



Full wwPDB X-ray Structure Validation Report ⓘ

Oct 18, 2022 – 02:02 pm BST

PDB ID : 7PRL
Title : MUC2 D1 with Cu(II)
Authors : Reznik, N.; Fass, D.
Deposited on : 2021-09-22
Resolution : 2.48 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.31.2
buster-report : 1.1.7 (2018)
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0267
CCP4 : 7.1.010 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.31.2

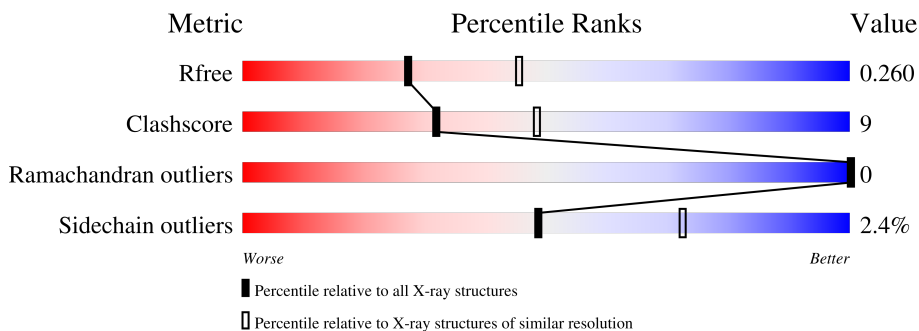
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.48 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	5857 (2.50-2.46)
Clashscore	141614	6594 (2.50-2.46)
Ramachandran outliers	138981	6469 (2.50-2.46)
Sidechain outliers	138945	6471 (2.50-2.46)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	395	 74% 15% • 10%

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 2868 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

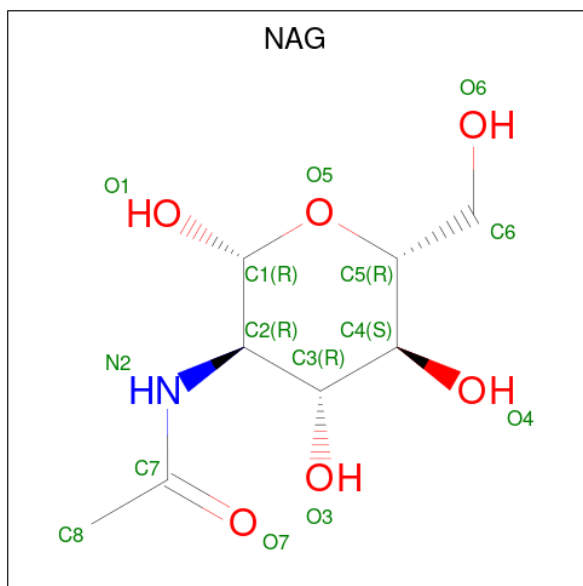
- Molecule 1 is a protein called Mucin-2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	357	2750	1701	481	531	37	0	7	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	390	HIS	-	expression tag	UNP A0A0G2JR65
A	391	HIS	-	expression tag	UNP A0A0G2JR65
A	392	HIS	-	expression tag	UNP A0A0G2JR65
A	393	HIS	-	expression tag	UNP A0A0G2JR65
A	394	HIS	-	expression tag	UNP A0A0G2JR65
A	395	HIS	-	expression tag	UNP A0A0G2JR65

- Molecule 2 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			6	3	3		
3	A	1	Total	C	O	0	0
			6	3	3		
3	A	1	Total	C	O	0	0
			6	3	3		
3	A	1	Total	C	O	0	0
			6	3	3		

- Molecule 4 is COPPER (II) ION (three-letter code: CU) (formula: Cu) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Cu	0	0
			1	1		

- Molecule 5 is water.

