WORLDWIDE wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 20, 2024 – 11:46 PM EDT

PDB ID	:	7H9J
Title	:	Group deposition for crystallographic fragment screening of Chikungunya virus nsP3 macrodomain – Crystal structure of Chikungunya virus nsP3 macrodomain in complex with Z57376600 (CHIKV_MacB-x1454)
Authors	:	Aschenbrenner, J.C.; Fairhead, M.; Godoy, A.S.; Balcomb, B.H.; Capkin, E.; Chandran, A.V.; Dolci, I.; Golding, M.; Koekemoer, L.; Lithgo, R.M.; Marples, P.G.; Ni, X.; Oliva, G.; Thompson, W.; Tomlinson, C.W.E.; Wild, C.; Winokan, M.; Xavier, M.-A.E.; Fearon, D.; von Delft, F.
Deposited on	:	2024-04-26
Resolution	:	1.88 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 (i) 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 (i)) were used in the production of this report:

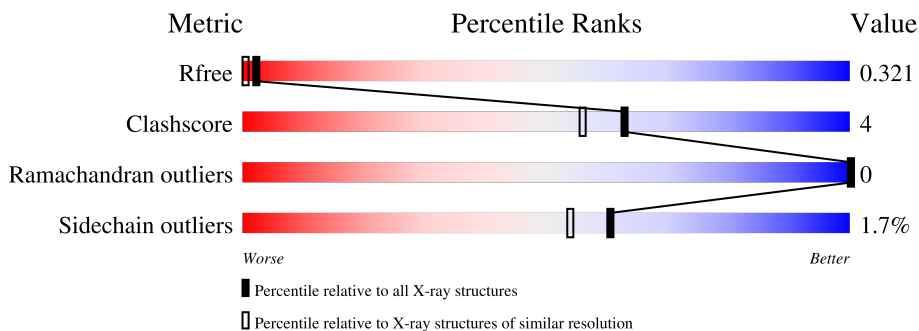|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)

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 1.88 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	Percentile Ranks	Value
Rfree		0.321
Clashscore		4
Ramachandran outliers		0
Sidechain outliers		1.7%

Worse ... *Better*

▌ Percentile relative to all X-ray structures

⫿ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	9470 (1.90-1.86)
Clashscore	141614	10282 (1.90-1.86)
Ramachandran outliers	138981	10152 (1.90-1.86)
Sidechain outliers	138945	10152 (1.90-1.86)

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 <=5%

Mol	Chain	Length	Quality of chain
1	A	163	91% 9%
1	B	163	85% 15%
1	C	163	91% 9%
1	D	163	89% 9% •

Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

WORLDWIDE
wwPDB
PROTEIN DATA BANK

2   Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 5633 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 Non-structural protein 3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	163	Total 1262	C 787	N 220	O 245	S 10	0	2	0
1	B	163	Total 1263	C 787	N 220	O 247	S 9	0	2	0
1	C	163	Total 1274	C 794	N 225	O 246	S 9	0	3	0
1	D	159	Total 1249	C 778	N 220	O 243	S 8	0	3	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP Q8JUX6
A	-1	ALA	-	expression tag	UNP Q8JUX6
A	0	MET	-	expression tag	UNP Q8JUX6
A	77	THR	SER	conflict	UNP Q8JUX6
B	-2	GLY	-	expression tag	UNP Q8JUX6
B	-1	ALA	-	expression tag	UNP Q8JUX6
B	0	MET	-	expression tag	UNP Q8JUX6
B	77	THR	SER	conflict	UNP Q8JUX6
C	-2	GLY	-	expression tag	UNP Q8JUX6
C	-1	ALA	-	expression tag	UNP Q8JUX6
C	0	MET	-	expression tag	UNP Q8JUX6
C	77	THR	SER	conflict	UNP Q8JUX6
D	-2	GLY	-	expression tag	UNP Q8JUX6
D	-1	ALA	-	expression tag	UNP Q8JUX6
D	0	MET	-	expression tag	UNP Q8JUX6
D	77	THR	SER	conflict	UNP Q8JUX6

- Molecule 2 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula: C_2H_6OS).

DMS

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total 4	C 2	O 1	S 1	0	0
2	A	1	Total 4	C 2	O 1	S 1	0	0
2	A	1	Total 4	C 2	O 1	S 1	0	0
2	A	1	Total 4	C 2	O 1	S 1	0	0
2	A	1	Total 4	C 2	O 1	S 1	0	0
2	A	1	Total 4	C 2	O 1	S 1	0	0
2	B	1	Total 4	C 2	O 1	S 1	0	0
2	B	1	Total 4	C 2	O 1	S 1	0	0
2	B	1	Total 4	C 2	O 1	S 1	0	0
2	B	1	Total 4	C 2	O 1	S 1	0	0
2	C	1	Total 4	C 2	O 1	S 1	0	0
2	C	1	Total 4	C 2	O 1	S 1	0	0
2	C	1	Total 4	C 2	O 1	S 1	0	0
2	D	1	Total 4	C 2	O 1	S 1	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	D	1	Total	C	O	S	0	0
			4	2	1	1		
2	D	1	Total	C	O	S	0	0
			4	2	1	1		
2	D	1	Total	C	O	S	0	0
			4	2	1	1		

