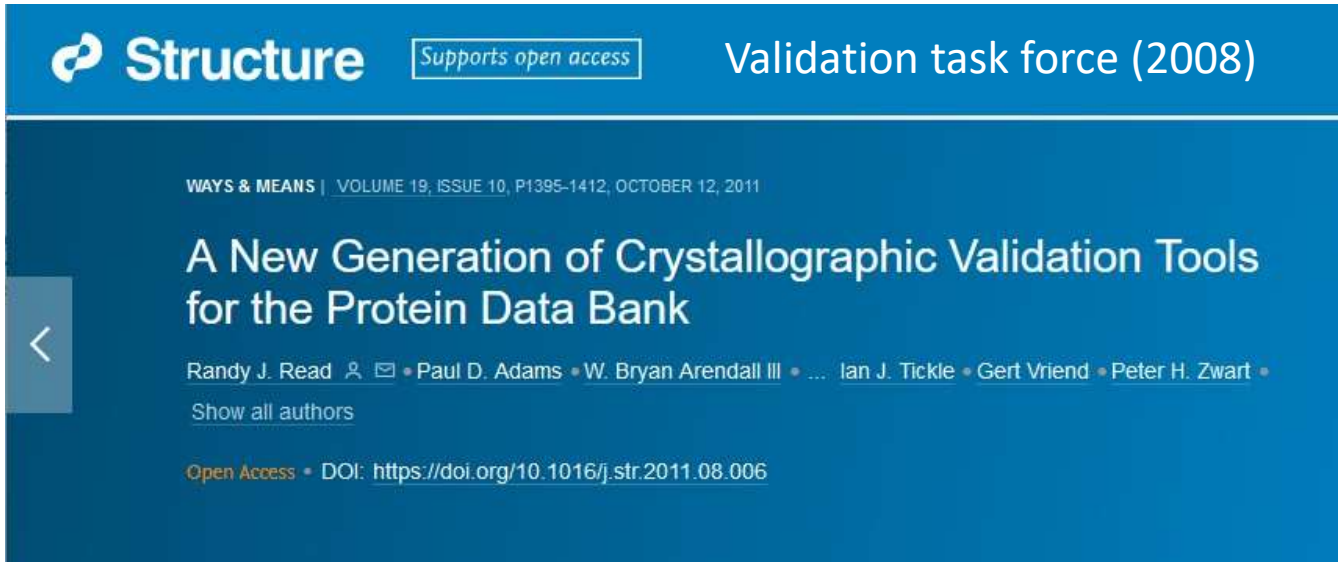
Data Validation & Deposition

Colin McVey

13th July 2022

Validation

X-ray data to final model



https://www.rcsb.org/docs/additional-resources/structure-validation-and-quality

https://validate.wwpdb.org

Modern day Crystallography

The crystallographer ideally has knowledge of:

- history of the protein sample
- the biology of the system
- Solid foundation in chemistry & physics

Understands :-

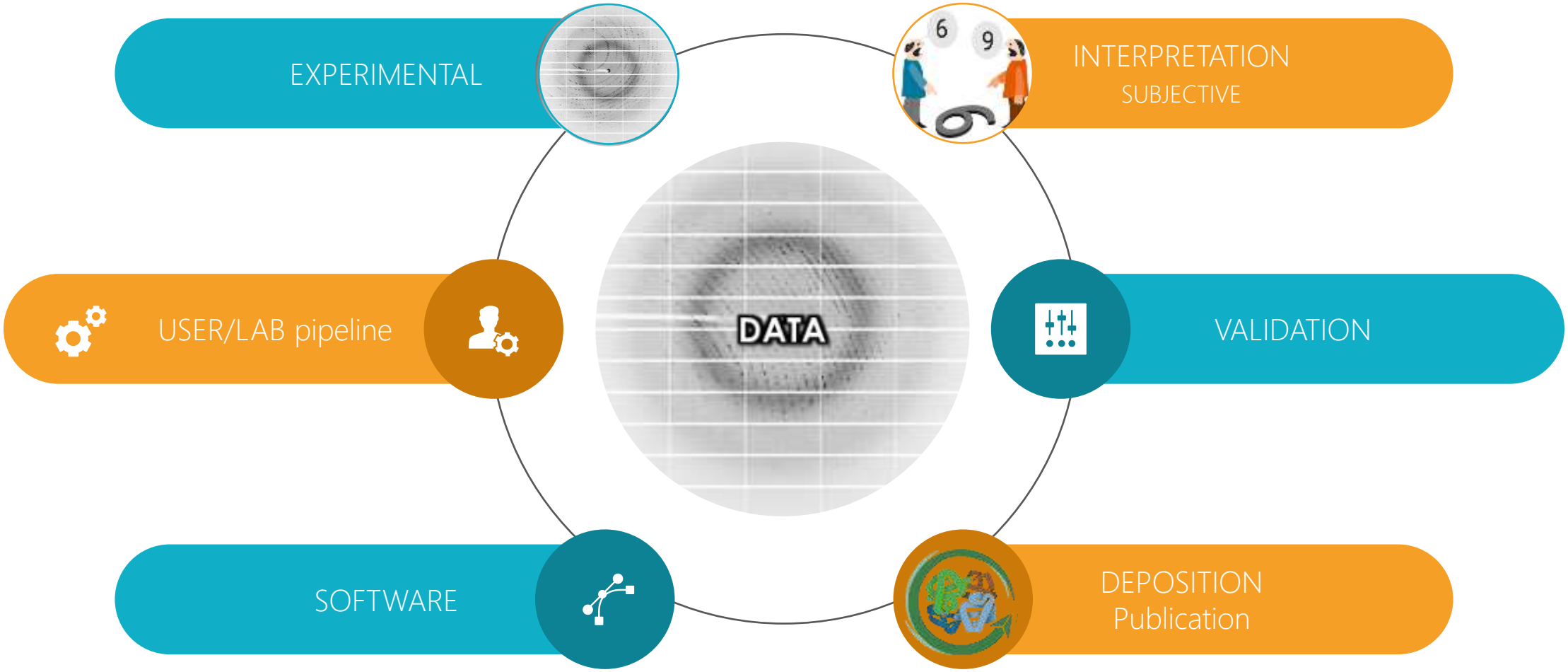
- data collection and processing, space groups, (t)NCS, twinning
- refinement process and software
- Experience in map interpretation

Read the relevant literature and remember

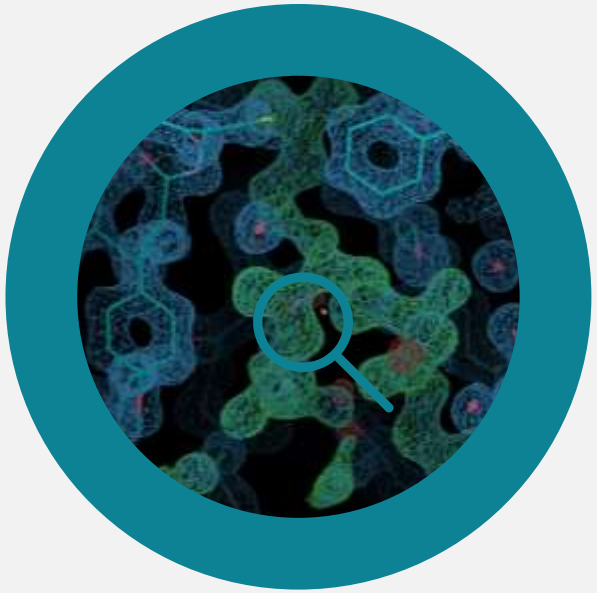
Modern day Structural biologist

"Jack of all trades master of none"

Data Analysis



DATA Scrutiny



PHASE DETERMINATION

IMPROVEMENT

Anomalous scatter location
Molecular Replacement
Impact of Alphafold

MODEL REFINEMENT

PHENIX-REFMAC-BUSTER

Different software & approaches
Use different formula to compute R-factor, which makes comparison of R-factors between programs nonsensical.

MAP INTERPRETATION

SUBJECTIVE

Resolution, R-factor, CC* (overfitting), $CC_{1/2}$ (resolution cutoff)
Free R-factor → indicate overfitting

Why are mistakes made

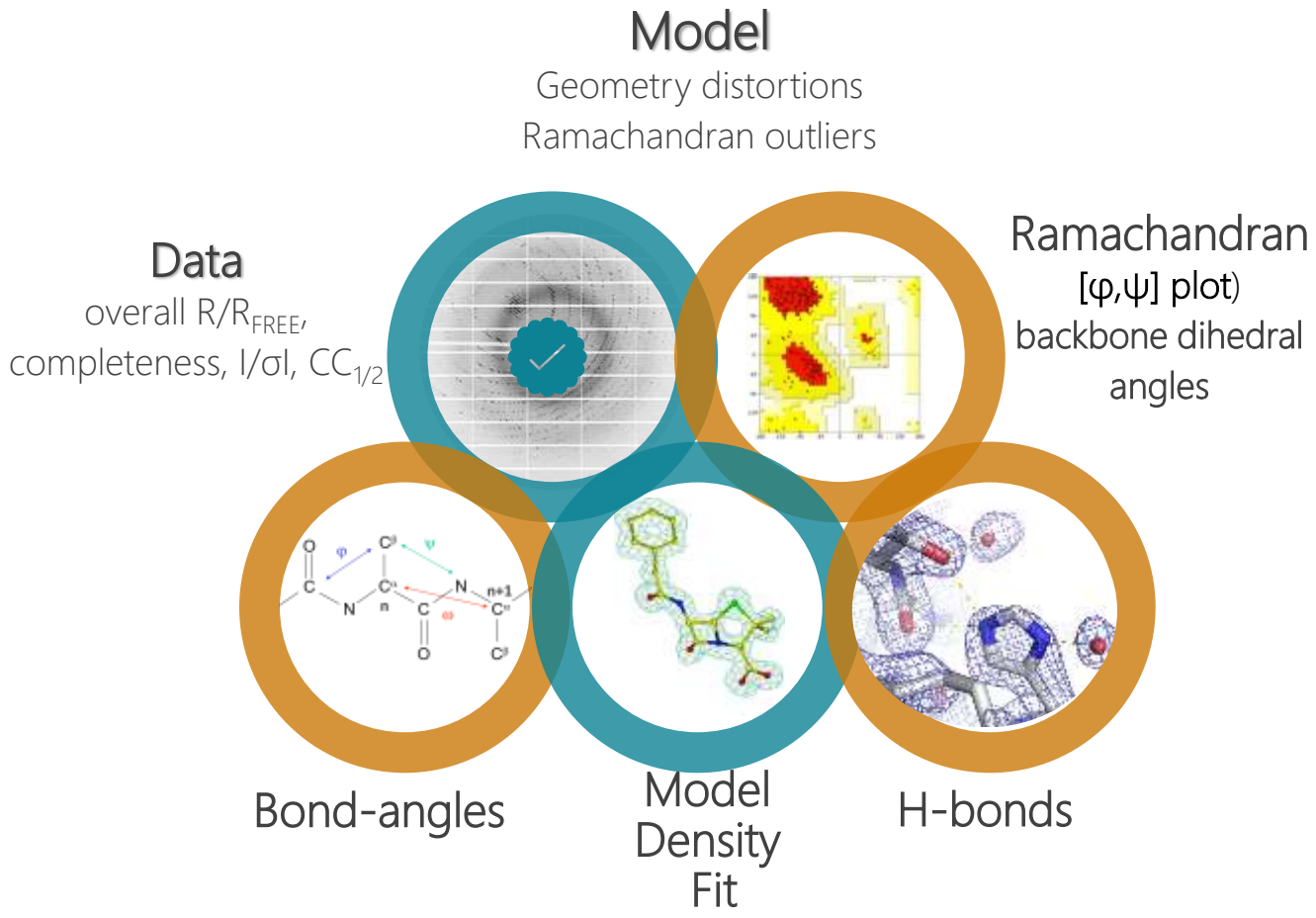
Limitations of the X-ray data

- Incomplete
- Weak
- Resolution
- Space and time averaged
- Phase errors

The human element

- Subjectivity involved in map interpretation and refinement (even at 1.26 Å resolution!)
- Inexperienced people do the work, use of black box approach
- Not everybody is a good chemist
- Even experienced people make mistakes

What to Validate



When to Validate

Throughout

What to Validate

- A good model makes sense
  - Physical
    - Close packing, no bad contacts, crystal solvent content
  - Chemical
    - correct local geometry, bonds, angles, etc
  - Crystallographic
    - R-factors, reasonable pattern of B-factor variation, model to density map fit, Space Group symmetry
  - Statistically
    - No under-modelling (under-refinement), no over-fitting (over-modelling)
    - Model global quality stats are expected to be in agreement with corresponding values found in similar structures (pdb slider plots)

