



Roberto Marabini



Marcos Grajera



Pablo Conesa

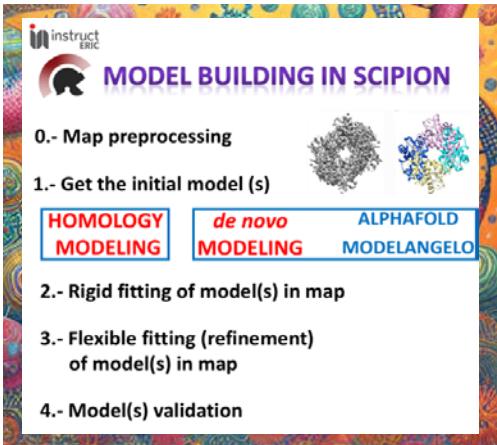


Marta Martínez

I2PC COURSE ON SINGLE PARTICLE ANALYSIS AND MODELLING BY CRYO-EM

CNB-CSIC, Madrid, June 16th – June 20th, 2025

I2PC Course on Single Particle analysis and modelling by Cryo-EM. Madrid, June 16 - 20, 2025



Model Building Agenda

June 18th

- 14:00 - 14:45: [General introduction to model building in cryo-EM](#) (M)
 - Example 1: Naive model with [Coot](#)
- 14:45 - 15:15: Introduction to the [data](#) used in the practical sessions (human_haemoglobin) (details [here](#)) (M)
 - Review Kahoot [students, teacher](#) (R 000)
- 15:15:16:00
 - [Homology modeling workflow](#) (R)
 - [Import map and sequence, Input data](#) (R)
 - [Map preprocessing](#) (M)

TAKE HOME MESSAGE: Try to improve the visual aspect of your map in terms of connectivity and/or details before tracing it.

- 16:00 - 16:30: coffee break
- 16:30 - 18:30: Getting a first estimate of the model "de novo" ([AlphaFold](#) & [ModelAngelo](#)) and by [sequence homology](#). (R-M)

TAKE HOME MESSAGE: Use the structure of a good homologous sequence if you have it, otherwise try ModelAngelo if your map's resolution is good enough, alternatively try to predict the structure with AlphaFold. Remember to double check the model with the experimental data.

- 18:00: End of the day

June 19th

- 09:30-09:45: Review yesterday's work (R)
- 09:45-10:30: Structure analysis: [Structure comparison](#) (M)

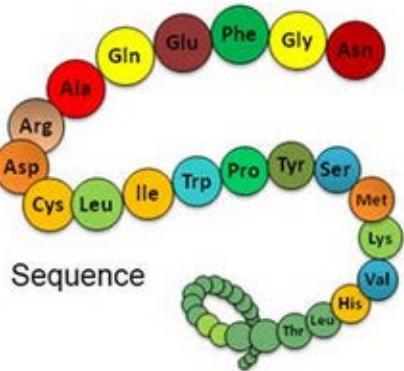
TAKE HOME MESSAGE: Models are always wrong... but some are useful.

- 10:30-11:00: [Rigid Fitting of Initial Models](#) (R)
- 11:00-11:30: Coffee break
- 11:30-13:00: [Flexible fitting \(Coot\)](#) and [Validation](#)
 - 11:30-12:15 [Coot Tutorial](#) (R)
 - 12:15-13:00: [Flexible fitting \(Phenix & Refmac\)](#) and [Validation](#) (M-M)
- 13:00-14:00: Lunch Break
 - 14:00-15:00: [Flexible fitting \(Phenix & Refmac\)](#) and [Validation](#) (M-M) (cont)
 - 15:00-15:30: Review Kahoot [students, teacher](#) (R: 0001 - revision)
- 15:30-16:00: [Model building of human Hgb beta subunit](#) (R-M)
- 16:00-16:30: Coffee break
- 16:30-17:00: [Model building of human Hgb beta subunit](#) (R-M) (cont)
- 17:00-18:00 Building the model of the whole protein
 - [Chain merge & symmetrization](#) (R)
- 18:00:End of the day

June 20th

- 09:30-09:45: Review yesterday's work (R)
- 9:45-10:30: [Submission of structures to EMDB](#). Understanding the validation report. [Haemoglobin PDB full validation report](#) (R)
 - [Comparison validation scores PDB structure](#)(M)
- 10:30-11:00: Course Questionnaire (R)
- 11:00-11:30: Coffee break
- 11:30-12:30: Structure analysis: [chain contacts](#) Structural search with [DALI](#) Structural search with [FOLDSEEK](#) (M)
- 13:00-14:00: Lunch
- 14:00: End of the Course
- XX:XX: * Review Kahoot [students, teacher](#) (R: Review Model Building)

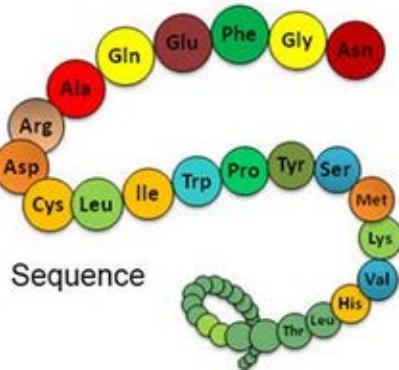
Our ultimate goal: To characterize the structure of macromolecules



Why is it important to characterize the structure of macromolecules (protein complexes)?

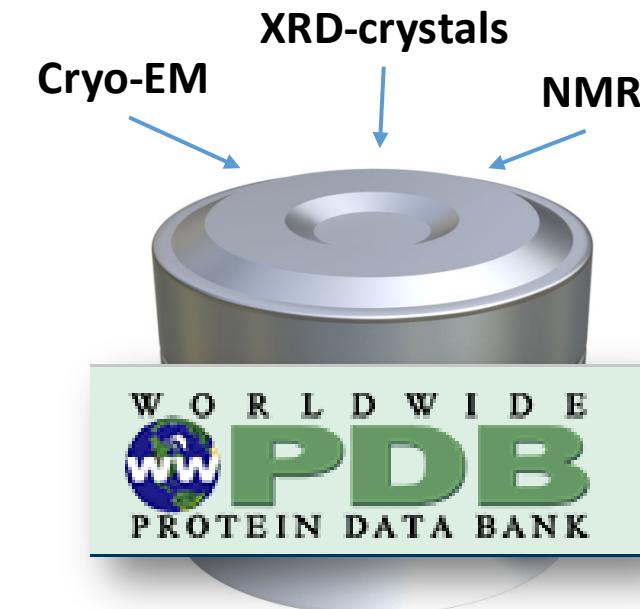
- Clues about function and dynamics (sensible domains/motifs)
- Clues about interacting partners (complexes)
- Deleterious effect of mutations (diseases)
- Designing of inactivating drugs
- Eventually: Basic knowledge to train AI methods
 - Predict structures from sequence (AlphaFold)
 - Design new proteins

Our ultimate goal: To characterize the structure of proteins



Why is it important to characterize the structure of macromolecules (protein complexes)?

- Clues about function and dynamics (sensible domains/motifs)
 - Clues about interacting partners (complexes)
 - Deleterious effect of mutations (diseases)
 - Designing of inactivating drugs
-
- Eventually: Basic knowledge to train AI methods
 - Predict structures from sequence (AlphaFold)
 - Design new proteins



Q-scores

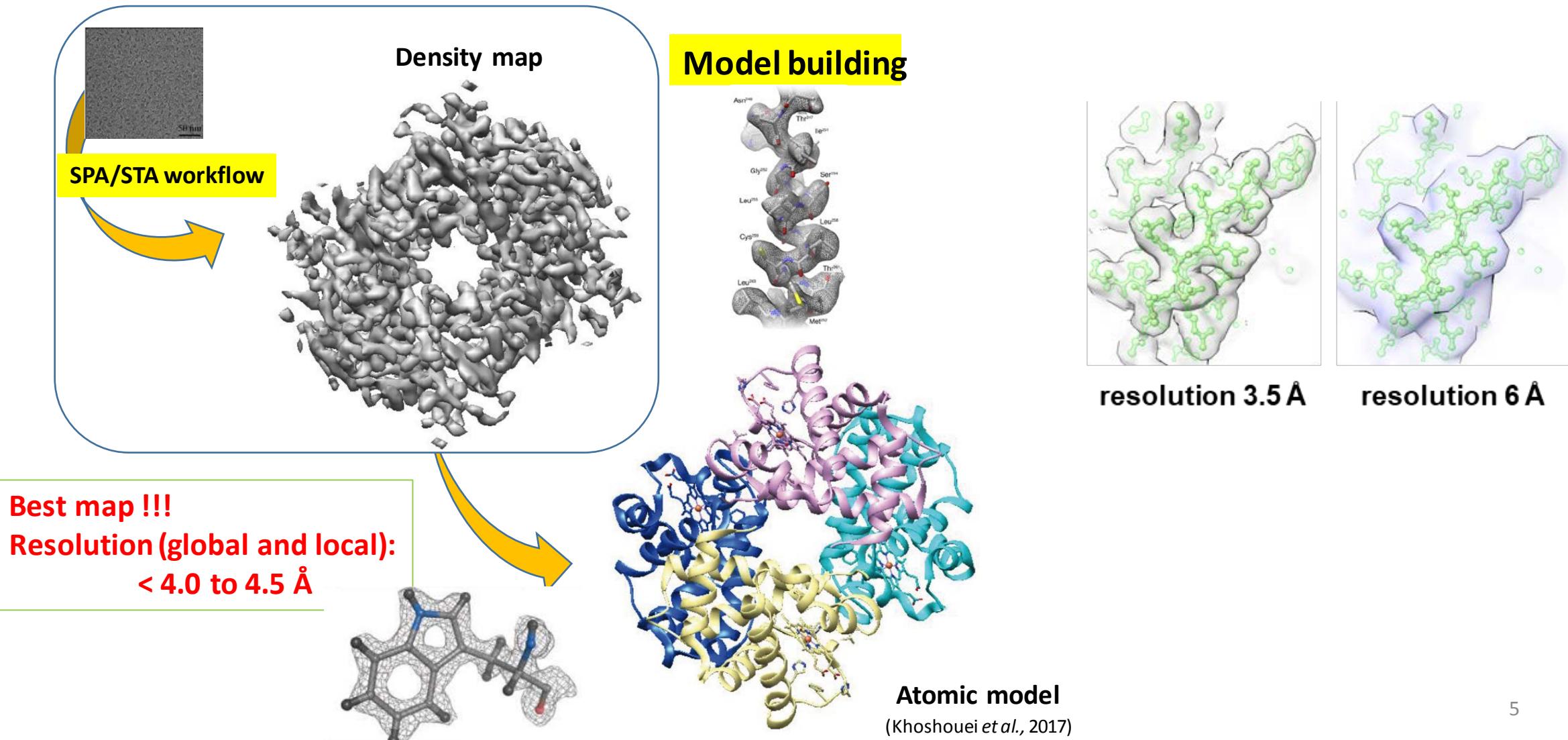
237,777
Structures
 from the
PDB archive

Computed
Structure
Models (CSM)
 ModelArchive

Integrative modeling: combines spatial restraints derived from a variety of biophysical techniques to determine the 3D structure of complex ⁴ macromolecular assemblies.