- Molecule 3 is 2-AMINO-2-HYDROXYMETHYL-PROPANE-1,3-DIOL (three-letter code: TRS) (formula: $C_4H_{12}NO_3$).
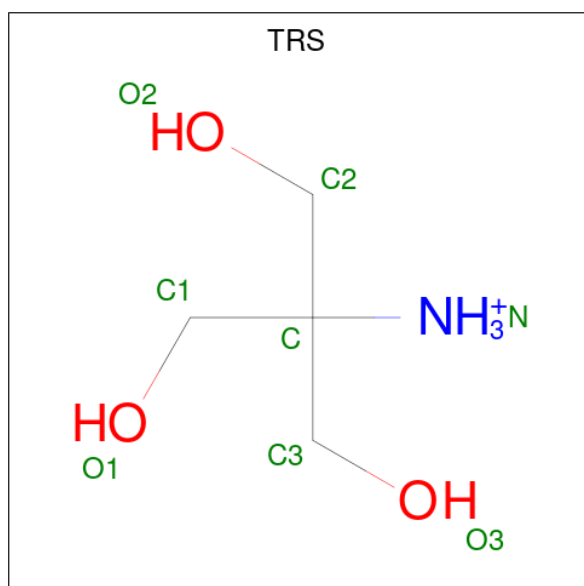
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			8	4	1	3		
3	D	1	Total	C	N	O	0	0
			8	4	1	3		

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	3	Total	Cl	0	0
			3	3		
4	B	2	Total	Cl	0	0
			2	2		
4	C	3	Total	Cl	0	0
			3	3		
4	D	2	Total	Cl	0	0
			2	2		

- Molecule 5 is 5-bromo-6-methylpyridin-2-amine (three-letter code: 4JO) (formula: $C_6H_7BrN_2$) (labeled as "Ligand of Interest" by depositor).
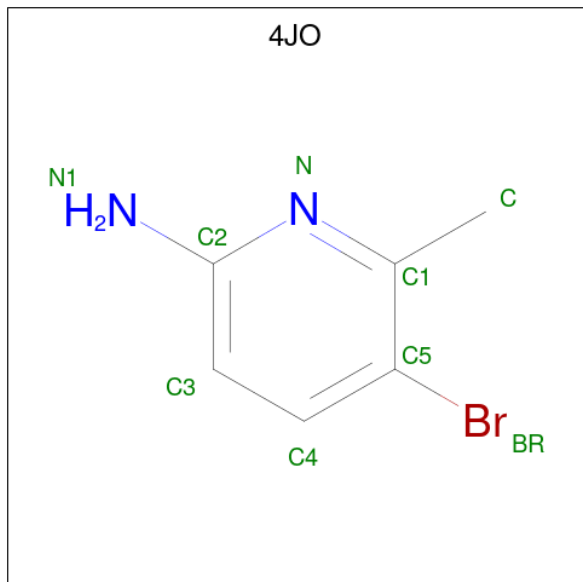
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	Br	C	N	0	0
			9	1	6	2		
5	B	1	Total	Br	C	N	0	0
			9	1	6	2		
5	C	1	Total	Br	C	N	0	0
			9	1	6	2		
5	D	1	Total	Br	C	N	0	0
			9	1	6	2		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	140	Total	O	0	1
			141	141		
6	B	101	Total	O	0	0
			101	101		
6	C	118	Total	O	0	0
			118	118		
6	D	95	Total	O	0	0
			95	95		

3 Residue-property plots ⓘ

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: Non-structural protein 3

Chain A:

• Molecule 1: Non-structural protein 3

Chain B:
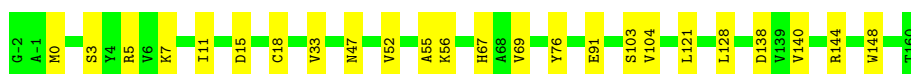
• Molecule 1: Non-structural protein 3

Chain C:

• Molecule 1: Non-structural protein 3

Chain D:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, α, β, γ	87.55Å 87.55Å 85.70Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	39.01 – 1.88 38.98 – 1.88	Depositor EDS
% Data completeness (in resolution range)	94.8 (39.01-1.88) 94.8 (38.98-1.88)	Depositor EDS
R_{merge}	0.33	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.78 (at 1.88Å)	Xtriage
Refinement program	REFMAC 5.8.0267, REFMAC5	Depositor
R, R_{free}	0.241 , 0.309 0.287 , 0.321	Depositor DCC
R_{free} test set	2869 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	15.4	Xtriage
Anisotropy	0.570	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.24 , 49.3	EDS
L-test for twinning[2]	$< |L| > = 0.52, < L^2 > = 0.35$	Xtriage
Estimated twinning fraction	0.000 for -h,-k,l 0.040 for h,-h-k,-l 0.003 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	5633	wwPDB-VP
Average B, all atoms (Å^2)	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 24.56 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 3.7427e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5   Model quality ⓘ

5.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: CL, DMS, TRS, 4JO

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	0.75	0/1285	0.88	0/1739
1	B	0.71	0/1286	0.89	0/1741
1	C	0.71	0/1297	0.81	0/1754
1	D	0.71	0/1272	0.84	0/1721
All	All	0.72	0/5140	0.86	0/6955

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 ⓘ

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	1262	0	1256	11	0
1	B	1263	0	1253	16	0
1	C	1274	0	1272	10	0
1	D	1249	0	1240	8	0
2	A	24	0	36	5	0
2	B	16	0	24	0	0
2	C	12	0	18	0	0
2	D	16	0	24	0	0
3	A	8	0	12	0	0

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	D	8	0	12	0	0
4	A	3	0	0	0	0
4	B	2	0	0	0	0
4	C	3	0	0	0	0
4	D	2	0	0	0	0
5	A	9	0	7	1	0
5	B	9	0	7	1	0
5	C	9	0	7	1	0
5	D	9	0	7	1	0
6	A	141	0	0	1	0
6	B	101	0	0	3	0
6	C	118	0	0	3	0
6	D	95	0	0	1	0
All	All	5633	0	5175	45	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 4.