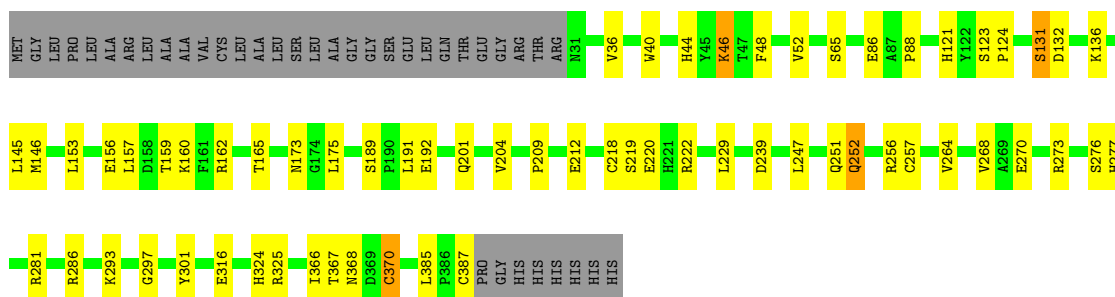
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	79	Total	O	0	0
			79	79		

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Mucin-2

Chain A:  74% 15% 10%



4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	95.90Å 141.21Å 68.52Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.17 – 2.48 49.17 – 2.48	Depositor EDS
% Data completeness (in resolution range)	99.2 (49.17-2.48) 92.8 (49.17-2.48)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.02 (at 2.48Å)	Xtrriage
Refinement program	PHENIX 1.14_3260	Depositor
R, R_{free}	0.186 , 0.260 0.186 , 0.260	Depositor DCC
R_{free} test set	850 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å ²)	36.9	Xtrriage
Anisotropy	0.893	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	2868	wwPDB-VP
Average B, all atoms (Å ²)	53.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.32% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CU, NAG, GOL

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.41	0/2834	0.60	0/3852

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2750	0	2542	47	0
2	A	14	0	13	0	0
3	A	24	0	32	2	0
4	A	1	0	0	0	0
5	A	79	0	0	9	0
All	All	2868	0	2587	47	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 9.

All (47) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:370:CYS:HB3	1:A:387:CYS:SG	2.18	0.83
1:A:370:CYS:CB	1:A:387:CYS:HG	1.92	0.82
1:A:370:CYS:CB	1:A:387:CYS:SG	2.70	0.79
1:A:136:LYS:HG2	1:A:146:MET:HG2	1.66	0.78
1:A:370:CYS:HG	1:A:387:CYS:HG	0.78	0.77
1:A:145:LEU:HD11	1:A:153:LEU:HD11	1.67	0.76
1:A:212:GLU:OE2	1:A:256:ARG:NH2	2.18	0.74
1:A:286:ARG:HH12	3:A:405:GOL:H32	1.52	0.73
1:A:293:LYS:H	3:A:403:GOL:H32	1.59	0.67
1:A:46:LYS:HG2	1:A:52:VAL:HG22	1.79	0.65
1:A:160:LYS:NZ	5:A:503:HOH:O	2.29	0.64
1:A:276[B]:SER:OG	1:A:325:ARG:O	2.18	0.62
1:A:175:LEU:O	5:A:501:HOH:O	2.16	0.59
1:A:293:LYS:HD3	1:A:301:TYR:CE1	2.38	0.58
1:A:173:ASN:OD1	5:A:501:HOH:O	2.17	0.58
1:A:121:HIS:HD2	5:A:532:HOH:O	1.88	0.56
1:A:123:SER:HB2	1:A:124:PRO:HD2	1.88	0.56
1:A:239:ASP:OD1	1:A:239:ASP:N	2.40	0.55
1:A:370:CYS:O	1:A:385:LEU:N	2.36	0.53
1:A:65:SER:O	1:A:165:THR:HA	2.09	0.51
1:A:252[A]:GLN:OE1	1:A:256:ARG:NH1	2.44	0.50
1:A:252[A]:GLN:HG3	1:A:256:ARG:HD2	1.93	0.50
1:A:264:VAL:O	1:A:268:VAL:HG23	2.12	0.49
1:A:159:THR:O	1:A:162:ARG:HG3	2.12	0.49
1:A:48:PHE:CZ	1:A:157:LEU:HD12	2.48	0.49
1:A:257:CYS:HB3	5:A:518:HOH:O	2.13	0.48
1:A:222:ARG:HB2	1:A:251:GLN:OE1	2.14	0.47
1:A:252[A]:GLN:HG3	1:A:256:ARG:HH11	1.79	0.47
1:A:131:SER:OG	1:A:132:ASP:N	2.49	0.46
1:A:370:CYS:HB2	1:A:385:LEU:O	2.15	0.45
1:A:191:LEU:HD11	1:A:209:PRO:HG2	1.99	0.44
1:A:229:LEU:HD12	1:A:247:LEU:HD23	1.98	0.44
1:A:366:ILE:HD12	1:A:367:THR:H	1.82	0.44
1:A:218:CYS:O	1:A:251:GLN:HB3	2.17	0.43
1:A:316:GLU:HG3	5:A:553:HOH:O	2.17	0.43
1:A:247:LEU:O	1:A:251:GLN:HG3	2.17	0.43
1:A:36:VAL:HG12	1:A:156:GLU:HG2	2.00	0.43
1:A:281:ARG:HE	1:A:281:ARG:HB3	1.59	0.43
1:A:44:HIS:ND1	5:A:507:HOH:O	2.36	0.42
1:A:273:ARG:NH2	5:A:502:HOH:O	2.23	0.42
1:A:201:GLN:NE2	1:A:204:VAL:HG21	2.36	0.41
1:A:297:GLY:HA3	5:A:549:HOH:O	2.19	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:40:TRP:CD2	1:A:270:GLU:HG2	2.56	0.41
1:A:86:GLU:HG2	1:A:256:ARG:O	2.21	0.41
1:A:277:HIS:HA	1:A:324:HIS:CE1	2.56	0.41
1:A:189:SER:OG	1:A:192:GLU:HG2	2.21	0.40
1:A:220:GLU:OE1	1:A:220:GLU:N	2.55	0.40

