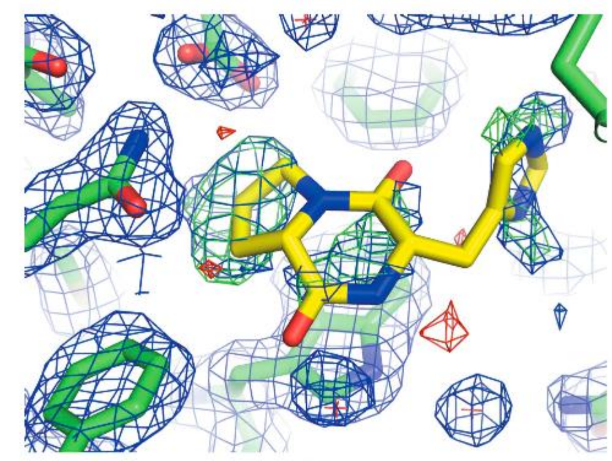
Structure Validation



Martyn Winn & Andrea Thorn

Weds 29th Sept 2021

Funded by
DFG
Bundesministerium für Bildung und Forschung
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG
Coronavirus Structural Task Force
www.insidecorona.net

Science and Technology Facilities Council
Scientific Computing
Research Complex at Harwell
CCP4
CCP-EM
CCPBioSim

Acknowledgements

- CCP-EM team – www.ccpem.ac.uk
- Coronavirus Structural Task Force - https://insidecorona.net/
- CCP4 team – www.ccp4.ac.uk
- PDBe / EMDB teams
- UKVALID consortium

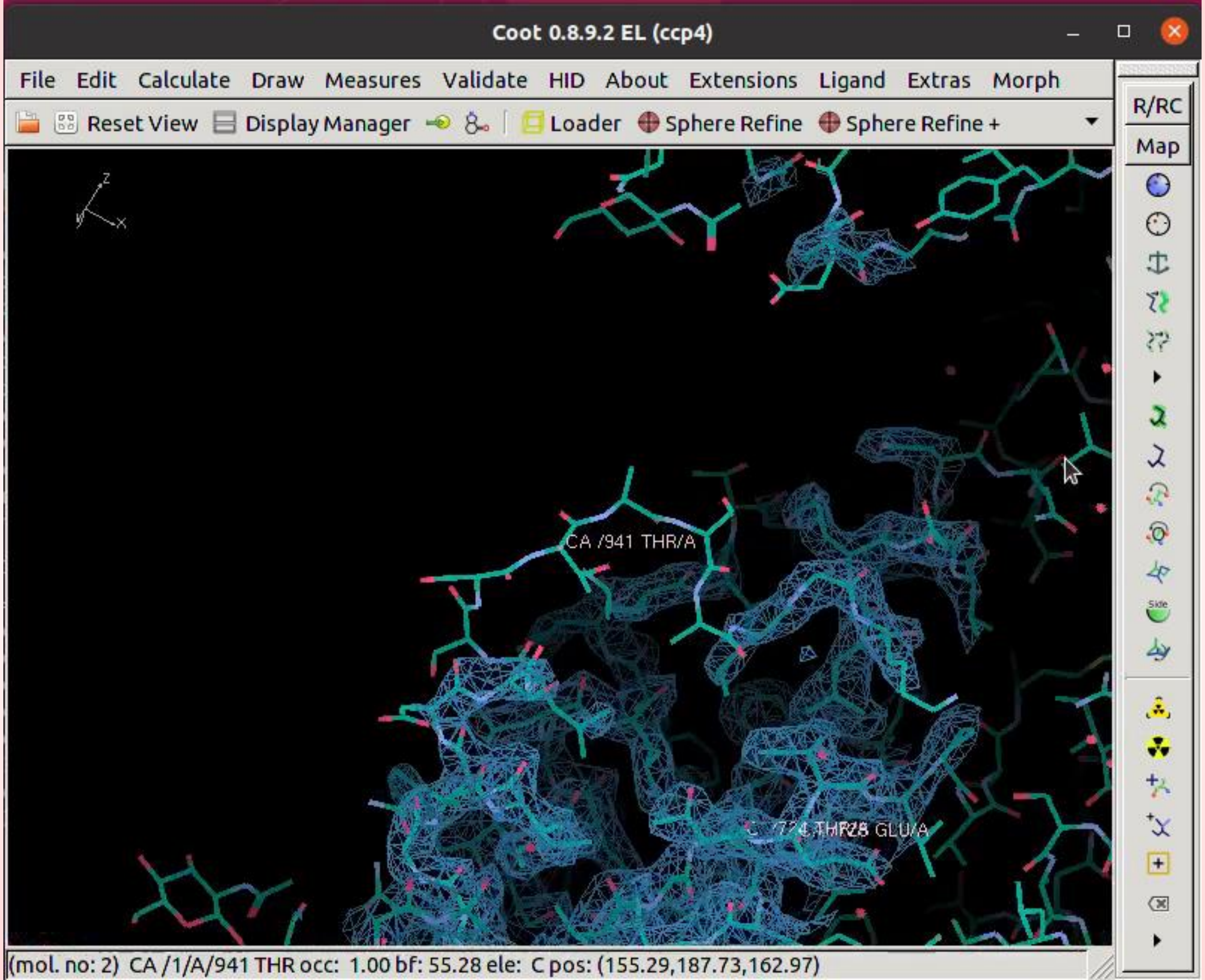
… plus many others in the structural biology community.

Experimental Structural Biology

Methods

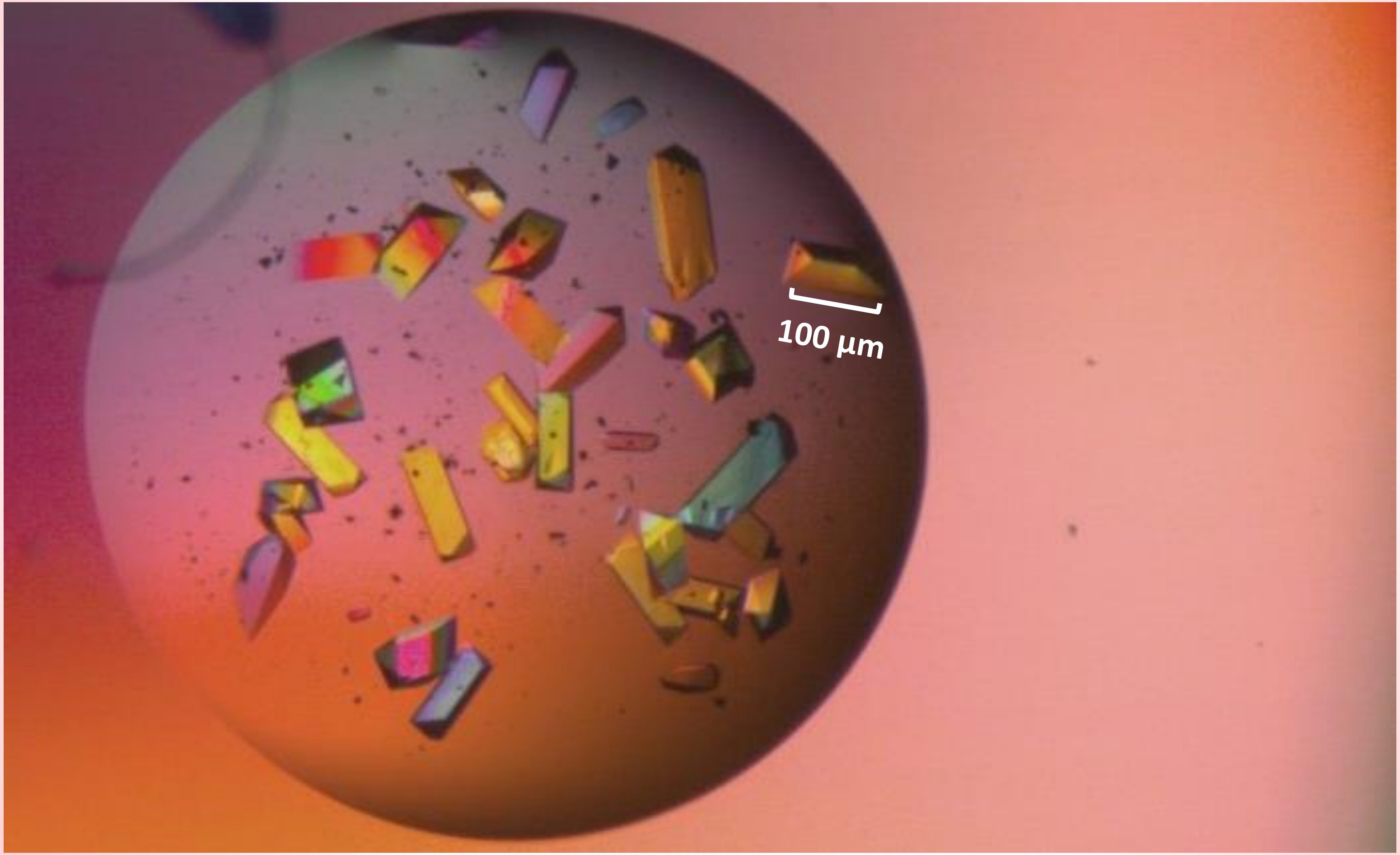
- NMR
- Crystallography
- Cryo-EM

In all of these methods, a model is fitted onto measured data in order to explain them.



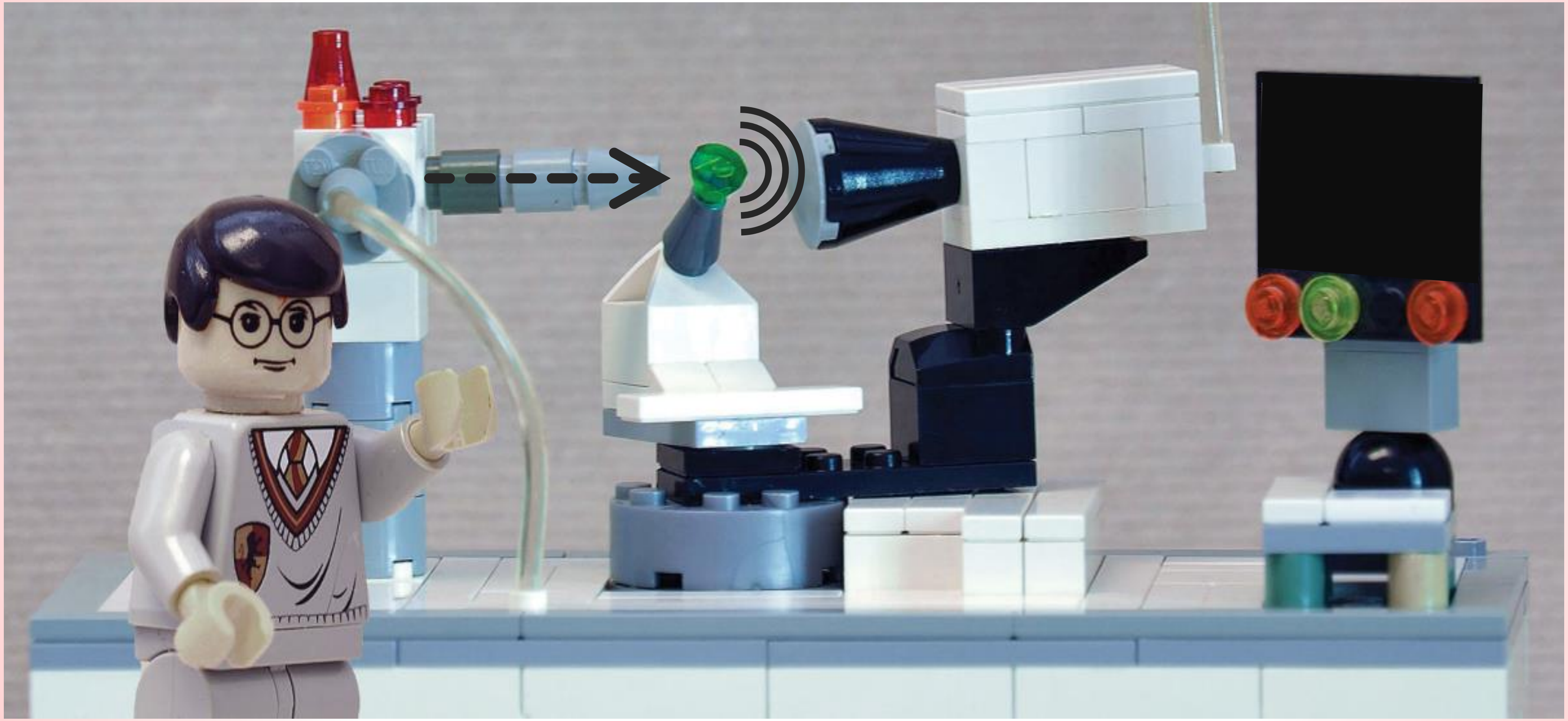
```
(mol. no: 2) CA /1/A/941 THR occ:  1.00 bf: 55.28 ele:  C pos: (155.29,187.73,162.97)
```

Crystallography: Specimen

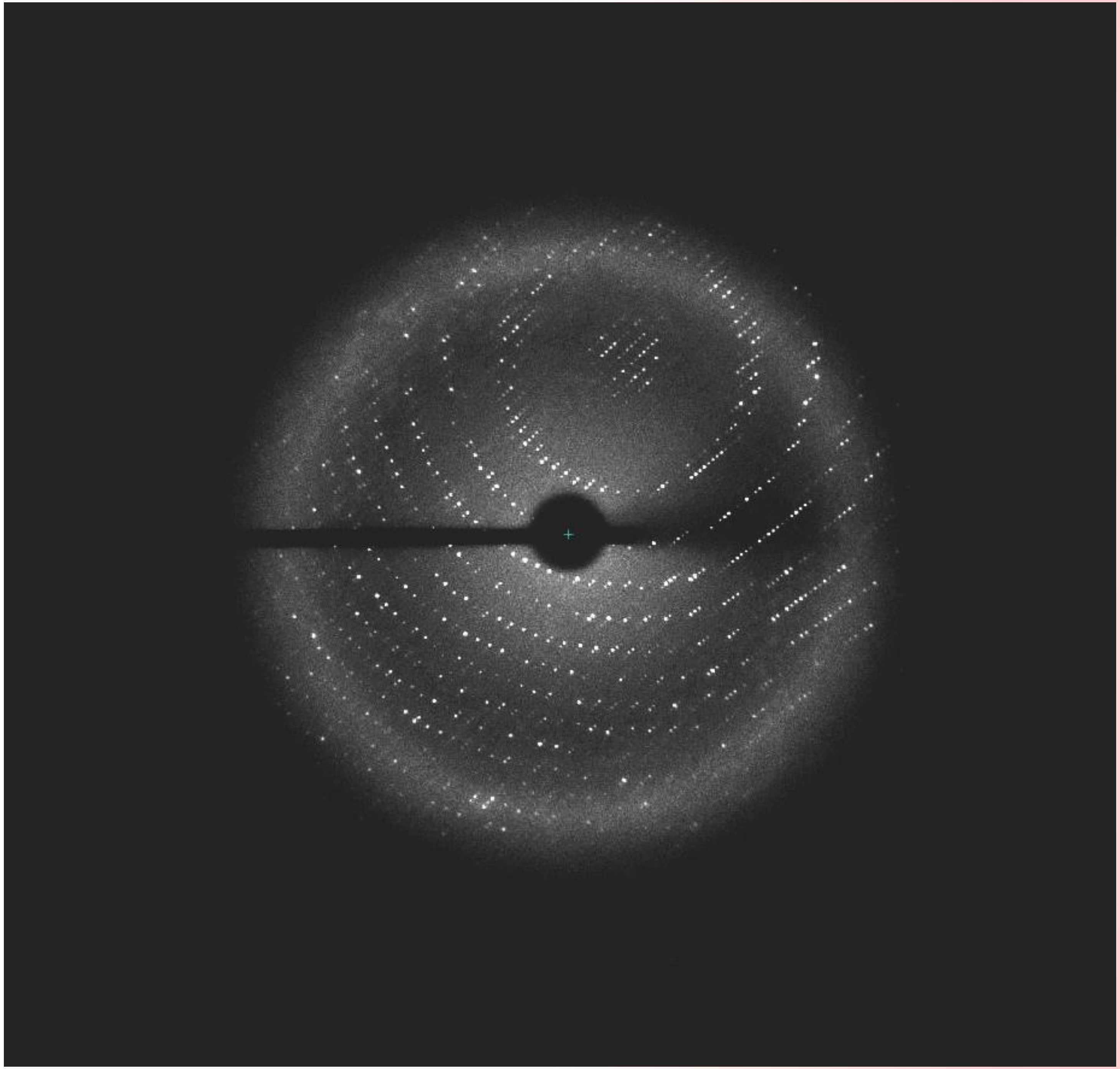


Picture: Imperial College London

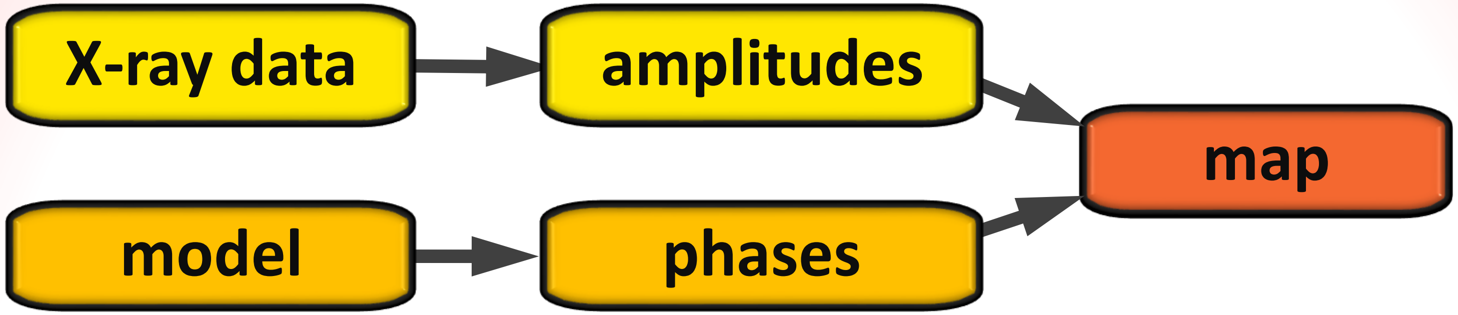
Crystallography: Experiment



Diffraction pattern on detector. Amplitude is recorded in spots – phase is lost!



Phase problem in crystallography



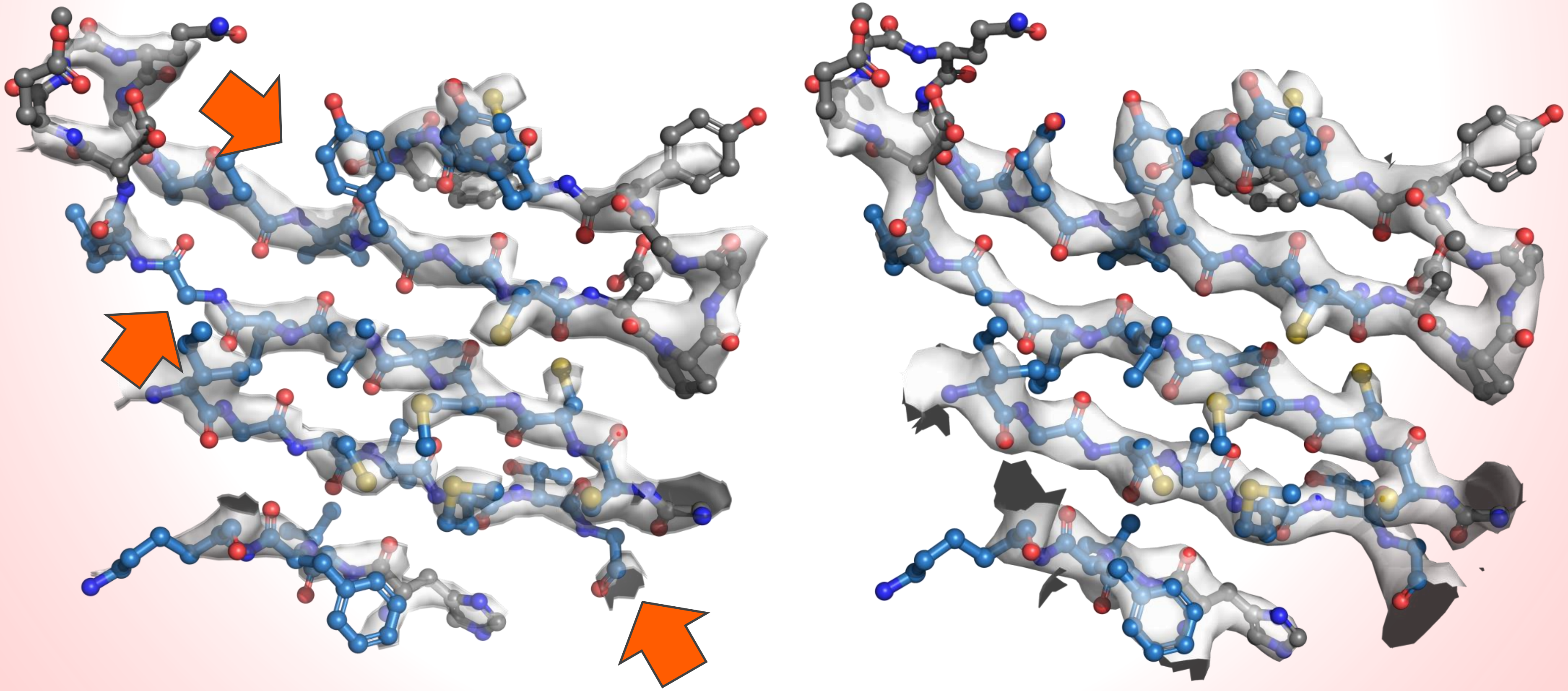
X-ray data

amplitudes

map

model

phases



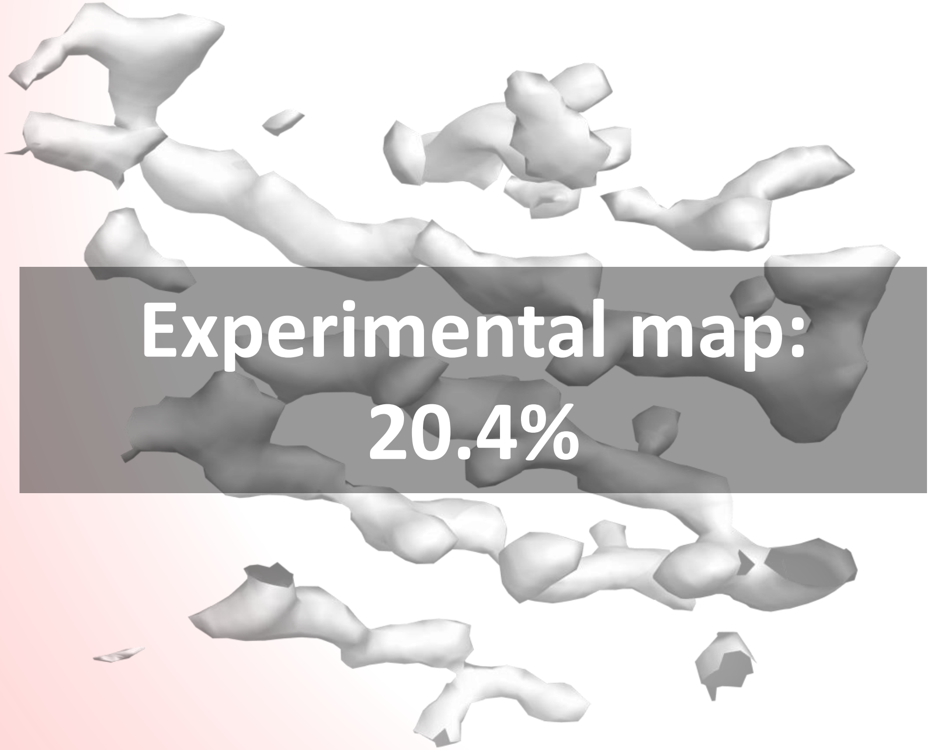
Does my structure fit the data? .