The worst 5 of 45 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:B:91[A]:GLU:OE2	6:B:301:HOH:O	1.93	0.86
1:C:80:GLU:OE1	6:C:301:HOH:O	2.11	0.67
1:C:144:ARG:HG3	5:C:207:4JO:BR	2.51	0.66
1:B:121:LEU:HD12	1:B:148:TRP:CD2	2.31	0.65
1:A:144:ARG:HG2	5:A:210:4JO:BR	2.58	0.58

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	163/163 (100%)	161 (99%)	2 (1%)	0	100	100
1	B	163/163 (100%)	160 (98%)	3 (2%)	0	100	100
1	C	164/163 (101%)	160 (98%)	4 (2%)	0	100	100
1	D	160/163 (98%)	158 (99%)	2 (1%)	0	100	100
All	All	650/652 (100%)	639 (98%)	11 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	136/134 (102%)	134 (98%)	2 (2%)	65	59
1	B	136/134 (102%)	134 (98%)	2 (2%)	65	59
1	C	137/134 (102%)	133 (97%)	4 (3%)	42	32
1	D	136/134 (102%)	135 (99%)	1 (1%)	84	83
All	All	545/536 (102%)	536 (98%)	9 (2%)	60	54

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	159	ARG
1	D	145	ASP
1	B	15	ASP
1	C	0	MET
1	C	31	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	157	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 33 ligands modelled in this entry, 10 are monoatomic - leaving 23 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	$\#|Z| > 2$
2	DMS	B	201	-	3,3,3	0.22	0	3,3,3	0.13	0
5	4JO	A	210	-	8,9,9	0.18	0	7,12,12	0.43	0
2	DMS	D	203	-	3,3,3	0.26	0	3,3,3	0.14	0
2	DMS	A	206	-	3,3,3	0.35	0	3,3,3	0.05	0
2	DMS	B	202	-	3,3,3	0.22	0	3,3,3	0.09	0
2	DMS	A	202	-	3,3,3	0.12	0	3,3,3	0.14	0
2	DMS	D	205	-	3,3,3	0.30	0	3,3,3	0.05	0
2	DMS	C	201	-	3,3,3	0.25	0	3,3,3	0.08	0
5	4JO	D	208	-	8,9,9	0.19	0	7,12,12	0.38	0
5	4JO	C	207	-	8,9,9	0.20	0	7,12,12	0.36	0
3	TRS	A	204	-	7,7,7	0.25	0	9,9,9	0.34	0
2	DMS	B	203	-	3,3,3	0.19	0	3,3,3	0.12	0
3	TRS	D	202	-	7,7,7	0.18	0	9,9,9	0.23	0
2	DMS	A	203	-	3,3,3	0.28	0	3,3,3	0.14	0
2	DMS	D	204	-	3,3,3	0.21	0	3,3,3	0.12	0
2	DMS	D	201	-	3,3,3	0.22	0	3,3,3	0.12	0
2	DMS	A	205	-	3,3,3	0.32	0	3,3,3	0.15	0
5	4JO	B	207	-	8,9,9	0.20	0	7,12,12	0.39	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	DMS	C	205	-	3,3,3	0.27	0	3,3,3	0.03	0
2	DMS	C	206	-	3,3,3	0.27	0	3,3,3	0.05	0
2	DMS	A	201	-	3,3,3	0.35	0	3,3,3	0.15	0
2	DMS	B	204	-	3,3,3	0.24	0	3,3,3	0.10	0
2	DMS	A	211	-	3,3,3	0.27	0	3,3,3	0.12	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
5	4JO	D	208	-	-	-	0/1/1/1
5	4JO	C	207	-	-	-	0/1/1/1
5	4JO	A	210	-	-	-	0/1/1/1
3	TRS	A	204	-	-	2/9/9/9	-
3	TRS	D	202	-	-	4/9/9/9	-
5	4JO	B	207	-	-	-	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	D	202	TRS	N-C-C3-O3
3	D	202	TRS	C3-C-C1-O1
3	D	202	TRS	N-C-C1-O1
3	A	204	TRS	C2-C-C3-O3
3	D	202	TRS	C2-C-C1-O1

There are no ring outliers.

