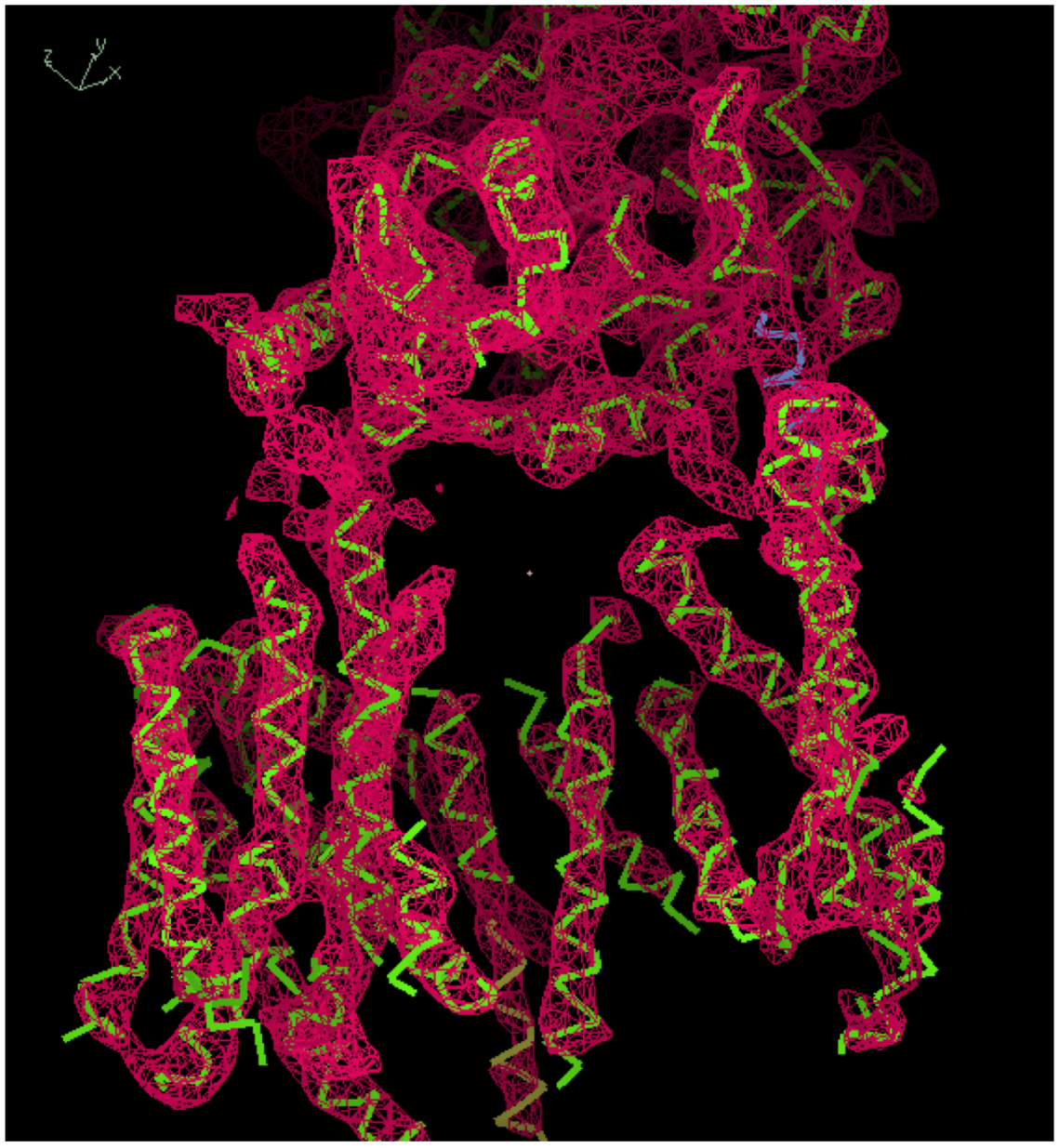
Apply molecular symmetry and re-refine



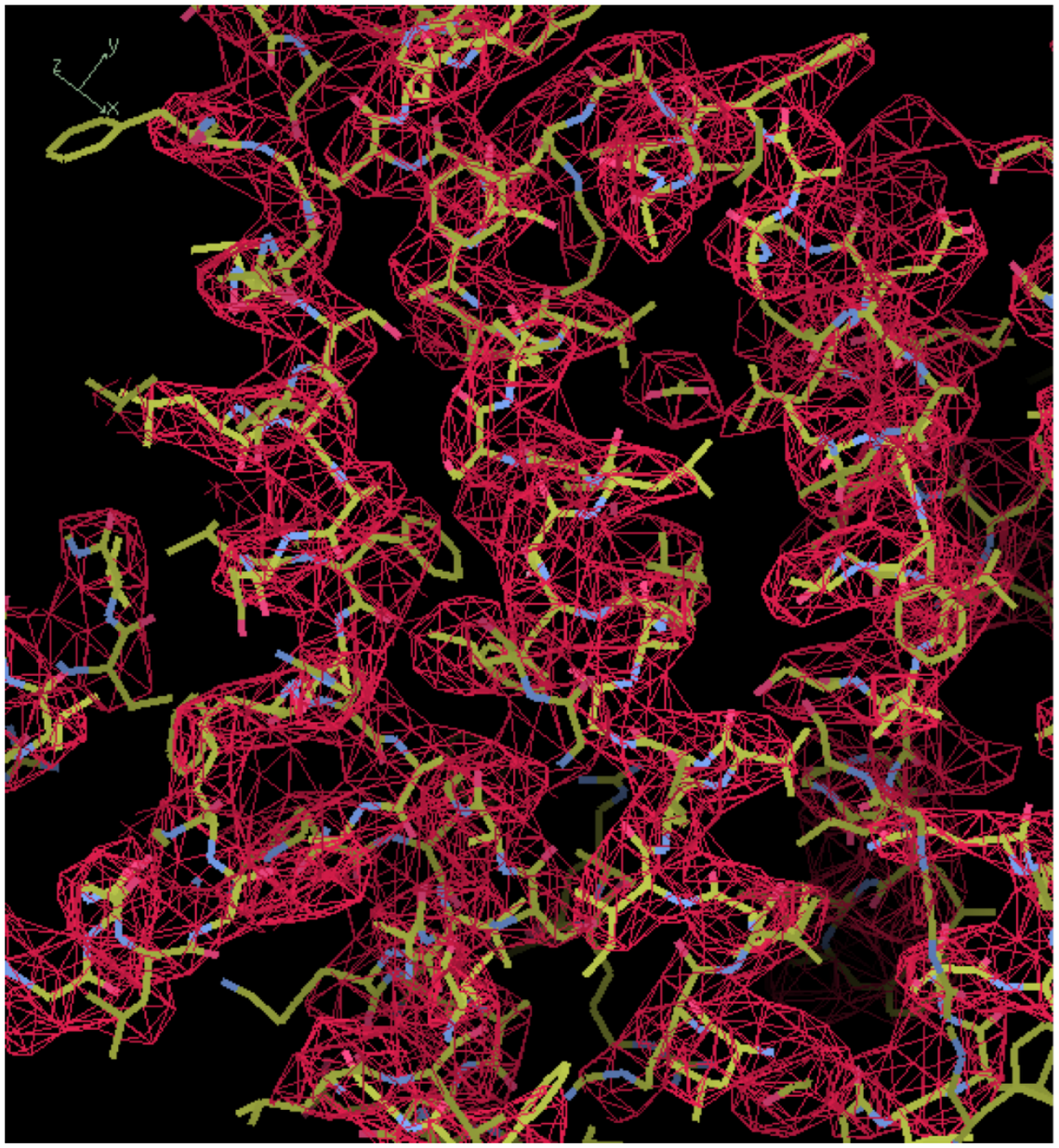
30S Ribosome (1j5e, 2.9 Å)

