Phenix

Molecular replacement and model-building using distant homology models as templates

ACA Trueblood Award Lecture
July 24, 2013
Honolulu, Hawaii

Tom Terwilliger
Los Alamos National Laboratory

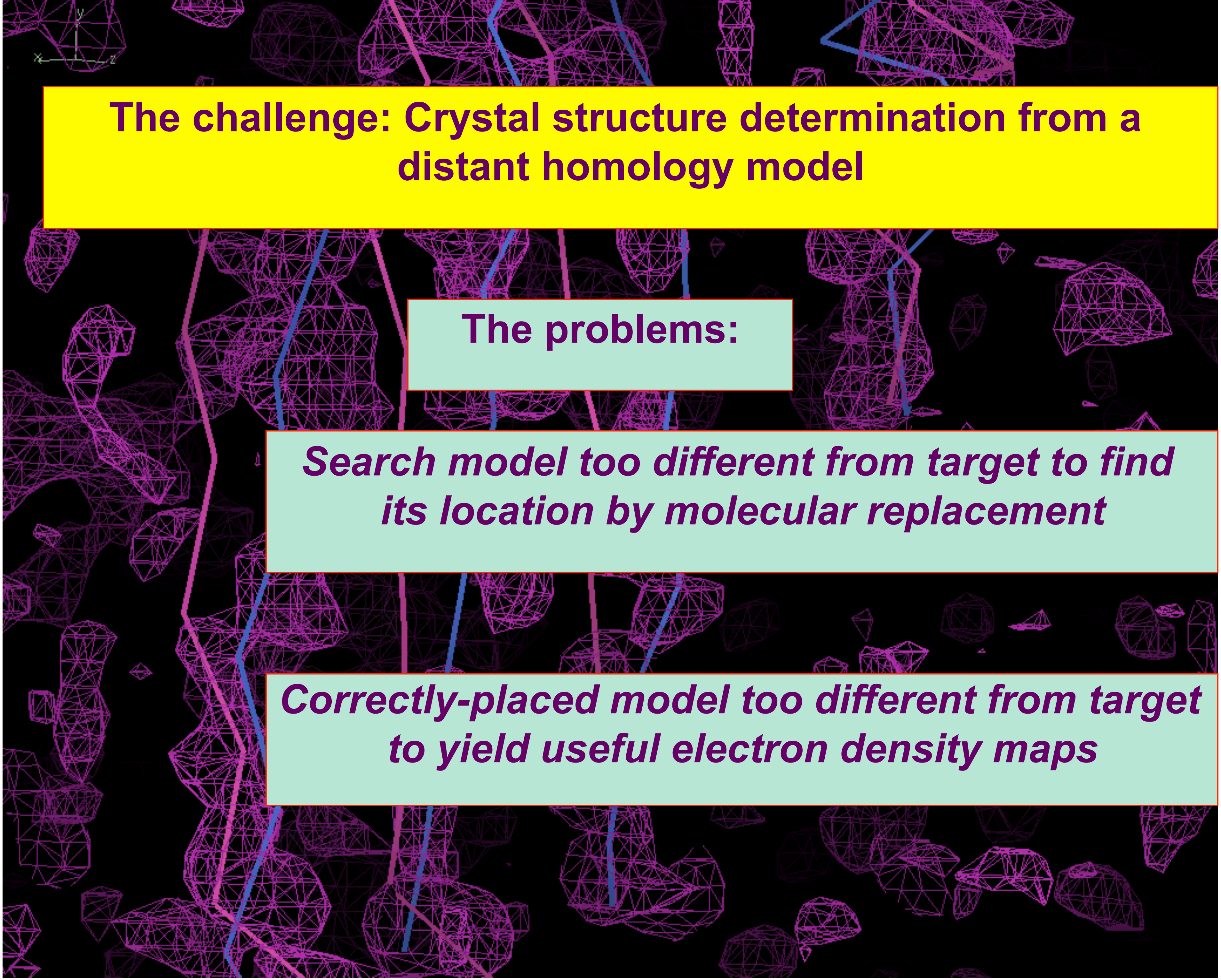
Kenneth Trueblood
1920-1998

Los Alamos
NATIONAL LABORATORY

BERKELEY LAB

UNIVERSITY OF
CAMBRIDGE

The challenge: Crystal structure determination from a distant homology model

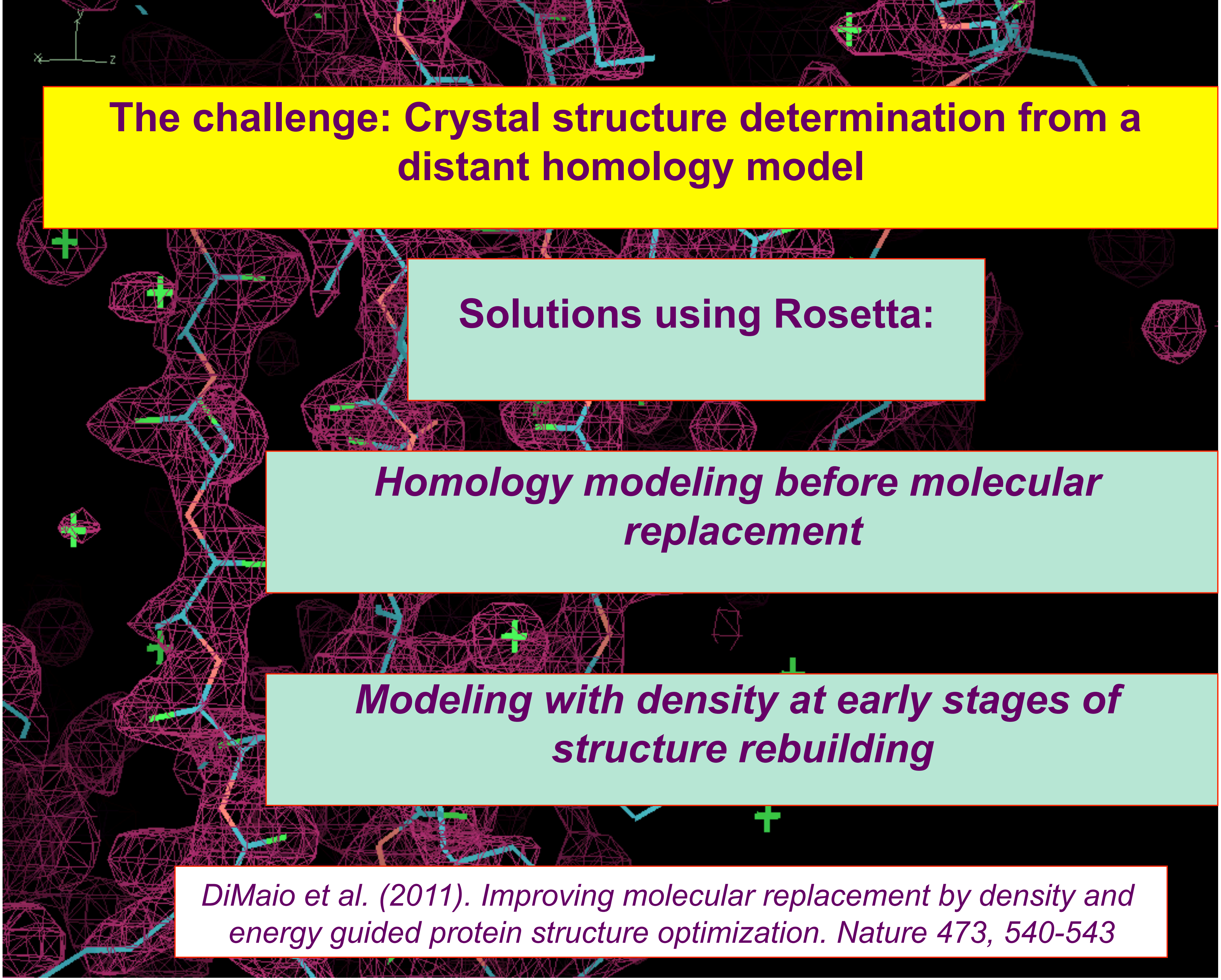


The problems:

Search model too different from target to find its location by molecular replacement

Correctly-placed model too different from target to yield useful electron density maps

The challenge: Crystal structure determination from a distant homology model



Solutions using Rosetta:

Homology modeling before molecular replacement

Modeling with density at early stages of structure rebuilding

DiMaio et al. (2011). Improving molecular replacement by density and energy guided protein structure optimization. Nature 473, 540-543

Complementarity of PHENIX and Rosetta model-building

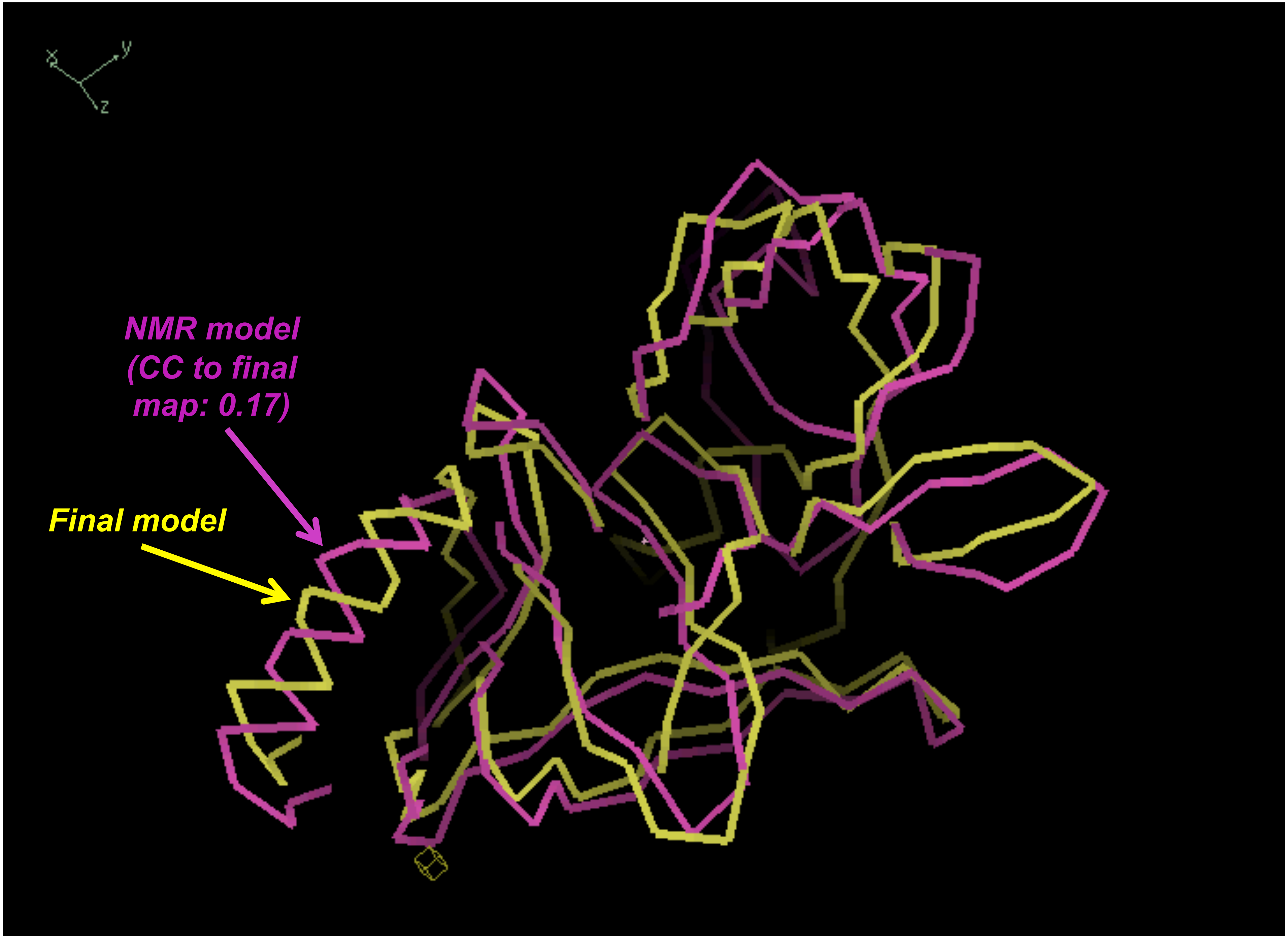
Characteristic	*Crystallographic model-building (PHENIX)*	*Structure-modeling (Rosetta)*
Optimization	Interpretation of patterns of density	Creating physically plausible models
Model-building approach	Density search for regular secondary structure	Ab-initio modeling or homology modeling
Fragment libraries	3-residue fragment library	3- and 9-residue libraries
Model-building target	Fit to density	Rosetta force field (optional density term)
Refinement target	Structure-factor likelihood refinement target	Rosetta force field (optional density term)

Phenix

Generating improved homology models for molecular replacement with Rosetta

ag9603
NMR model, 100% identity
1.7 Å data

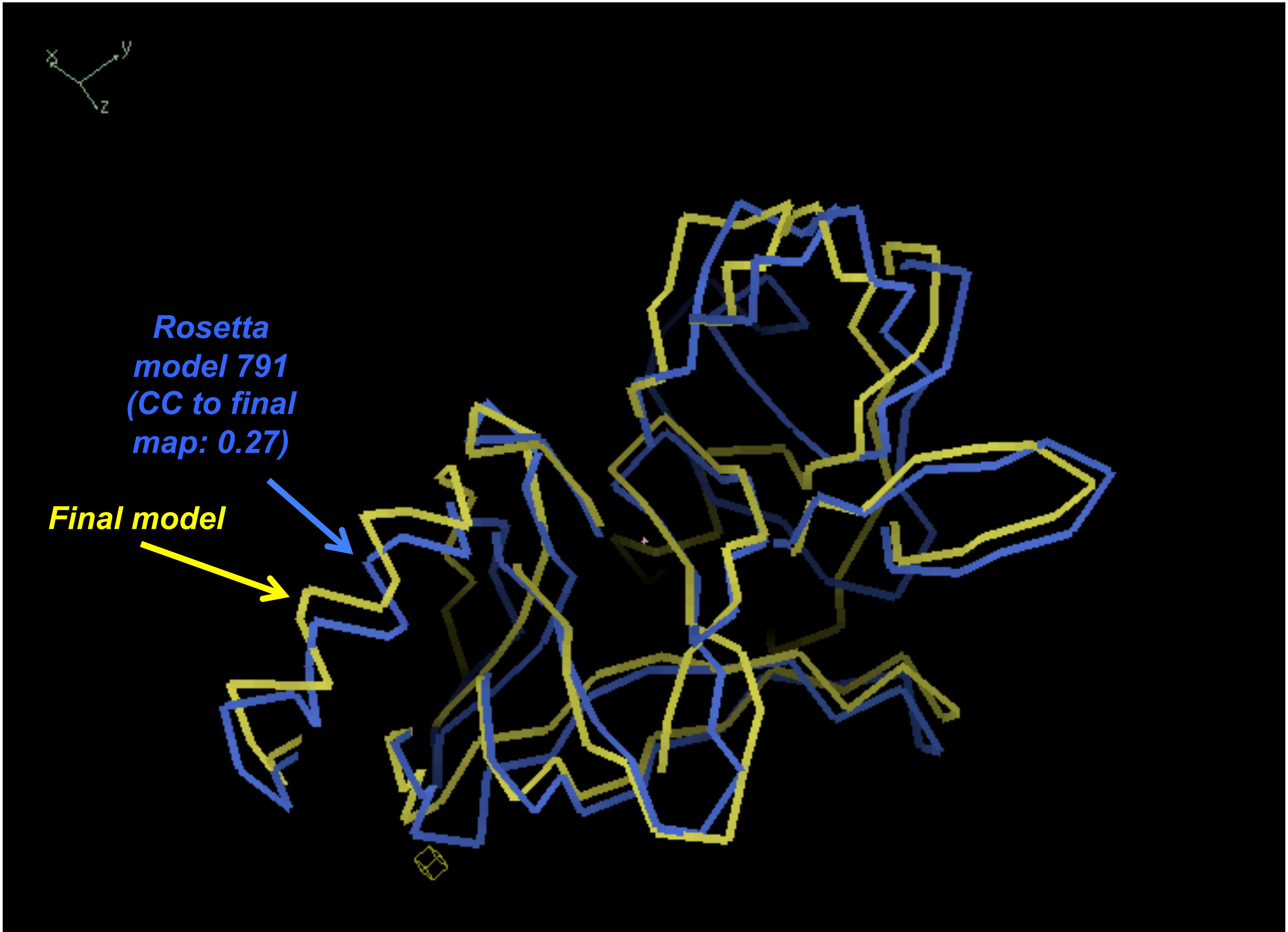
NMR model of of ag9603



NMR model (CC to final map: 0.17)

Final model

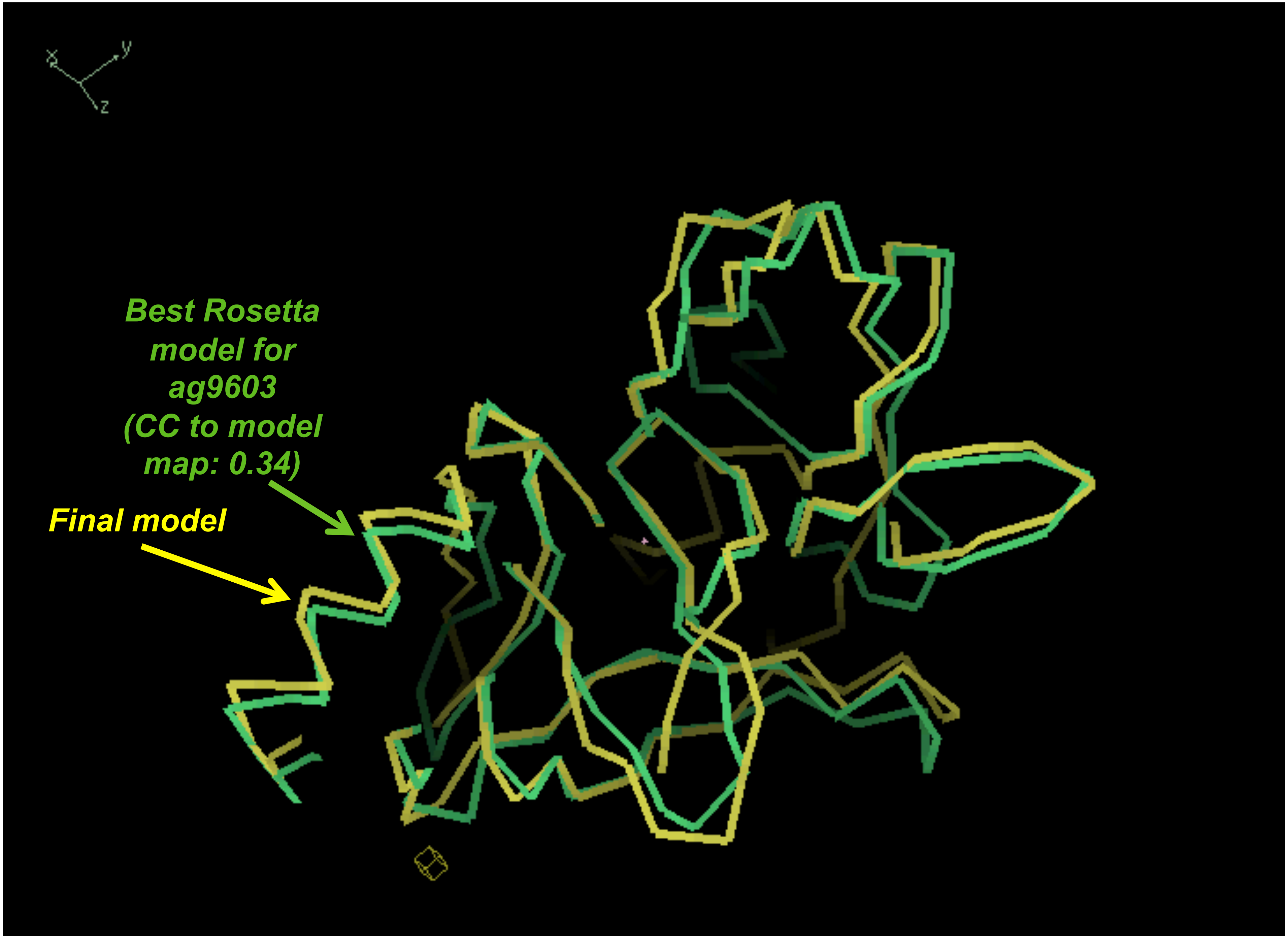
Typical Rosetta model of of ag9603



Rosetta model 791 (CC to final map: 0.27)

Final model

Best Rosetta model of of ag9603

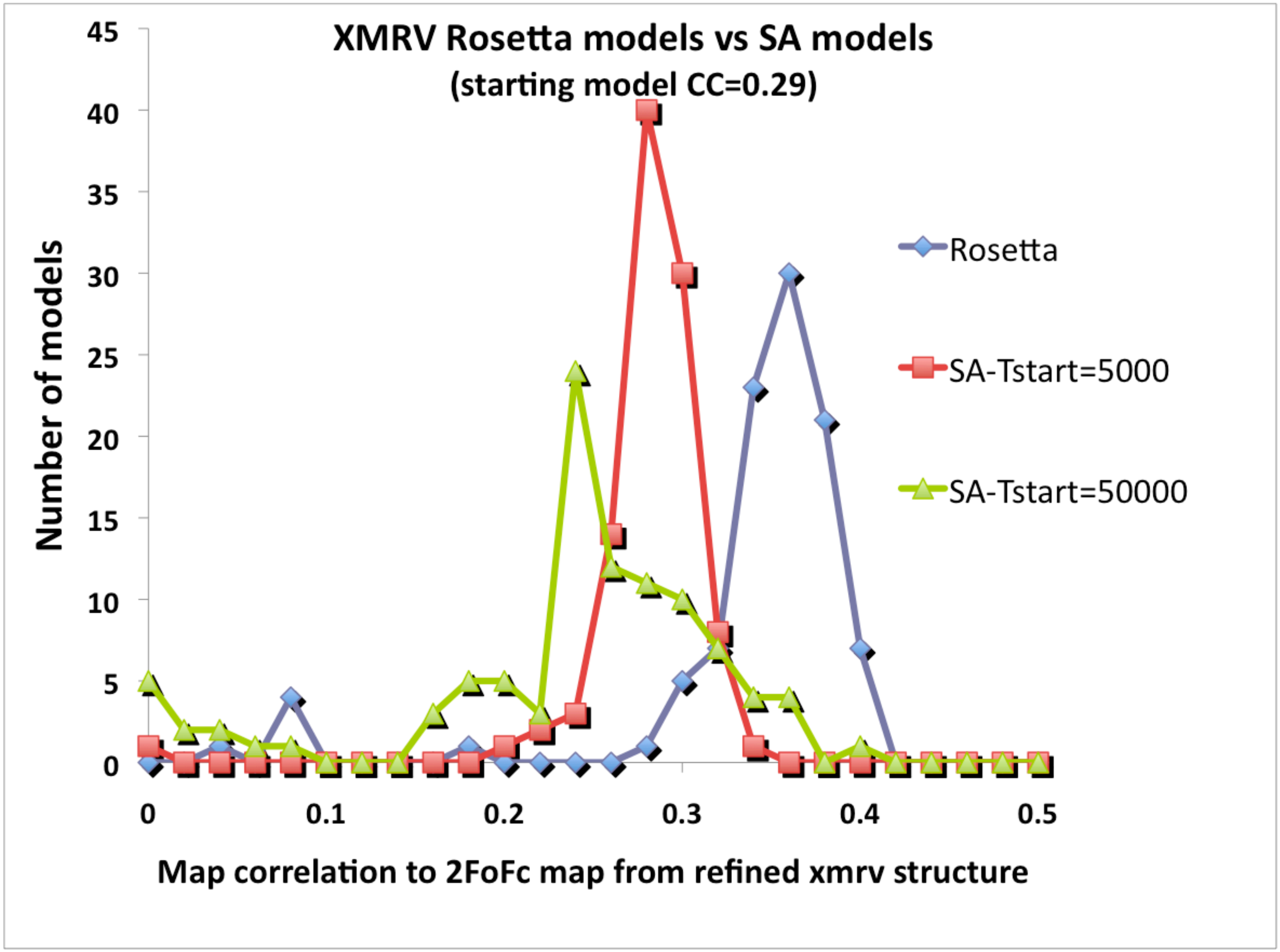


Improving models with Rosetta using density

Density fit as part of optimization target
Modeling segments not in template

Comparison with simulated annealing

Rosetta models vs SA models

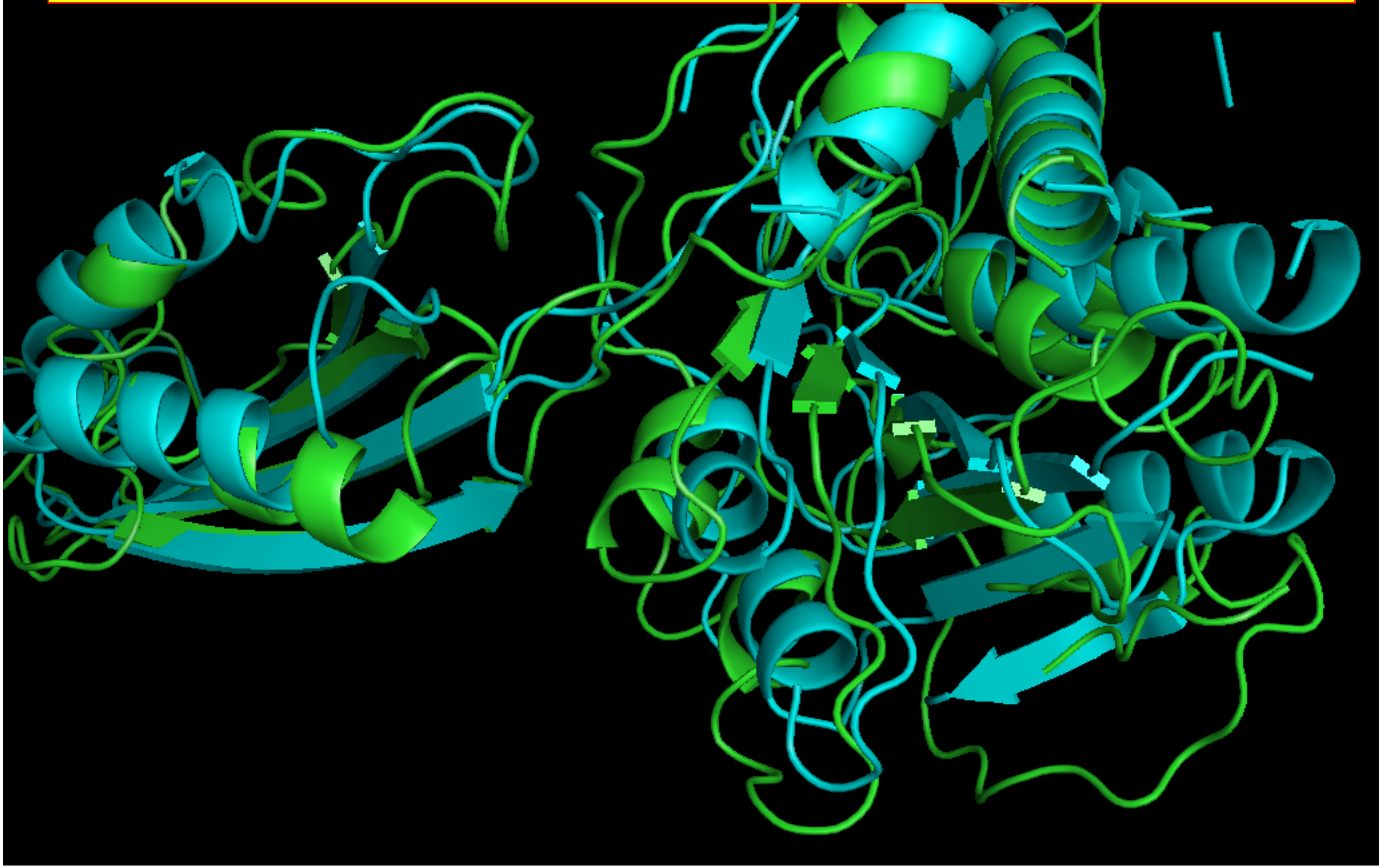


Integrating Rosetta modeling in Phenix

phenix.mr_rosetta

Molecular replacement
Rebuilding with Rosetta
Rebuilding with phenix.autobuild

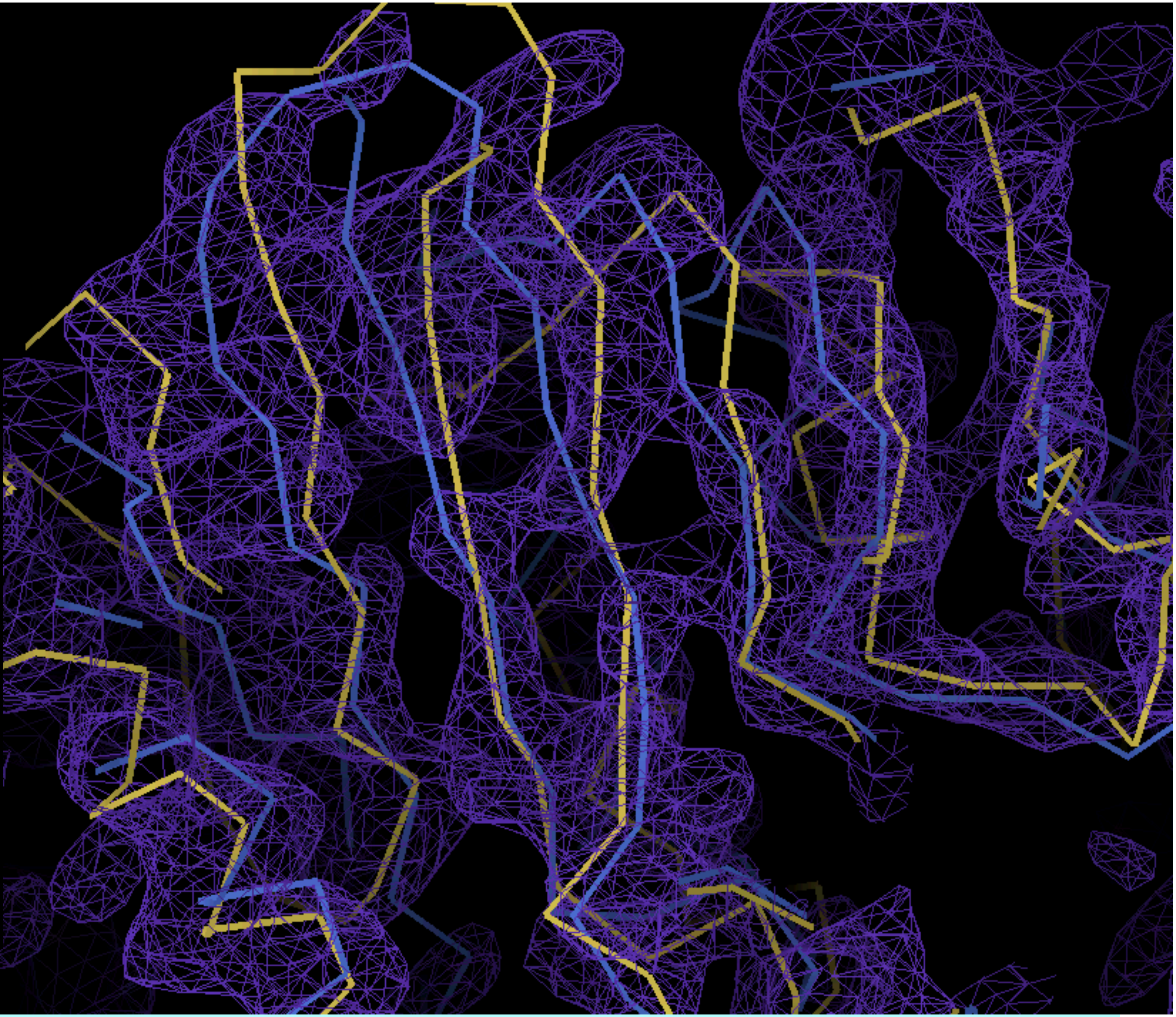
mr_rosetta example:
HP3342, 22% identity template, 3.2 Å data, B=87 A^2



