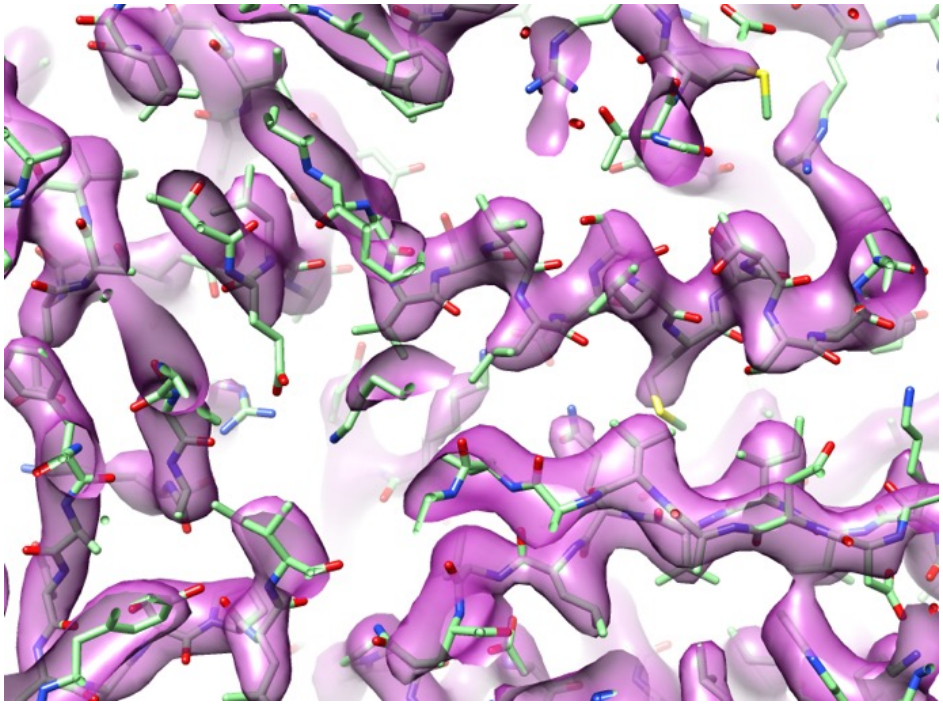
Cryo-EM Model-building

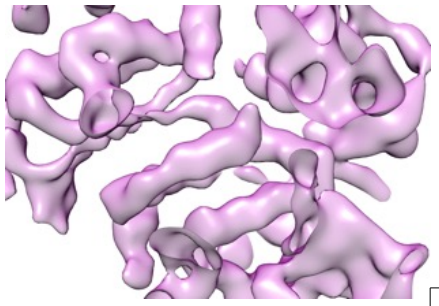


Phenix workshop

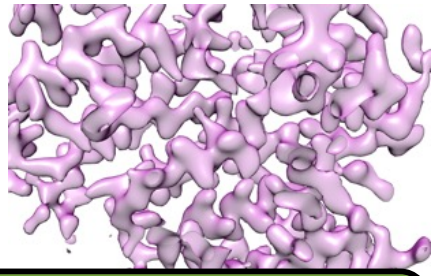
Sept. 15, 2020

Tom Terwilliger, New Mexico Consortium
Li-Wei Hung, Los Alamos National Laboratory
Pavel Afonine, Paul Adams, Dorothee Liebschner, Lawrence Berkeley National Laboratory

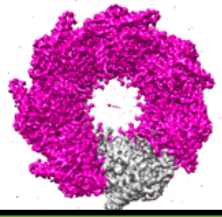
Cryo-EM tools in Phenix



Starting map



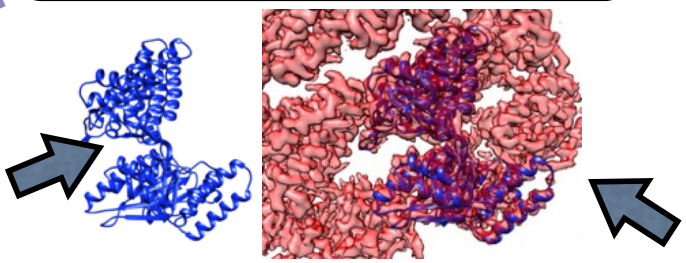
Map improvement



Map symmetry

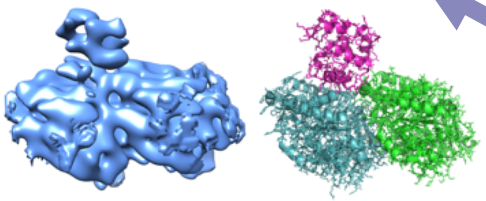
Map manipulation

Extract unique part



Docking, model-building

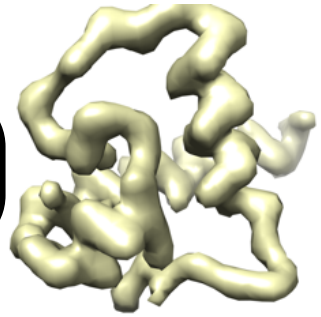
Refinement



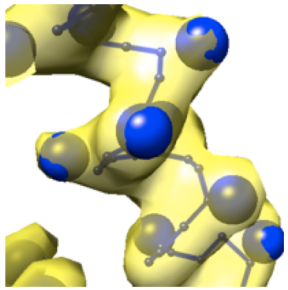
Validation

Map-to-model: Cryo-EM model building

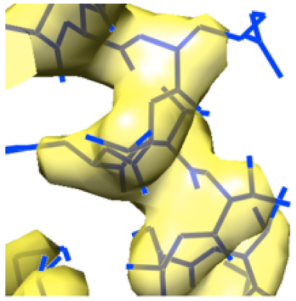
Isolate density for a chain



Identify C_α and C_β positions from side-chain density



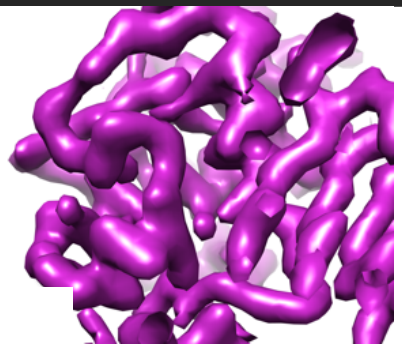
Construct and refine all-atom model



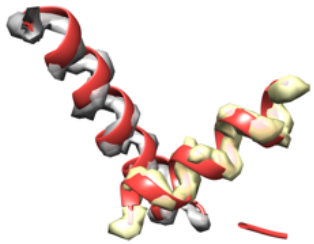
Terwilliger et al., (2018) Nature Methods 15, 905–908
Terwilliger et al., (2020) Protein Science 29, 87-99

Isolating density for a chain

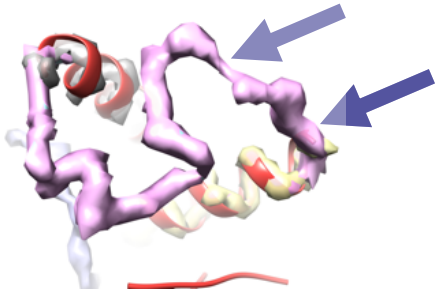
Trace chain the way a person does



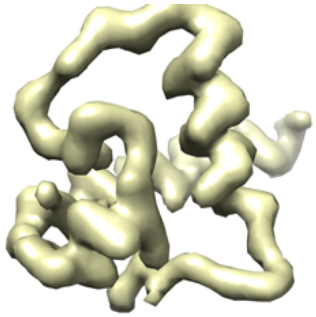
Find secondary structure (helices, strands)



