WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Aug 16, 2023 – 12:10 PM EDT

PDB ID	:	2A5I
Title	:	Crystal structures of SARS coronavirus main peptidase inhibited by an aza-peptide epoxide in the space group C2
Authors	:	Lee, T.-W.; Cherney, M.M.; Huitema, C.; Liu, J.; James, K.E.; Powers, J.C.; Eltis, L.D.; James, M.N.
Deposited on	:	2005-06-30
Resolution	:	1.88 Å(reported)

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

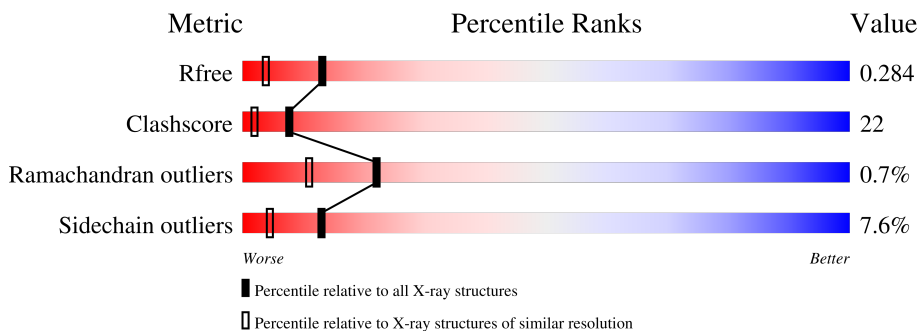|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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		0.284
Clashscore		22
Ramachandran outliers		0.7%
Sidechain outliers		7.6%

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	306	67% 28% • •

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

| Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	EDO	A	308[A]	-	-	X	-
4	GOL	A	309	-	X	X	-

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 2752 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 3C-like peptidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	306	Total 2436	C 1544	N 411	O 454	S 27	0	15	0

- Molecule 2 is (5S,8S,14R)-ETHYL 11-(3-AMINO-3-OXOPROPYL)-8-BENZYL-14-HYDR OXY-5-ISOBUTYL-3,6,9,12-TETRAOXO-1-PHENYL-2-OXA-4,7,10,11-TETRAAZAPEN TADECAN-15-OATE (three-letter code: AZP) (formula: $C_{32}H_{43}N_5O_9$).
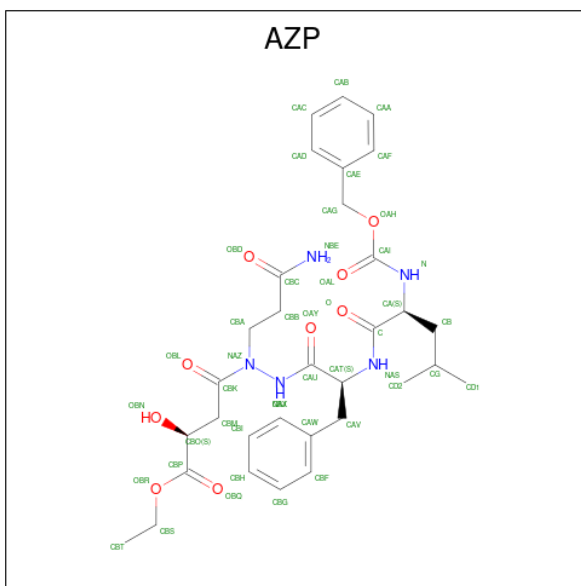
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total 46	C 32	N 5	O 9	0	0

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).

EDO

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 8	C 4	O 4	0	1

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

GOL

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

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	256	Total 256	O 256	0	0

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: 3C-like peptidase
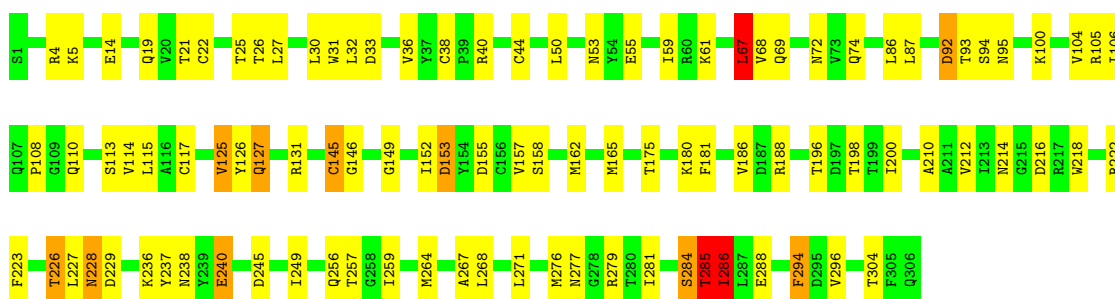
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	106.70Å 83.68Å 52.86Å 90.00° 105.65° 90.00°	Depositor
Resolution (Å)	40.00 – 1.88 32.32 – 1.87	Depositor EDS
% Data completeness (in resolution range)	94.2 (40.00-1.88) 93.4 (32.32-1.87)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.17 (at 1.87Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.198 , 0.242 0.263 , 0.284	Depositor DCC
R_{free} test set	1730 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	32.8	Xtriage
Anisotropy	0.652	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 55.9	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	2752	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.41% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[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: GOL, AZP, EDO