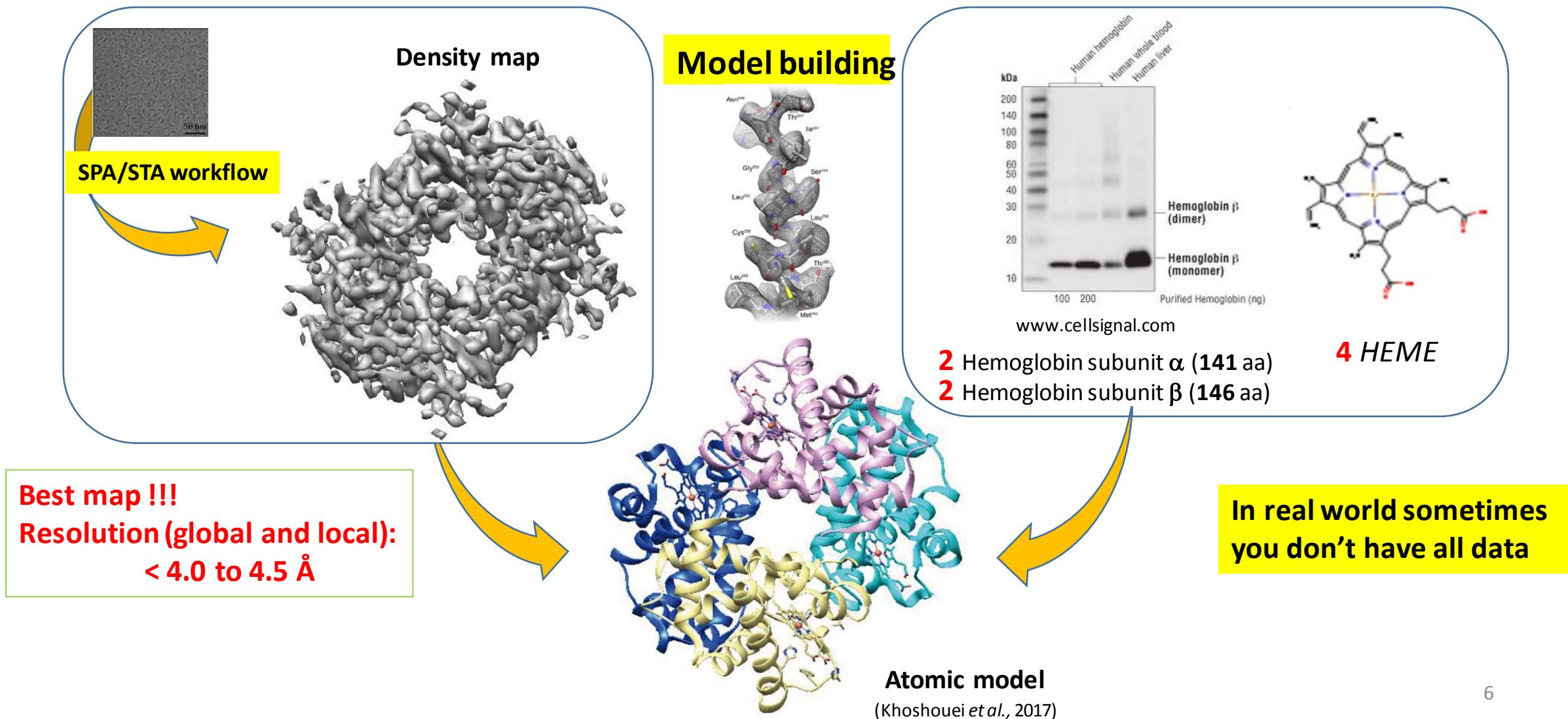
Modeling proteins from cryo-EM data

Approach to the problem: Atomic interpretation of the map



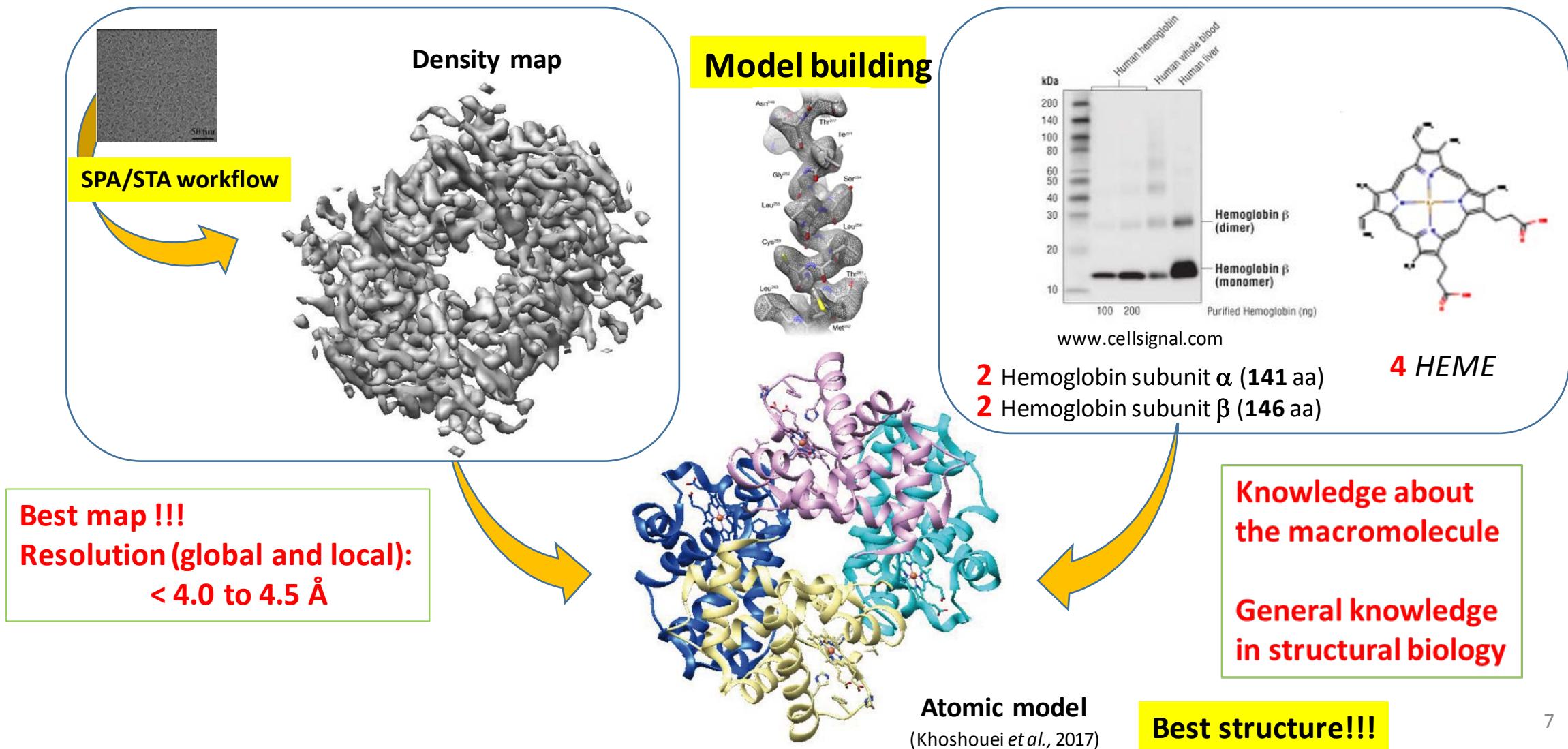
Modeling proteins from cryo-EM data

Approach to the problem: Atomic interpretation of the map



Modeling proteins from cryo-EM data

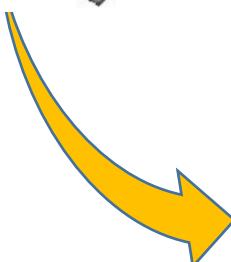
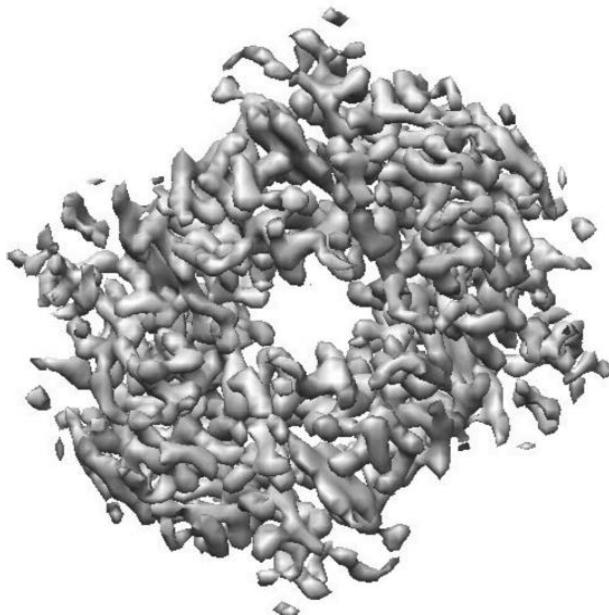
Approach to the problem: Atomic interpretation of the map



Modeling proteins from cryo-EM data

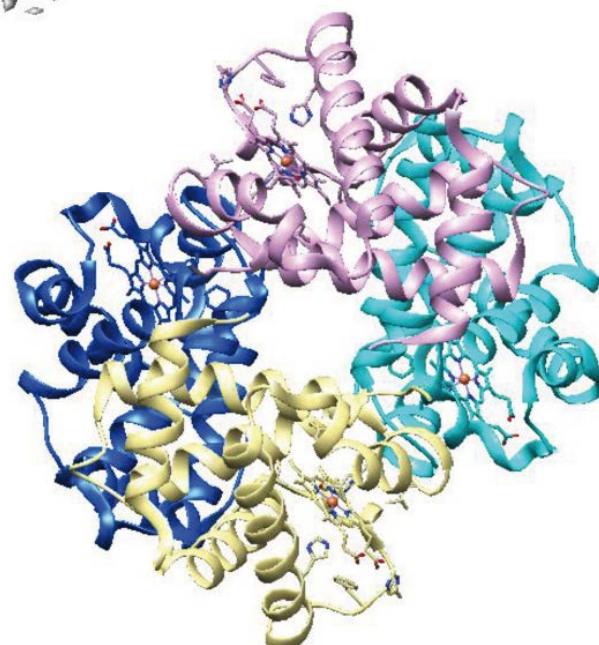
Approach to the problem: Atomic interpretation of the map

Electron density map



Atomic model

(Khoshouei *et al.*, 2017)



"All models are wrong, but some are useful"

(George P. Box)



5NI1

| pdb_00005ni1



CryoEM structure of haemoglobin at 3.2 Å determined with the Volta phase plate

PDB DOI: <https://doi.org/10.2210/pdb5NI1/pdb> Entry: 5NI1 supersedes: 5ME2

EM Map EMD-3488: EMDB EMDataResource

Classification: OXYGEN TRANSPORT

Organism(s): Homo sapiens

Expression System: Homo sapiens

Mutation(s): No

Deposited: 2017-03-22 Released: 2017-04-12

Deposition Author(s): Khoshouei, M., Radjainia, M., Bunker, R., Baumeister, W., Danev, R.