NOT REALLY

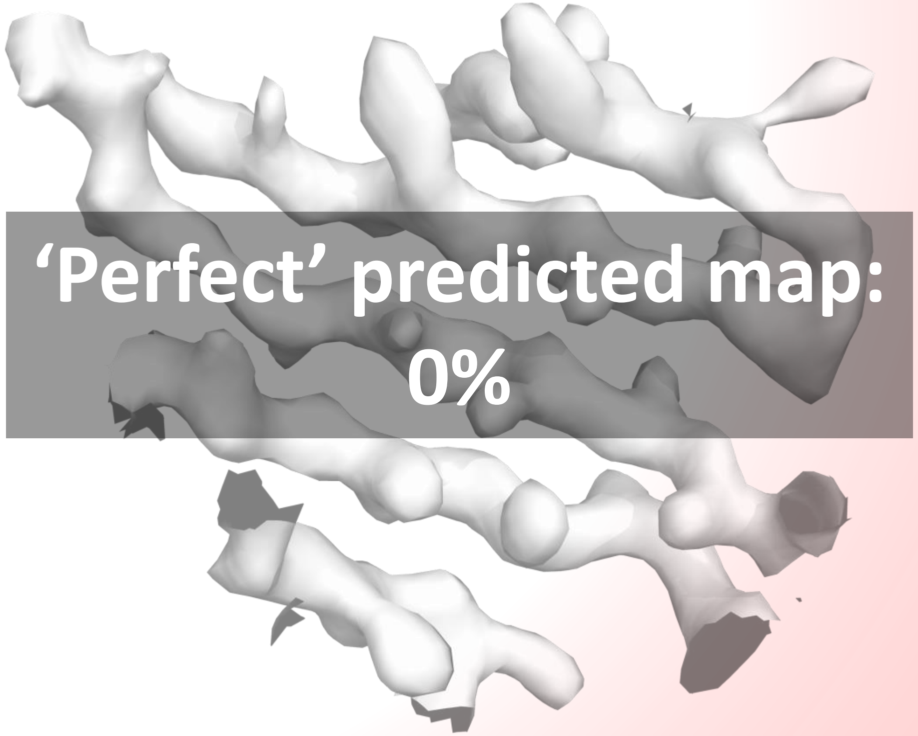
$$R = \frac{100\cdot\left(\sum\left|\underbrace{|F_{obs}|}_{\text{DATA}}-\underbrace{|F_{calc}|}_{\text{MODEL}}\right|\right)}{\sum|F_{obs}|}$$

Typical R-values:

Small molecules	**5%**
Macromolecules	**20%**

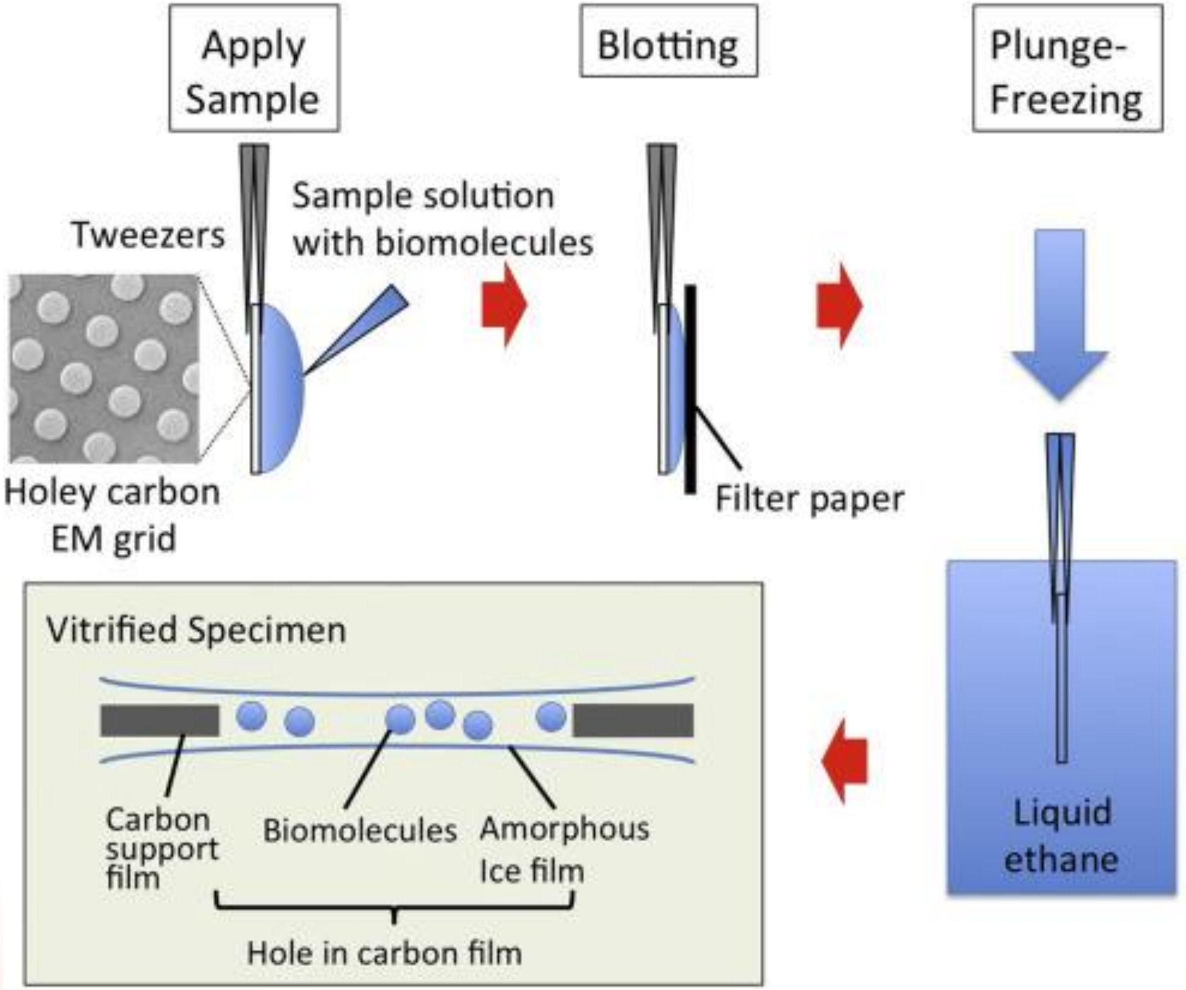


Experimental map:
20.4%

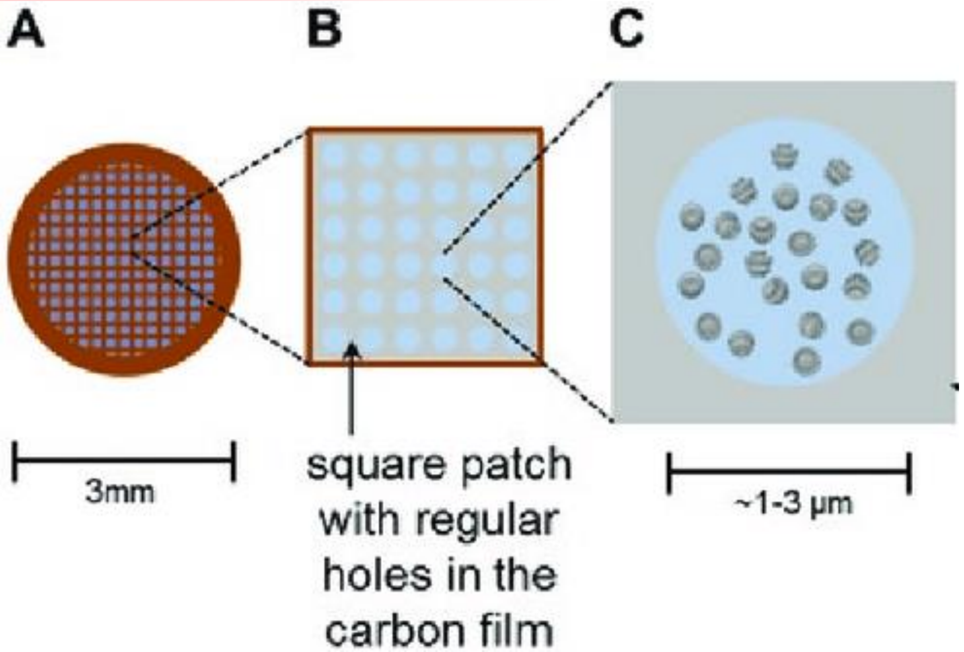


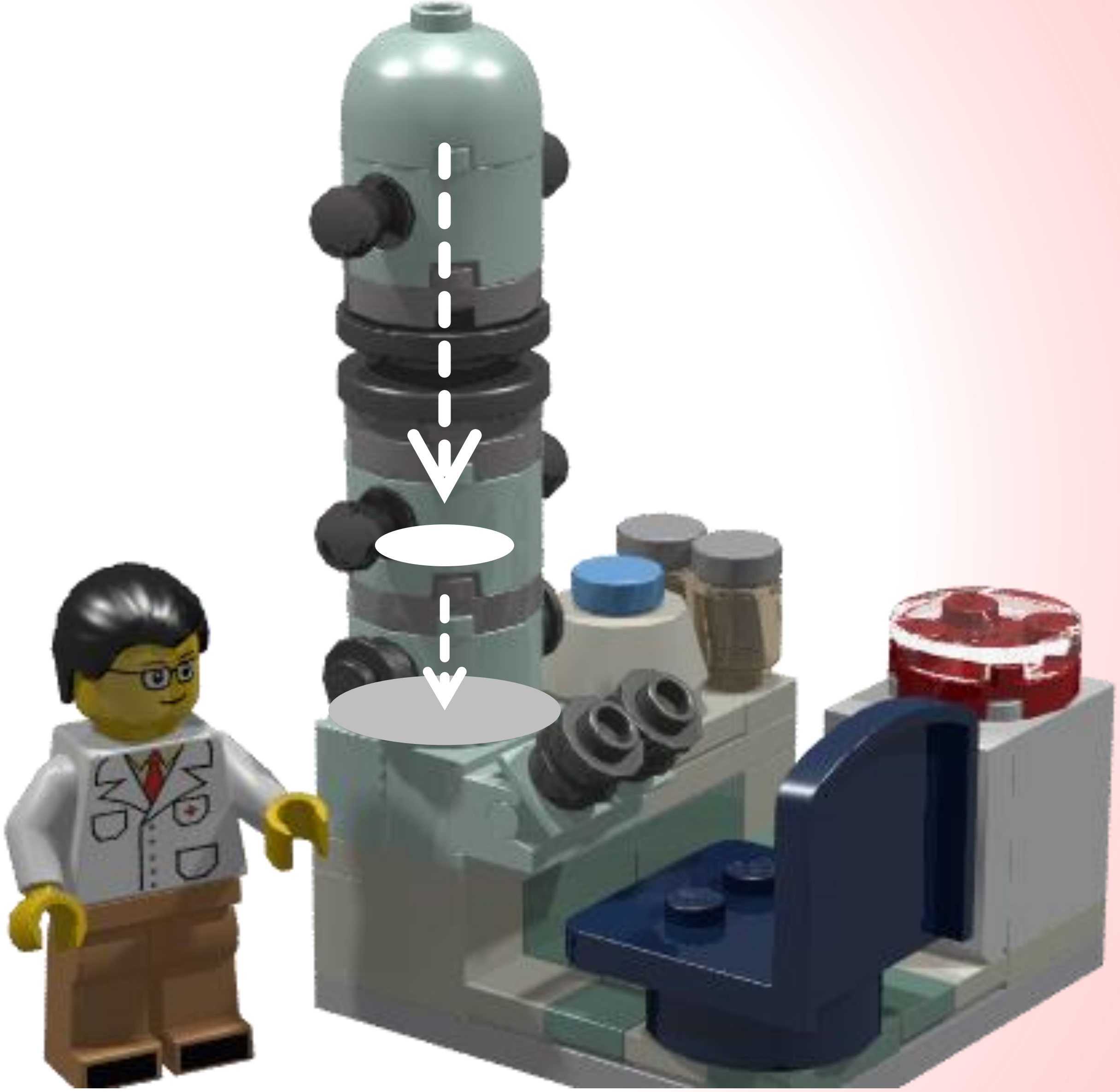
'Perfect' predicted map:
0%

Electron Cryo Microscopy: specimen



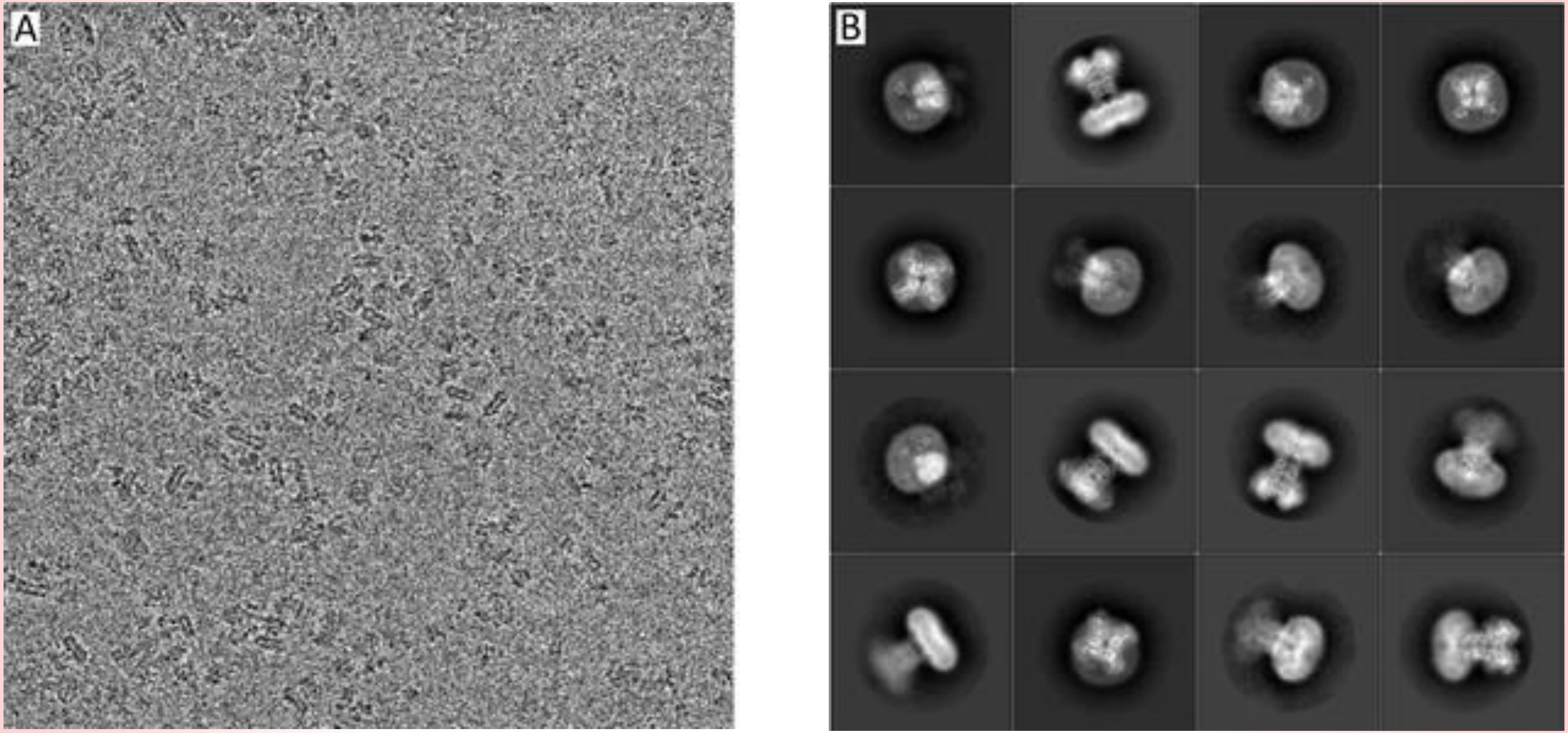
Electron Cryo Microscopy: Experiment





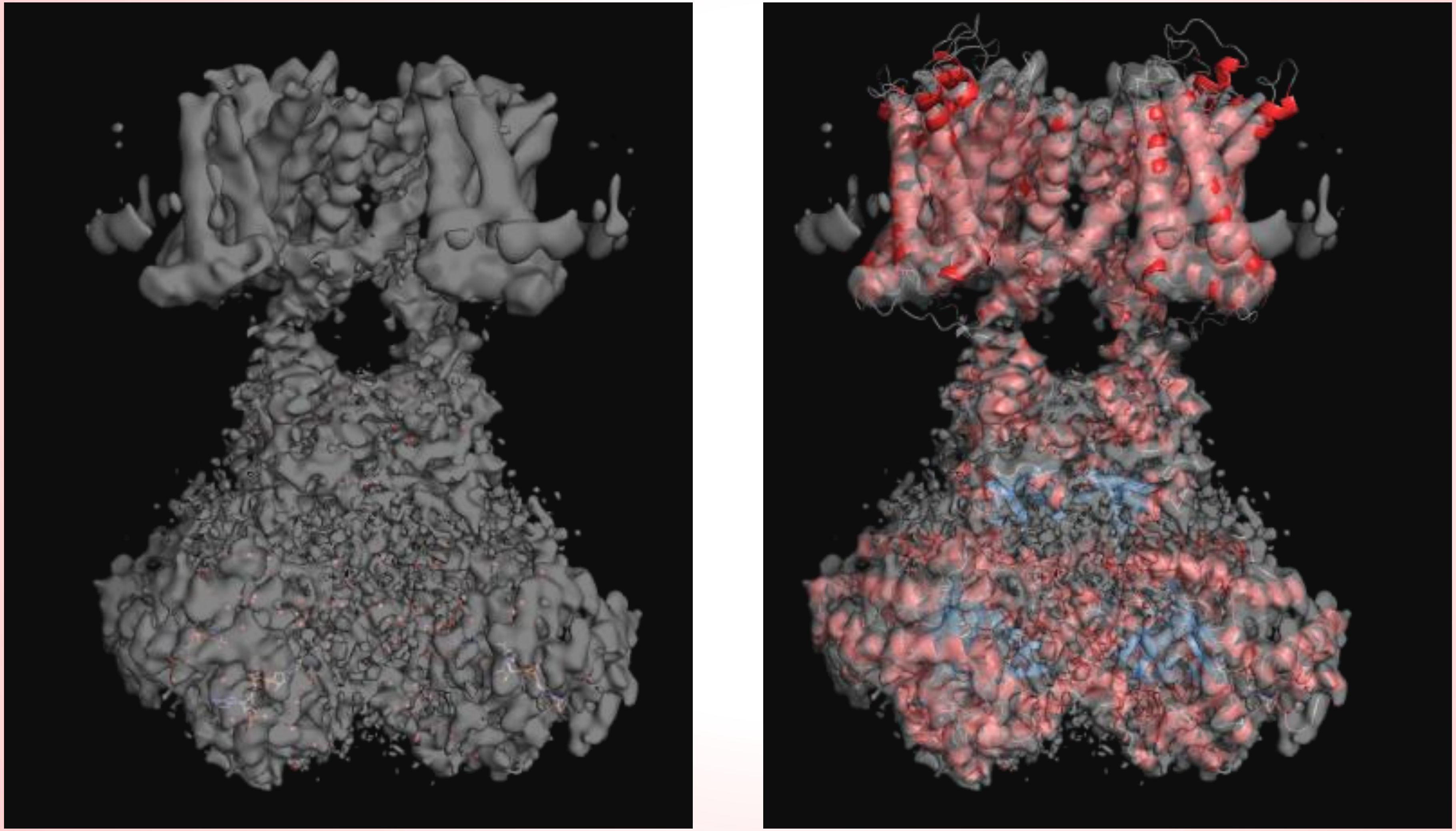
Pictures by E.V. Orlova and lego.com: Vaccine heroes

Electron Cryo Microscopy



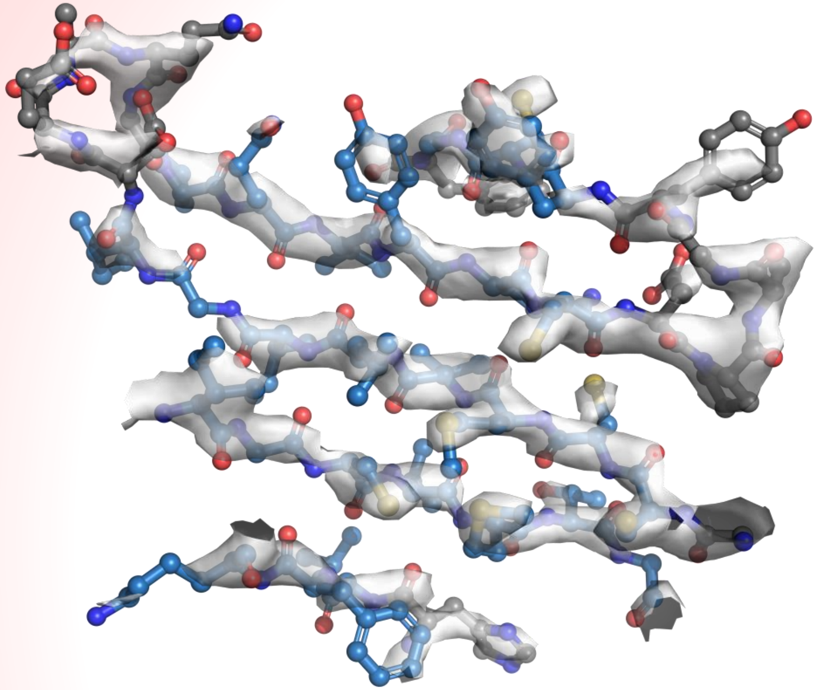
Matthies, D., Bae, C., Toombes, G.E., Fox, T., Bartesaghi, A., Subramaniam, S., Swartz, K.J. (2018) Life 2018;7:e37558

Electron Cryo Microscopy

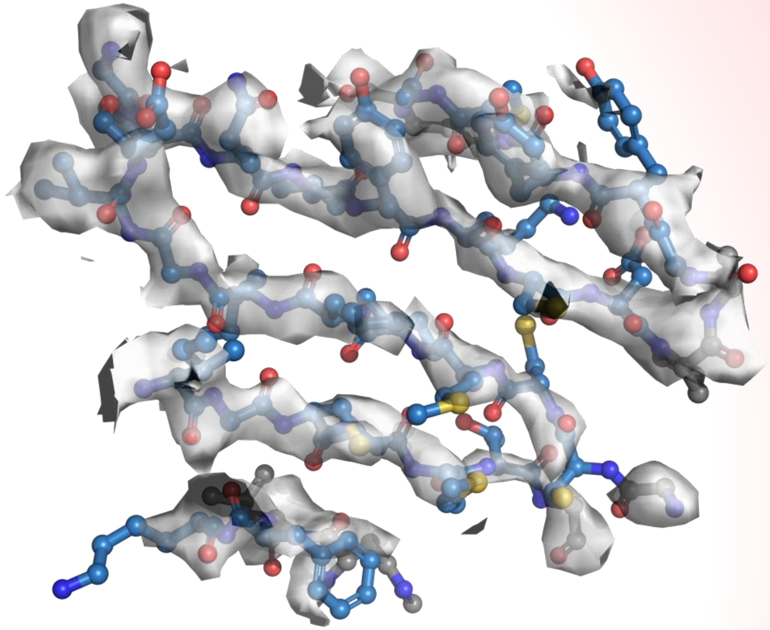


Matthies, D., Bae, C., Toombes, G.E., Fox, T., Bartesaghi, A., Subramaniam, S., Swartz, K.J. (2018) Life 2018;7:e37558

Differences to crystallography



X-ray map at 2.6 Å



EM map at 3.5 Å

- Phases are part of the data: map not made with model
- Electrons instead of X-rays
- Error model…? Validation?

Crystallographic modelling tools are suboptimal for Cryo-EM.

Finding starting structures

Where do I find a starting structure?

- Experimental structure from the Protein Data Bank
- Homology model based on template(s) from the PDB
- Sequence-based structure prediction

In many cases, you have a choice:

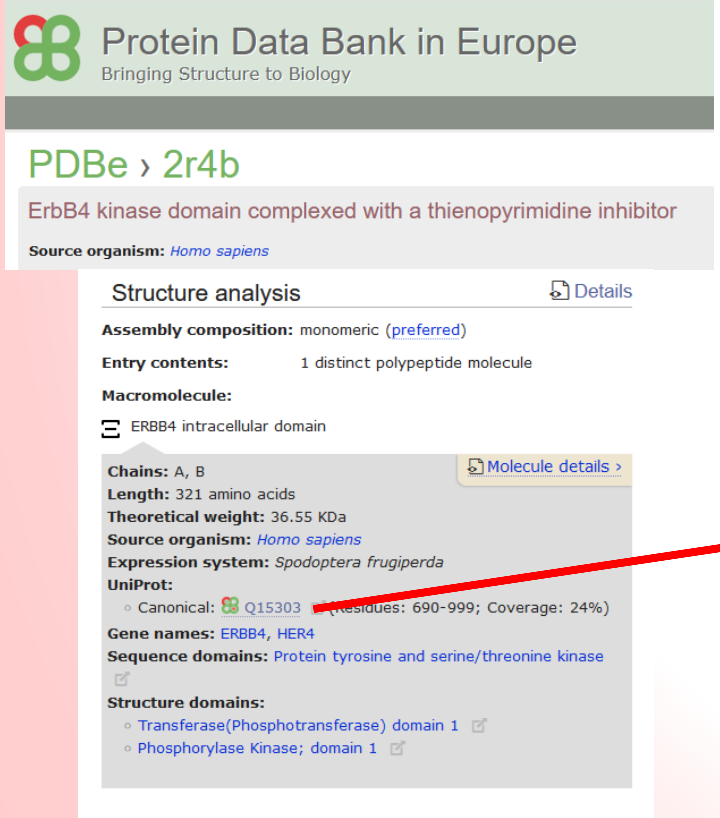
- How do you find all possibilities?
- Which is best in terms of coverage, conformation, quality?

Do not blindly take first available structure!