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

Building at Low Resolution

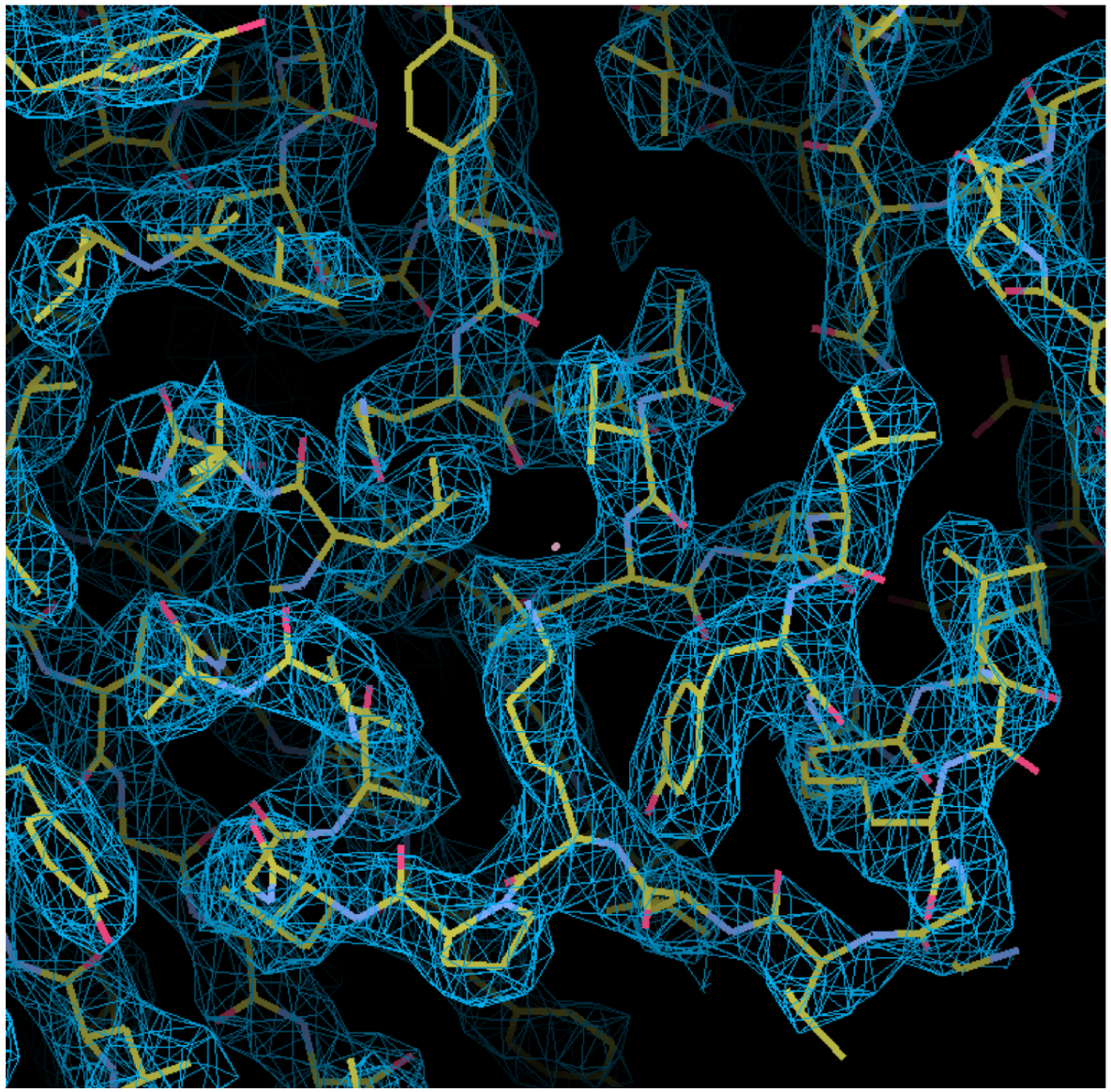


Gamma-secretase at 4.5 Å
(autobuilt model; emd_2677)