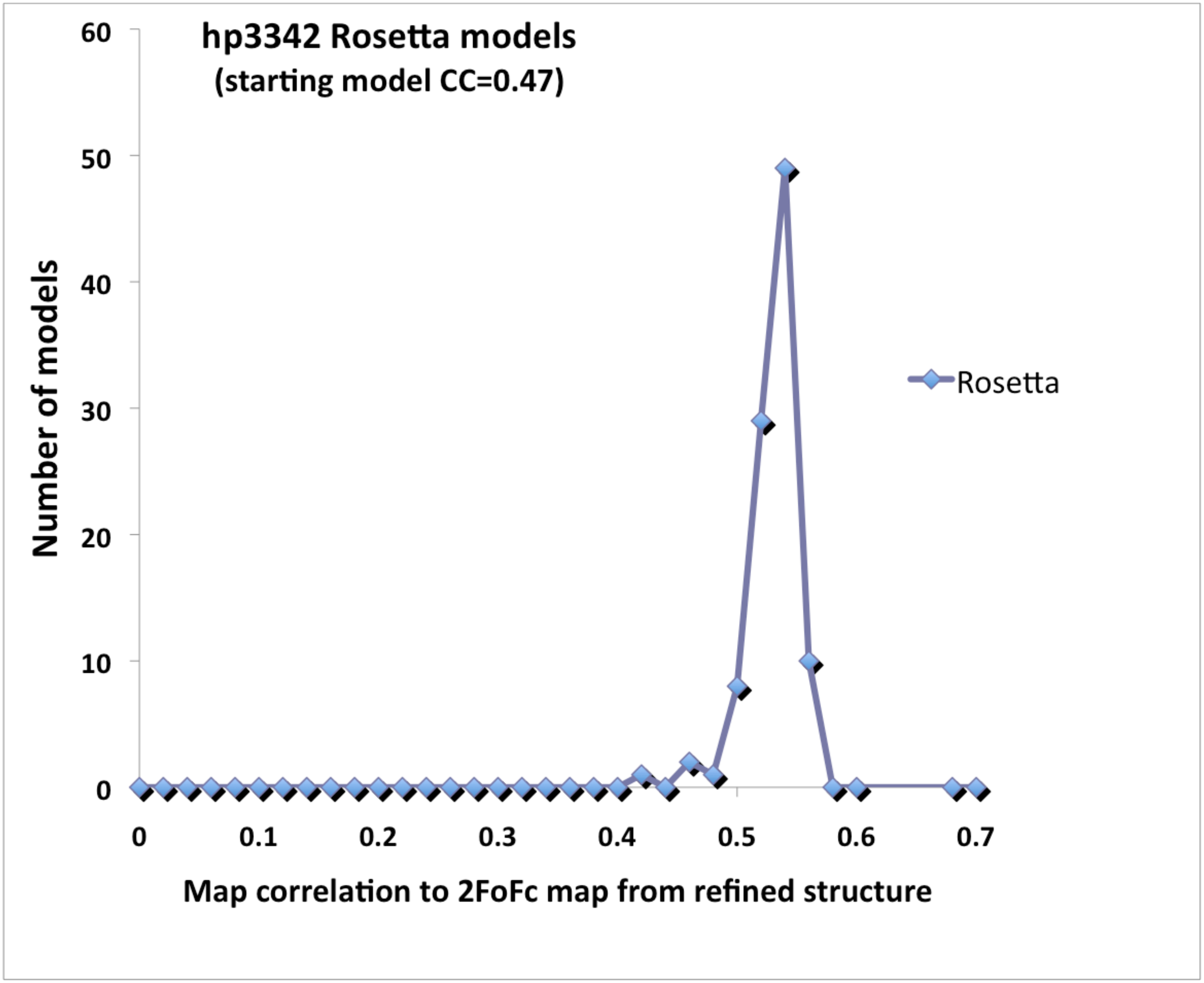
hp3342, 22%
identity
template
(1vgy),
3.2 Å data

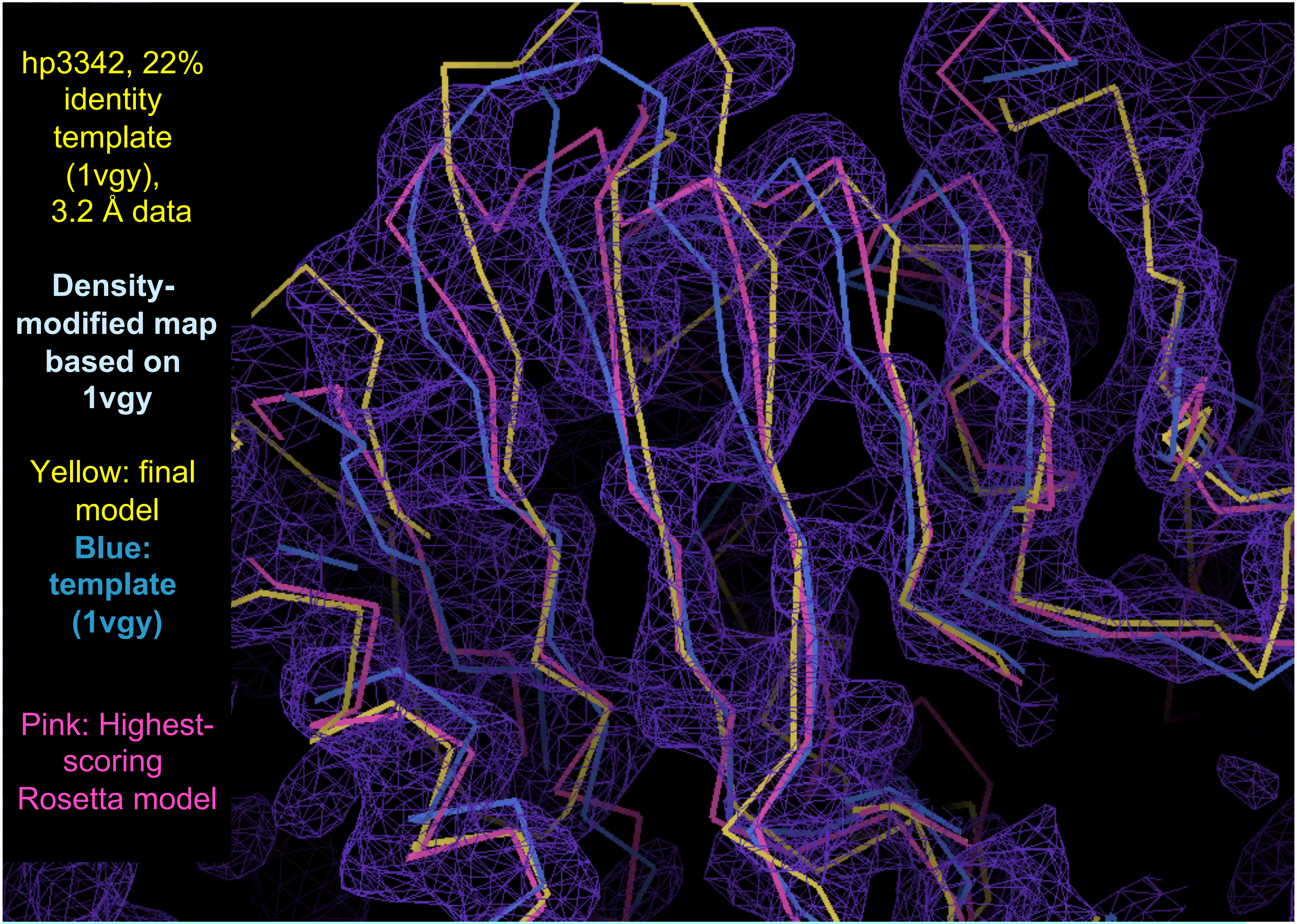
**Density-
modified map
based on
1vgy**

Yellow: final
model
**Blue:
template
(1vgy)**



MR_rosetta example: hp3342





hp3342, 22% identity template (1vgy), 3.2 Å data

Density-modified map based on 1vgy

Yellow: final model

Blue: template (1vgy)

Pink: Highest-scoring Rosetta model

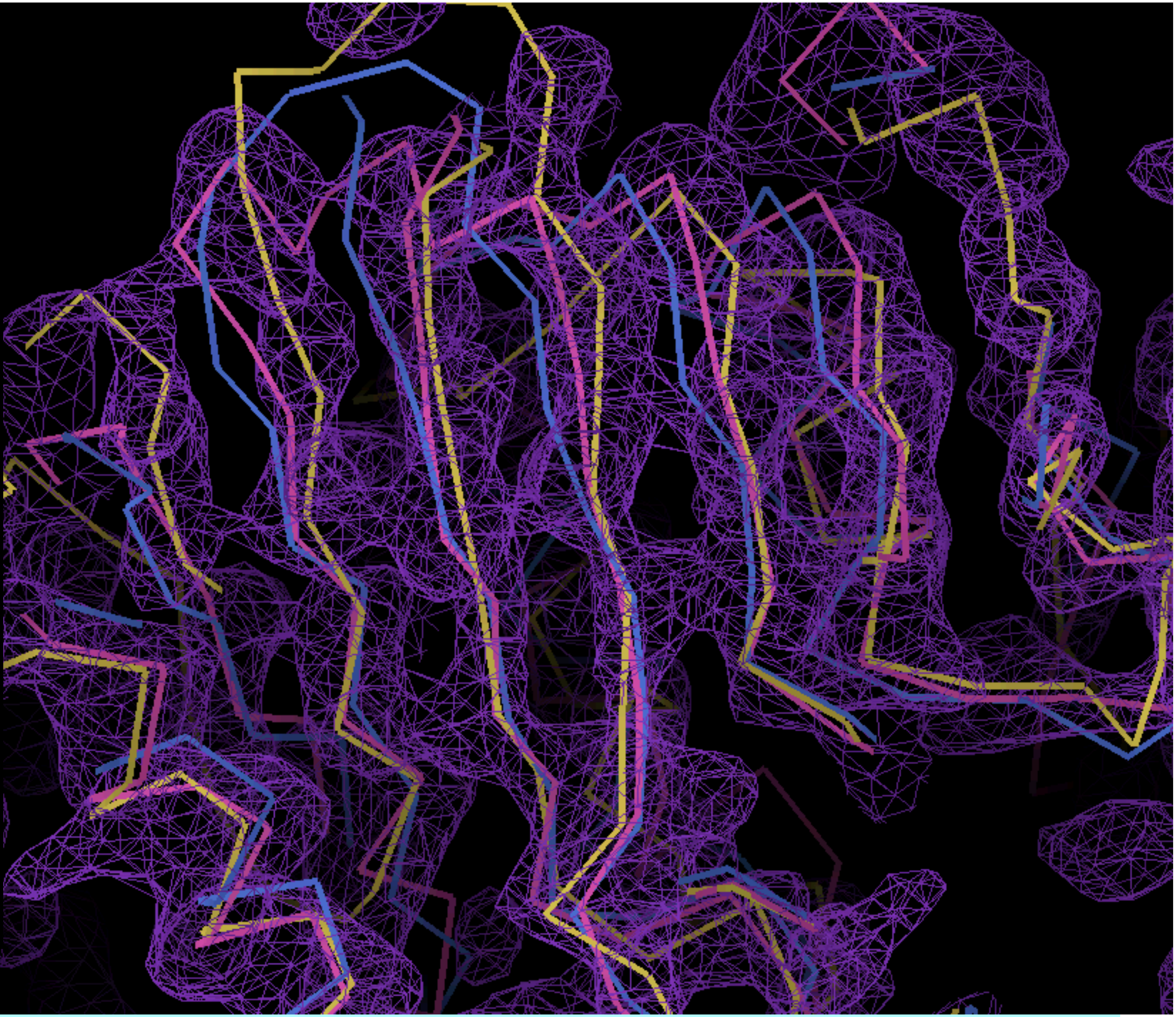
hp3342, 22% identity template (1vgy), 3.2 Å data

Density-modified map based on Rosetta model

Yellow: final model

Blue: template (1vgy)

Pink: Highest-scoring Rosetta model

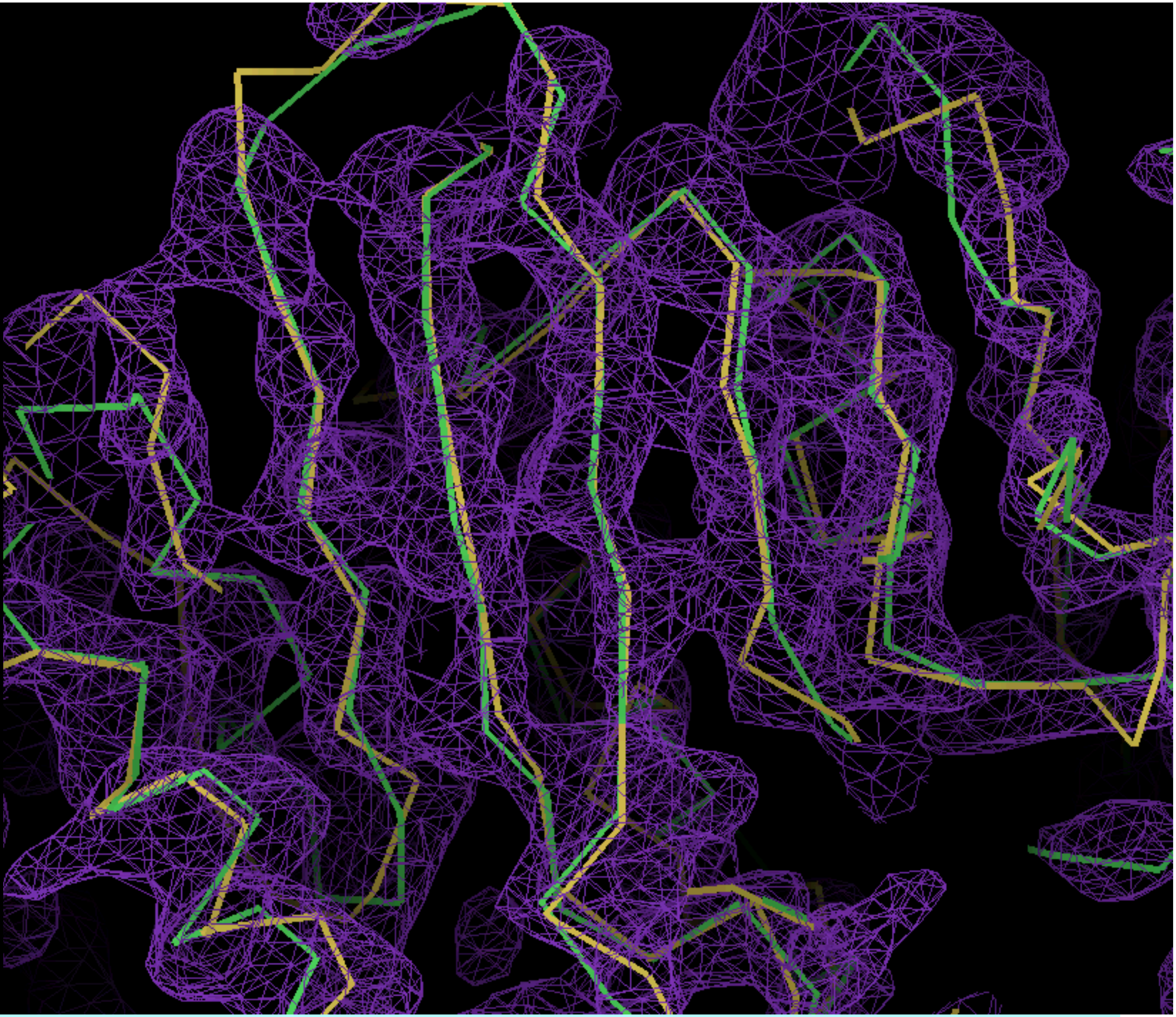


hp3342, 22% identity template (1vgy), 3.2 Å data

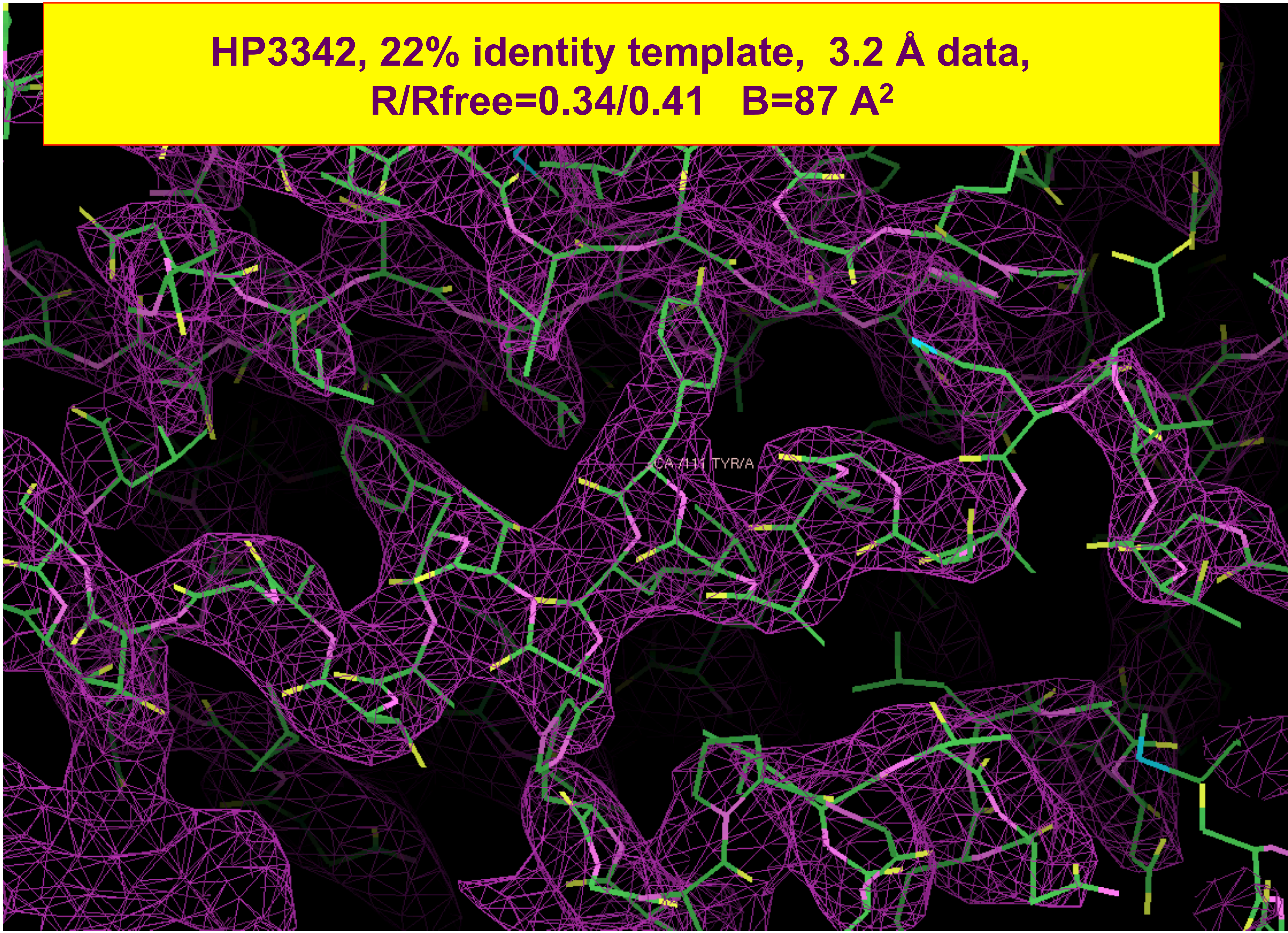
Density-modified map based on Rosetta model

Yellow: final model

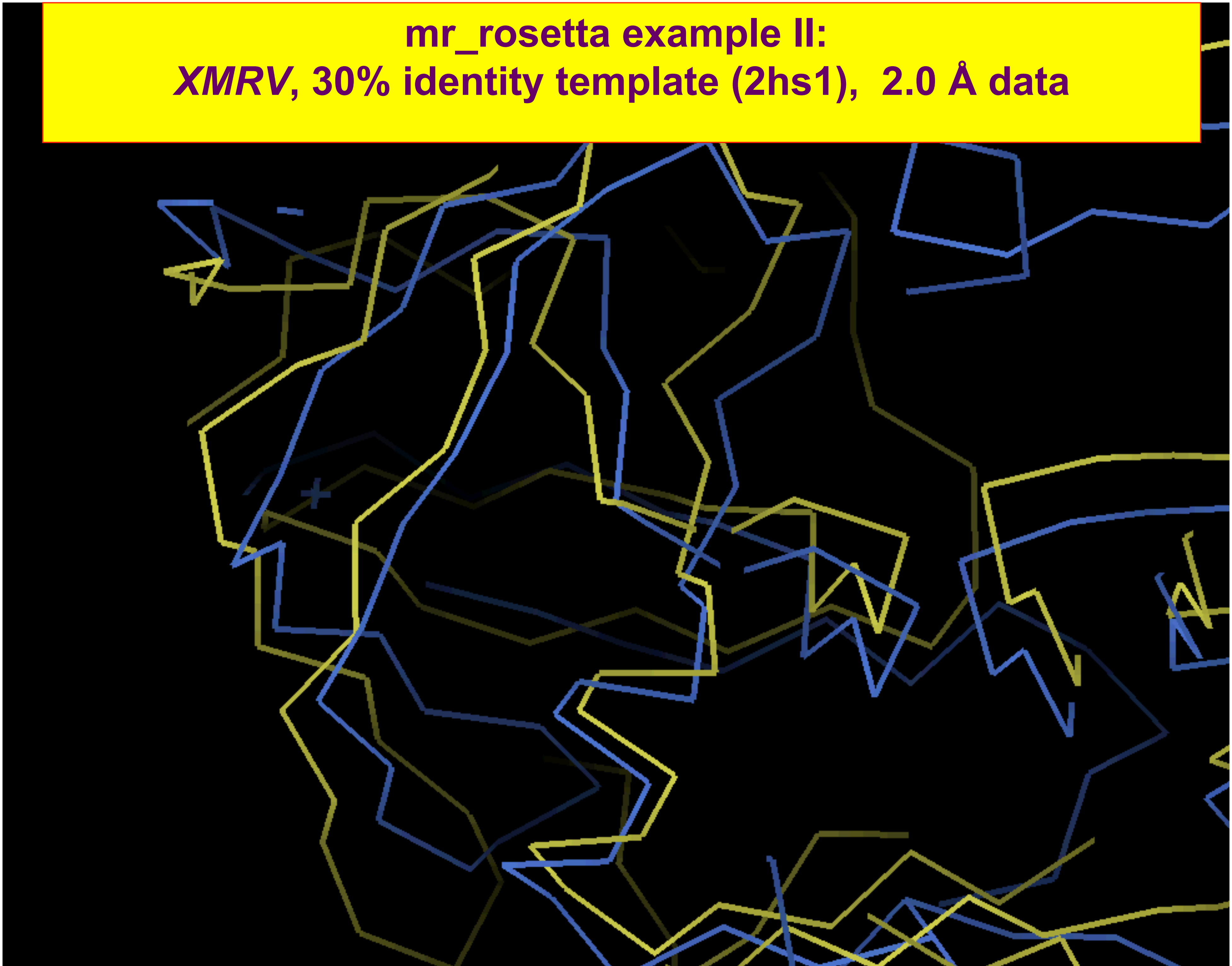
Green: autobuild model



HP3342, 22% identity template, 3.2 Å data, R/Rfree=0.34/0.41 B=87 A^2



mr_rosetta example II:
XMRV, 30% identity template (2hs1), 2.0 Å data

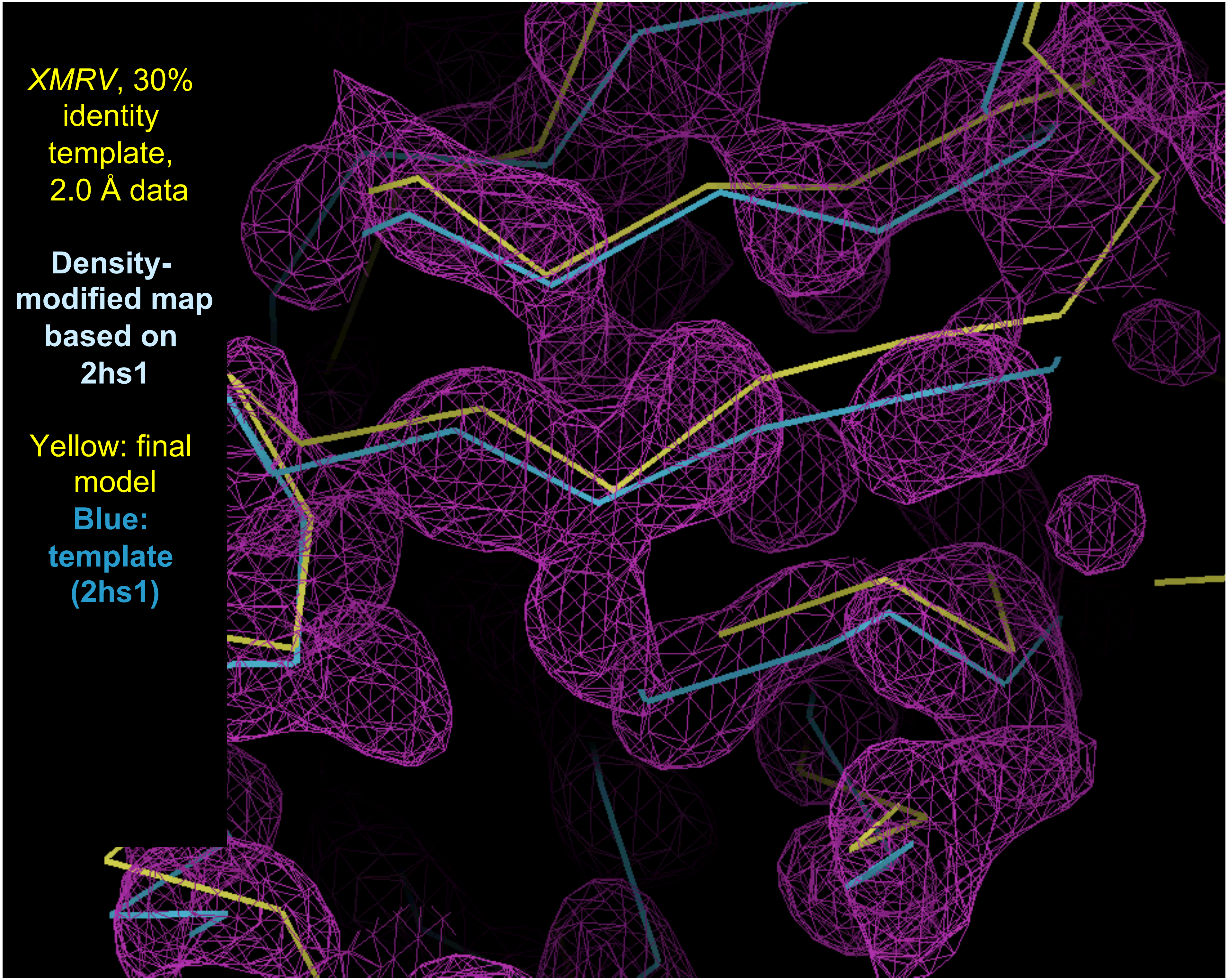


XMRV, 30% identity template, 2.0 Å data

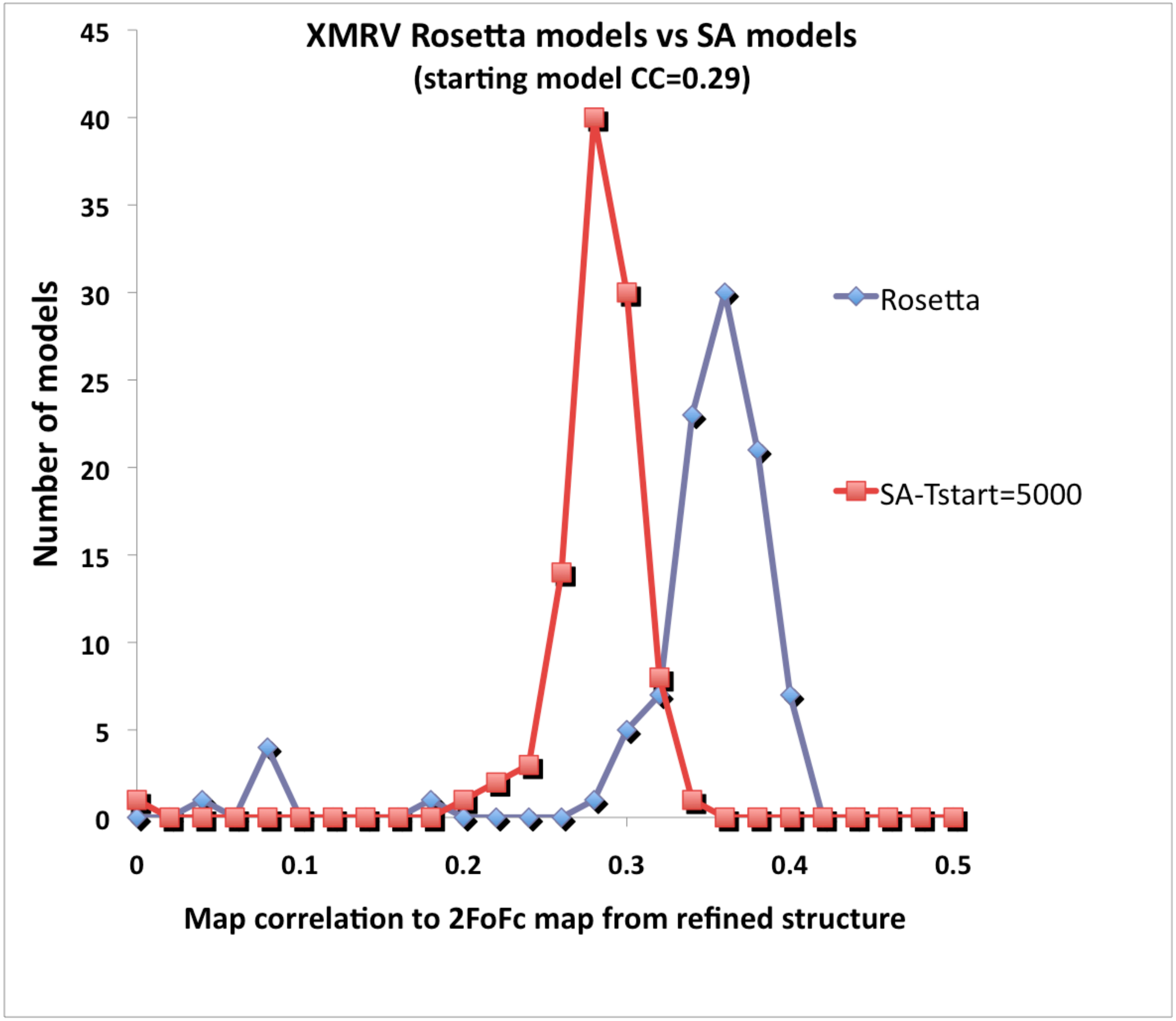
Density-modified map based on 2hs1

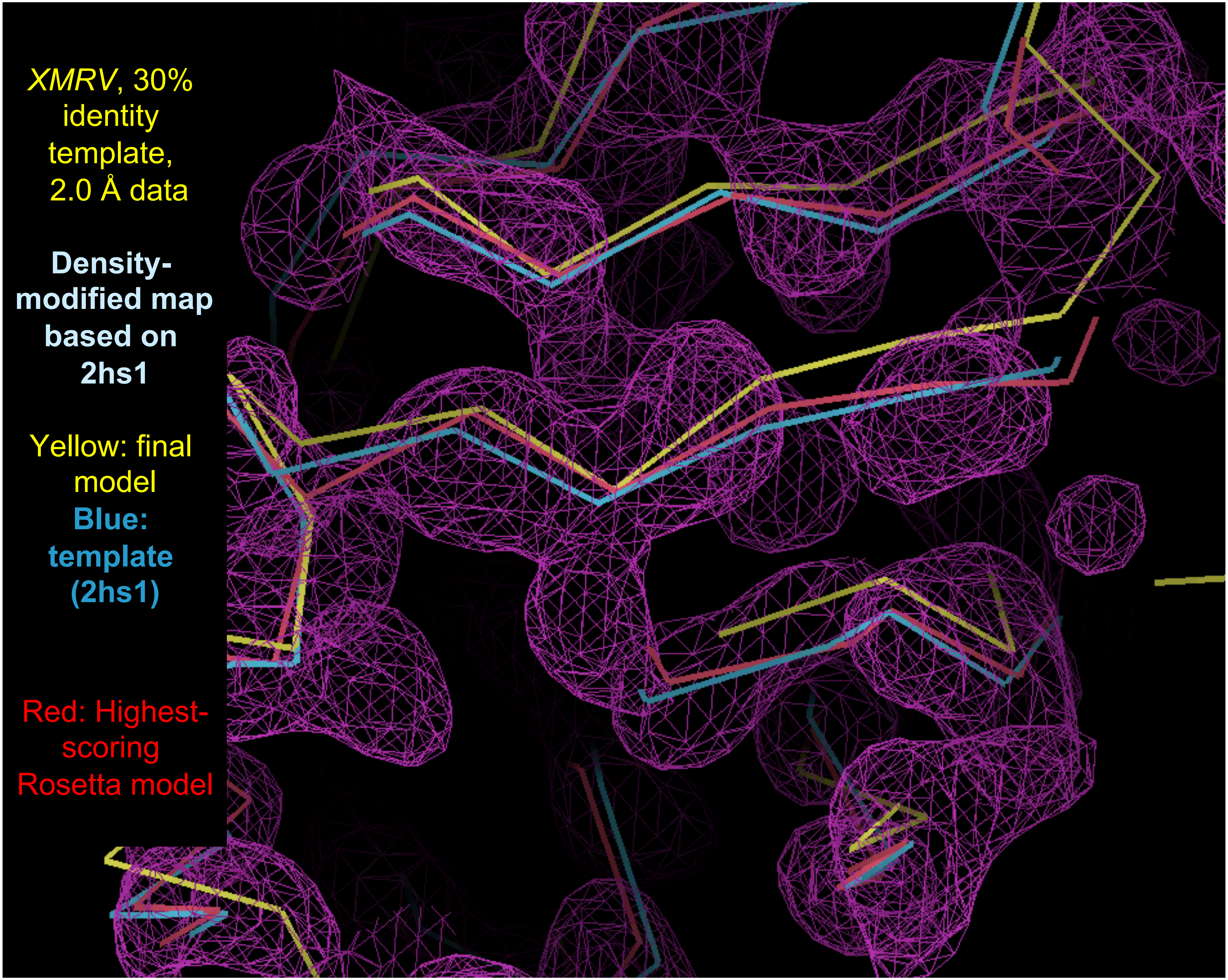
Yellow: final model

Blue: template (2hs1)



MR_rosetta example: *XMRV*





XMRV, 30% identity template, 2.0 Å data

Density-modified map based on 2hs1

Yellow: final model

Blue: template (2hs1)