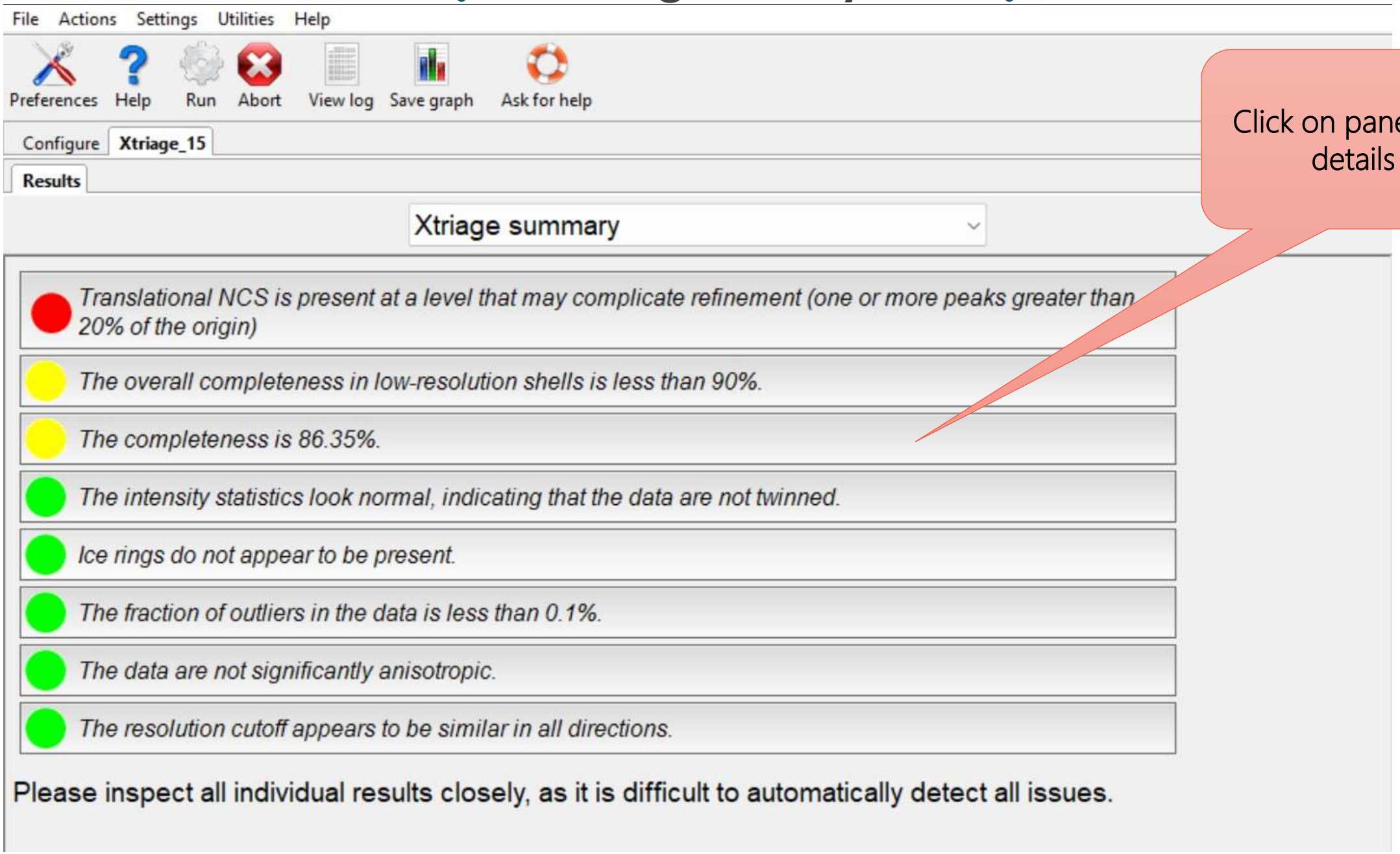
Why Validate

- Problems detected in initial stages - save a lot of time later
- Subjectivity in data and map interpretation
  - User interpret the maps: experience, skills acquired
  - pressure to meet deadlines
- Software development – developer errors
  - Programs may contain bugs
- Subjectivity in model building and refinement
  - Model parameterization, target weights, starting points
  - Quality of data - impact interpretation
- Deception or honest user error

Validation Software

- **XTRIAGE** (Zwart et. al, 2005) – included in Phenix
- Xtriage is a set of routines that tries to assess the quality of the experimental data and establish knowledge of certain X-ray data pathologies
  - **https://phenix-online.org/documentation/reference/xtriage.html**
- **PROCHECK** (Laskowski et. al, 1998) – included in CCP4i2
  - checks the stereochemical quality of a protein structure
  - 'Validation and Deposition' module
- **SFCHECK** (Vagin et al., 1999) – included in CCP4i2
  - Assess the agreement between the atomic model and X-ray data
  - 'Validation and Deposition' module
- **MOLPROBITY** (Chen et al, 2010) – bundled in CCP4i2 & PHENIX→COOT
  - http://molprobity.biochem.duke.edu/

Xtriage Analysis



Xtriage Data Analysis

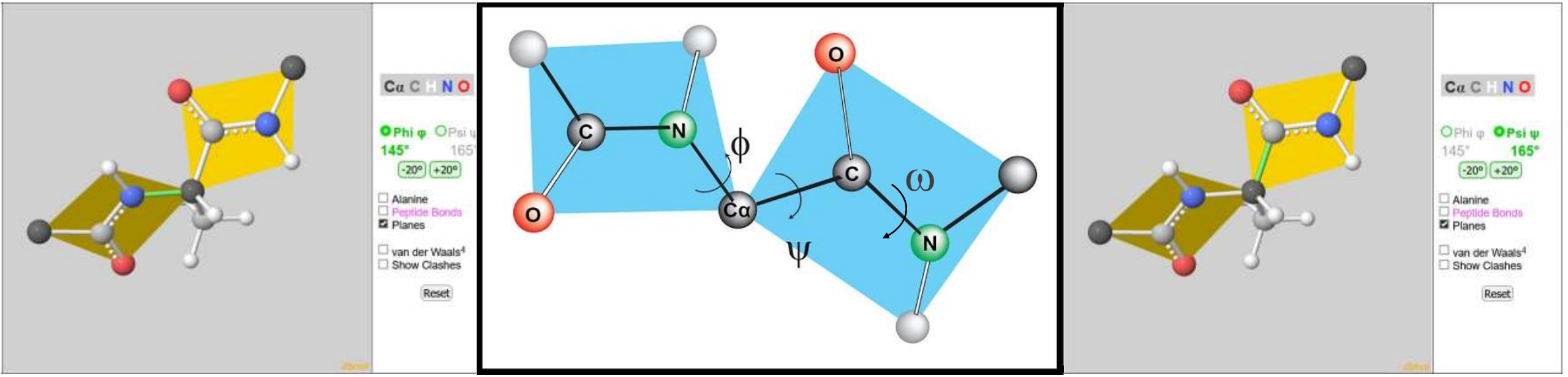
- Matthews coefficient probabilities
- Completeness by resolution
- Wilson plot analysis
- Detection of translational NCS (tNCS)
- Analysis of systematic absences and combination of tNCS with current space group
- Anomalous signal from measurability analysis
- Symmetry and twinning analyses
- Alternative point - group symmetry (can be detected on the basis of an R-value analyses)

Reading: https://phenix-online.org/documentation/reference/xtriage.html

Model Validation Steps

- Main-chain conformation: **Ramachandran plot** outliers?
- Side-chain conformations: unusual **rotamers**?
- **Stereochemistry**: how much does the model geometry **deviate** from the "ideal" values used in the restrained refinement?
- **Peptide dihedral angle** deviations from planarity: large deviations (> 15°) unlikely.
- ***cis*-peptide bonds**: unlikely except if involving Proline residues.
- Protein model **completeness**: are there missing atoms/residues, ligands? Where?

Main-chain Geometry

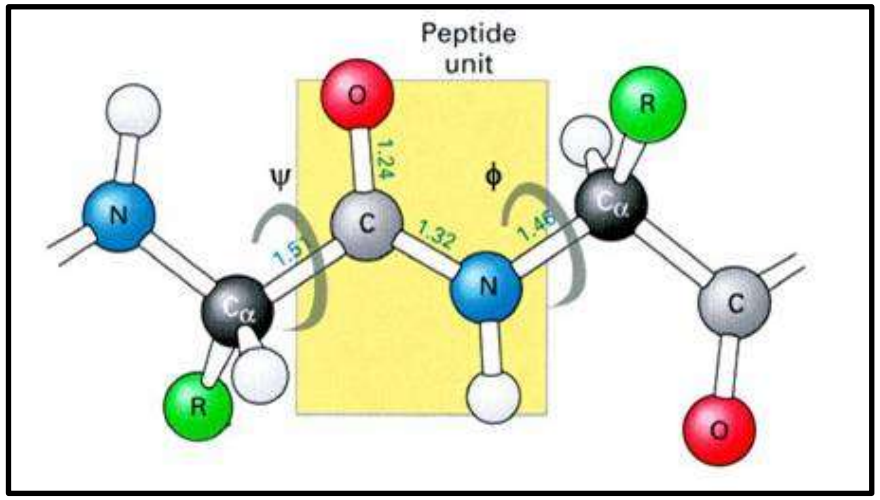


Phi (Φ; C, N, C_α, C) and psi (Ψ; N, C_α, C, N) are on either side of the C_α atom and omega (ω; C_α, C, N, C_α) describes the angle of the peptide bond. Not all combinations of **φ,ψ dihedral angles** are energetically favourable for non-glycine amino acid residues.

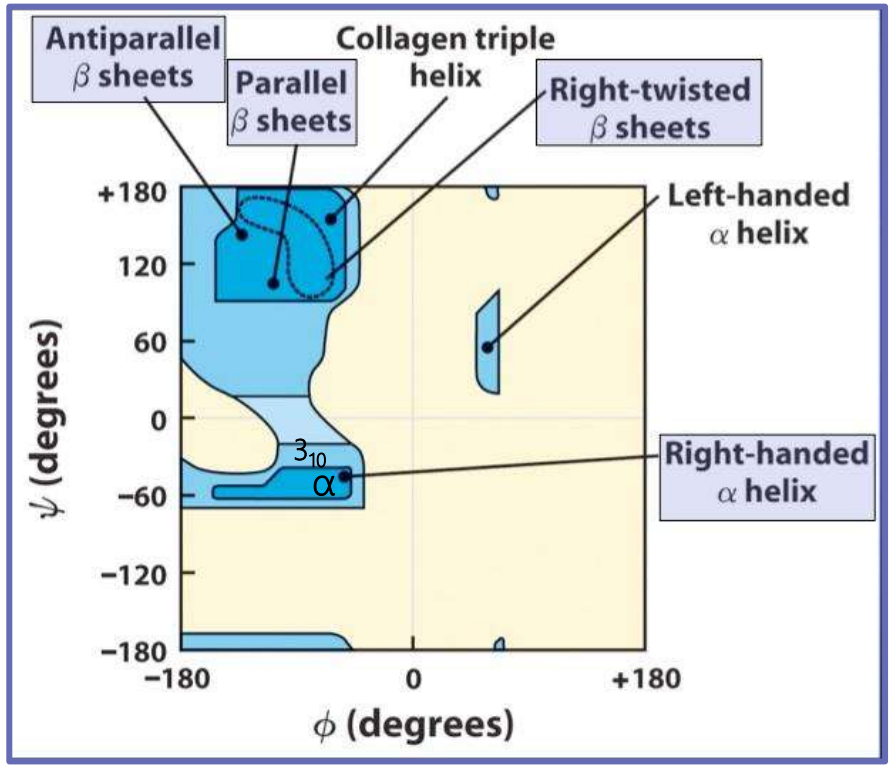
The ω (**peptide bond**) dihedral angles are **planar** due to the partial double bond character of the C-N bond. Trans peptide bonds (ω = 180°) are more favourable. **Cis** (ω = 0°) peptide bonds are sometimes observed for **Proline residues**.

Main-chain Geometry

- Ramachandran plot
  - φ and ψ angles
  - Compare to the PDB



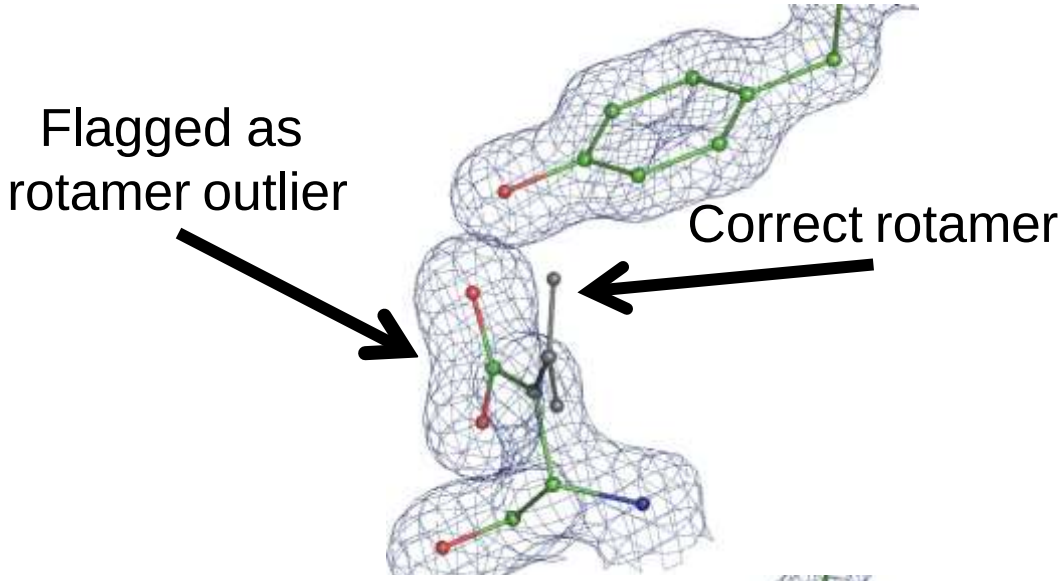
- Different implementations
  - MolProbity and CCP4/Phenix: preferred, okay, outlier → integrated with coot
  - Good for finding specific problems

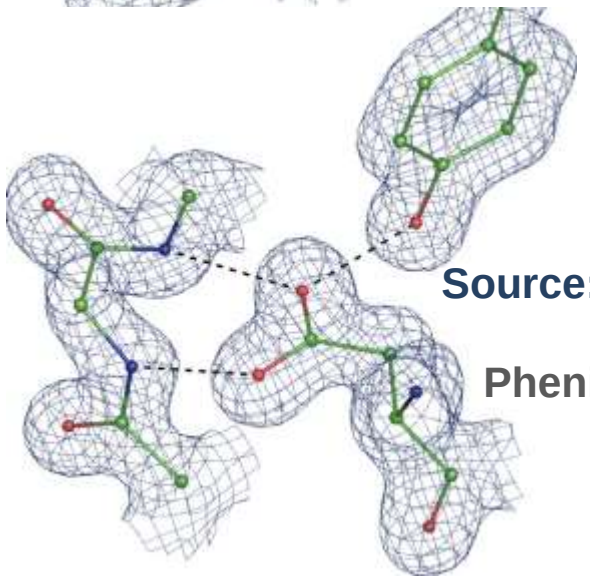


Side-chain Geometry

Side chain torsion angles

- Steric hindrance causes discrete rotamers
- ~ 15% of the side chains cannot be represented by known library of rotamer conformations
- Outliers are typically fitting errors…
- …or false positives, outliers have to be explained