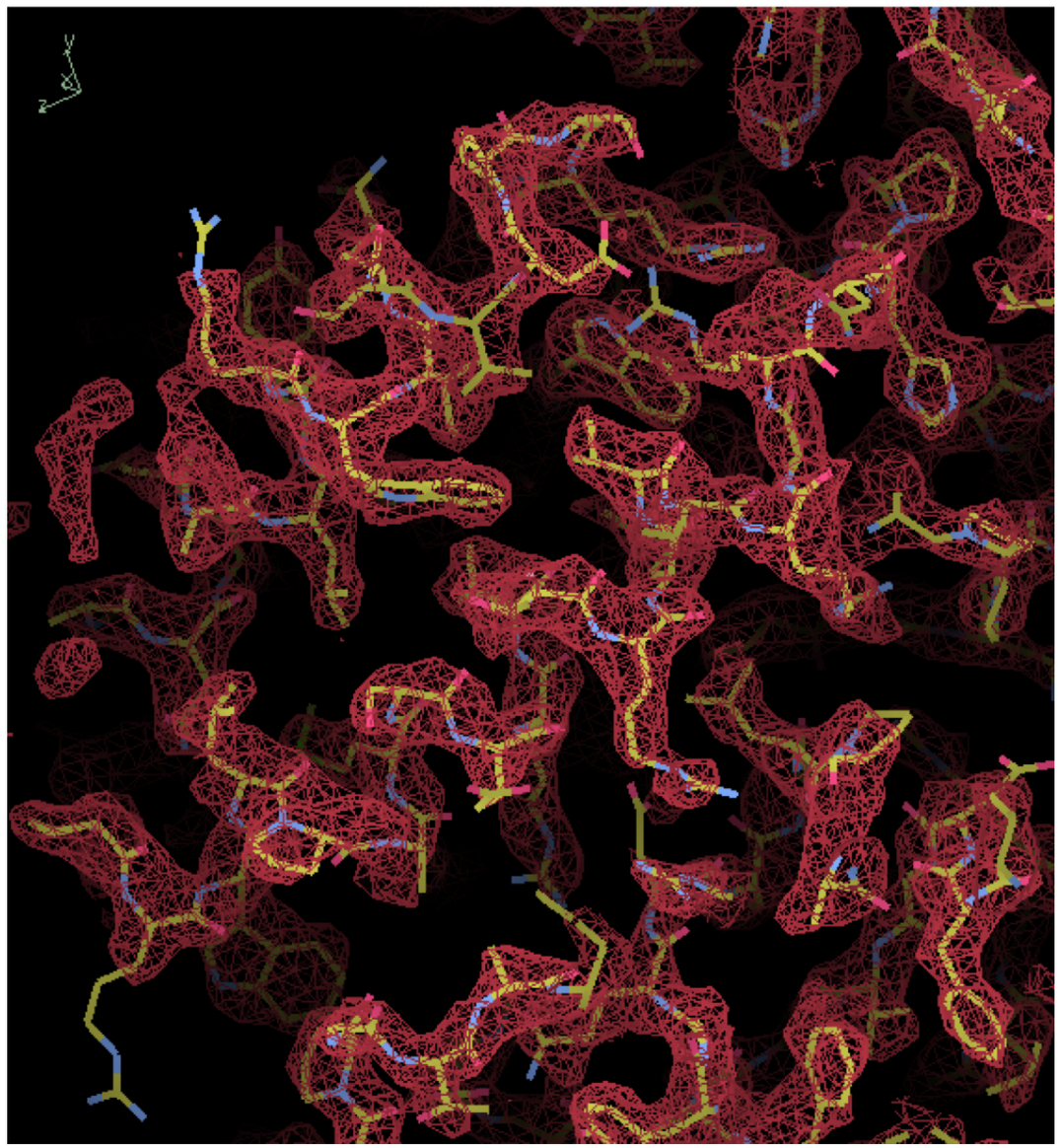


Gamma-secretase structure at 3.4 Å
(autobuilt model; emd_3061)

Building at Medium/High Resolution



Proteasome at 2.8 Å
(autobuilt model; emd_6287)

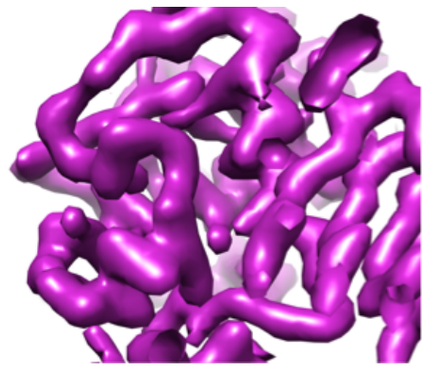


Beta-galactosidase at 2.2 Å
(autobuilt model; emd_2984)

Autobuilding Performance

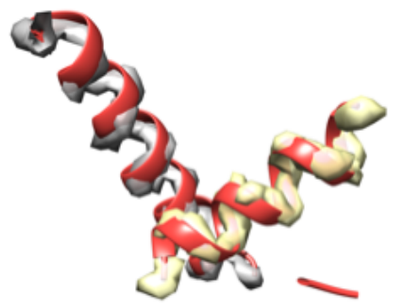
Model Building Version 2

Trace chain the way a person does

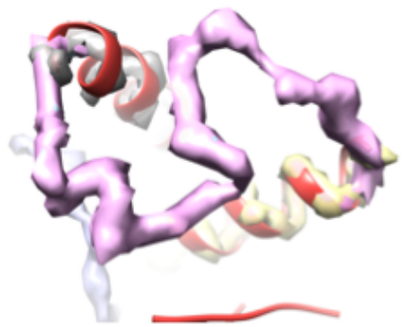


Find secondary structure

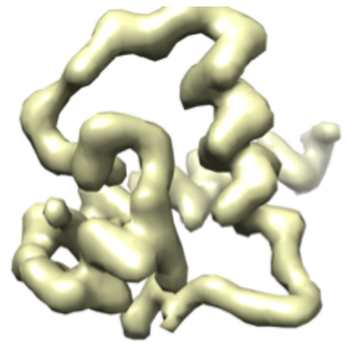
Find clear regions of density



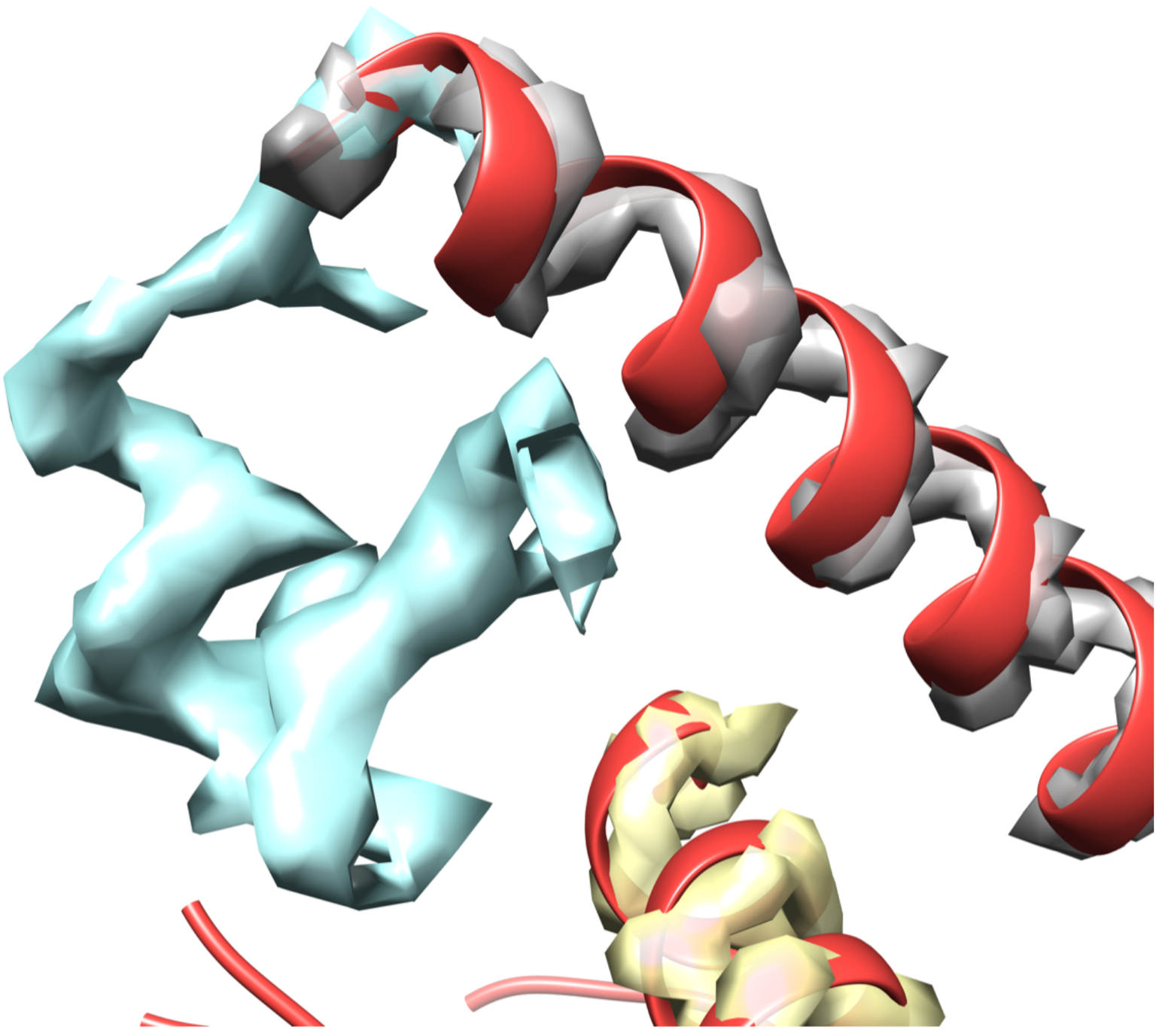
Adjust contour level until a region just connects to another