PDB-KB

1. Start from known structure (kinase domain of ErbB4)

https://www.ebi.ac.uk/pdbe/entry/pdb/2r4b



Protein Data Bank in Europe

Bringing Structure to Biology

PDBe › 2r4b

ErbB4 kinase domain complexed with a thienopyrimidine inhibitor

Source organism: *Homo sapiens*

Structure analysis 	Details

Assembly composition: monomeric (preferred)

Entry contents: 1 distinct polypeptide molecule

Macromolecule:

ERBB4 intracellular domain

Molecule details ›

Chains: A, B

Length: 321 amino acids

Theoretical weight: 36.55 KDa

Source organism: *Homo sapiens*

Expression system: *Spodoptera frugiperda*

UniProt:

- Canonical: Q15303 (Residues: 690-999; Coverage: 24%)

Gene names: ERBB4, HER4

Sequence domains: Protein tyrosine and serine/threonine kinase

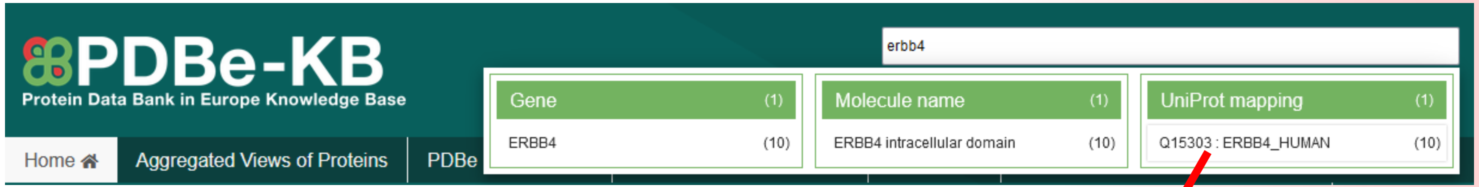
Structure domains:

- Transferase(Phosphotransferase) domain 1
- Phosphorylase Kinase; domain 1

2. Check for other structures

1. Start from protein name or Uniprot ID

https://www.ebi.ac.uk/pdbe/pdbe-kb



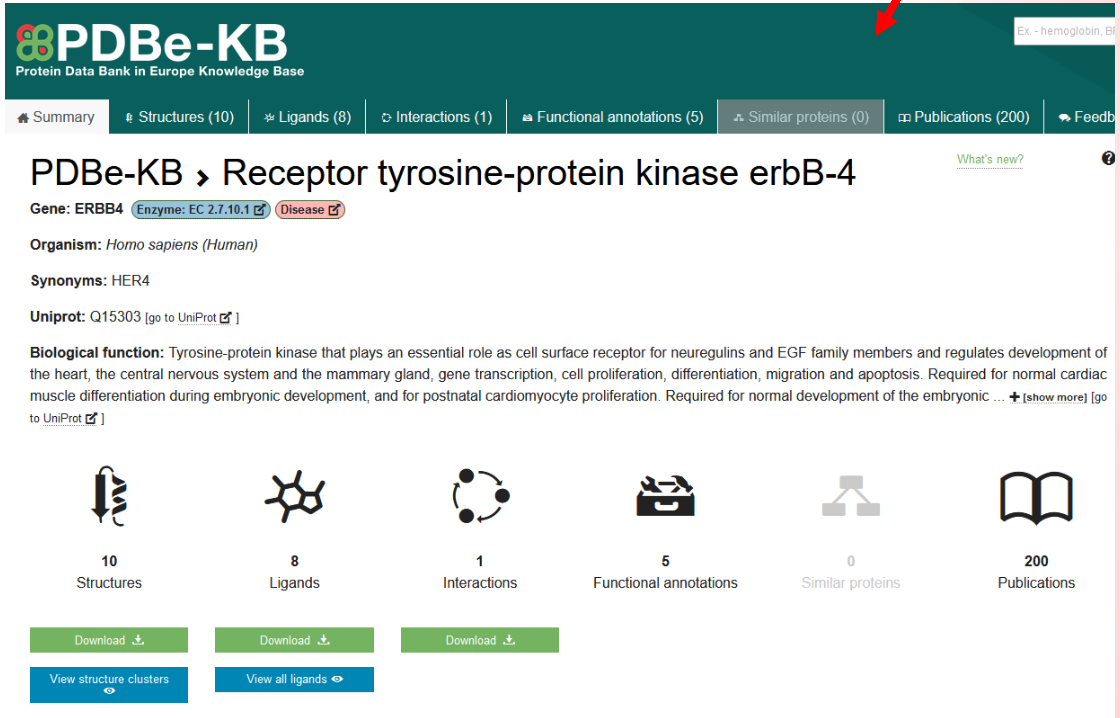
PDBe-KB

Protein Data Bank in Europe Knowledge Base

erbb4

Gene (1)	Molecule name (1)	UniProt mapping (1)
ERBB4 (10)	ERBB4 intracellular domain (10)	Q15303 : ERBB4_HUMAN (10)

2. Check for all structures



PDBe-KB

Protein Data Bank in Europe Knowledge Base

PDBe-KB › Receptor tyrosine-protein kinase erbB-4

Gene: ERBB4 Enzyme: EC 2.7.10.1 Disease

Organism: *Homo sapiens (Human)*

Synonyms: HER4

Uniprot: Q15303 [go to UniProt]

Biological function: Tyrosine-protein kinase that plays an essential role as cell surface receptor for neuregulins and EGF family members and regulates development of the heart, the central nervous system and the mammary gland, gene transcription, cell proliferation, differentiation, migration and apoptosis. Required for normal cardiac muscle differentiation during embryonic development, and for postnatal cardiomyocyte proliferation. Required for normal development of the embryonic ... + [show more] [go to UniProt]

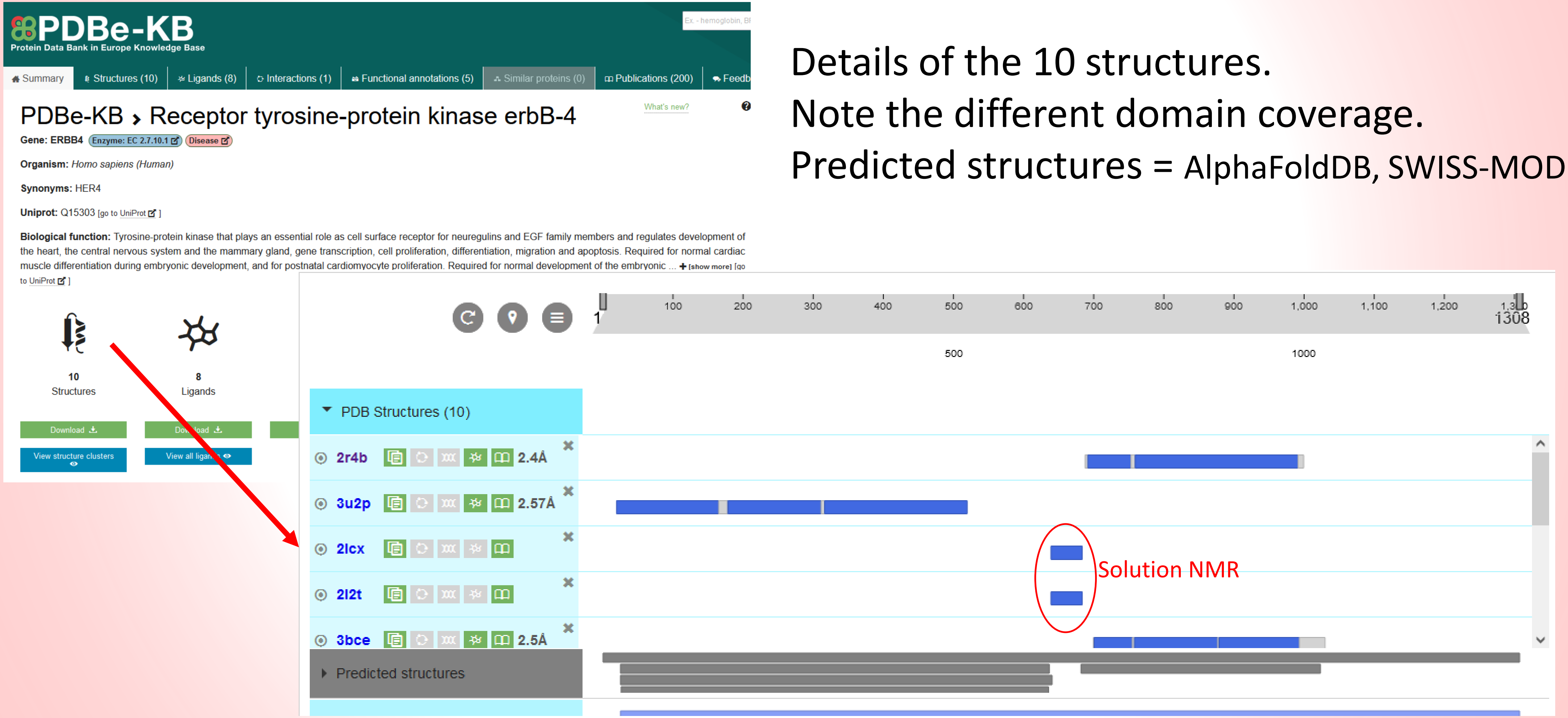
10 Structures	8 Ligands	1 Interactions	5 Functional annotations	0 Similar proteins	200 Publications
Download View structure clusters	Download View all ligands	Download			

Structures

Details of the 10 structures.

Note the different domain coverage.

Predicted structures = AlphaFoldDB, SWISS-MODEL



PDBe-KB
Protein Data Bank in Europe Knowledge Base

PDBe-KB › Receptor tyrosine-protein kinase erbB-4

Gene: ERBB4 Enzyme: EC 2.7.10.1 Disease

Organism: *Homo sapiens (Human)*

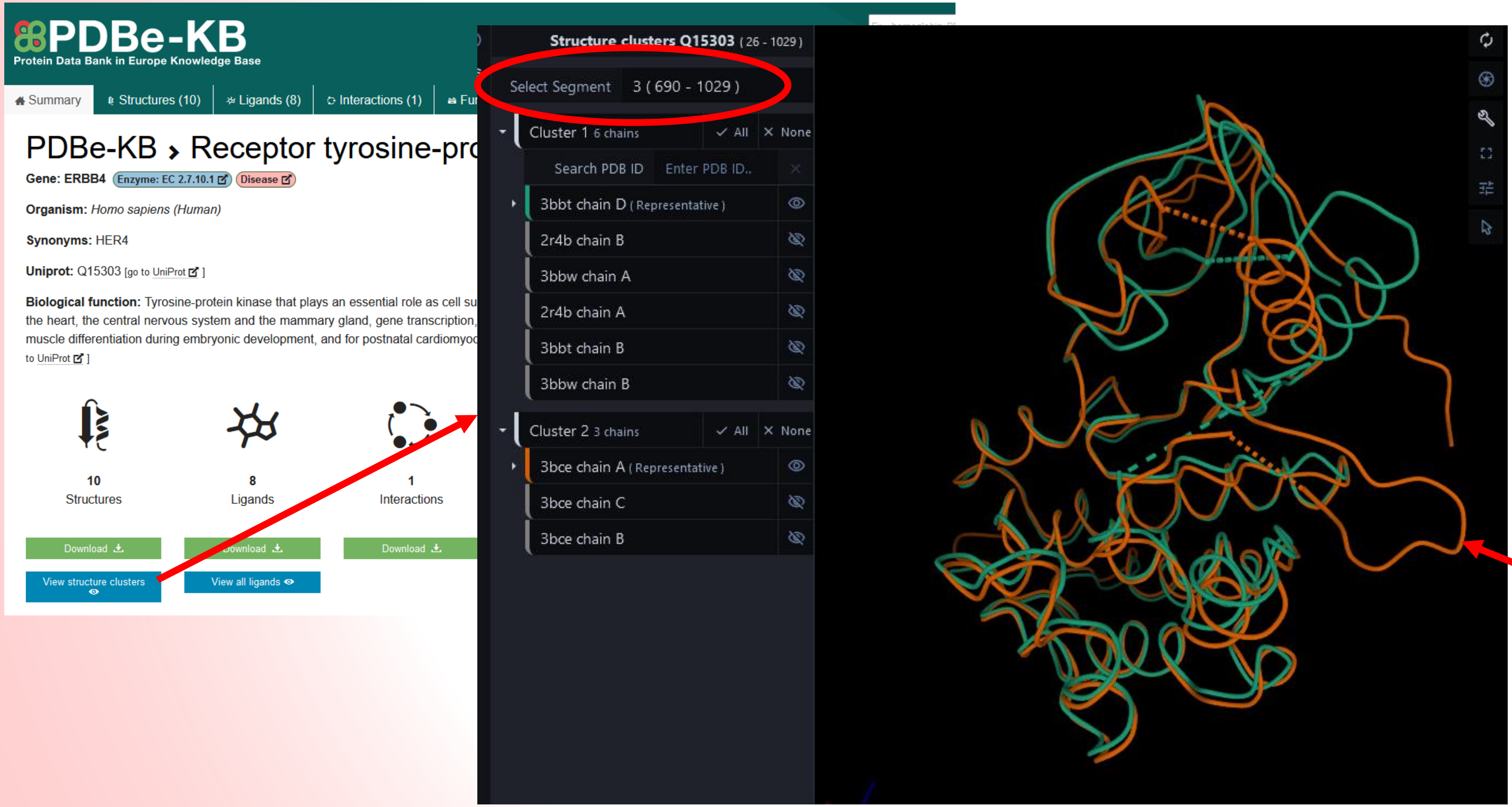
Synonyms: HER4

Uniprot: Q15303 [go to UniProt]

Biological function: Tyrosine-protein kinase that plays an essential role as cell surface receptor for neuregulins and EGF family members and regulates development of the heart, the central nervous system and the mammary gland, gene transcription, cell proliferation, differentiation, migration and apoptosis. Required for normal cardiac muscle differentiation during embryonic development, and for postnatal cardiomyocyte proliferation. Required for normal development of the embryonic ... + [show more] [go to UniProt]

Structure clusters

Choose Segment 3 = kinase domain



PDBe-KB › Receptor tyrosine-pro

Gene: ERBB4 Enzyme: EC 2.7.10.1 Disease

Organism: *Homo sapiens (Human)*

Synonyms: HER4

Uniprot: Q15303 [go to UniProt]

Biological function: Tyrosine-protein kinase that plays an essential role as cell su
the heart, the central nervous system and the mammary gland, gene transcription,
muscle differentiation during embryonic development, and for postnatal cardiomyoc
to UniProt]

Structure clusters Q15303 (26 - 1029)

Select Segment 3 (690 - 1029)

Cluster 1 6 chains ✓ All ✕ None

Search PDB ID Enter PDB ID..

- 3bbt chain D (Representative)
- 2r4b chain B
- 3bbw chain A
- 2r4b chain A
- 3bbt chain B
- 3bbw chain B

Cluster 2 3 chains ✓ All ✕ None

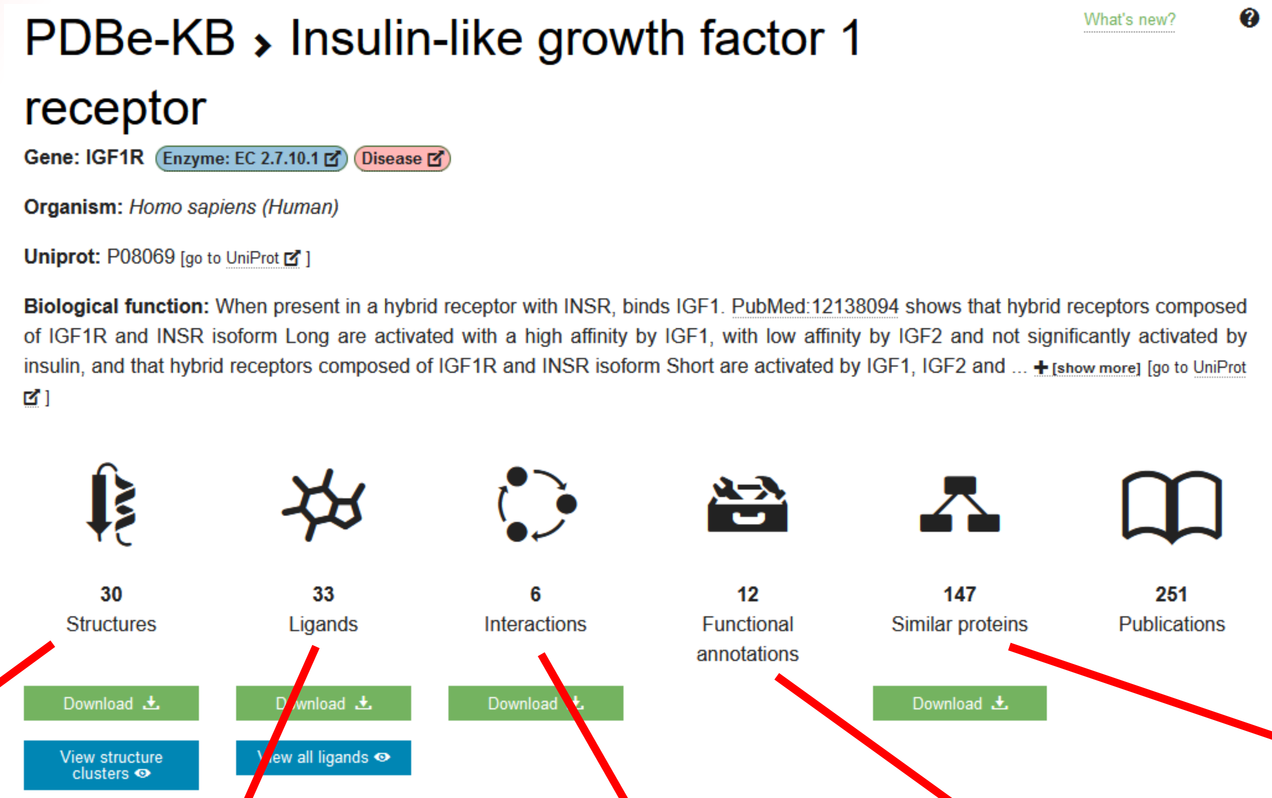
- 3bce chain A (Representative)
- 3bce chain C
- 3bce chain B

Two cluster representatives: active and inactive conformations.

Activation loop of activated structure.

Toggle individual structures on and off.
Identify residues of interest.

Complementary data



PDBe-KB › Insulin-like growth factor 1 receptor

Gene: IGF1R Enzyme: EC 2.7.10.1 Disease

Organism: *Homo sapiens (Human)*

Uniprot: P08069 [go to UniProt]

Biological function: When present in a hybrid receptor with INSR, binds IGF1. PubMed:12138094 shows that hybrid receptors composed of IGF1R and INSR isoform Long are activated with a high affinity by IGF1, with low affinity by IGF2 and not significantly activated by insulin, and that hybrid receptors composed of IGF1R and INSR isoform Short are activated by IGF1, IGF2 and ... + [show more] [go to UniProt]

PDB entries for this protein

Observed ligands (including glycans, ions) and where they bind.

Observed interactions (protein, NA) and interface residues

Domains, sequence variants, etc

Other PDB entries with > 90% sequence identity (UniRef90)

How to choose a structure (1st attempt)

1. Use PDBe-KB to find list for protein of interest
   - for ErbB4 there are 10
2. Check the domain coverage and unmodelled loops
   - there are 4 for the kinase domain, with similar completeness
3. Check the conformation
   - for the inactive conformation, there are 3 structures / 6 chains
4. Look at any ligands or interactions
5. Check for mutations
6. Look at global resolution

With a short-list, move to more detailed quality assessment (next topic)

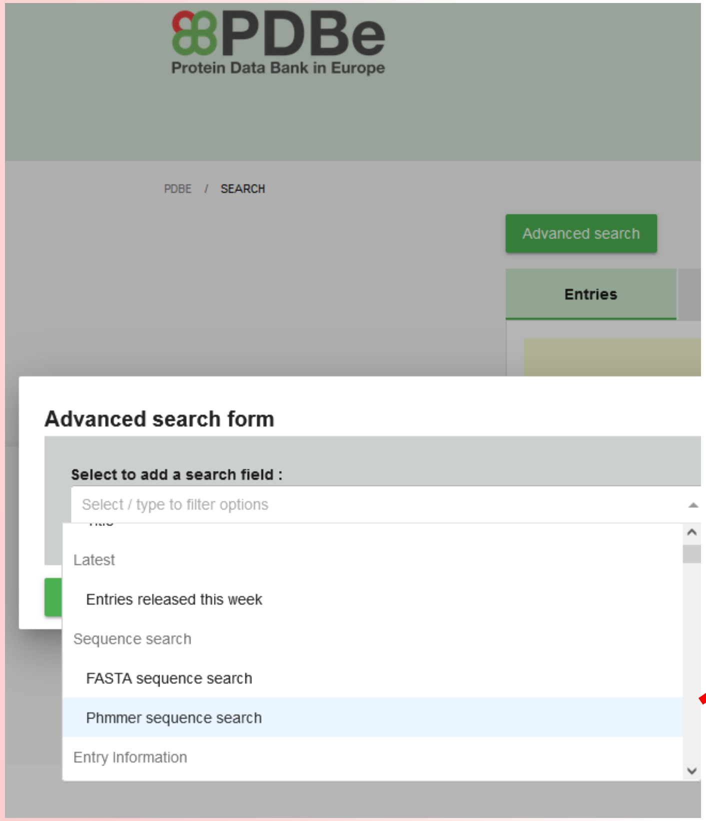
Exercise: investigate ErbB4 ectodomain structures

Homology models

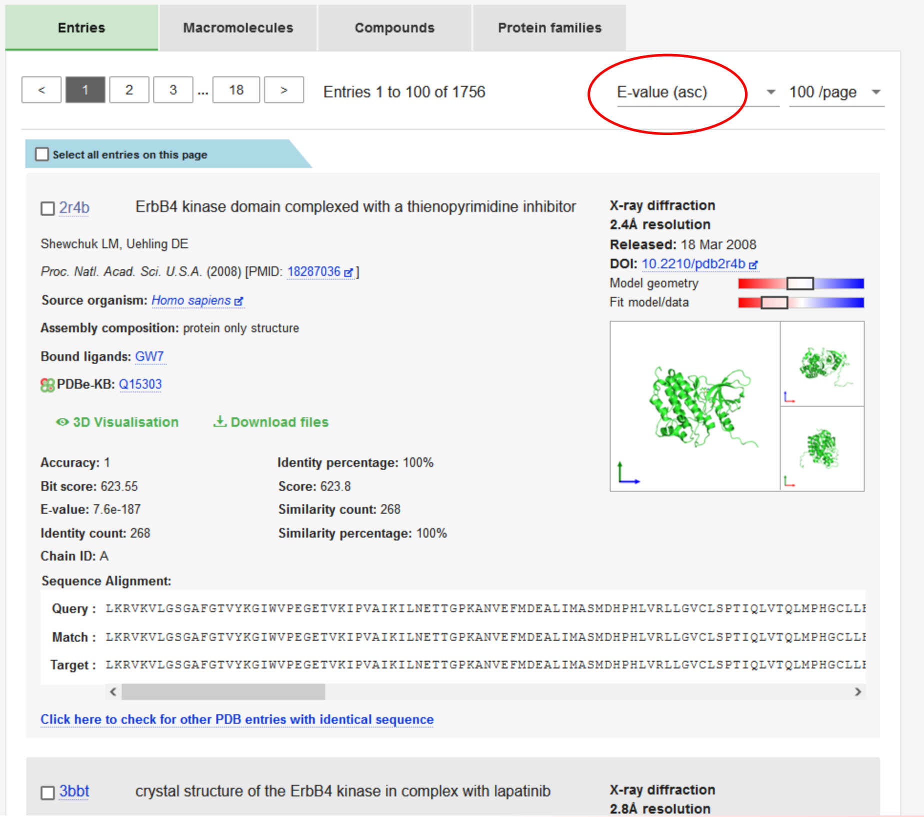
Sequence-based search of the PDB.

https://www.ebi.ac.uk/pdbe/entry/search/index/?advancedSearch:true=

ErbB4, then EGFR, ErbB2, ErbB3, etc



Search on ErbB4 kinase domain sequence



Entries 1 to 100 of 1756

☐ 2r4b ErbB4 kinase domain complexed with a thienopyrimidine inhibitor

Shewchuk LM, Uehling DE

Proc. Natl. Acad. Sci. U.S.A. (2008) [PMID: 18287036]

Source organism: *Homo sapiens*

Assembly composition: protein only structure

Bound ligands: GW7

PDBe-KB: Q15303

X-ray diffraction
2.4Å resolution
Released: 18 Mar 2008
DOI: 10.2210/pdb2r4b
Model geometry
Fit model/data

Accuracy: 1
Bit score: 623.55
E-value: 7.6e-187
Identity count: 268
Chain ID: A

Identity percentage: 100%
Score: 623.8
Similarity count: 268
Similarity percentage: 100%

Sequence Alignment:

```
Query :  LKRVKVLGSGAFGTVYKGIWVPEGETVKIPVAIKILNETTGPKANVEFMDEALIMASMDHPHLVRLLGVCLSPTIQLVTQLMPHGCLL
Match :  LKRVKVLGSGAFGTVYKGIWVPEGETVKIPVAIKILNETTGPKANVEFMDEALIMASMDHPHLVRLLGVCLSPTIQLVTQLMPHGCLL
Target : LKRVKVLGSGAFGTVYKGIWVPEGETVKIPVAIKILNETTGPKANVEFMDEALIMASMDHPHLVRLLGVCLSPTIQLVTQLMPHGCLL
```

Click here to check for other PDB entries with identical sequence

☐ 3bbt crystal structure of the ErbB4 kinase in complex with lapatinib

X-ray diffraction
2.8Å resolution

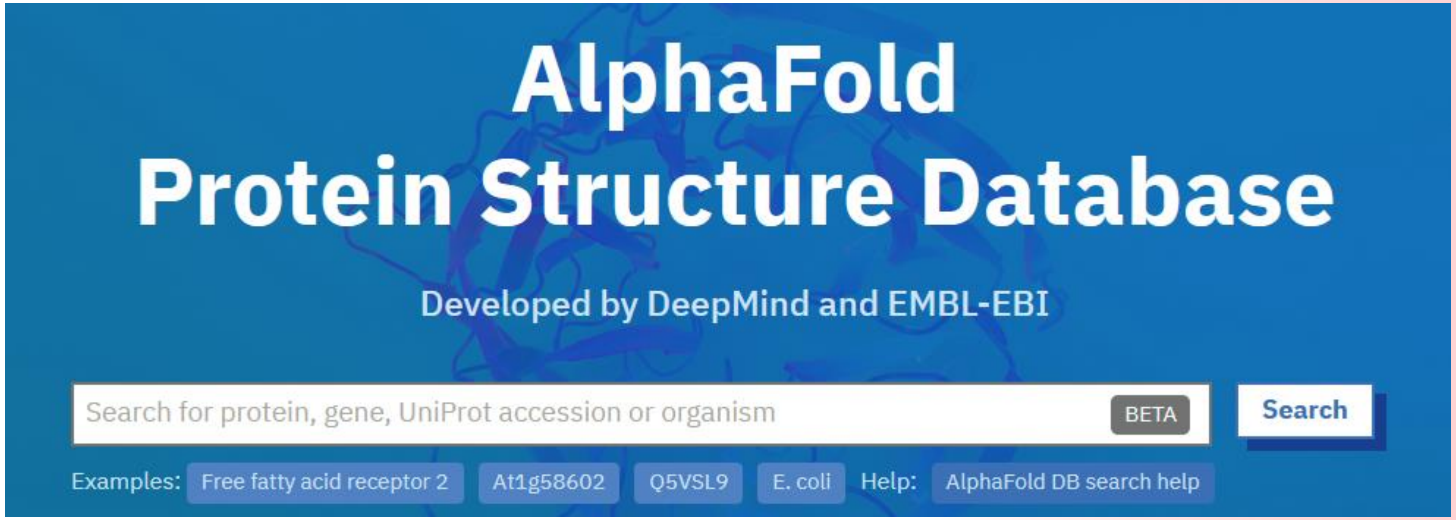
Template(s) for homology modelling

Or use available services (SWISS-MODEL, Phyre2, etc)

Do novo models

What if there is no suitable starting model in the PDB ?
De novo modelling from sequence improving all the time.