Red: Highest-scoring Rosetta model

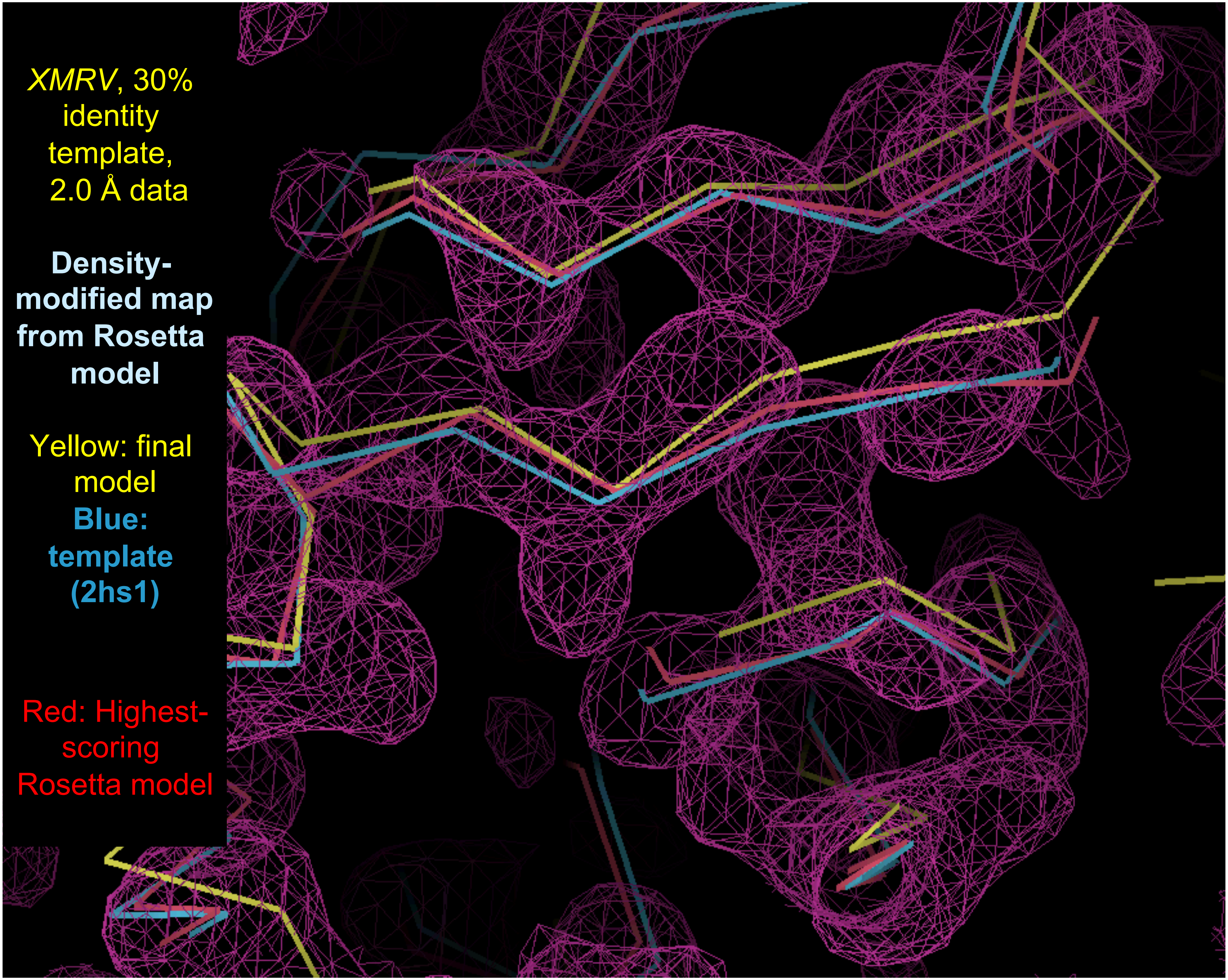
XMRV, 30% identity template, 2.0 Å data

Density-modified map from Rosetta model

Yellow: final model

Blue: template (2hs1)

Red: Highest-scoring Rosetta model



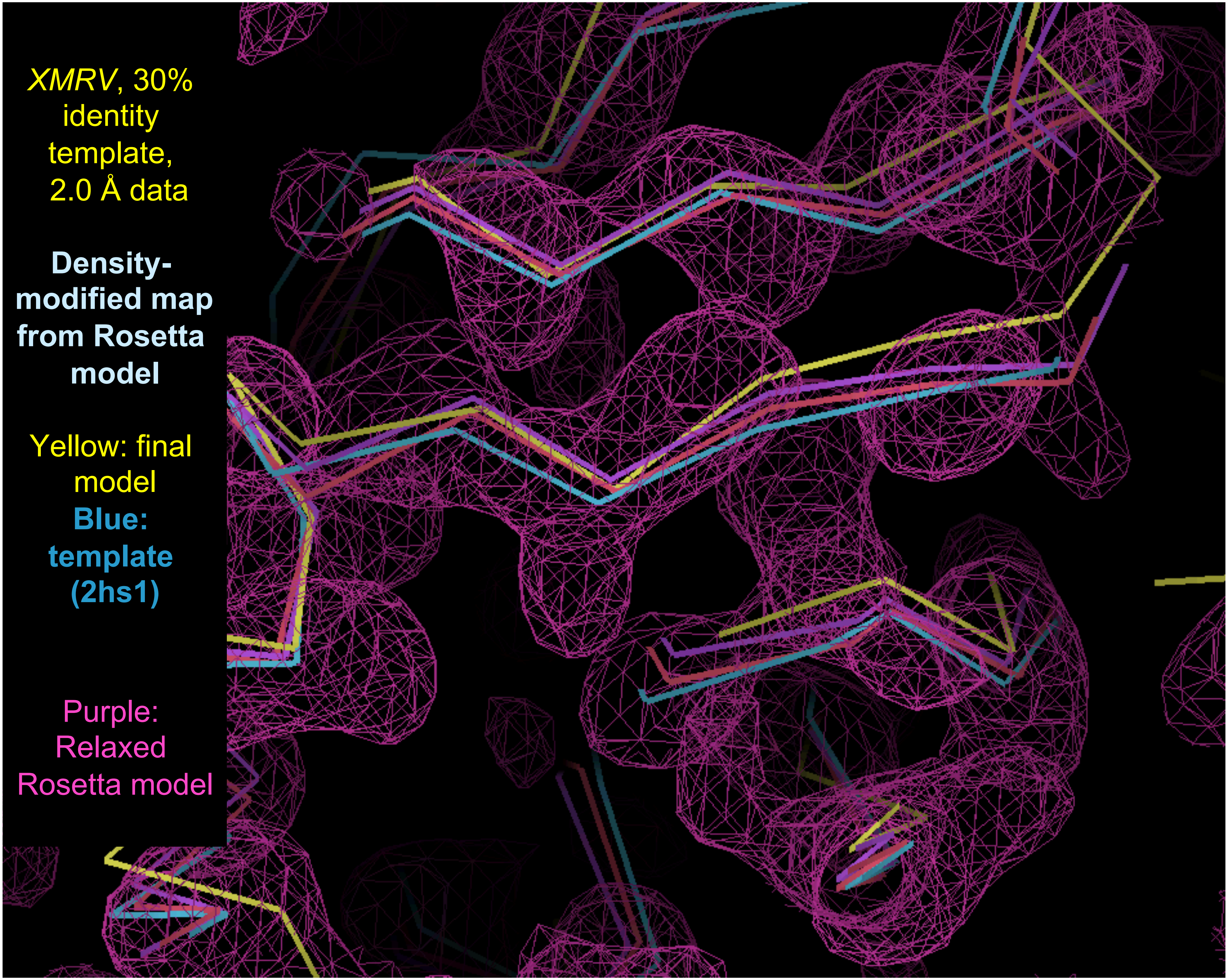
XMRV, 30% identity template, 2.0 Å data

Density-modified map from Rosetta model

Yellow: final model

Blue: template (2hs1)

Purple: Relaxed Rosetta model



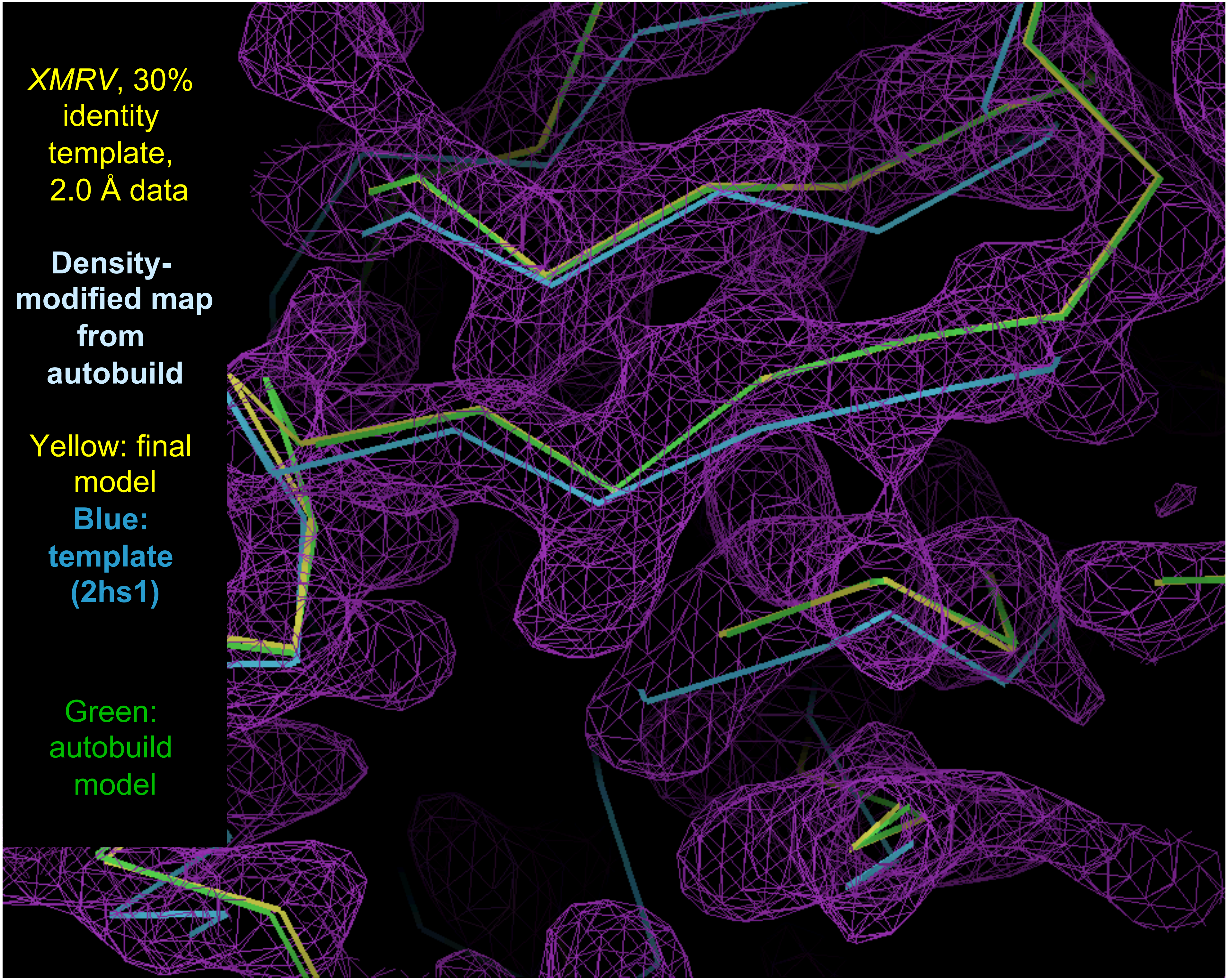
XMRV, 30% identity template, 2.0 Å data

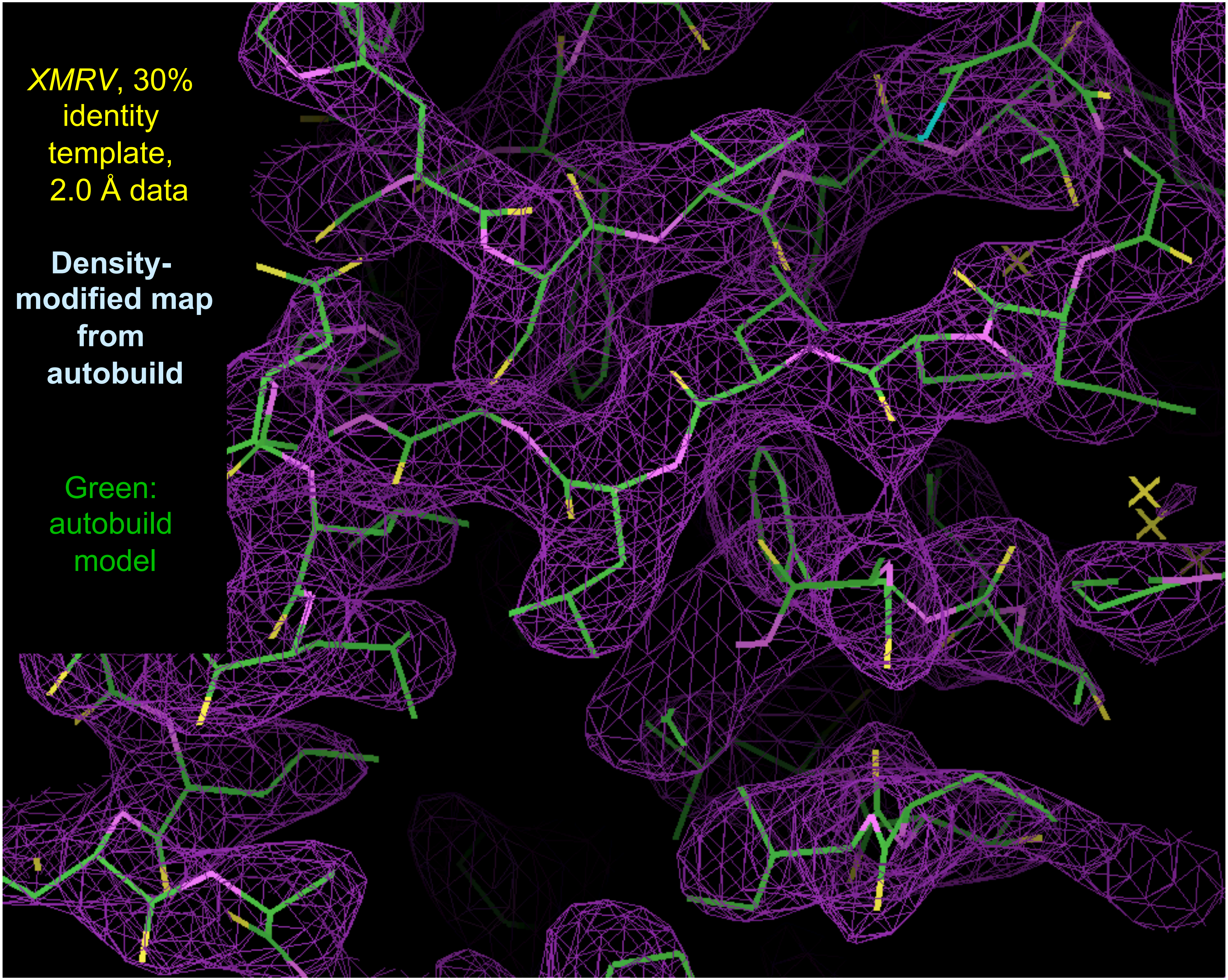
Density-modified map from autobuild

Yellow: final model

Blue: template (2hs1)

Green: autobuild model





XMRV, 30% identity template, 2.0 Å data

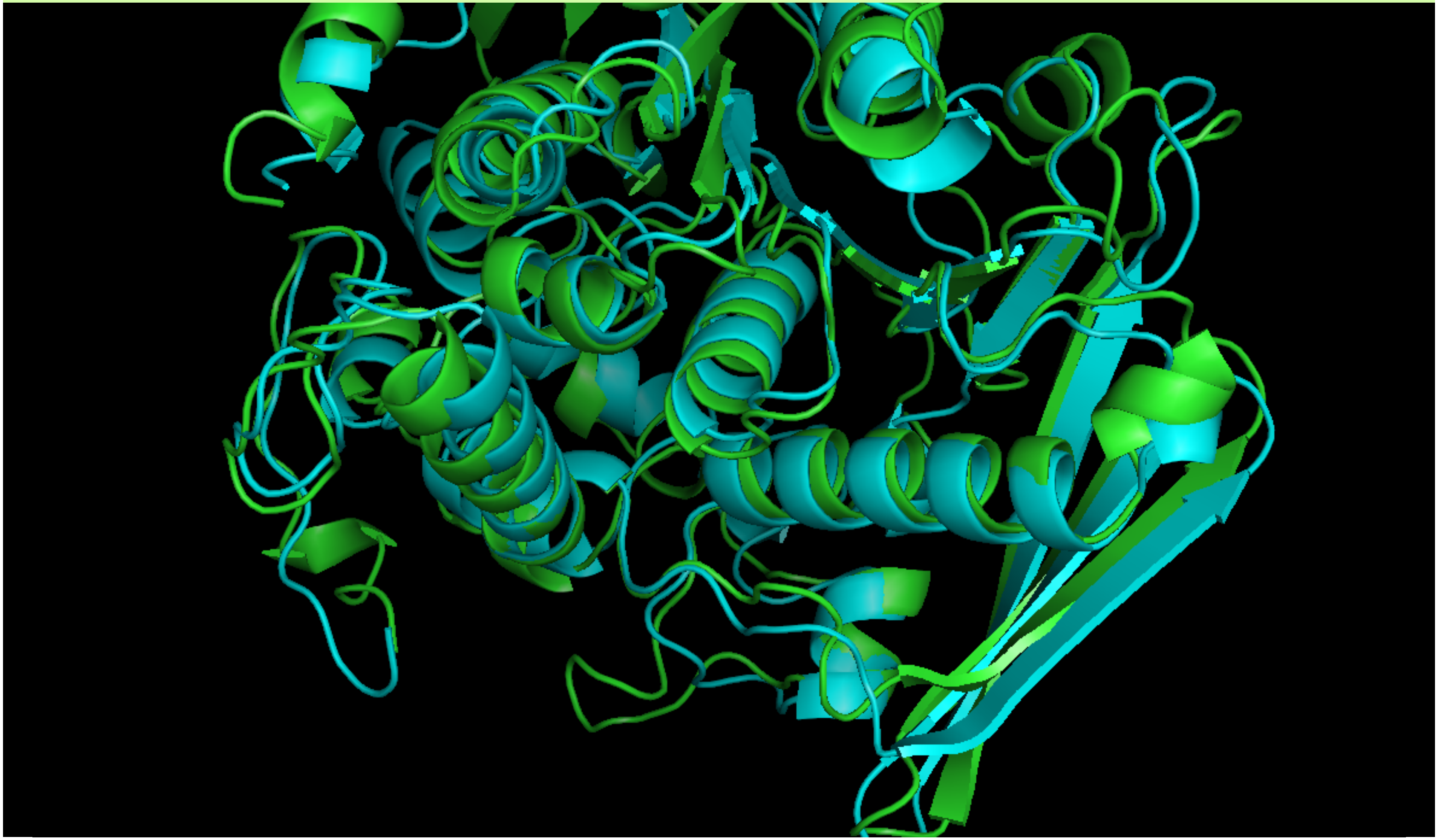
Density-modified map from autobuild

Green: autobuild model

Structure determination of cab55348 (*using template supplied by user)*

1.9 A, 28% sequence identity (AutoMR alone fails with R/Rfree=0.47/0.53)

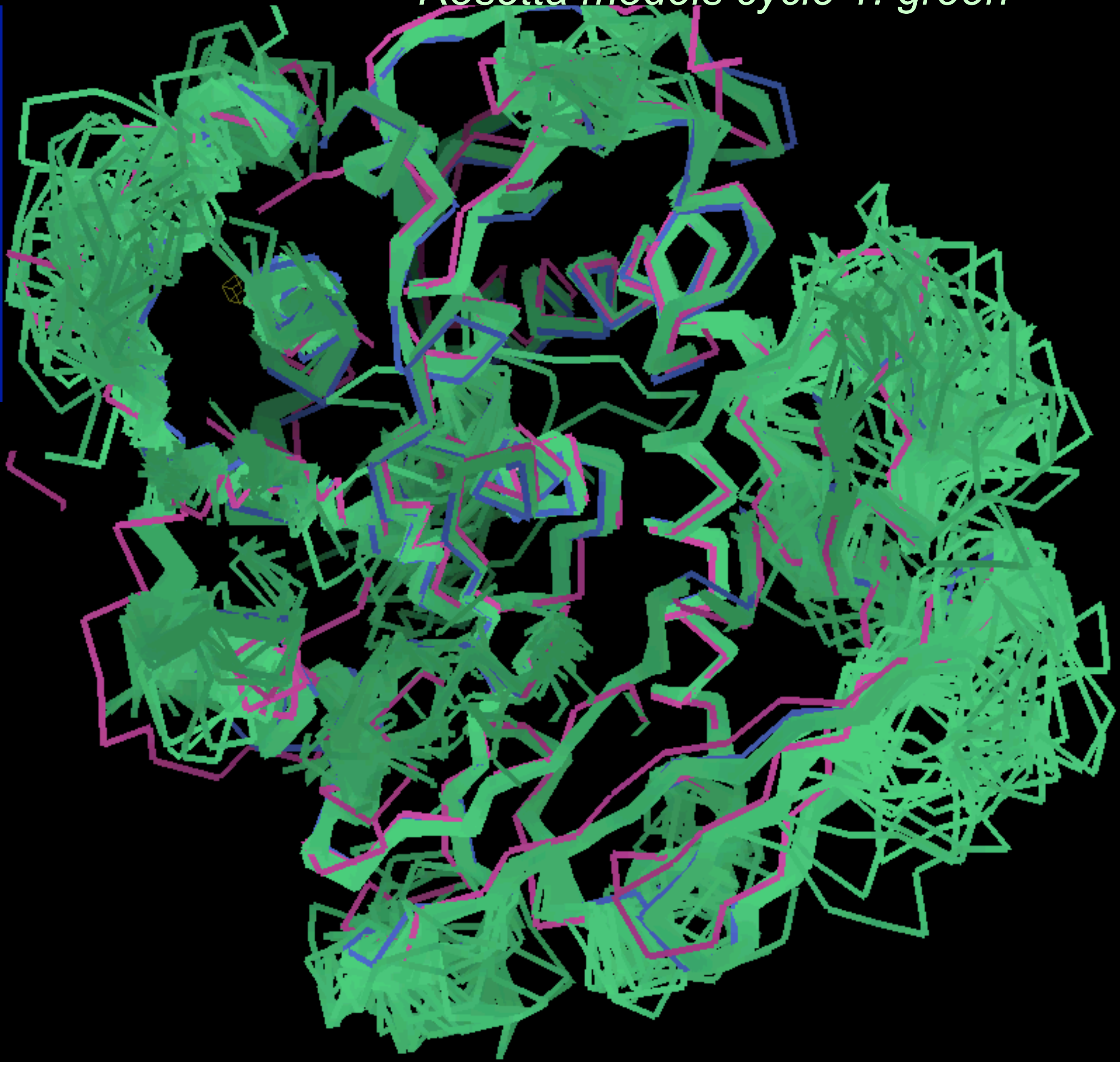
MR model: blue, Final model: green



MR model : blue   Final model: pink

Rosetta models cycle 1: green

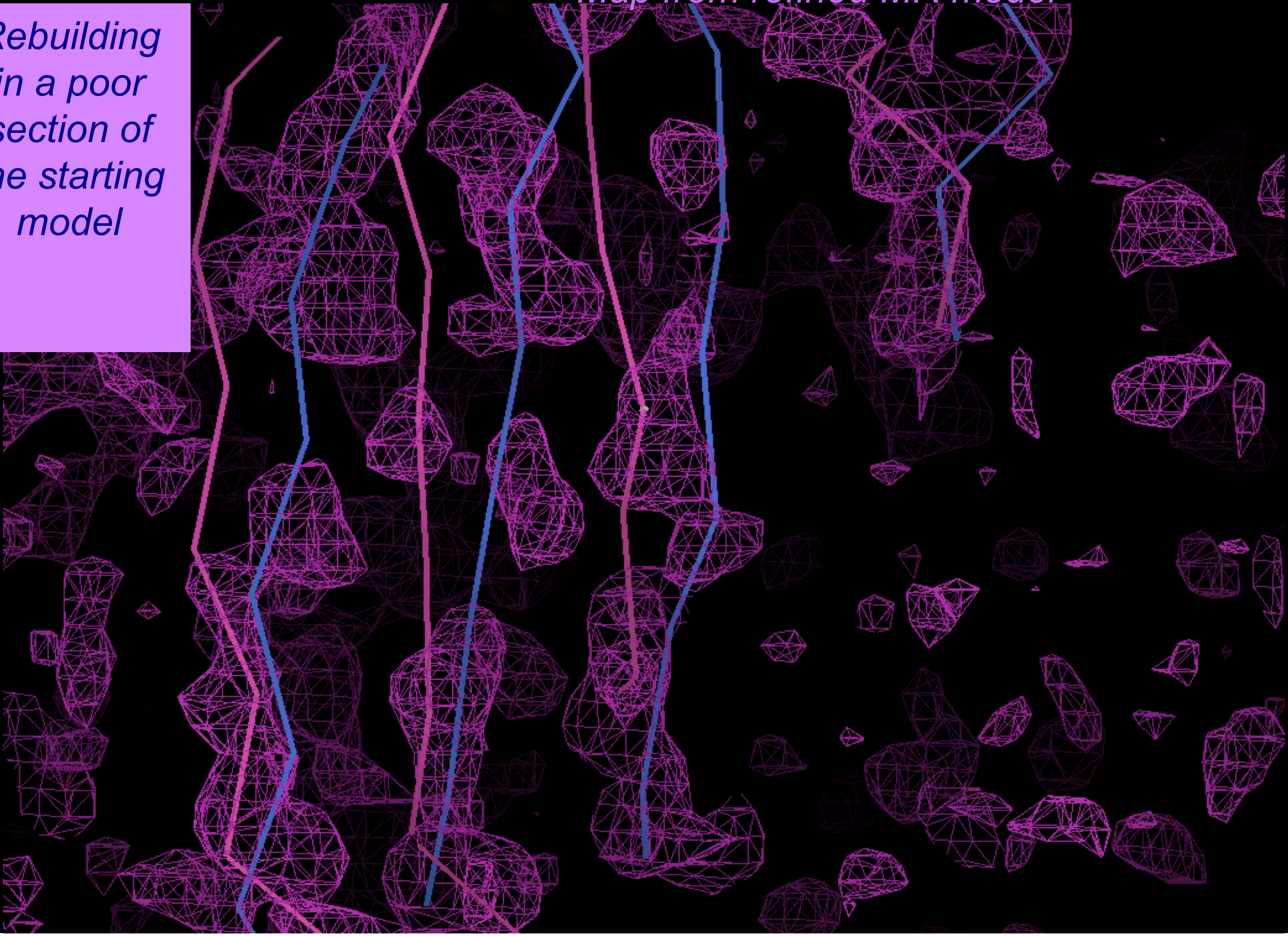
Sample Rosetta models in cycles 1 and 2,



MR model : blue Final model: pink

Map from refined MR model

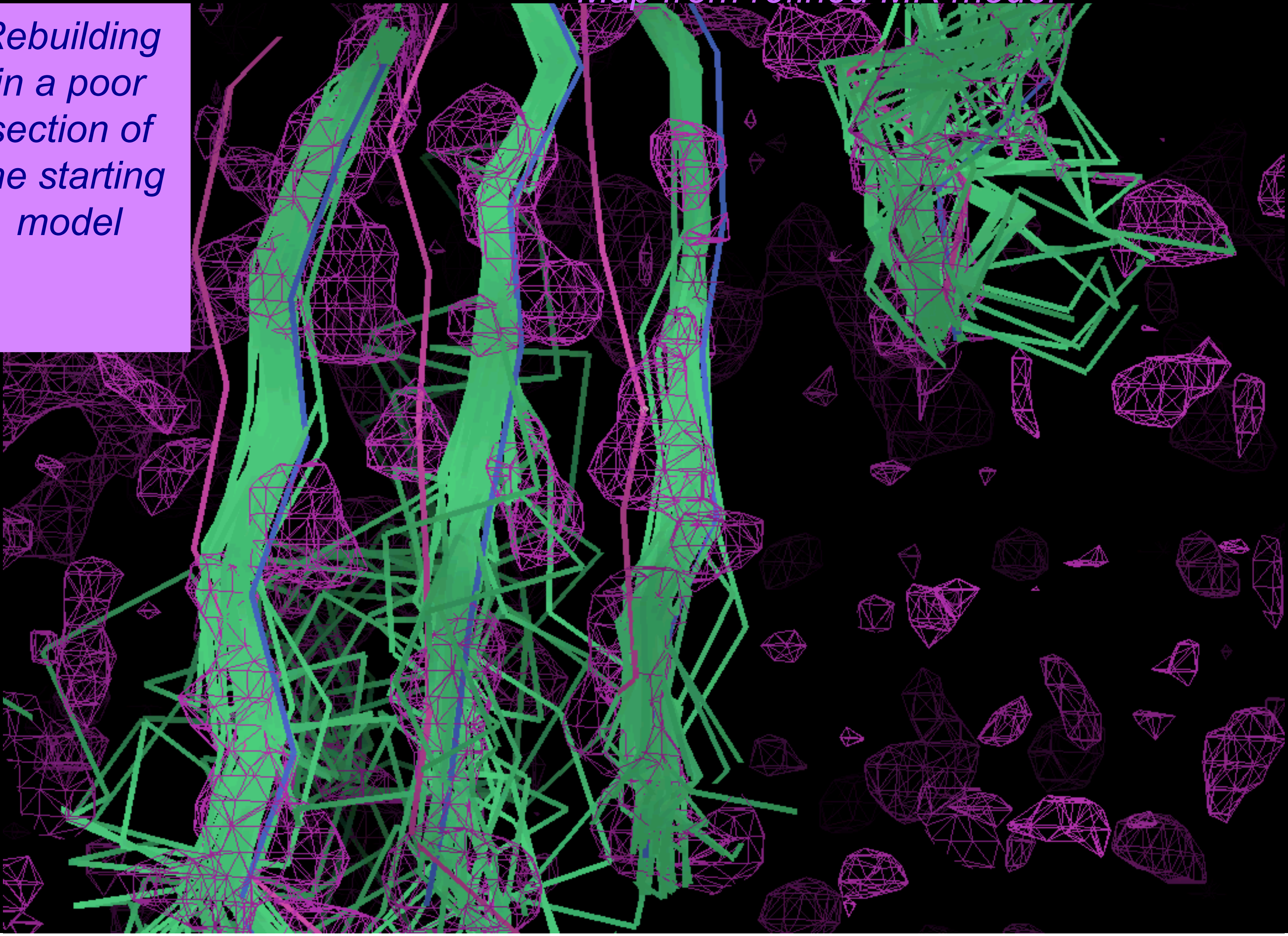
Rebuilding in a poor section of the starting model