Source:Pavel Afonine

Phenix developer

Sidechain Dilemma

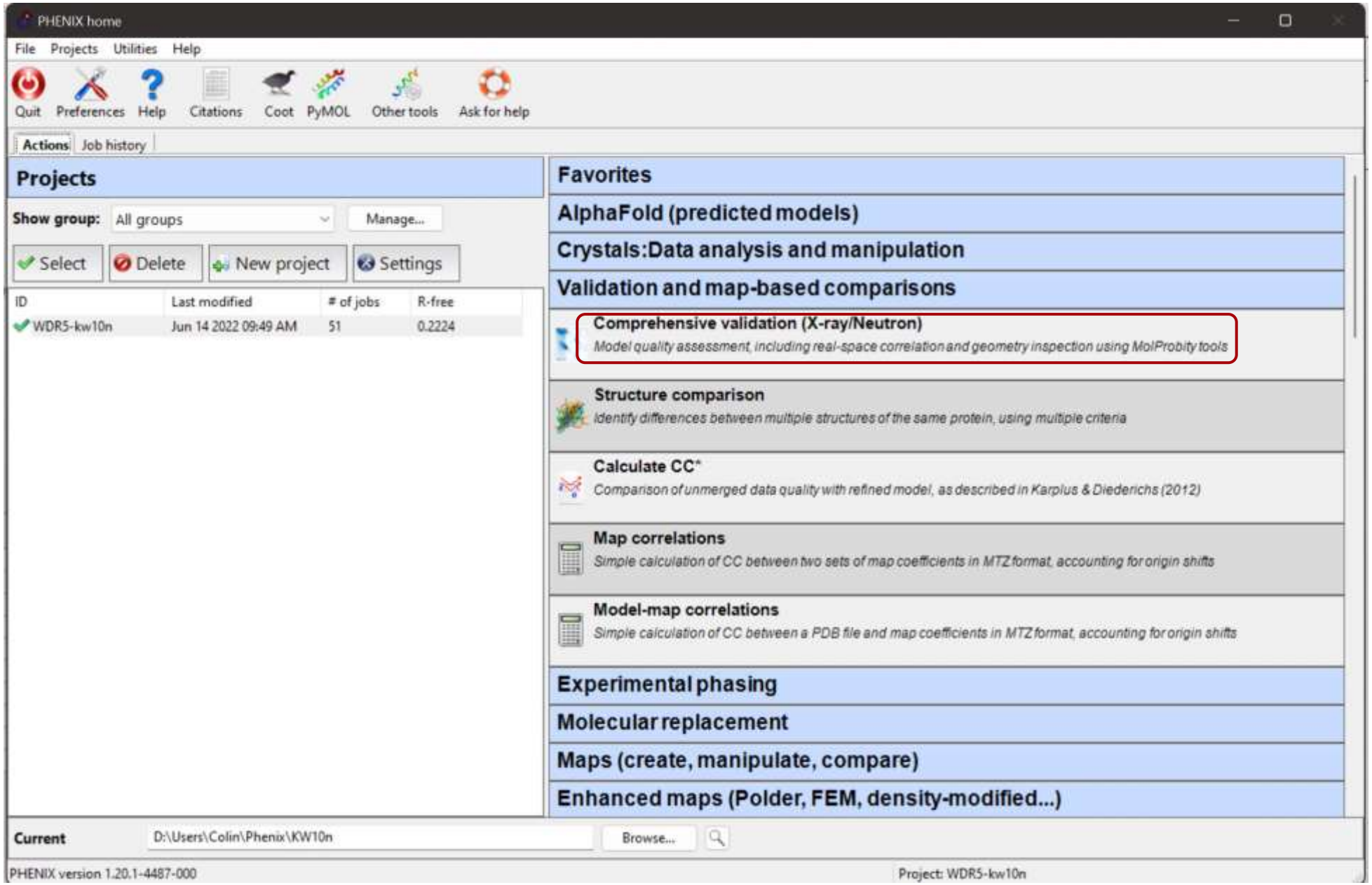
I can't see density for a sidechain beyond the C-beta atom. How should I model it?

Opinions differ in the crystallography community on the proper approach to side chain disorder

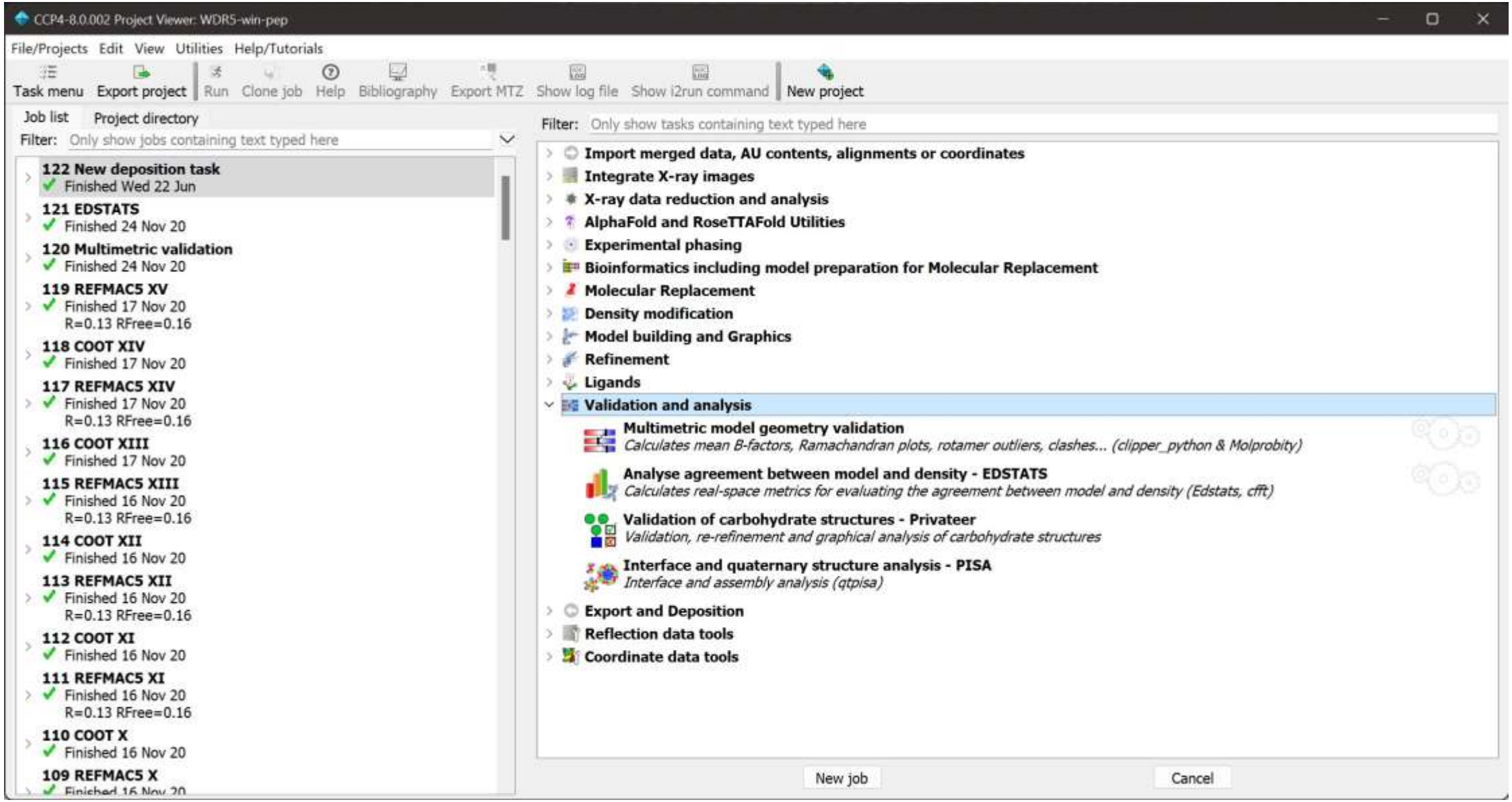
Significant discussions on PHENIX and CCP4 mailing lists:

- Delete all atoms not visible in density, but leave the residue name alone.
- Pick an appropriate rotamer, and let the B-factors rise to account for disorder. This avoids truncated sidechains that may be mistaken for other residues, and is more realistic when interpreting surface electrostatics.
- A third approach, setting the occupancy of missing atoms to zero but leaving them in the model, this is strongly disfavoured, as the resulting positions and B-factors are entirely theoretical (not immediately obvious to a non crystallographers).