Find connections: adjust contour level until a region just connects to one other



Iterate to build up chain



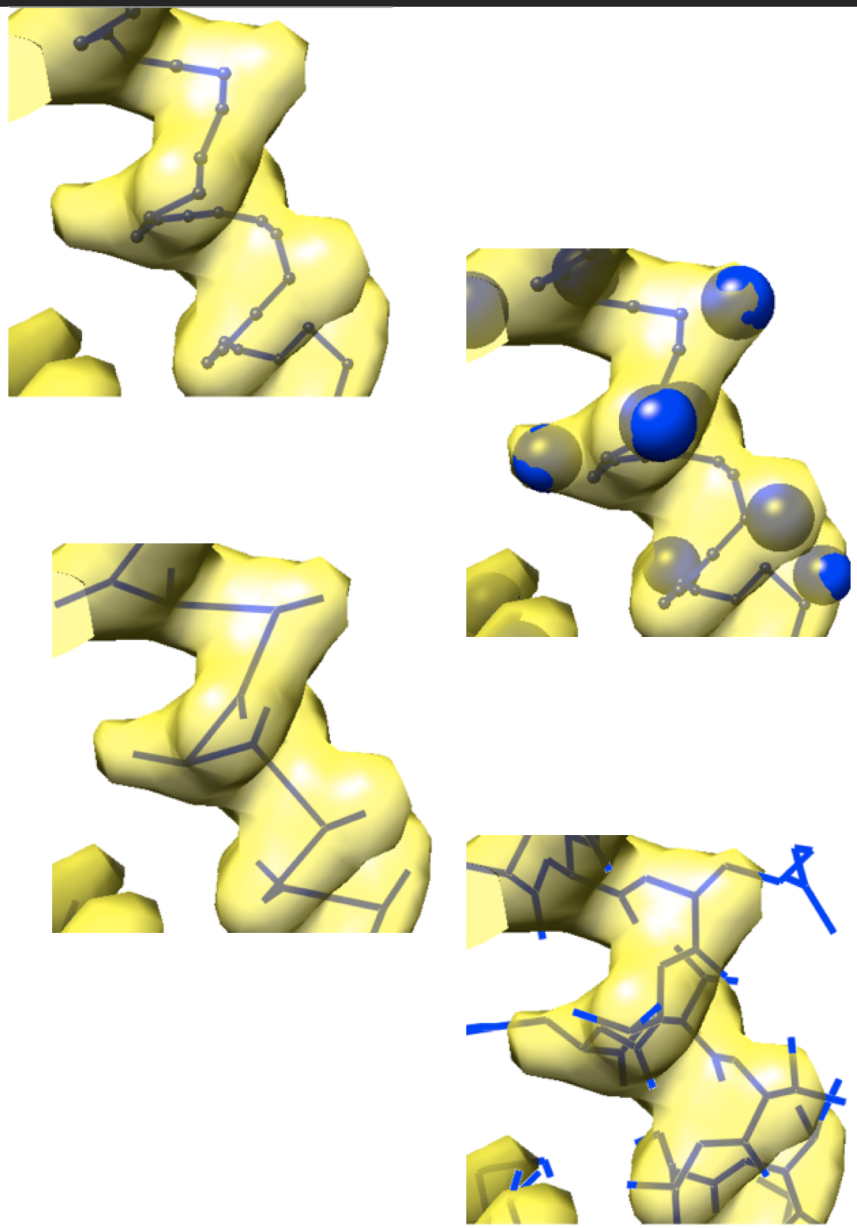
Finding C_α and C_β positions

Trace chain path through high density

Find C_β positions from side-chain density

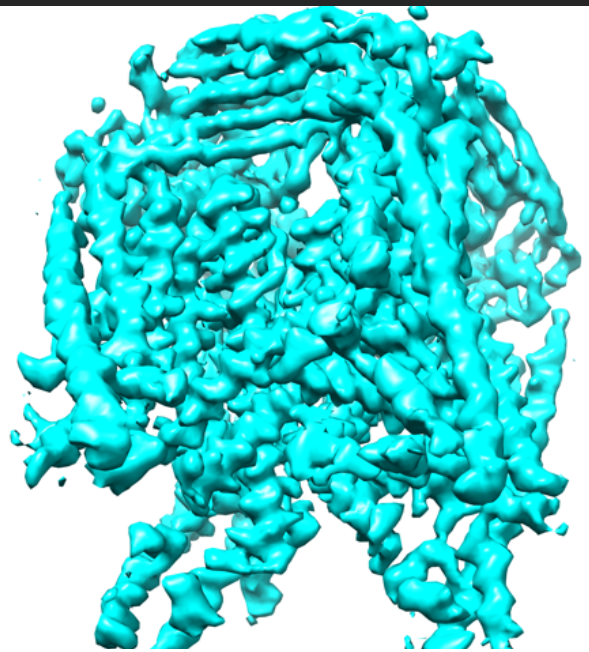
Choose C_α positions 3.8 Å apart and next to C_β positions

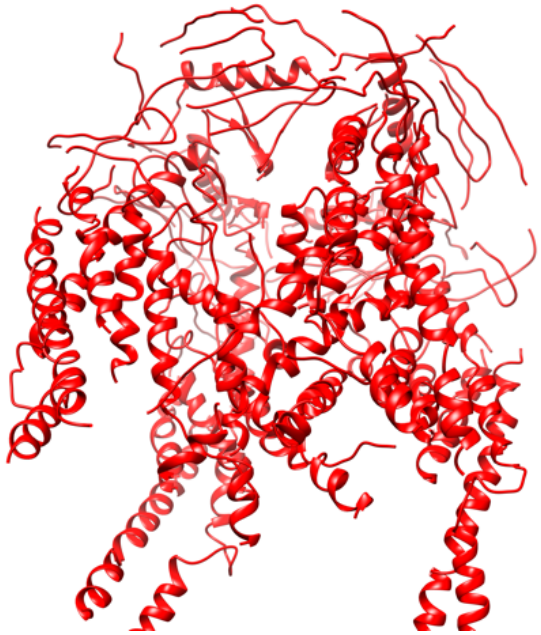
Construct all-atom model with Pulchra* and refine

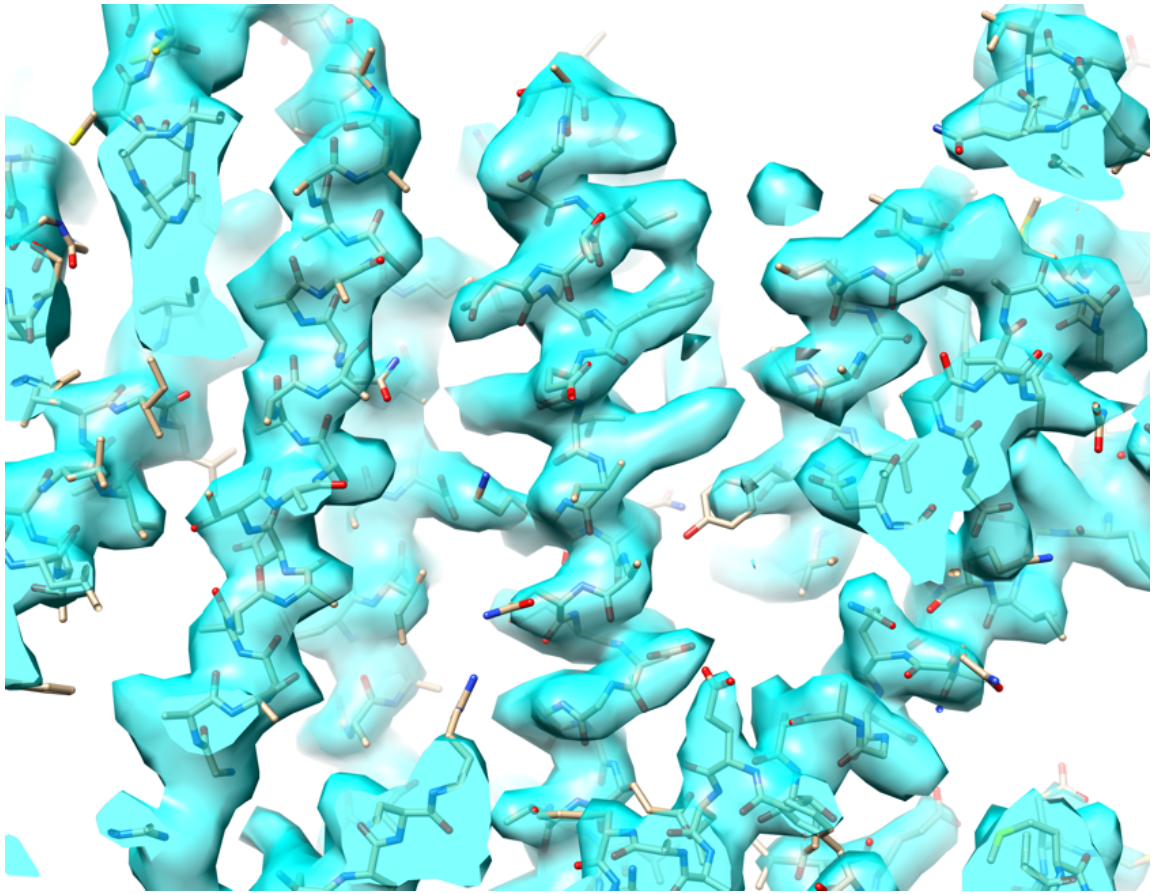


*Rotkiewicz & Skolnick (2008). J. Comp. Chem. 29, 1460.