Experimental Data Snapshot

Method: ELECTRON MICROSCOPY

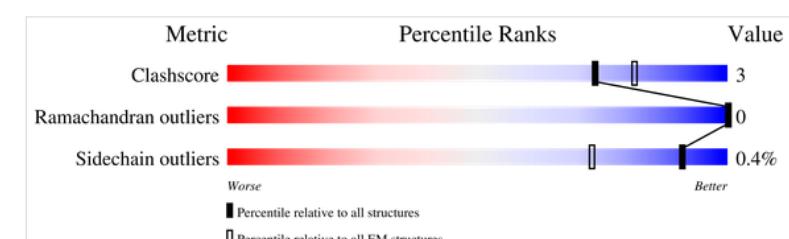
Resolution: 3.20 Å

Aggregation State: PARTICLE

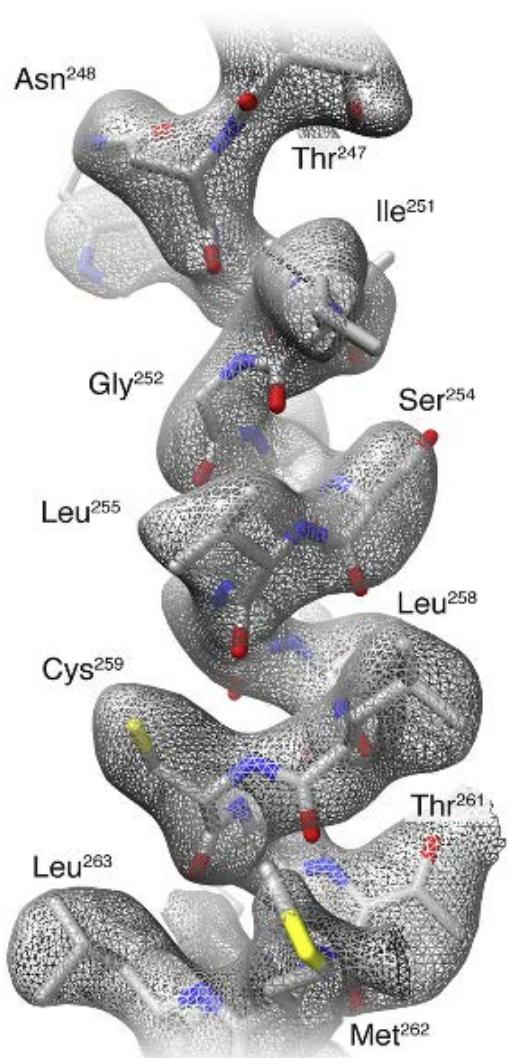
Reconstruction Method: SINGLE PARTICLE

wwPDB Validation

[3D Report](#) [Full Report](#)



Modeling proteins from cryo-EM data



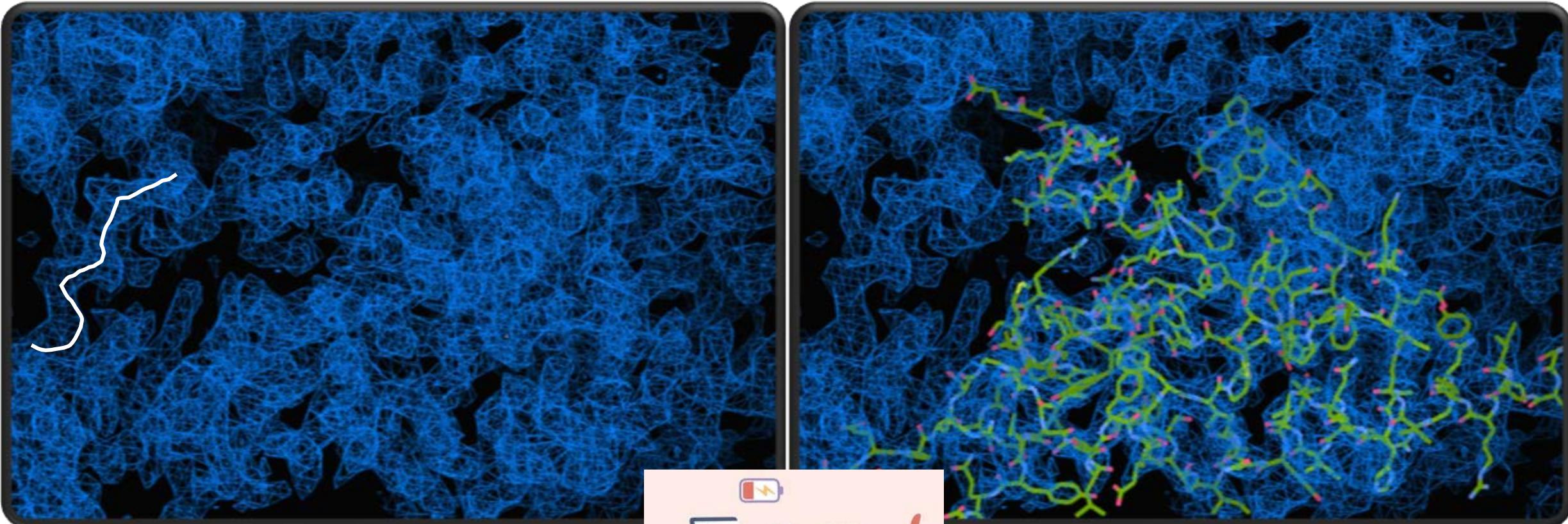
Model building

Basic steps (workflow)

Tools

Modeling proteins from cryo-EM data

https://github.com/I2PC/scipion-courses/wiki/modeling_with_coot



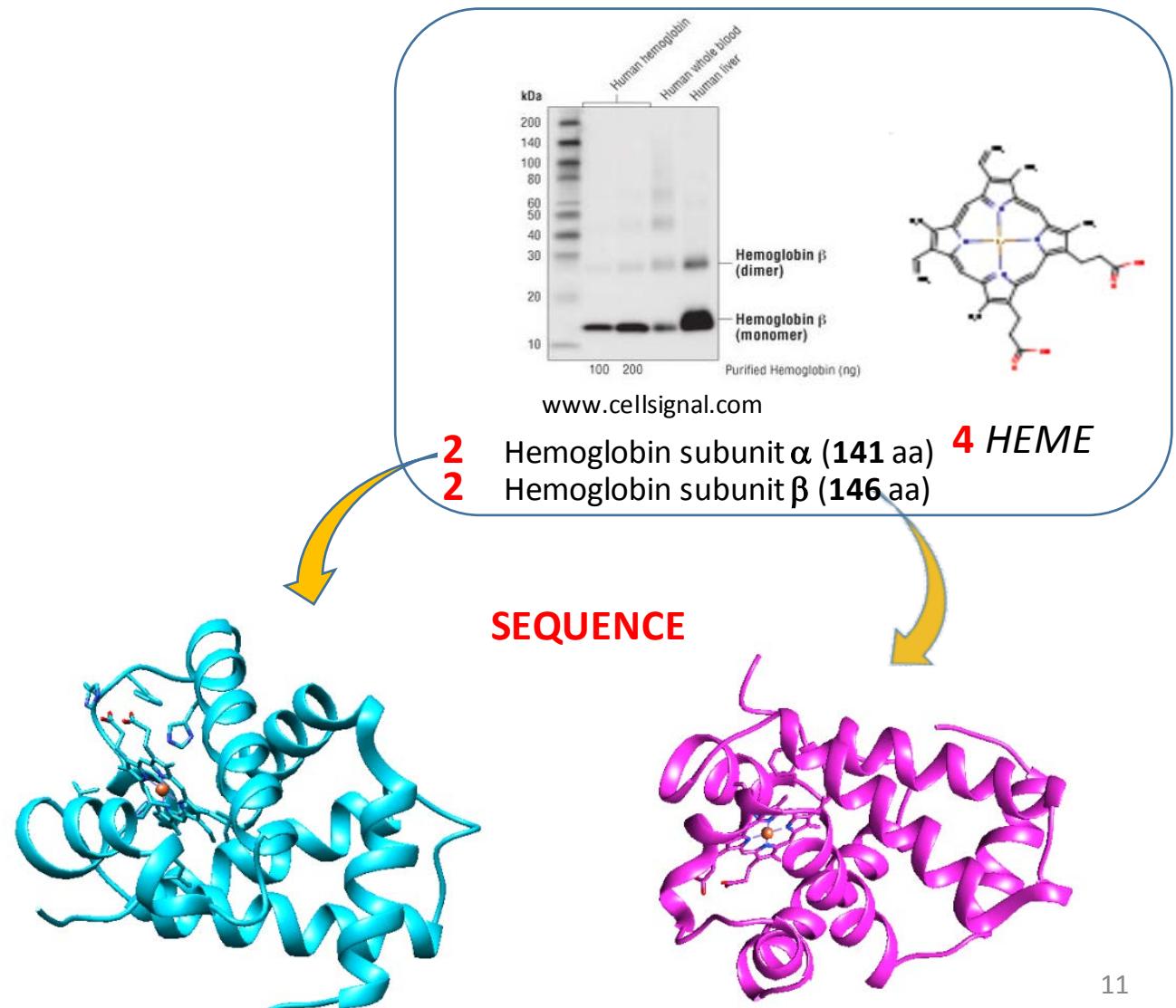
SEED



Step 1.- Getting the initial model(s): Structure prediction

de novo
modeling

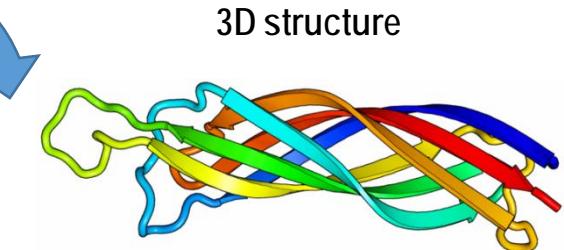
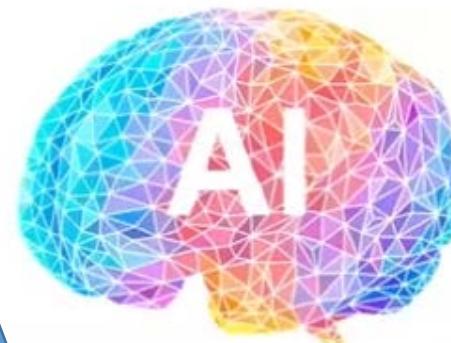
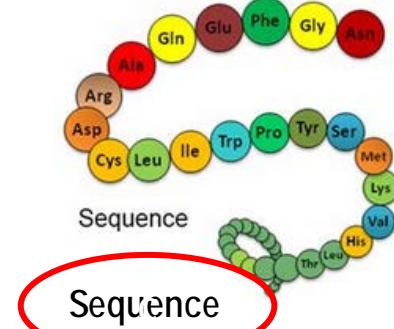
AlphaFold
ModelAngelo



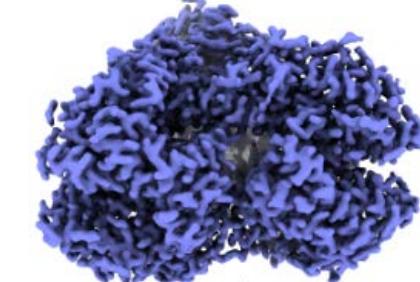
Step 1.- Getting the initial model(s): Structure prediction

de novo
modeling

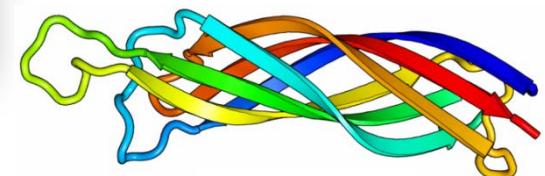
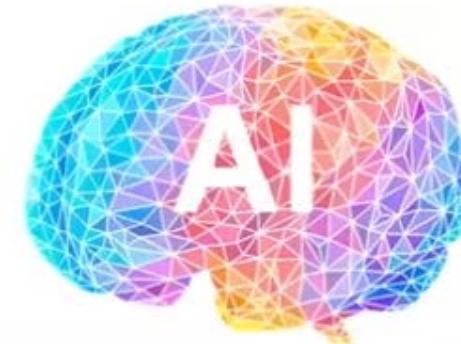
AlphaFold
ModelAngelo