Iterate to build up a connected chain



Model Building Version 2



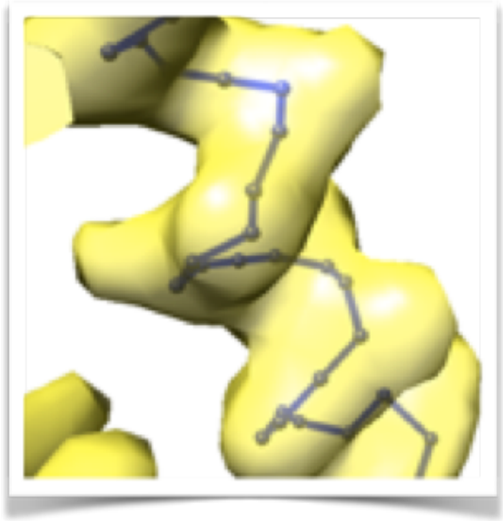
Finding C_α and C_β positions

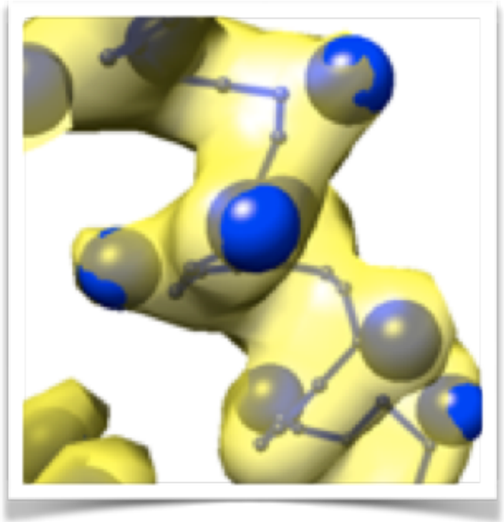
Trace chain path through high density

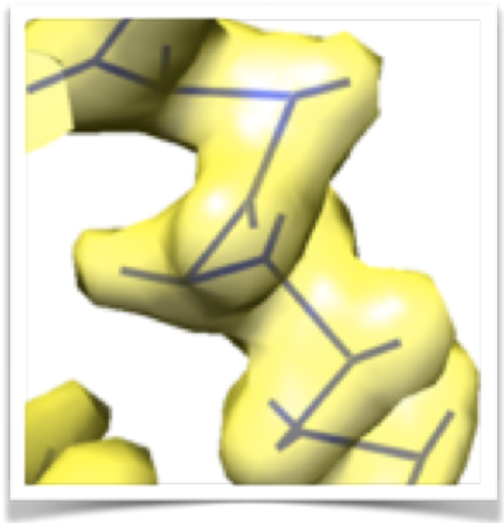
Find C_β positions from side-chain density

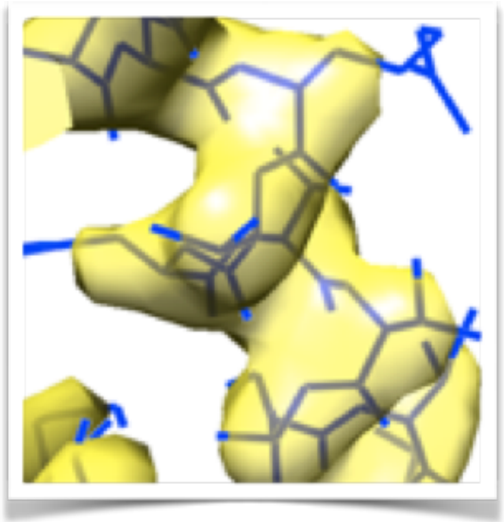
Choose C_α positions 3.8 Å apart and next to C_β positions