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	362/395 (92%)	346 (96%)	16 (4%)	0	100 100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	304/334 (91%)	296 (97%)	8 (3%)	46 70

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	46	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	88	PRO
1	A	131	SER
1	A	219	SER
1	A	252[A]	GLN
1	A	252[B]	GLN
1	A	368	ASN
1	A	370	CYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	201	GLN
1	A	274	GLN
1	A	372	GLN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 6 ligands modelled in this entry, 1 is monoatomic - leaving 5 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	GOL	A	402	-	5,5,5	0.98	0	5,5,5	0.96	0
2	NAG	A	401	1	14,14,15	0.43	0	17,19,21	0.53	0
3	GOL	A	404	-	5,5,5	0.95	0	5,5,5	1.01	0
3	GOL	A	403	-	5,5,5	1.04	0	5,5,5	0.89	0
3	GOL	A	405	-	5,5,5	0.97	0	5,5,5	0.87	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	GOL	A	402	-	-	1/4/4/4	-
2	NAG	A	401	1	-	4/6/23/26	0/1/1/1
3	GOL	A	404	-	-	4/4/4/4	-
3	GOL	A	403	-	-	0/4/4/4	-
3	GOL	A	405	-	-	4/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (13) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	404	GOL	O1-C1-C2-C3
3	A	405	GOL	O1-C1-C2-O2
3	A	405	GOL	C1-C2-C3-O3
2	A	401	NAG	C8-C7-N2-C2
2	A	401	NAG	O7-C7-N2-C2
3	A	405	GOL	O2-C2-C3-O3
3	A	404	GOL	C1-C2-C3-O3
3	A	405	GOL	O1-C1-C2-C3
3	A	404	GOL	O2-C2-C3-O3
3	A	404	GOL	O1-C1-C2-O2
2	A	401	NAG	O5-C5-C6-O6
2	A	401	NAG	C4-C5-C6-O6
3	A	402	GOL	C1-C2-C3-O3

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	403	GOL	1	0
3	A	405	GOL	1	0