ModelAngelo



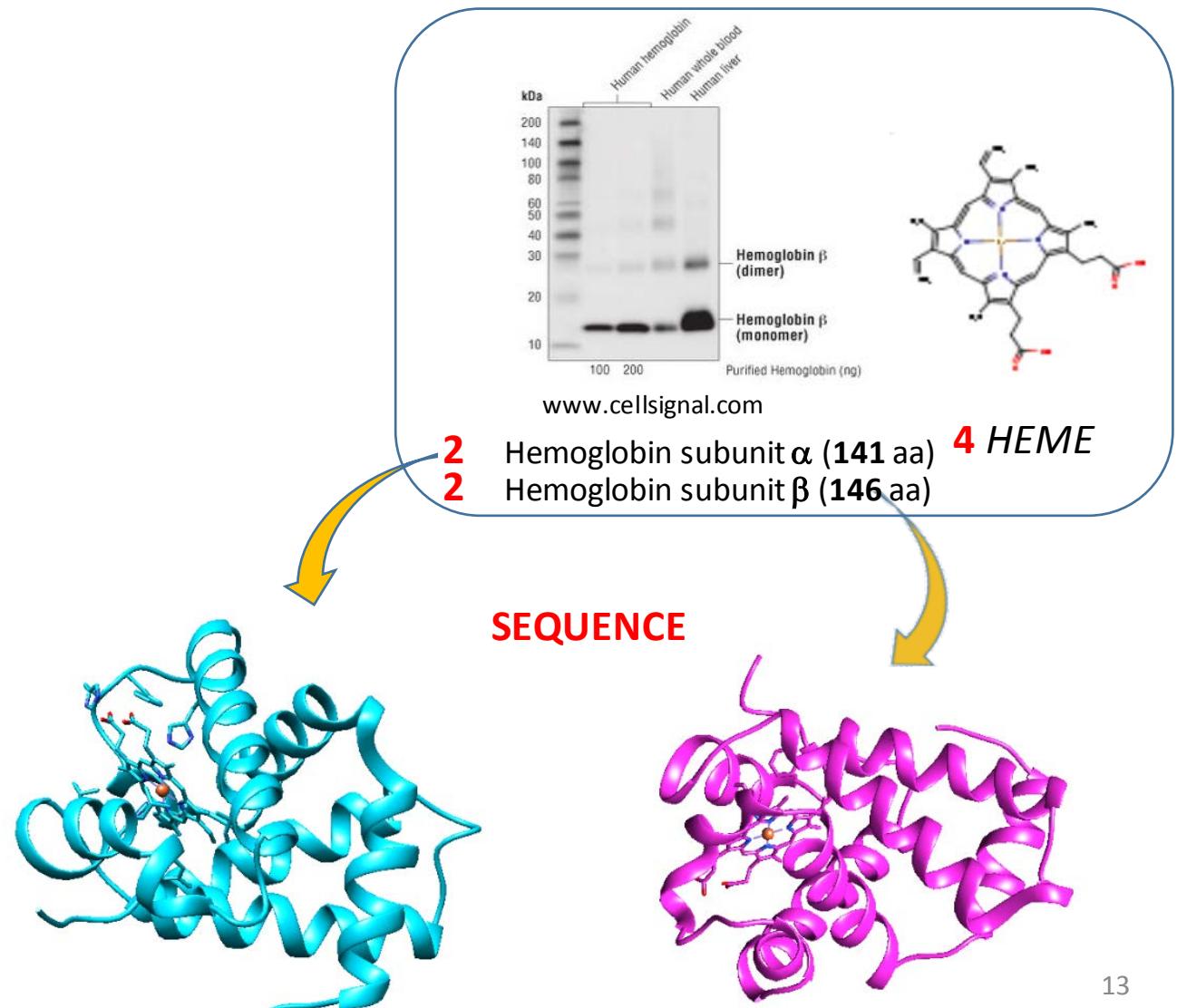
WORLDWIDE
PDB
PROTEIN DATA BANK



Step 1.- Getting the initial model(s): Structure prediction

de novo
modeling

AlphaFold
ModelAngelo



Step 1.- Getting the initial model(s): Structure prediction

de novo
modeling

AlphaFold
ModelAngelo

(template optional)

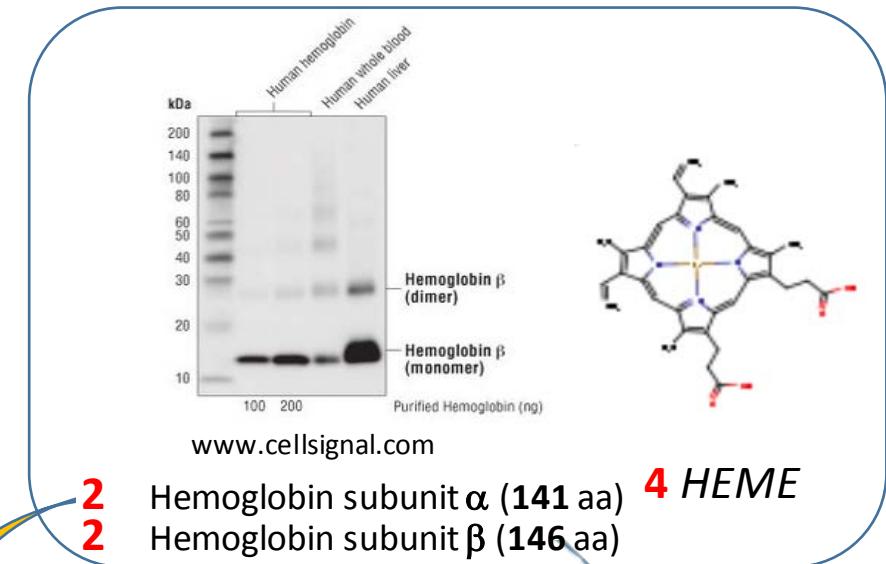
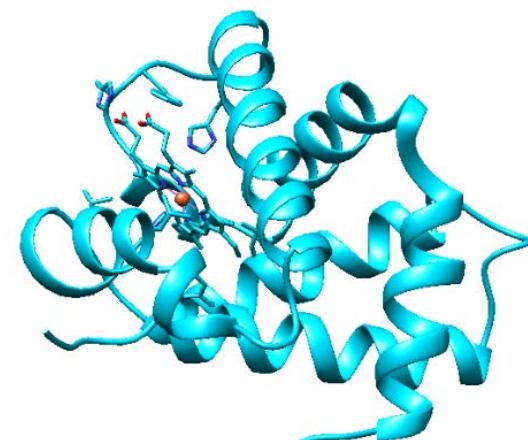
initial model ≠ real model

1st clue:
Skeleton of α carbons
and
many side chains

Homology
modeling

Our proteins borrow
the structure
of similar proteins

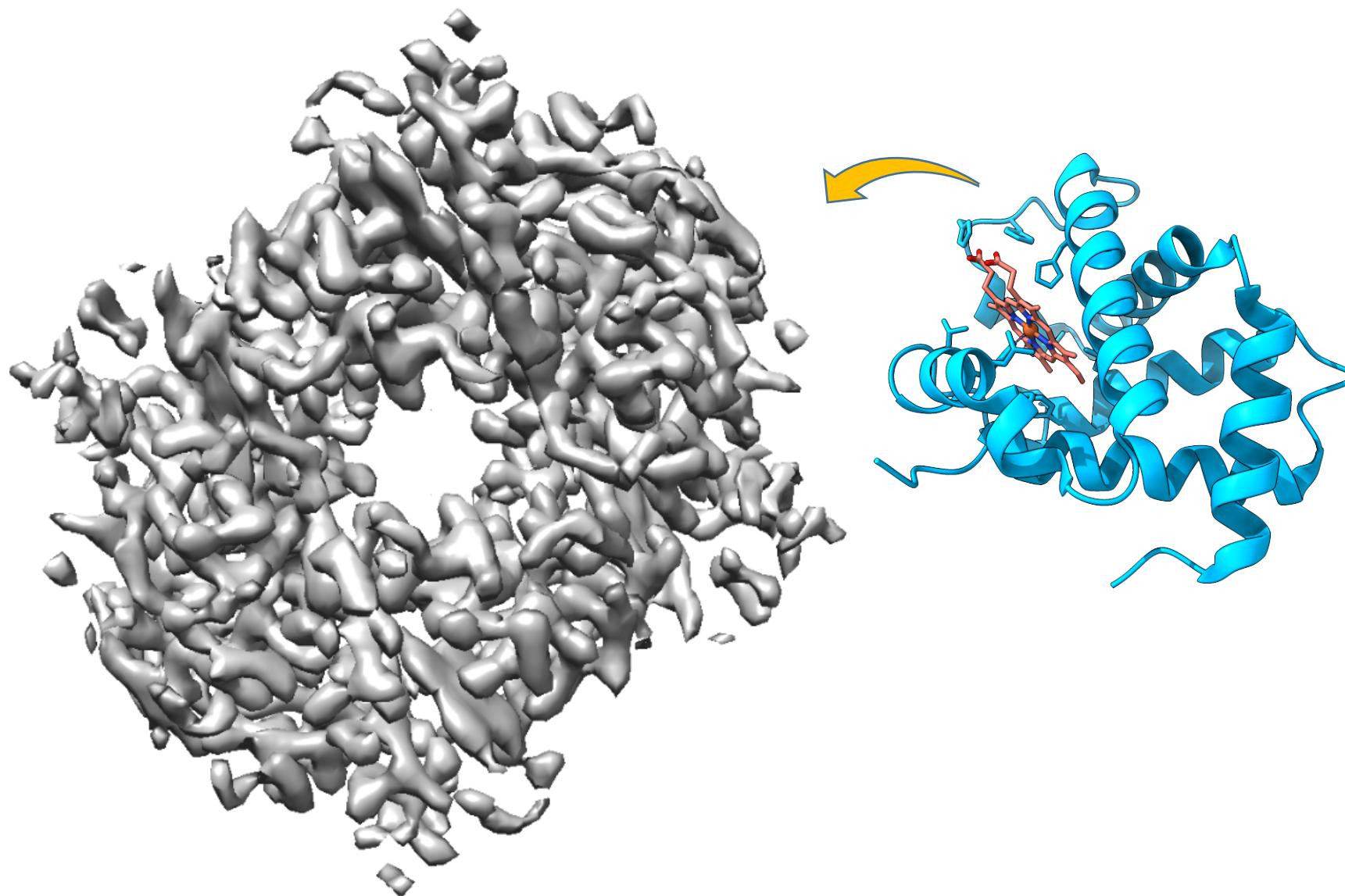
(template)