TRPML3 channel (4.1 Å, 78% built, 1.3 Å rmsd)

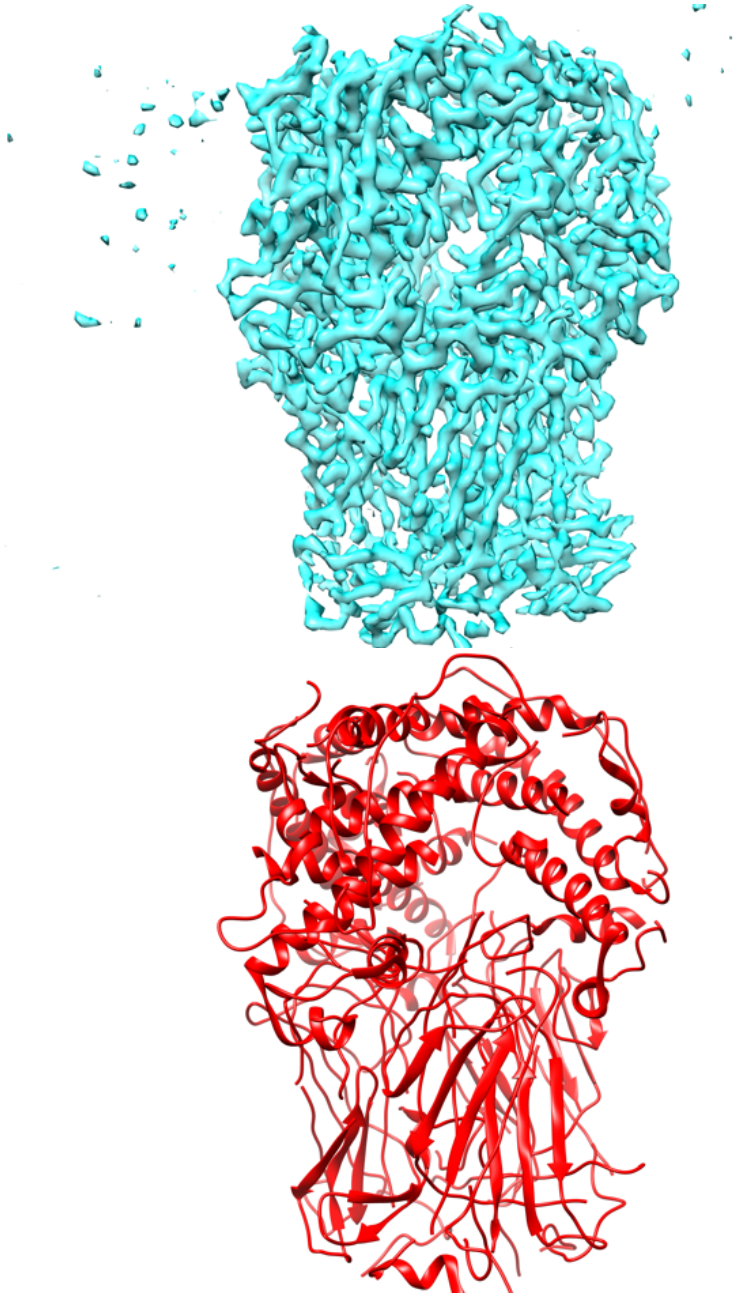


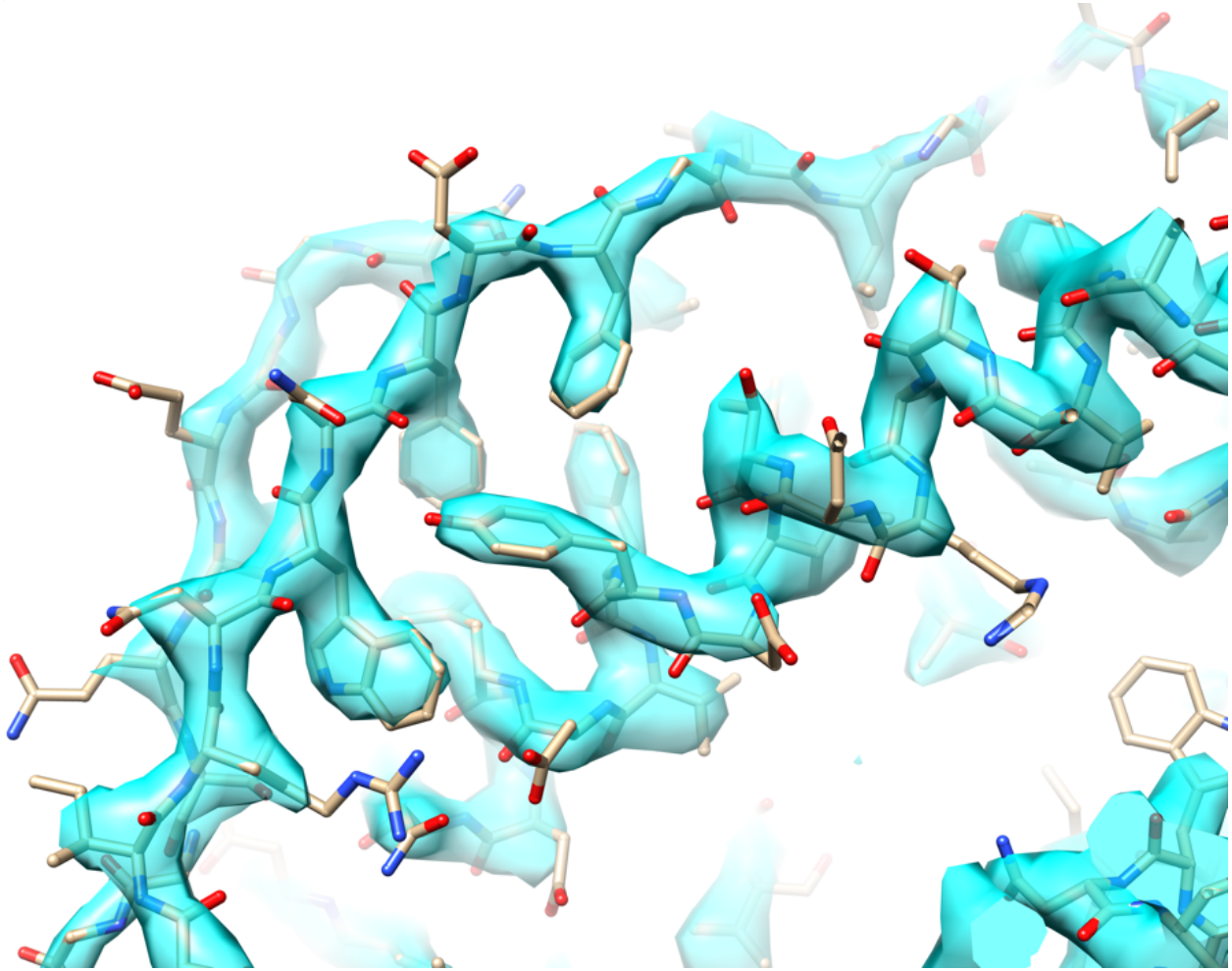




Data from Zhou, X. et al. (2017) Nat. Struct. Mol. Biol. 24: 1146

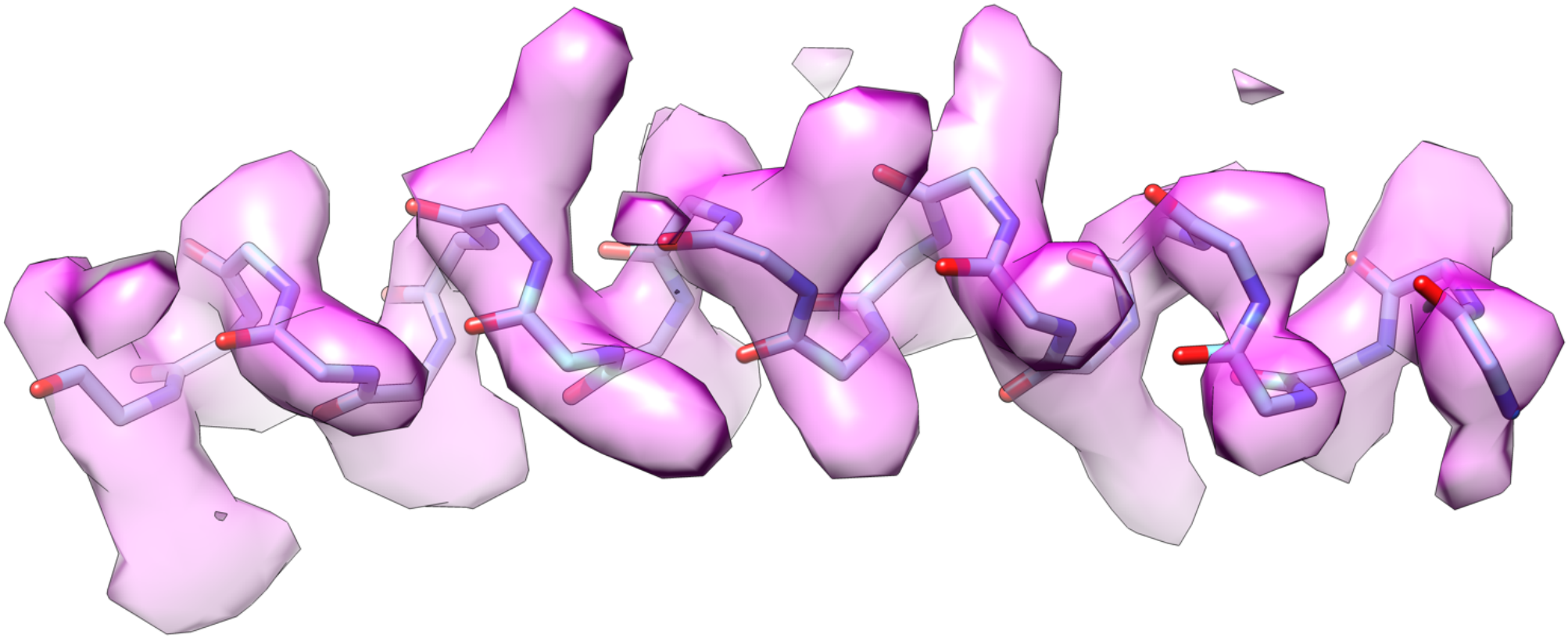
Rotavirus VP6 (2.6 Å, 100% built, 0.9 Å rmsd)