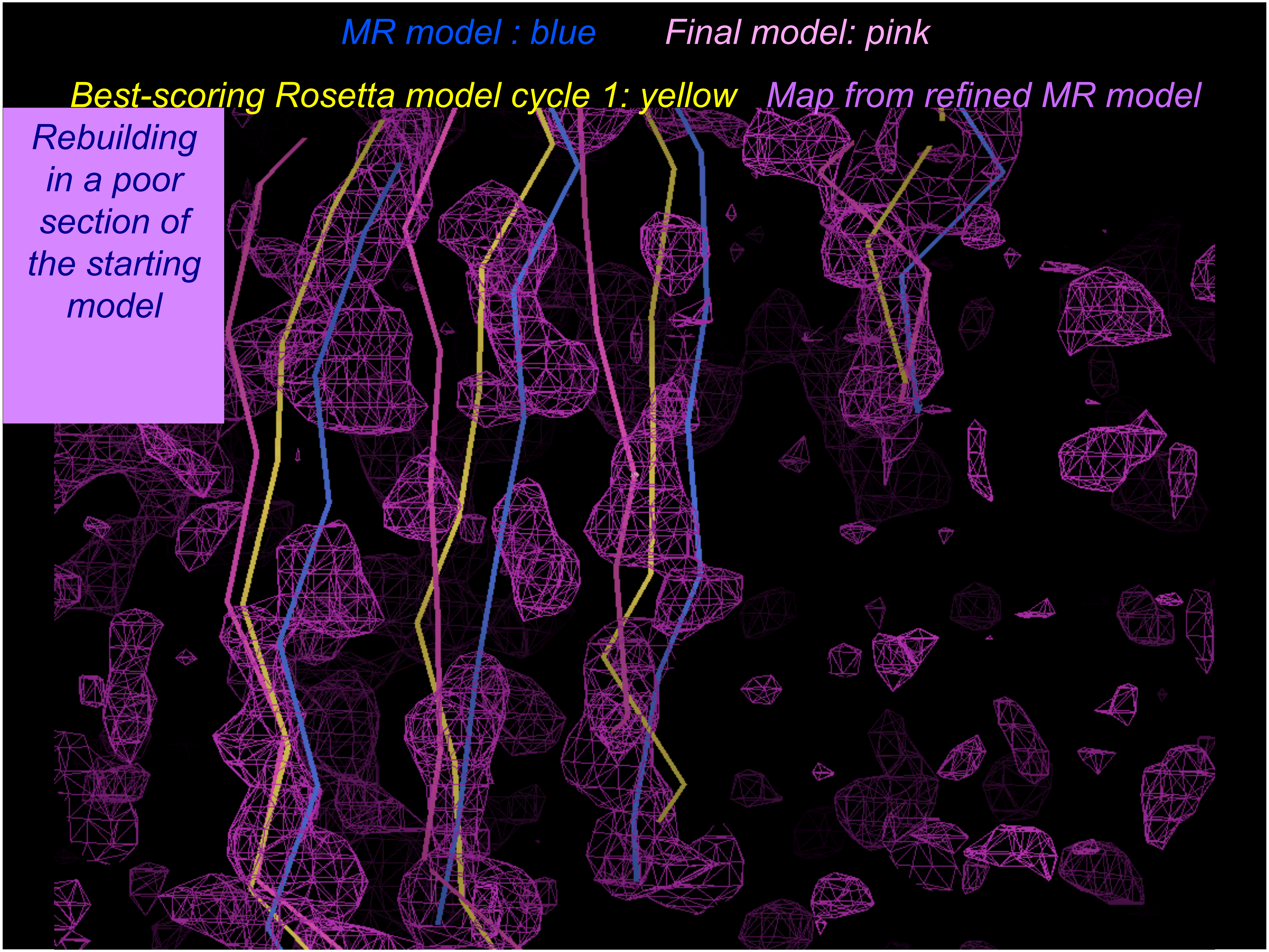


MR model : blue   Final model: pink   Rosetta models cycle 1: green

Map from refined MR model

Rebuilding in a poor section of the starting model





MR model : blue

Final model: pink

Best-scoring Rosetta model cycle 1: yellow

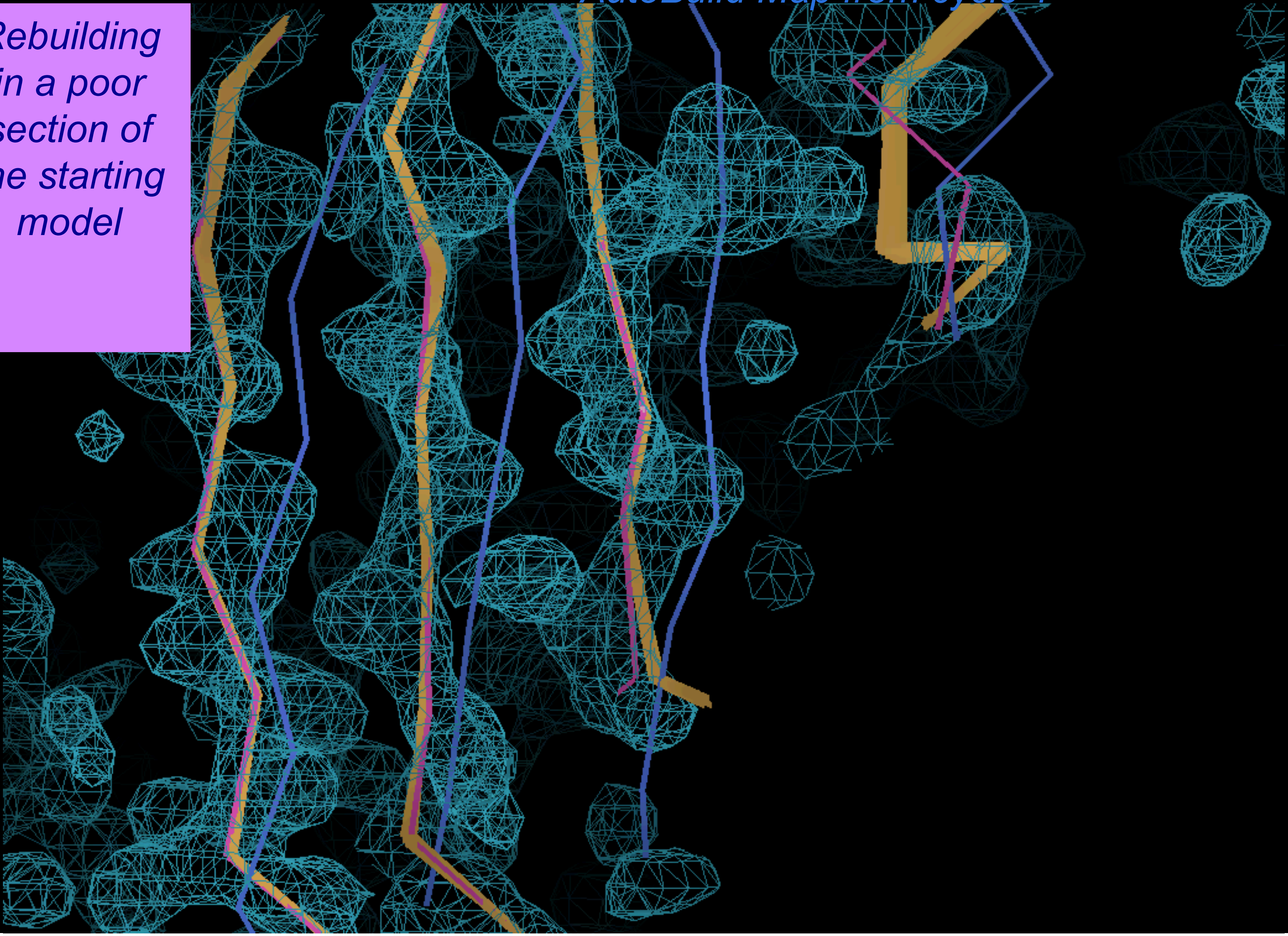
Map from refined MR model

Rebuilding in a poor section of the starting model

MR model : blue   Final model: pink   Rosetta models cycle 2: yellow

AutoBuild Map from cycle 1

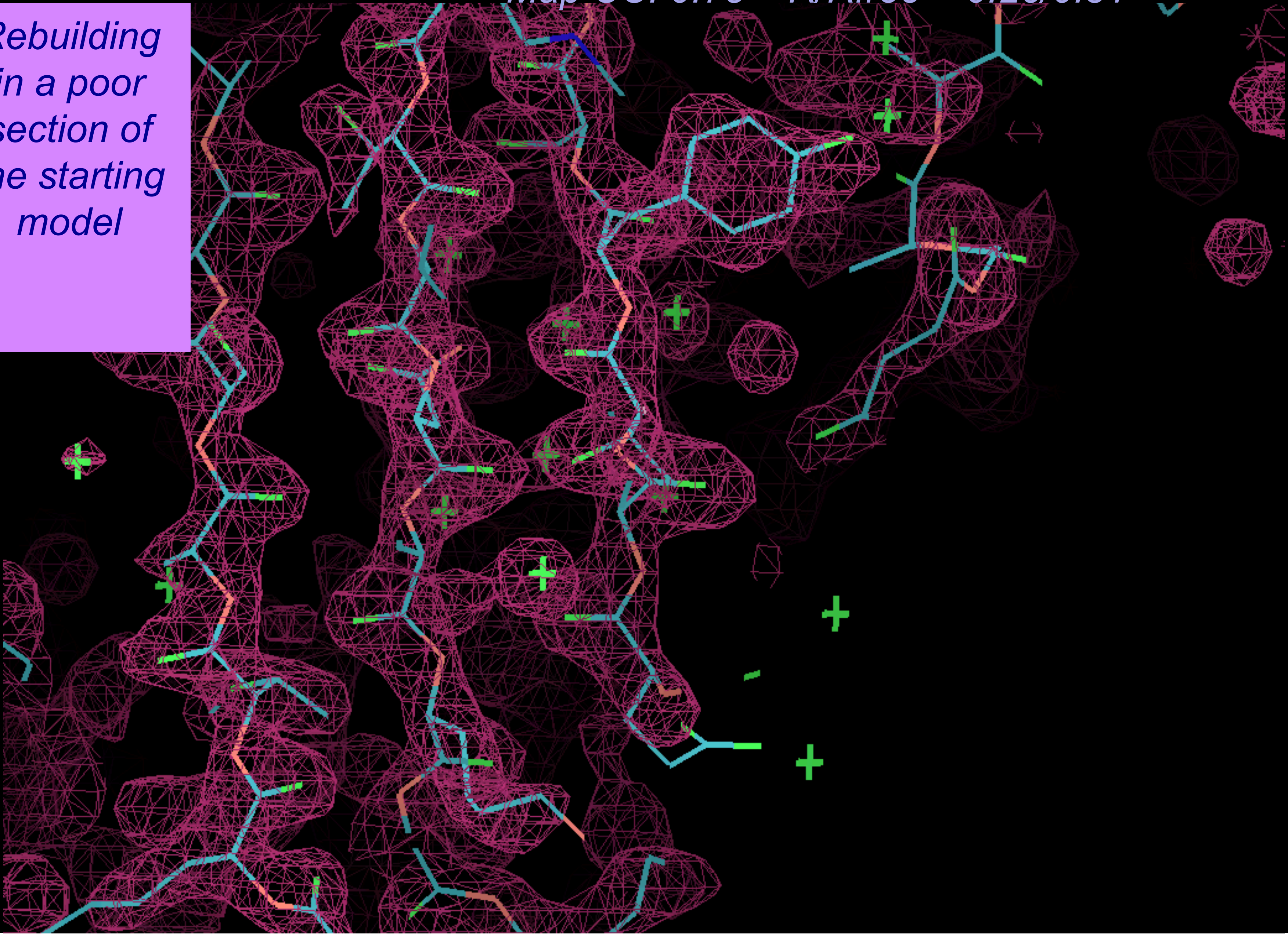
Rebuilding in a poor section of the starting model



AutoBuild model cycle 2

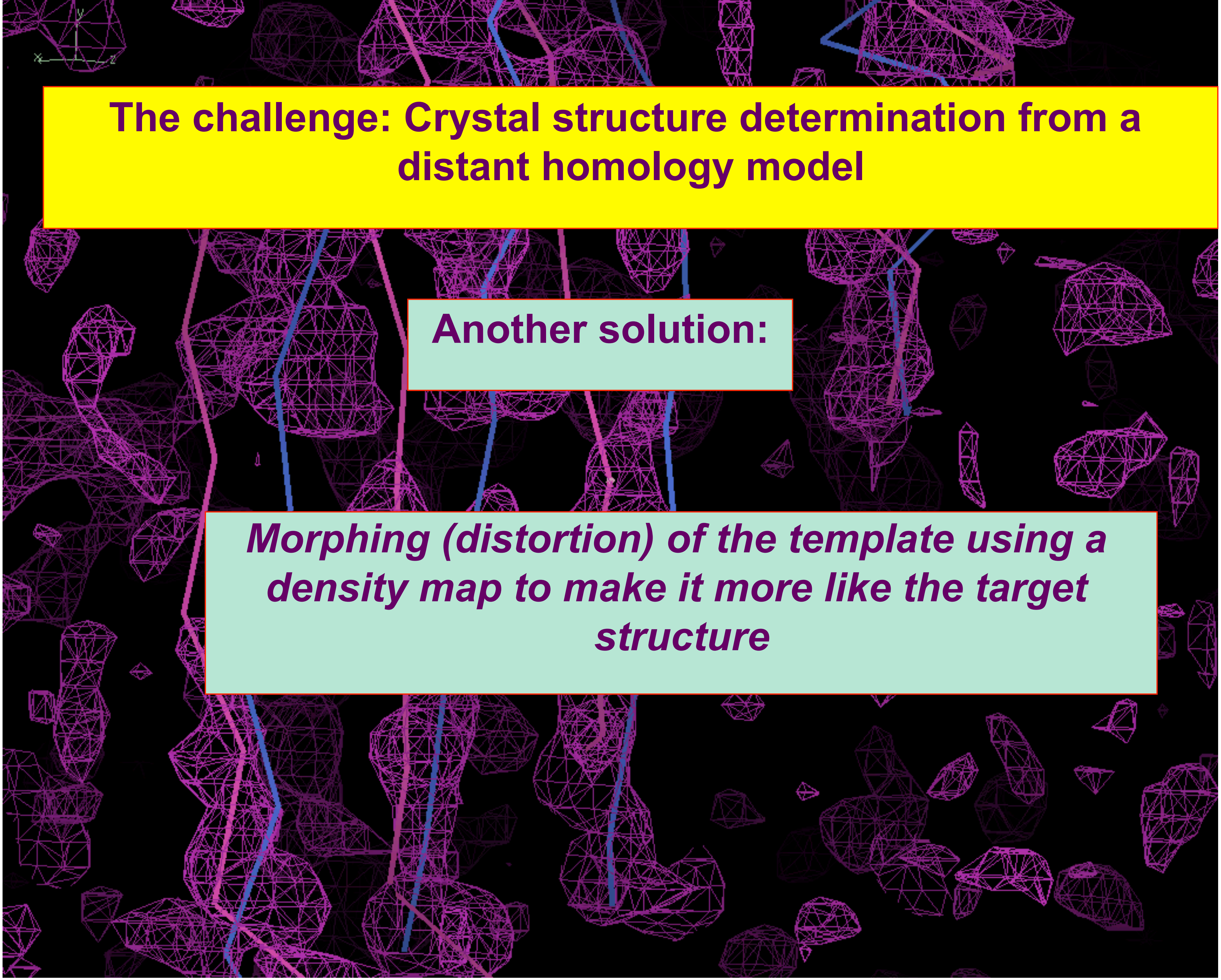
Map CC: 0.78 R/Rfree = 0.26/0.31

Rebuilding in a poor section of the starting model



mr_rosetta rebuilding starting with placed templates

structure	dmin	% ident	ncs	Free R	template **(Structural genomics structures)**
ag9603a	1.7	100	2	0.27	NMR model
cab55348	1.9	31	1	0.23	Unpublished structure
xmrv	2.0	30	2	0.34	HIV protease
fk4430	2.1	22	1	0.29	Nudix hydrolase (MCSG)
thiod	2.1	22/15	1	0.30	Thioredoxin;protein disulfide isomerase
bfr258e	2.2	19	2	0.28	Glutathione-S-transferase
niko	2.5	27	2	0.31	Carboxyvinyltransferase (RIKEN)
estan	2.5	18	1	0.25	Alpha-amylase
fj6376	2.7	21	4	0.30	Domain of unknown function 364 (JCSG)
pc02153	2.8	29	1	0.44	Prephanate dehydrogenase (NYSGC)
pc0265	2.9	29	2	0.39	Xanthine dehydrogenase
tirap	3.0	22	1	0.42	MYD88 (NMR model, NESG)
hp3342	3.2	20	1	0.42	Succinyl diaminopimelate desuccinylase (SGX)



The challenge: Crystal structure determination from a distant homology model

Another solution:

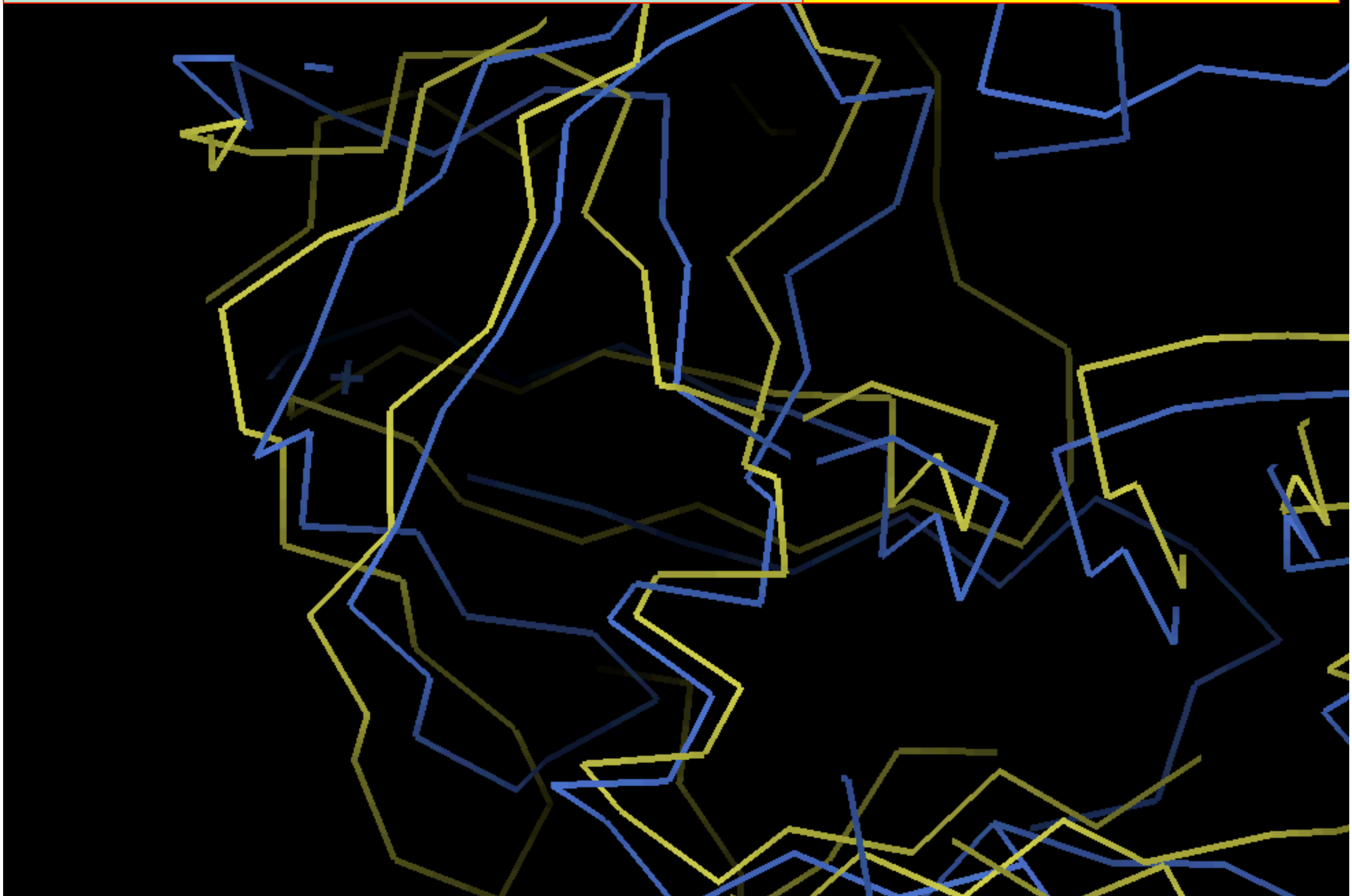
Morphing (distortion) of the template using a density map to make it more like the target structure

Related structures often have high local similarity

ag9603; approximate NMR model as template in pink

Related structures often have high local similarity

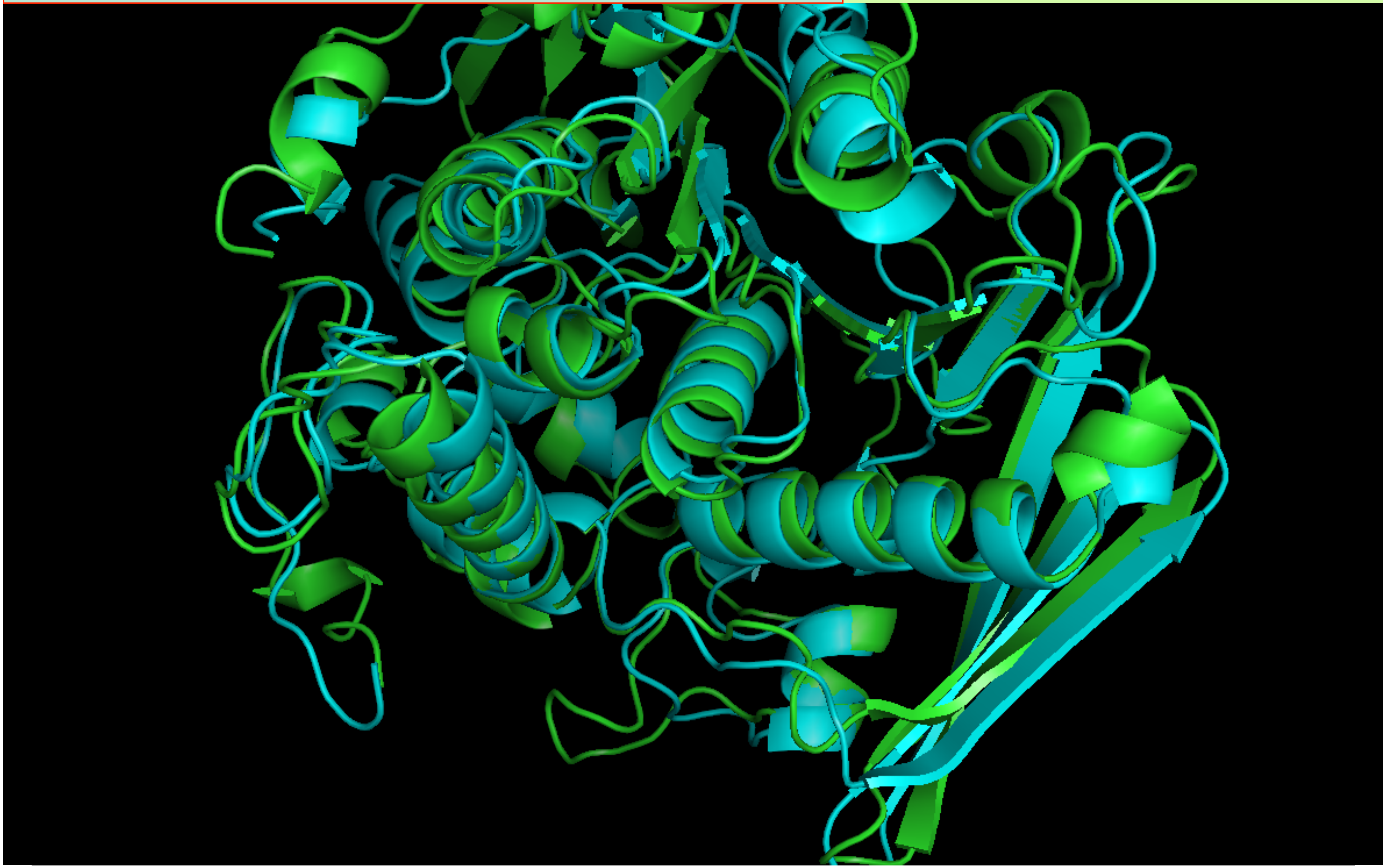
XMRV PR, 30% identity template (2hs1) in blue



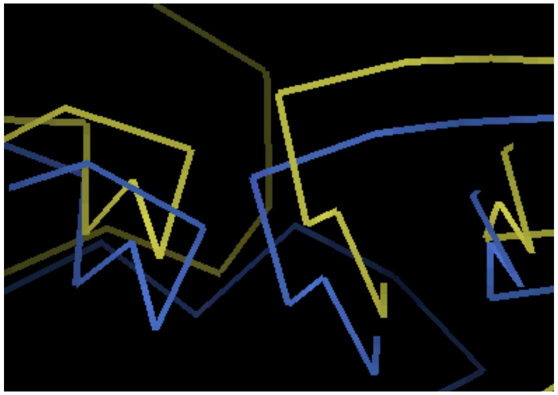
Related structures often have high local similarity

cab55348

32% identical template (Cip2) in blue



Taking advantage of local similarities of homologous structures



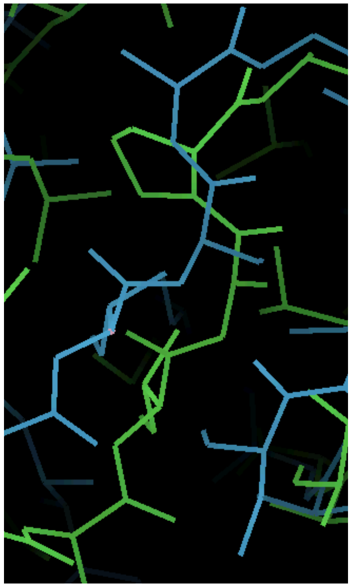
Rigid-body refinement of segments

Fragment searches (FFFEAR, ESSENS)

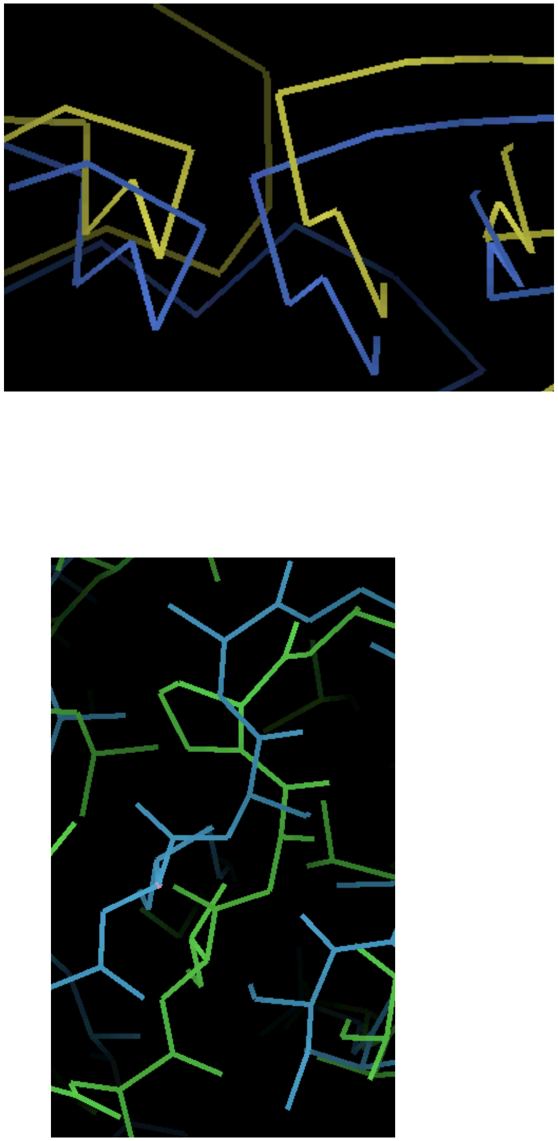
DEN or jelly-body refinement

Rosetta modeling

Morphing



Morphing



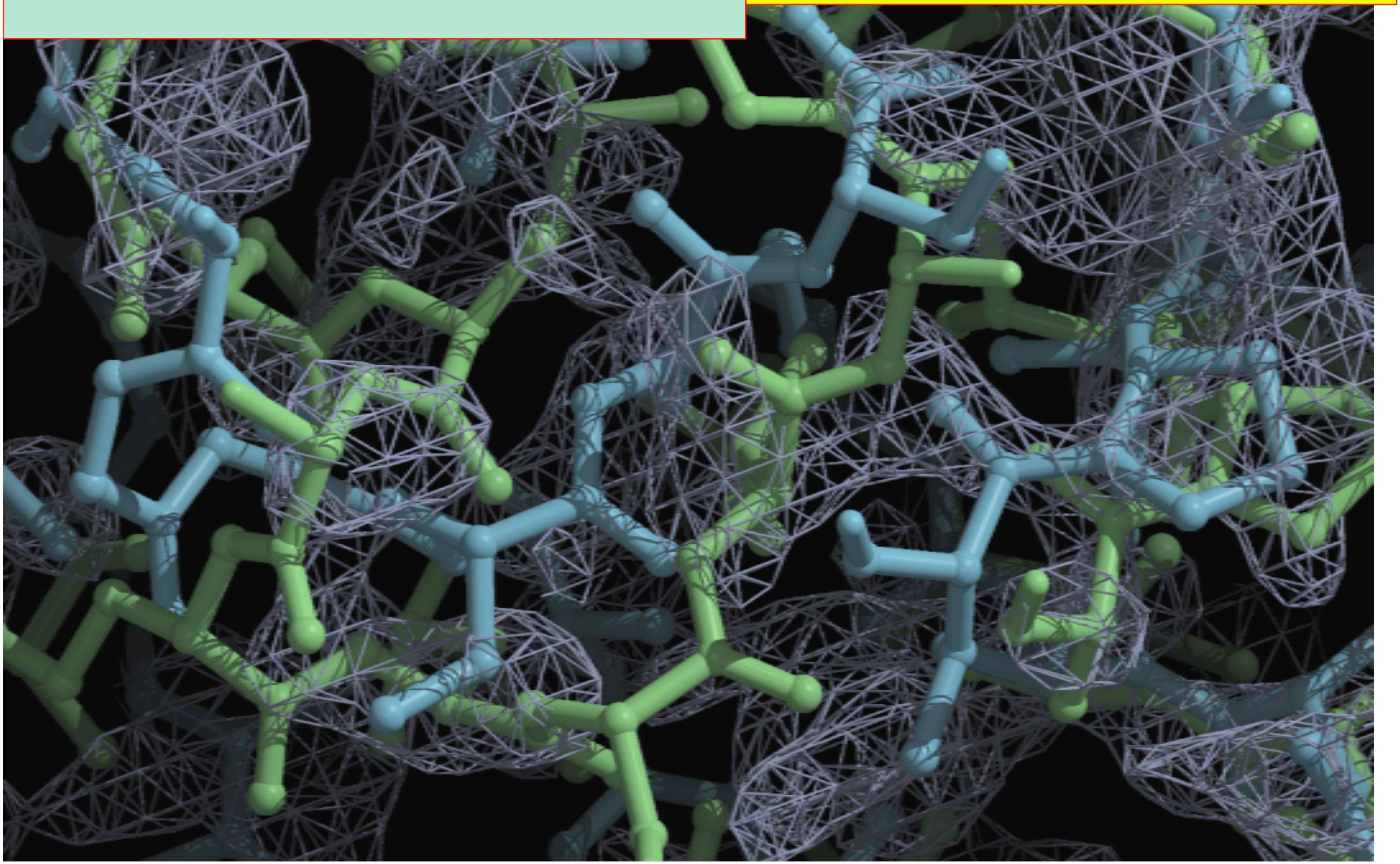
Local structures may superimpose very closely

The position of a large group of atoms can be identified accurately with a poor map

Relationship between structures may be a simple distortion

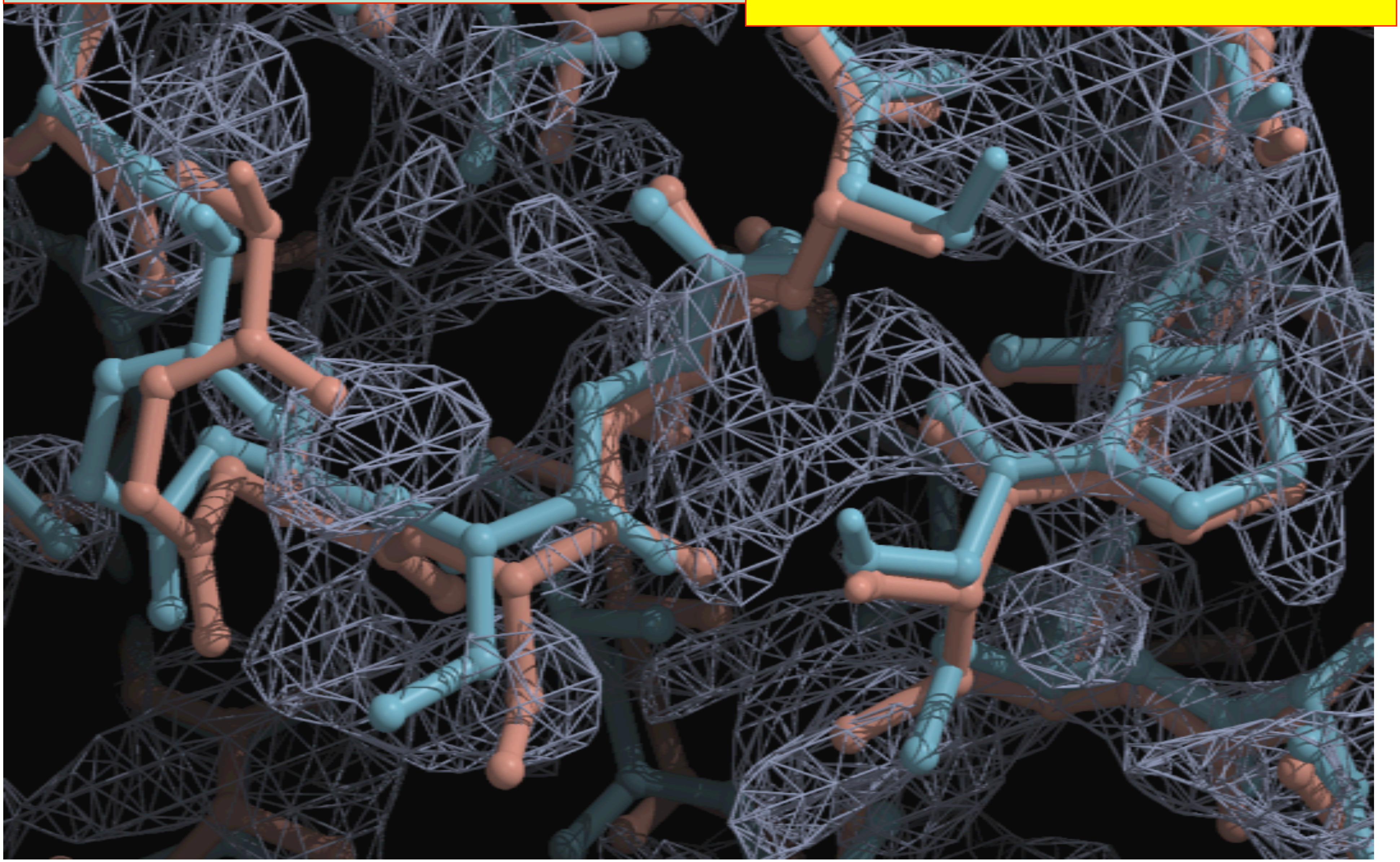
A challenging morphing problem:
How can we use this map to identify the shifts needed?

cab55342: final model green
3PIC (32% identity) in blue

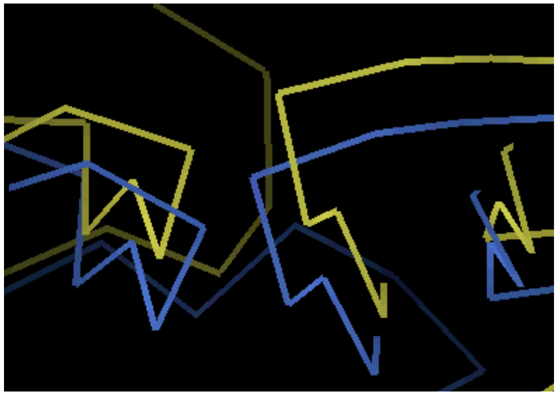


Standard refinement does not move the structure very much..