7 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	210	4JO	1	0
2	A	206	DMS	1	0
2	A	202	DMS	3	0
5	D	208	4JO	1	0
5	C	207	4JO	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	205	DMS	1	0
5	B	207	4JO	1	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand 4JO A 210

Bond lengths   Bond angles

Torsions   Rings

Ligand 4JO D 208

Bond lengths

Bond angles

Torsions

Rings

Ligand 4JO C 207

Bond lengths

Bond angles

Torsions

Rings

Ligand 4JO B 207

Bond lengths

Bond angles

Torsions

Rings
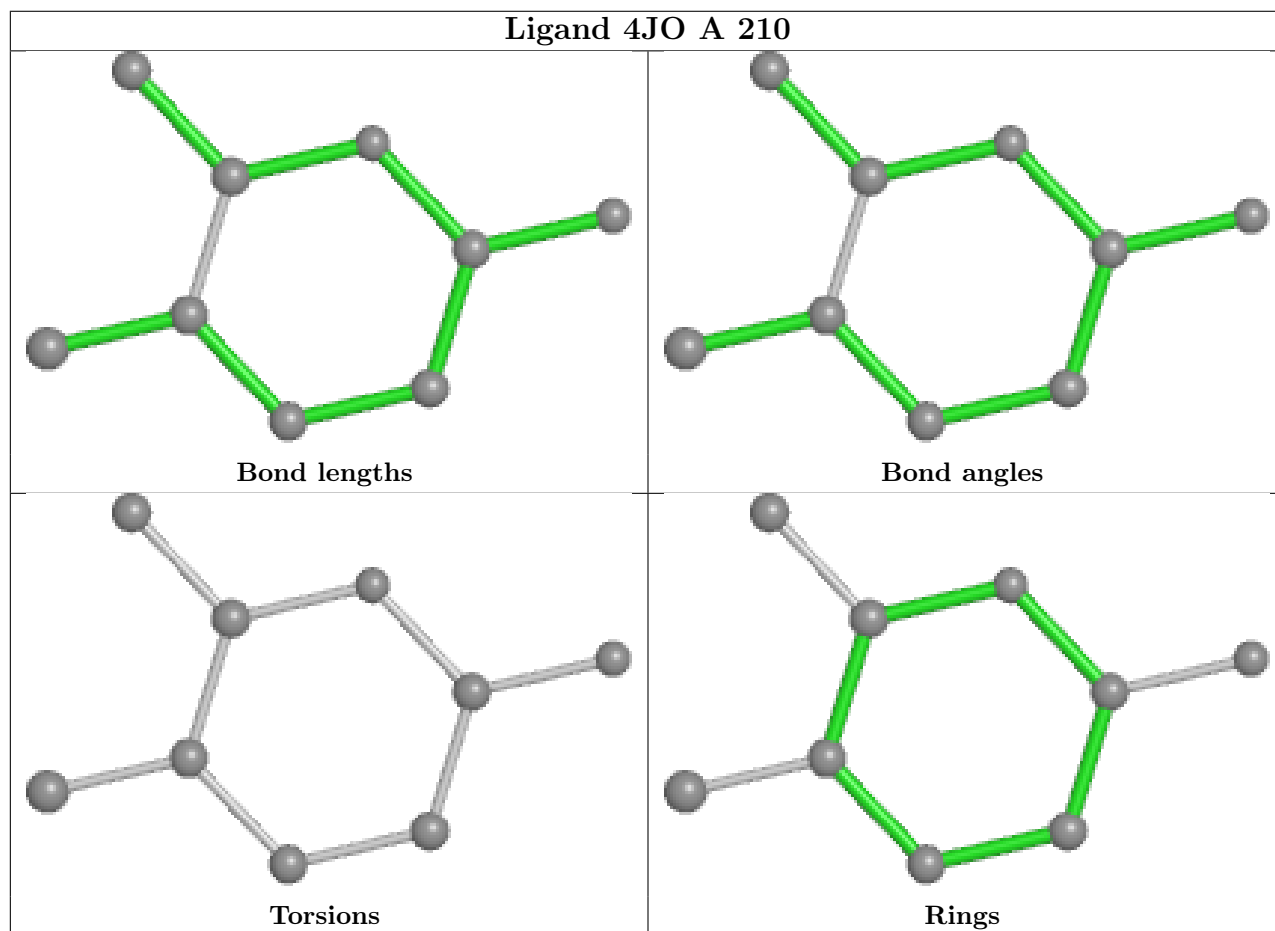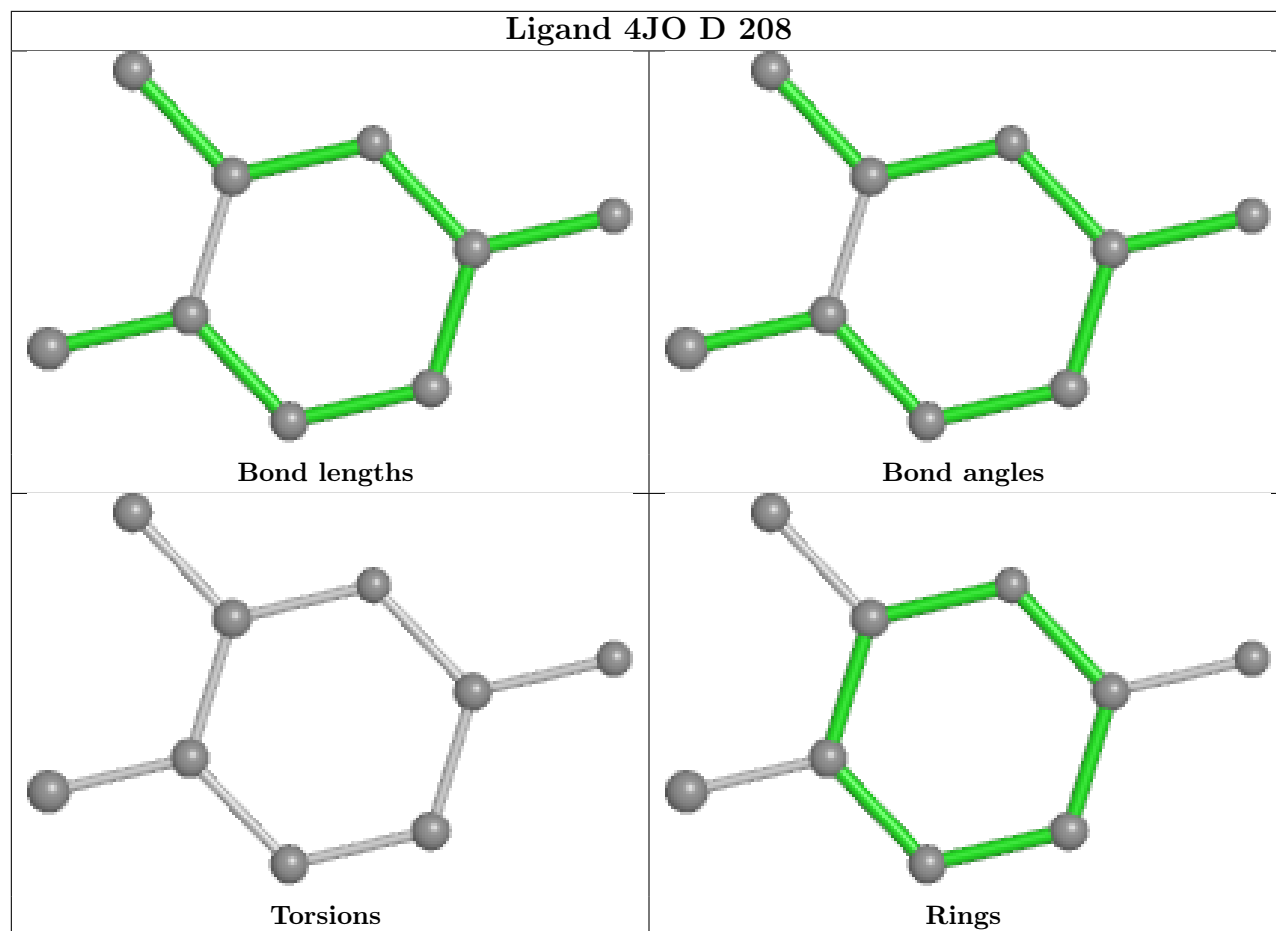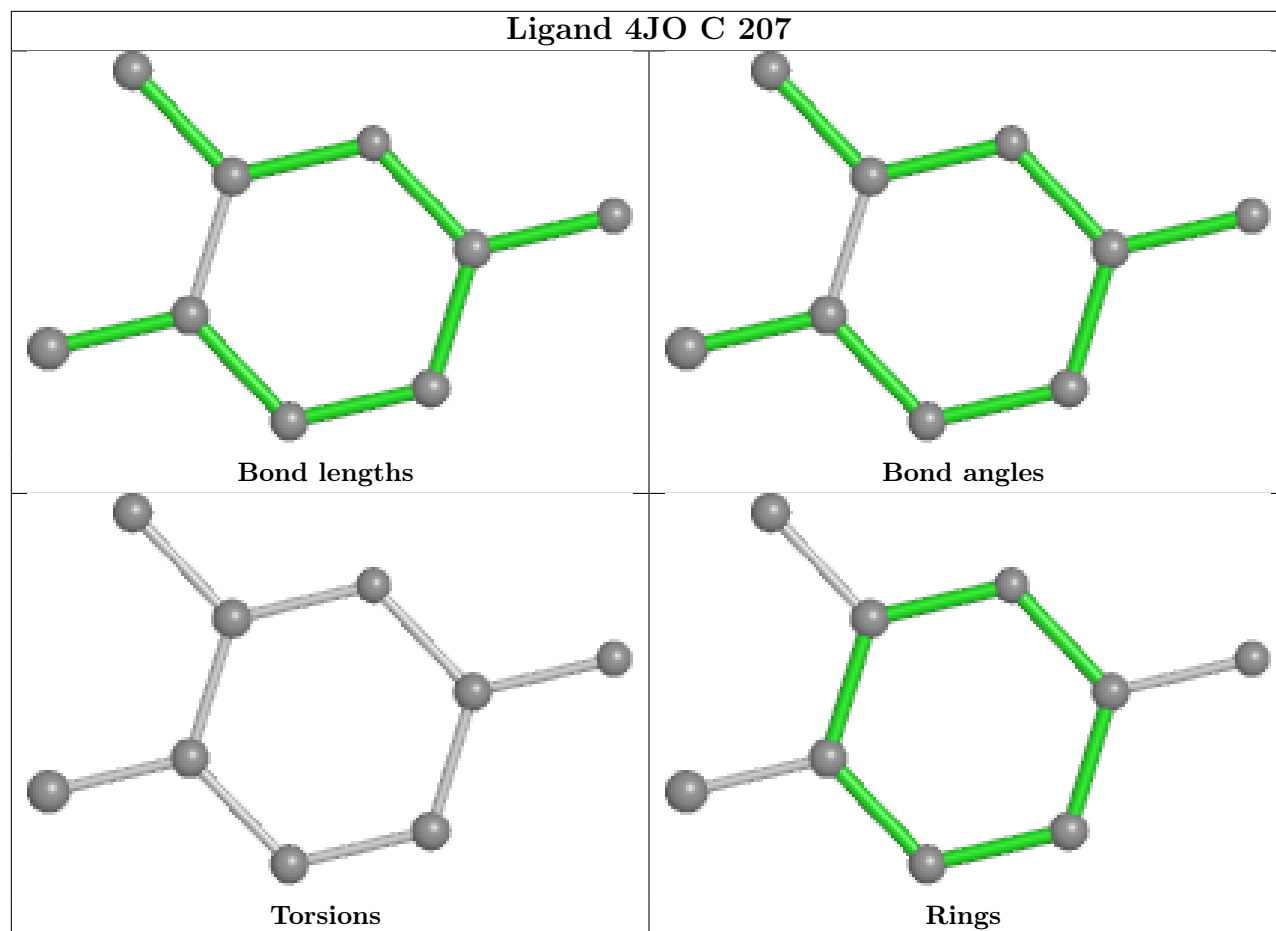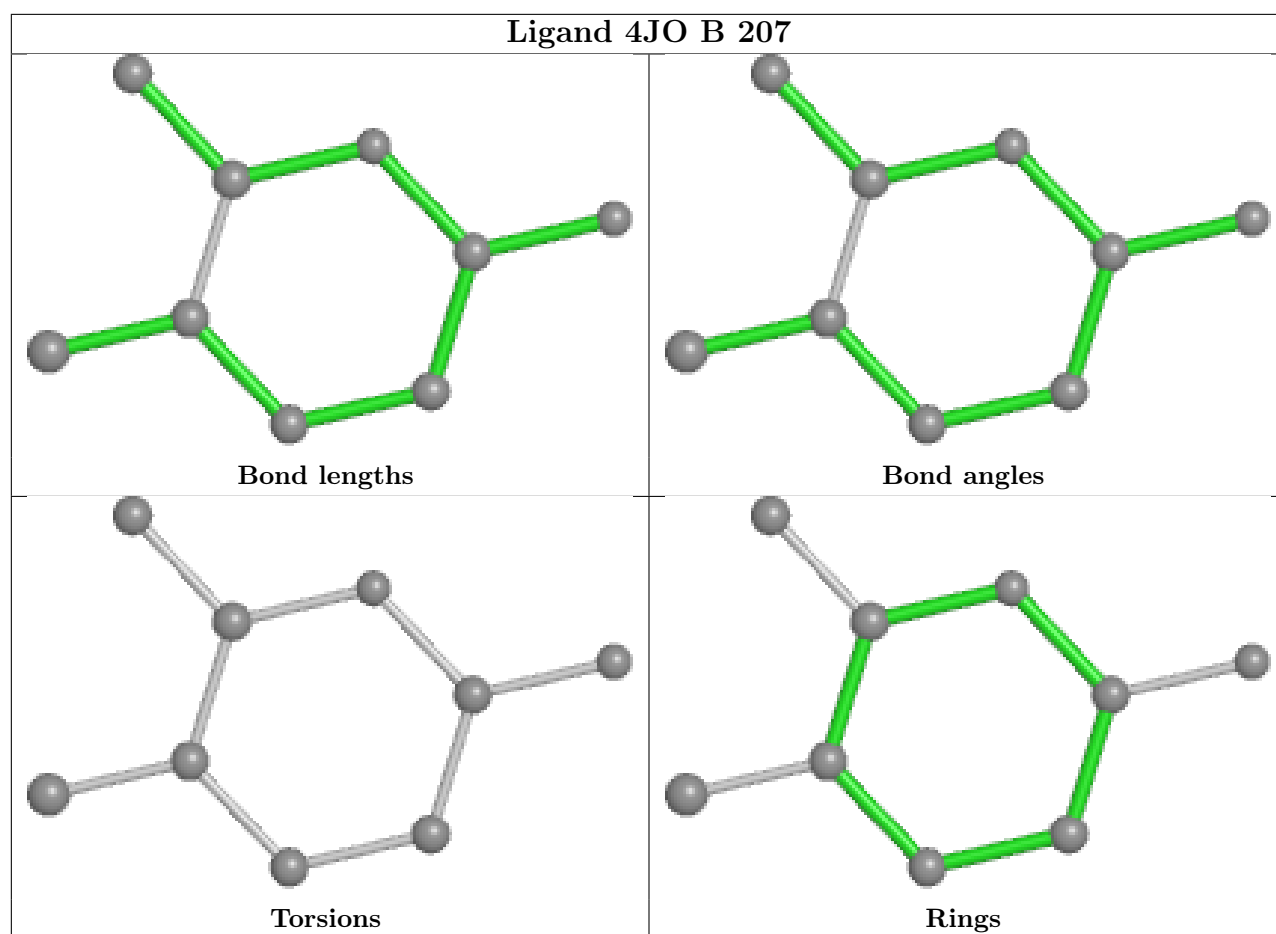
5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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.