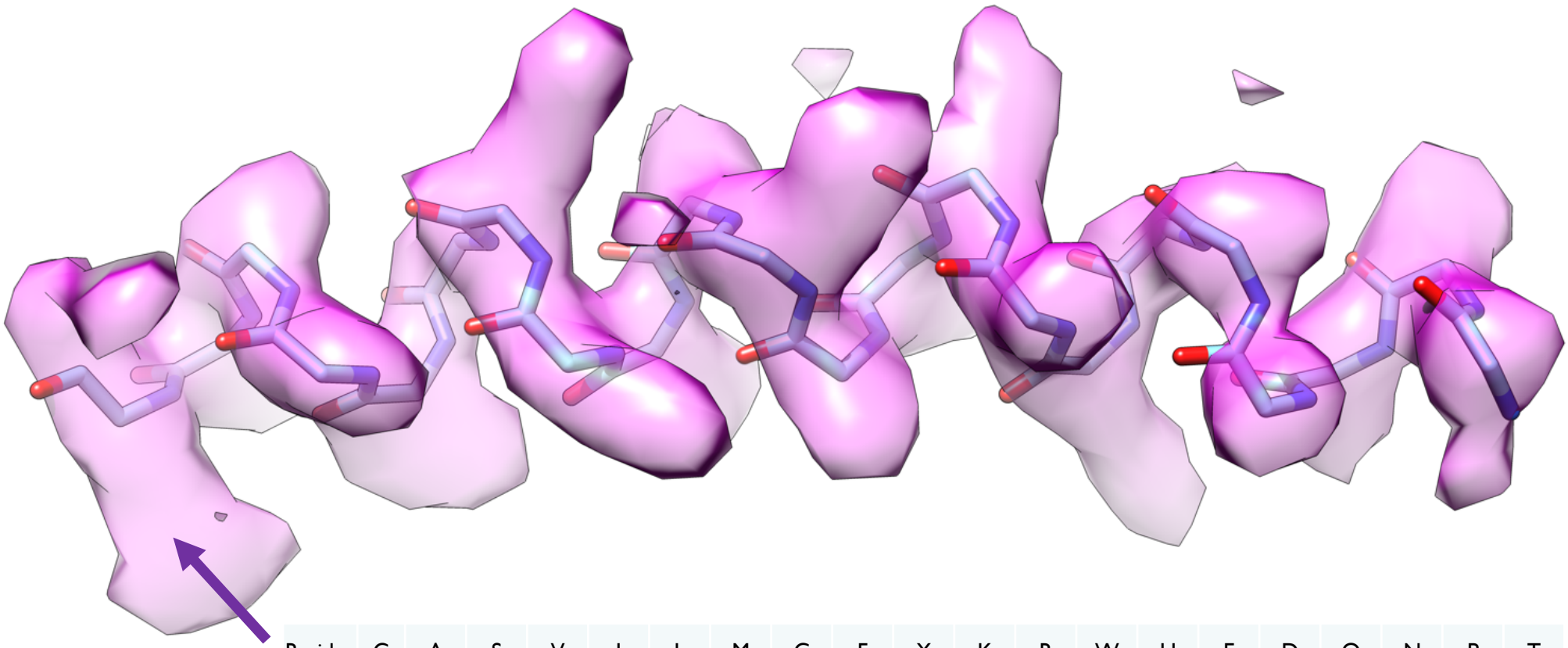
Data from Grant and Grigorieff, eLife 2015;4:e06980

Sequence from density in a map



Can we identify what part of the sequence goes with this helix?