Construct all-atom model with Pulchra* and refine



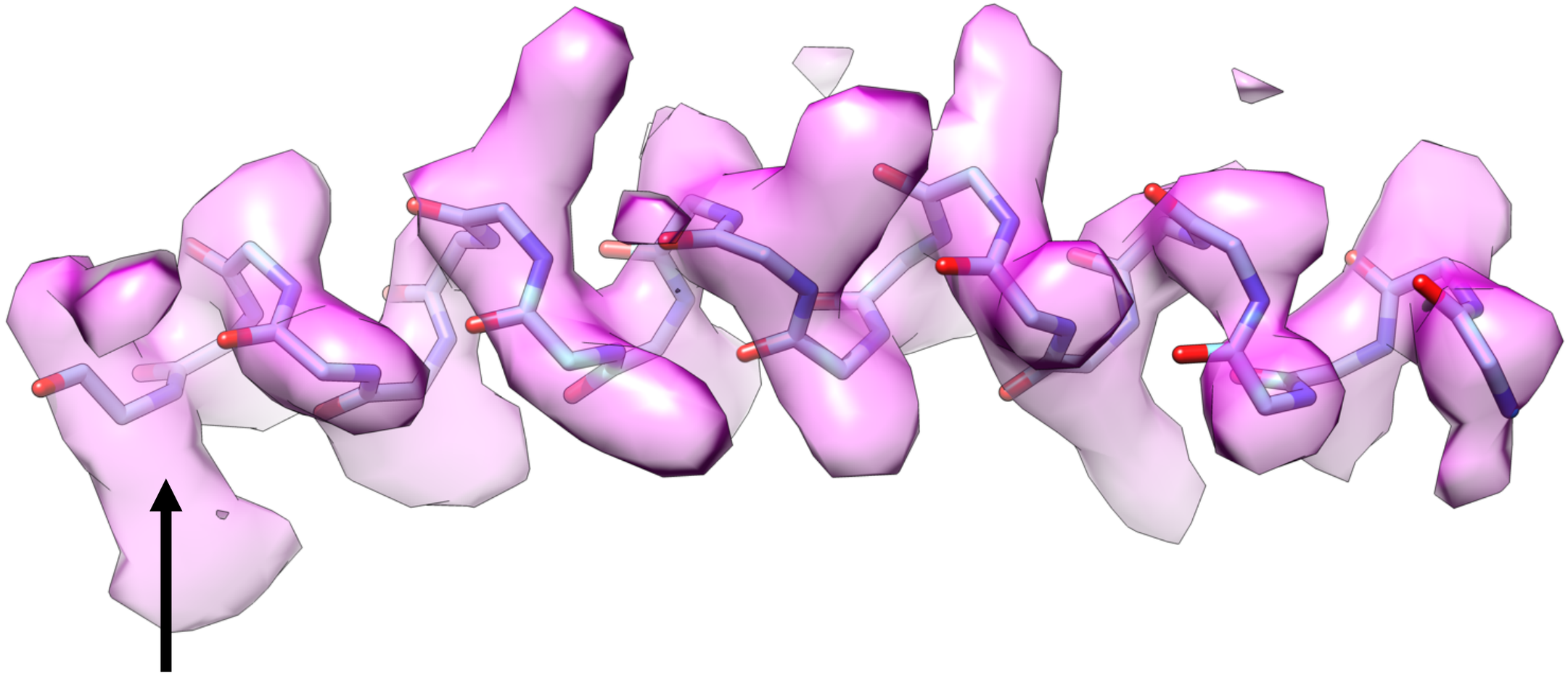






*Rotkiewicz & Skolnick (2008). J. Comp. Chem. 29, 1460.

Sequence Assignment

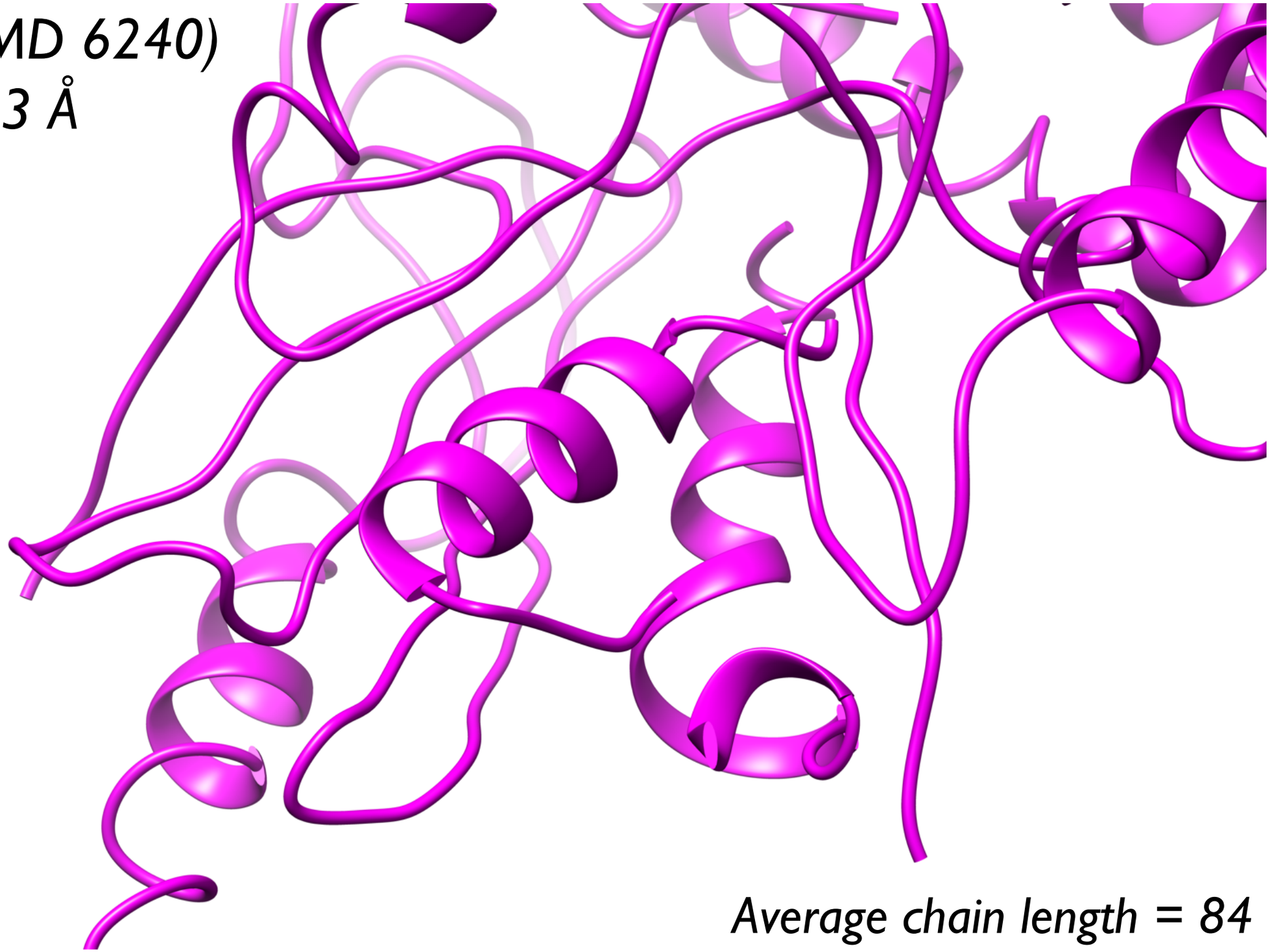


Residue	G	A	S	V	I	L	M	C	F	Y	K	R	W	H	E	D	Q	N	P	T
CC	0.30	0.50	0.53	0.47	0.58	0.62	0.68	0.59	0.83	0.77	0.71	0.69	0.70	0.82	0.65	0.64	0.60	0.60	0.35	0.47
Prob	3	0	0	0	0	0	1	0	40	23	5	5	4	9	2	2	1	0	2	0