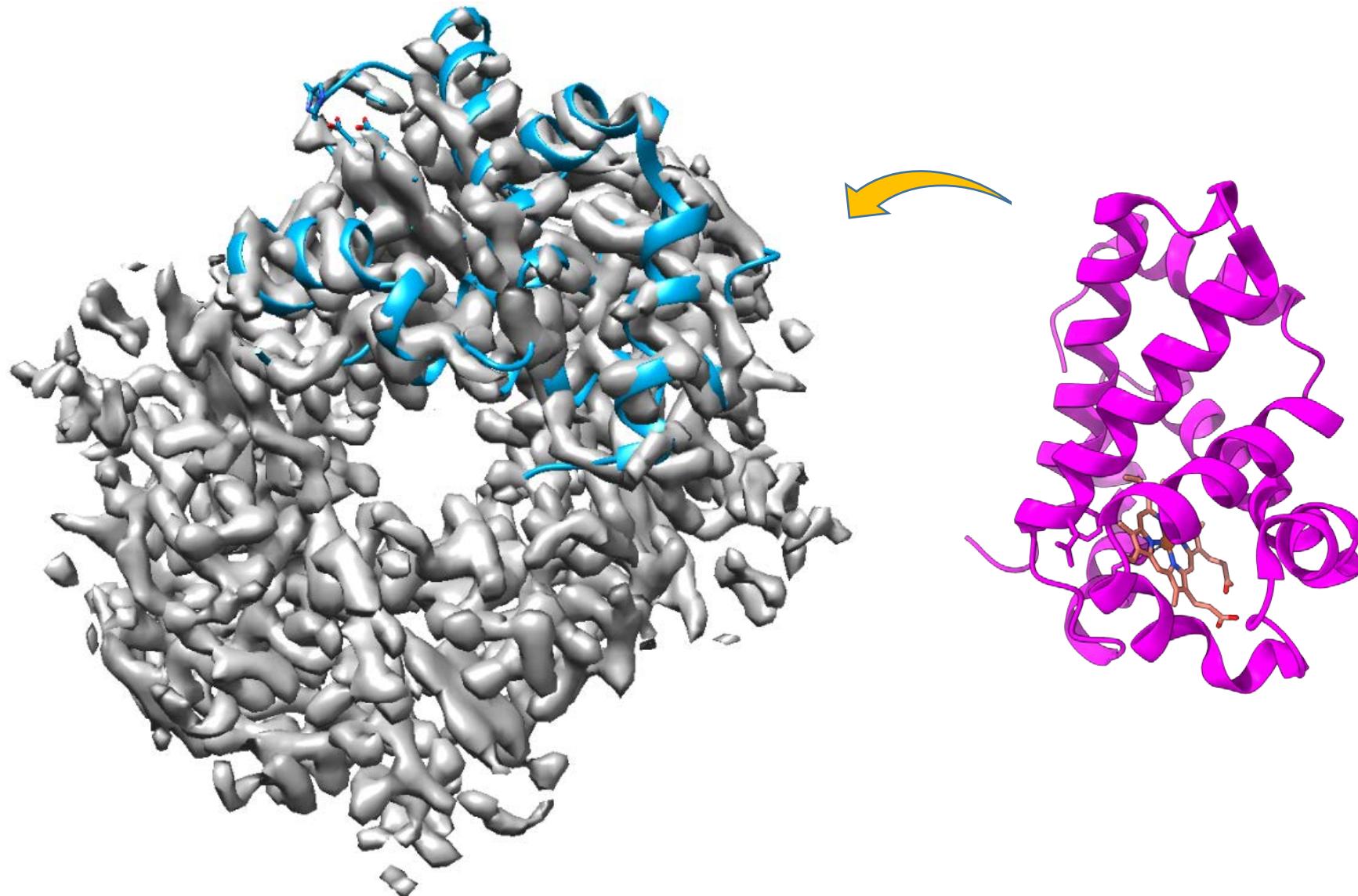
SEQUENCE



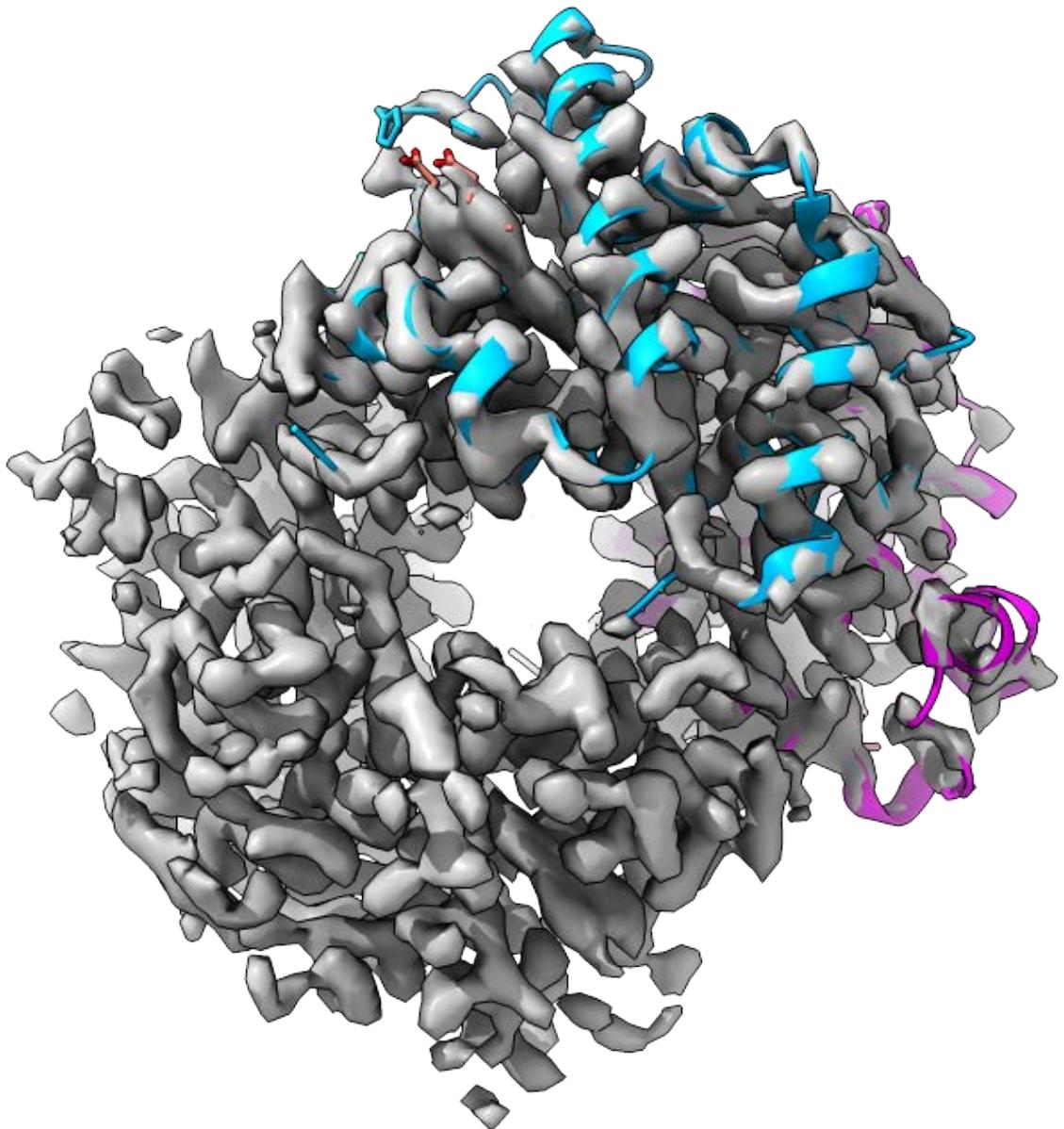
Step 2.- Rigid fitting of model(s) in map



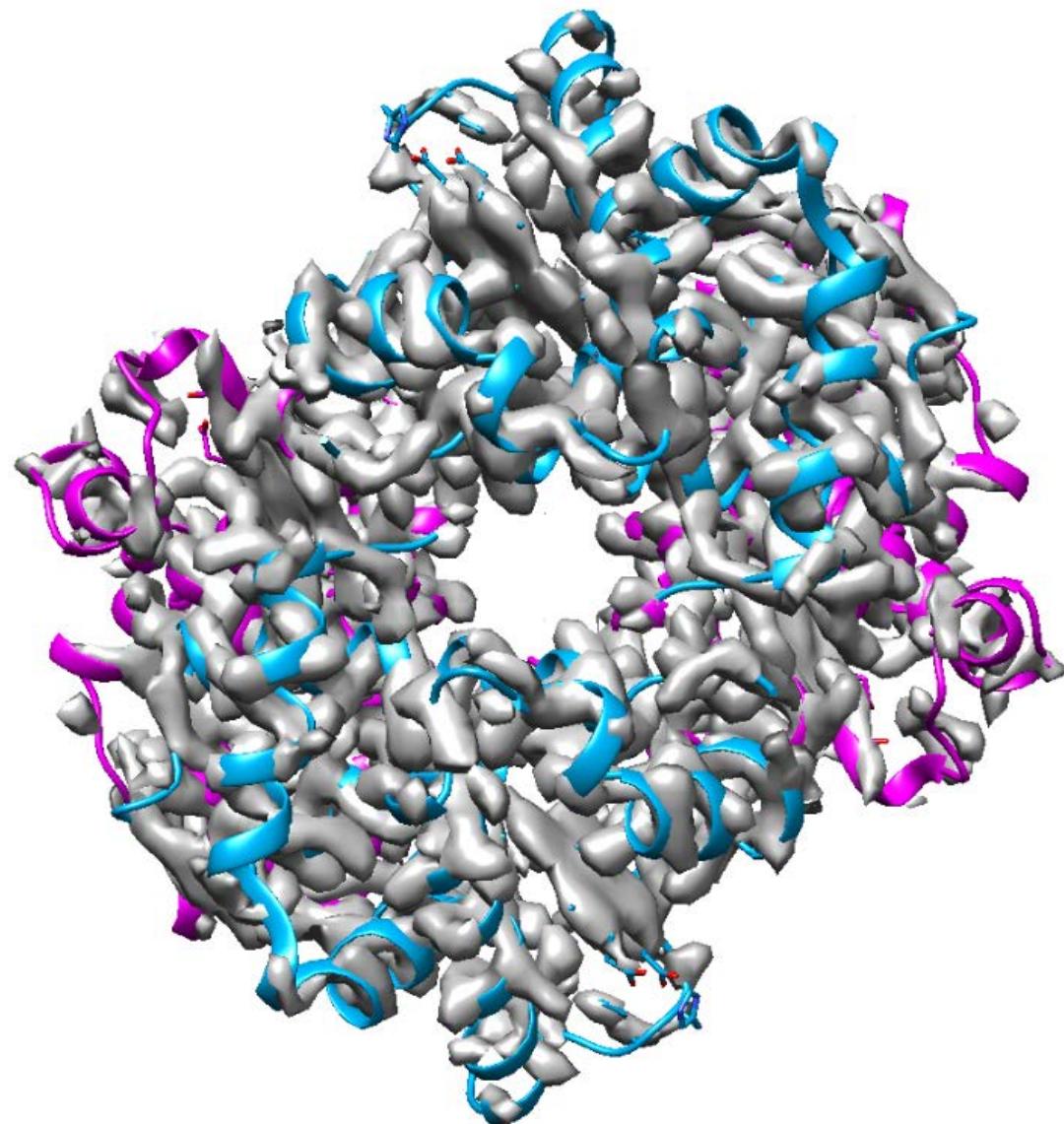
Step 2.- Rigid fitting of model(s) in map