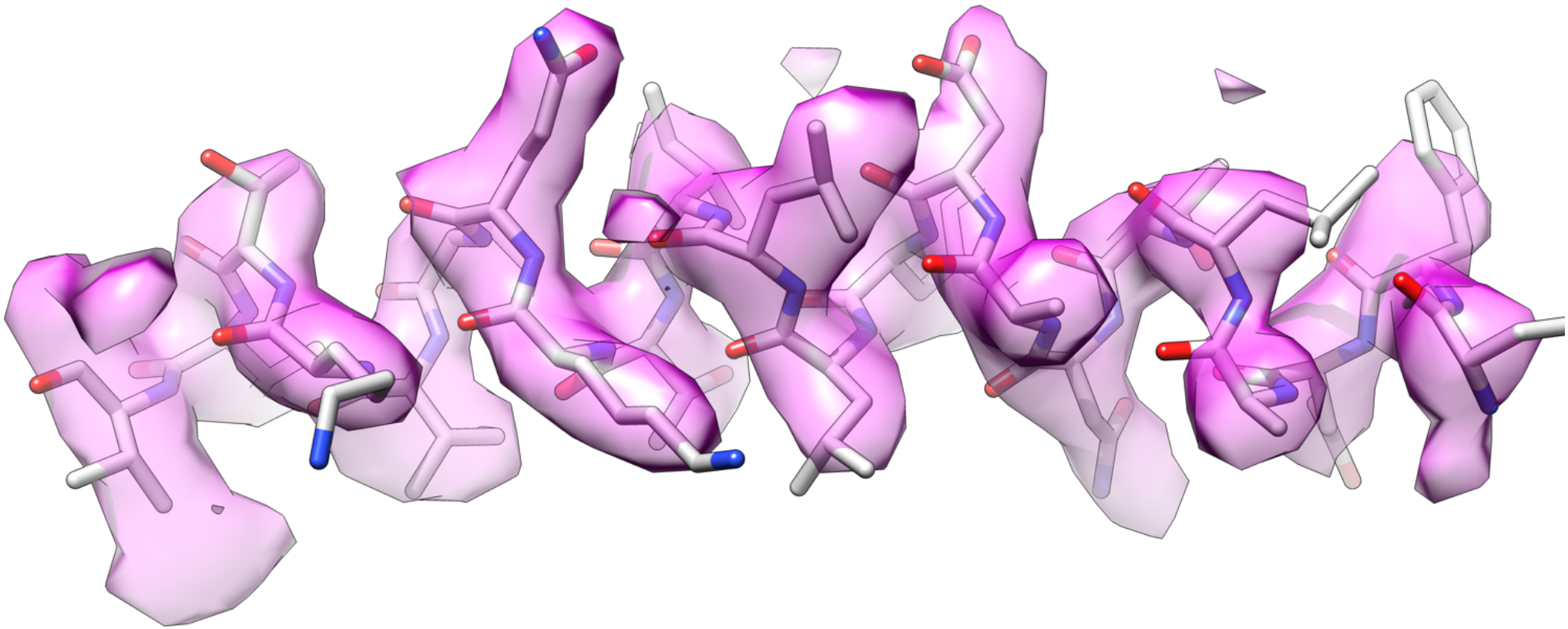
Sequence assignment



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

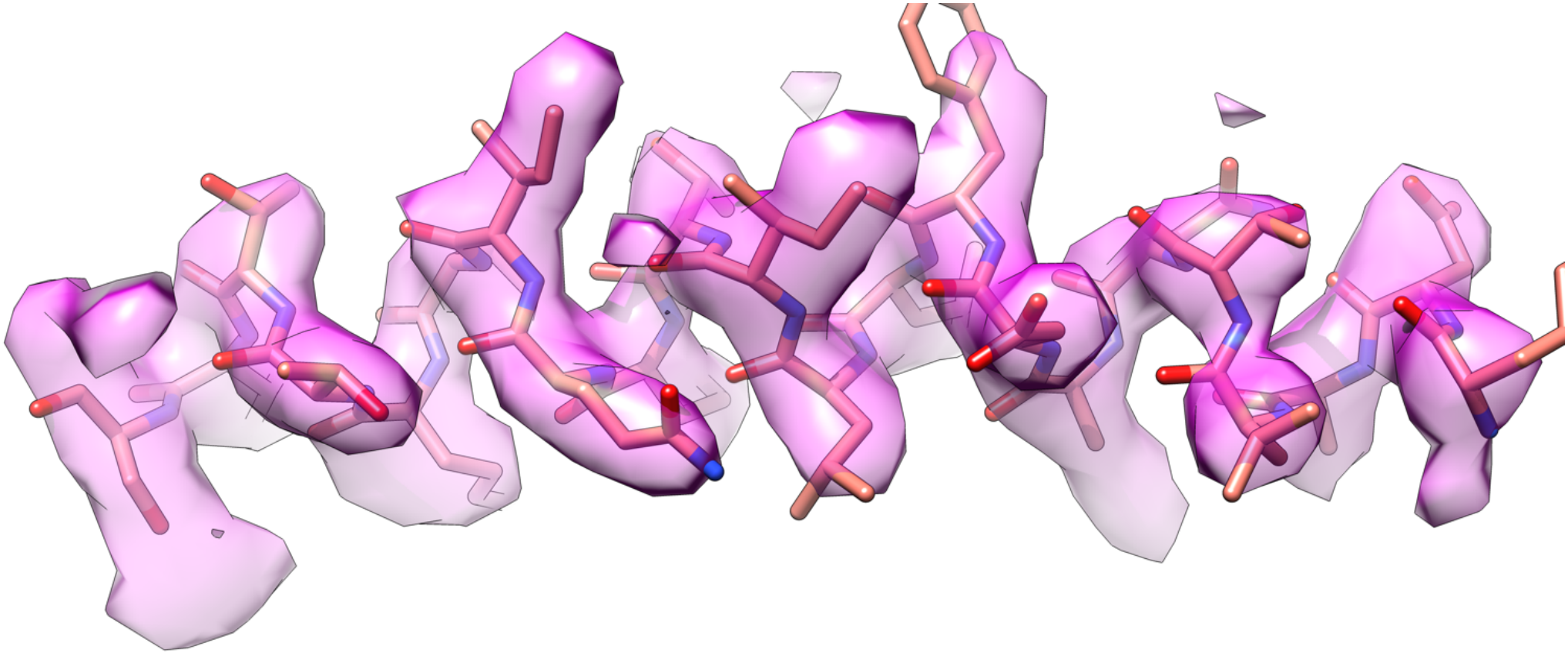
Compatibility of density at position 21 with each amino acid