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

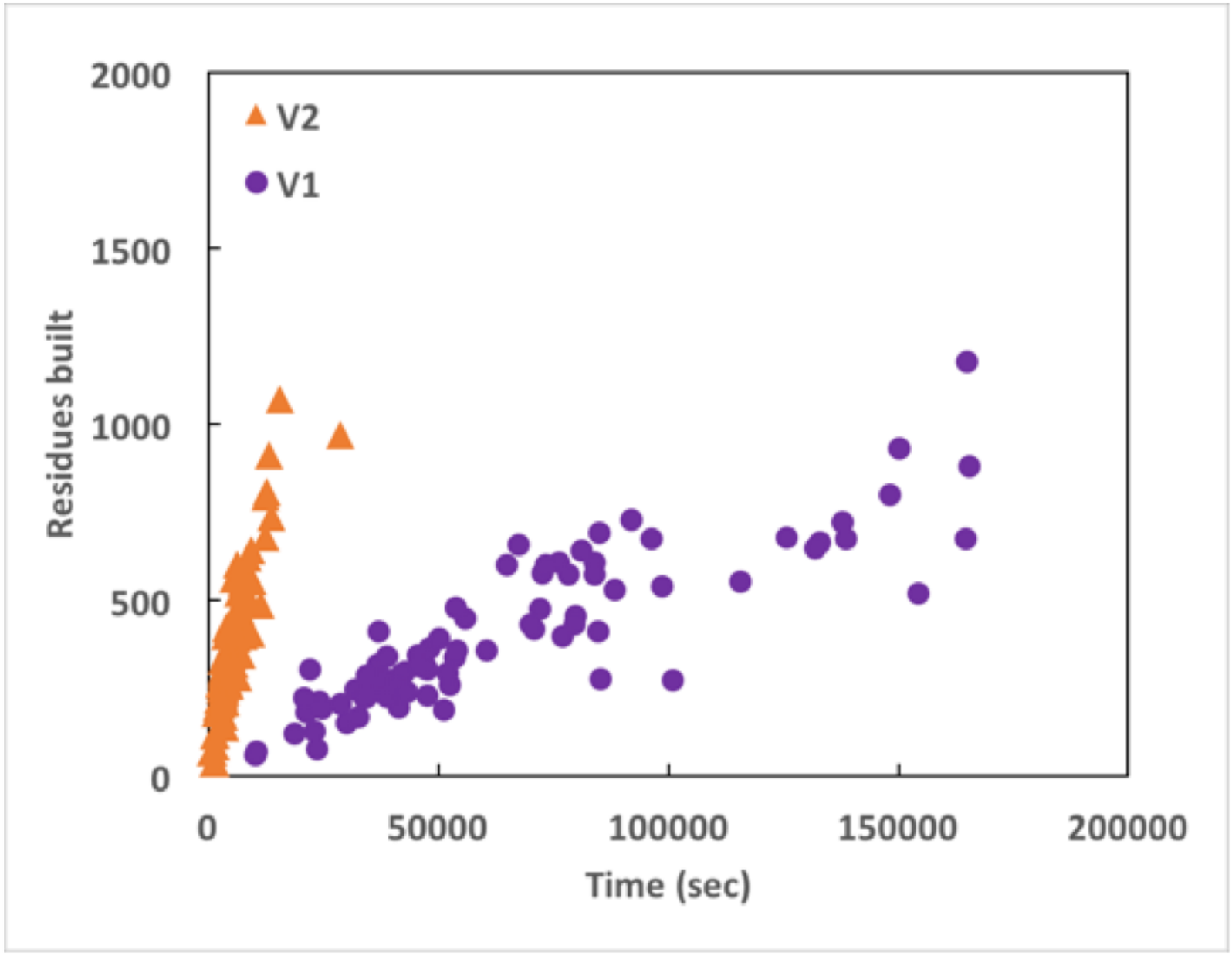
Improved Connectivity

3j9e (EMD 6240)
3.3 Å



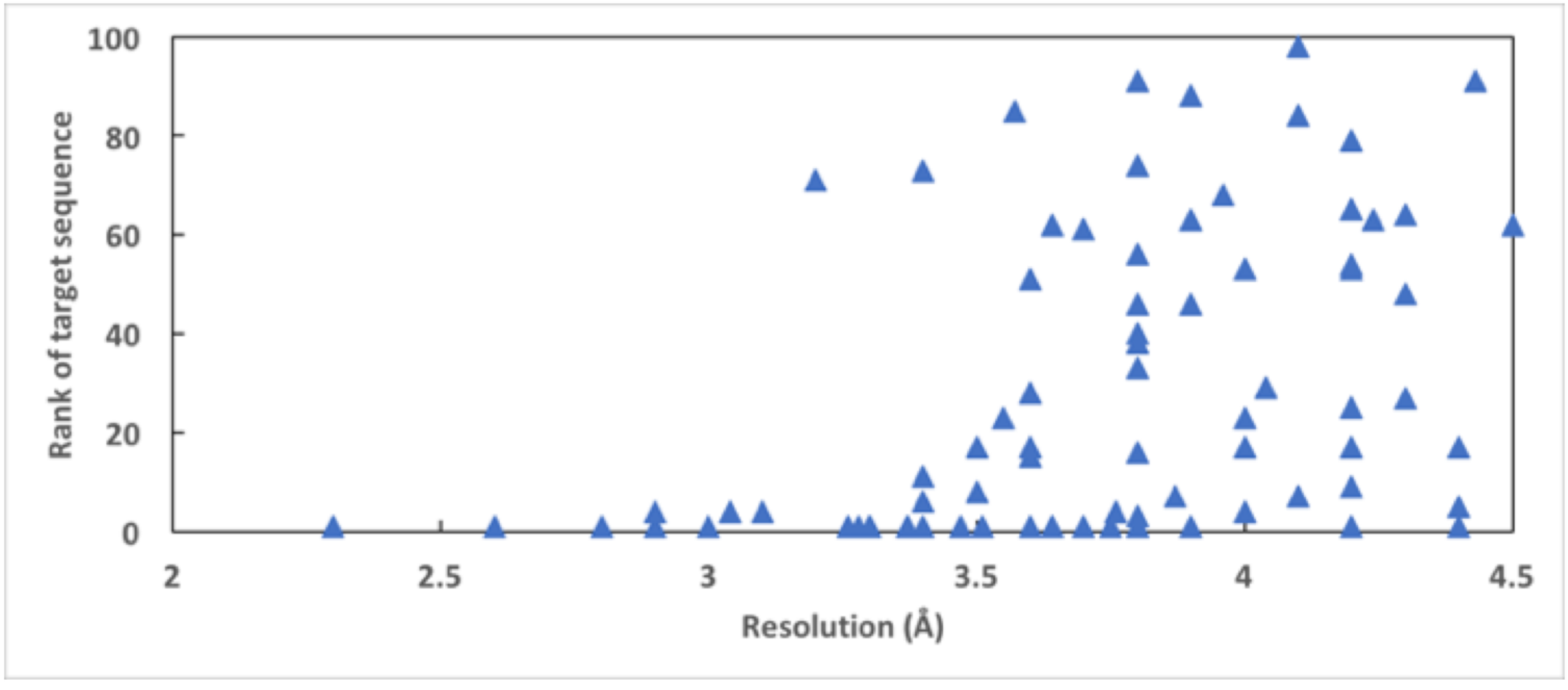
Average chain length = 84

Improved Performance



What's The Molecule?