cab55342:
3PIC (32% identity) in blue
Refined template in orange



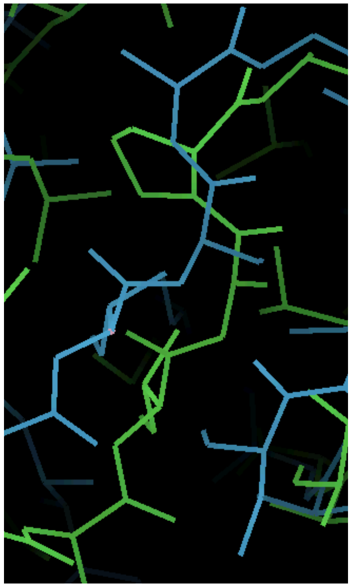
Steps in morphing



A. Identify local translation to apply to one C_α atom and nearby atoms

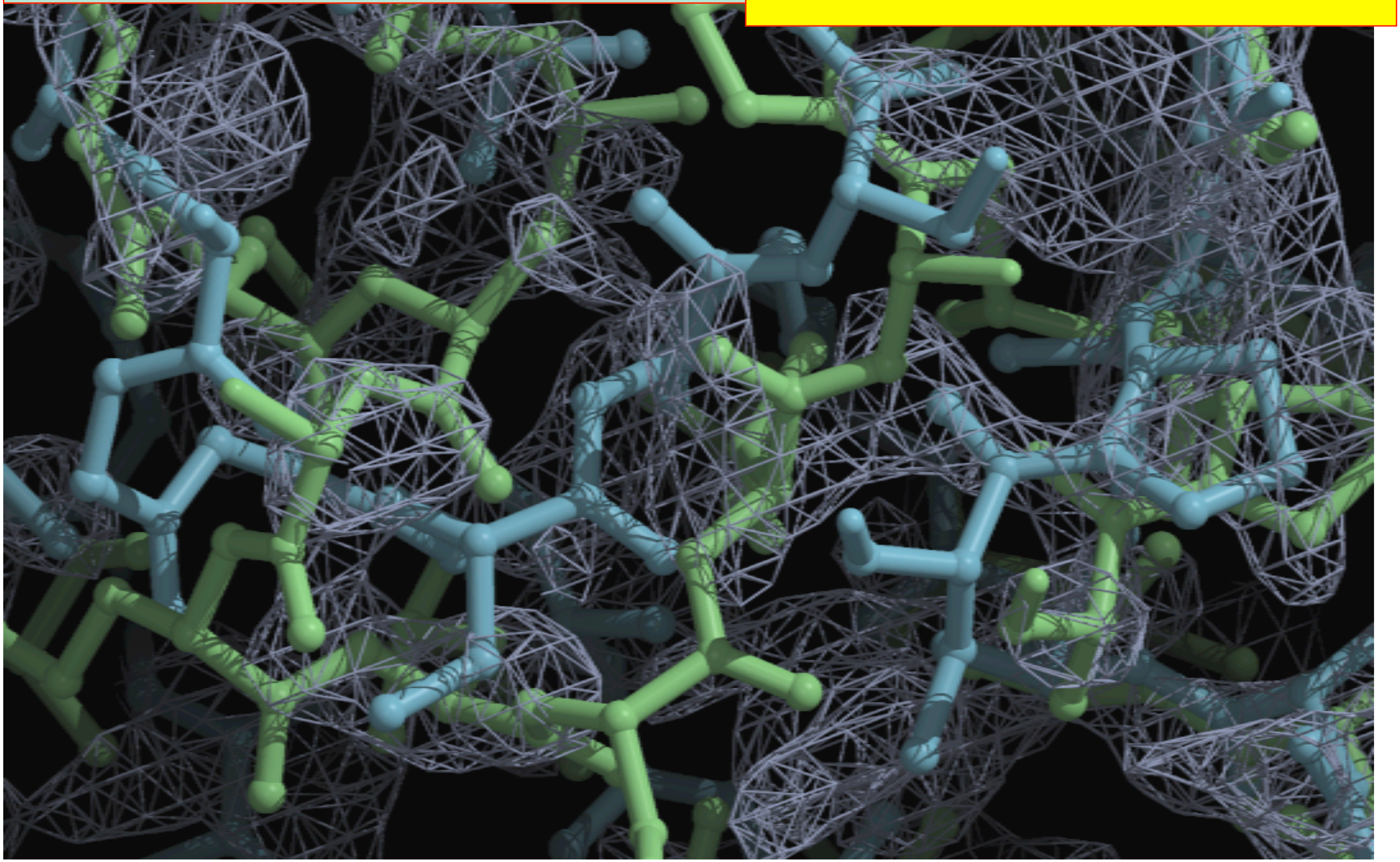
B. Smooth the local translations in window of 10 residues

C. Apply the smoothed translation to all atoms in the residue



Identify local translation to apply to one C_α atom and nearby atoms

cab55342: final model (green)
3PIC (32% identity, blue)
prime-and-switch map (blue)

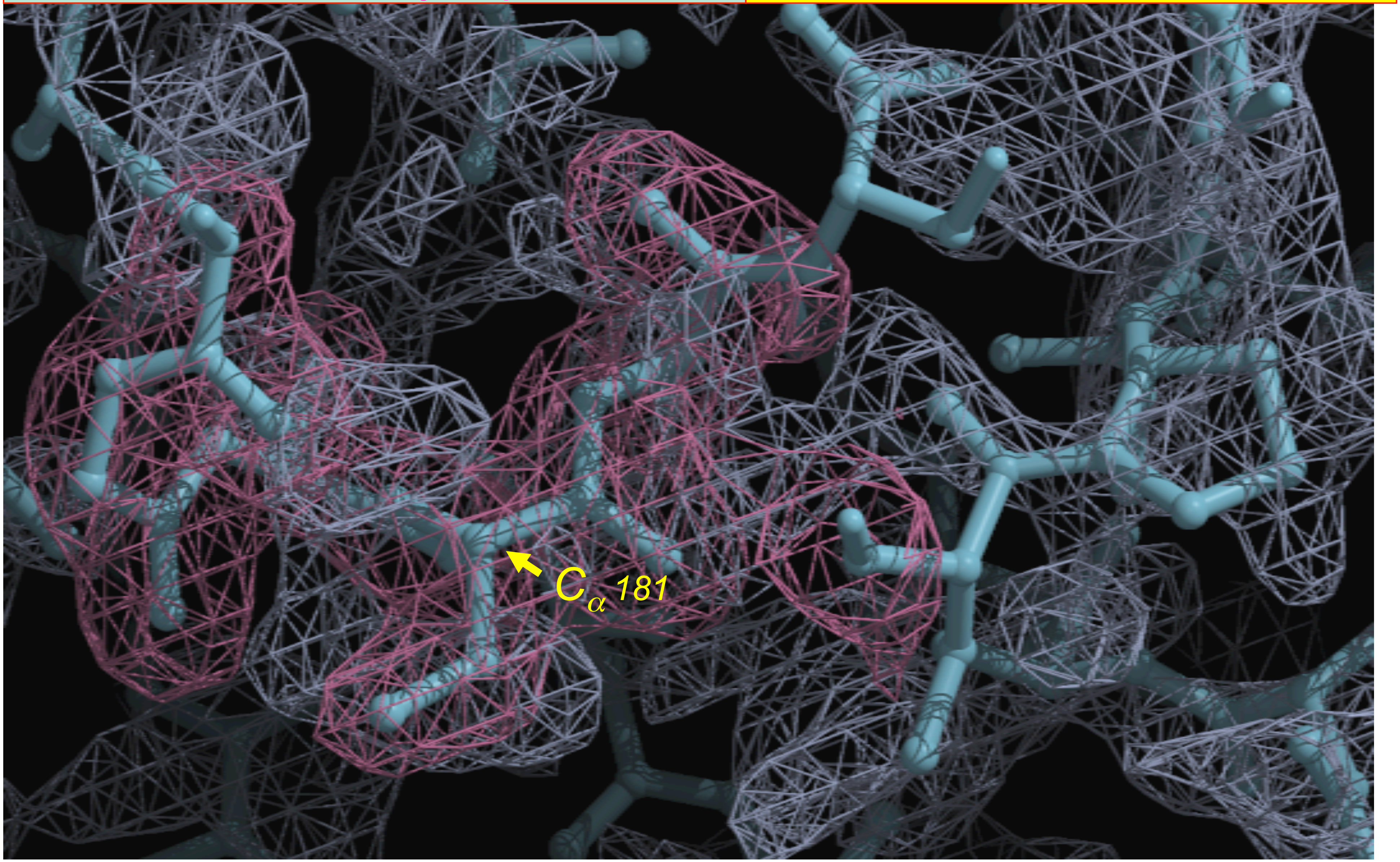


Identify local translation to apply to one C_α atom and nearby atoms

Model density in pink

cab55342:

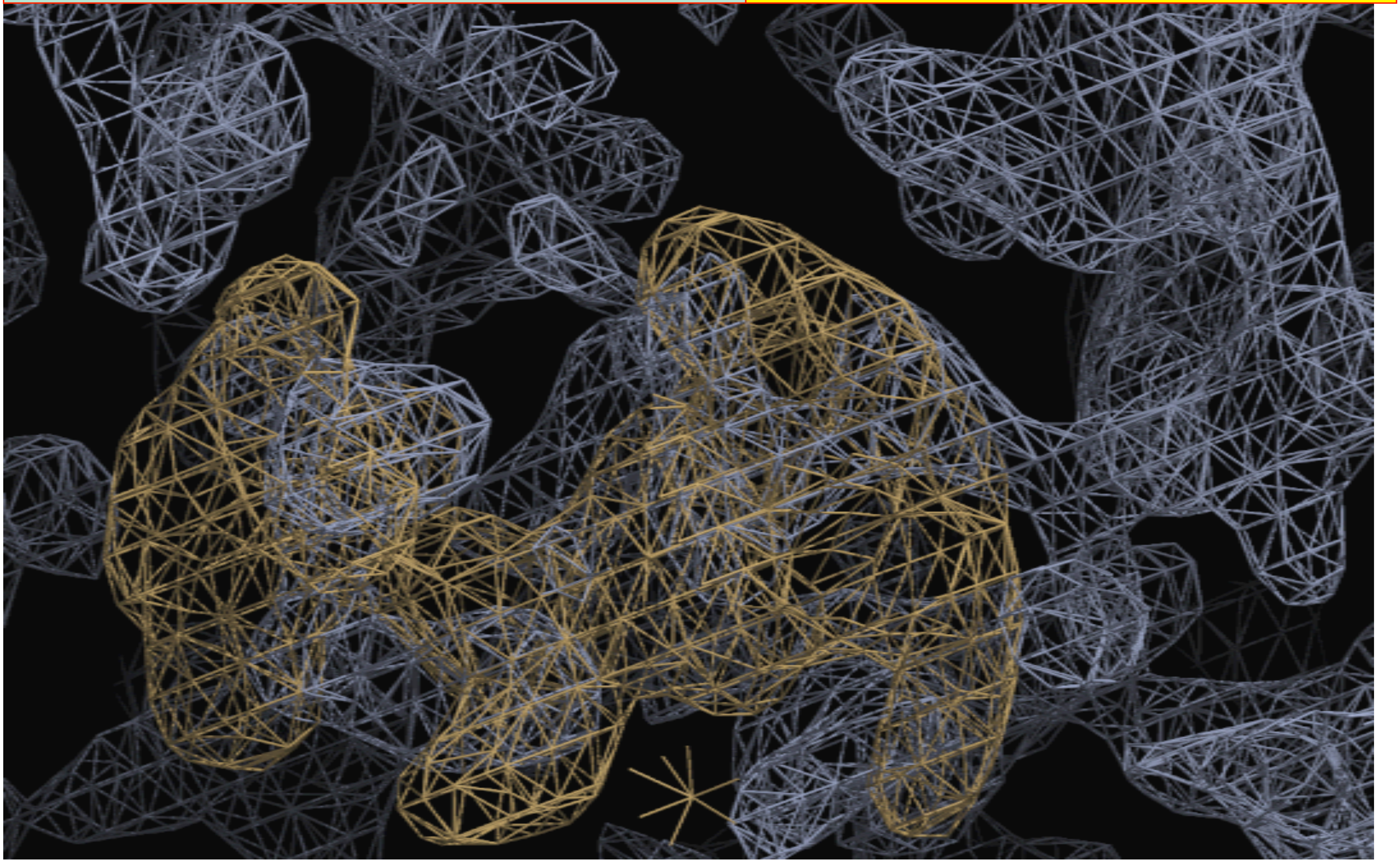
3PIC (32% identity, blue)



Identify local translation to apply to one C_α atom and nearby atoms

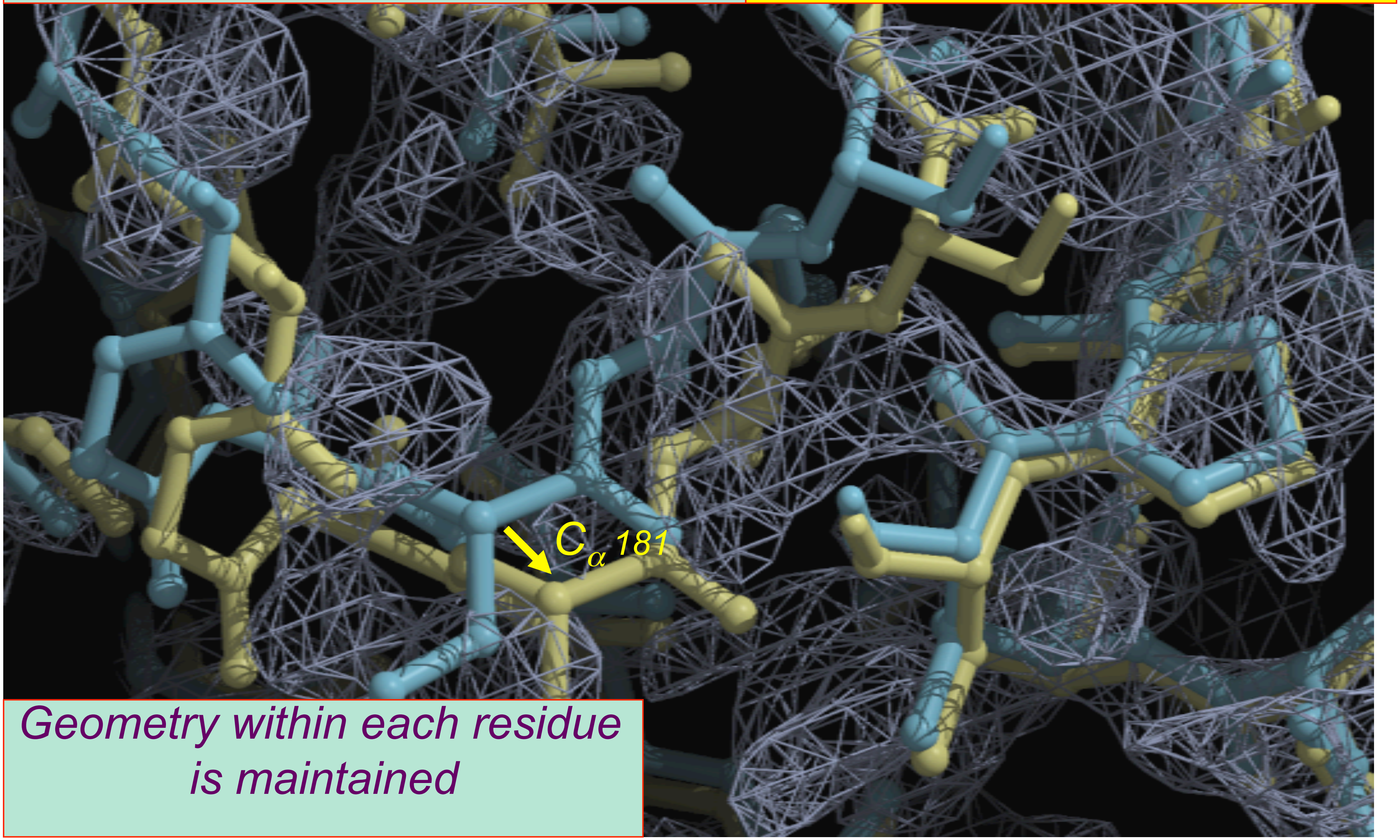
Model density offset to match map

cab55342:
3PIC (32% identity, blue)



Smooth offset over nearby residues and apply to all atoms in the residue

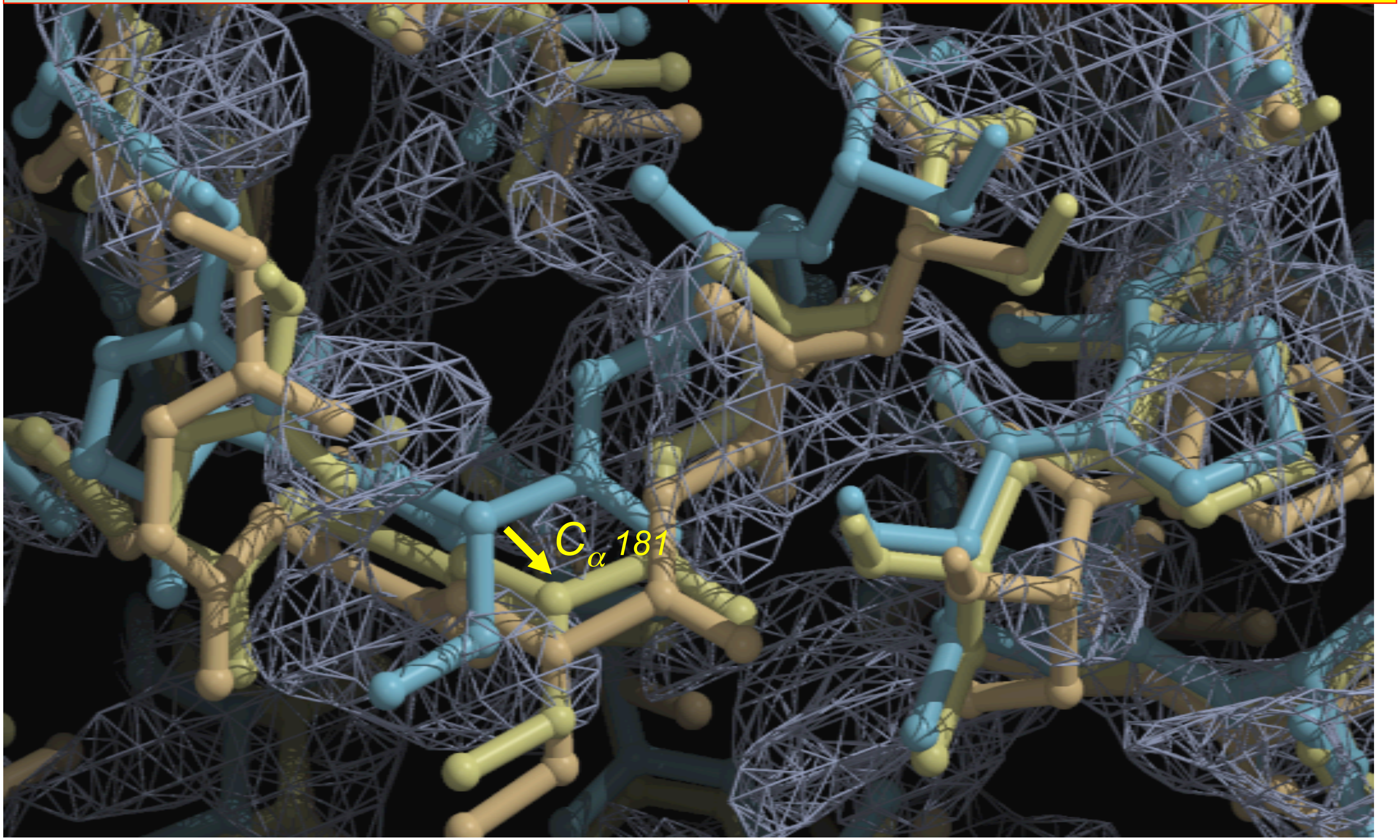
cab55342:
3PIC (32% identity, blue)
Morphed model (yellow)



Geometry within each residue is maintained

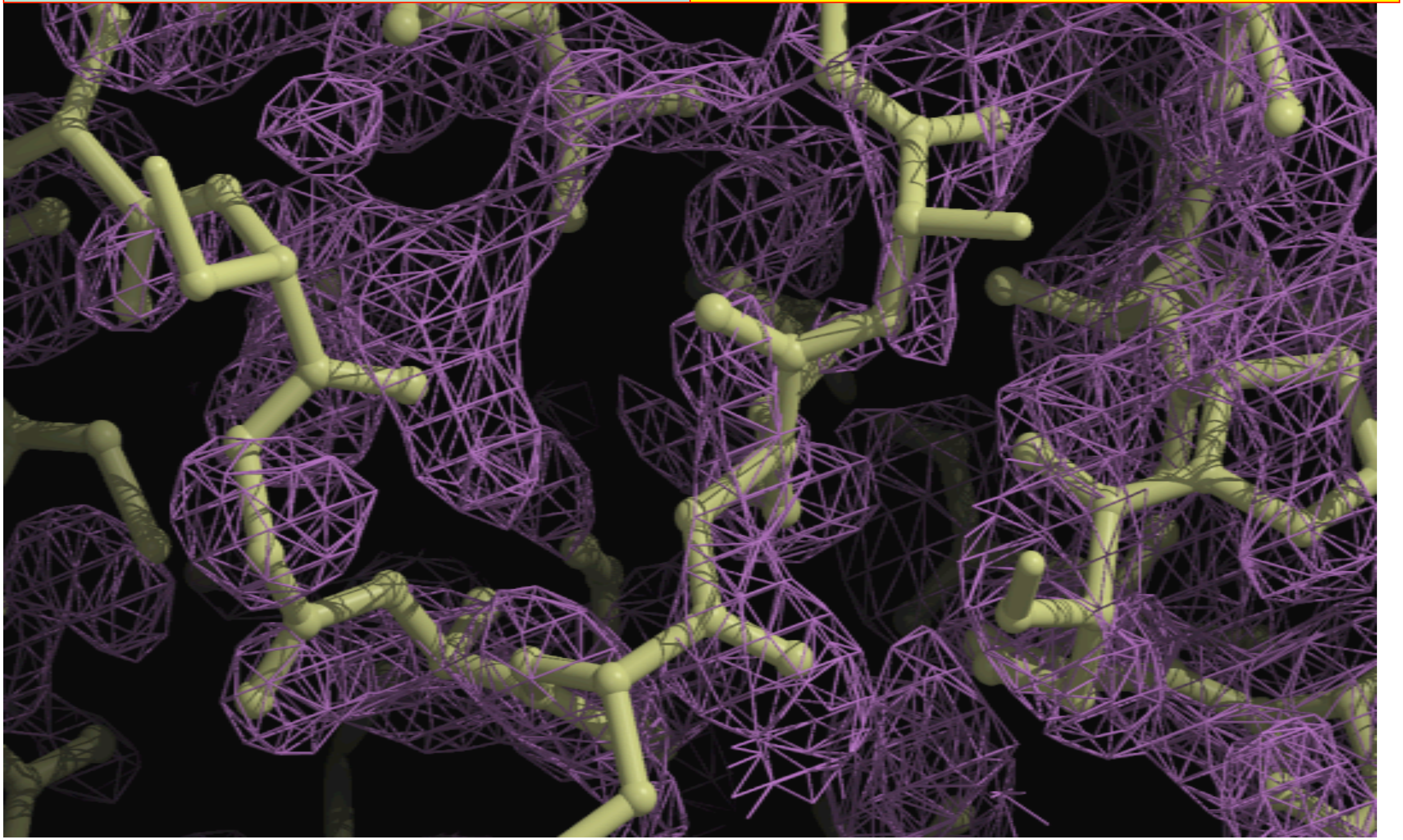
Refine morphed model

3PIC (32% identity) in blue
Morphed model (yellow)
Refined morphed model (orange)



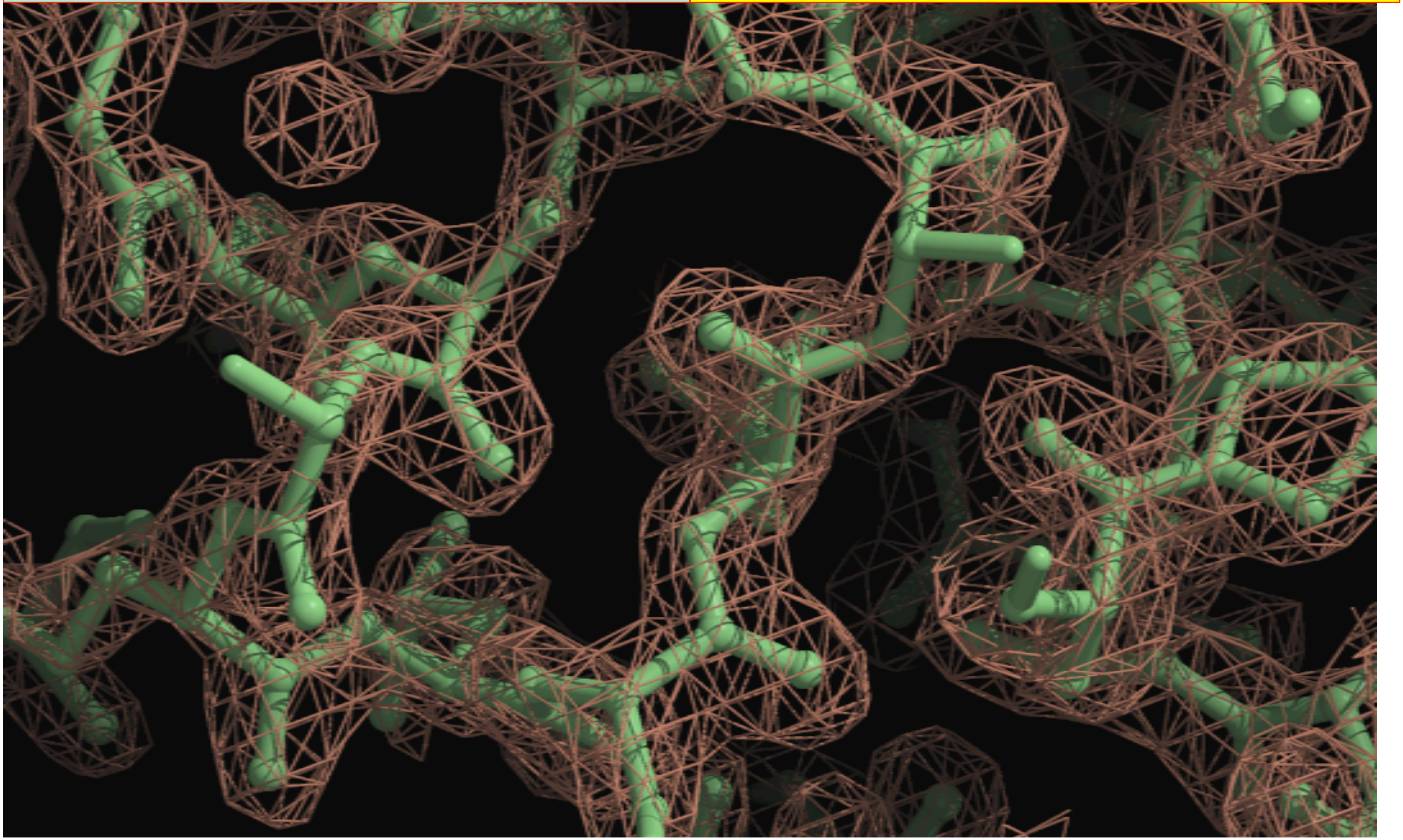
Get new map
Repeat morphing 6 times...

3PIC (32% identity) blue
Refined morphed model (yellow)
prime-and-switch map (purple)



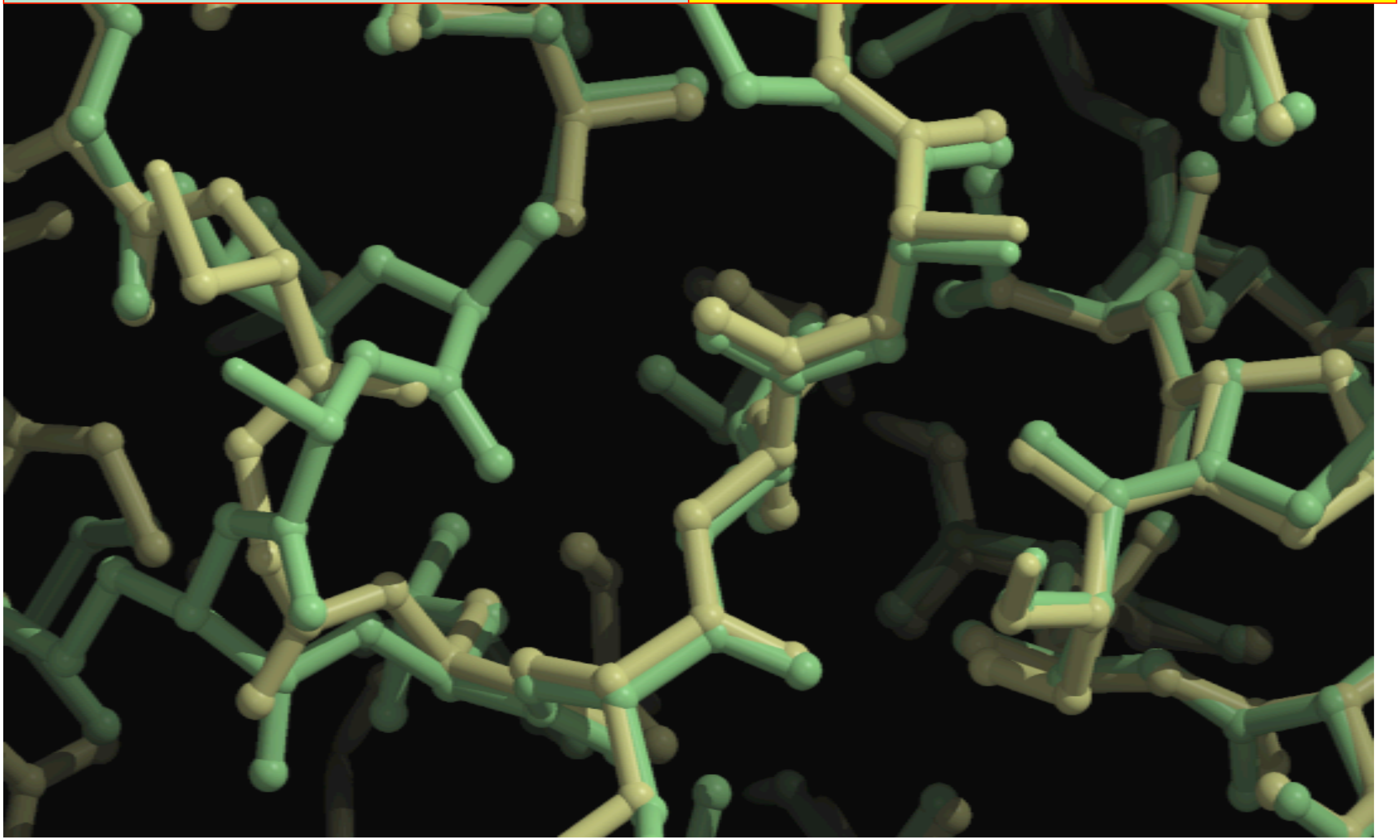
Autobuilding starting with morphed model

cab55342
Autobuild model
Density-modified map



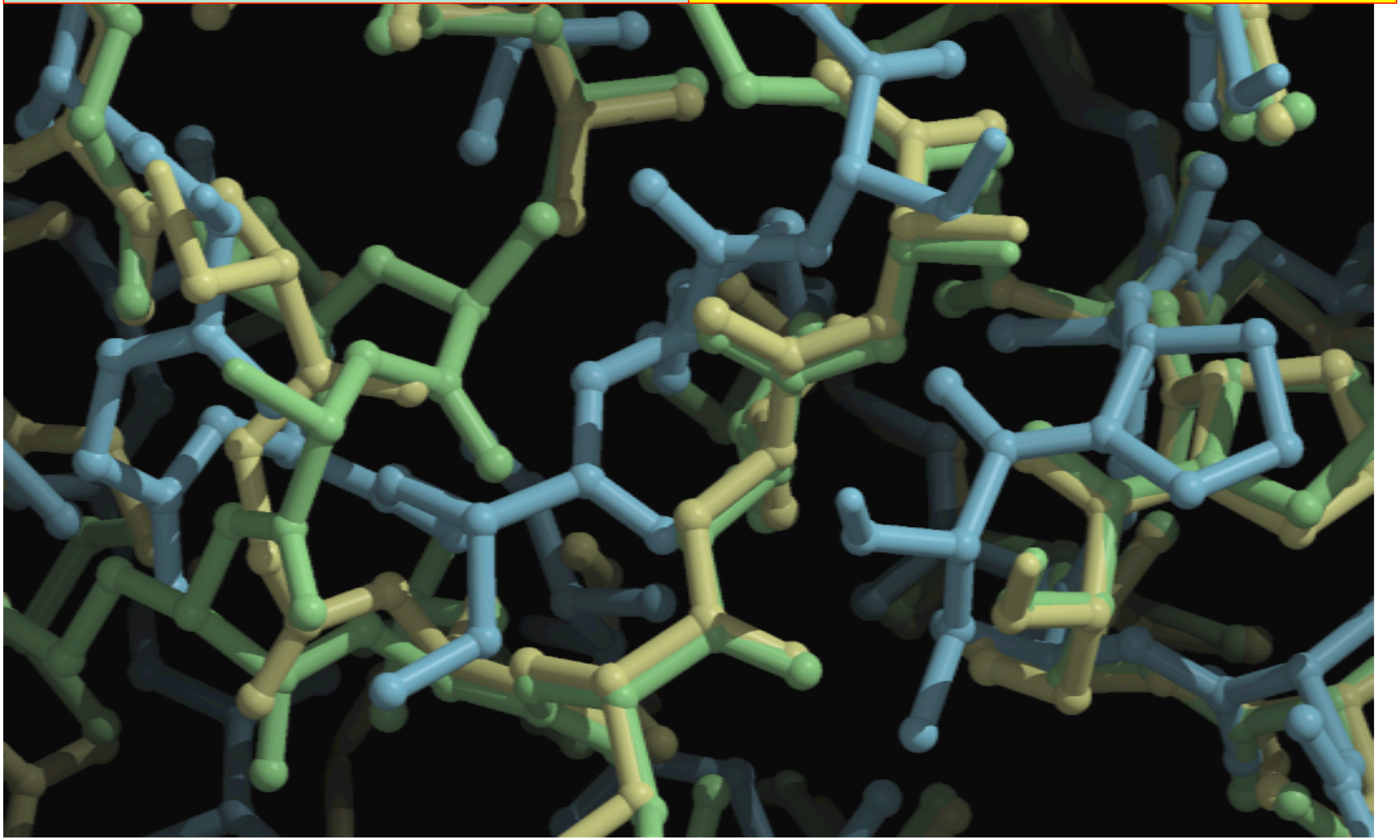
Autobuilding starting with morphed model

cab55342
Morphed model (yellow)
Autobuild model (green)