Sequence assignment: Test all alignments



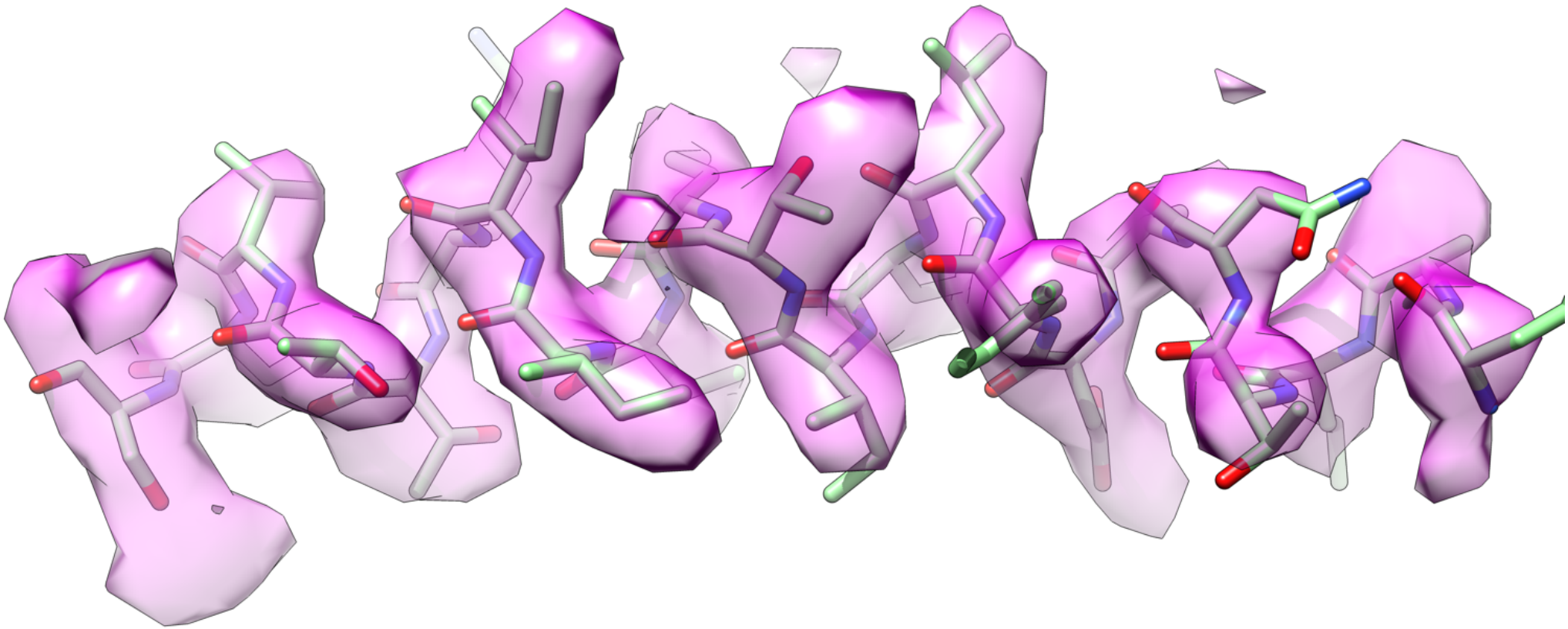
Offset: -3 Score: -4.3

Testing all alignments…



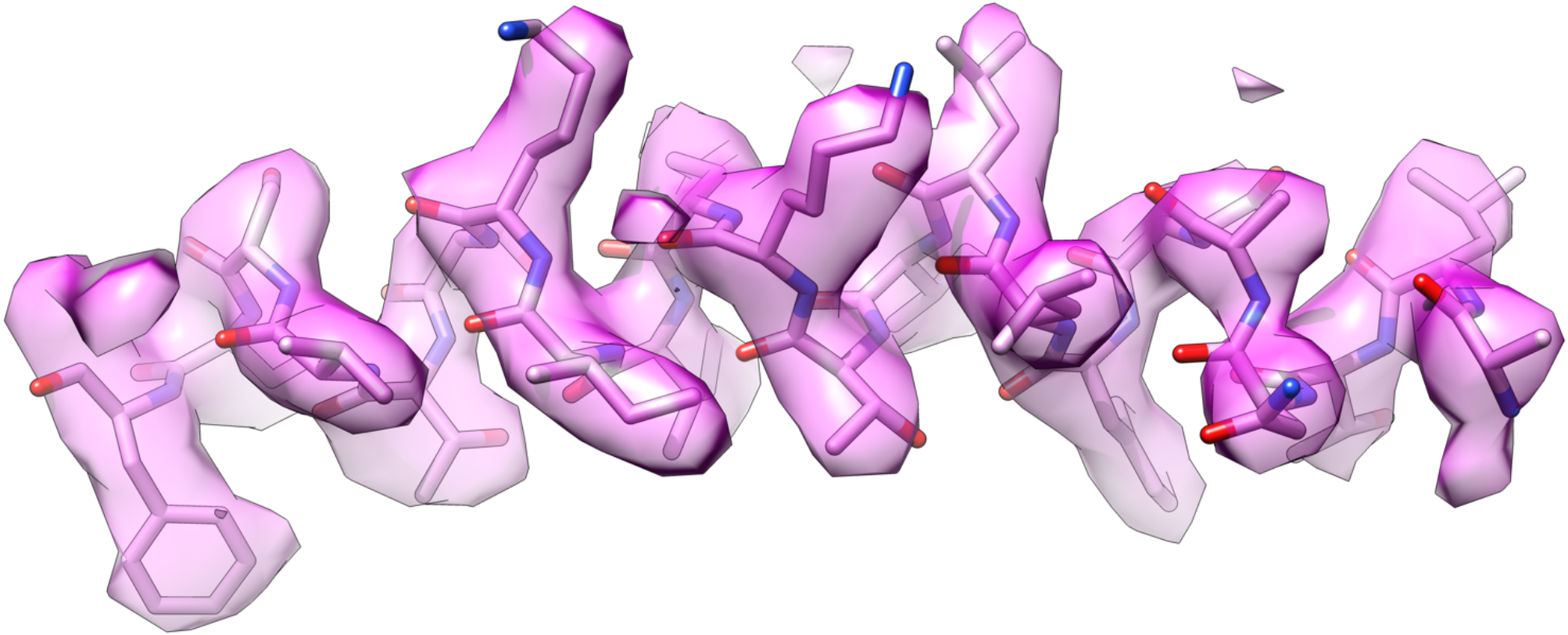
Offset: -2 Score: -0.4

Testing all alignments…



Offset: -1 Score: 4.0

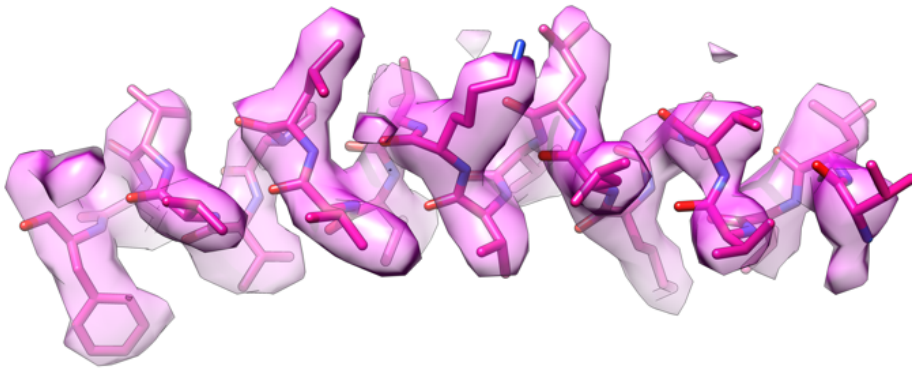
Testing all alignments…

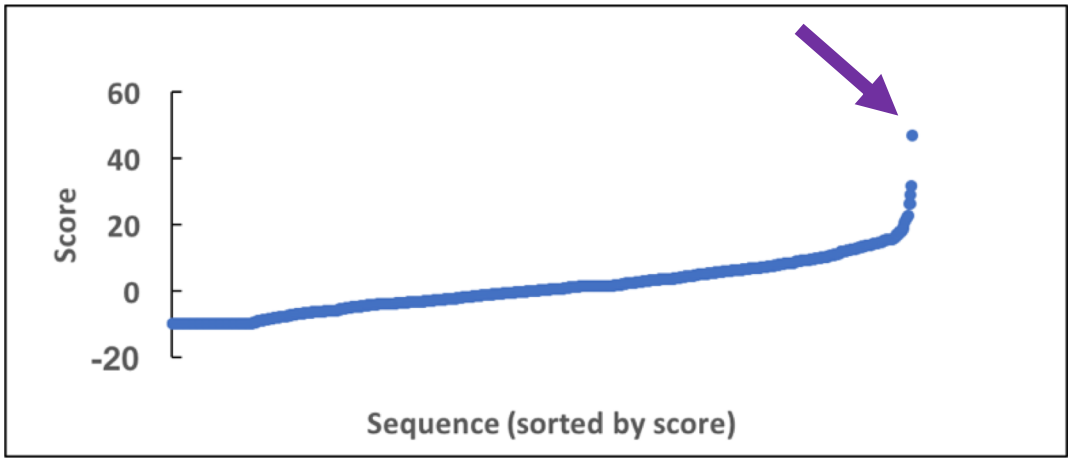


Offset: 0 Score: 25.4

Identifying a protein from a map

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



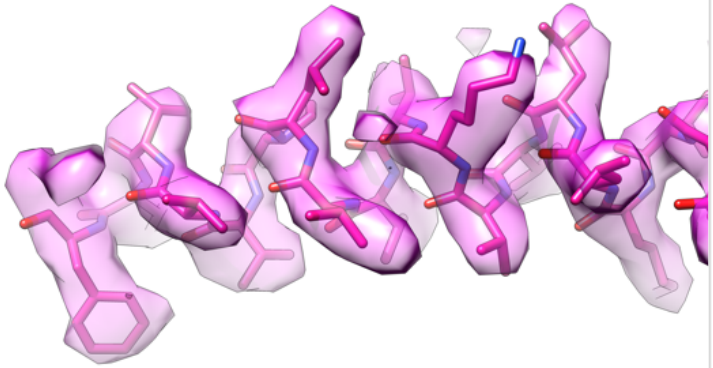


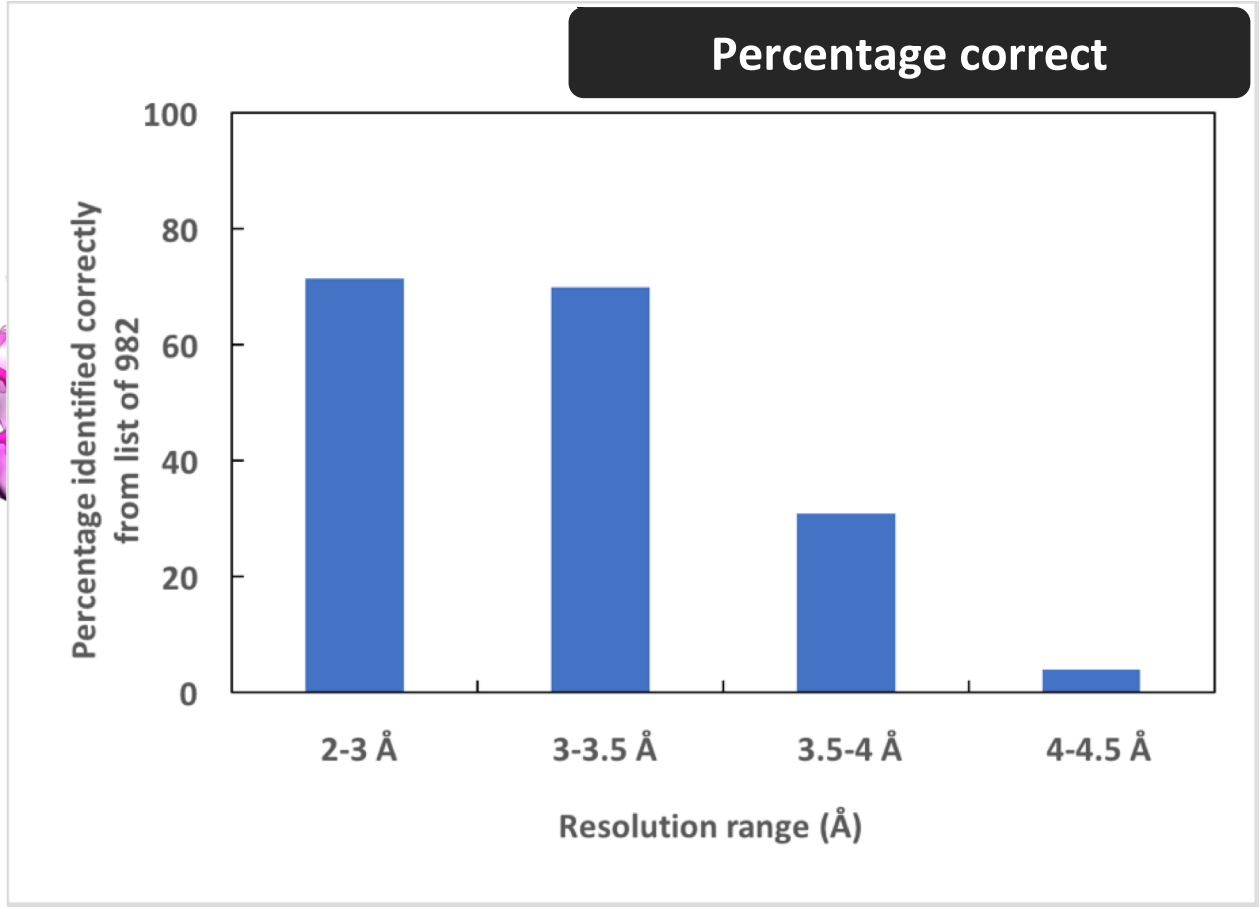
Sequence alignment of full chain against 883 sequences

ID of correct sequence (*P. falciparum* glutamine synthetase) with Z-score of 6.0