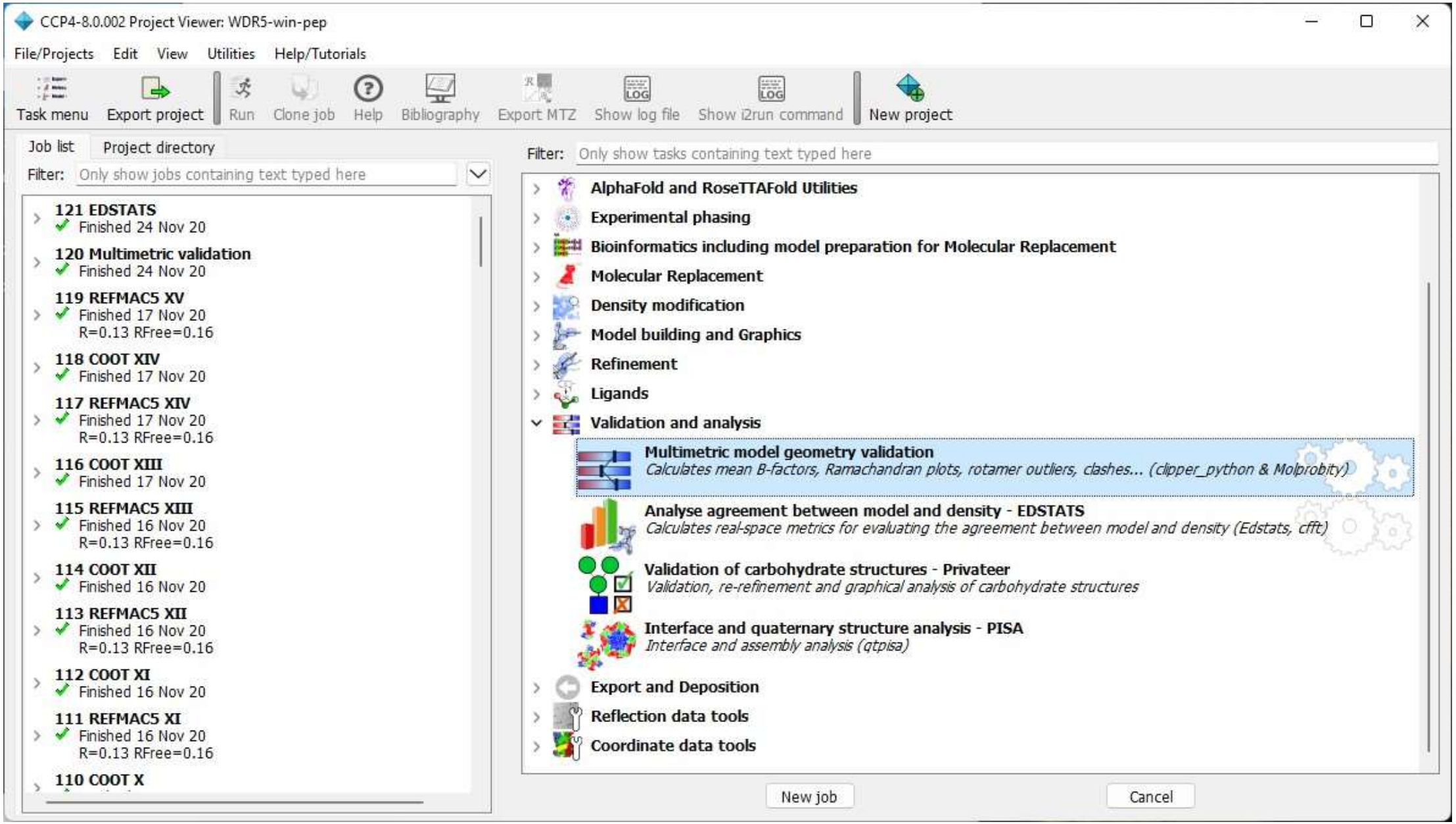
Validation Phenix



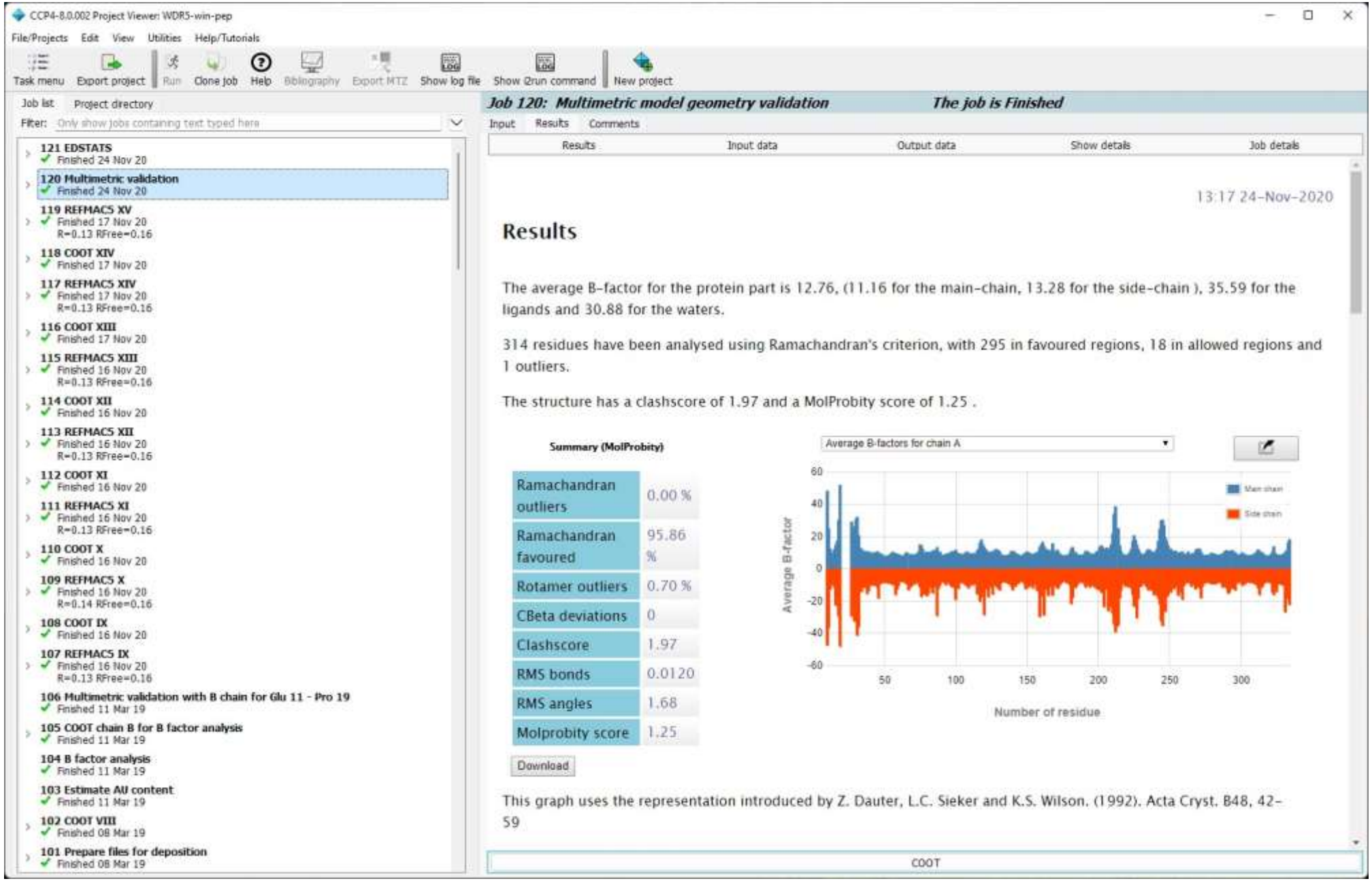
Validation CCP4i2



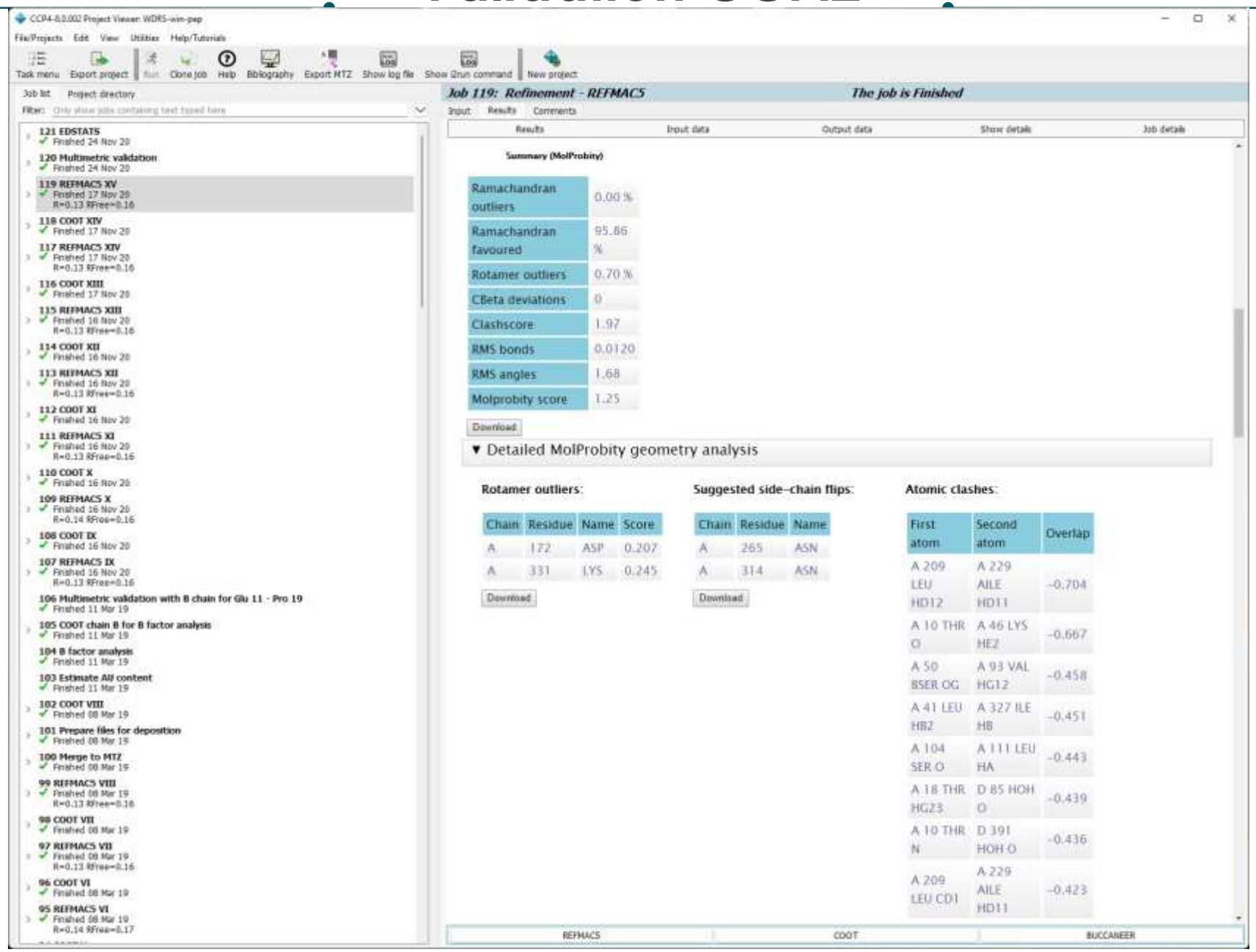
Validation CCPi2



Validation CCPi2

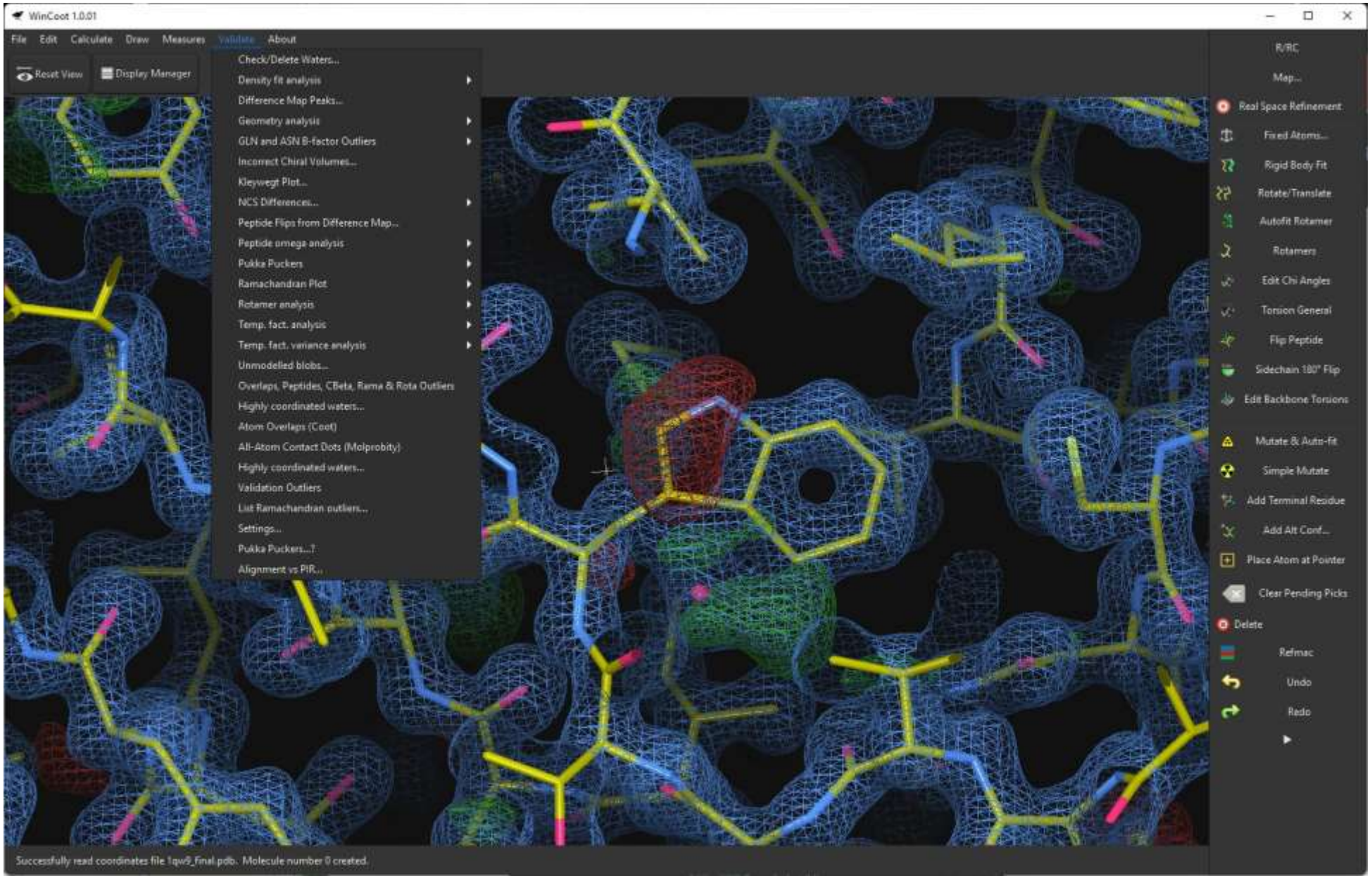


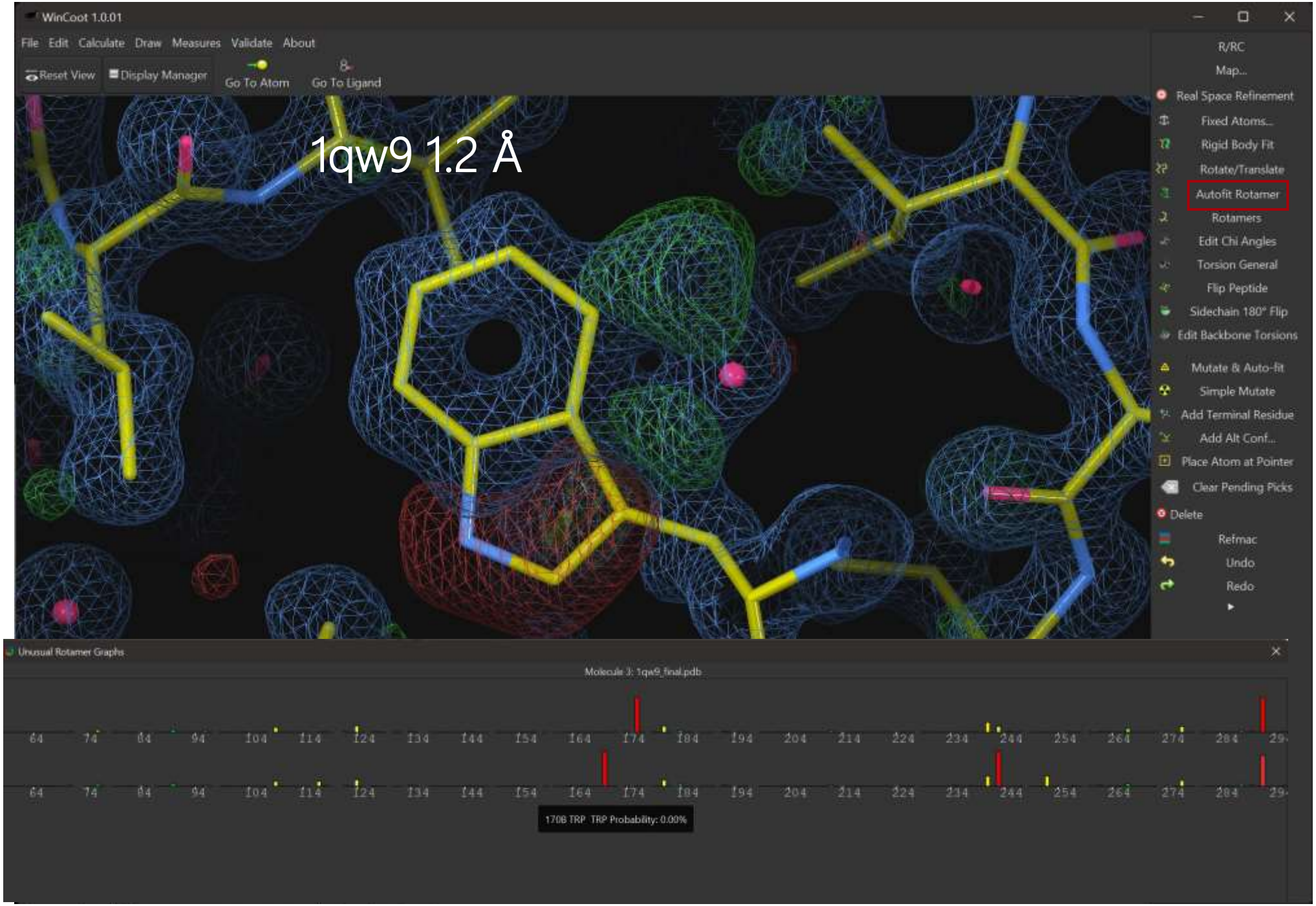
Validation CCPi2

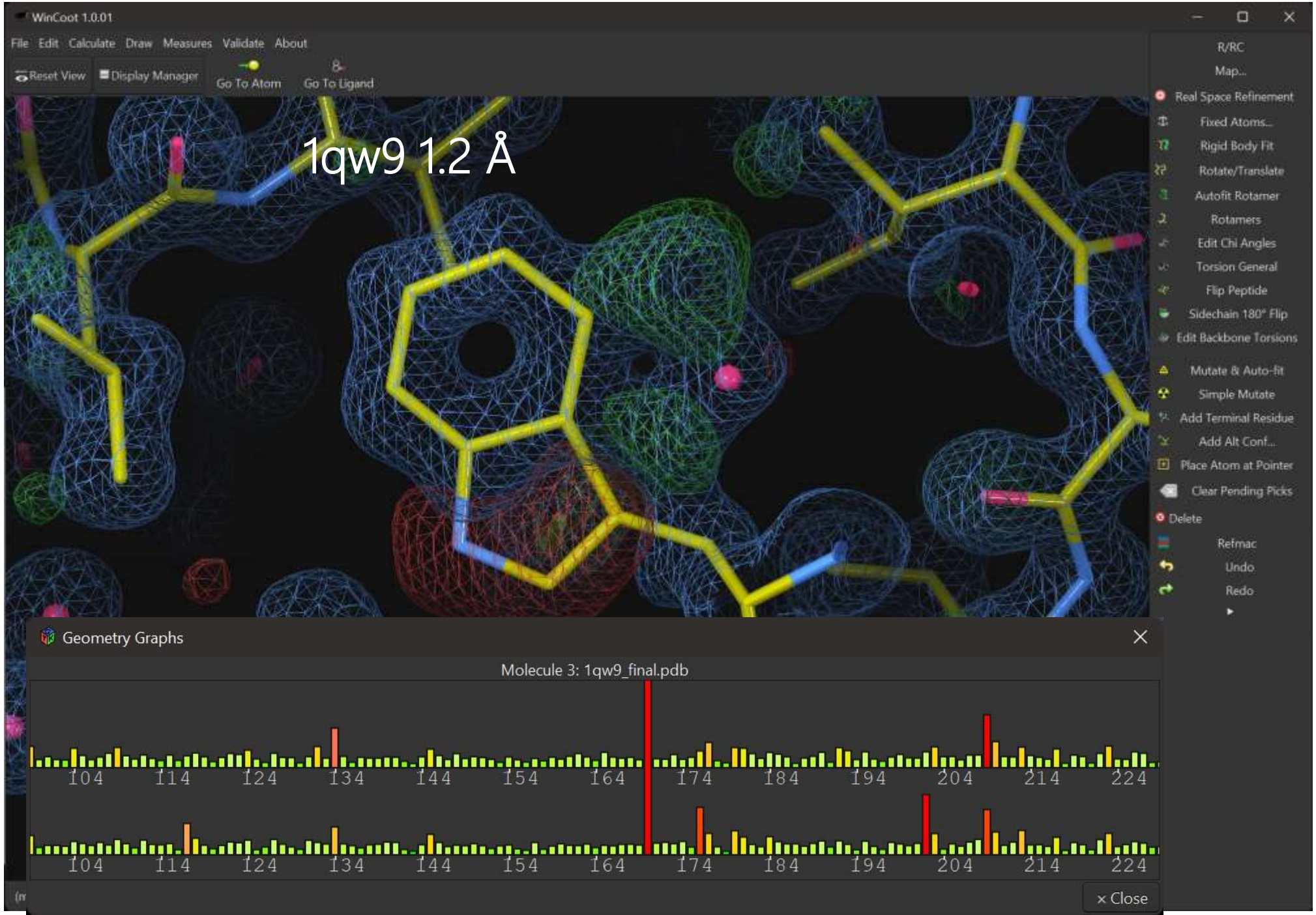


Coot Validate

Throughout




WinCoot 1.0.01
File Edit Calculate Draw Measures Validate About
Reset View
Display Manager
Go To Atom
Go To Ligand
1qw9 1.2 Å
R/RC
Map...
Real Space Refinement
Fixed Atoms...
Rigid Body Fit
Rotate/Translate
Autofit Rotamer
Rotamers
Edit Chi Angles
Torsion General
Flip Peptide
Sidechain 180° Flip
Edit Backbone Torsions
Mutate & Auto-fit
Simple Mutate
Add Terminal Residue
Add Alt Conf...
Place Atom at Pointer
Clear Pending Picks
Delete
Refmac
Undo
Redo
Unusual Rotamer Graphs
Molecule 3: 1qw9_final.pdb
64 74 84 94 104 114 124 134 144 154 164 174 184 194 204 214 224 234 244 254 264 274 284
1708 TRP TRP Probability: 0.00%

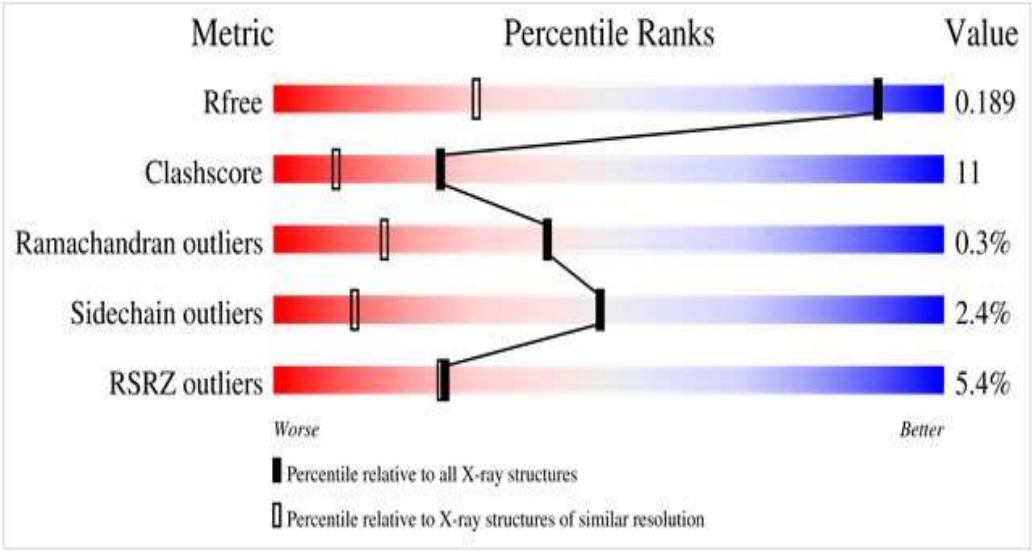

WinCoot 1.0.01
File Edit Calculate Draw Measures Validate About
Reset View
Display Manager
Go To Atom
Go To Ligand
1qw9 1.2 Å
R/RC
Map...
Real Space Refinement
Fixed Atoms...
Rigid Body Fit
Rotate/Translate
Autofit Rotamer
Rotamers
Edit Chi Angles
Torsion General
Flip Peptide
Sidechain 180° Flip
Edit Backbone Torsions
Mutate & Auto-fit
Simple Mutate
Add Terminal Residue
Add Alt Conf...