Step 2.- Rigid fitting of model(s) in map

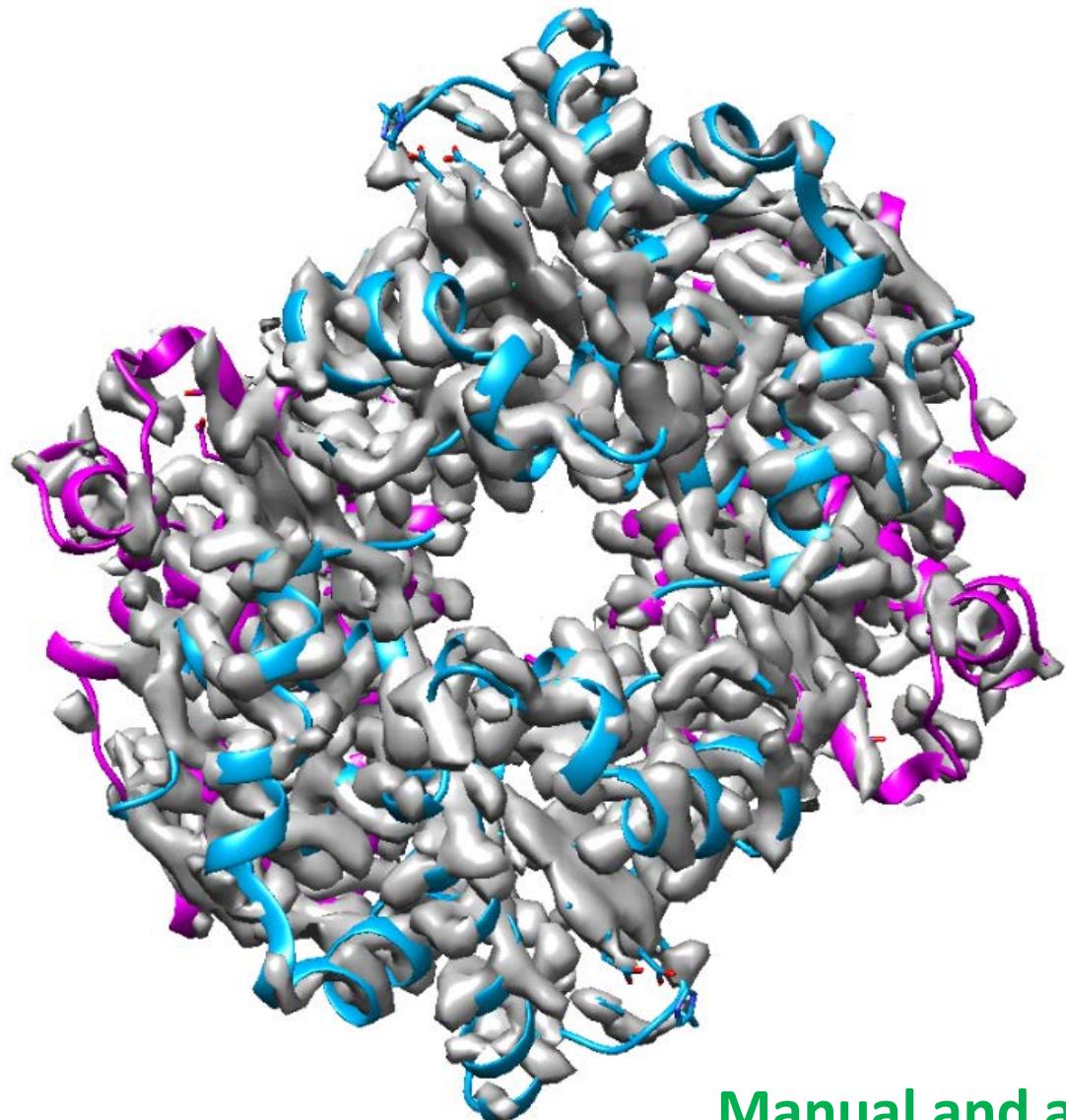


Step 2.- Rigid fitting of model(s) in map



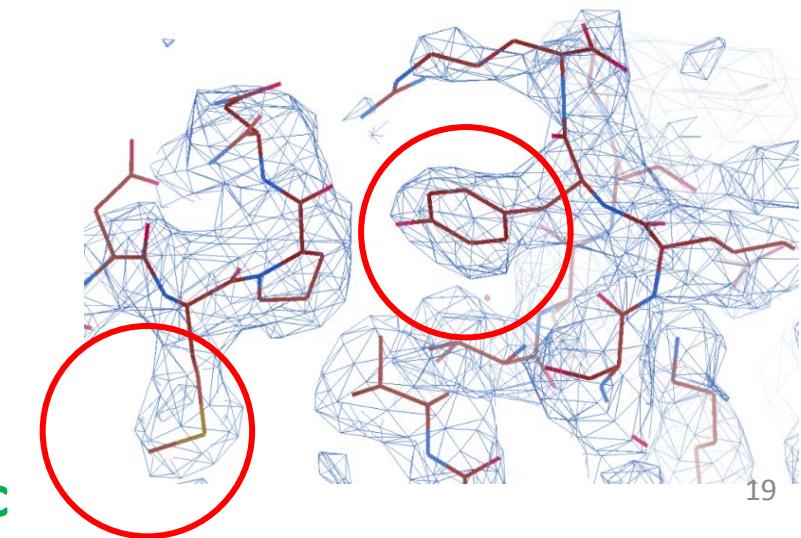
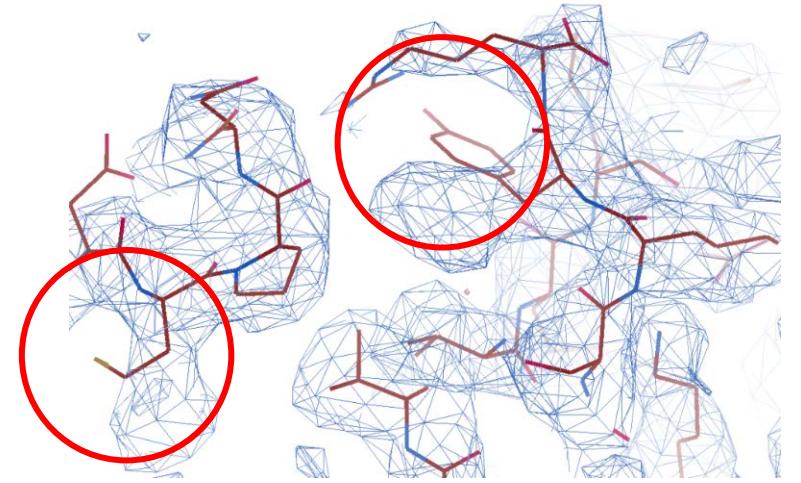
Strategy

Step 3.- Flexible fitting (refinement) of model(s) in map



Manual and automatic

Getting the final model(s)



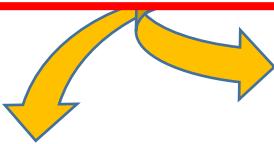
Step 4.- Model(s) validation

Is this structure reliable?



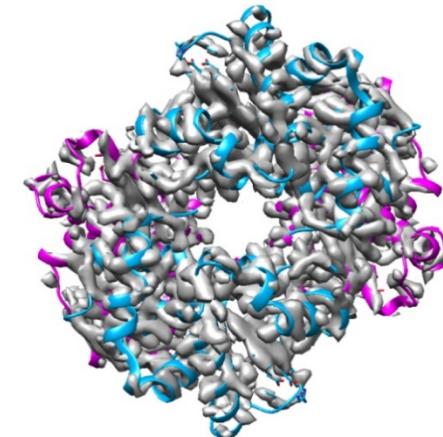
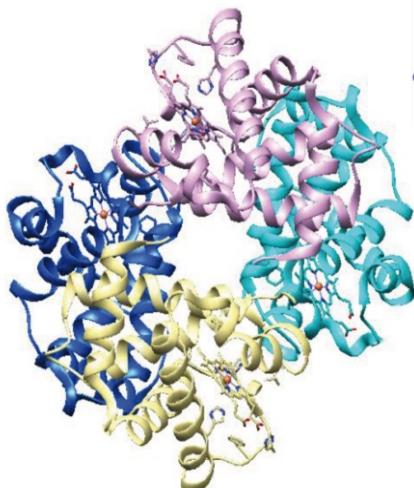
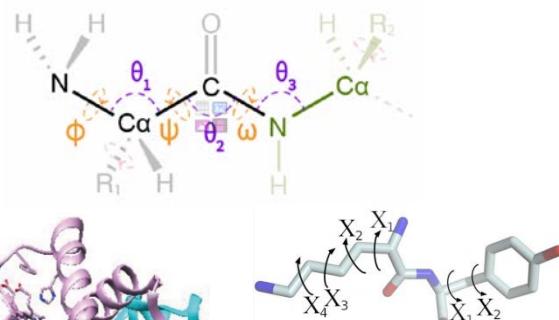
Is this structure similar
to the structures of the database?
(Bond lengths, angles, contacts ...)

Q (Submission)

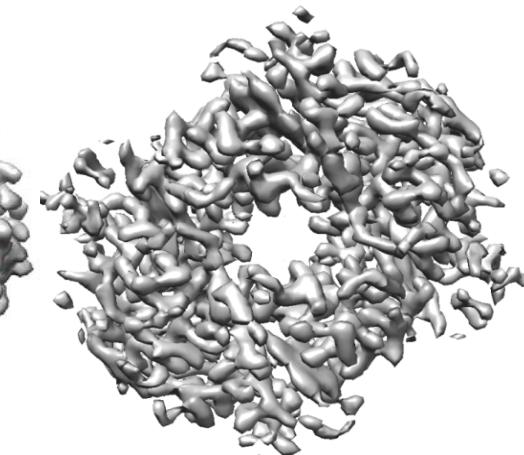
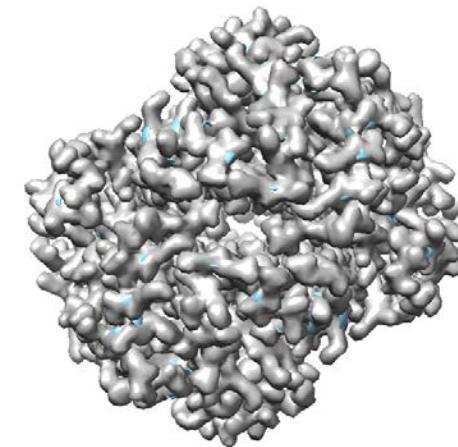


Map and model similarity

Geometry

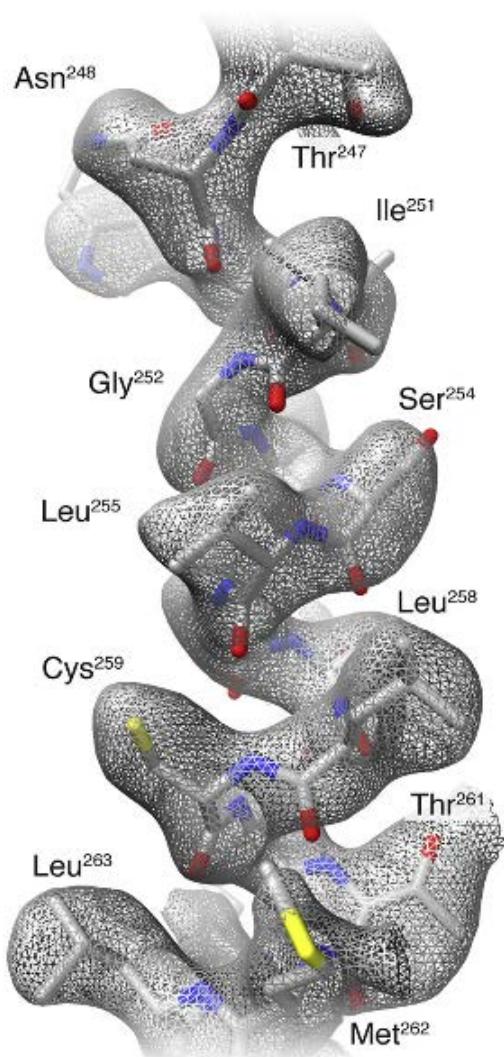


Map/model
fitting



Is the map derived from the model similar
to the original map?

Modeling proteins from cryo-EM data



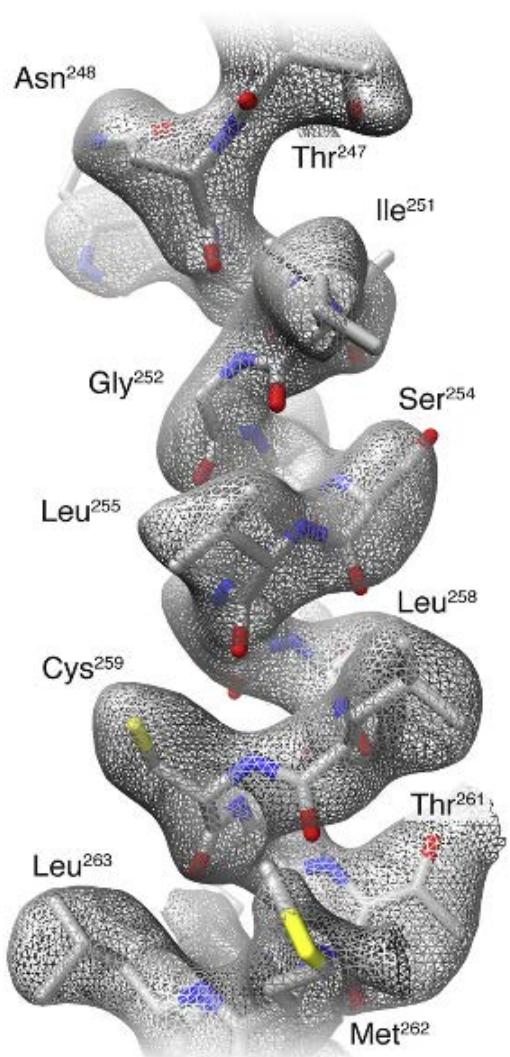
Model building

Basic steps

- 1.- Get the initial model (s)**
- 2.- Rigid fitting of model(s) in map**
- 3.- Flexible fitting (refinement) of model(s) in map**
- 4.- Model(s) validation**