AlphaFold database.
Predictions for 21 model organisms.
Discussion later …


AlphaFold
Protein Structure Database

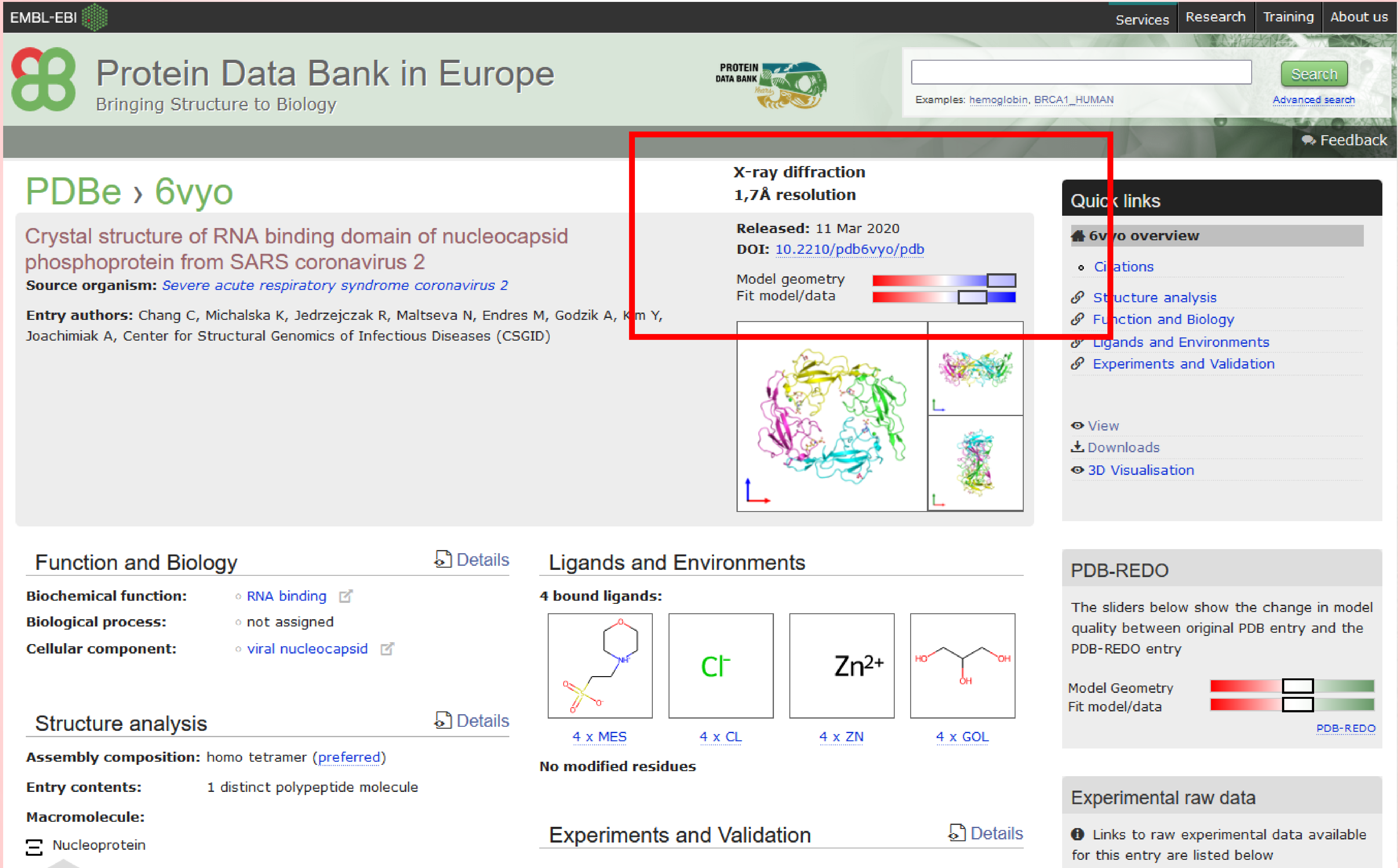
Developed by DeepMind and EMBL-EBI

Other servers …
RoseTTAFold https://robetta.bakerlab.org
GREMLIN gremlin.bakerlab.org
PConsFam pconsfam.bioinfo.se

Finding starting structures
Exercise 1a

Explore available structures of receptor tyrosine kinase ErbB4, using PDBe-KB, with focus on ectodomain.

Validation Report and common metrics for crystallography



Protein Data Bank in Europe

Bringing Structure to Biology

Examples: hemoglobin, BRCA1_HUMAN

PDBe › 6vyo

X-ray diffraction
1,7Å resolution

Released: 11 Mar 2020
DOI: 10.2210/pdb6vyo/pdb

Model geometry
Fit model/data

Crystal structure of RNA binding domain of nucleocapsid phosphoprotein from SARS coronavirus 2

Source organism: *Severe acute respiratory syndrome coronavirus 2*

Entry authors: Chang C, Michalska K, Jedrzejczak R, Maltseva N, Endres M, Godzik A, Kim Y, Joachimiak A, Center for Structural Genomics of Infectious Diseases (CSGID)

Function and Biology

Biochemical function:	RNA binding
Biological process:	not assigned
Cellular component:	viral nucleocapsid

Structure analysis

Assembly composition: homo tetramer (preferred)

Entry contents: 1 distinct polypeptide molecule

Macromolecule:

Nucleoprotein

Ligands and Environments

4 bound ligands:

4 x MES
Cl⁻ 4 x CL
Zn²⁺ 4 x ZN
4 x GOL

No modified residues

Experiments and Validation

PDB-REDO

The sliders below show the change in model quality between original PDB entry and the PDB-REDO entry

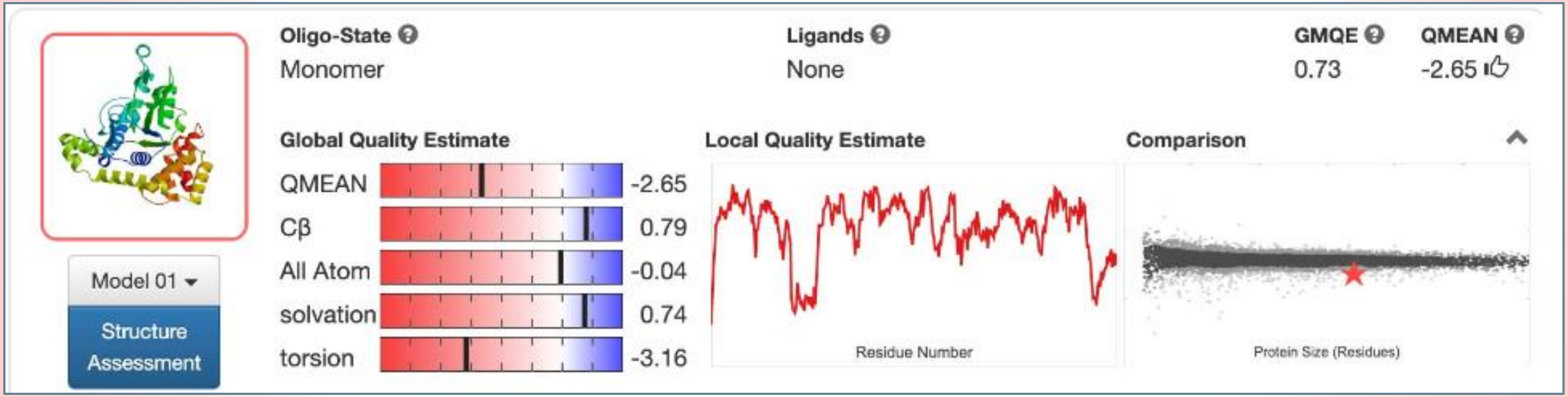
Model Geometry
Fit model/data

Experimental raw data

Links to raw experimental data available for this entry are listed below

Wait, I know those sliders!

Qualitative Model Energy Analysis used by SwissModel



Oligo-State
Monomer

Ligands
None

GMQE
0.73

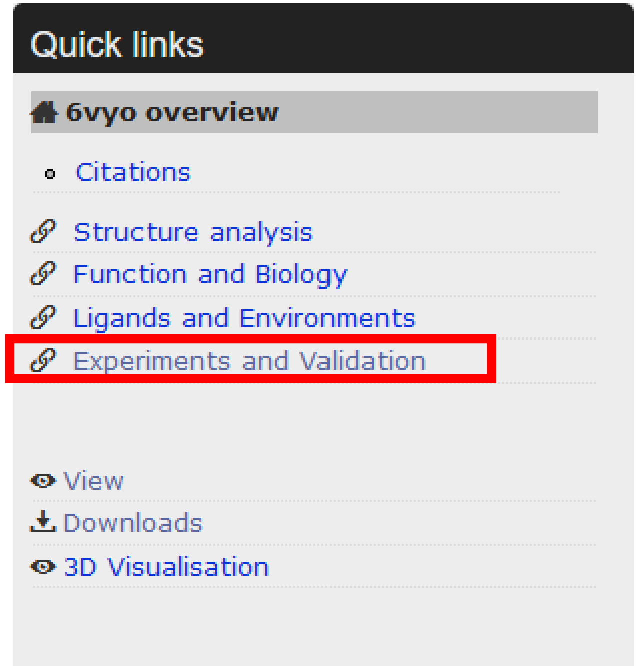
QMEAN
-2.65

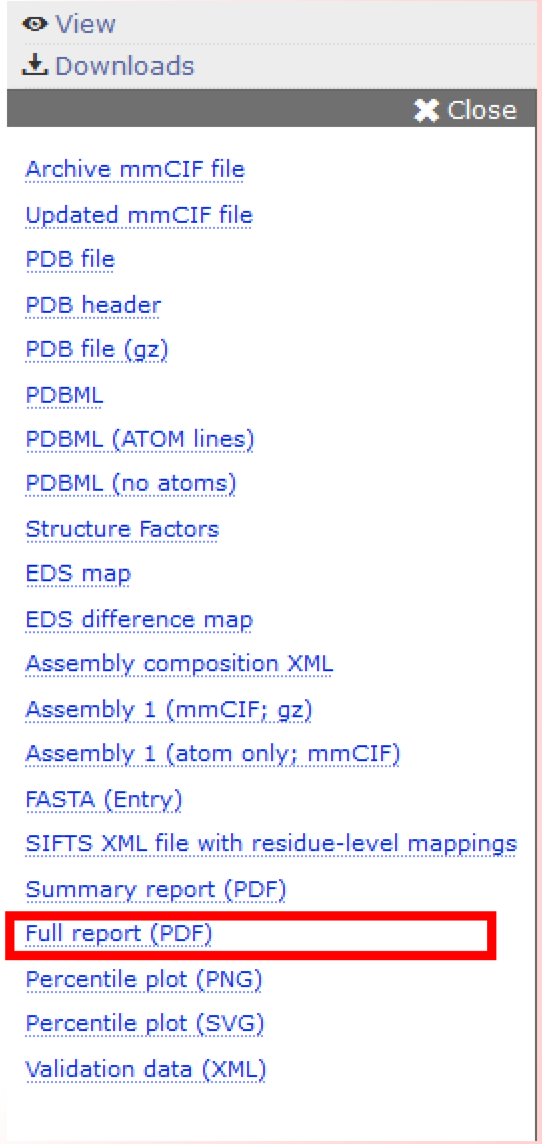
Global Quality Estimate

QMEAN	-2.65
Cβ	0.79
All Atom	-0.04
solvation	0.74
torsion	-3.16

Picture: Salam Al-Karadaghi, Lund University CC-BY-NC

Where to find PDB validation?





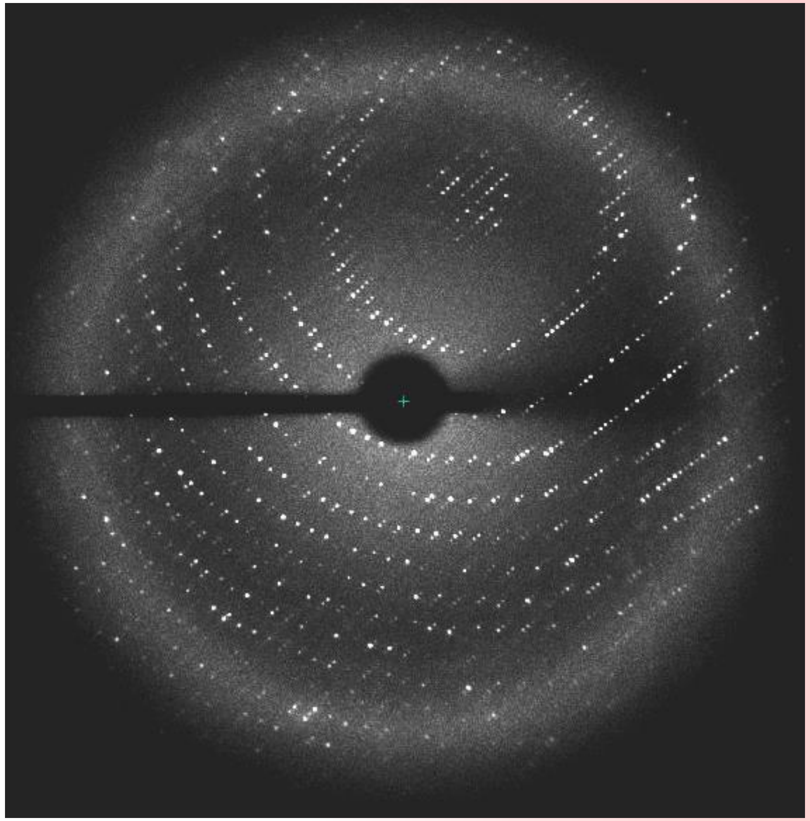
What to look for?

- **Sequence completeness** (see previous section)
- **Resolution**: How much information was there to begin with?
- **R-values**: How well does the model fit the data?
- **Geometric validation** criteria:
  - Ramachandran /CaBLAM
  - Clashes
  - Side chain conformations = rotamers

Resolution: How much information was there to begin with?

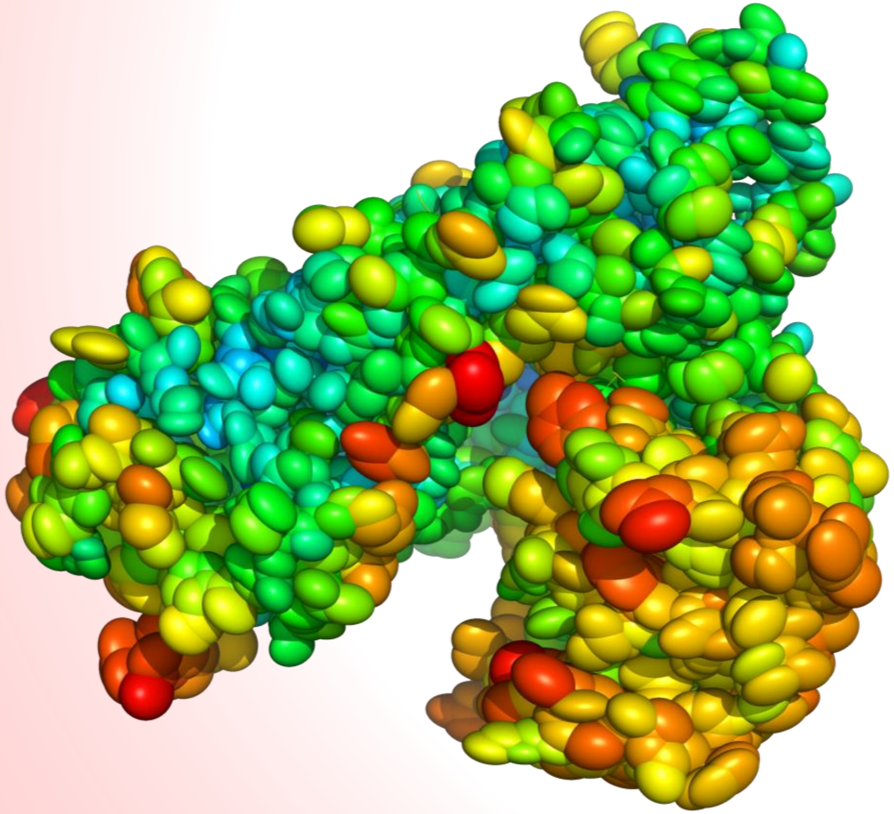
- Chosen by experimentalist:
  small differences irrelevant
- Data quality can still be bad

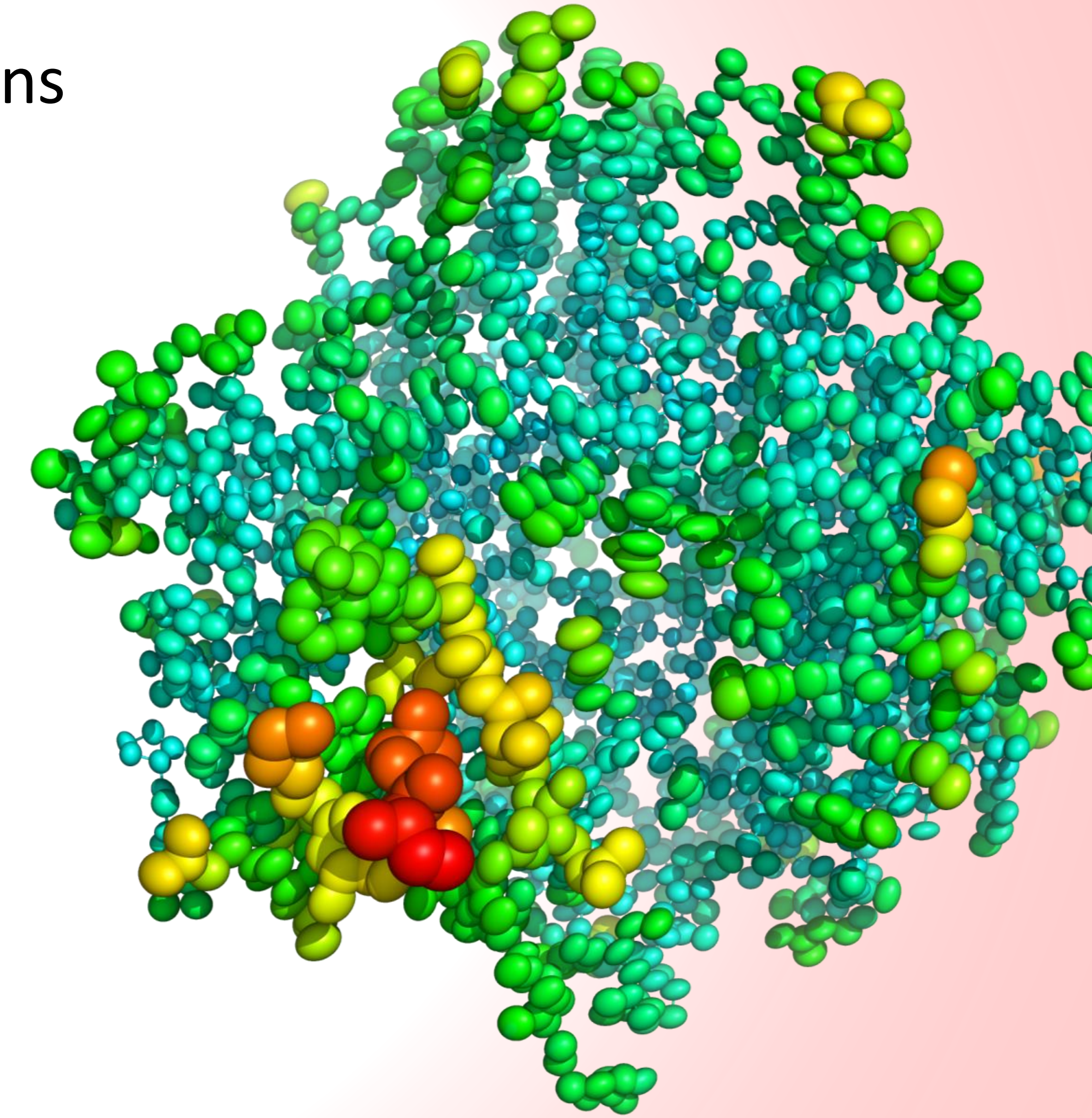
< 1.7 Å	individual atomic positions
1.7 - 2.6 Å	conformations mostly be correct ideal geometry information is used
2.6 - 3.7 Å	fold correct errors in conformation & register rotamer libraries are used
> 3.7 Å	fold can be determined individual coordinates meaningless



B factors

- Uncertainty of individual atomic positions
- Can be isotropic or anisotropic



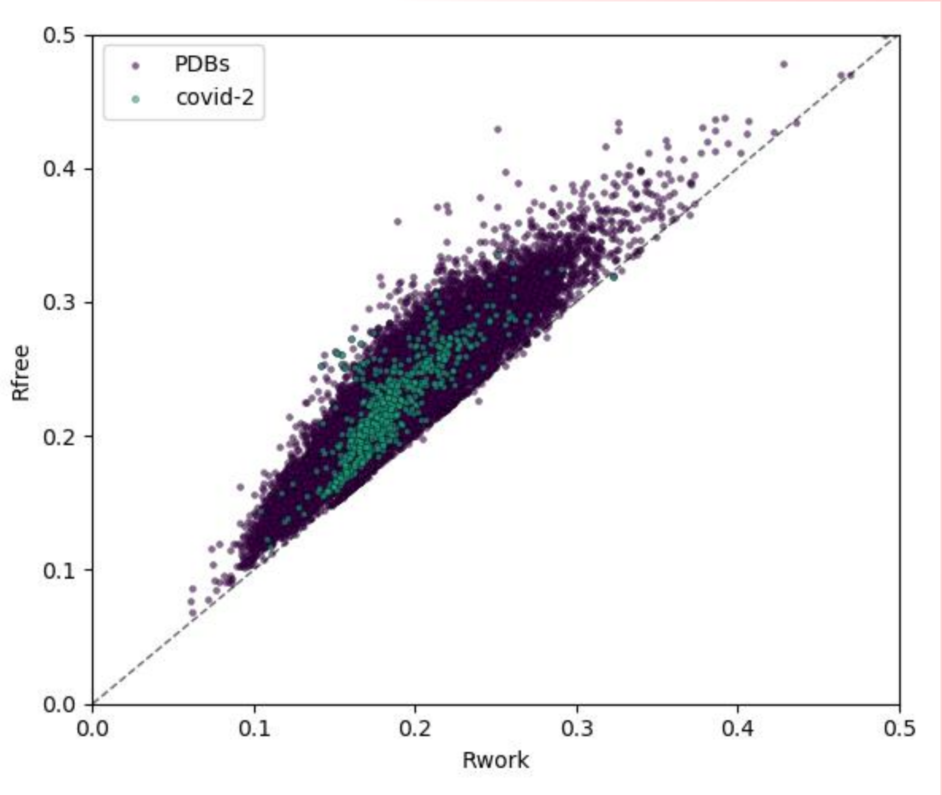


Pictures courtesy of Nick Pearce

R values

- Residual value
- Discrepancy between diffraction data and model
- The lower the better!
- R_{free} semi-independent criterion, must always be higher than the R-value
- Typical R-values:

  0.24, or 24% at 2-3 Å resolution
- Geometry can still be very low



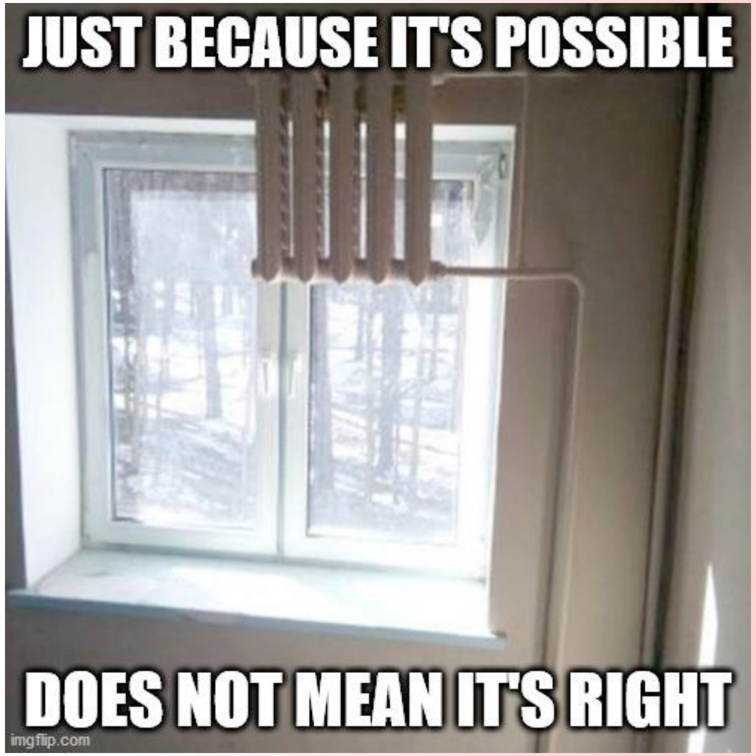
R-values of SARS-CoV-2 structures (cyan) and all X-ray crystallography structures from the last five years (purple dots)

Pictures courtesy of Yunyun Gao / Coronavirus Structural Task Force

Geometric validation

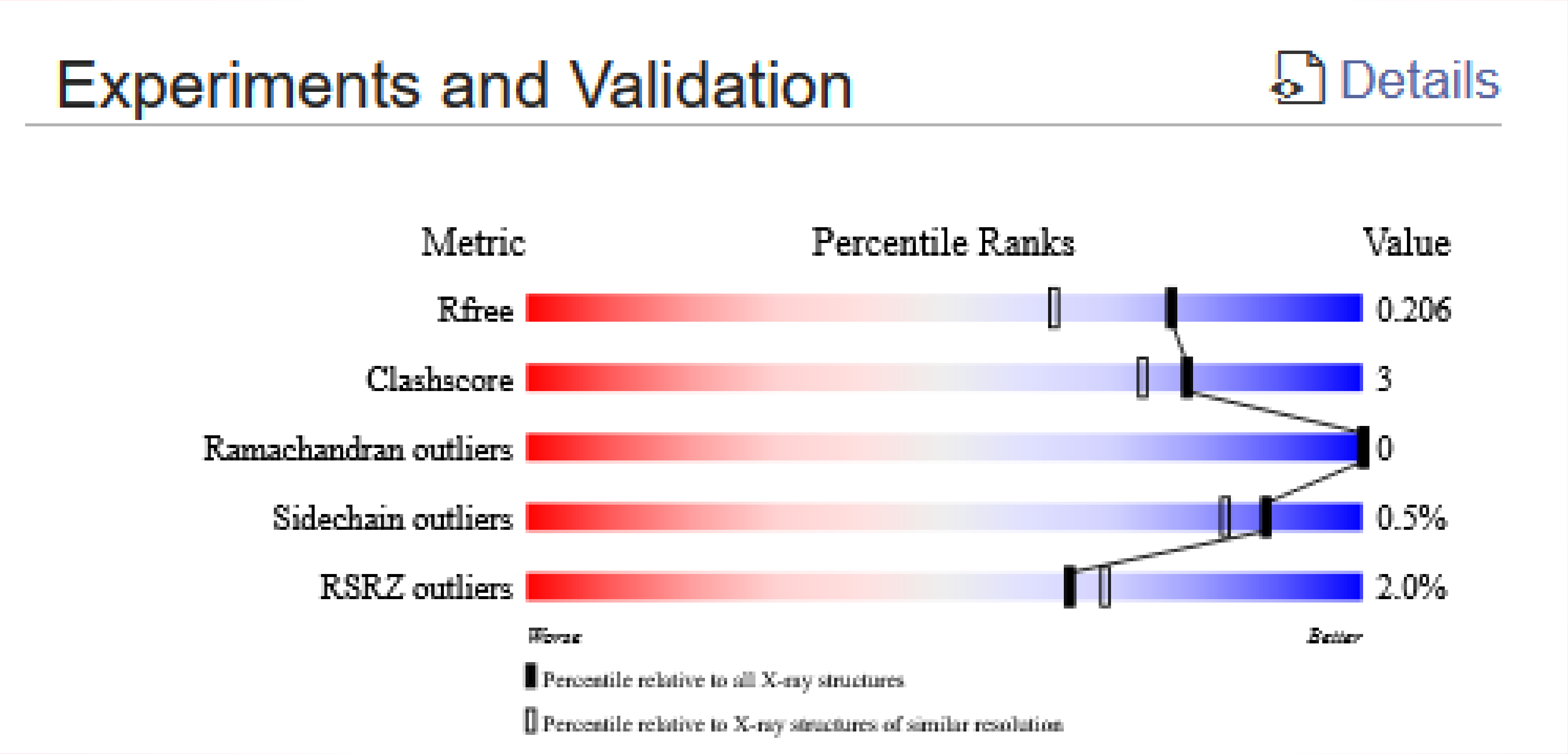
Very similar for X-ray and Cryo-EM as the same models are used!