Place Atom at Pointer
Clear Pending Picks
Delete
Refmac
Undo
Redo
Geometry Graphs
Molecule 3: 1qw9_final.pdb
104 114 124 134 144 154 164 174 184 194 204 214 224
104 114 124 134 144 154 164 174 184 194 204 214 224
× Close

Slider Plots

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 1.20 Å
R-Value Free: 0.179
R-Value Work: 0.160
R-Value Observed: 0.173

wwPDB Validation 3D Report | Full Report



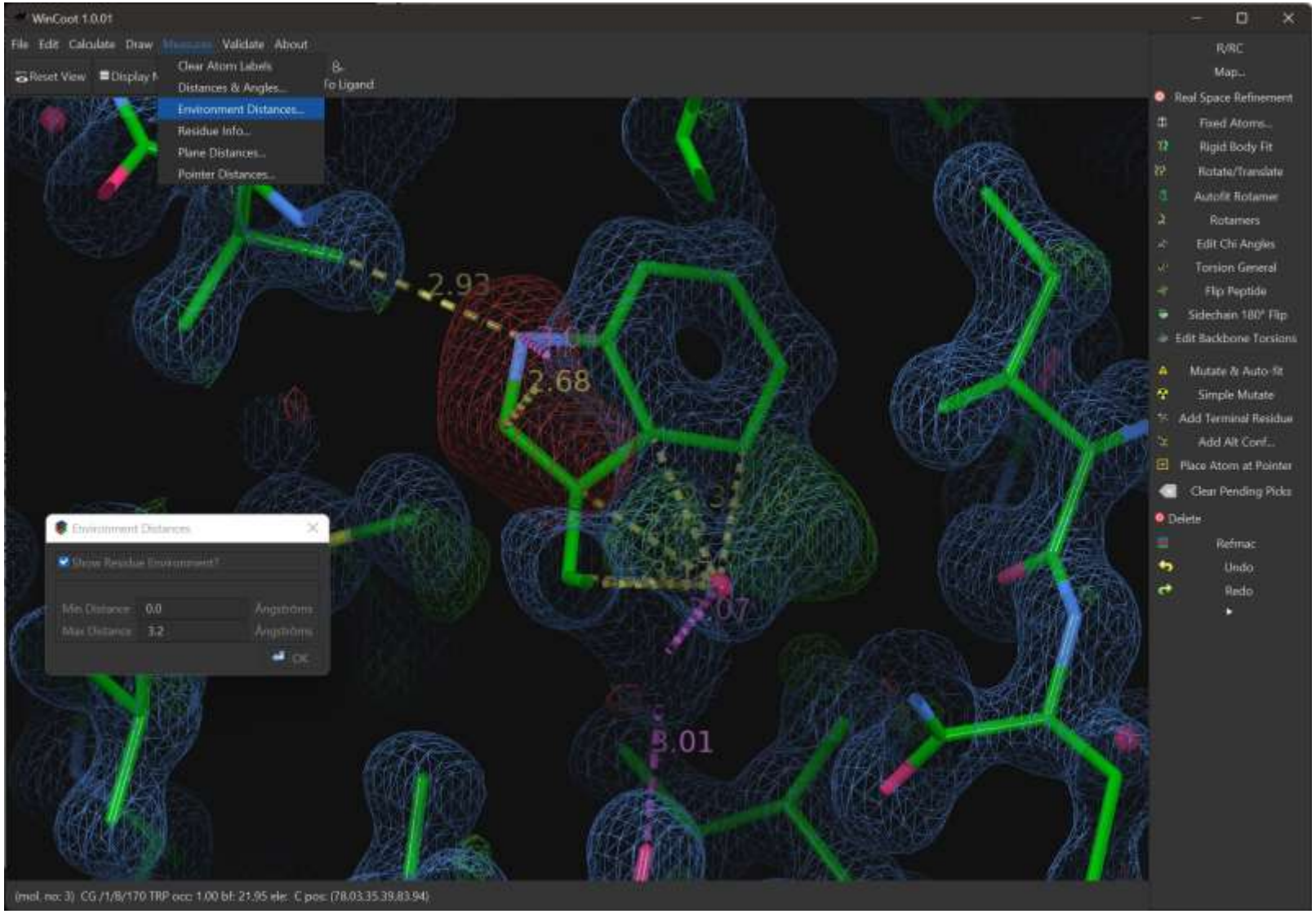
Ligand Structure Quality Assessment

Published EMBO 2003

Homo 6-mer ligand studies seemed to ignore basic side chain model building (using "O") protocol for a high resolution structure!

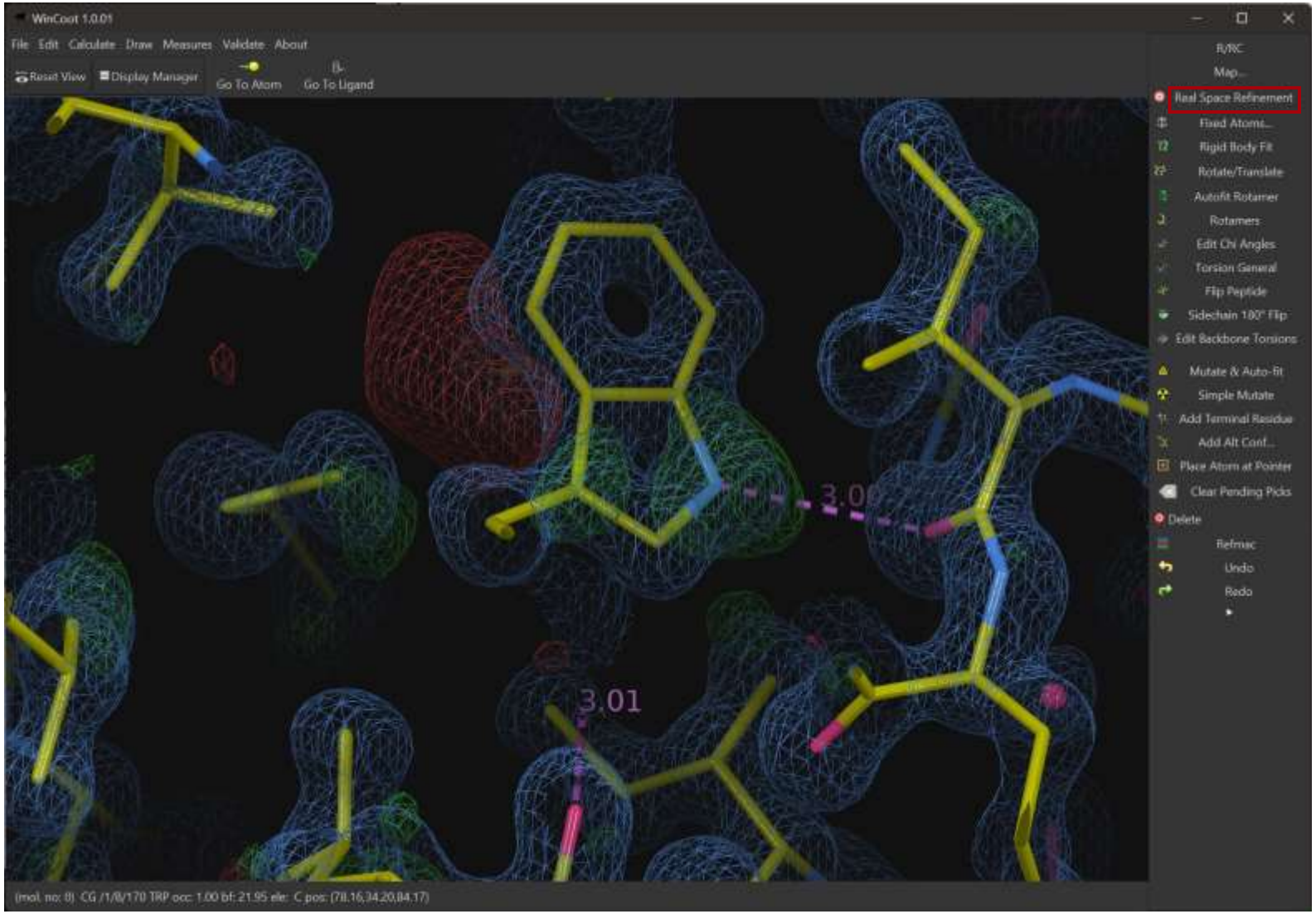
Hydrogen Bond environment

- Asn, Gln, and His flips
  - Detected by MolProbity, corrected in Phenix
  - Also use common sense, check with coot>measures>environment distances
- Buried unsatisfied H-bond donors and acceptors indicate (subtle) errors
- Waters should also make hydrogen bonds
- explicit riding hydrogen atoms can improve geometry at any resolution



Hydrogen Bond environment

- Asn, Gln, and His flips
  - Detected by MolProbity, corrected in Phenix
  - Also use common sense, check with coot>measures>environment distances
- Buried unsatisfied H-bond donors and acceptors indicate (subtle) errors
- Waters should also make hydrogen bonds
- explicit riding hydrogen atoms can improve geometry at any resolution



When should I refine with hydrogens?