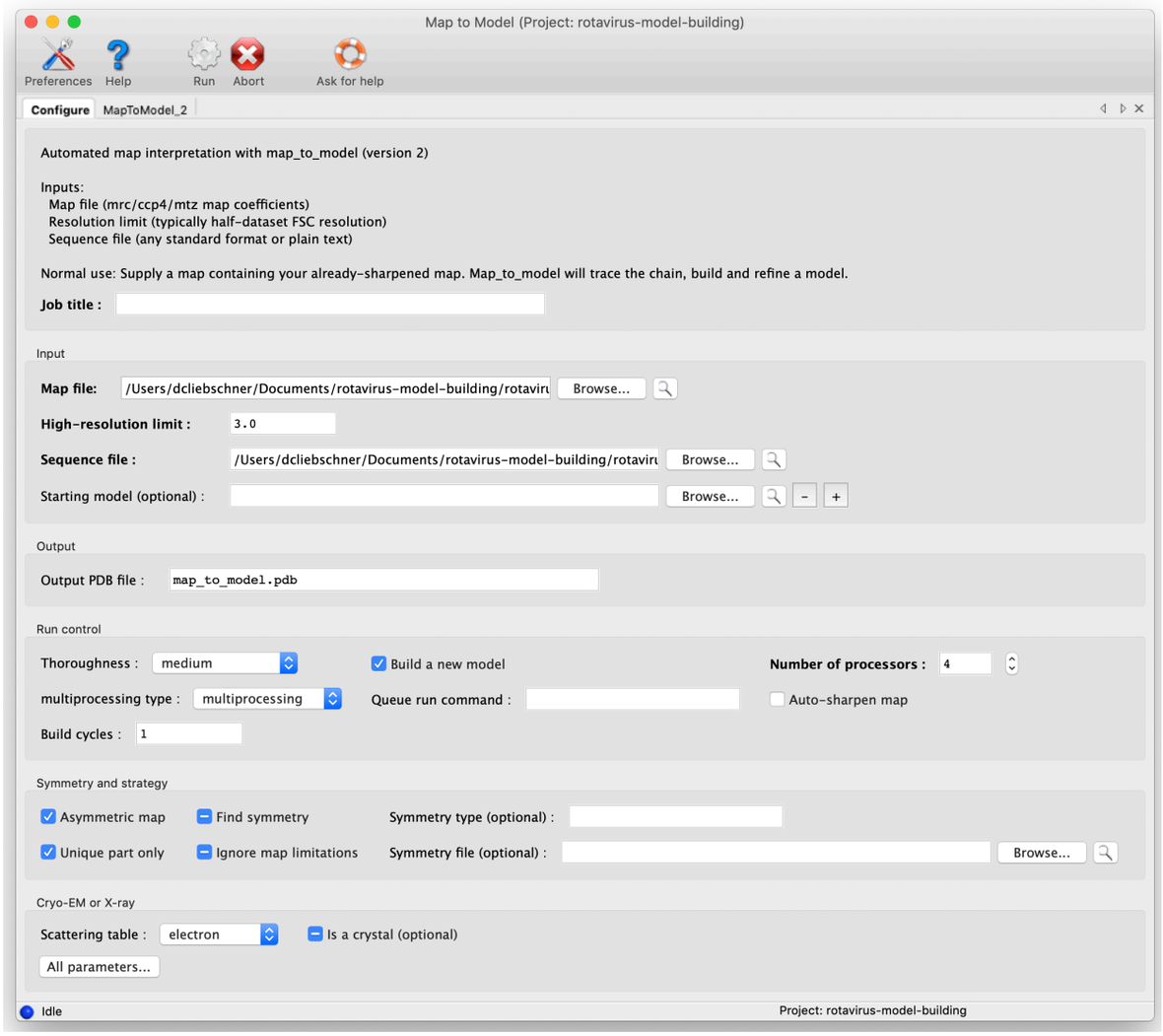
- Use the highest side chain probabilities to determine a sequence (from the map)
- Search the sequence database to identify the molecule

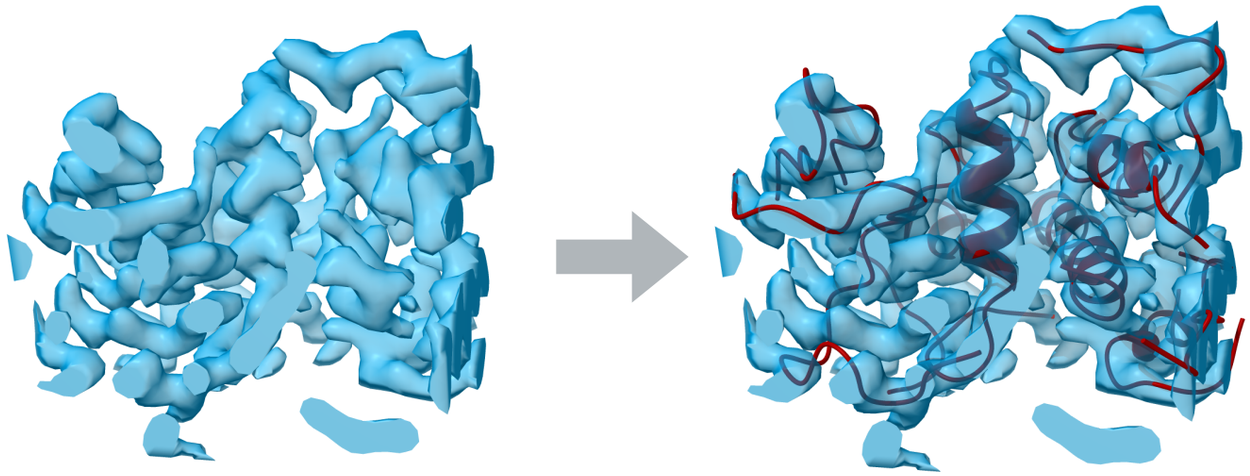


With Xiaorun Li, Chi-min Ho & Hong Zhou, UCLA

Automated model building

Map to Model (GUI)
phenix.map_to_model





Model built into the map

Map,
Sequence,
Resolution

Conclusions

Automated model building is possible, but can be improved

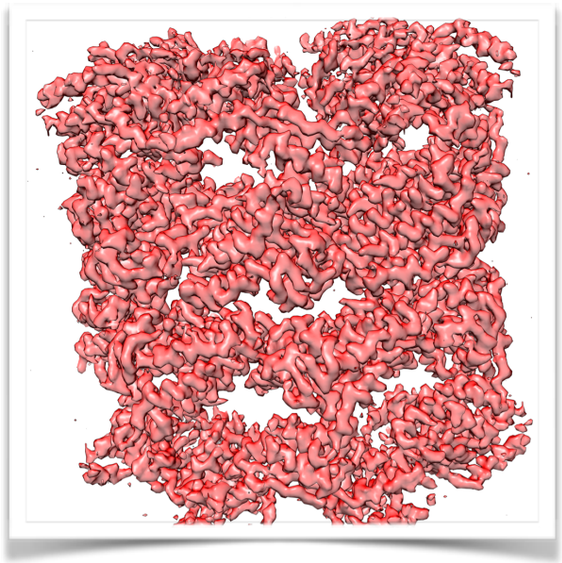
- Include information from secondary structure prediction, evolution etc.
- Combine structure-modeling tools (Rosetta) with Phenix model-building