- Ramachandran / CaBLAM
- Clashscore
- Uncommon rotamers
- Wrong rotamer, glycosylation, ligand geometry

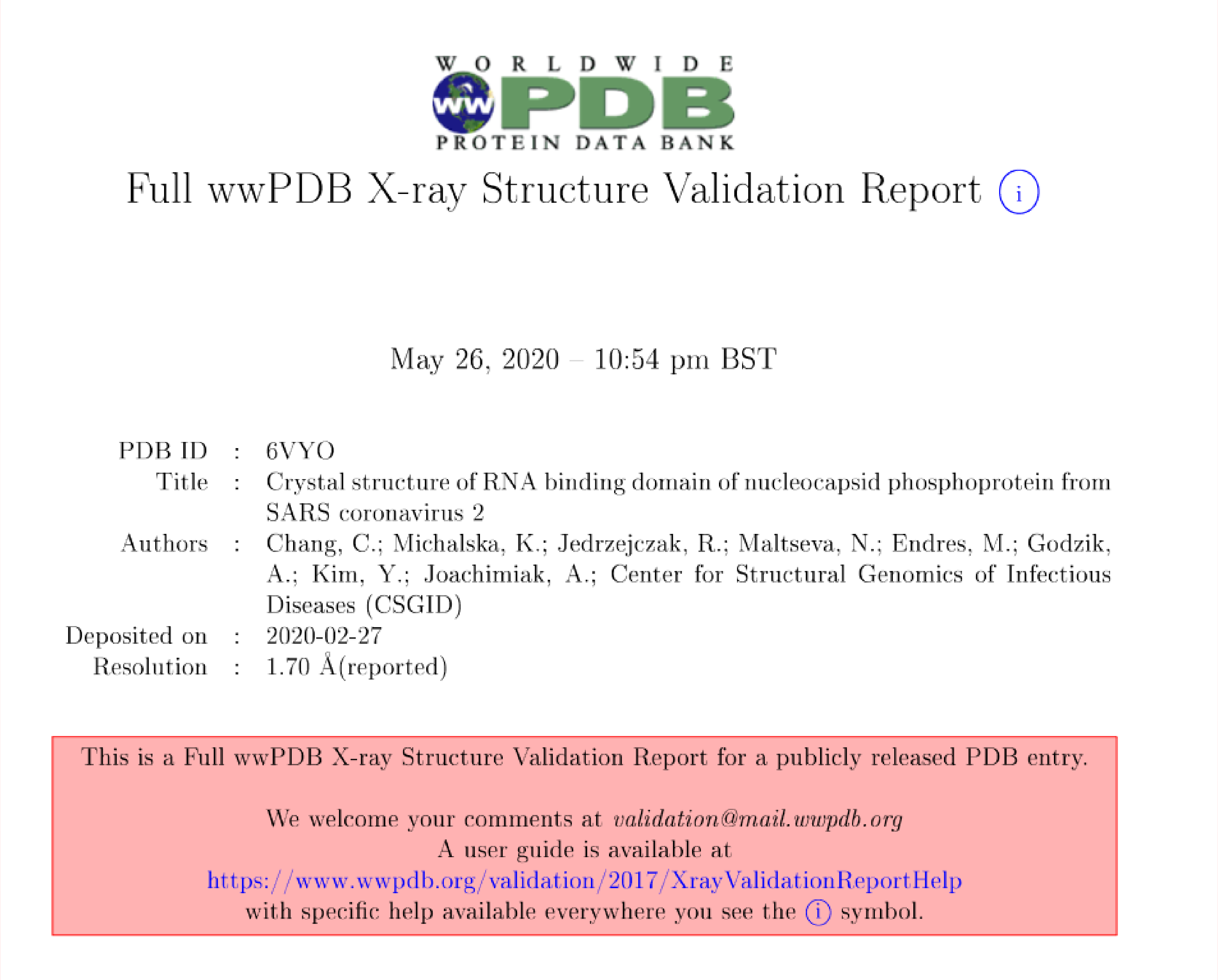


CoA Picture courtesy of Bernhard Rupp

Back to PDBe



Experiments and Validation



WORLDWIDE
wwPDB
PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report ⓘ

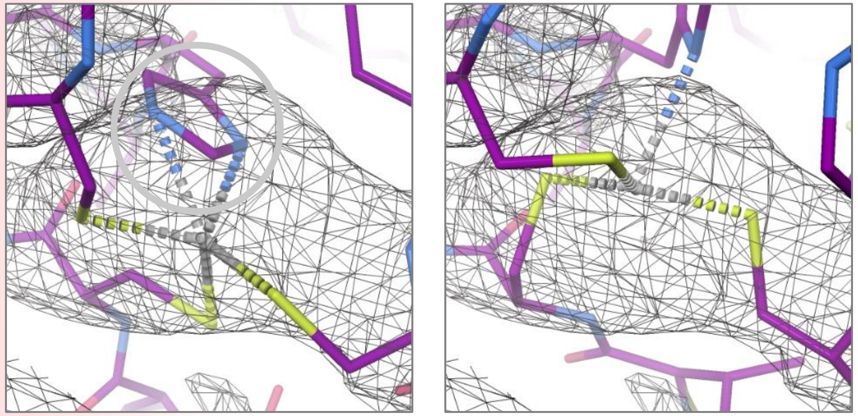
May 26, 2020 – 10:54 pm BST

PDB ID : 6VYO
Title : Crystal structure of RNA binding domain of nucleocapsid phosphoprotein from SARS coronavirus 2
Authors : Chang, C.; Michalska, K.; Jedrzejczak, R.; Maltseva, N.; Endres, M.; Godzik, A.; Kim, Y.; Joachimiak, A.; Center for Structural Genomics of Infectious Diseases (CSGID)
Deposited on : 2020-02-27
Resolution : 1.70 Å(reported)

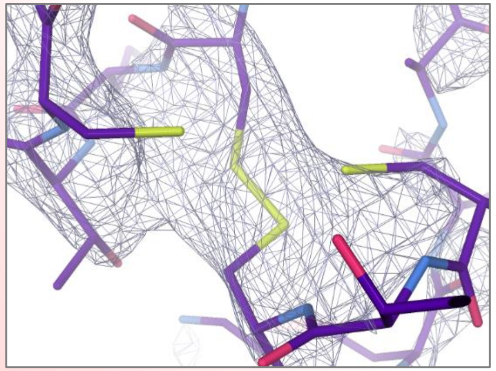
Beware!

Nothing really replaces looking at the structure with expertise

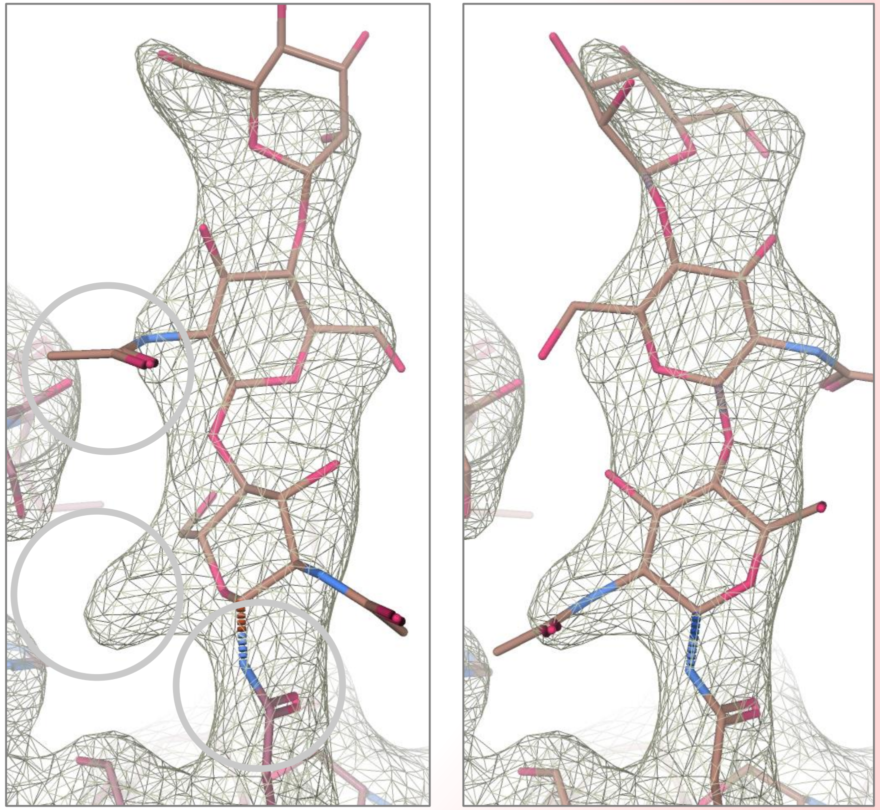
-> Train this!



PDB 5c8t nsp14-nsp10
histidine-zinc coordination



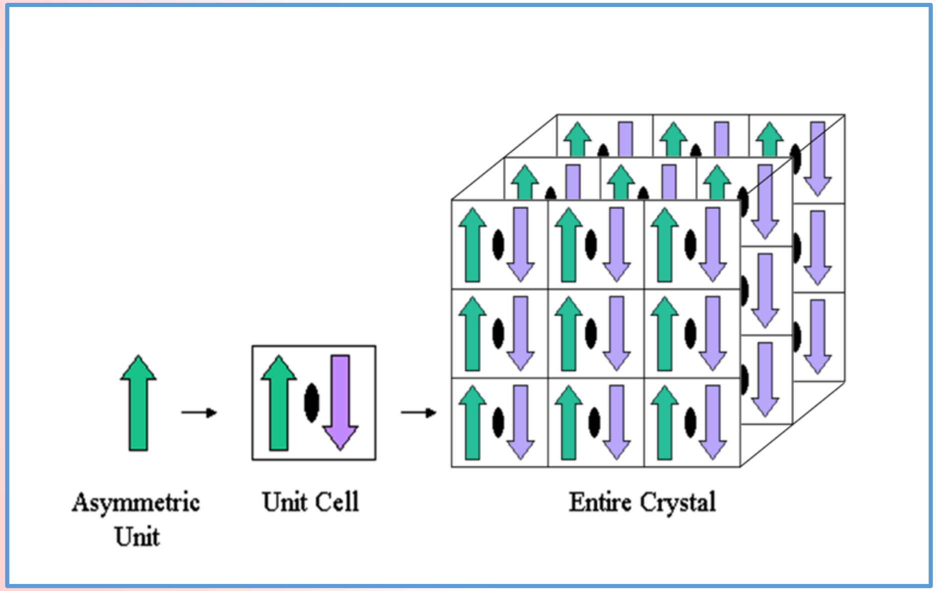
Papain-like protease
(PDB 6w9c)



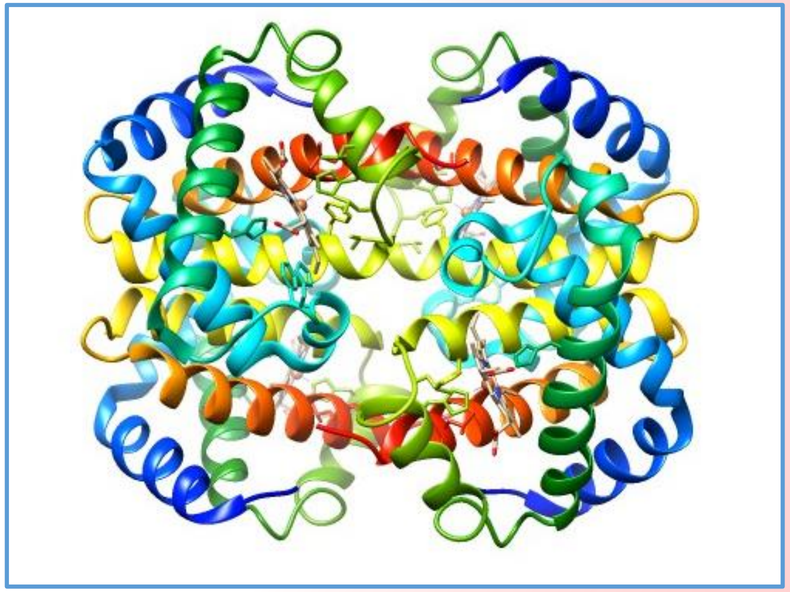
PDB 6vw1 spike bound to ACE2
glycosidic bond chirality

Beware #2: Biological assembly vs. asymmetric unit

- Whole structure may not be present upon opening a pdb file
- It contains only the “asymmetric unit” (ASU)
- Need to get “biological assembly” by generating symmetry mates



ASU: 2 chains
half a hemoglobin



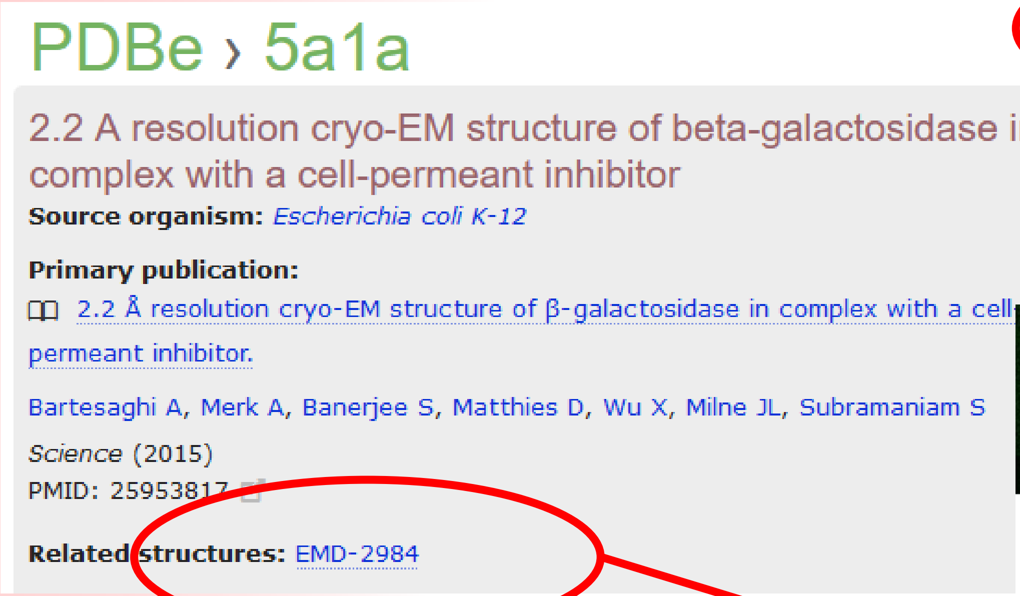
Biological assembly: 4 chains
a whole hemoglobin

https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/biological-assemblies

Validation Report and common metrics for cryoEM

CryoEM structures

16/9/2021
16,623 entries in EMDB
8,351 have fitted model in PDB



PDBe › 5a1a

2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor

Source organism: *Escherichia coli K-12*

Primary publication:

📖 2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor.

Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JL, Subramaniam S

Science (2015)
PMID: 25953817

Related structures: EMD-2984

Electron Microscopy
2.2Å resolution

Released: 06 May 2015
DOI: 10.2210/pdb5a1a/pdb

Link to EMDB

EMD-2984
Single-particle
2.2 Å

Deposition: 26/04/2015
Map released: 06/05/2015
Last modified: 14/10/2015

EMD-2984

2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor

Sample Organism: Escherichia coli K-12, unidentified
Sample: Escherichia coli beta-galactosidase bound to phenylethyl beta-D-thiogalactopyranoside (PETG)
Fitted models: 5a1a
Raw data: EMPIAR-10061

Deposition Authors: Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne J, Subramaniam S

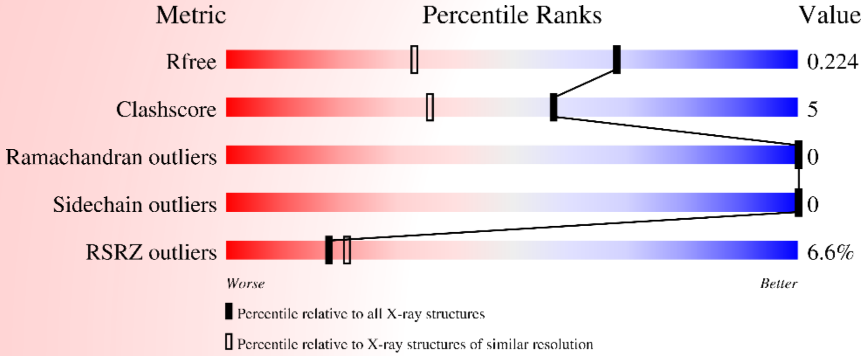
📖 **2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor**

Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JL, Subramaniam S
(2015) Science , 348 , 1147 - 1151

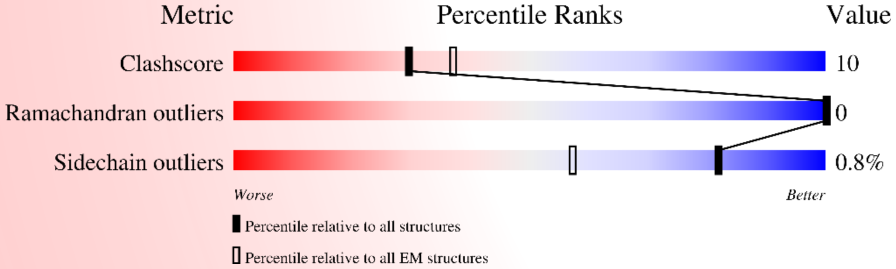
PUBMED: 25953817

PDB contains atomistic model,
EMDB contains EM map.

CryoEM - PDB entry page



xtal



cryoEM

No indication whether model agrees with the experiment!

Validation under development

Tier 1

Validation Analysis pages – for experts

Tier 2

Validation tab on EMDB page

Tier 3

Validation Report from PDB page

Tier 3: Validation Report (PDB)

Overall quality at a glance

Entry composition

Residue-property plots

Experimental information

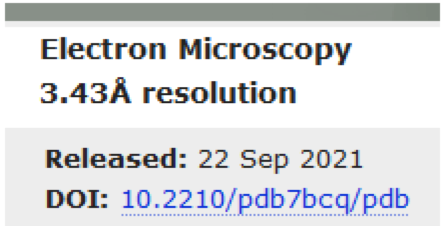
Model quality – geometry, clashes, Ramachandran, rotamers (*as xtal*)

Map visualisation – projections, central slices, largest variance slices, surface

Map analysis – map value dist^n, volume estimate, RAPS

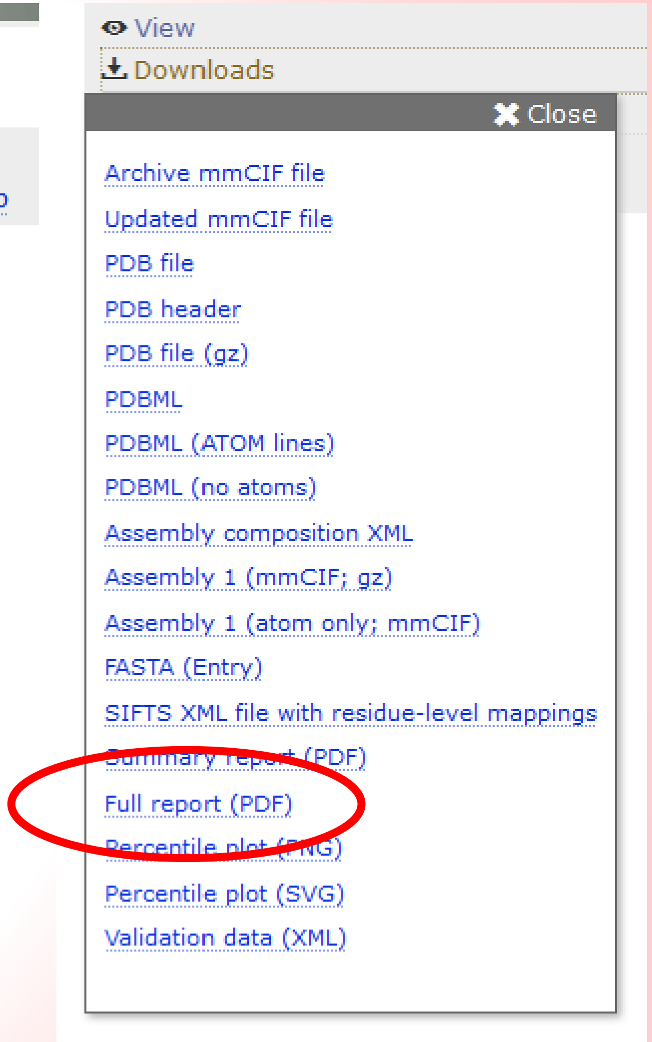
Fourier-Shell correlation – resolution estimates

Map-model fit – map-model overlay, atom inclusion



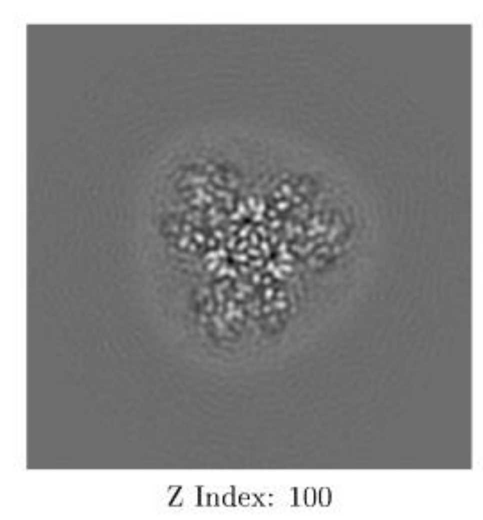
Electron Microscopy
3.43Å resolution

Released: 22 Sep 2021
DOI: 10.2210/pdb7bcq/pdb

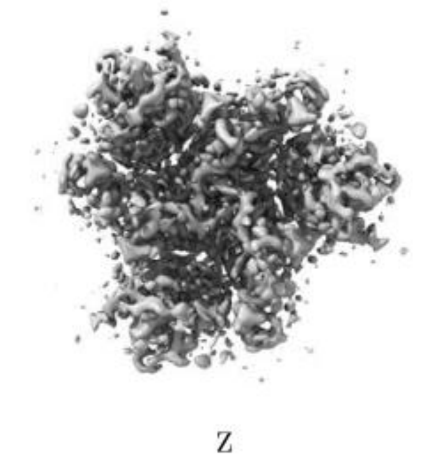


Archive mmCIF file
Updated mmCIF file
PDB file
PDB header
PDB file (gz)
PDBML
PDBML (ATOM lines)
PDBML (no atoms)
Assembly composition XML
Assembly 1 (mmCIF; gz)
Assembly 1 (atom only; mmCIF)
FASTA (Entry)
SIFTS XML file with residue-level mappings
Summary report (PDF)
Full report (PDF)
Percentile plot (PNG)
Percentile plot (SVG)
Validation data (XML)

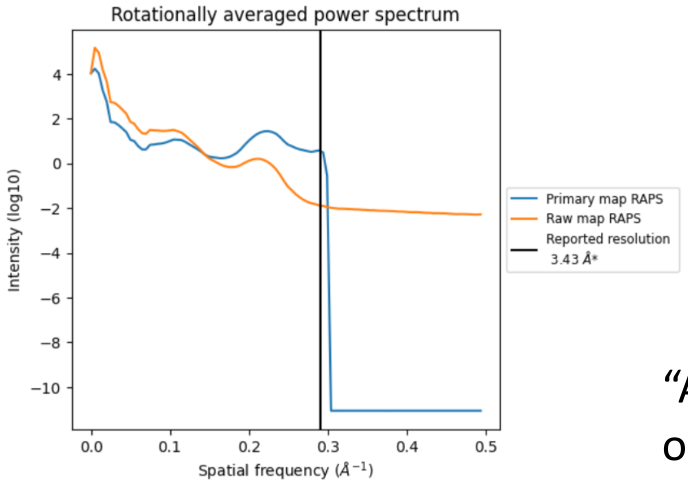
Some examples from Validation Report



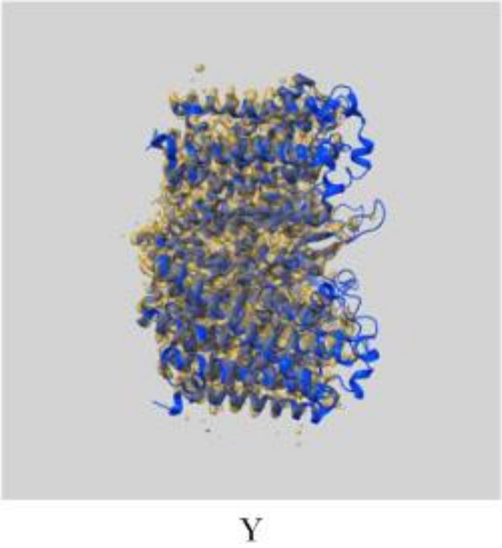
Largest variance slice EMD-12142



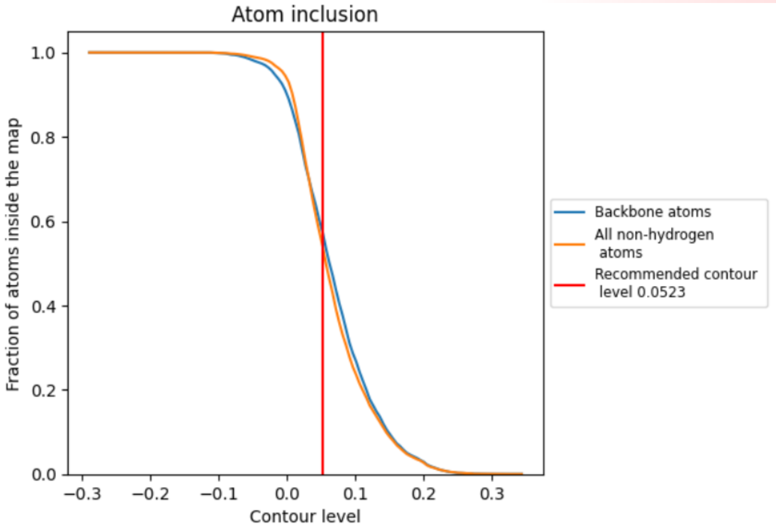
Surface view EMD-12142



RAPS:
Map probably sharpened.
High res terms truncated.