This is largely a matter of personal preference. Using explicit riding hydrogen atoms can improve geometry at any resolution; at higher resolutions, approximately 2 Angstrom or better, they will generally improve R-free as well. At atomic resolution (1.5 A or better) they should always be part of the final model. Note that at unless you have true subatomic resolution (0.9 A or better), the hydrogens should always be refined as "riding", meaning that their coordinates are defined by the heavy atoms, not individually refined against X-ray data.

Validation – overview

	POSITIVE	NEGATIVE
INTERNAL	STRENGTH › Experienced lab members in crystallographic analysis. › Good research environment. › Robust software protocols.	WEAKNESS › User inexperience. › Poor research environment – teaching or expertise. › Black box automated crystallography › Busy boss or senior scientist.
EXTERNAL	REVIEW › Reviewers of your paper. › PDB deposition (software & biocuration staff). › Researchers in the field (those who may come across your structure and use it)	CONSEQUENCE › Unnoticed problems. › Likely to be discovered especially more obvious errors – retraction of data › May be negative consequences for collaborators & research community

Software Tools

- Many different software tools
  - General: PDB-REDO, MolProbity, PDB-validation server
  - Special purpose:
    - PDB-care, (PDB CArbohydrate REsidue check), also Privateer
    - CheckMyMetal, CheckMyBlob web servers.
  - Tools may check the same things differently

DEPOSITION

OneDep, a unified system for deposition, biocuration, and validation of experimentally determined X-ray, NMR, and 3DEM macromolecular structures,

https://deposit-pdbe.wwpdb.org/deposition

PDB Deposition

SAMPLE SDEQUENCE

Sequence should
- Contain all residues
- Include expression tags and disordered residues.

Any mismatch should correspond to mutation, variant or expression tag in your sequence.

LIGAND

Are your ligands already in Chemical Component Dictionary (CCD)?

☑ The 3-character code for your ligand should match that in the Dictionary

☒ During deposition you can provide a SMILES or upload a 2D chemical diagram

PREPARE DATA

Generate coordinate file in PDBx/mmCIF format
- Output directly from refinement program suites (e.g., PHENIX, CCP4i2)
- Ensures that the maximum metadata is captured for deposition.

VALIDATE VALIDATE

wwPDB provides a standalone validation server or via CCP4i2 interface.

DEPOSIT STRUCTURE

Deposition to OneDep system.

A session ID is provided to enable you to continue your deposition at a later date. PDB ID will be provided upon completion of your deposition

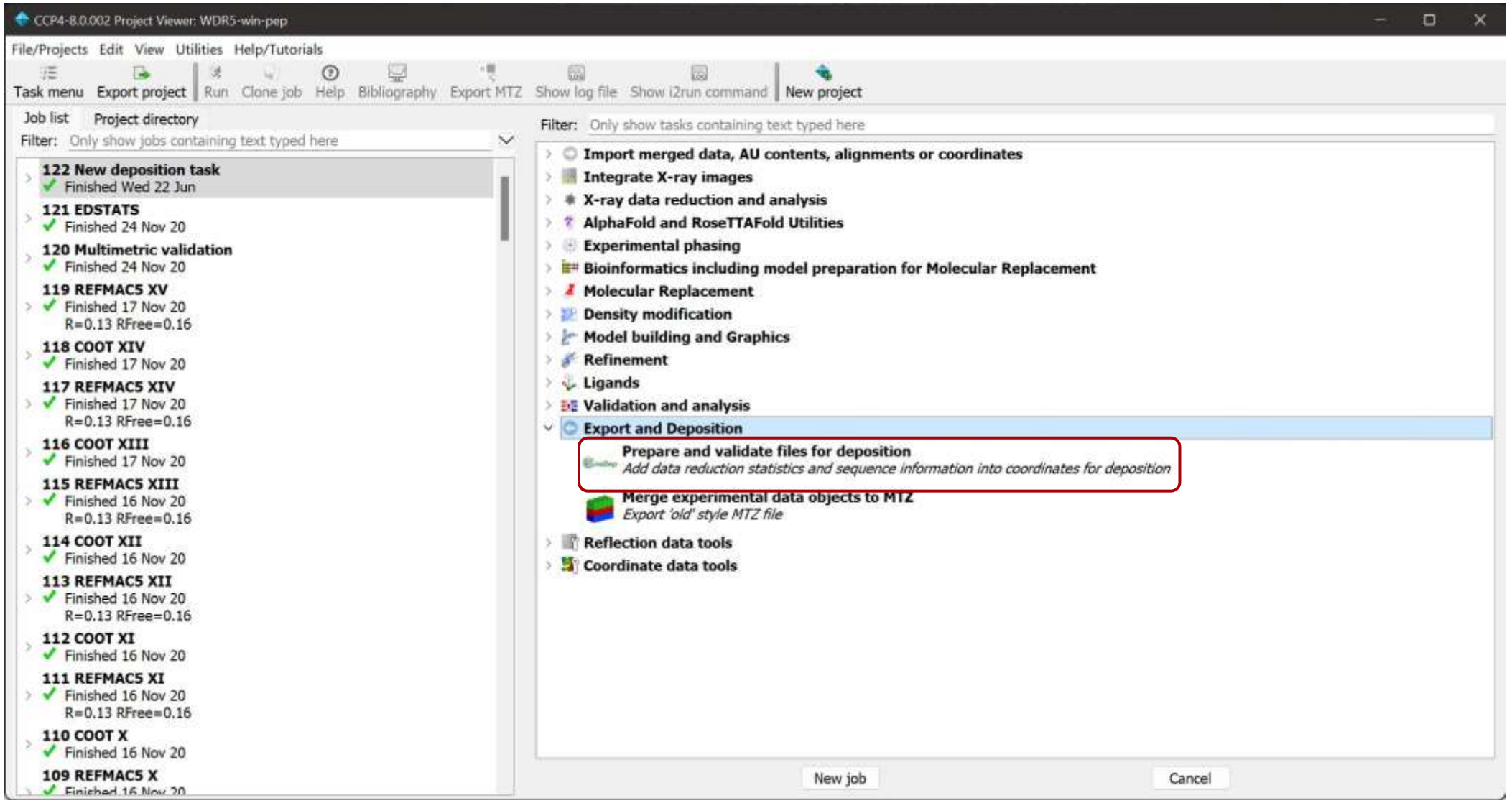
Less form filling !

Export & Deposition

PDBx/mmCIF explicitly documents all relationships between common data items (e.g. atom and residue identifiers) which permits software applications to evaluate and validate referential integrity with any PDB entry, and maps information between the residue sequences of the experimental sample and the model coordinates.

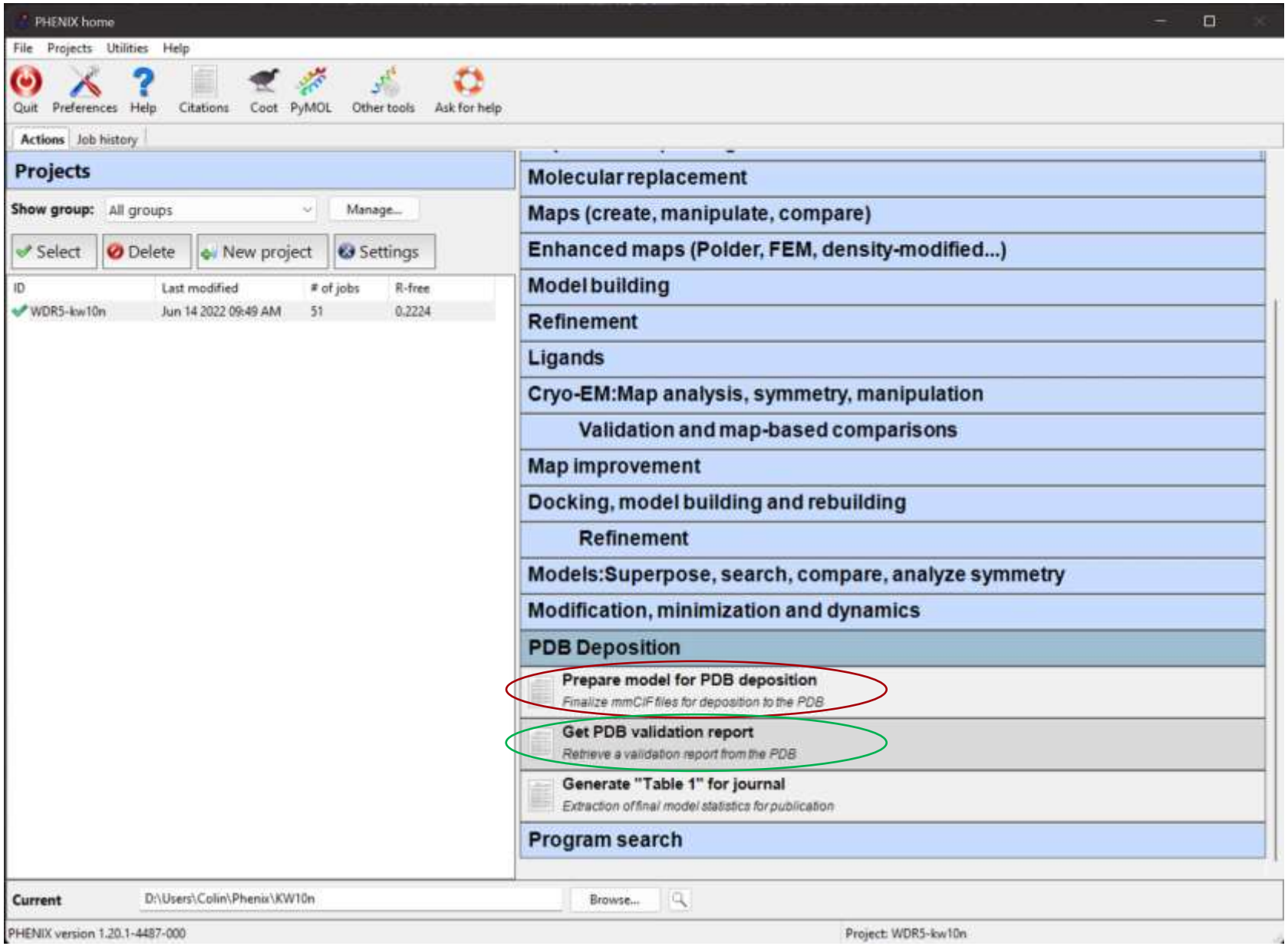
CCP4i2 Deposition

v7.1 onwards

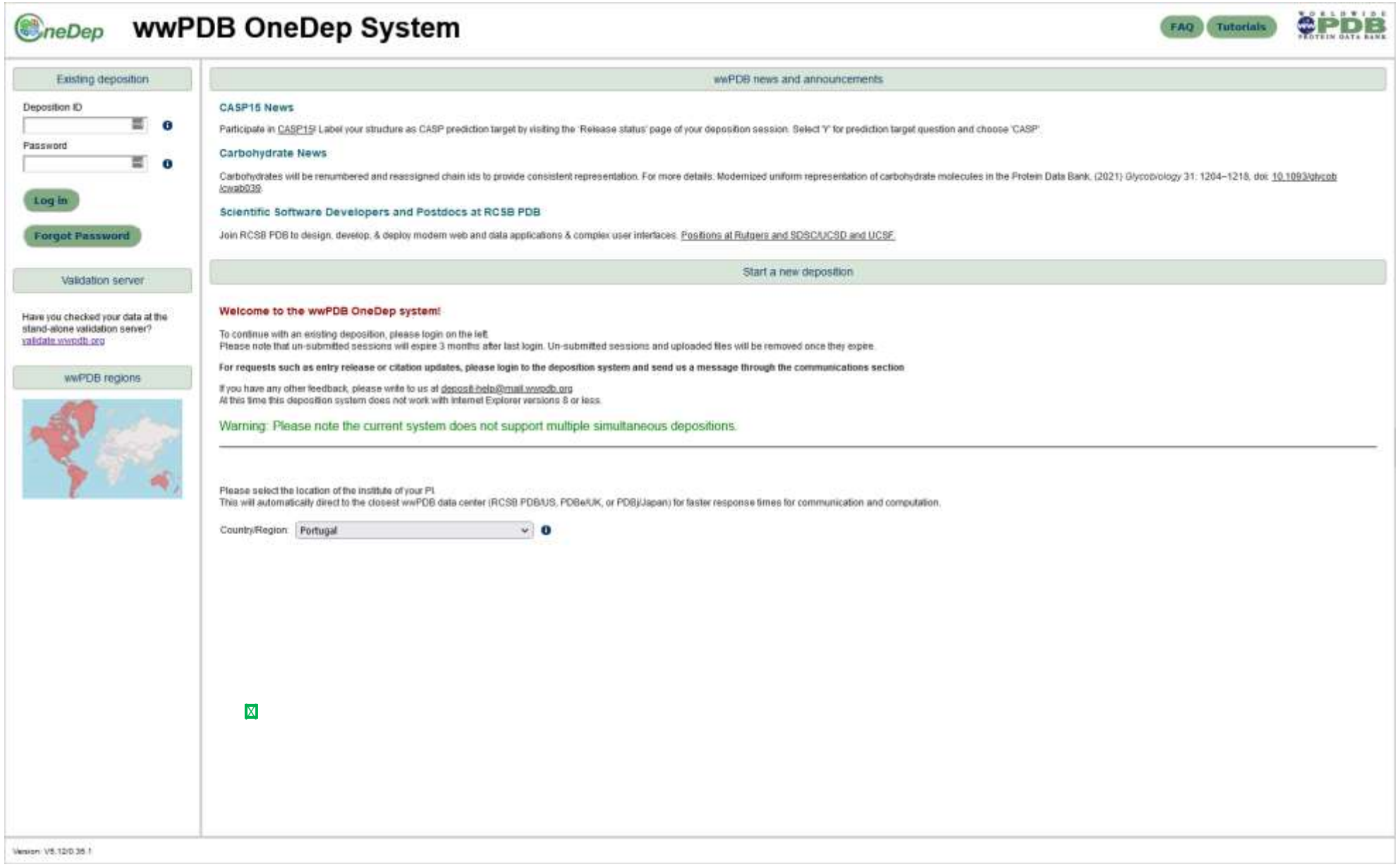


Youtube: CCP4 SW2020 - Whats new - Improved Validation and deposition in CCP4i2 - speaker: Huw Jenkins