Electron density around 4JO A 210:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)
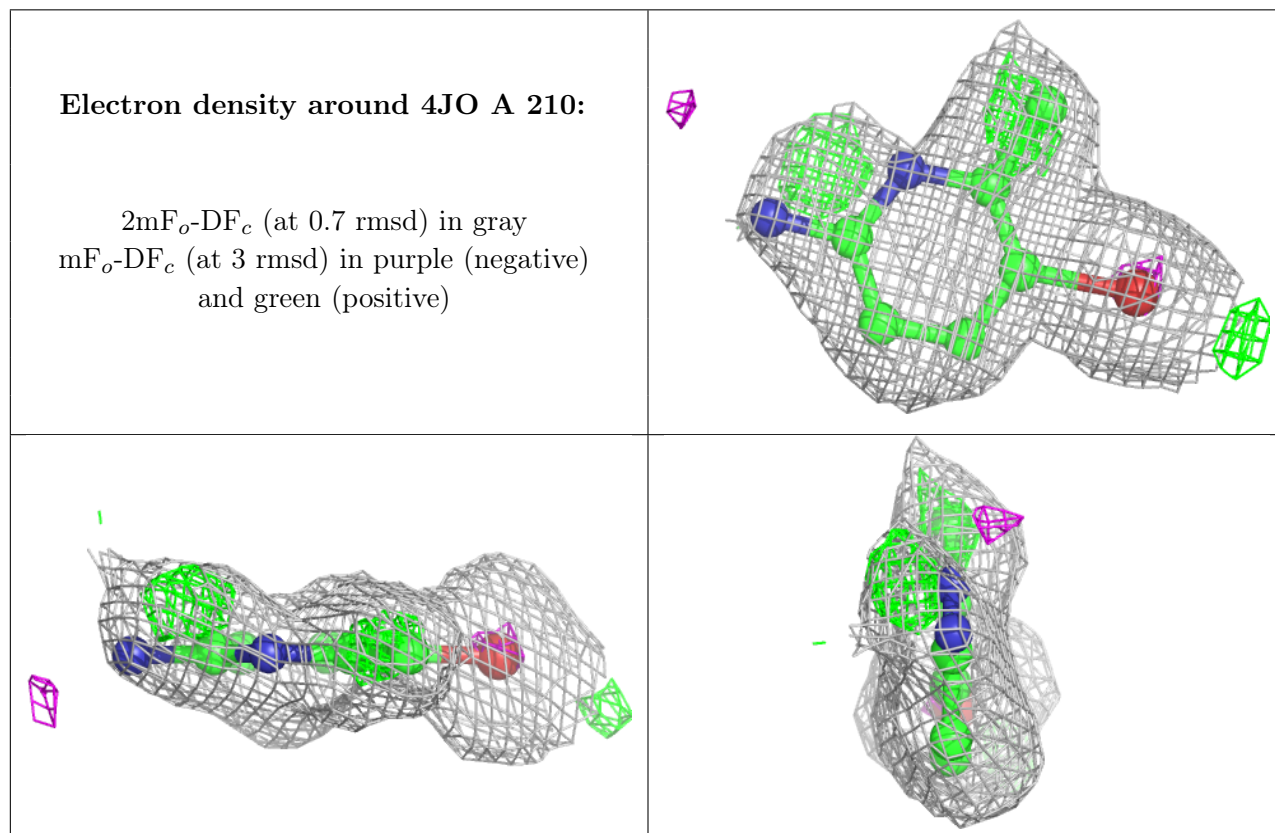
Electron density around 4JO B 207:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)
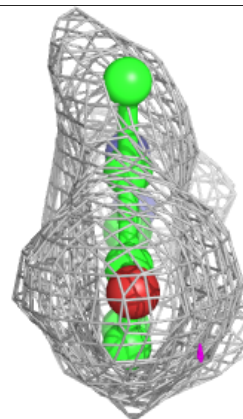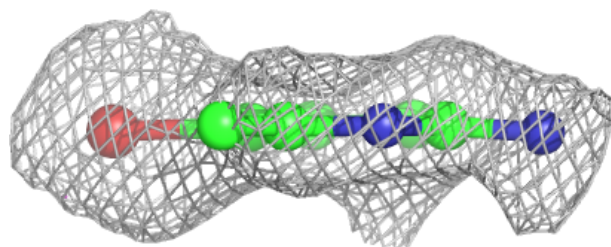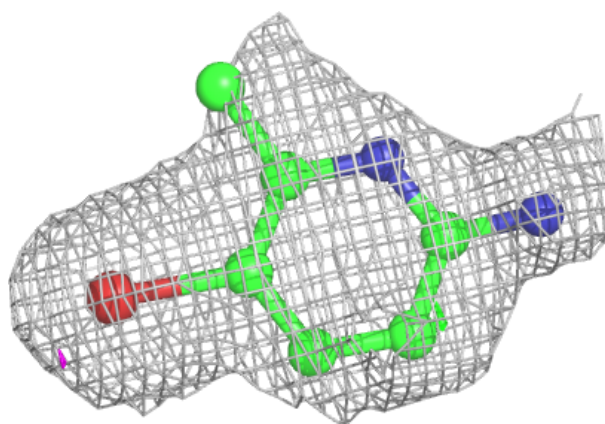
Electron density around 4JO C 207:

2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative)
and green (positive)

Electron density around 4JO D 208:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)
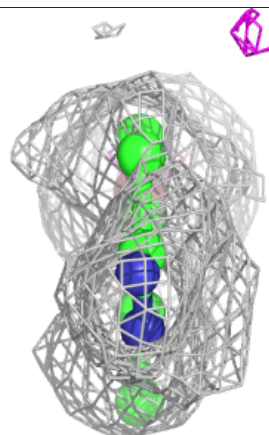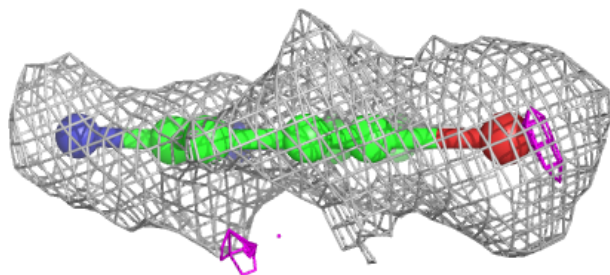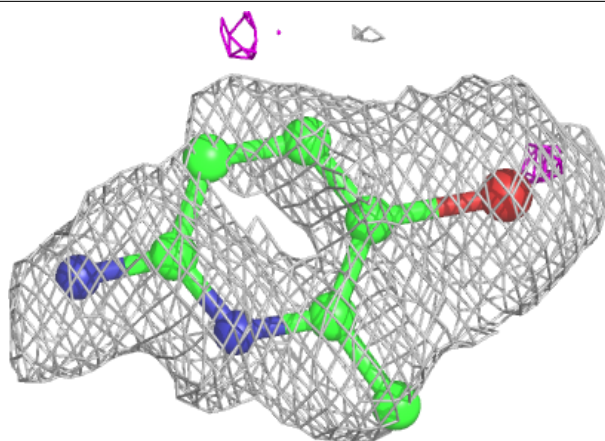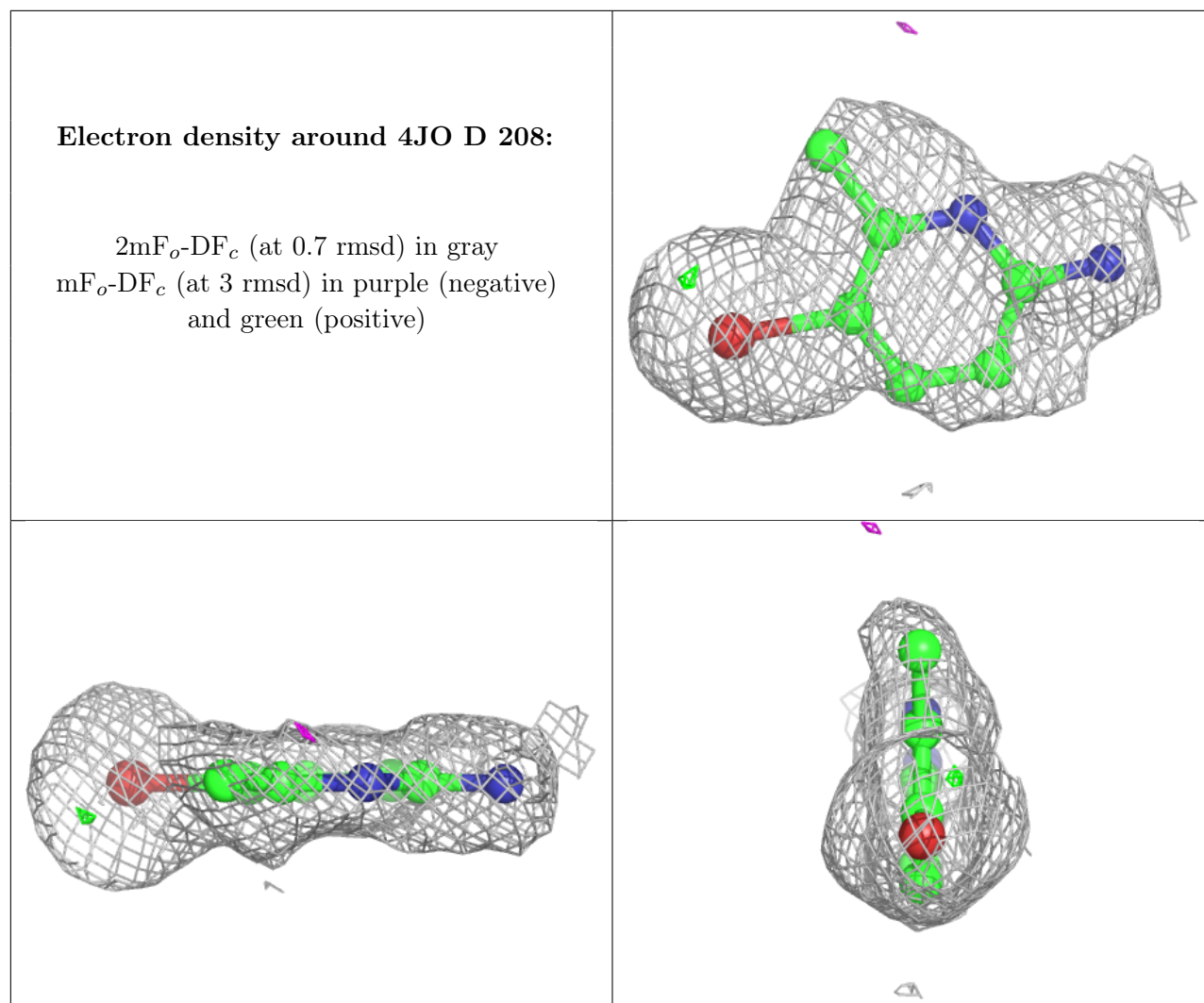
6.5 Other polymers (i)

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