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

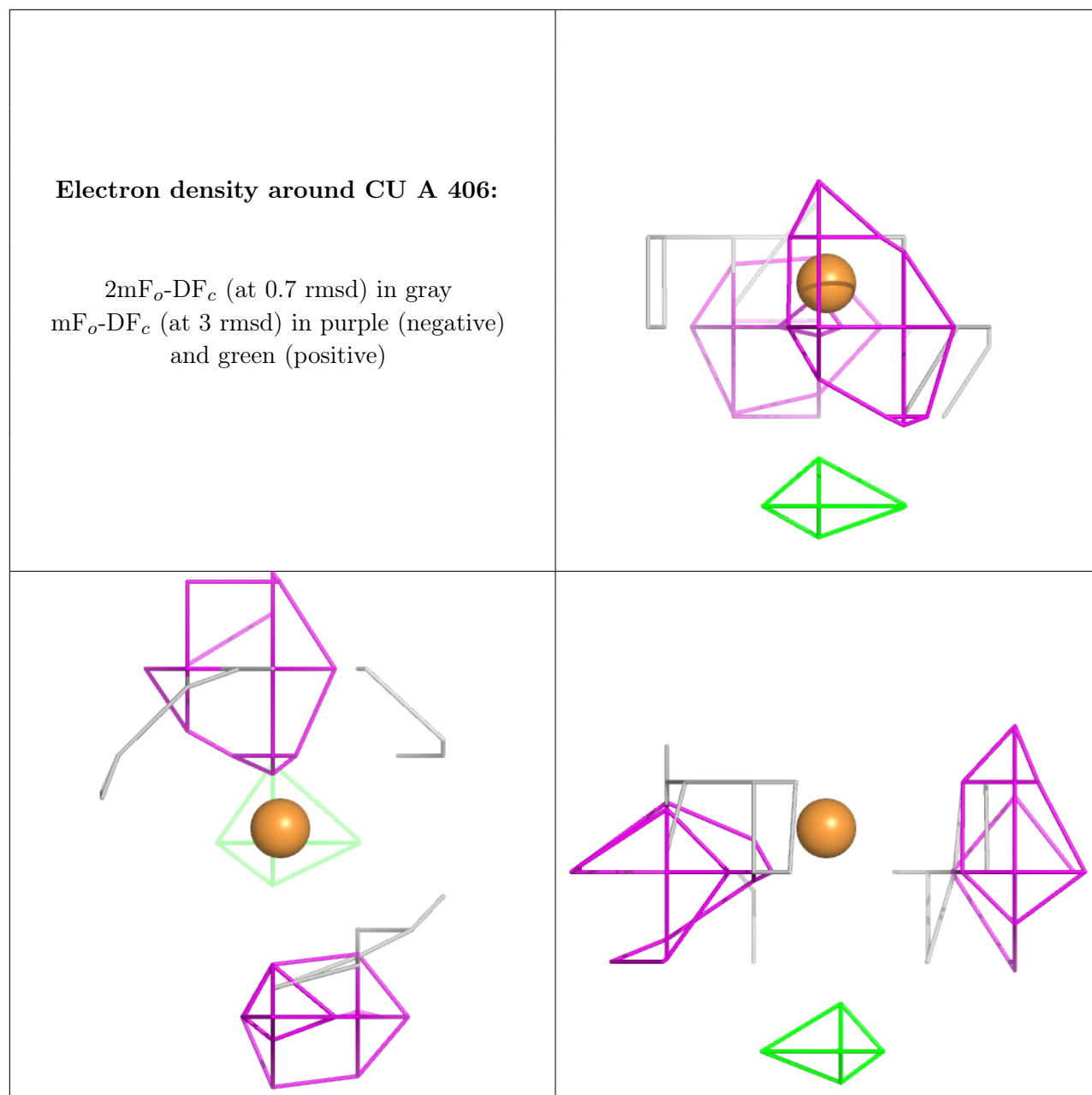
6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.



6.5 Other polymers [i](#)

Unable to reproduce the depositors R factor - this section is therefore empty.