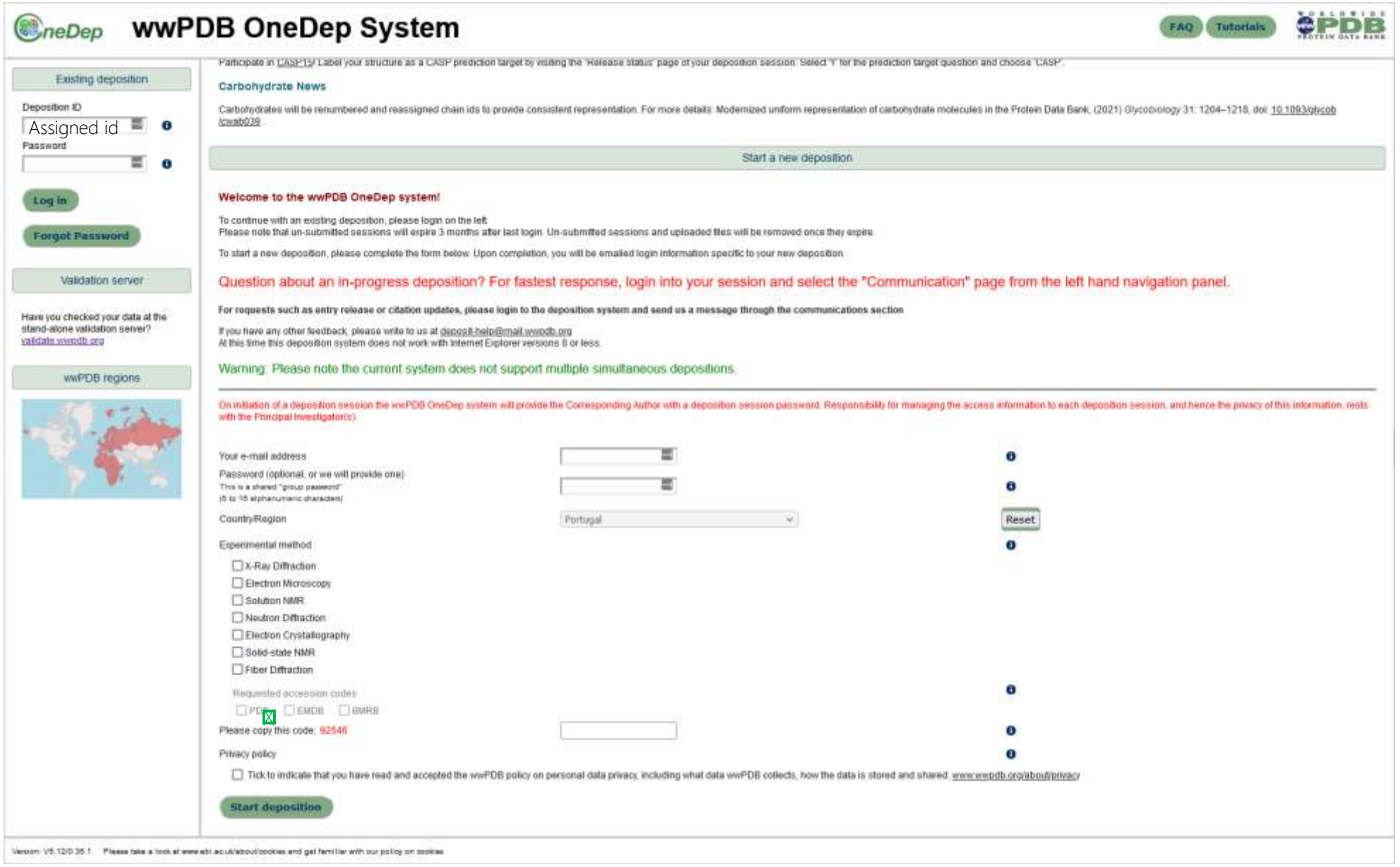
Phenix Deposition



OneDep


OneDep
wwPDB OneDep System
FAQ
Tutorials
Existing deposition
Deposition ID
Password
Log in
Forgot Password
Validation server
Have you checked your data at the stand-alone validation server?
validate.wwpdb.org
wwPDB regions
wwPDB news and announcements
CASP15 News
Participate in CASP15! Label your structure as CASP prediction target by visiting the 'Release status' page of your deposition session. Select 'Y' for prediction target question and choose 'CASP'.
Carbohydrate News
Carbohydrates will be renumbered and reassigned chain ids to provide consistent representation. For more details: Modernized uniform representation of carbohydrate molecules in the Protein Data Bank. (2021) Glycobiology 31: 1204–1218. doi: 10.1093/glycob/cwab039.
Scientific Software Developers and Postdocs at RCSB PDB
Join RCSB PDB to design, develop, & deploy modern web and data applications & complex user interfaces. Positions at Rutgers and SDSC/UCSD and UCSF.
Start a new deposition
Welcome to the wwPDB OneDep system!
To continue with an existing deposition, please login on the left.
Please note that un-submitted sessions will expire 3 months after last login. Un-submitted sessions and uploaded files will be removed once they expire.
For requests such as entry release or citation updates, please login to the deposition system and send us a message through the communications section.
If you have any other feedback, please write to us at deposit-help@mail.wwpdb.org
At this time this deposition system does not work with Internet Explorer versions 8 or less.
Warning: Please note the current system does not support multiple simultaneous depositions.
Please select the location of the institute of your PI.
This will automatically direct to the closest wwPDB data center (RCSB PDB/US, PDBe/UK, or PDBj/Japan) for faster response times for communication and computation.
Country/Region: Portugal

OneDep


wwPDB OneDep System
FAQ
Tutorials
Existing deposition
Deposition ID
Assigned id
Password
Log in
Forgot Password
Validation server
Have you checked your data at the stand-alone validation server?
validate.wwpdb.org
wwPDB regions
Participate in CASP15! Label your structure as a CASP prediction target by visiting the 'Release status' page of your deposition session. Select 'Y' for the prediction target question and choose 'CASP'.
Carbohydrate News
Carbohydrates will be renumbered and reassigned chain ids to provide consistent representation. For more details: Modernized uniform representation of carbohydrate molecules in the Protein Data Bank. (2021) Glycobiology 31: 1204–1218. doi: 10.1093/glycob/cwab039
Start a new deposition
Welcome to the wwPDB OneDep system!
To continue with an existing deposition, please login on the left.
Please note that un-submitted sessions will expire 3 months after last login. Un-submitted sessions and uploaded files will be removed once they expire.
To start a new deposition, please complete the form below. Upon completion, you will be emailed login information specific to your new deposition.
Question about an in-progress deposition? For fastest response, login into your session and select the "Communication" page from the left hand navigation panel.
For requests such as entry release or citation updates, please login to the deposition system and send us a message through the communications section.
If you have any other feedback, please write to us at deposit-help@mail.wwpdb.org
At this time this deposition system does not work with Internet Explorer versions 8 or less.
Warning: Please note the current system does not support multiple simultaneous depositions.
On initiation of a deposition session the wwPDB OneDep system will provide the Corresponding Author with a deposition session password. Responsibility for managing the access information to each deposition session, and hence the privacy of this information, rests with the Principal Investigator(s).
Your e-mail address
Password (optional, or we will provide one)
This is a shared "group password"
(6 to 16 alphanumeric characters)
Country/Region
Portugal
Reset
Experimental method
X-Ray Diffraction
Electron Microscopy
Solution NMR
Neutron Diffraction
Electron Crystallography
Solid-state NMR
Fiber Diffraction
Requested accession codes
PDB
EMDB
BMRB
Please copy this code: 92546
Privacy policy
Tick to indicate that you have read and accepted the wwPDB policy on personal data privacy, including what data wwPDB collects, how the data is stored and shared. www.wwpdb.org/about/privacy
Start deposition

wwPDB Deposition: D_1292112360 -- ID: 7BCY(REL)

FAQ | Tutorial

Deposition locked

All items

Mandatory items

Navigation

- Instructions
- Communication
- Review
  - Assembly Review
  - Validation reports
- Re-upload files
- Admin
  - Contact information
  - Grant information
  - Release status
  - Entry title & author
  - Citation information
- Macromolecules
  - 1) WD repeat-containing protein
  - 2) ORF 73
- Data collection
  - Crystal Information
  - Collection Source
  - Software Used
  - Collection Statistics
- Refinement
  - Refinement
- Ligands
- Related entries
- Summary & conditions

Log out

Communication with the wwPDB

This page shows you the current messages associated with this deposition and allows you to send a new message to the wwPDB.

- All messaging on this page will be archived permanently with this deposition.
- This page can ONLY be used to communicate information about this deposition.
- This page shows:
  - All messages as a list.
  - A box for you to create and send a new message.
  - If you are communicating changes to be made in the annotation of your file, please do so preferably using CIF format.

Old messages

2021-11-30 15:24:11	auto	Release of PDB ID 7BCY	Dear Depositors, This message...
2021-11-30 14:27:27	depositor - McVey, Colin (principal investigator/group leader)	Release with citation update	Dear pdb Please could you rele...
2020-12-22 12:09:57	RG	Acknowledgement of Structure Approval	Dear Depositors, Thank you fo...
2020-12-22 12:05:00	depositor - McVey, Colin (principal investigator/group leader)	Approval with corrections	Dear Genevieve, Thank you for...
2020-12-22 02:18:18	GE	PDB ID 7BCY - Validation report and processed files are ready for your review	Dear Depositors, Thank you fo...
2020-12-21 15:20:12	depositor - McVey, Colin (principal investigator/group leader)	Other corrections	Hi, Could I change the title t...

Last message from annotator

Time stamp: 2021-11-30 15:24:11

Sender: auto

Subject: Release of PDB ID 7BCY

Full Text:

Dear Depositors,

This message is to inform you that your structure PDB ID 7BCY (Deposition ID D_1292112360) and the associated experimental data which were deposited with release instructions, 'HPUB', will be released on 8 December 2021. This is the next available release date.

Attachments:

D_1292112360_model-review_P1.cif.V2
D_1292112360_model-annotate_P1.pdb.V2
D_1292112360_sf-annotate_P1.cif.V2
D_1292112360_val-report-annotate_P1.pdf.V1
D_1292112360_val-report-full-annotate_P1.pdf.V1
D_1292112360_val-data-annotate_P1.xml.V1
D_1292112360_val-report-wwpdb-2fo-fc-edmap-coef-annotate_P1.cif.V1
D_1292112360_val-report-wwpdb-fo-fc-edmap-coef-annotate_P1.cif.V1
D_1292112360_val-report-annotate_P1.pdf.V1
D_1292112360_val-report-full-annotate_P1.pdf.V1
D_1292112360_val-data-annotate_P1.xml.V1
D_1292112360_val-report-wwpdb-2fo-fc-edmap-coef-annotate_P1.cif.V1
D_1292112360_val-report-wwpdb-fo-fc-edmap-coef-annotate_P1.cif.V1

Deposition locked

All items

Mandatory items

Navigation

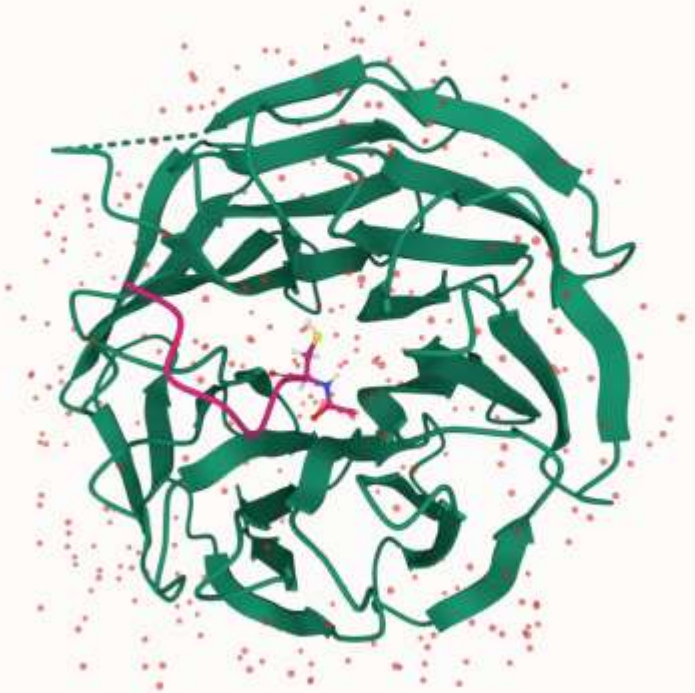
- Instructions
- Communication
- Review
  - Assembly Review
  - Validation reports
- Re-upload files
- Admin
  - Contact information
  - Grant information
  - Release status
  - Entry title & author
  - Citation information
- Macromolecules
  - 1) WD repeat-containing protein
  - 2) ORF 73
- Data collection
  - Crystal Information
  - Collection Source
  - Software Used
  - Collection Statistics
- Refinement
  - Refinement
- Ligands
- Related entries
- Summary & conditions

Log out

Review

This page shows the annotated assemblies in 3D. Use the menu on the right hand side of the 3D viewer to change between the assemblies.

Sequence of D_12921123... Chain 1: WD repeat... A



Structure Tools

Structure

D_1292112360 | X-ray structure of W...

Type Assembly

Asm Id 1: Author And Software ...