Tools

Modeling proteins from cryo-EM data



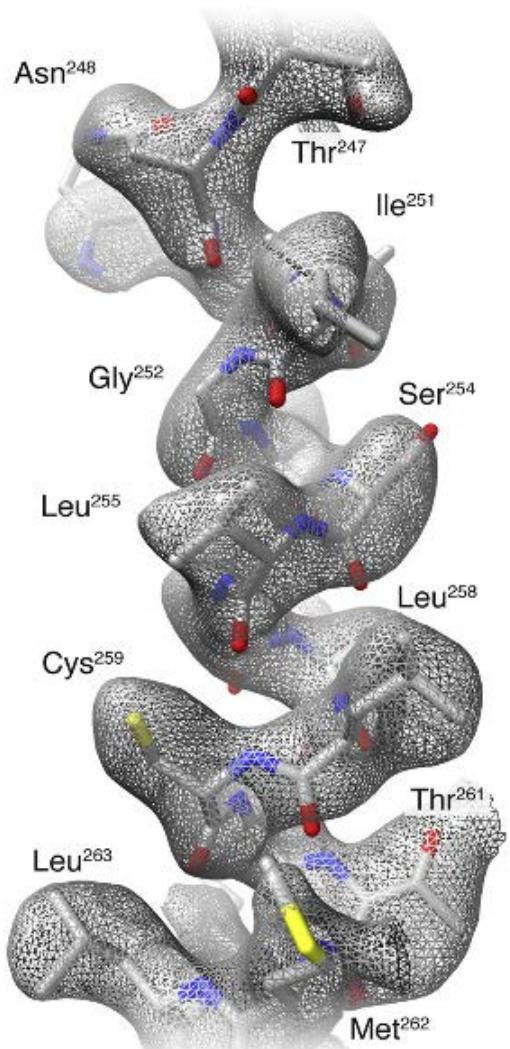
Model building

Basic steps

- 1.- Get the initial model (s)
- 2.- Rigid fitting of model(s) in map
- 3.- Flexible fitting (refinement) of model(s) in map
- 4.- Model(s) validation

Tools

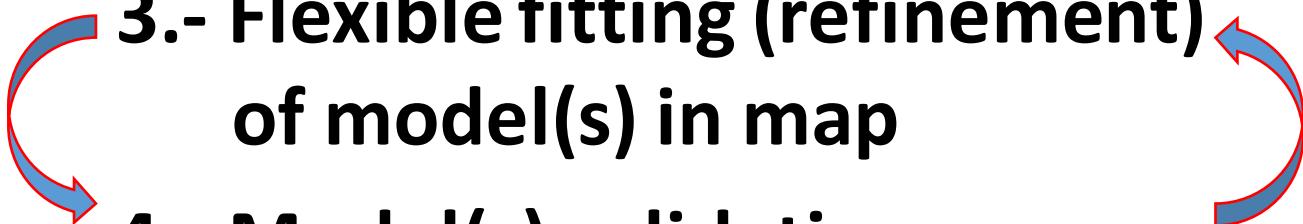
Modeling proteins from cryo-EM data



Model building

Basic steps

- 0.- Map preprocessing**
- 1.- Get the initial model (s)**
- 2.- Rigid fitting of model(s) in map**
- 3.- Flexible fitting (refinement)
of model(s) in map**
- 4.- Model(s) validation**



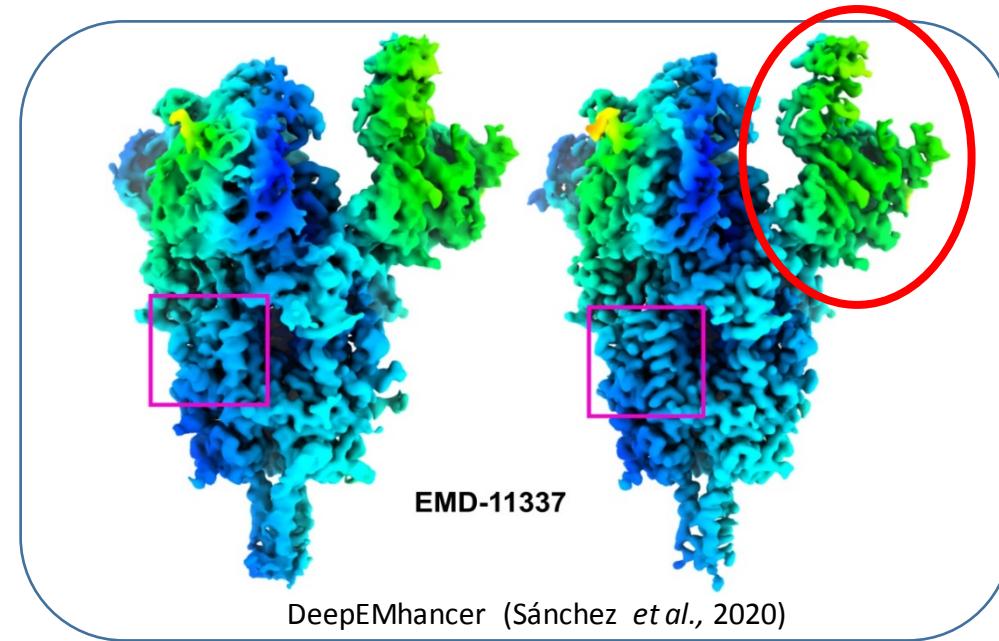
Tools

Step 0.- Map preprocessing

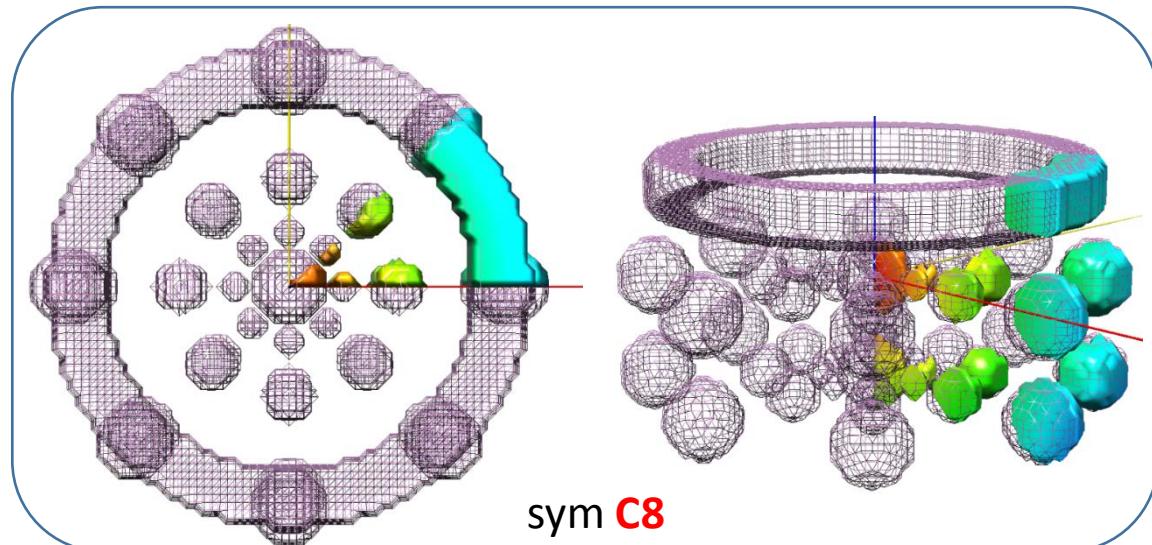
Map optimization
(sharpening, improving
density)

Global and/or locally

Extract asymmetric unit
(geometrical)



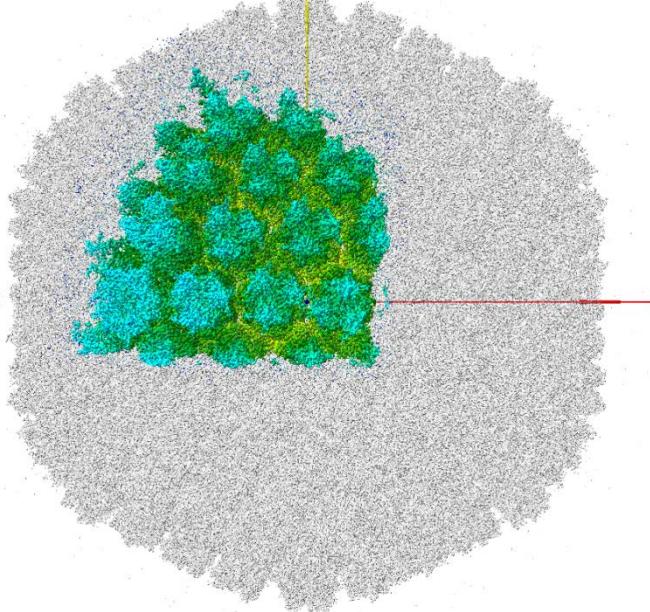
EMD-11337
DeepEMhancer (Sánchez *et al.*, 2020)



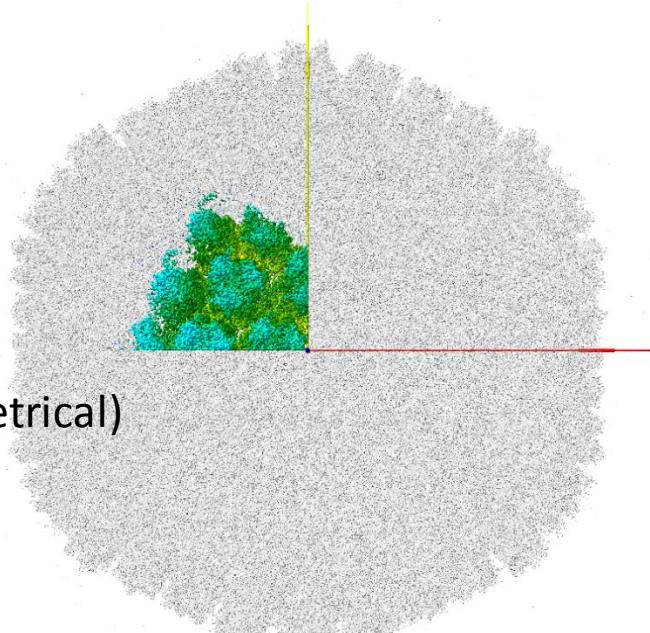
sym C8

Asymmetric unit

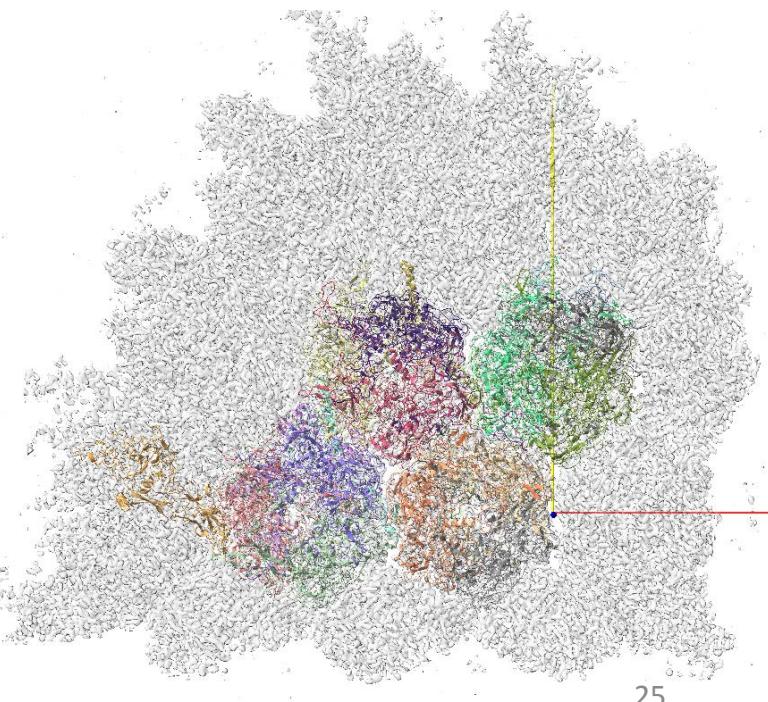
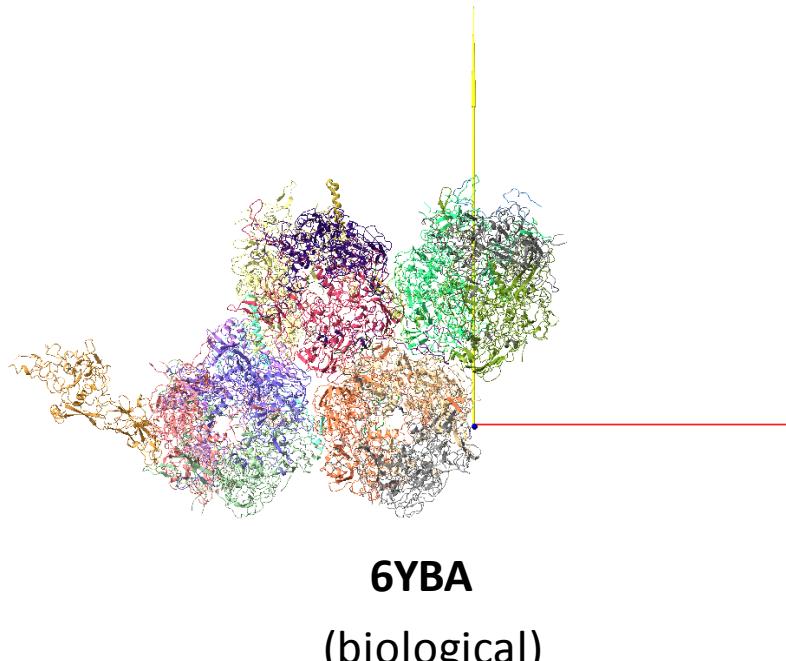
(geometrical)