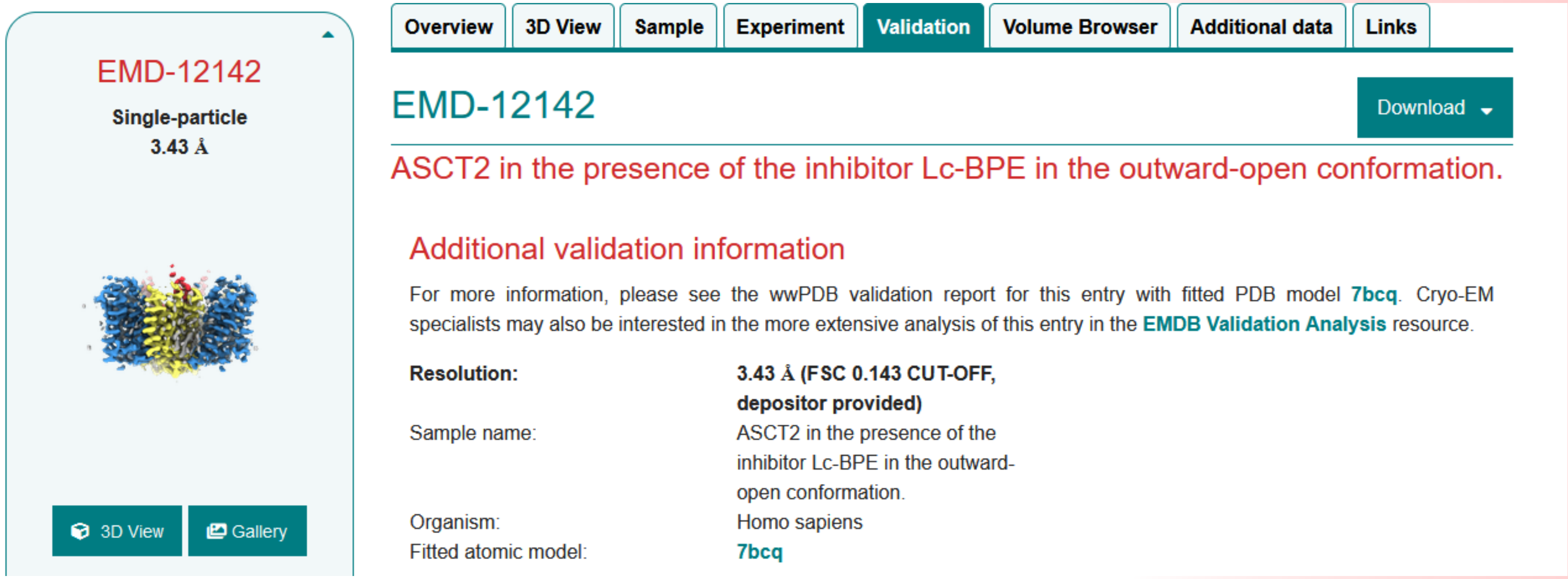


“At the recommended contour level, 57% of all backbone atoms, 54% of all non-hydrogen atoms, are inside the map.”



Tier 2: Validation tab on EMDB page

- **Resolution** (and how determined)
- Recommended **contour** level
- **Map visualisation** – projections, central slices, largest variance slices, surface
- **Map analysis** – map value distn, volume estimate, RAPS
- **Fourier-Shell correlation** – resolution estimates
- **Map-model fit** – map-model overlay, atom inclusion
- **Symmetry** information



EMD-12142

ASCT2 in the presence of the inhibitor Lc-BPE in the outward-open conformation.

Additional validation information

For more information, please see the wwPDB validation report for this entry with fitted PDB model **7bcq**. Cryo-EM specialists may also be interested in the more extensive analysis of this entry in the **EMDB Validation Analysis** resource.

Resolution:	**3.43 Å (FSC 0.143 CUT-OFF, depositor provided)**
Sample name:	ASCT2 in the presence of the inhibitor Lc-BPE in the outward-open conformation.
Organism:	Homo sapiens
Fitted atomic model:	**7bcq**

Compared to PDB validation report:

- Additional map visualisations
- Atom inclusion by residue
- Point group symmetry

Perhaps easier to browse.

Tier 1: Common metrics for cryoEM

Model only:

Molprobity: ideal geometry, clashes, CaBLAM

Side chain packing powerful but rarely used.

Model to map:

CC, Q-score, SMOC, 3D-Strudel, FSC-Q

Many: local ones are the most useful.

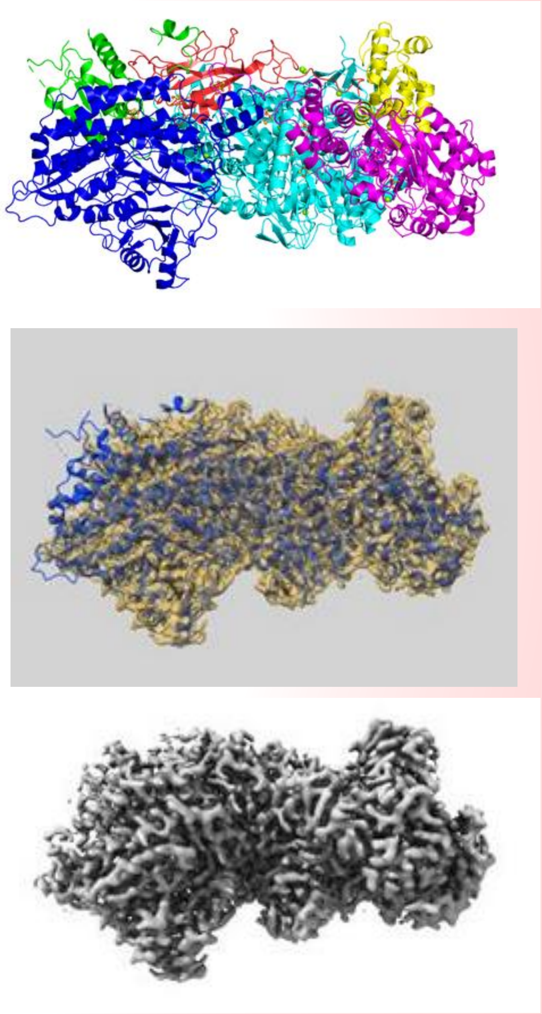
Problem: hard to put these on absolute scale.

Map only:

Assess overfitting, orientation bias, particle inclusion

Not relevant here, but aware that map is derived

7awt



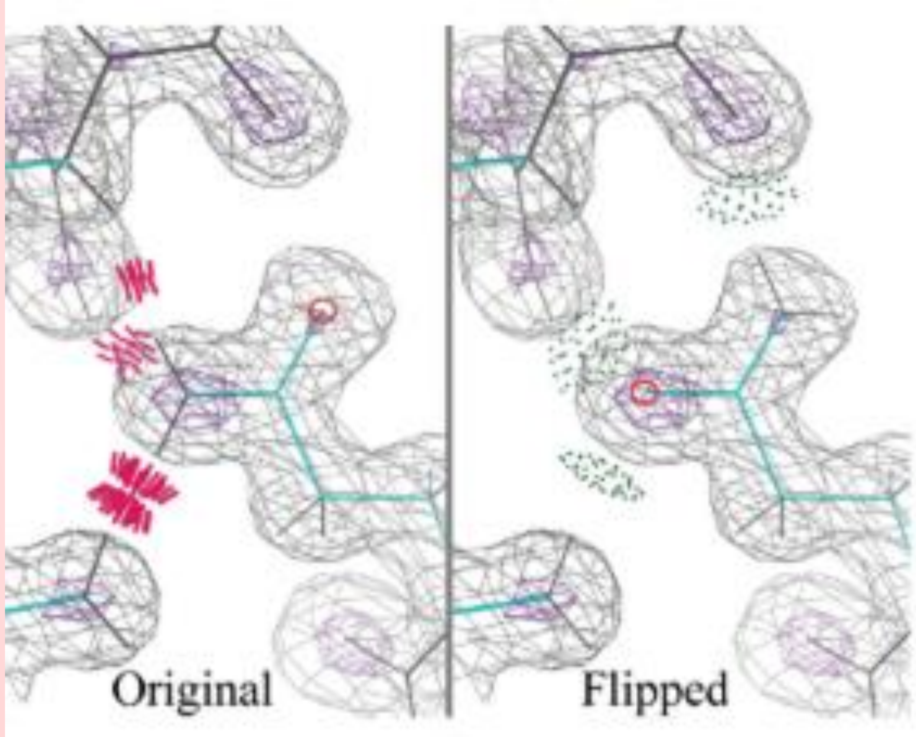
EMD-11930

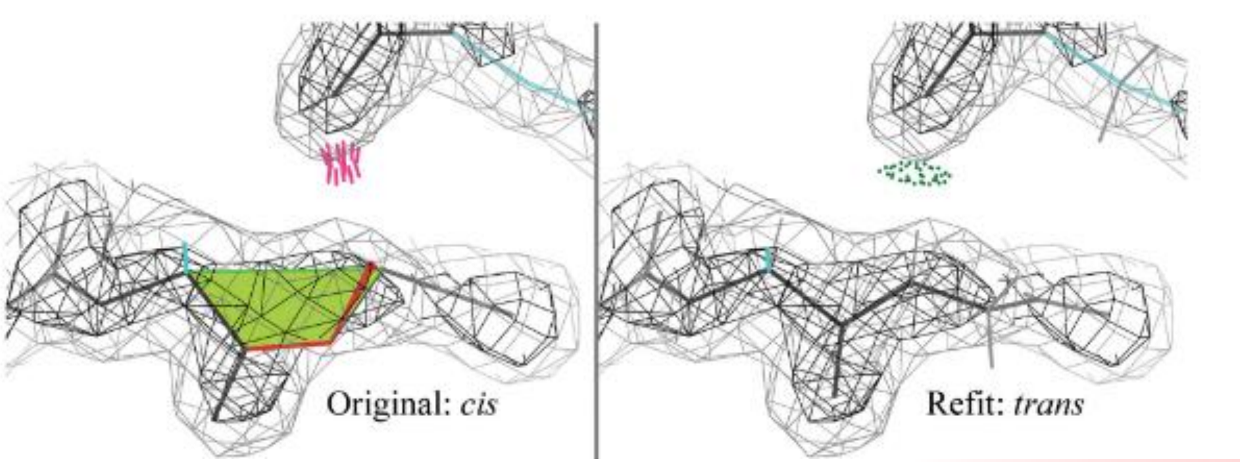
NADH quinone oxidoreductase

Geometry problems

Richardson et al. 2018; 2cn3

Clashes





Often too many cis-peptides

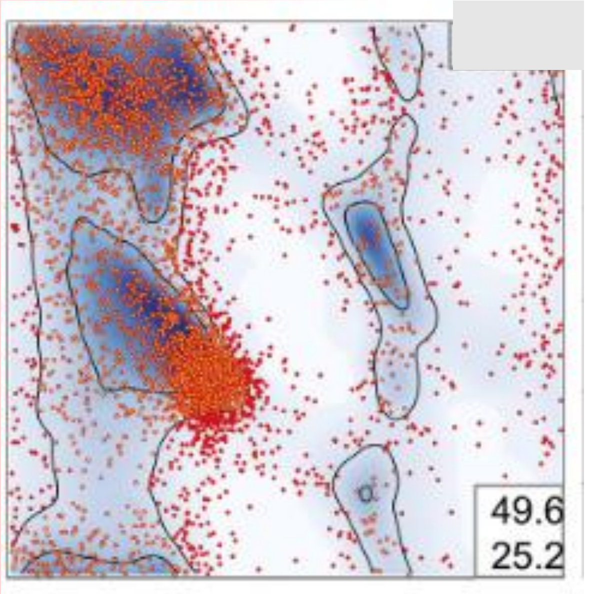
Cis-Pro : 5-6%

Cis-nonPro: ~0.05%

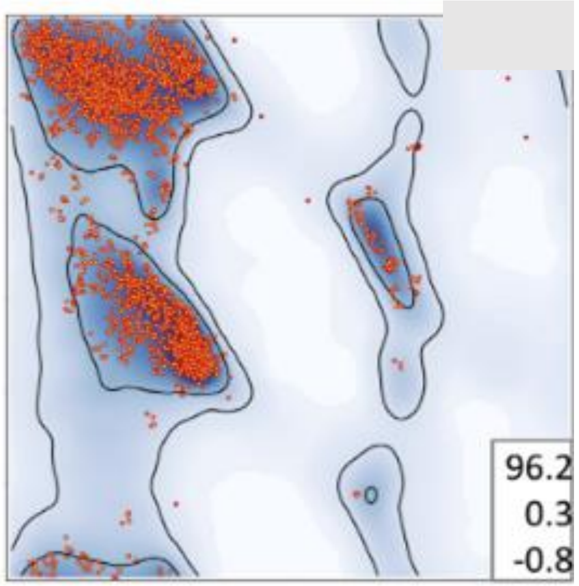
Ramachandran use / abuse

Ramachandran outliers

Ramachandran z-scores

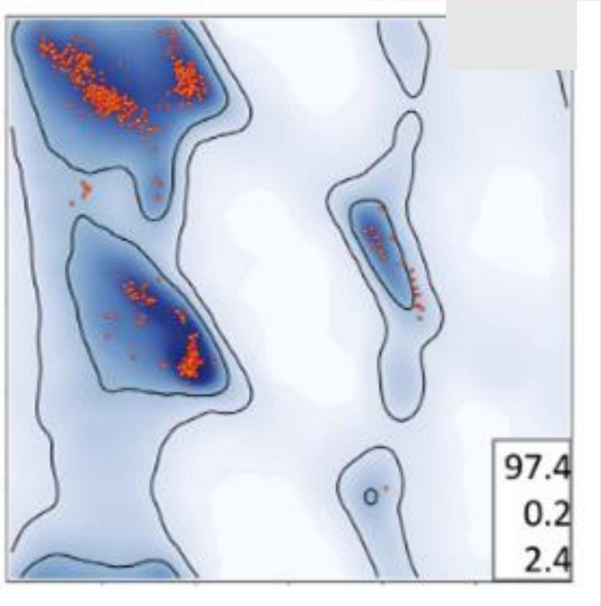


Clearly major problems!



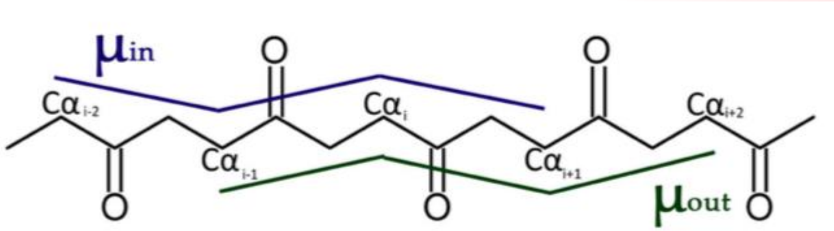
Xtal 1.5Å

Sobolev et al. Structure (2020)

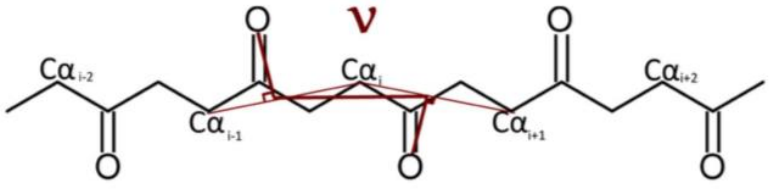


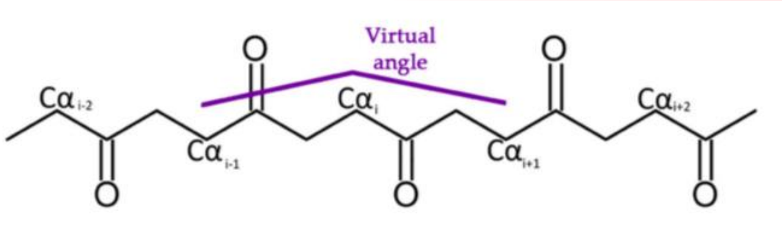
CryoEM 3.2Å
Few outliers but strange distribution.
Tight Ramachandran restraints used.
Clashscore 131.

CaBLAM

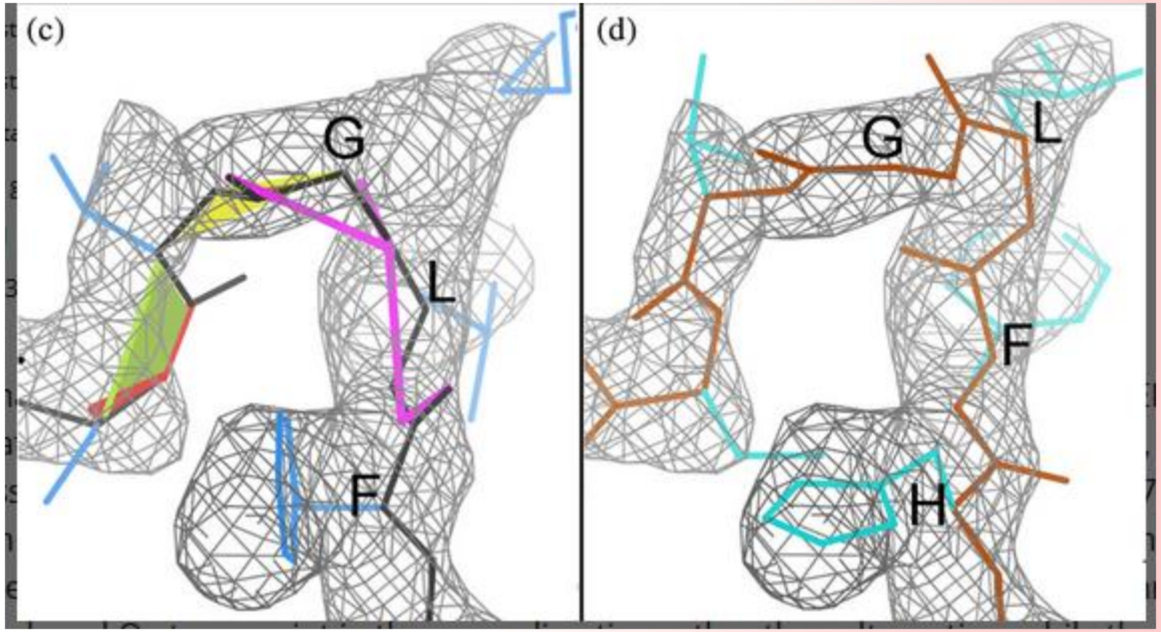


Backbone geometry check. Useful for low resolution.





Structures at 3–4 Å typical of modern cryoEM “often score much better on validation than they really are” “CaBLAM was designed to directly measure incompatibility of modeled CO directions with the local Cα trace.”
- Prisant et al. Protein Science (2020)

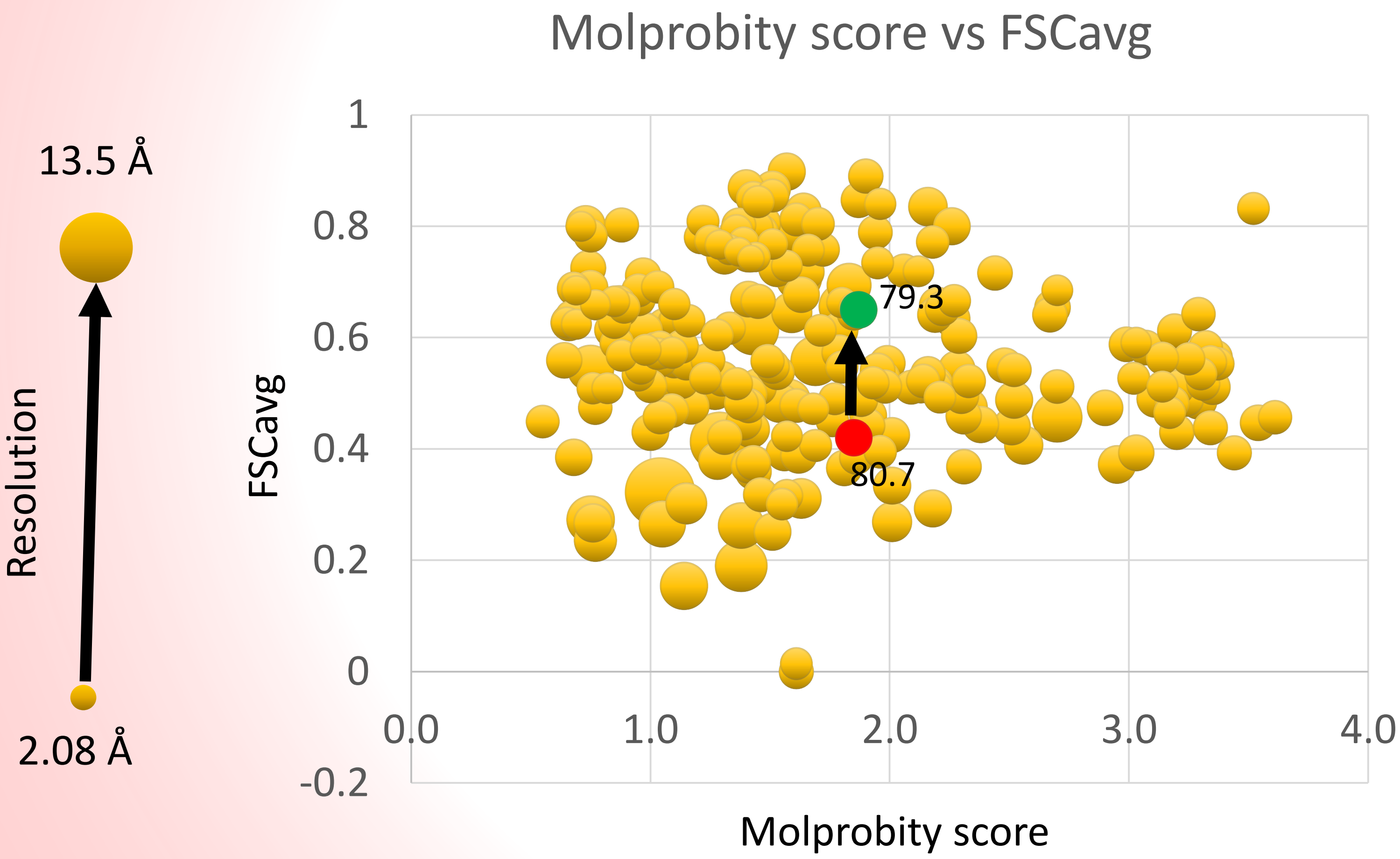


CaBLAM outliers flagged in magenta.

incorrect

correct

Perfect geometry or agree with experiment?



Be aware that models with good geometry (low Molprobity score) may not actually agree with experimental data (FSCavg).

Can be possible to improve agreement without distorting geometry with careful refinement.

Picture courtesy of Agnel Joseph

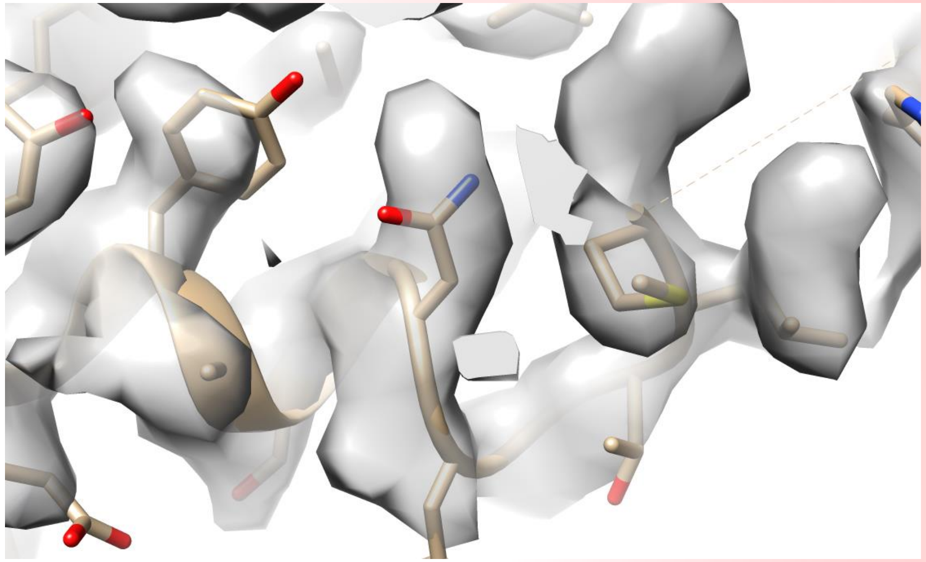
Register errors

Minor errors can be corrected, but some harder to recover from.

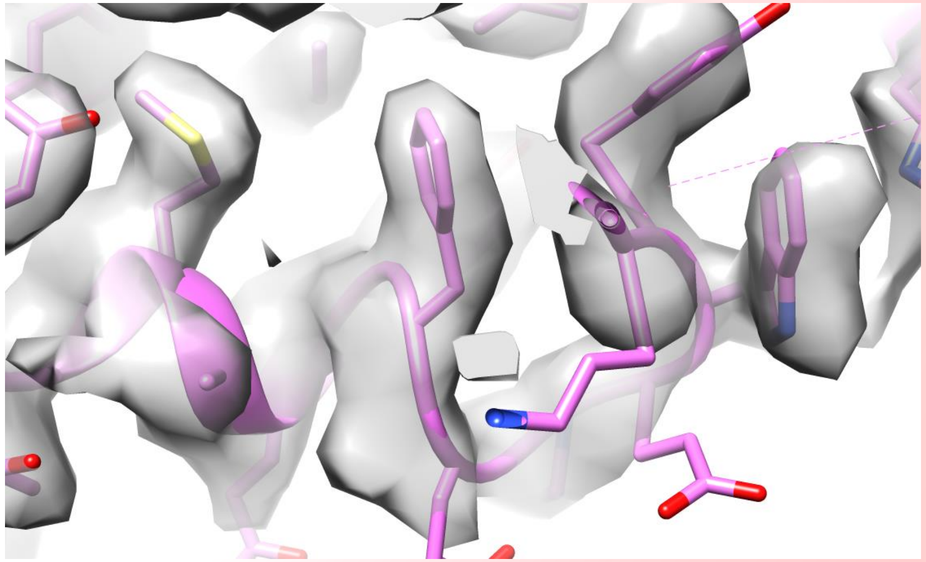
Example 2.5Å SARS-CoV2 RNA pol EMD_30210 / 7bv2. Disconnected C-terminal sequence built in wrong place (later exercise).