Dynamic Bonds Off

Nothing Focused

Measurements

+ Add

Components D_1292112360

Preset + Add

Polymer Cartoon

Ligand Ball & Stick

Water Ball & Stick

Unit Cell P 1

Deposition locked

All items

Mandatory items

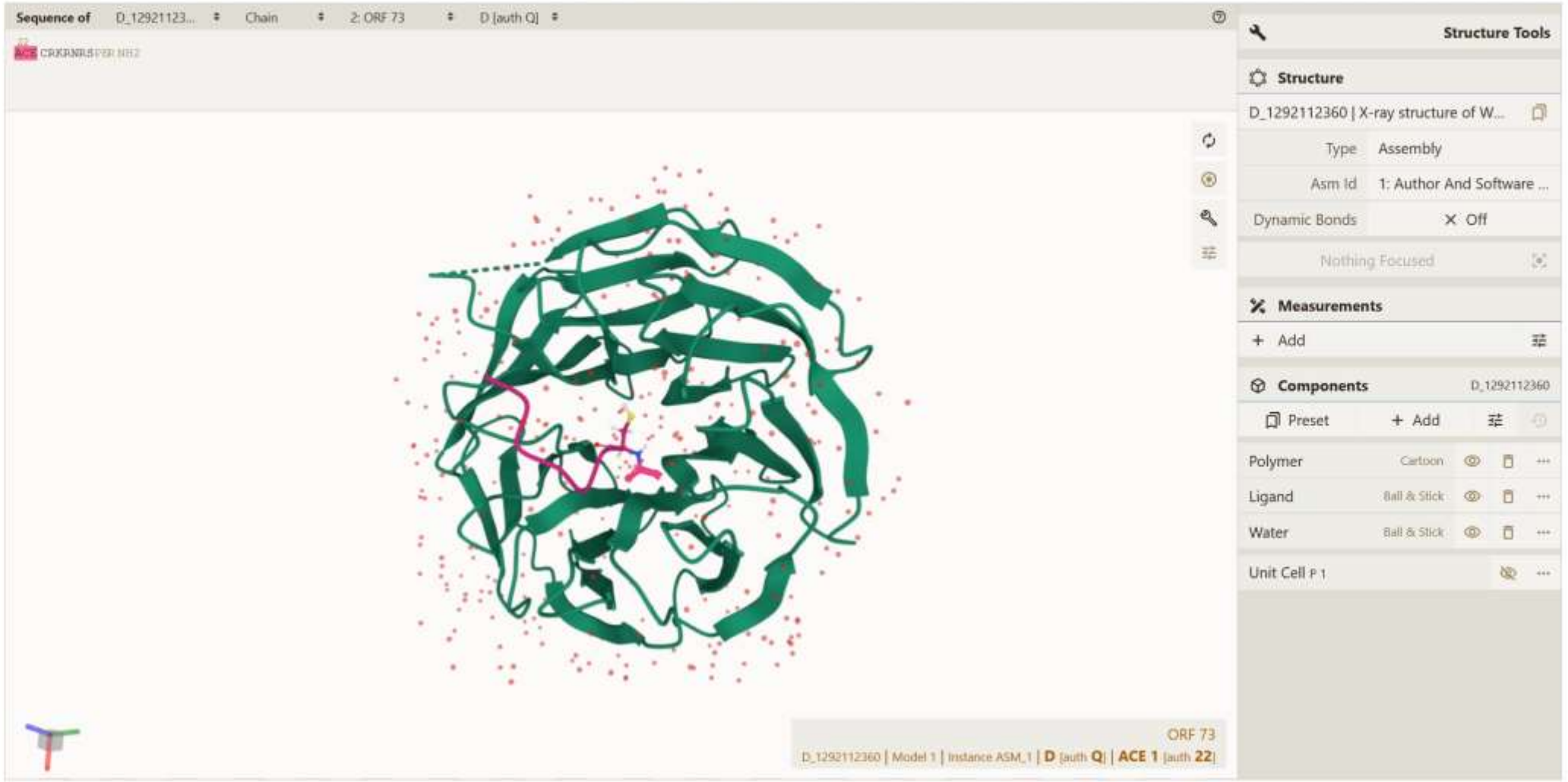
Navigation

- Instructions
- Communication
- Review
  - Assembly Review
  - Validation reports
- Re-upload files
- Admin
  - Contact information
  - Grant information
  - Release status
  - Entry title & author
  - Citation information
- Macromolecules
  - 1) WD repeat-containing proteir
  - 2) ORF 73
- Data collection
  - Crystal Information
  - Collection Source
  - Software Used
  - Collection Statistics
- Refinement
  - Refinement
- Ligands
- Related entries
- Summary & conditions

Log out

Review

This page shows the annotated assemblies in 3D. Use the menu on the right hand side of the 3D viewer to change between the assemblies.



Deposition locked

All items

Mandatory items

Navigation

- Instructions
- Communication
- Review
  - Assembly Review
  - Validation reports
- Re-upload files
- Admin
  - Contact information
  - Grant information
  - Release status
  - Entry title & author
  - Citation information
- Macromolecules
  - 1) WD repeat-containing protein
  - 2) ORF 73
- Data collection
  - Crystal Information
  - Collection Source
  - Software Used
  - Collection Statistics
- Refinement
  - Refinement
- Ligands
- Related entries
- Summary & conditions

Log out

Sequence

Input the sequence of this molecule using standard one-letter codes. Please include the complete sequence including tags, linkers, unobserved regions and mutations. Non-standard residues should be input using the three-letter code in parenthesis, e.g. (MSE).

Sample sequence in one letter code*: (ACE)CRKRNRSPER(NH2)

The sequence alignment is valid

Refresh the alignment from sample sequence provided

The following is the alignment between the sample sequence and sequence from the coordinates. **The sample sequence is shown at the top of the alignment display.**

Important: Please address any discrepancy between the sample sequence and coordinate sequence by either providing a correct sample sequence or re-uploading new coordinates.

```
ok : Sample sequence aligned with Coordinates - all chains :

Molecule : 2
Chain : P
------------10------
(ACE)CRKRNRSPER(NH2)
( | )|||||||---( - )
(ACE)CRKRNRS...(...)

Chain : Q
------------10------
(ACE)CRKRNRSPER(NH2)
( | )|||||||---( - )
(ACE)CRKRNRS...(...)
```

Compound details: N-terminal acetylation and C-terminal amidation of LANA peptide

Source

How was the molecule obtained: ○ Recombinantly expressed ○ Purified from natural source ◉ Chemically synthesized

Deposition locked

All items

Mandatory items

Navigation

- Instructions
- Communication
- Review
  - Assembly Review
  - Validation reports
- Re-upload files
- Admin
  - Contact information
  - Grant information
  - Release status
  - Entry title & author
  - Citation information
- Macromolecules
  - 1) WD repeat-containing proteir
  - 2) ORF 73
- Data collection
  - Crystal Information
  - Collection Source
  - Software Used
  - Collection Statistics
- Refinement
  - Refinement
- Ligands
- Related entries
- Summary & conditions

Log out

Crystal Information

Crystal Information

Crystal ID*: 1

Crystallization details

Method*: VAPOR DIFFUSION, SITTING DROP

pH: 5.5

pH range: 5.5 - 6.5

Temperature (K)*: 294

Details about the temperature:

Crystallisation conditions*: 100 mM Bis-Tris pH 5.5, 50 mM (NH4)2SO4, 25% PEG3350

Crystal details

Matthews Coefficient: 1.94

Measured density:

Percentage Solvent: 36.7

Description of crystal morphology: Rod-shaped

A description of the quality and habit of the crystal. The crystal dimensions should not normally be reported here; use instead the specific items in the EXPTL_CRYSTAL category relating to size for the gross dimensions of the crystal and data items in the EXPTL_CRYSTAL_FACE category to describe the relationship between individual faces.

Continue to next section

wwPDB Deposition: D_1292112295 -- ID: 7BED(REL)

FAQ | Tutorial

Deposition locked

All items

Mandatory items

Navigation

- Instructions
- Communication
- Review
  - Assembly Review
  - Validation reports
- Re-upload files
- Admin
  - Contact information
  - Grant information
  - Release status
  - Entry title & author
  - Citation information
- Macromolecules
  - 1) WD repeat-containing protein
- Data collection
  - Crystal Information
  - Collection Source
  - Software Used
  - Collection Statistics
- Refinement
  - Refinement
- Ligands
- Related entries
- Summary & conditions

Log out

Data collection 1

Diffraction ID*: 1

Data collection temperature (K)*: 100

Data collection temperature details:

Serial crystallography experiment: N Y

Diffraction Radiation

Protocol*: LAUE MAD SINGLE WAVELENGTH

Monochromator: Double crystal monochr

Scattering type*: electron neutron x-ray

Diffraction Source

Source type*: SYNCHROTRON

Source details*: DIAMOND BEAMLINE IC

Wavelength or list of wavelengths used for this experiment (Å)*: 0.9795

Diffraction Detector

Detector*: DECTRIS PILATUS3 6M

Detector type*: PIXEL

Date of collection*: 2016-01-25

Optics: DCM (Double Crystal Mc

Continue to next section

Deposition locked

All items

Mandatory items

Navigation

- Instructions
- Communication
- Review
  - Assembly Review
  - Validation reports
- Re-upload files
- Admin
  - Contact information
  - Grant information
  - Release status
  - Entry title & author
  - Citation information
- Macromolecules
  - 1) WD repeat-containing protein
- Data collection
  - Crystal Information
  - Collection Source
  - Software Used
  - Collection Statistics
- Refinement
  - Refinement
- Ligands
- Related entries
- Summary & conditions

Log out

Data Collection Statistics

Overall data quality

Reflections ID*:	1
Number of unique reflections measured*:	157708
Completeness for range (%)*:	99.3
Data redundancy*:	3.8
Resolution range high (Å)*:	1.26
Resolution range low (Å)*:	27.23
Rejection criteria (sigma(F)):	
Rejection criteria (sigma(I)):	
CC half:	0.999
CC star:	
Rmerge(I):	0.077
Rsym(I):	
Rpim(I):	0.045
Rmeas(I):	0.09
Rsplit:	0.068
Average I/sigma(I) for the data set*:	7.9
Chi squared:	0.57
I/Avg sigma(I):	9.0
Wilson B factor ($Å^2$):	10.07

Data quality in resolution shells

Reflections ID*:	1
Highest resolution shell - range high (Å)*:	1.26
Highest resolution shell - range low (Å)*:	1.28

Deposition locked

All items

Mandatory items

Navigation

- Instructions
- Communication
- Review
  - Assembly Review
  - Validation reports
- Re-upload files
- Admin
  - Contact information
  - Grant information
  - Release status
  - Entry title & author
  - Citation information
- Macromolecules
  - 1) WD repeat-containing protein
- Data collection
  - Crystal information
  - Collection Source
  - Software Used
  - Collection Statistics
- Refinement
  - Refinement
- Ligands
- Related entries
- Summary & conditions

Log out

Refinement statistics

Resolution range high (Å)*:	1.26
Resolution range low (Å)*:	27.23
Data cutoff (sigma(F)):	1.33
Outlier cutoff high (rms(abs(F))):	
Outlier cutoff low (rms(abs(F))):	
Completeness (working+test) (%)*:	96.07
Number of Unique reflections refined against*:	157708
Cross-validation method*:	FREE R-VALUE
Free R value test set selection:	RANDOM
R value (working set):	0.1590
Free R value:	0.1713
R value (work + test):	0.1596
Free R value test set size (%):	4.78
Free R value test set count:	7538
Estimated error of free R value:	
Mean B value (overall A^2):	15.04
Details of the refinement:	Description of special aspects of the refinement process

Solvent

Solvent probe radius (Å):	1.1100
Ion probe radius (Å):	
Solvent shrinkage radius (Å):	0.9000
ksol:	

Summary

Prevention is better than cure

- Learn continually – there's always "one more thing" to learn. - Steve Jobs
- Information in the data (crystallization, assays, & interaction studies) *versus* the model
- Always be the first to question your own results
- Essential that validation is an integral part of the modelling process
  - Not just something you do when you deposit & publish!
- Can't substitute for common sense
- Keep learning – subscribe and interact on the CCP4bb

Finally

Validation is time consuming !
TEDIOUS !!

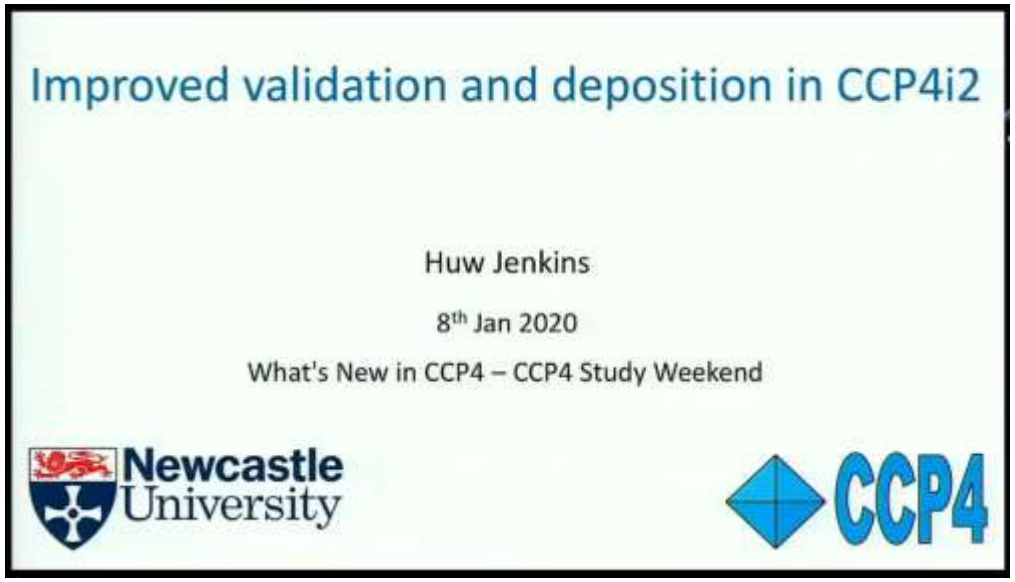
No pain no gain !

In the end you have better quality models
&
easier deposition process

Thank You

Useful links

CCP4 SW2020 - Whats new - Improved Validation and deposition in CCP4i2 - speaker: Huw Jenkins
https://www.youtube.com/watch?v=hd3pKgLpLM8&t=12s



- http://www.wwpdb.org/deposition/tutorial
- https://www.rcsb.org/docs/additional-resources/structure-validation-and-quality