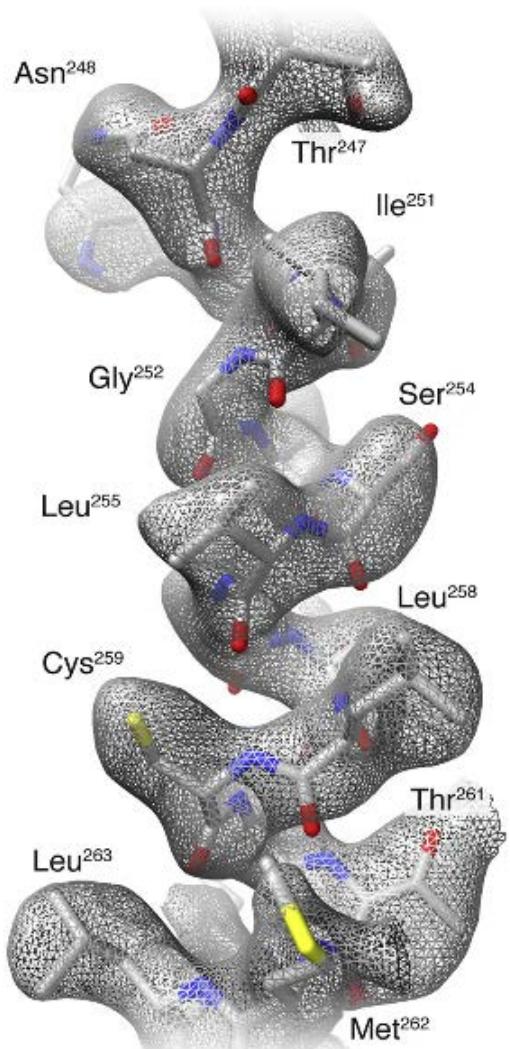
EMD-10768



6YBA
(biological)

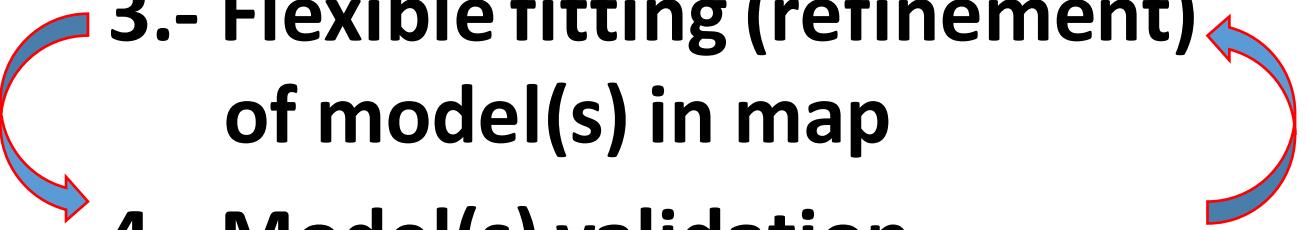


Modeling proteins from cryo-EM data



Model building

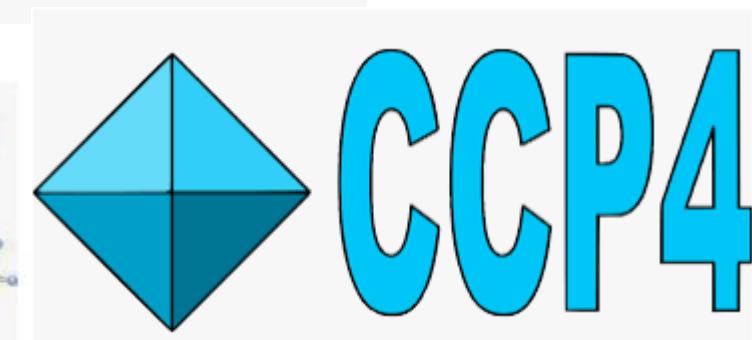
Basic steps

- 0.- Map preprocessing**
 - 1.- Get the initial model (s)**
 - 2.- Rigid fitting of model(s) in map**
 - 3.- Flexible fitting (refinement) of model(s) in map**
 - 4.- Model(s) validation**
- 

Tools

Tools

- * Borrowed from crystallography
- * Specifically designed for cryo-EM data



Programs & Suites

Interactive & Automatic

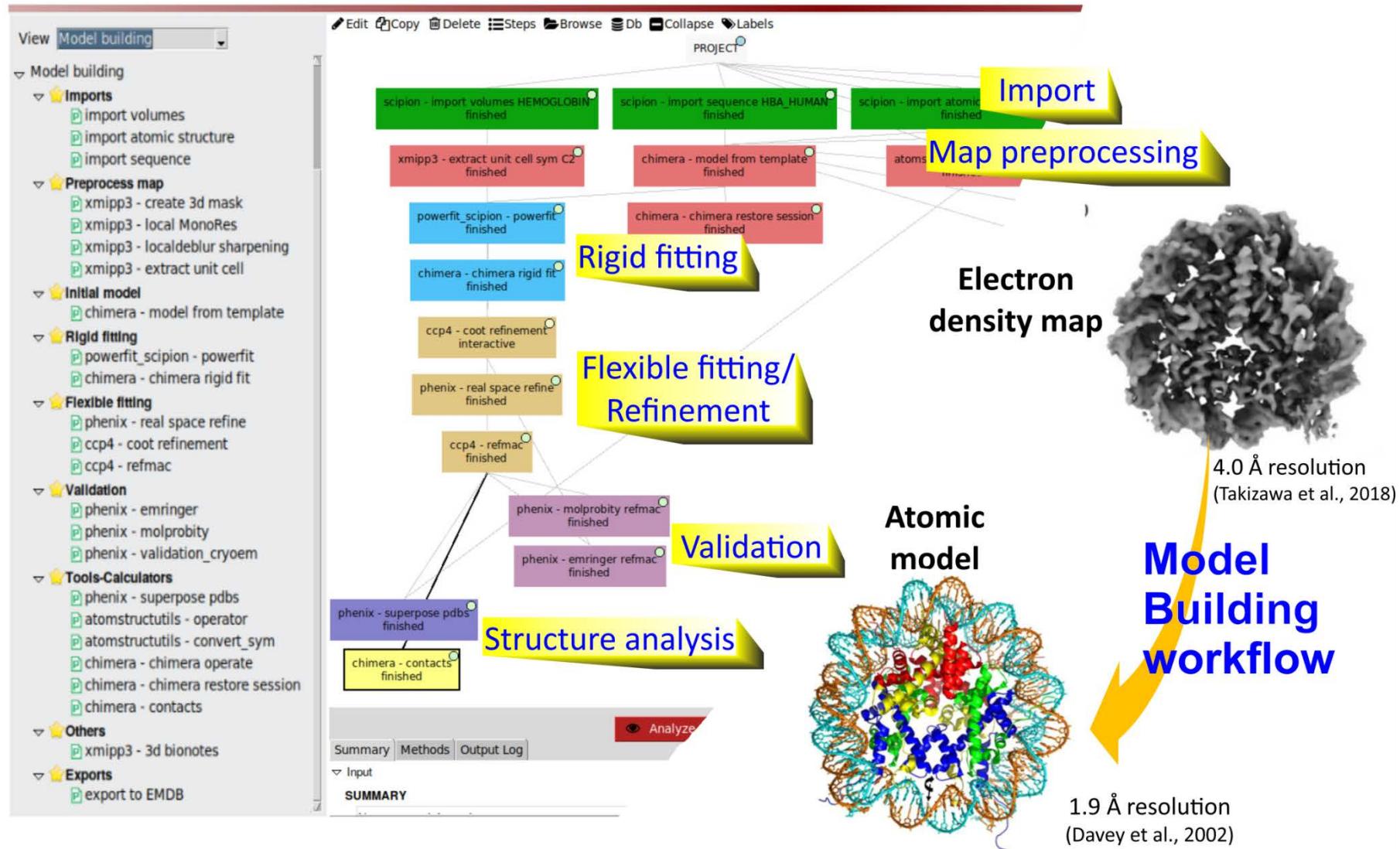


Project Help

 SCIPION

v2.0 (2019-04-08) Diocletian (release-2.0.0-fixes 03cc8683c4)

Demo Mc





Tutorials

- Single particle tutorials
- Tomography tutorials
- Model Building**
 - Model Building by Homology
 - Build initial model with AlphaFold
 - Build initial model with ModelAngelo

scipion-em.github.io/docs/docs/user/tutorials/modelBuilding/Homology/index.html

Aplicaciones Outlook.com - ruth... Facebook Correo :: Entrada Home - PubMed ... Programa de Forma... Búsqueda de conv... Cerveza Coron.

Graphical interface manuals
Image processing resources

Docs » Graphical interface manuals » Model Building Tutorial [View page source](#)

Model Building Tutorial

This tutorial shows a workflow aimed at obtaining high-quality atomic models from cryo-EM data by using scipion software framework.

Contents

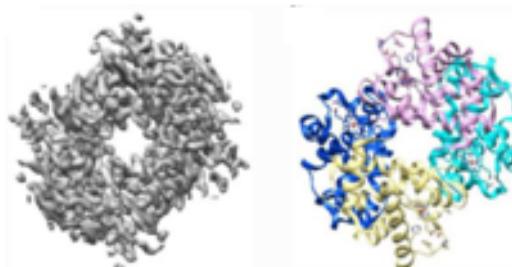
- 1. Revision History
- 2. Prelude
- 3. Introduction to Model building
- 4. Problem to solve: Haemoglobin
- 5. Input data description
- 6. Import Input data
- 7. 3D Map preprocessing
- 8. Structure Prediction by Sequence Homology. Searching for Homologues
- 9. Moving from sequence to atomic structure scenario
- 10. Merging 3D Maps and Atomic Structures: Rigid Fitting
- 11. Refinement: Flexible fitting
- 12. Structure validation and comparison
- 13. Building the asymmetric unit
- 14. The whole macromolecule
- 15. Summary of results and submission
- 16. A Note on Software Installation
- 17. How to solve some problems that you can find during the execution of the modeling workflow
- 18. TODO
- 19. References

Appendices



MODEL BUILDING IN SCIPION

0.- Map preprocessing



1.- Get the initial model (s)

**HOMOLOGY
MODELING**

***de novo*
MODELING**

**ALPHAFOLD
MODELANGELO**

2.- Rigid fitting of model(s) in map

3.- Flexible fitting (refinement)
of model(s) in map

4.- Model(s) validation