Standard metrics miss this, but can be identified with more sensitive tests (e.g. SMOC).

Deposited



Remodelled

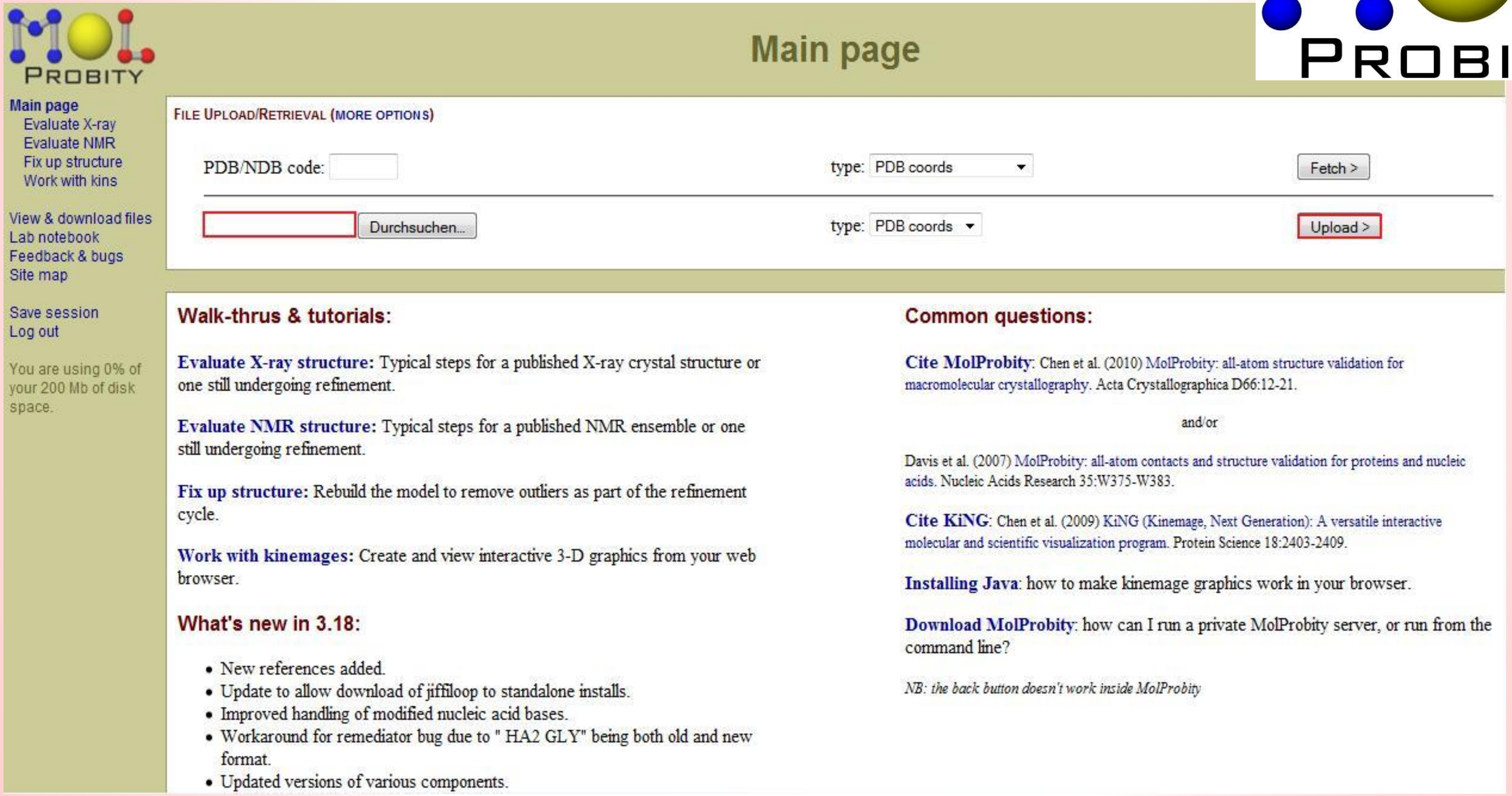


Break
Exercise 1b

Explore validation metrics for some of the structures from Exercise 1a.

Additional validation via web services

Molprobity



Main page

You are using 0% of your 200 Mb of disk space.

PDB/NDB code:

Walk-thrus & tutorials:

Evaluate X-ray structure: Typical steps for a published X-ray crystal structure or one still undergoing refinement.

Evaluate NMR structure: Typical steps for a published NMR ensemble or one still undergoing refinement.

Fix up structure: Rebuild the model to remove outliers as part of the refinement cycle.

Work with kinemages: Create and view interactive 3-D graphics from your web browser.

What's new in 3.18:

- New references added.
- Update to allow download of jiffiloop to standalone installs.
- Improved handling of modified nucleic acid bases.
- Workaround for remediator bug due to " HA2 GLY" being both old and new format.
- Updated versions of various components.

Common questions:

Cite MolProbity: Chen et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica D66:12-21.

and/or

Davis et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Research 35:W375-W383.

Cite KiNG: Chen et al. (2009) KiNG (Kinemage, Next Generation): A versatile interactive molecular and scientific visualization program. Protein Science 18:2403-2409.

Installing Java: how to make kinemage graphics work in your browser.

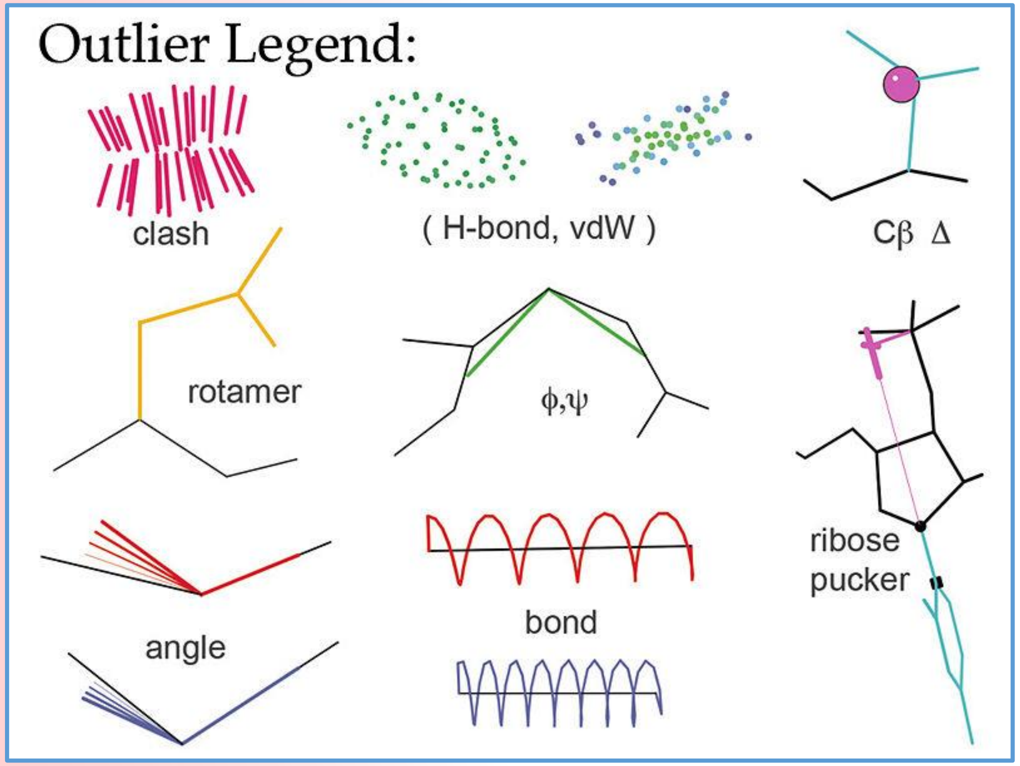
Download MolProbity: how can I run a private MolProbity server, or run from the command line?

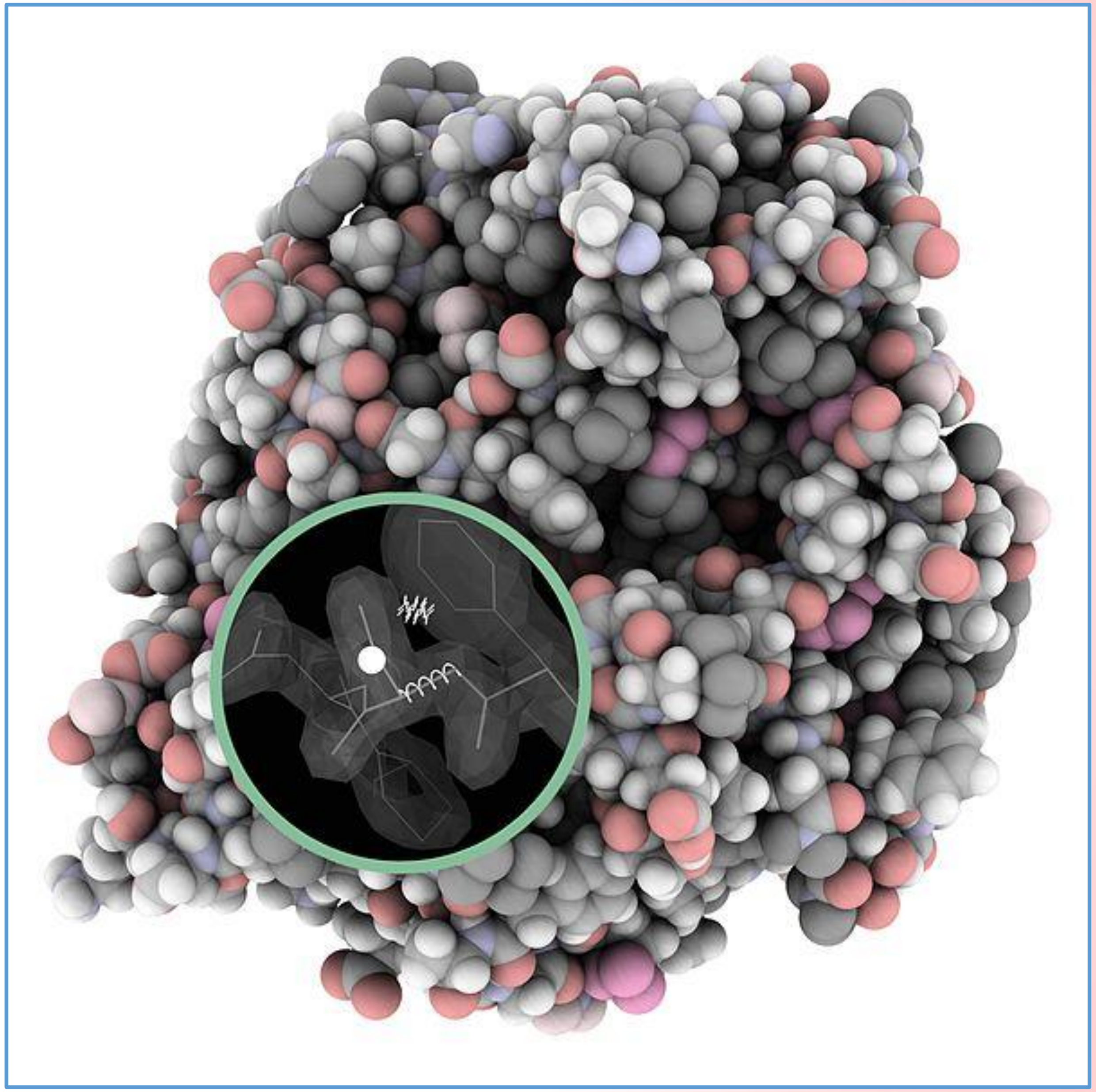
NB: the back button doesn't work inside MolProbity

http://molprobity.biochem.duke.edu/

Clashscore, all atoms	Number of steric overlaps bigger than 0.4 Å per 1000 atoms.
Poor rotamers	Percent of rotamers which are very rare.
Ramachandran outliers	Percent of residues with improbable backbone geometry.
Ramachandran favored	Percent of residues with probable backbone geometry.
C_β deviations	Percent of residues where C_β is not in the expected position.
MolProbity score	Mixed criterion consisting of Ramachandran plot, clashes and rotamers. It should ideally be higher than 90th percentile. **"Percentile":** how many structures of comparable resolution in the PDB below that level.
Residues with bad bonds	Percent residues with strange bonds.
Residues with bad angles	Percent residues with strange angles.

Molprobity in Coot



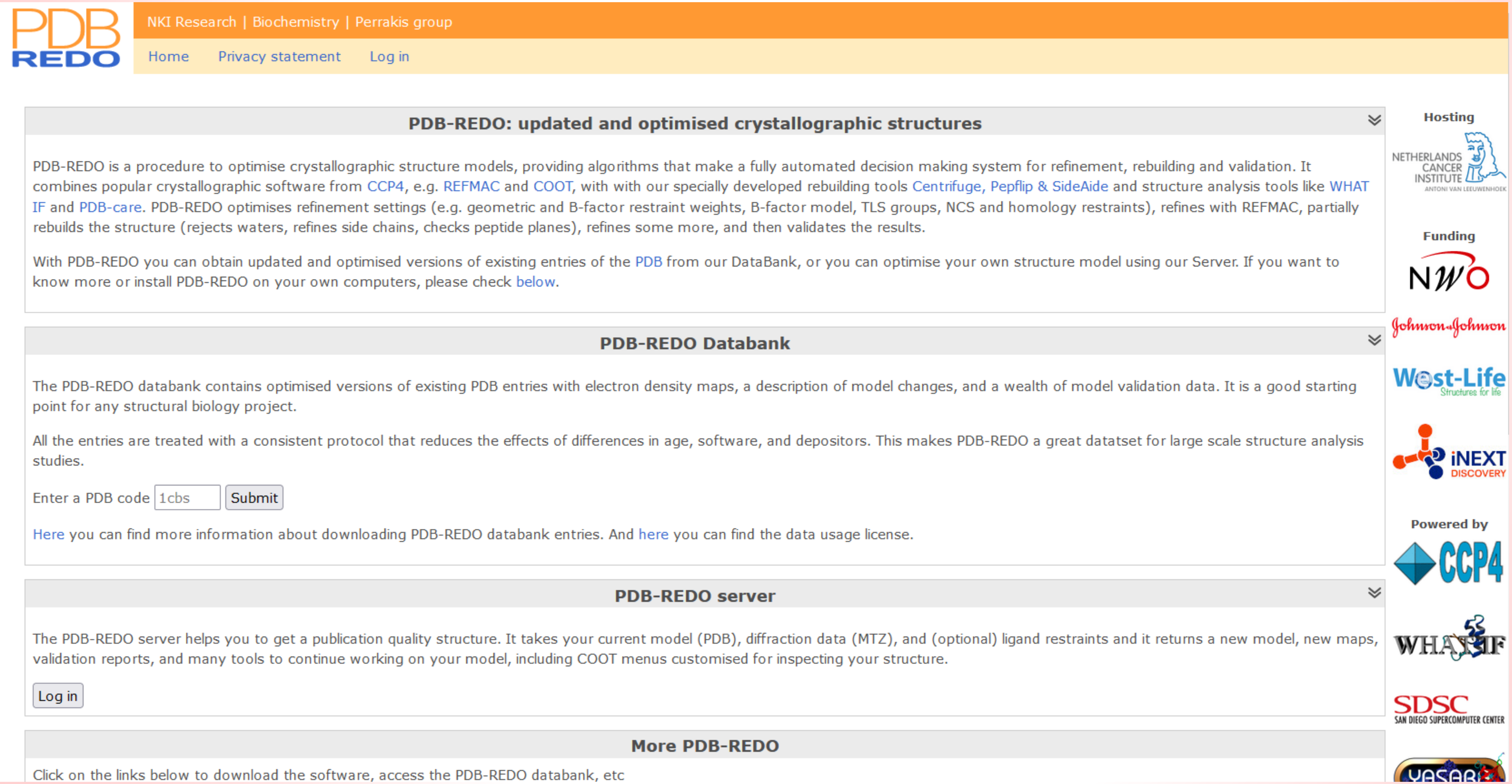


Pictures courtesy of Dave & Jane Richardson


98% MOLPROBITY SCORE
imgflip.com

PDB-REDO

https://pdb-redo.eu/



PDB-REDO: updated and optimised crystallographic structures

PDB-REDO is a procedure to optimise crystallographic structure models, providing algorithms that make a fully automated decision making system for refinement, rebuilding and validation. It combines popular crystallographic software from CCP4, e.g. REFMAC and COOT, with with our specially developed rebuilding tools Centrifuge, Pepflip & SideAide and structure analysis tools like WHAT IF and PDB-care. PDB-REDO optimises refinement settings (e.g. geometric and B-factor restraint weights, B-factor model, TLS groups, NCS and homology restraints), refines with REFMAC, partially rebuilds the structure (rejects waters, refines side chains, checks peptide planes), refines some more, and then validates the results.

With PDB-REDO you can obtain updated and optimised versions of existing entries of the PDB from our DataBank, or you can optimise your own structure model using our Server. If you want to know more or install PDB-REDO on your own computers, please check below.

PDB-REDO Databank

The PDB-REDO databank contains optimised versions of existing PDB entries with electron density maps, a description of model changes, and a wealth of model validation data. It is a good starting point for any structural biology project.

All the entries are treated with a consistent protocol that reduces the effects of differences in age, software, and depositors. This makes PDB-REDO a great datatset for large scale structure analysis studies.

Enter a PDB code 1cbs

Here you can find more information about downloading PDB-REDO databank entries. And here you can find the data usage license.

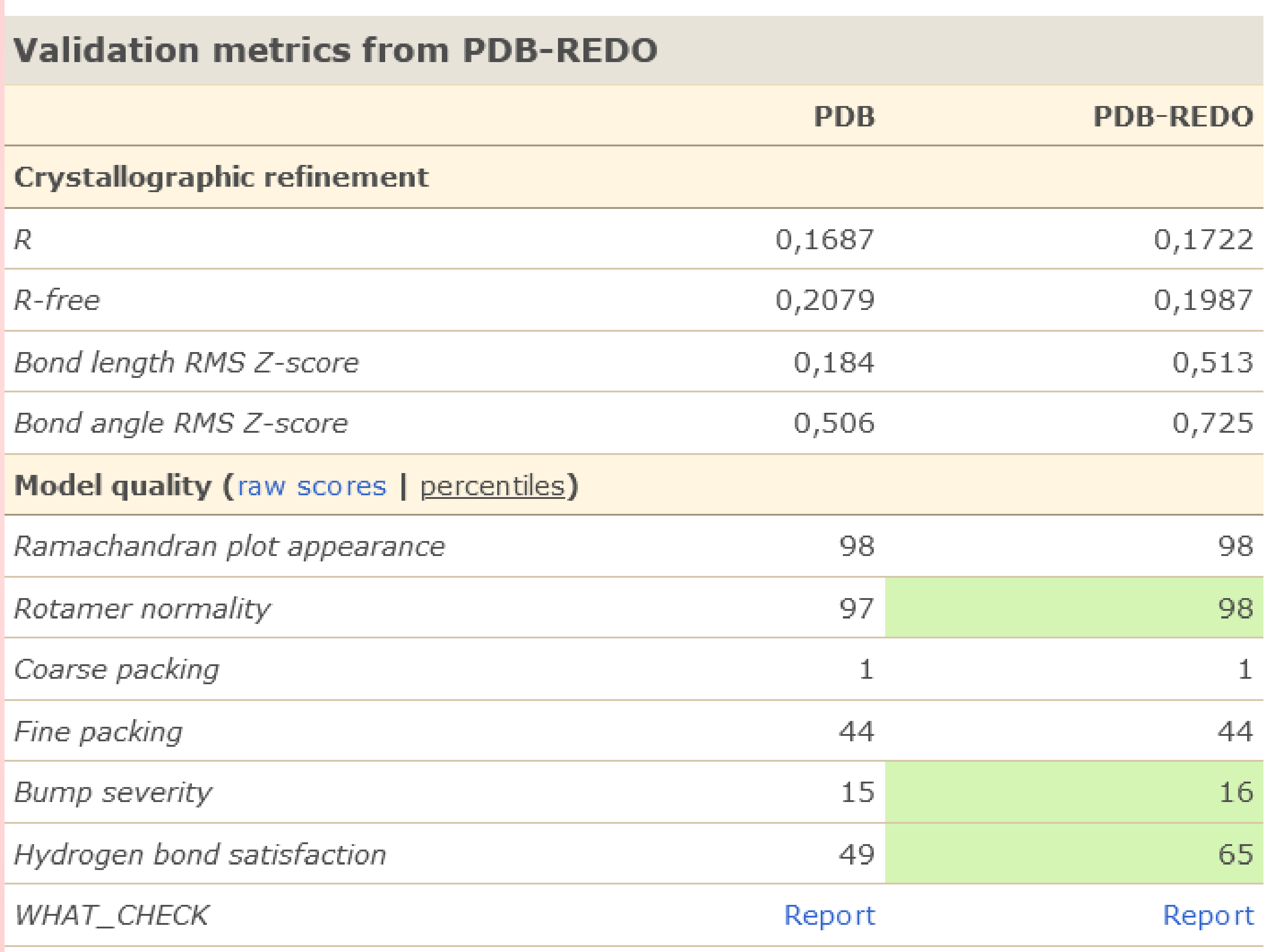
PDB-REDO server

The PDB-REDO server helps you to get a publication quality structure. It takes your current model (PDB), diffraction data (MTZ), and (optional) ligand restraints and it returns a new model, new maps, validation reports, and many tools to continue working on your model, including COOT menus customised for inspecting your structure.

More PDB-REDO

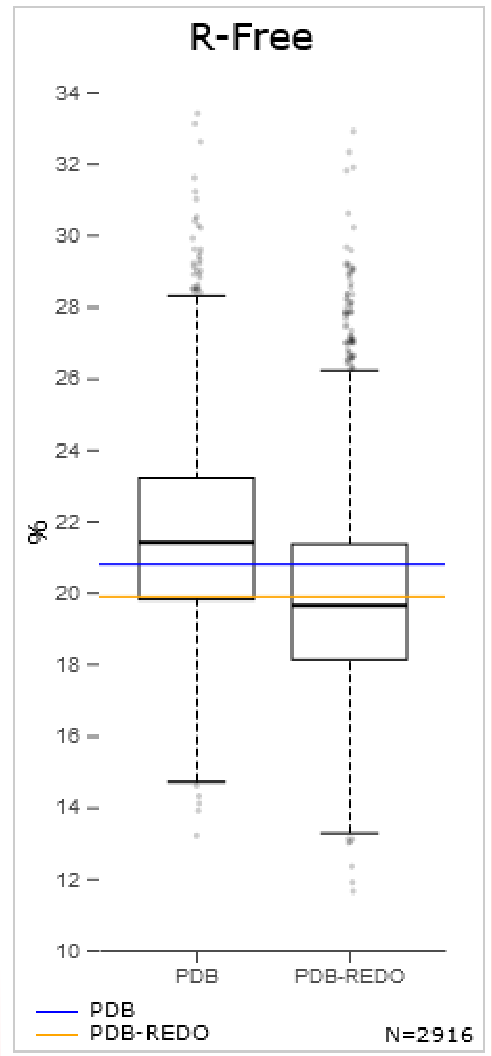
Click on the links below to download the software, access the PDB-REDO databank, etc

PDB-REDO report



Validation metrics from PDB-REDO

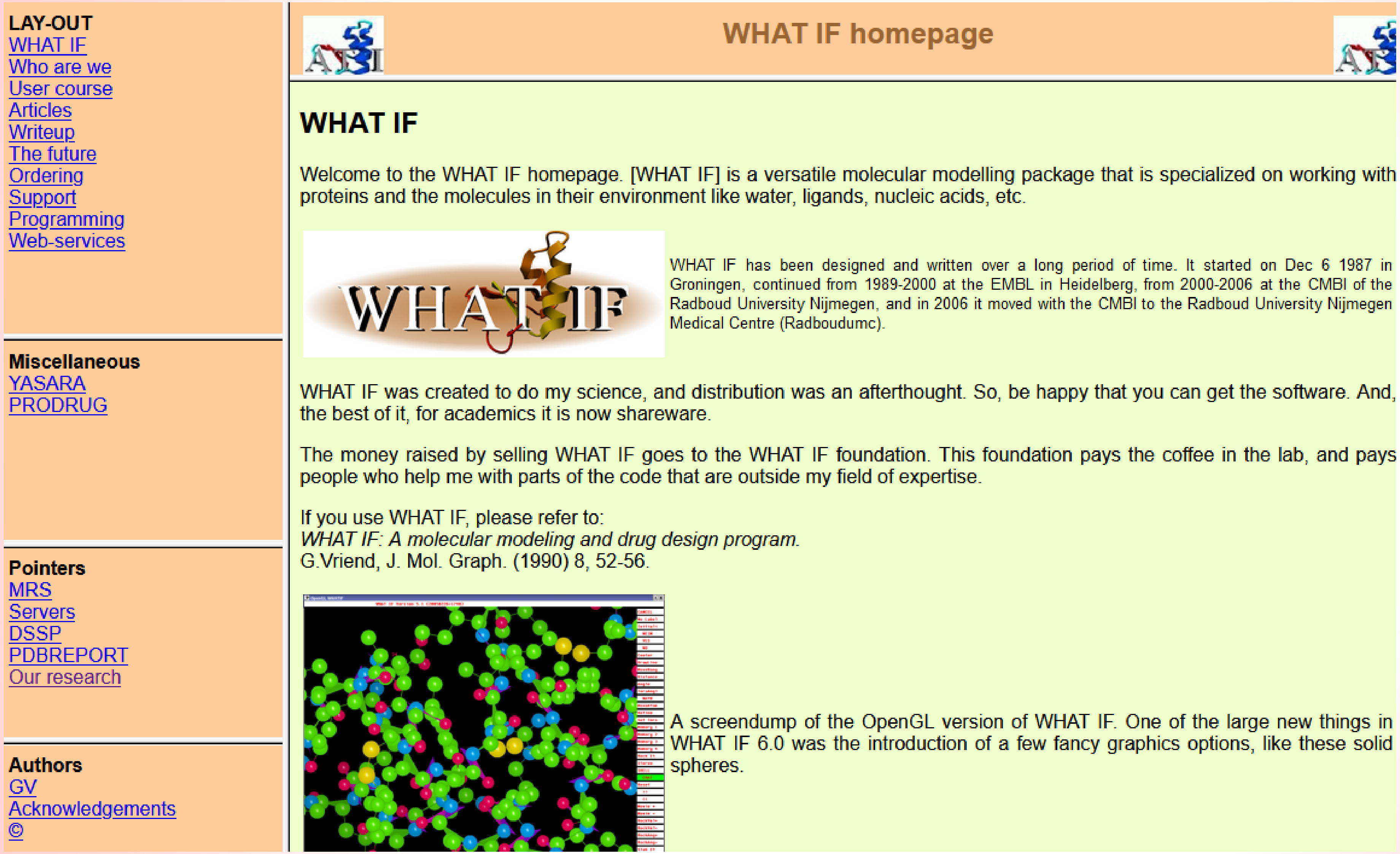
	PDB	PDB-REDO
Crystallographic refinement		
R	0,1687	0,1722
R-free	0,2079	0,1987
Bond length RMS Z-score	0,184	0,513
Bond angle RMS Z-score	0,506	0,725
Model quality (raw scores | percentiles**)**		
Ramachandran plot appearance	98	98
Rotamer normality	97	98
Coarse packing	1	1
Fine packing	44	44
Bump severity	15	16
Hydrogen bond satisfaction	49	65
WHAT_CHECK	Report	Report



What if…?