Autobuilding cab55342 starting with morphed model

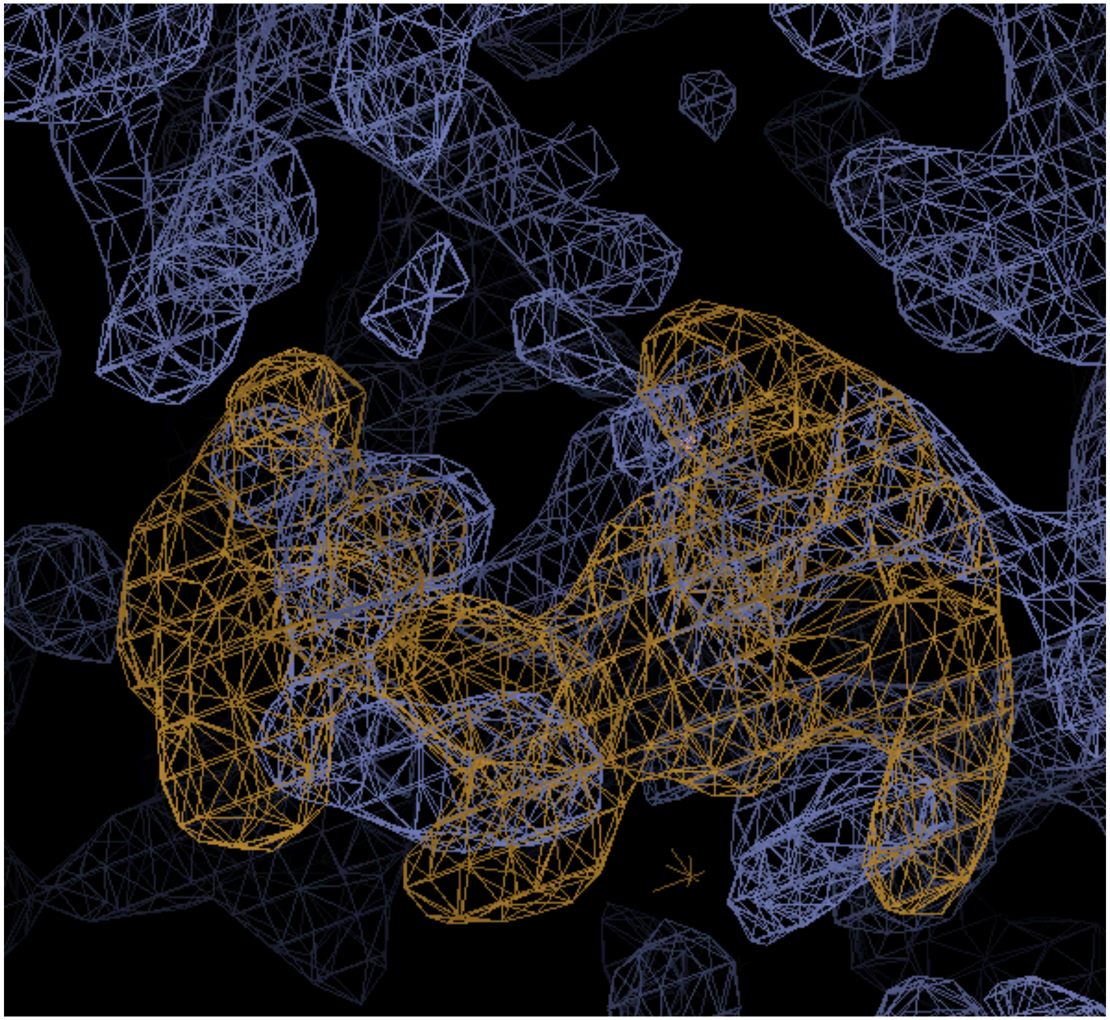
3PIC (32% identity, blue)
Morphed model (yellow)
Autobuild model (green)



What is the best map for morphing?

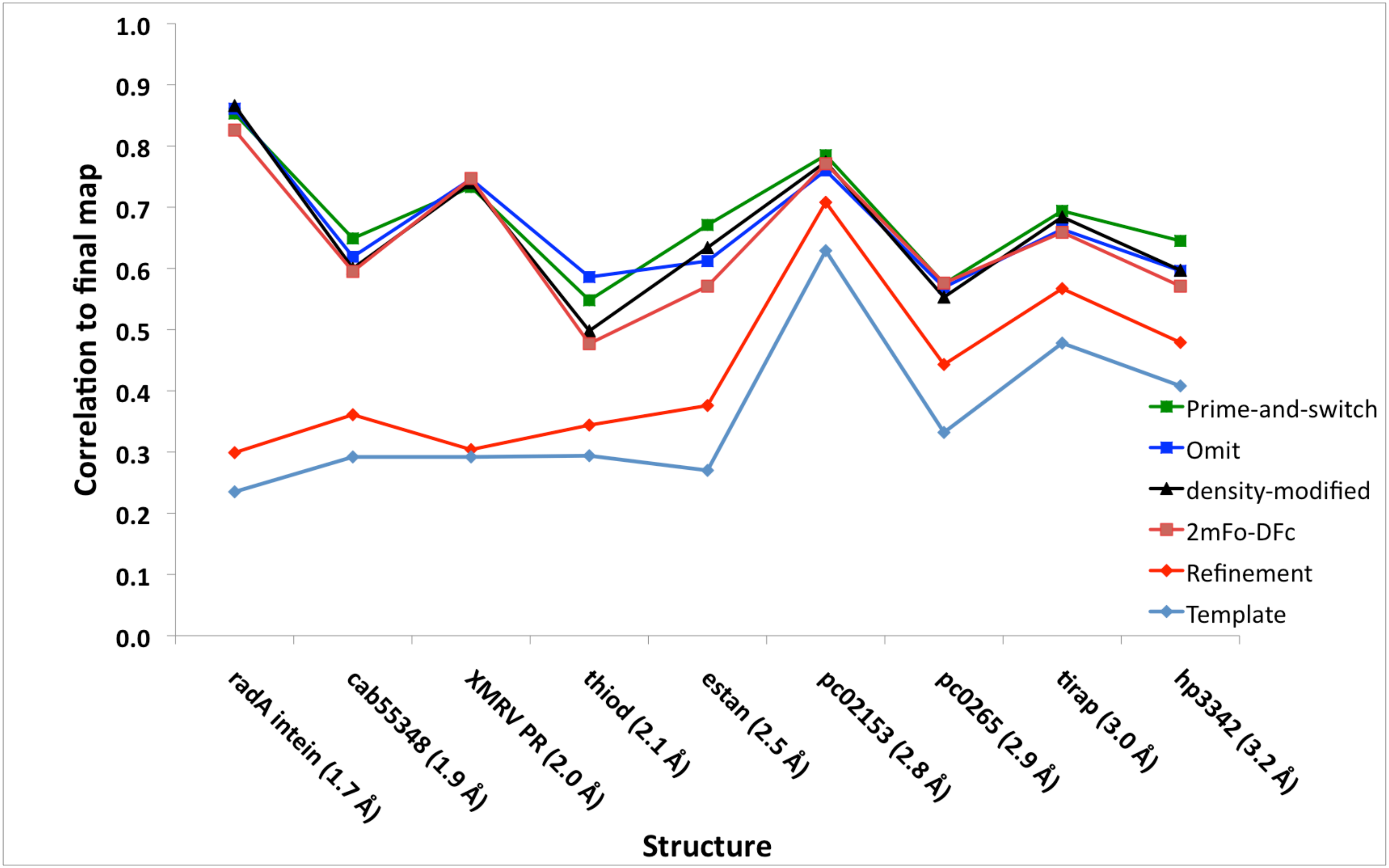
Test structures from DiMaio et al. (2011). *Improving molecular replacement by density and energy guided protein structure optimization. Nature* 473, 540-543.

(Structures that could be solved by AutoBuild excluded)

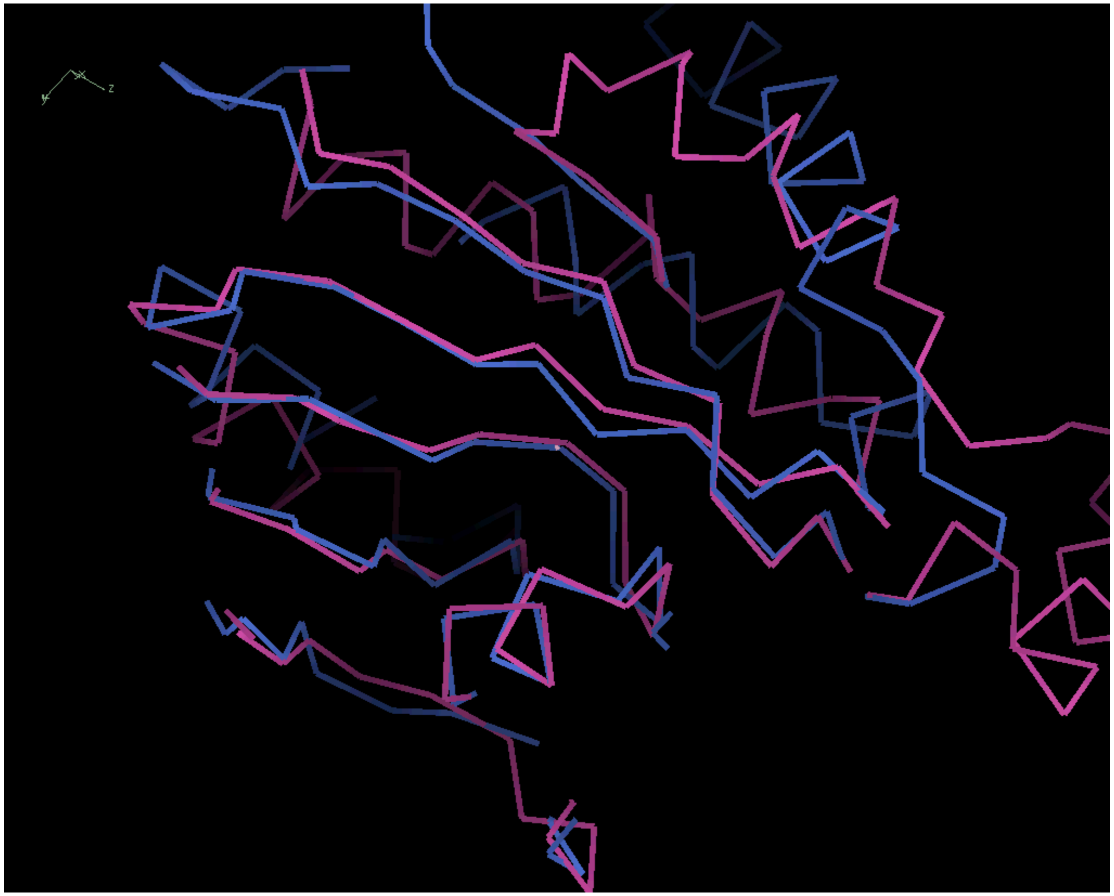


Which maps give the most useful morphing?

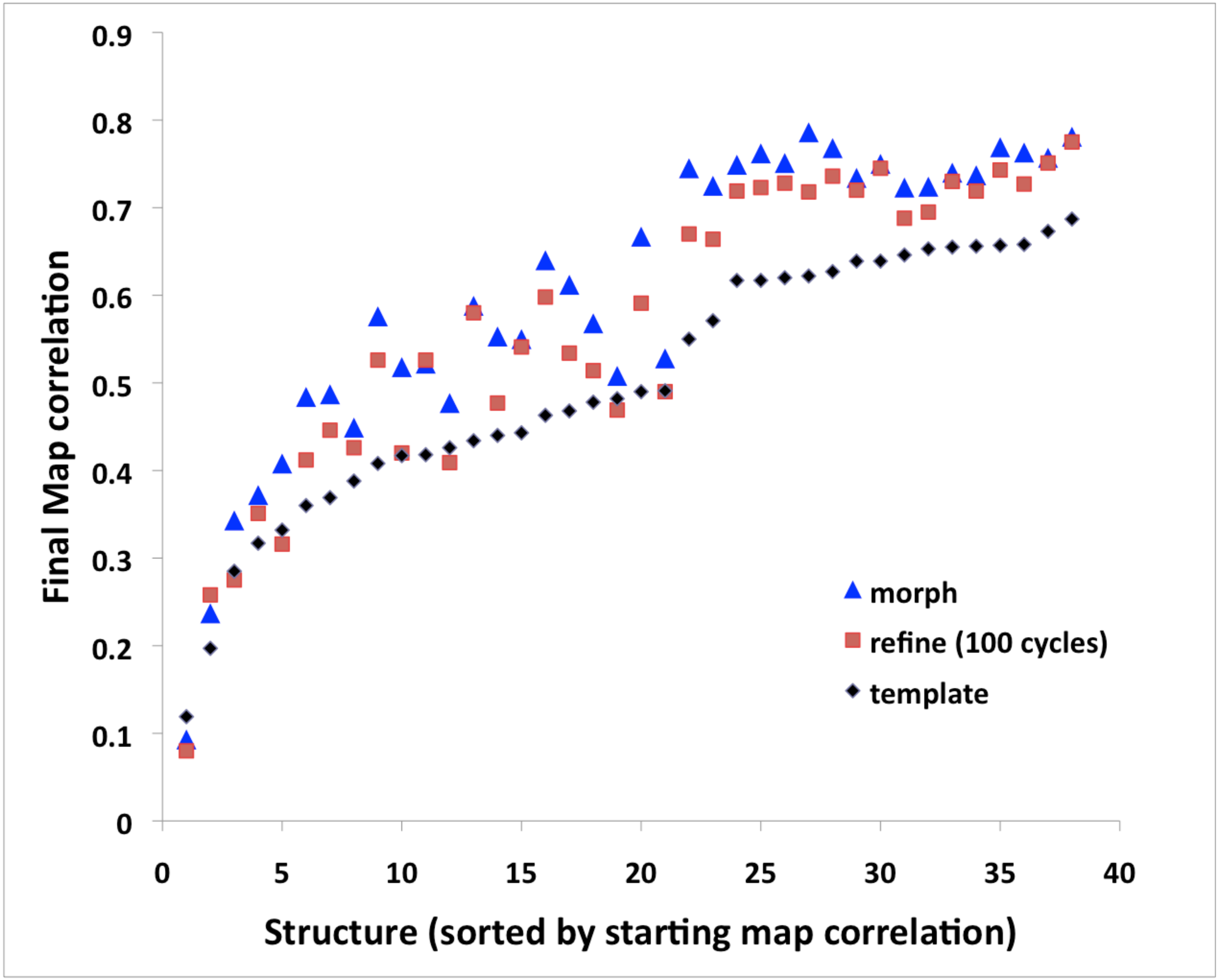
(Final map correlation after morphing using various maps)



Tests of morphing with
a series of templates
with varying similarity
to target structure

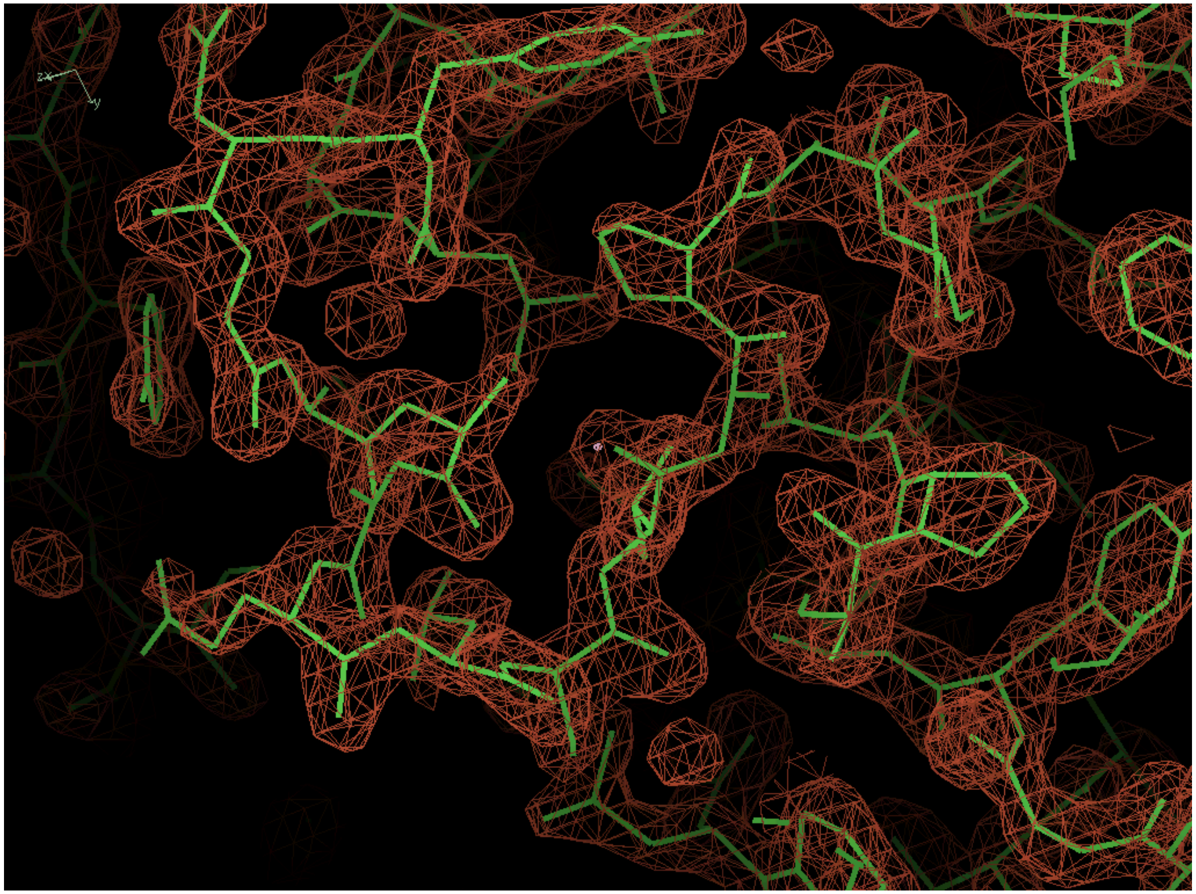


Morphing on a series of templates (1A2B; template sequence identity 7%-33%)

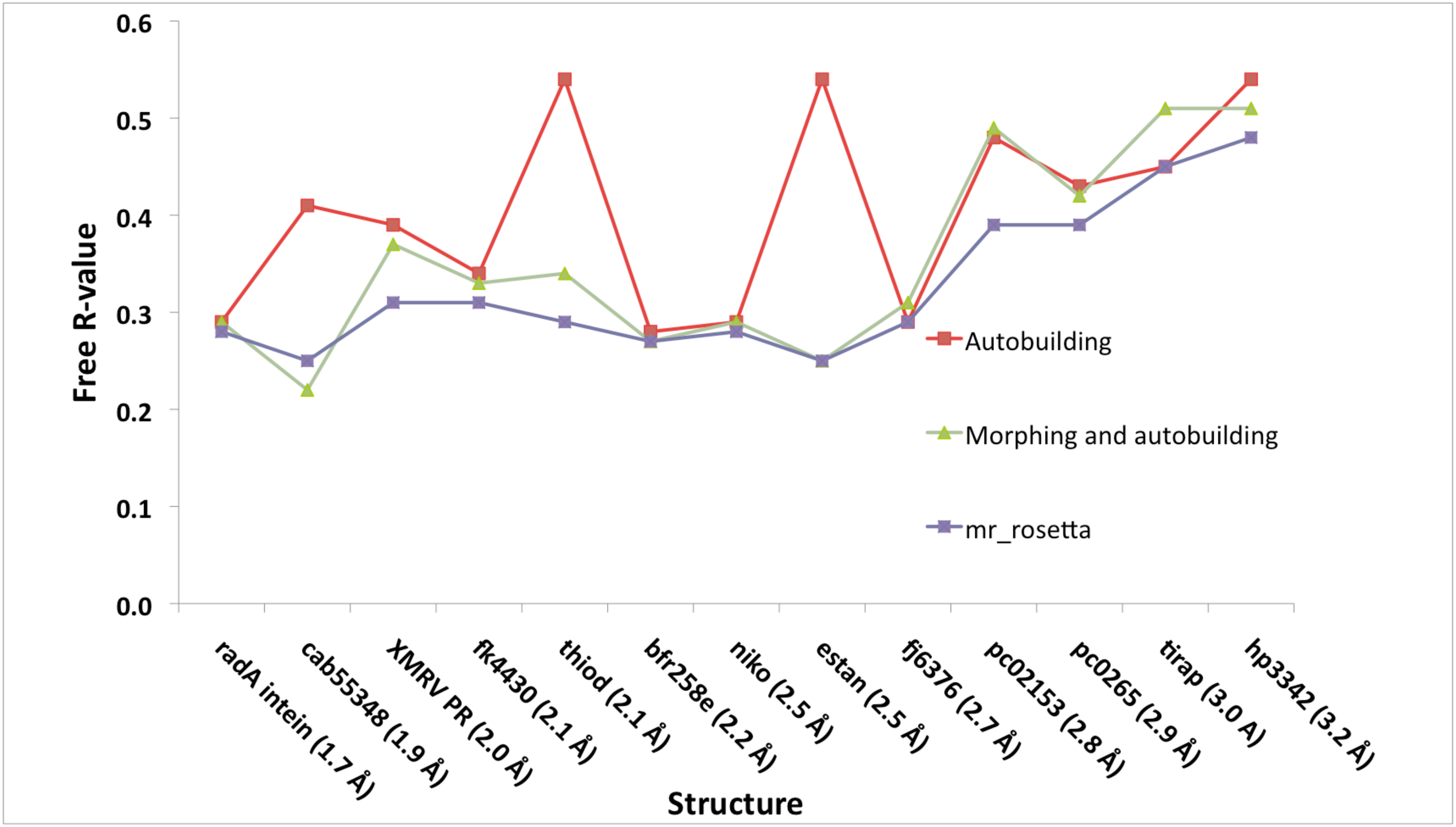


Tests of Autobuilding after morphing

Comparison with *phenix.mr_rosetta*

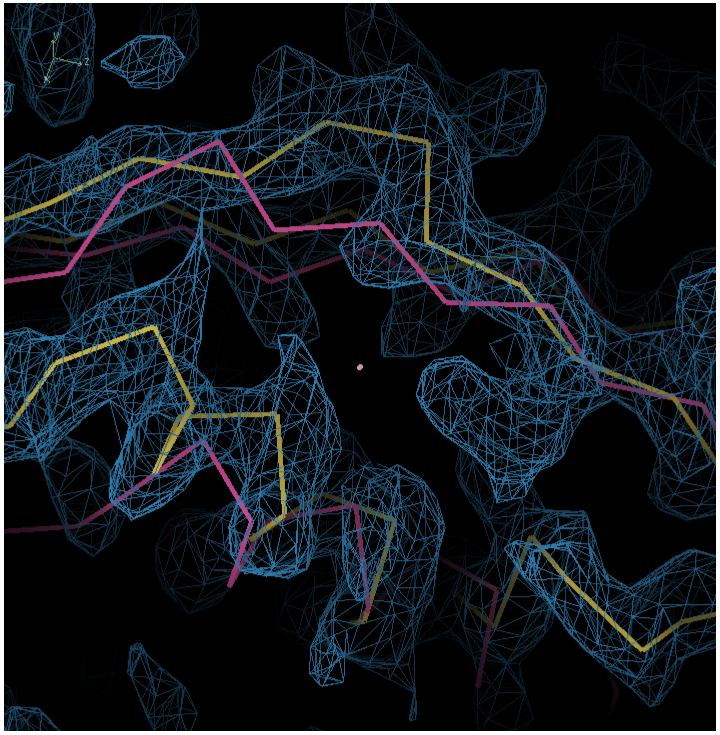


Morphing combined with autobuilding



Another MR problem:

A map at 3 Å or worse…
A homology model that fits the density in some places and not in others…



How do we decide what parts of the model to use?

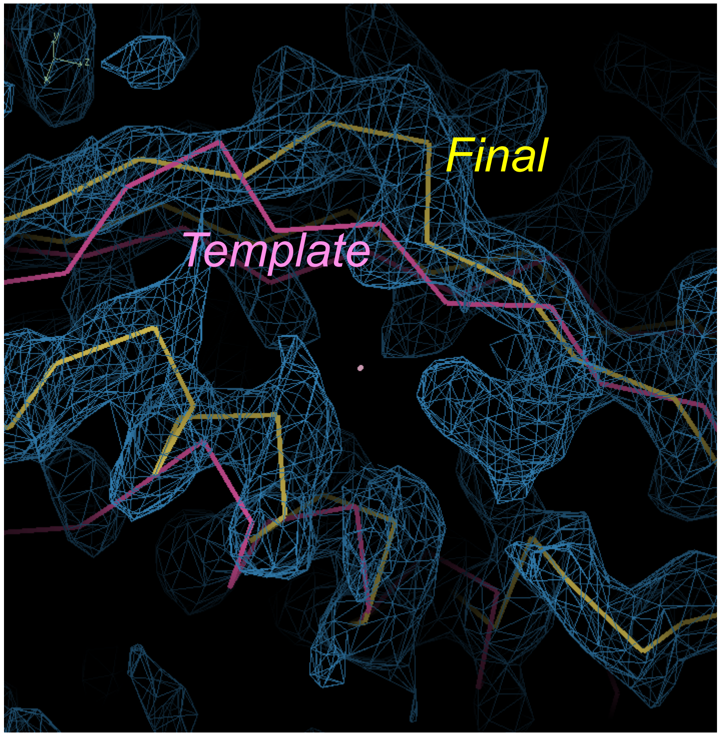
How do we assign the sequence to the model?

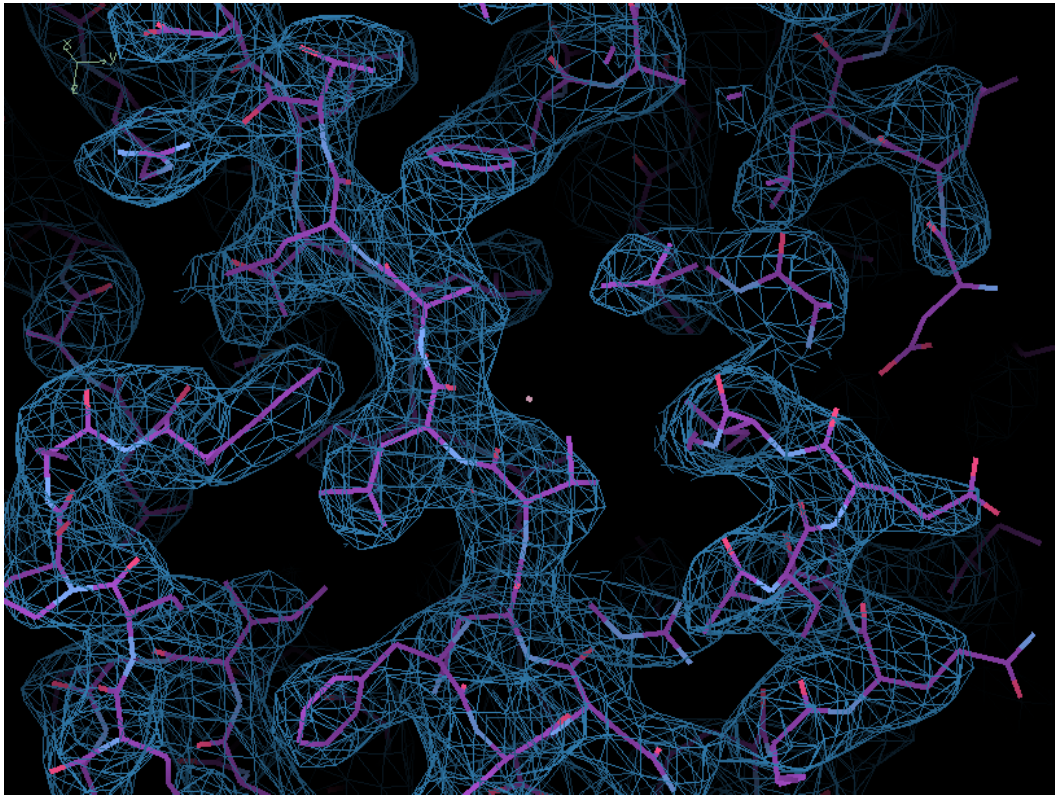
(making use of the connectivity of the template)

Cgl1109 (JCSG HP3342)
3.2 Å highly anisotropic data

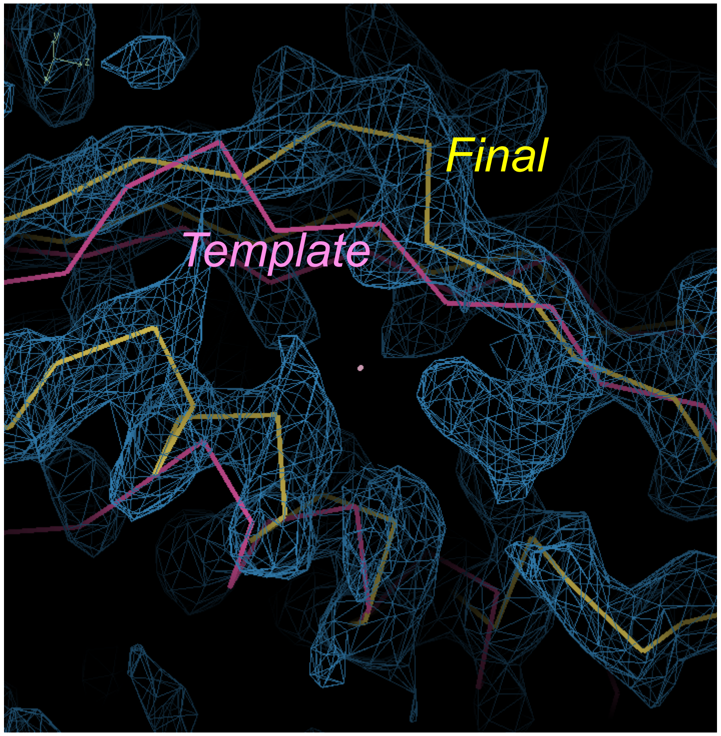
Putative dapE from *C. glutamicum, 267 residues*

Structure solved by Axel Brunger using DEN/Phenix autobuild





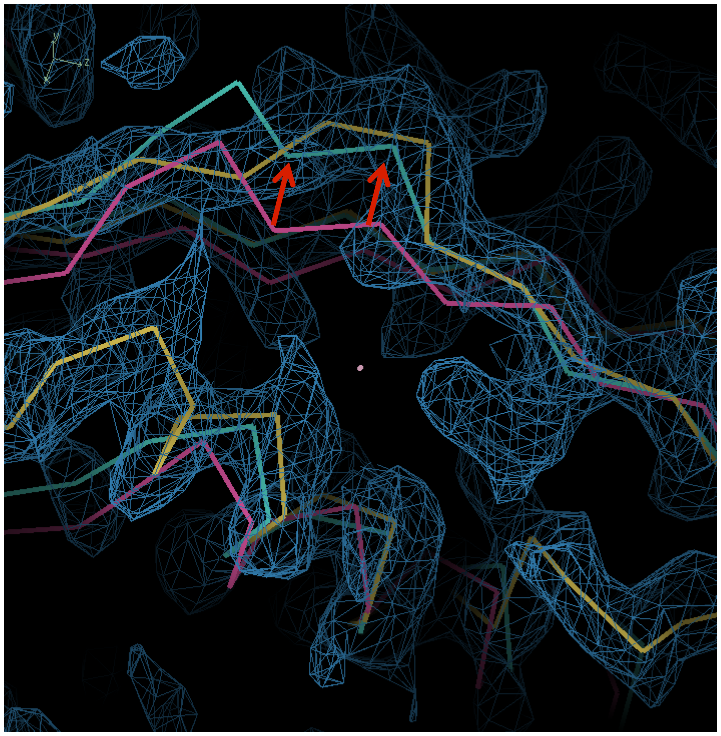
How do we decide what parts of the model to use?