Reliability of identifying a protein from a map

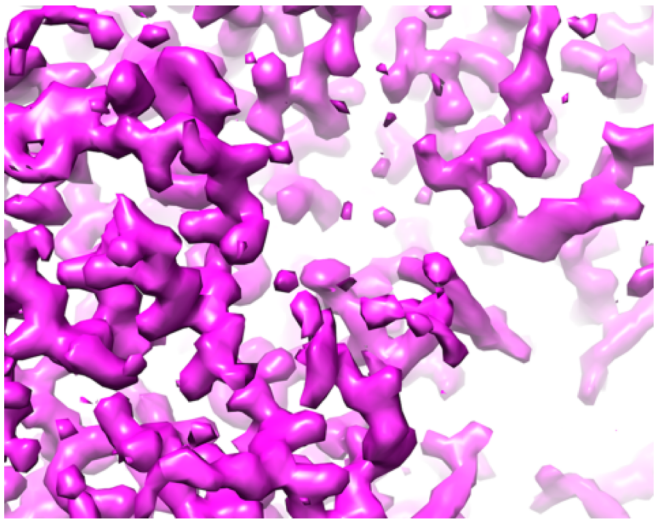
With Xiaorun Li, Chi-min Ho & Hong Zhou, UCLA

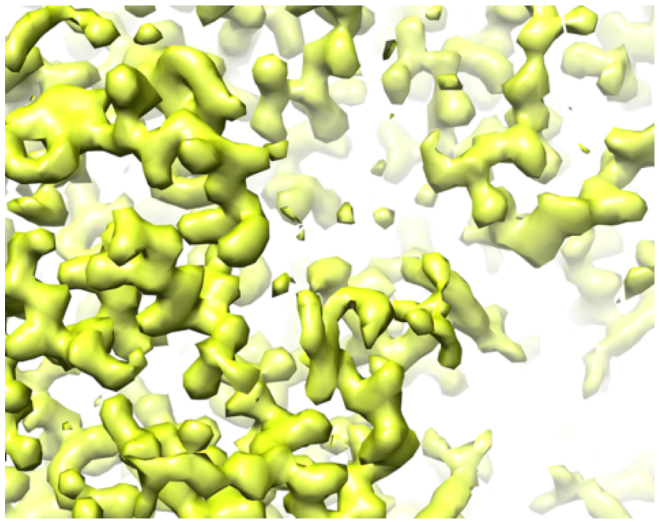


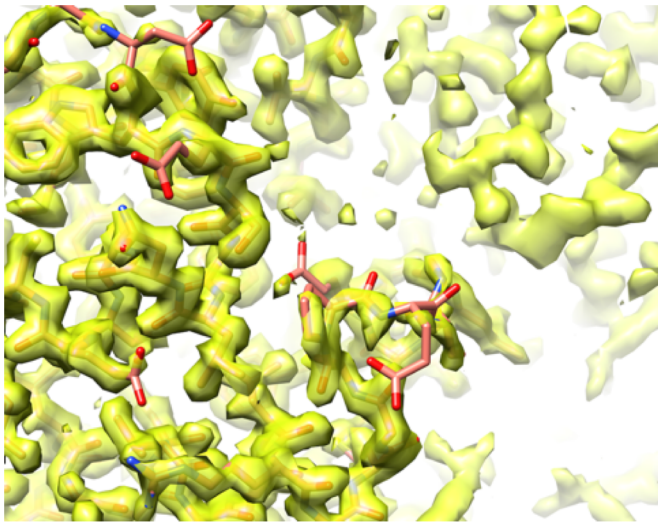


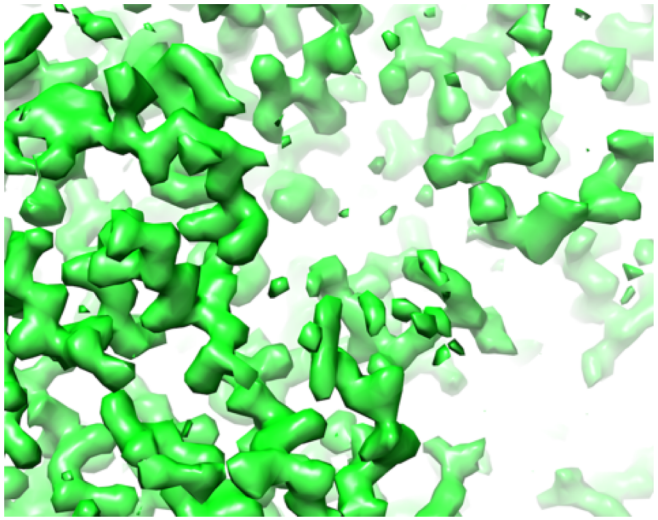
Build models for 92 maps from EMDB
Each model: Rank 982 sequences by Z-score of fit to density

Demo: model-building from a density-modified map



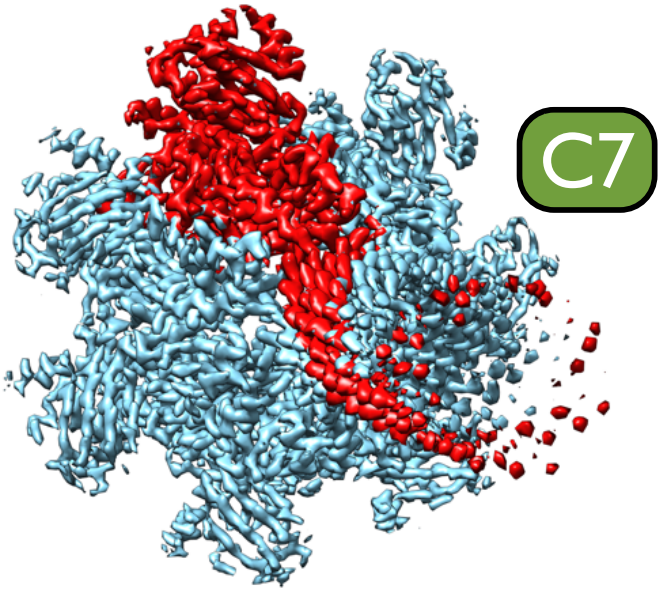


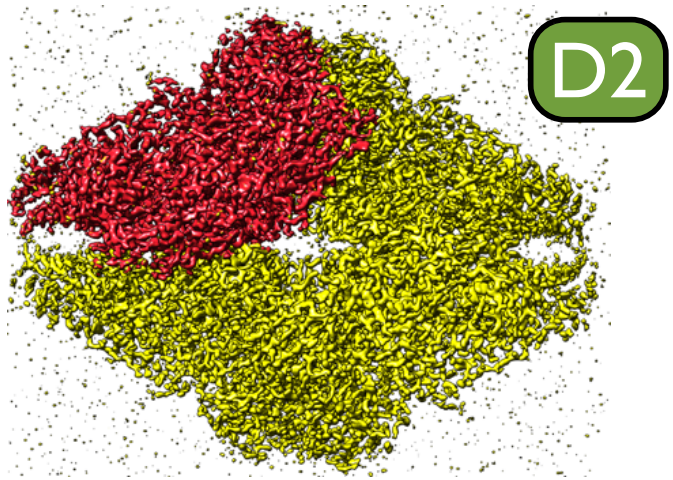




2 half-maps → Density-modified map → Model

Finding map symmetry with symmetry_from_map

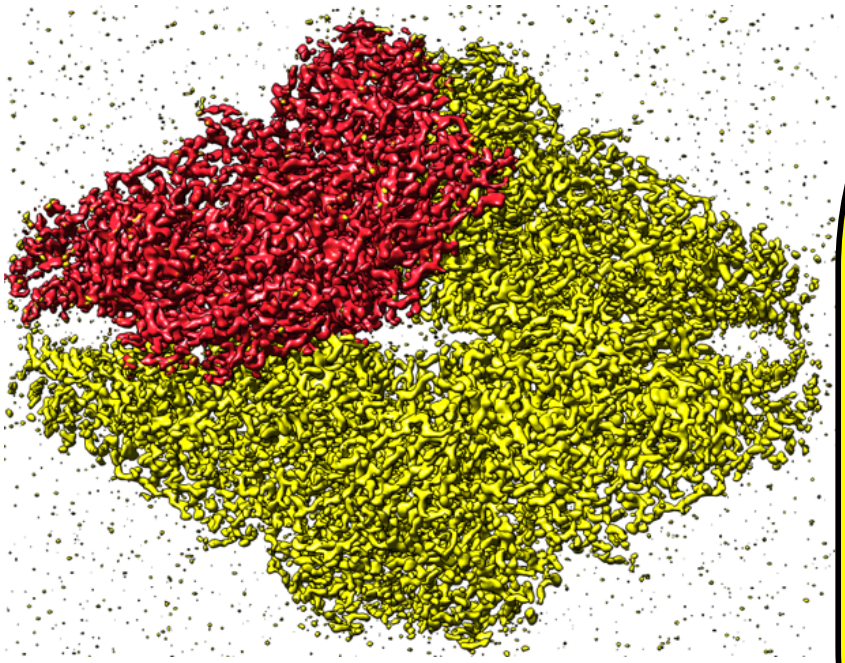




Procedure for finding symmetry:

- Test point group symmetries (e.g., C7, D2, I, O, T)
- Helical symmetry
- Principal rotation axes along z, x, y
- Score based on map correlation for symmetry-related points and number of operators

Extracting unique part of map with map_box (extract_unique option)

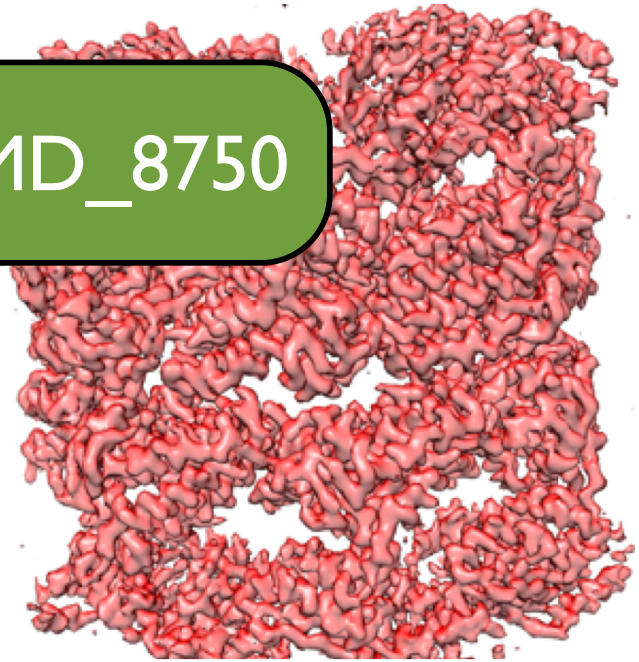


Procedure:

- Use symmetry of map
- Contour map at level that yields regions about 50 residues in size
- Group symmetry-related regions
- Choose one member of each group
- Optimize compactness and connectivity of unique part of map

Docking models with dock-in-map

EMD_8750



1ss8 chain A

Search procedure:

Pure translation

- low-res
- high-res

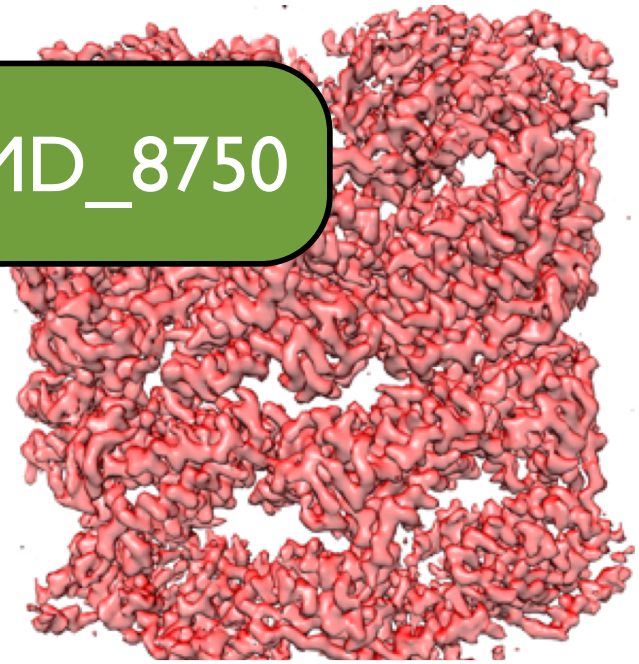
Rotation / translation

- low-res
- high-res

Score based on rigid-body refinement map-model correlation

Docking models with dock-in-map

EMD_8750



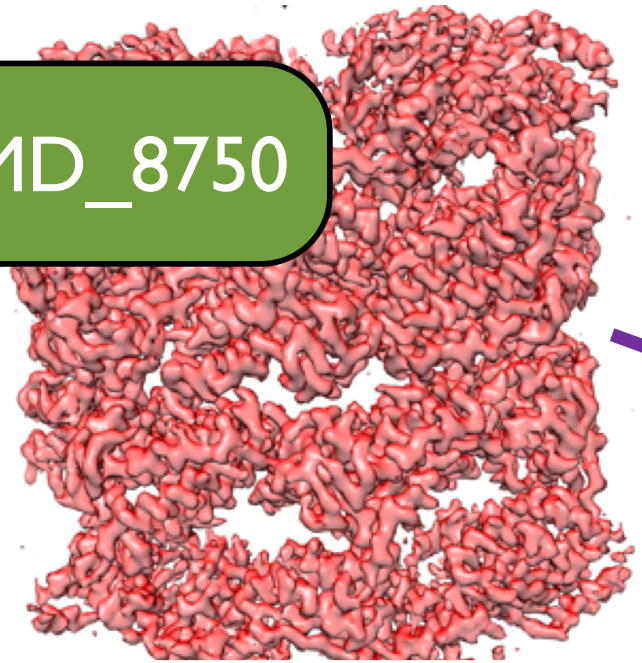
1ss8 chain A

Features

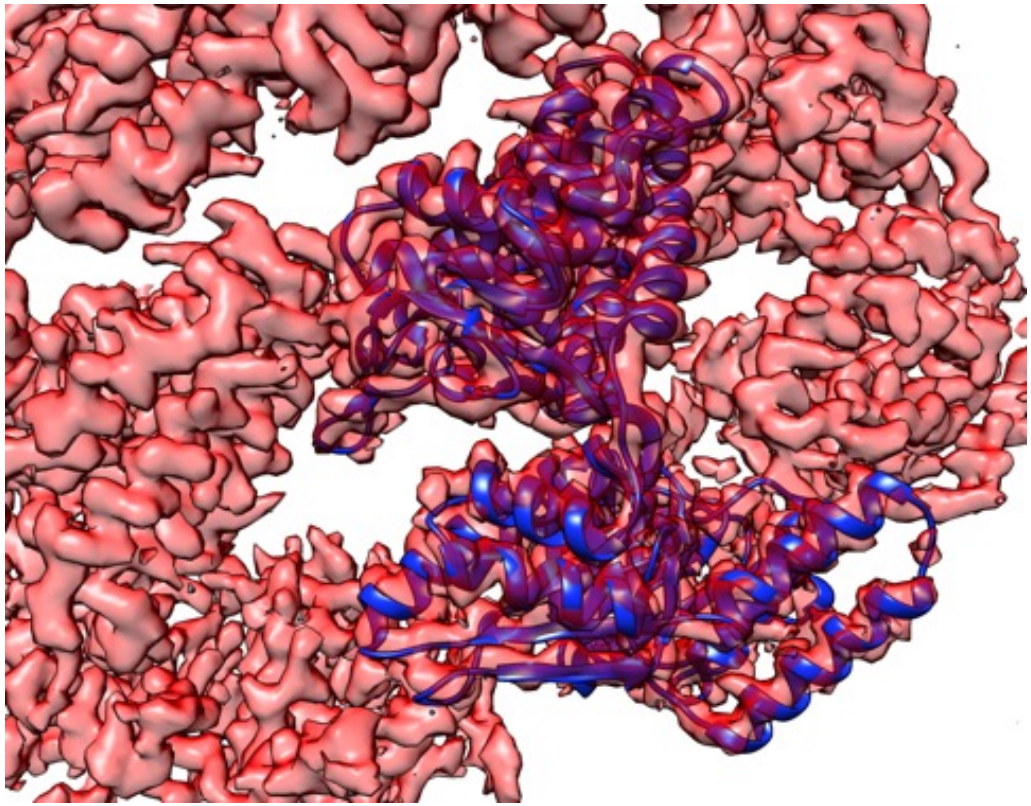
- Multiple chains
- Density search
- Symmetry
- Multiprocessing

Docking models with dock-in-map

EMD_8750



1ss8 chain A



1ss8 chain A docked in map

The *Phenix* Project

Lawrence Berkeley Laboratory

Paul Adams, Pavel Afonine, Nigel Moriarty, Billy Poon, Oleg Sobolev, Dorothee Liebschner, Christopher Schlicksup

Los Alamos National Laboratory

Tom Terwilliger, Li-Wei Hung

Duke University

Jane & David Richardson, Chris Williams, Bradley Hintze, Vincent Chen

Cambridge University

Randy Read, Airlie McCoy, Tristan Croll, Rob Oeffner, Massimo Sammito, Claudia Millán Nebot

University of Texas

Matthew Baker, Cory Hryc

An NIH/NIGMS funded Program Project