PDB REDO

WHAT IF



LAY-OUT
WHAT IF
Who are we
User course
Articles
Writeup
The future
Ordering
Support
Programming
Web-services

Miscellaneous
YASARA
PRODRUG

Pointers
MRS
Servers
DSSP
PDBREPORT
Our research

Authors
GV
Acknowledgements
©

WHAT IF homepage

WHAT IF

Welcome to the WHAT IF homepage. [WHAT IF] is a versatile molecular modelling package that is specialized on working with proteins and the molecules in their environment like water, ligands, nucleic acids, etc.

WHAT IF has been designed and written over a long period of time. It started on Dec 6 1987 in Groningen, continued from 1989-2000 at the EMBL in Heidelberg, from 2000-2006 at the CMBI of the Radboud University Nijmegen, and in 2006 it moved with the CMBI to the Radboud University Nijmegen Medical Centre (Radboudumc).

WHAT IF was created to do my science, and distribution was an afterthought. So, be happy that you can get the software. And, the best of it, for academics it is now shareware.

The money raised by selling WHAT IF goes to the WHAT IF foundation. This foundation pays the coffee in the lab, and pays people who help me with parts of the code that are outside my field of expertise.

If you use WHAT IF, please refer to:
WHAT IF: A molecular modeling and drug design program.
G.Vriend, J. Mol. Graph. (1990) 8, 52-56.

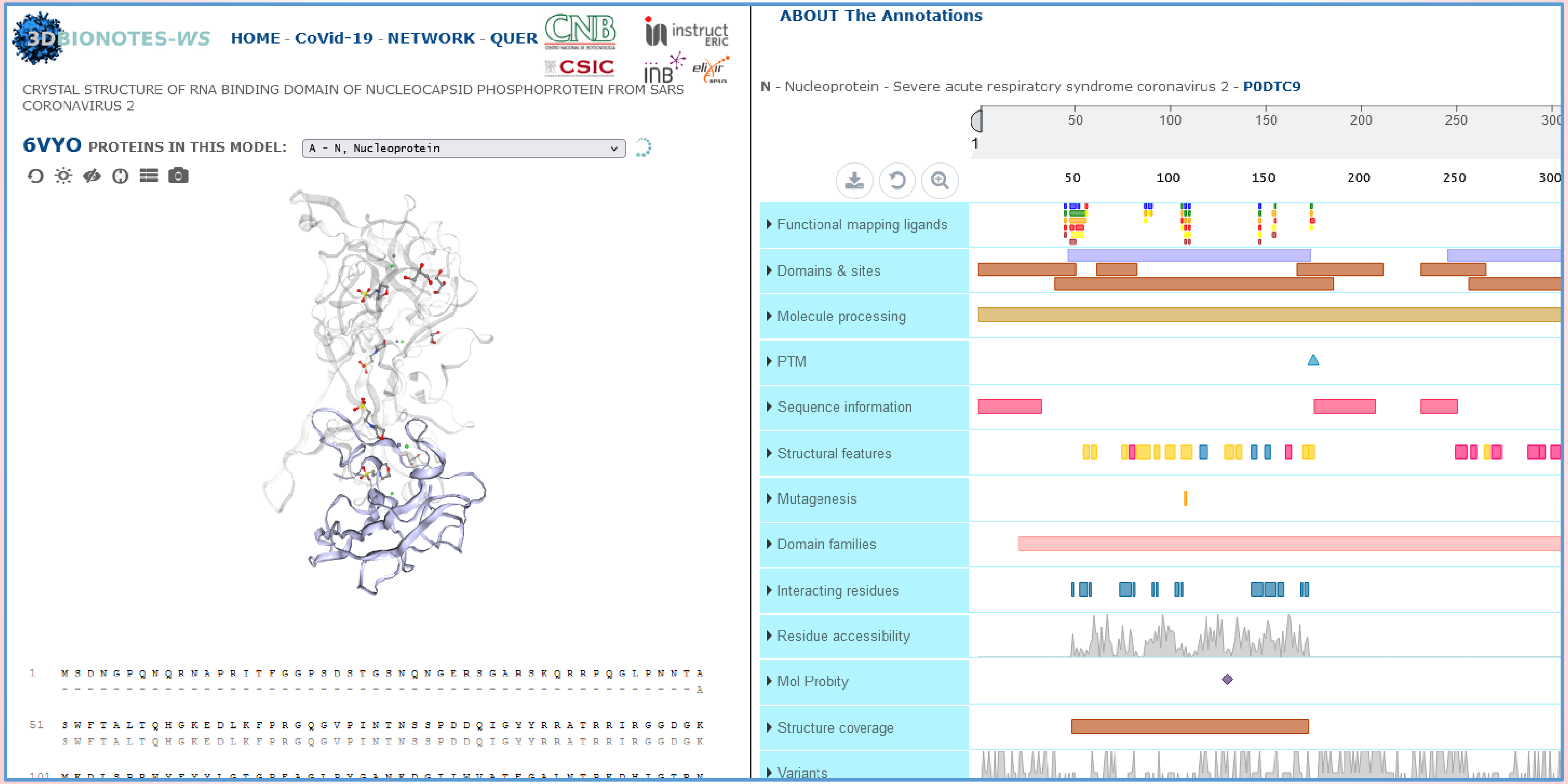
A screendump of the OpenGL version of WHAT IF. One of the large new things in WHAT IF 6.0 was the introduction of a few fancy graphics options, like these solid spheres.

https://swift.cmbi.umcn.nl/whatif/

WHAT CHECK

- Nomenclature
- Packing, symmetry axis collision, molecular contacts, accssibility
- Bond lengths & angles, planarity, chirality
- Peptide flips, omega angles
- Proline puckering, cysteine bridged
- Water / ion switches
- Average B factors
- Average structures from NMR

3D Bionotes



CRYSTAL STRUCTURE OF RNA BINDING DOMAIN OF NUCLEOCAPSID PHOSPHOPROTEIN FROM SARS CORONAVIRUS 2

N - Nucleoprotein - Severe acute respiratory syndrome coronavirus 2 - **P0DTC9**

```
1    M S D N G P Q N Q R N A P R I T F G G P S D S T G S N Q N G E R S G A R S K Q R R P Q G L P N N T A
     - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - A

51   S W F T A L T Q H G K E D L K F P R G Q G V P I N T N S S P D D Q I G Y Y R R A T R R I R G G D G K
     S W F T A L T Q H G K E D L K F P R G Q G V P I N T N S S P D D Q I G Y Y R R A T R R I R G G D G K
```

http://3dbionotes-ws.cnb.csic.es/ws

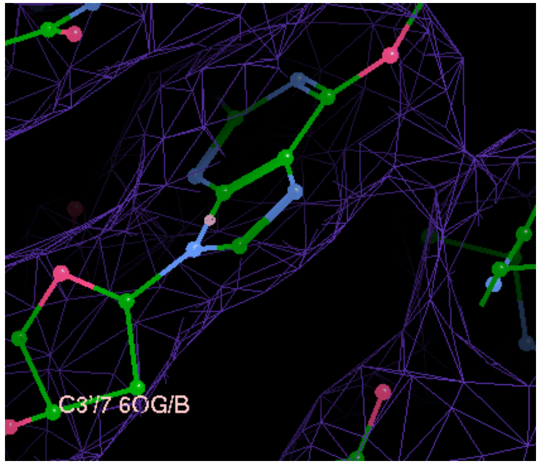
Looking at the density

Experimental maps

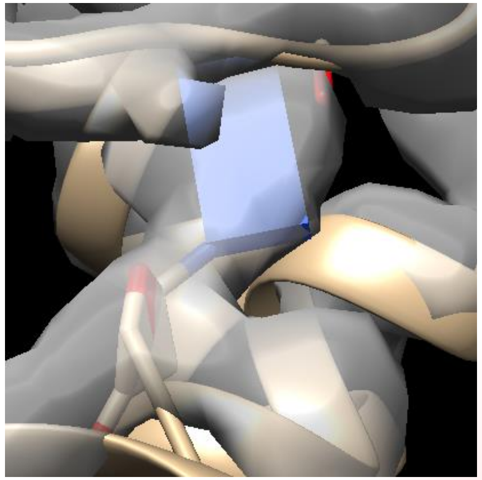
- Atomic model is only an interpretation of the experimental data.
- Should compare the model to the experimental map:
  - Xtal = electron density. CryoEM = electrostatic potential.

[In fact, maps are also not the original experimental data (the result of phasing in MX or reconstruction in cryoEM) but nevertheless better quality indicators than the atomic model.]

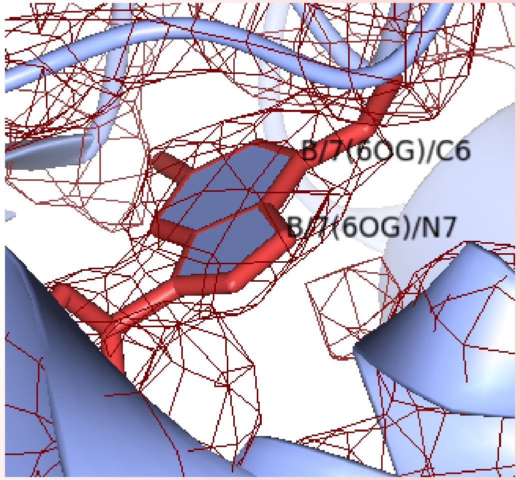
Metrics in Validation Report, but no substitute for looking!



Coot



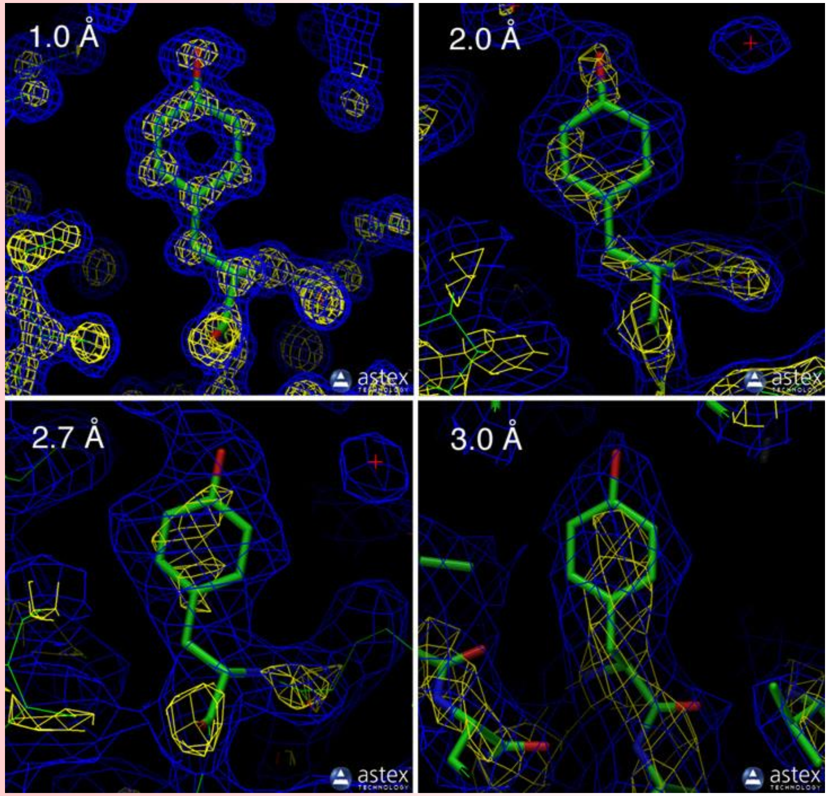
Chimera



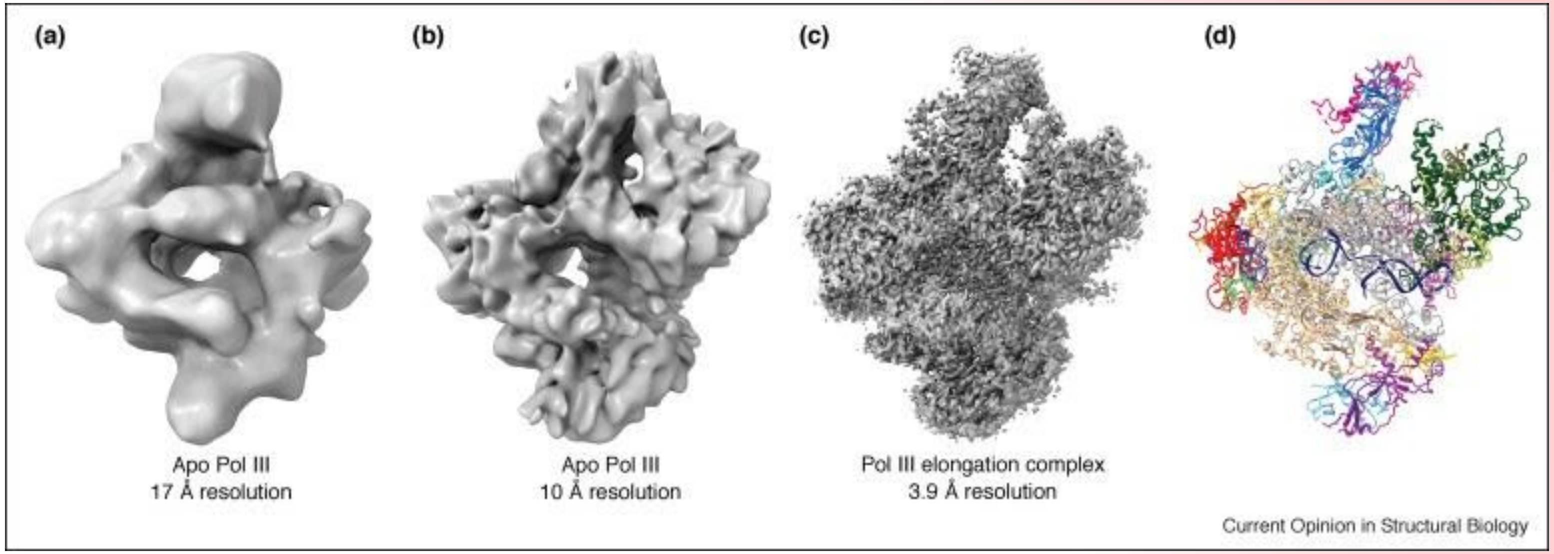
CCP4mg

Effect of resolution

Crystallographic maps



CryoEM maps



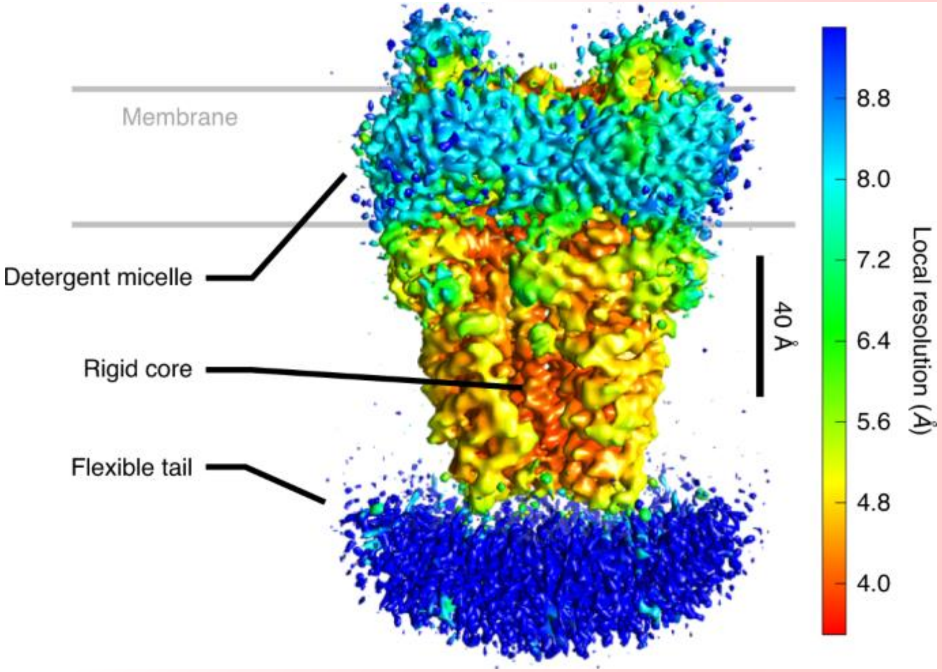
(a) Apo Pol III
17 Å resolution

(b) Apo Pol III
10 Å resolution

(c) Pol III elongation complex
3.9 Å resolution

Current Opinion in Structural Biology

- Local resolution can vary significantly across cryoEM structures.
- Not obvious from headline figure!

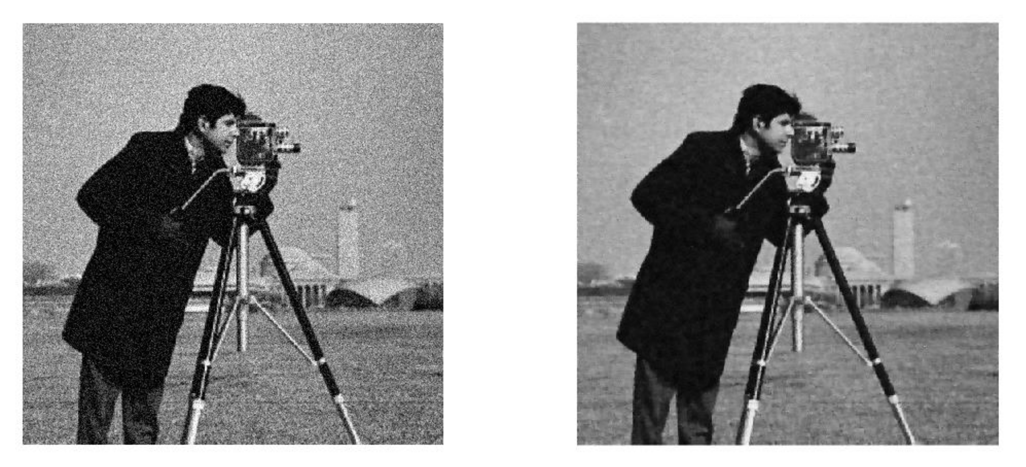


Map processing

Many standard techniques from image processing.
Can help model building / validation.

Sharpen – emphasise high resolution features

Blur – emphasise low resolution features

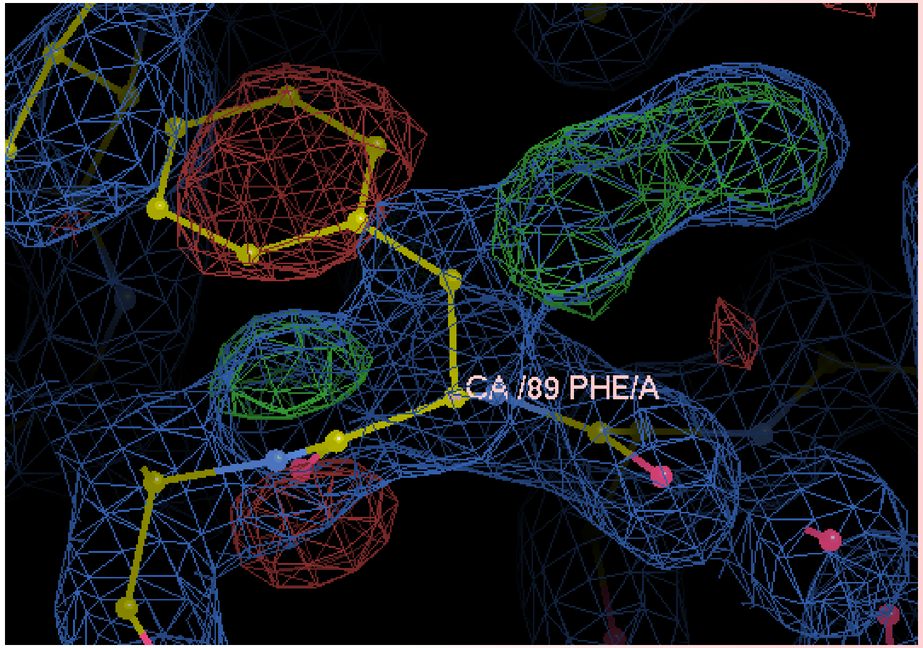


De-noise – separate out signal at all spatial frequencies

Crystallographic maps

- Phases are from atomic model
    - Main problem is bias, not noise
- Local sharpening / blurring via atomic B factors

- Main map – typically view at 1σ
- Difference maps:
    - Positive ⇒ unmodelled density
    - Negative ⇒ incorrect model

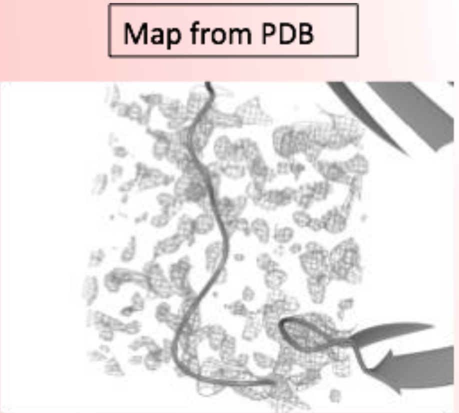


Coot tutorial

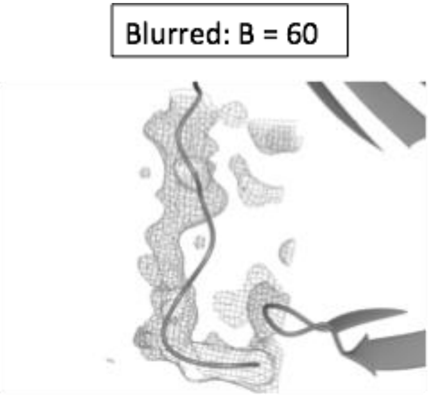
CryoEM maps

- Measured phases.
- Adjust high and low resolution components, locally or globally

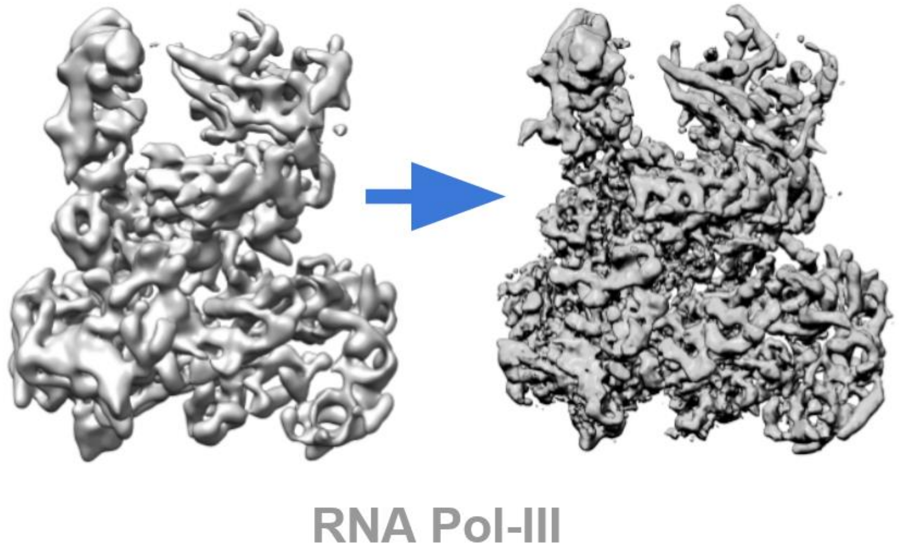
Map blurring



Original map

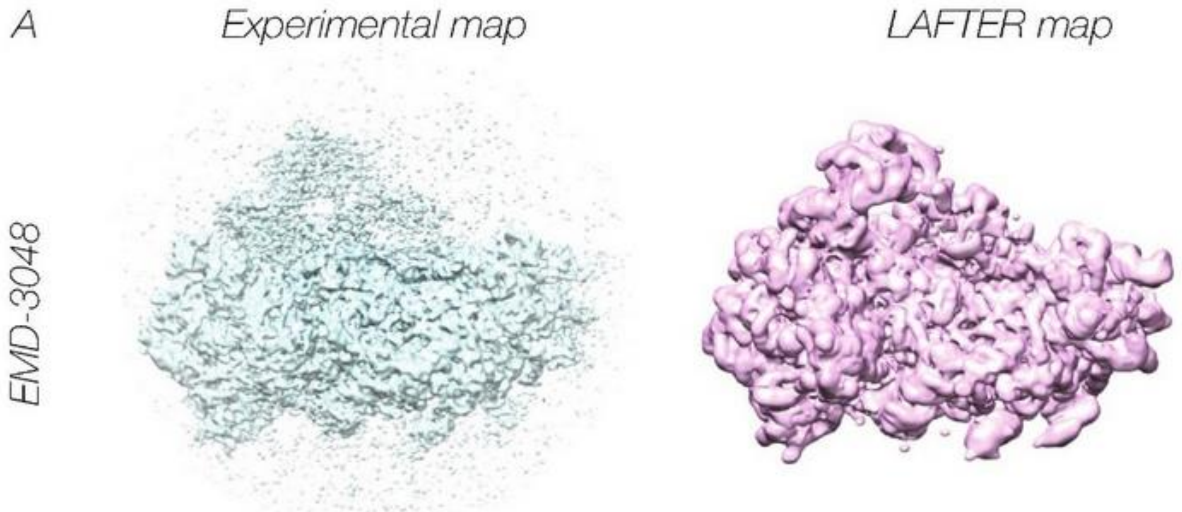


Apply global B factor



LocScale: local map sharpening

Jakobi et al. eLife (2017) 6:e27131



LAFTER: Map denoising

Looking at the density
Exercise 2

Recommended: use local installation of Coot
Backup: web-based Uglymol viewer

1t38 – xtal – protein-DNA complex at 3.2Å
7bv2 – cryoEM – SARS-Cov-2 RdRp complex at 2.5Å

Discussion