Morph model to optimally fit map (maintaining connectivity of model)

Choose residues to delete based on local fit to density map (create map with autobuild)

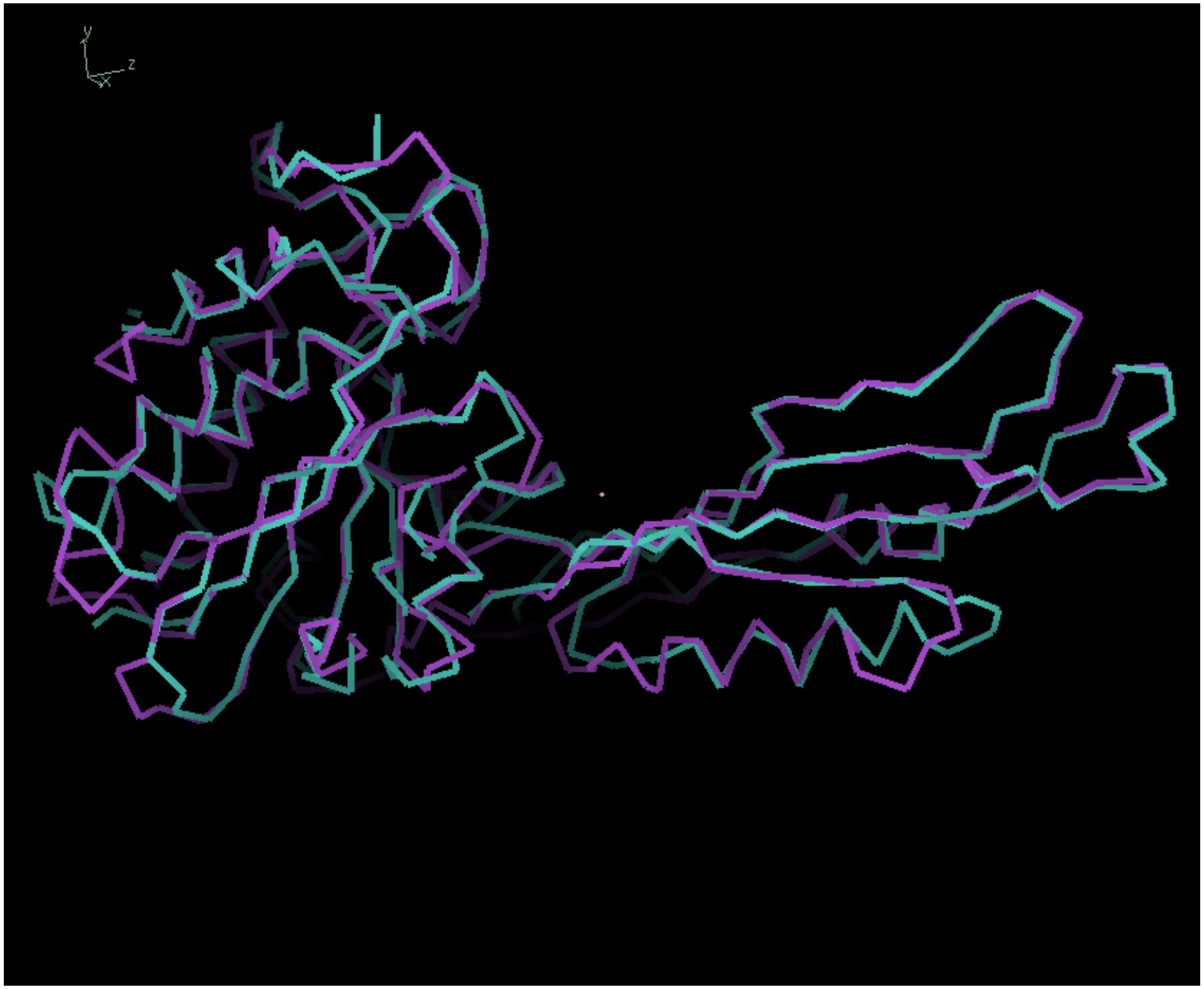
Morphing model to optimally fit map (maintaining connectivity of model)



Morph model by finding local distortions of model that improve fit to map

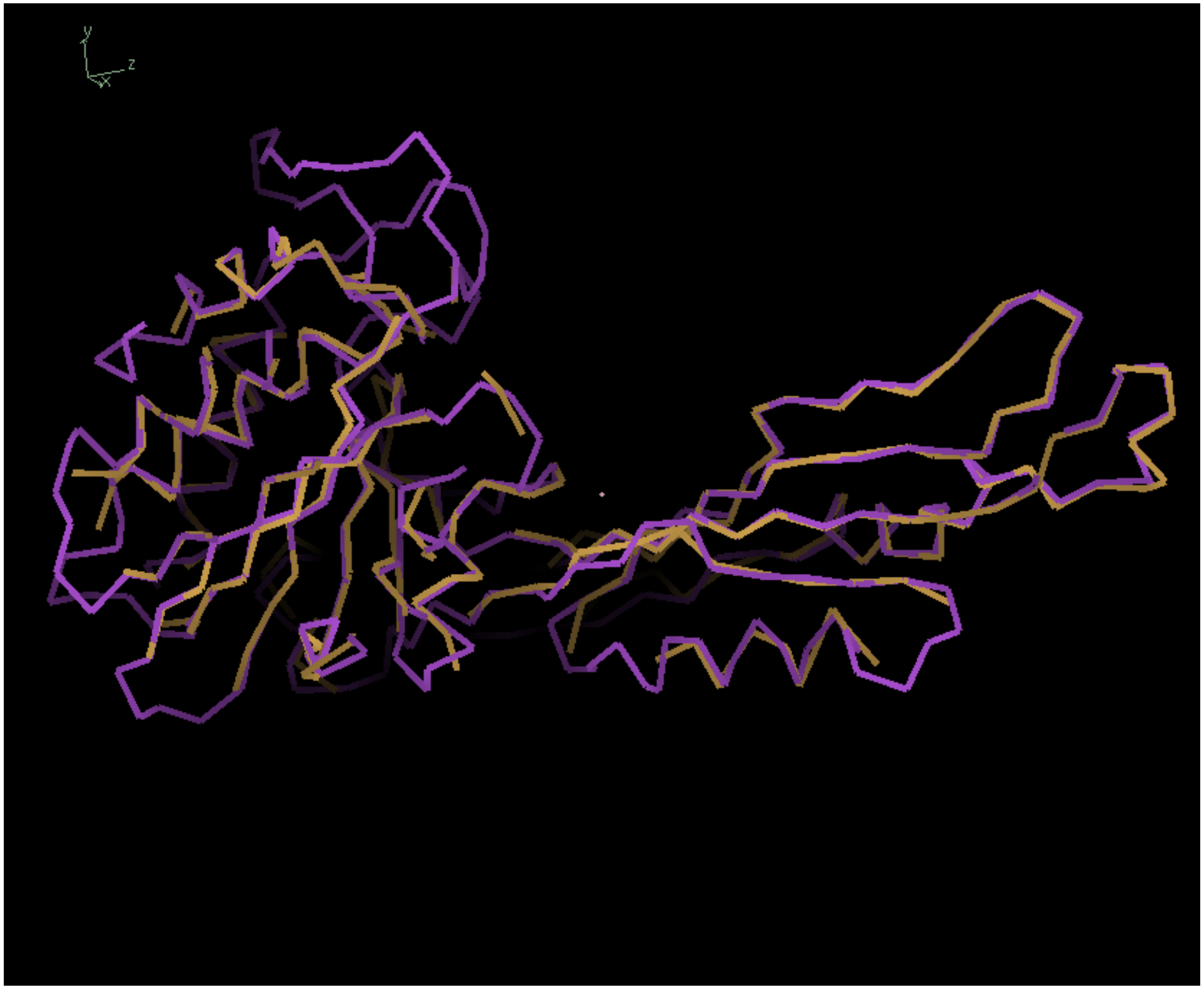
Create new map with autobuild starting from morphed model

Choose residues to delete based on fit to map (morphed and final models shown)



```
phenix.autobuild
data.mtz \
morph.pdb \
reject_weak=True \
min_weak_z=0.2 \
min_cc=0.4
```

Choose residues to delete based on fit to map (trimmed morphed model shown)



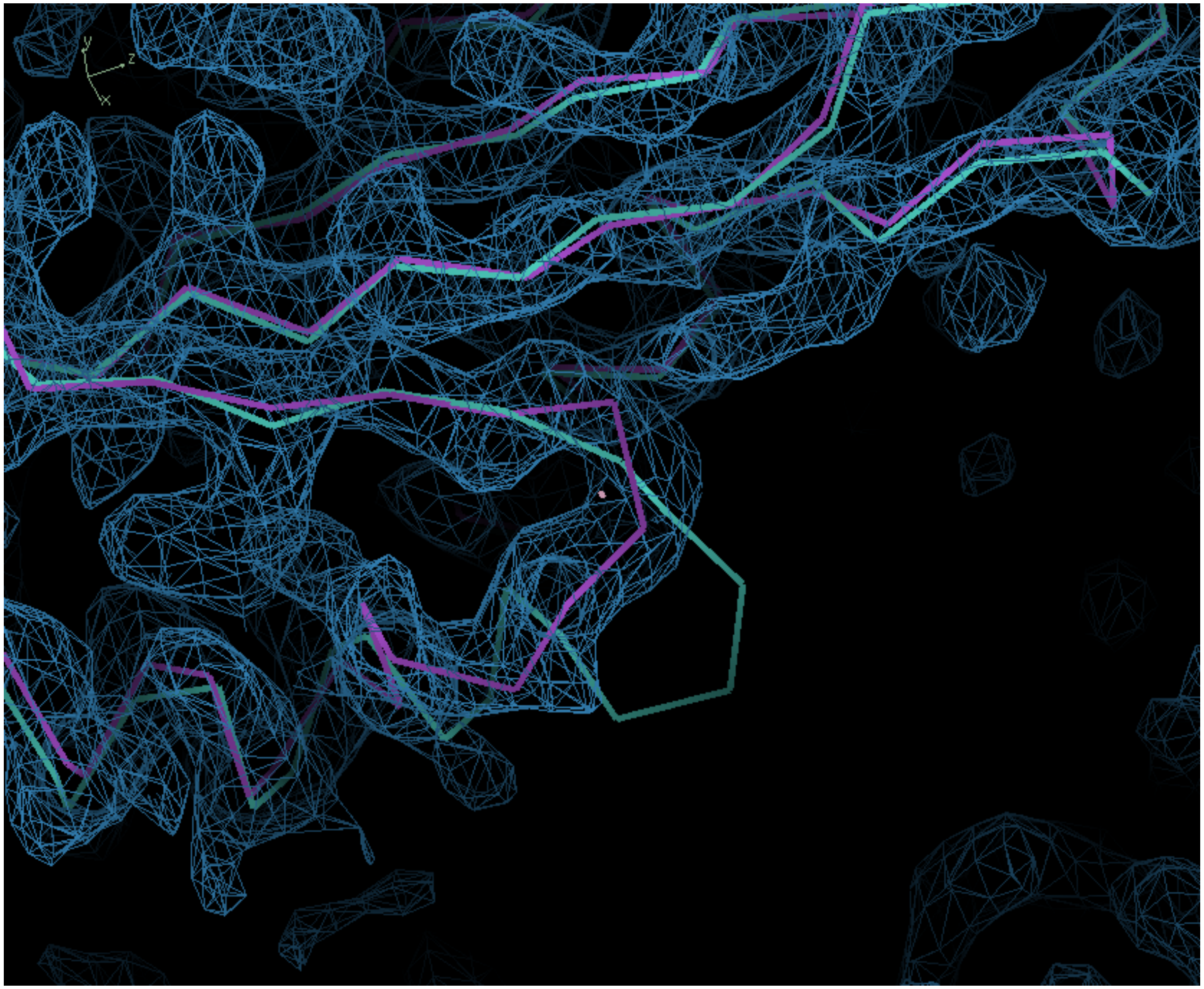
Remove if:

CC<0.4 or
$\rho < \rho_{main} - 0.2\sigma_{main}$

80/352 residues deleted

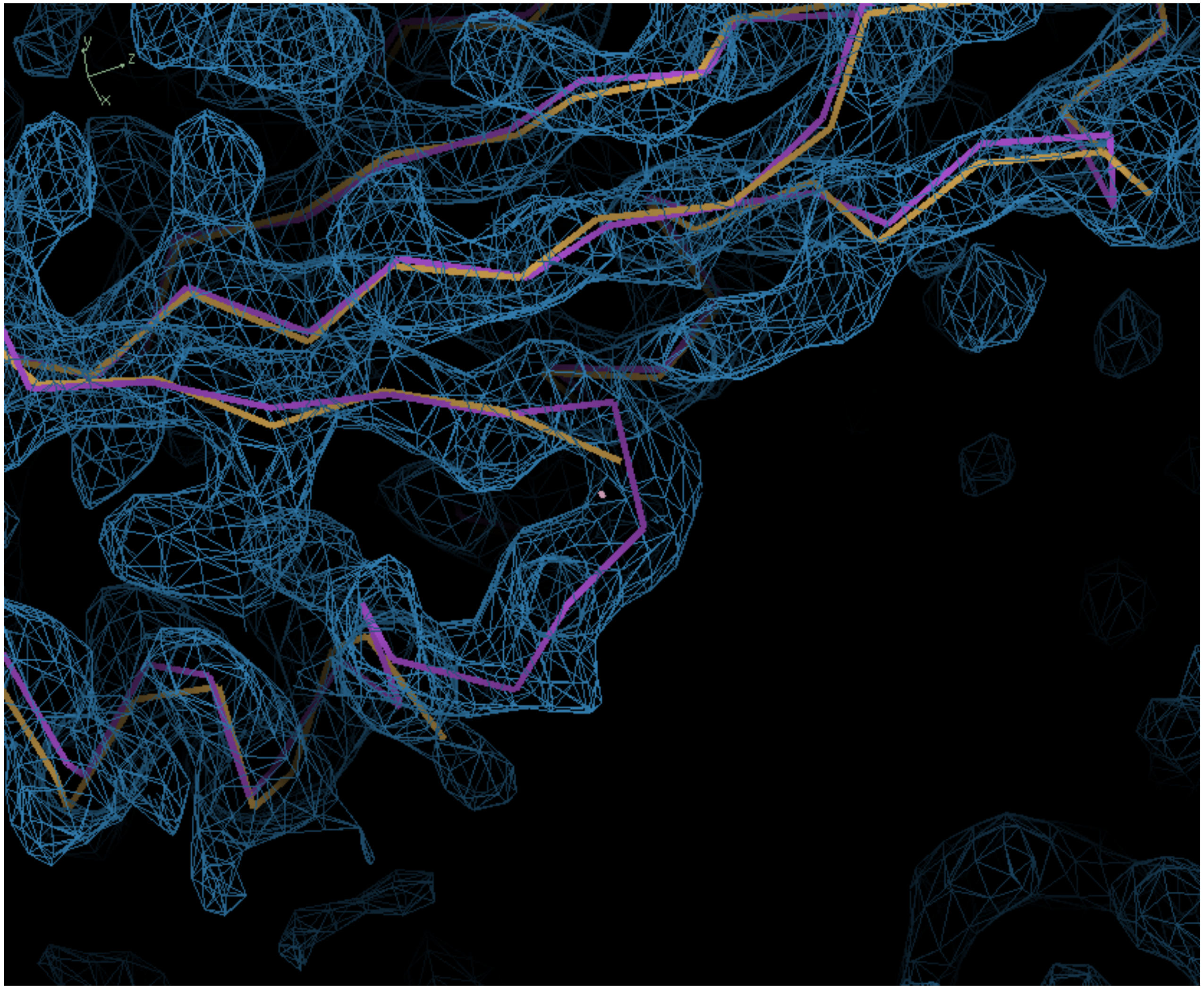
Remaining residues very close to final model

Choose residues to delete based on fit to map (closer view)



```
phenix.autobuild
data.mtz \
morph.pdb \
reject_weak=True \
min_weak_z=0.2 \
min_cc=0.4
```

Choose residues to delete based on fit to map (closer view)

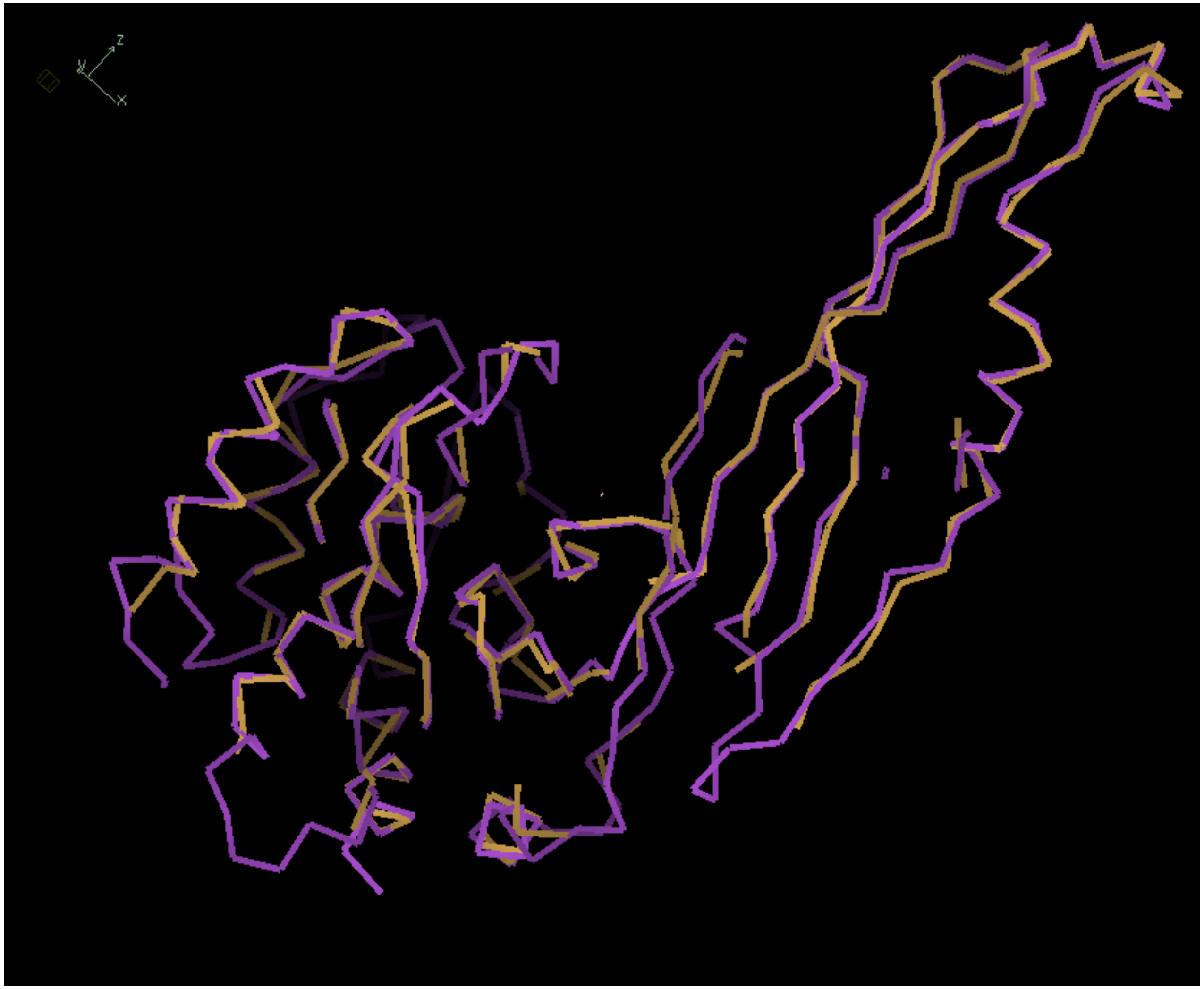


Remove if:

CC<0.4 or
$\rho < 0.5^* \rho_{main} + 0.2\sigma_{main}$

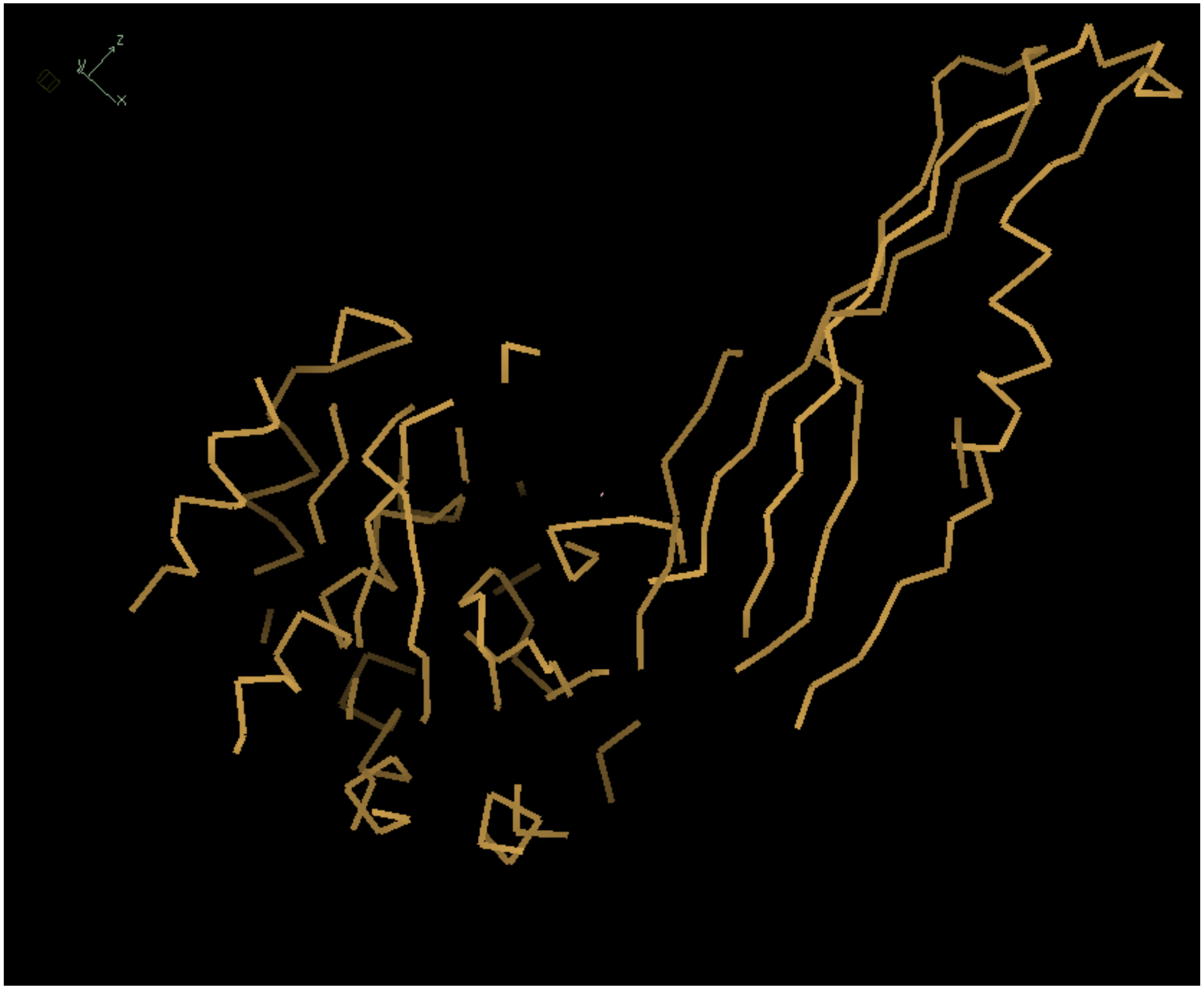
80/352 residues deleted

Trimmed model



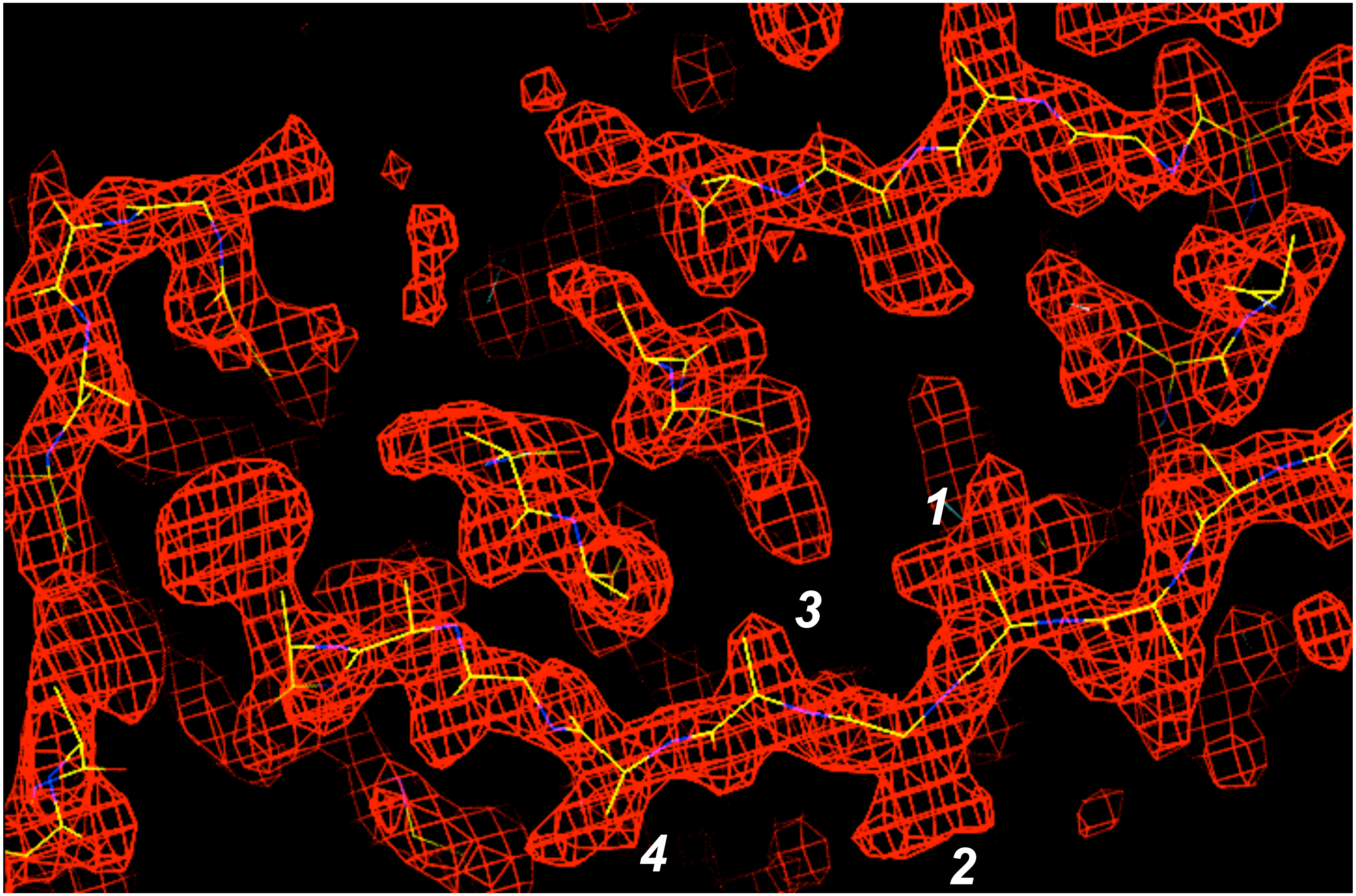
Trimmed model is very close to final model…

Trimmed model



…but sequence is not aligned…and connectivity is no longer obvious

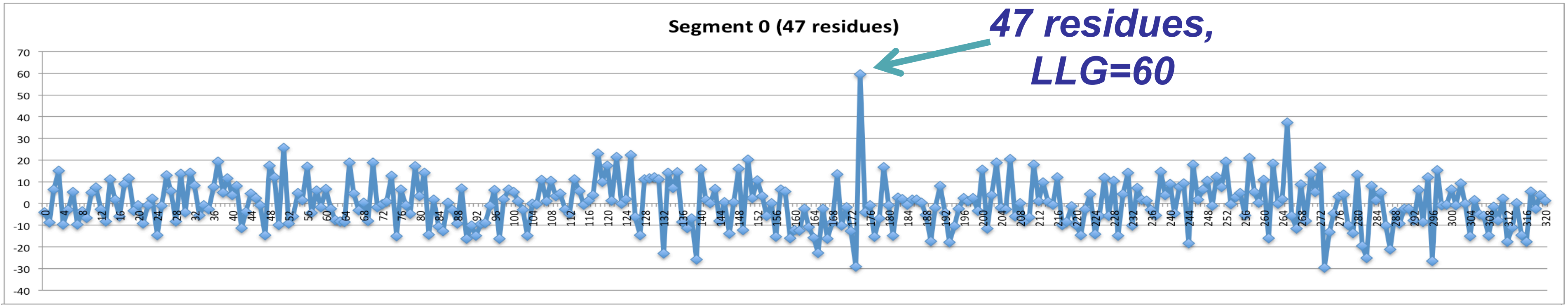
Probabilistic sequence assignment (*Resolve*)



Probability of each residue type at each position in sequence…

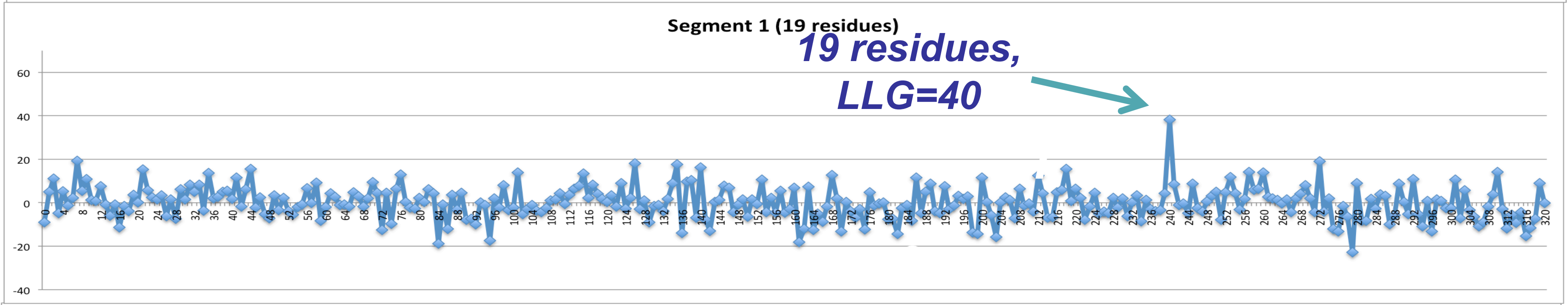
#	G	A	S	V	I	L	M	C	F	Y	K	R	W	H	E	D	Q	N	P	T
1	6	5	4	18	18	6	1	1	1	2	6	2	2	1	9	6	1	0	1	4
2	4	11	14	37	5	2	0	2	0	0	2	3	0	0	1	2	0	0	0	6
3	11	23	5	12	5	3	2	0	1	3	7	3	1	0	5	3	2	0	2	2
4	7	9	6	16	8	5	2	0	1	3	8	4	1	0	7	6	2	0	3	4
5	31	7	3	7	4	2	1	0	1	3	5	4	1	0	6	2	2	0	11	1
6	1	3	3	41	14	8	0	0	0	0	2	1	0	0	2	4	0	0	1	9
7	0	0	0	0	0	0	0	0	15	63	1	0	17	1	0	0	0	0	0	0
8	2	3	6	23	10	6	2	1	0	1	4	3	0	0	5	16	1	0	1	6
9	96	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

LLG for each possible start of a segment



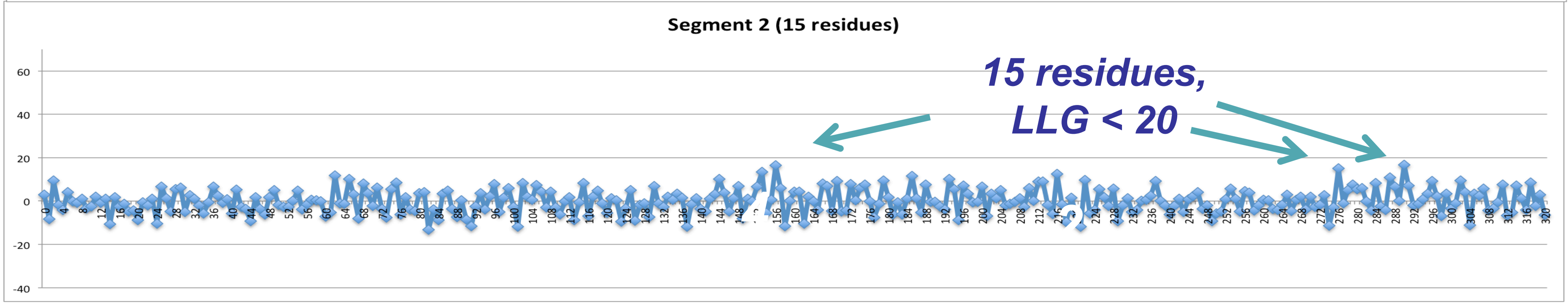
Segment 0 (47 residues)

47 residues, LLG=60



Segment 1 (19 residues)