Many challenges remain:

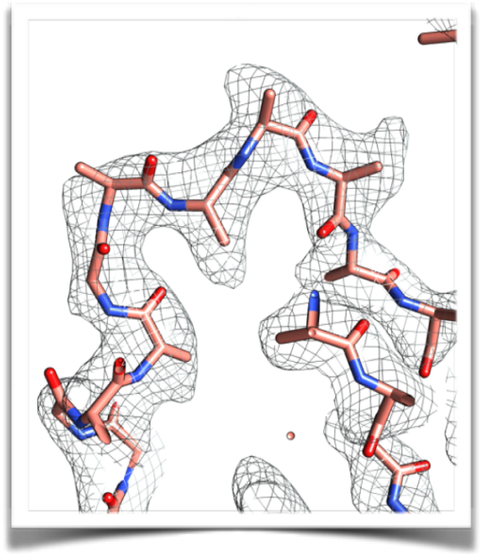
- Reliably accounting for uncertainty in magnification
- Local variation in resolution leads to uncertainties in interpretation

Title

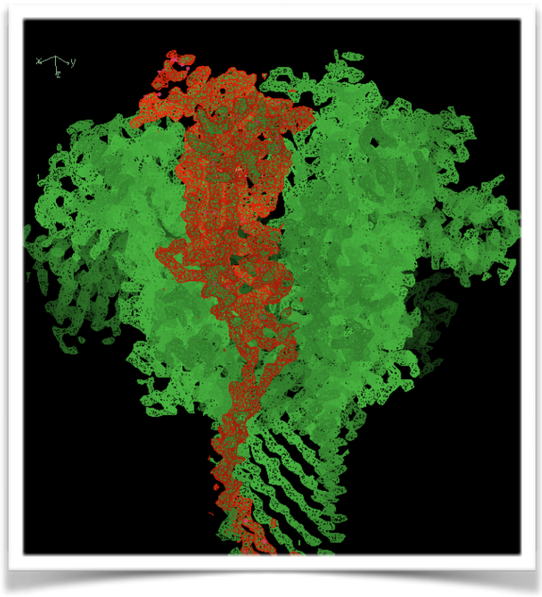
New Tools for Cryo-EM in *Phenix*



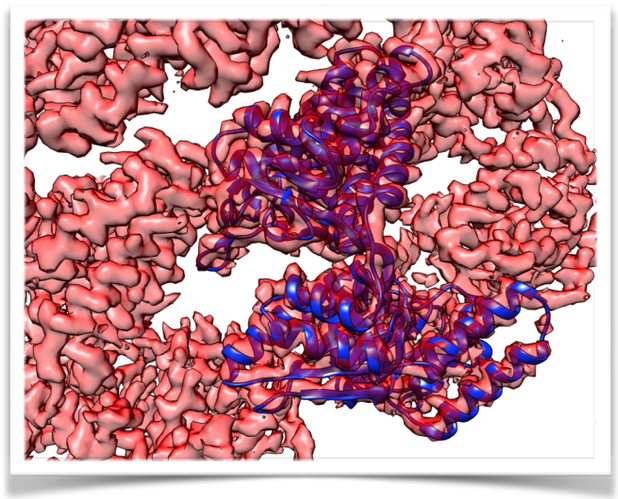
Symmetry from a map



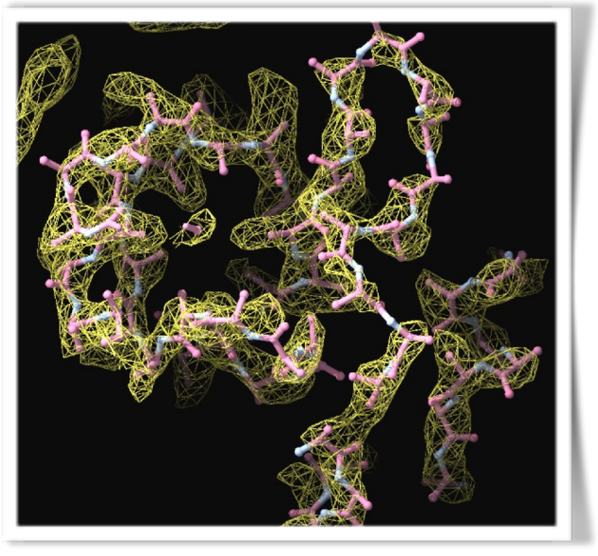
Automated map sharpening



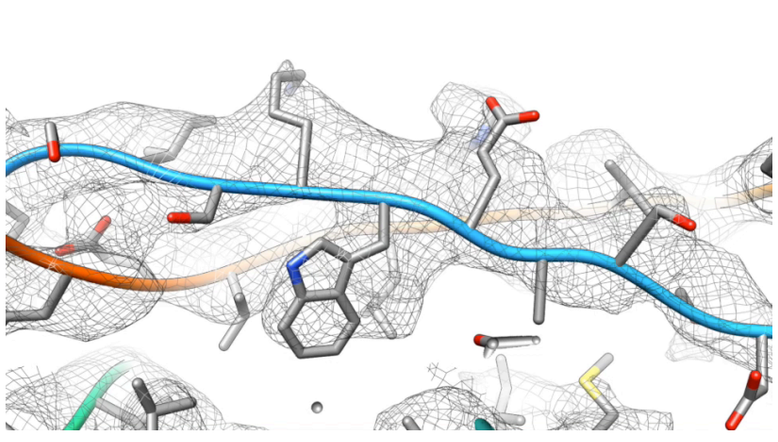
Map segmentation



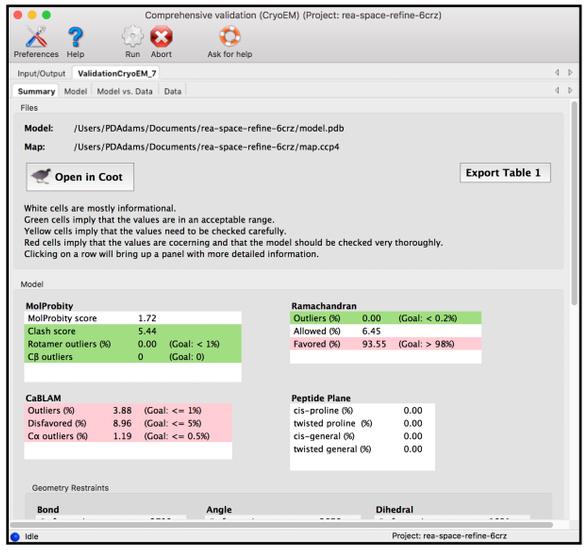
Rigid model docking



Automated model building



Real space refinement



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