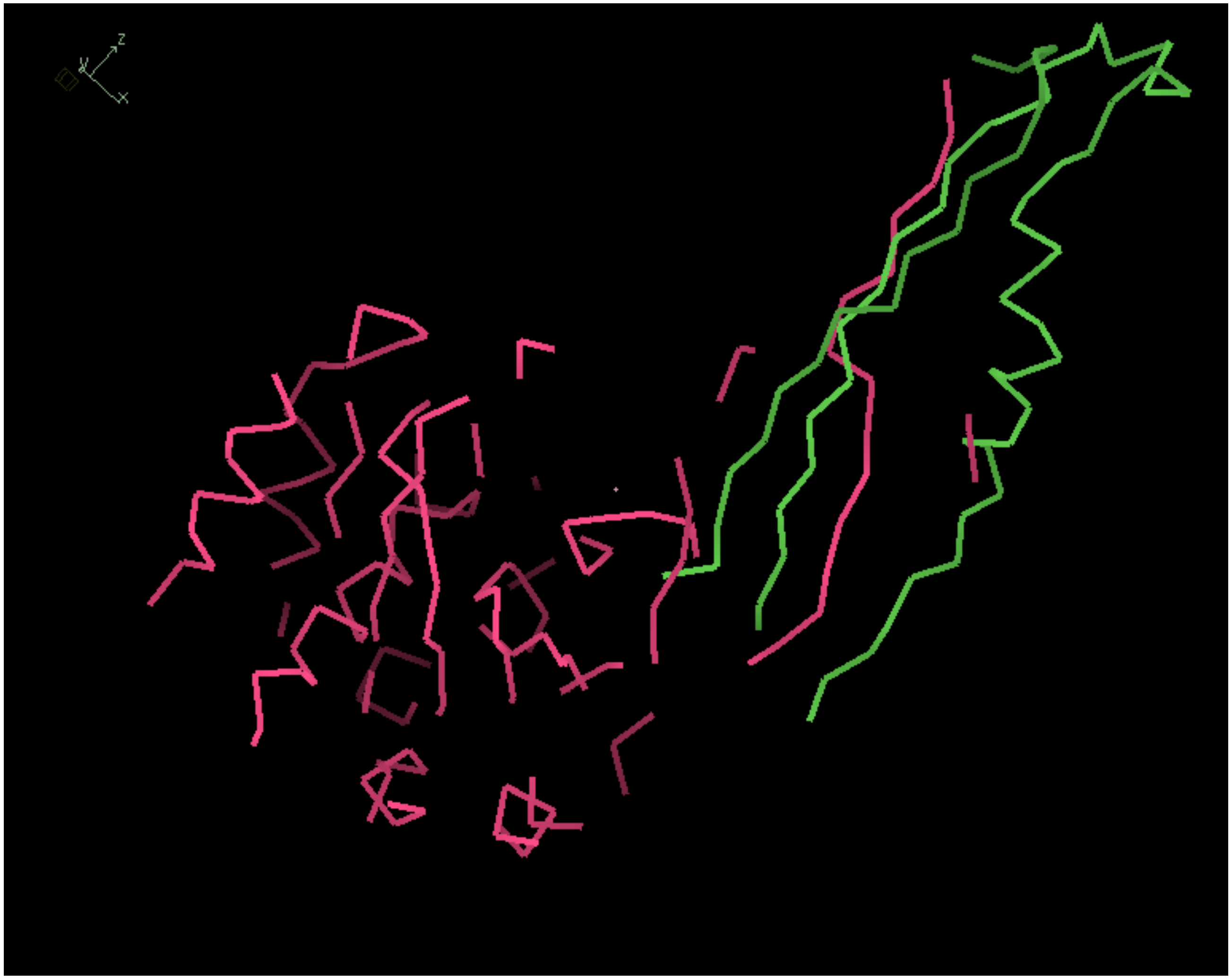
19 residues, LLG=40



Segment 2 (15 residues)

15 residues, LLG < 20

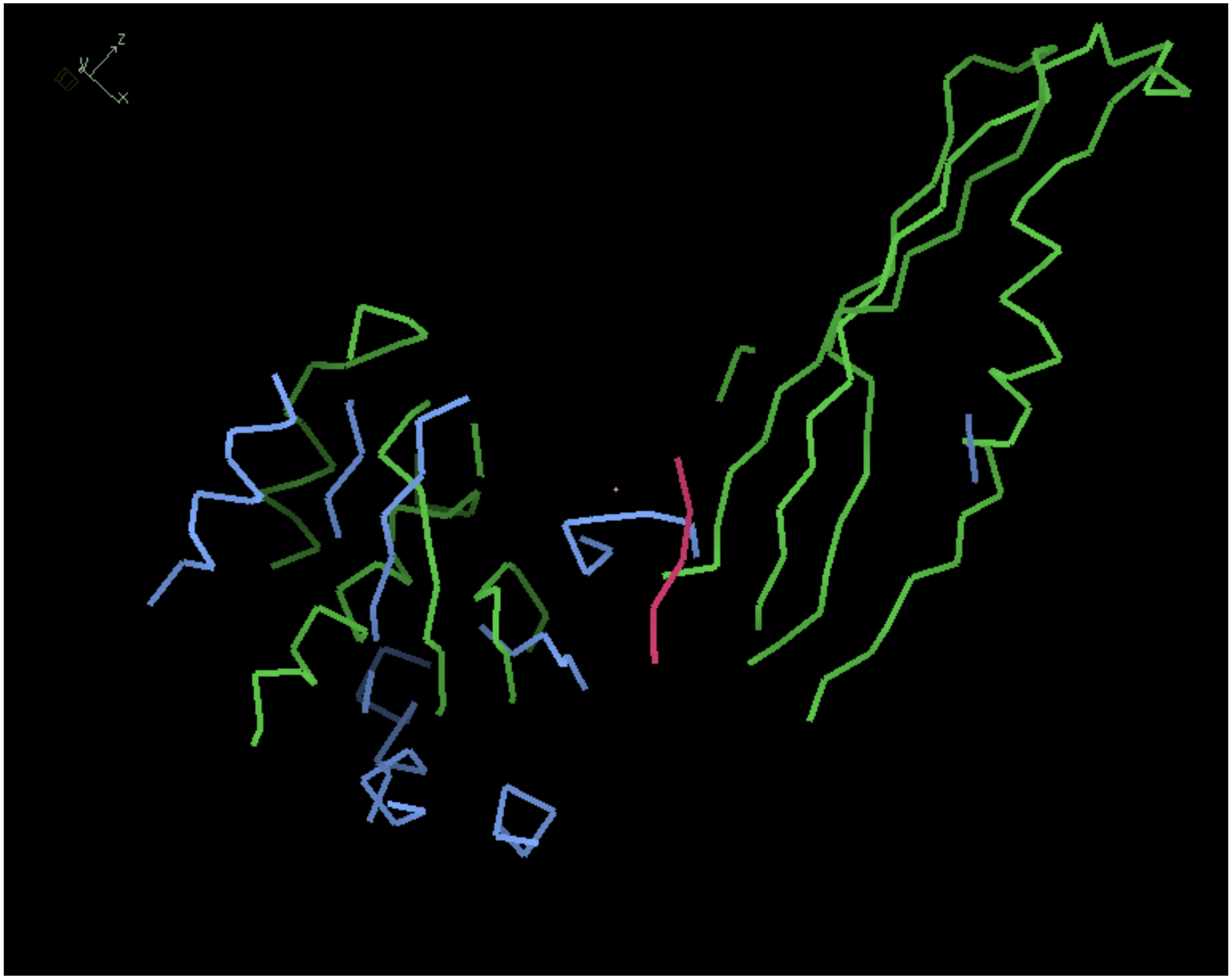
Sequence assignment using only fit of side-chains to density



69 residues assigned to sequence

206 not assigned

Sequence assignment not allowing overlap, and scoring for loops



164 residues assigned to sequence

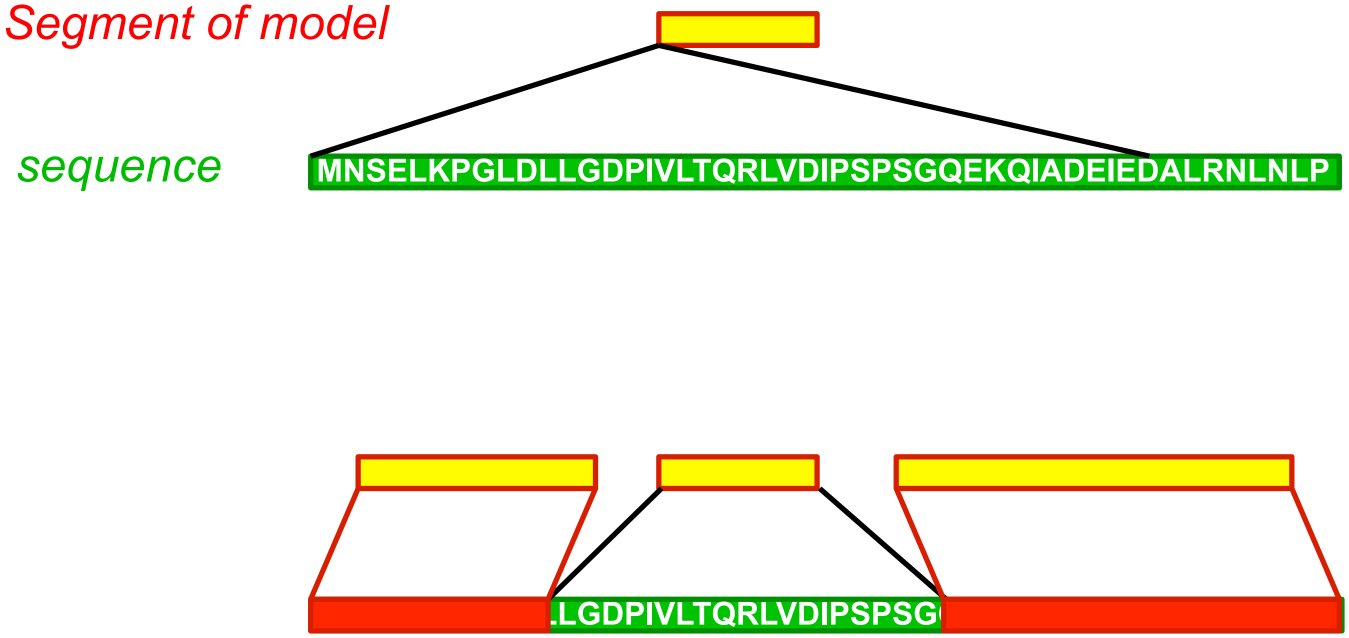
84 not assigned

6 incorrectly-assigned

If we start with a homology model…

We know the order of the segments

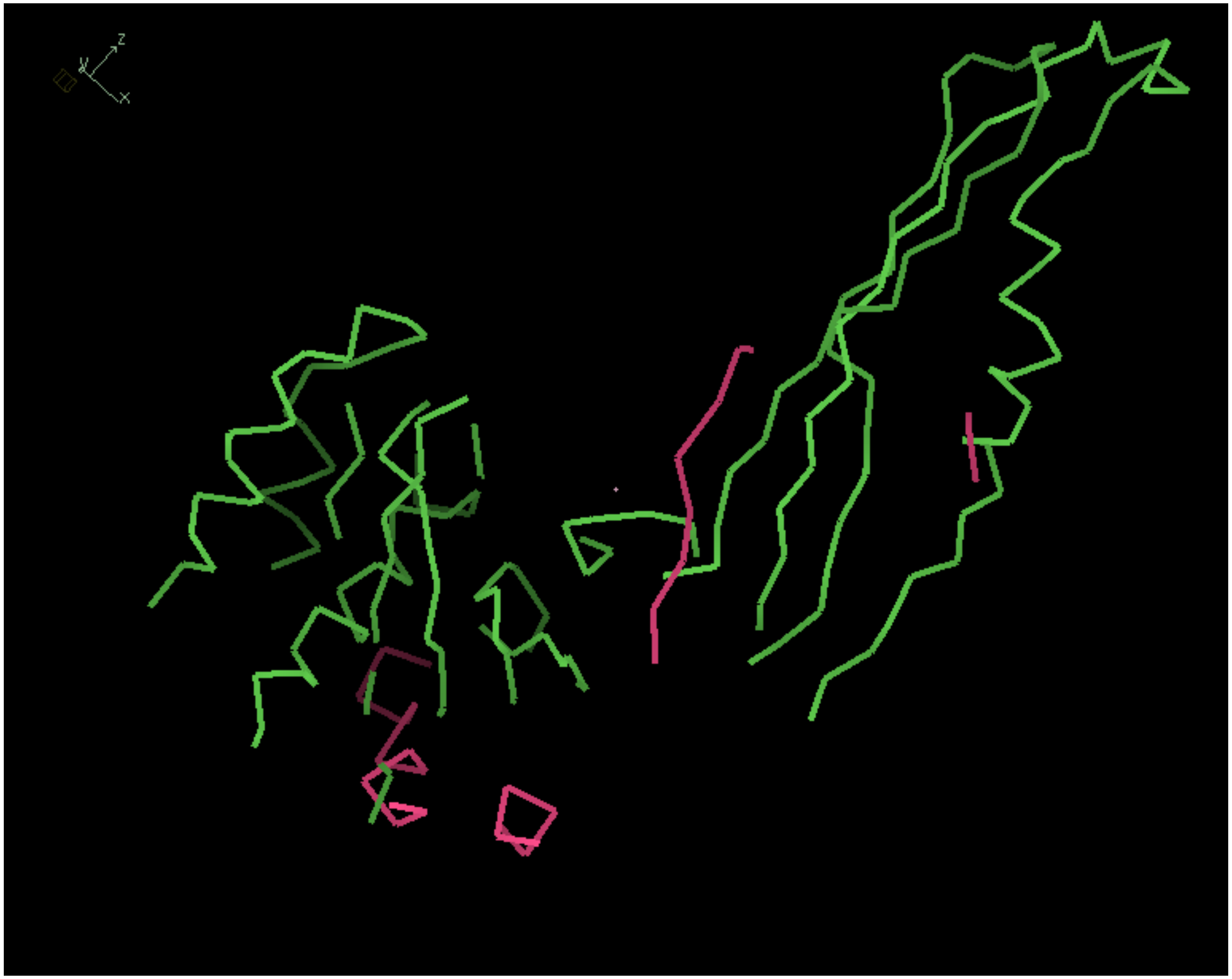
This vastly reduces the number of possible arrangements.



Order not known…
Segment can
go anywhere

Order known…
Many locations
Excluded by
Other segments

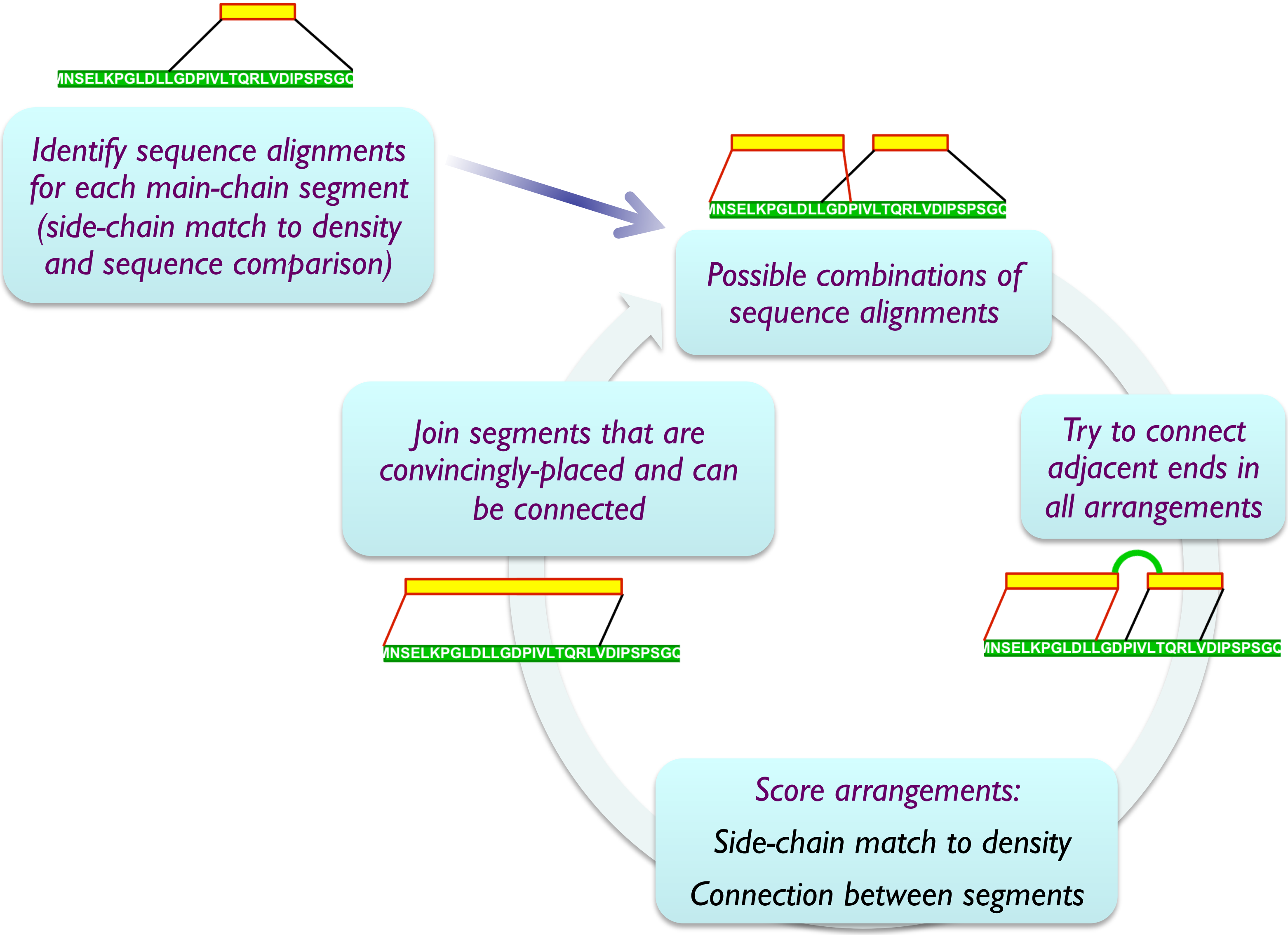
Including known order of segments



207 residues assigned to sequence

39 not assigned

Iterative sequence assignment



Identify sequence alignments for each main-chain segment (side-chain match to density and sequence comparison)

Possible combinations of sequence alignments

Try to connect adjacent ends in all arrangements

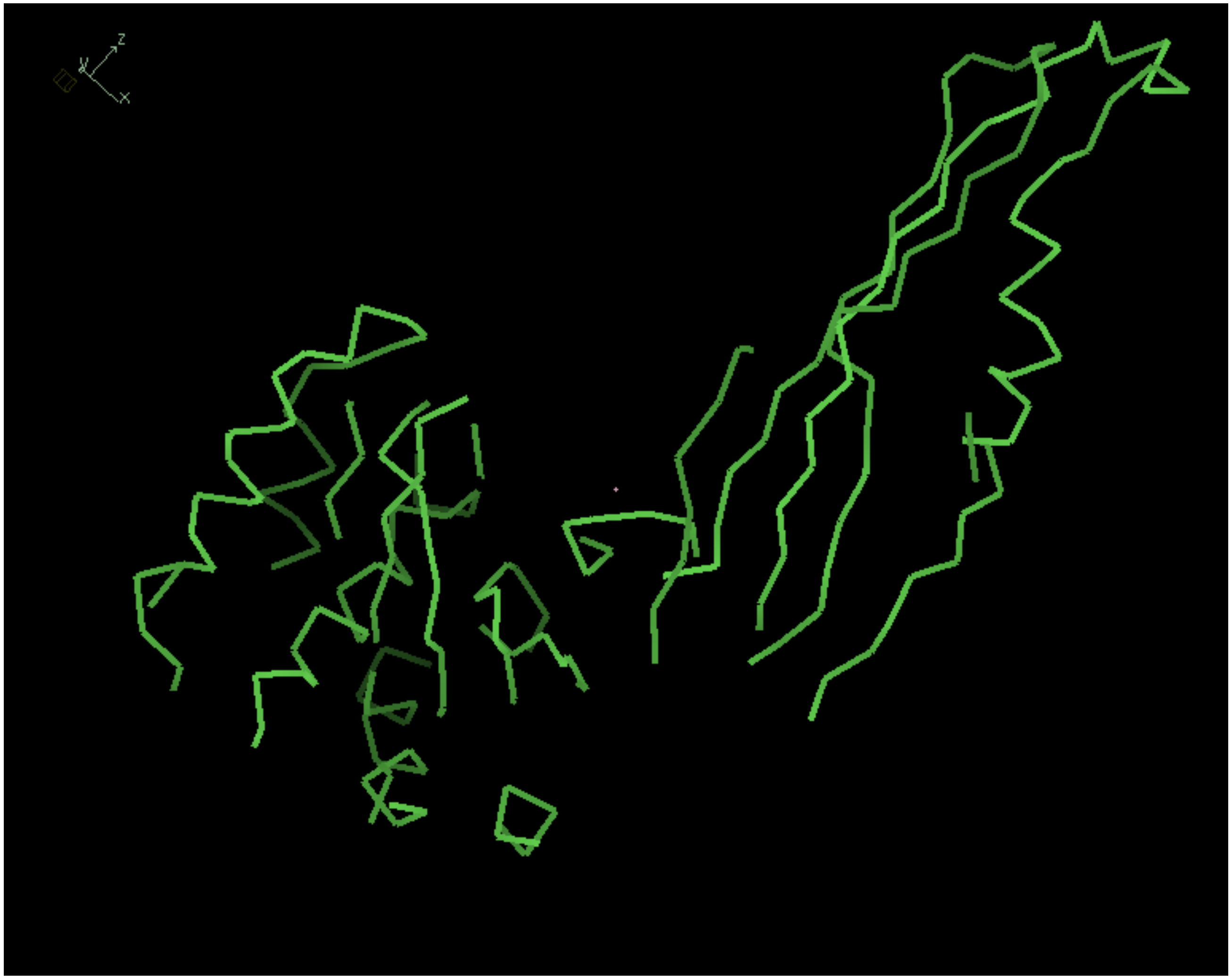
Score arrangements:

Side-chain match to density

Connection between segments

Join segments that are convincingly-placed and can be connected

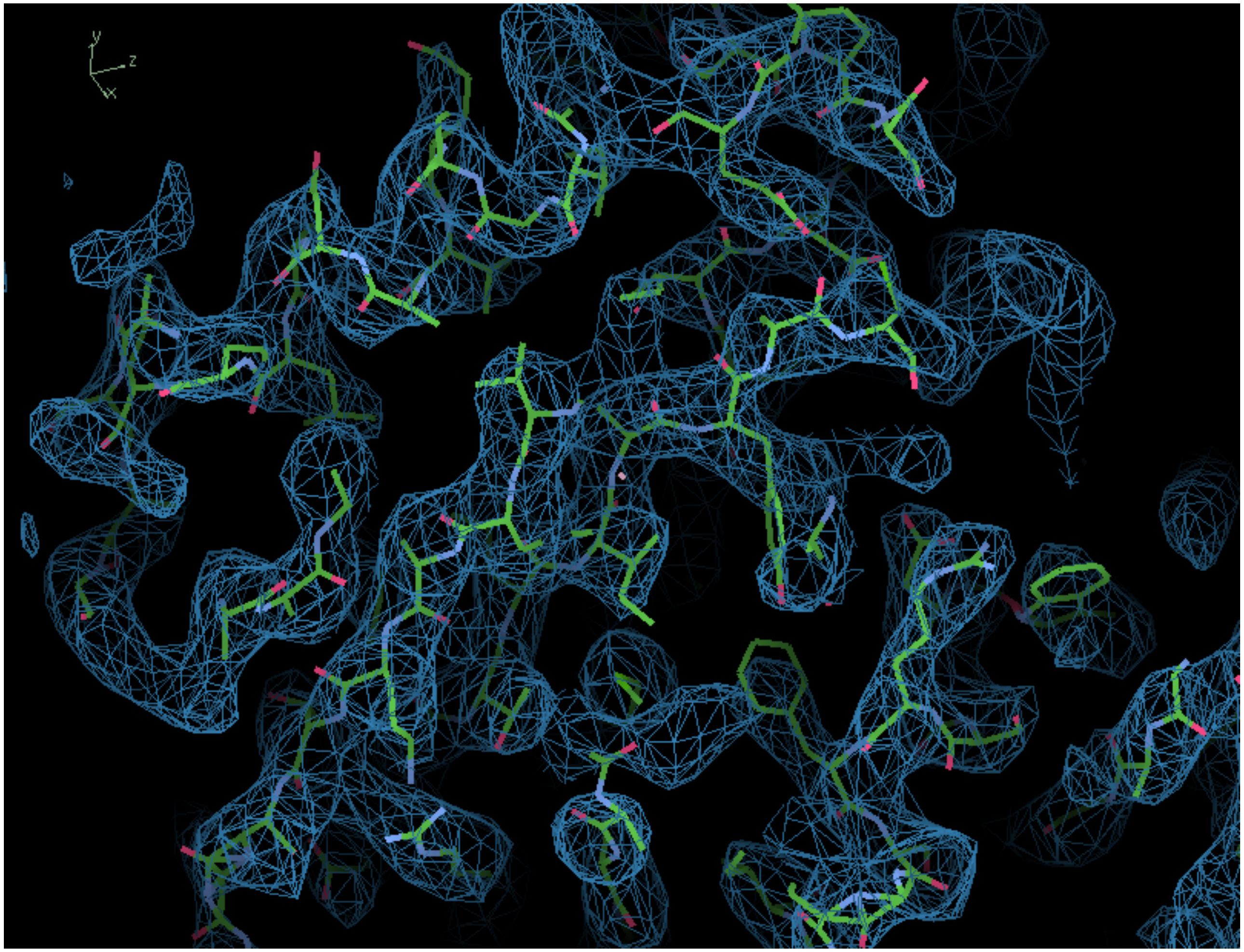
No overlap, loops, keep order of segments, iterate



262 residues assigned to sequence

0 not assigned

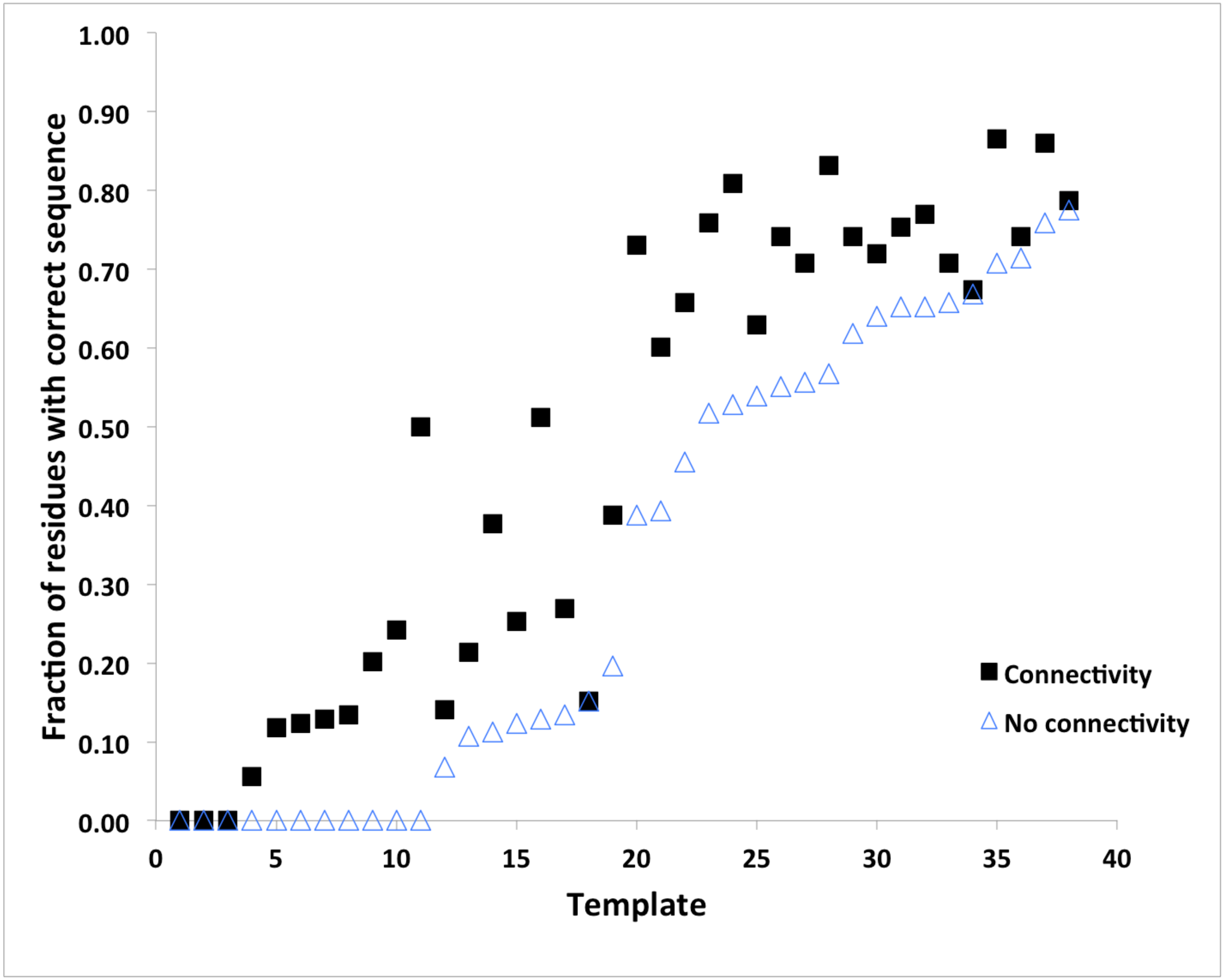
Result…



Fully correct assignment of all parts of starting model to sequence…

Morphing, then sequence assignment on a series of templates

(1A2B; template sequence identity 7%-33%)

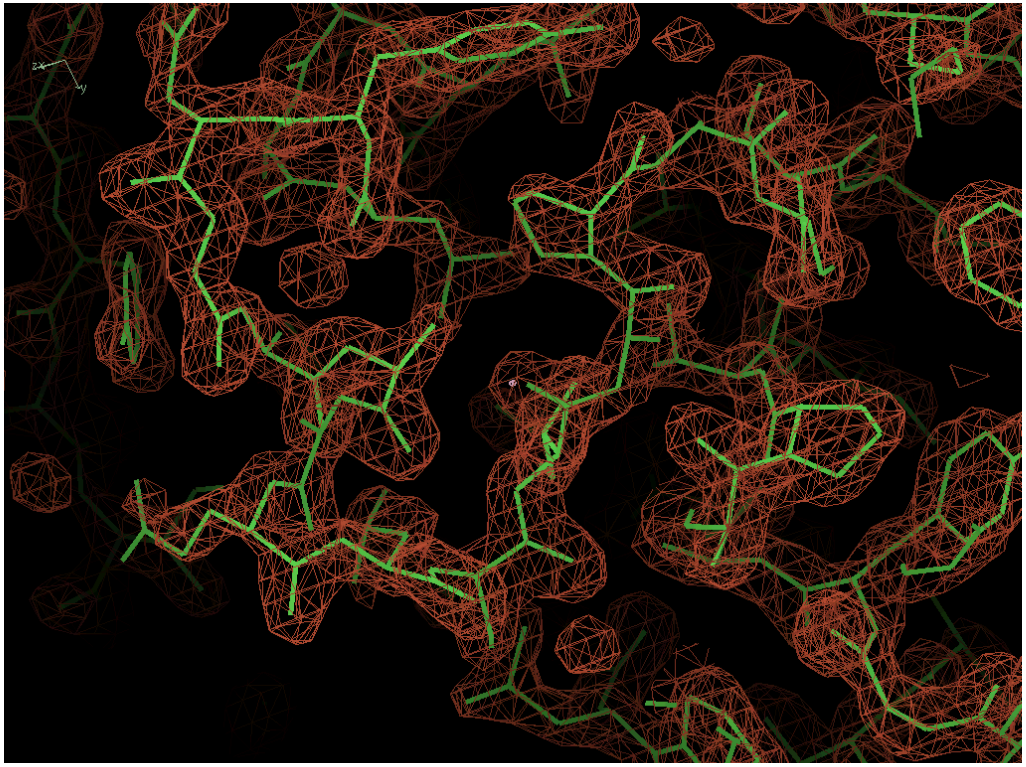


Applications for morphing

Molecular replacement templates that are close but distorted

Building models into experimental electron density maps when a distant related structure is available

Generalized mapping of one structure to another – can apply to coordinates or density



Thanks for data to...

Alex Wlodawer, NCI (XMRV PR)
Herb Axelrod, Debanu Das, JCSG (hp3342)
Gustav Oberdorfer, Ulrike Wagner, Univ. of Graz
Eugene Valkov, Univ. of Cambridge
Assaf Alon, Deborah Fass, Weizmann Institute of Science
Sergey M. Vorobiev, NESG
Hideo Iwai, Univ. of Helsinki
P. Raj Pokkuluri, Argonne National Laboratory

Scripts, documentation, and data for *phenix.morph_model* and *phenix.mr_rosetta* are available at...

http://www.phenix-online.org

Acknowledgements

Frank DiMaio, David Baker (Univ. of Washington)

Randy Read (Cambridge University)

Paul Adams, Pavel Afonine (Lawrence Berkeley National Laboratory)

Axel Brunger (Stanford University)

Li-Wei Hung (Los Alamos National Laboratory)

The PHENIX Project

Phenix

Lawrence Berkeley Laboratory

Paul Adams, Ralf Grosse-Kunstleve, Pavel Afonine, Nat Echols, Nigel Moriarty, Jeff Headd, Nicholas Sauter, Peter Zwart, Richard Gildea

Los Alamos National Laboratory

Tom Terwilliger, Li-Wei Hung

BERKELEY LAB

Los Alamos
NATIONAL LABORATORY

UNIVERSITY OF
CAMBRIDGE

Randy Read, Airlie McCoy, Gabor Bunkoczi, Rob Oeffner

Cambridge University

Duke University

Jane & David Richardson, Vincent Chen, Chris Williams, Bryan Arendall, Swati Jain, Bradley Hintze

An NIH/NIGMS funded Program Project