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	1.05	7/2544 (0.3%)	1.12	8/3451 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	2

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	240	GLU	CB-CG	-6.86	1.39	1.52
1	A	117	CYS	CB-SG	-6.77	1.70	1.82
1	A	288	GLU	CB-CG	-6.02	1.40	1.52
1	A	44	CYS	CB-SG	-5.97	1.72	1.81
1	A	38	CYS	CB-SG	5.14	1.91	1.82

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	92	ASP	CB-CA-C	-8.93	92.55	110.40
1	A	67	LEU	CA-CB-CG	7.53	132.61	115.30
1	A	67	LEU	CB-CG-CD2	-6.30	100.29	111.00
1	A	285	THR	N-CA-C	-5.84	95.24	111.00
1	A	115	LEU	CB-CG-CD1	5.83	120.91	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	146	GLY	Mainchain
1	A	92	ASP	Mainchain

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	2436	0	2391	103	0
2	A	46	0	42	5	0
3	A	8	0	12	5	0
4	A	6	0	7	6	0
5	A	256	0	0	15	1
All	All	2752	0	2452	108	1

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 22.

The worst 5 of 108 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:86:LEU:CD2	1:A:162[B]:MET:CE	2.23	1.17
1:A:264[B]:MET:HE3	1:A:267:ALA:HB3	1.34	1.10
1:A:86:LEU:HD22	1:A:162[B]:MET:HE1	1.24	1.09
1:A:276:MET:HE3	1:A:281:ILE:HD12	1.09	1.09
1:A:86:LEU:HD21	1:A:162[B]:MET:HE2	1.09	1.07

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:A:357:HOH:O	5:A:515:HOH:O[2_556]	2.15	0.05

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	318/306 (104%)	307 (96%)	9 (3%)	2 (1%)	25	14

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	277	ASN
1	A	286	ILE

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	278/263 (106%)	255 (92%)	23 (8%)	11	3

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	227	LEU
1	A	268	LEU
1	A	236	LYS
1	A	284	SER
1	A	127[A]	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	19	GLN
1	A	119	ASN

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 ⓘ

4 ligands are modelled in this entry.

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	#$|Z| > 2$
4	GOL	A	309	3	5,5,5	1.24	1 (20%)	5,5,5	1.88	2 (40%)
3	EDO	A	308[A]	4	3,3,3	0.56	0	2,2,2	0.14	0
2	AZP	A	307	1	45,47,47	1.68	3 (6%)	56,61,61	1.69	14 (25%)
3	EDO	A	308[B]	-	3,3,3	0.49	0	2,2,2	0.55	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
4	GOL	A	309	3	-	4/4/4/4	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	EDO	A	308[A]	4	-	1/1/1/1	-
2	AZP	A	307	1	-	8/51/53/53	0/2/2/2
3	EDO	A	308[B]	-	-	1/1/1/1	-

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	307	AZP	NAX-NAZ	6.09	1.47	1.40
2	A	307	AZP	OAH-CAI	5.64	1.46	1.35
2	A	307	AZP	OBR-CBP	5.15	1.43	1.33
4	A	309	GOL	O2-C2	-2.68	1.35	1.43

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	307	AZP	CAT-CAU-NAX	4.95	121.35	114.22
2	A	307	AZP	OBR-CBP-OBQ	-3.37	117.74	124.13
2	A	307	AZP	CBB-CBA-NAZ	-3.31	107.33	111.48
2	A	307	AZP	OBN-CBO-CBM	-3.02	102.63	110.05
2	A	307	AZP	OAH-CAI-N	2.92	116.44	110.50

There are no chirality outliers.

5 of 14 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	307	AZP	CBM-CBK-NAZ-CBA
2	A	307	AZP	OBN-CBO-CBP-OBR
4	A	309	GOL	O1-C1-C2-C3
4	A	309	GOL	C1-C2-C3-O3
2	A	307	AZP	OBL-CBK-NAZ-CBA

There are no ring outliers.

4 monomers are involved in 16 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	309	GOL	6	0
3	A	308[A]	EDO	4	0
2	A	307	AZP	5	0
3	A	308[B]	EDO	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 AZP A 307

Bond lengths

Bond angles

Torsions

Rings
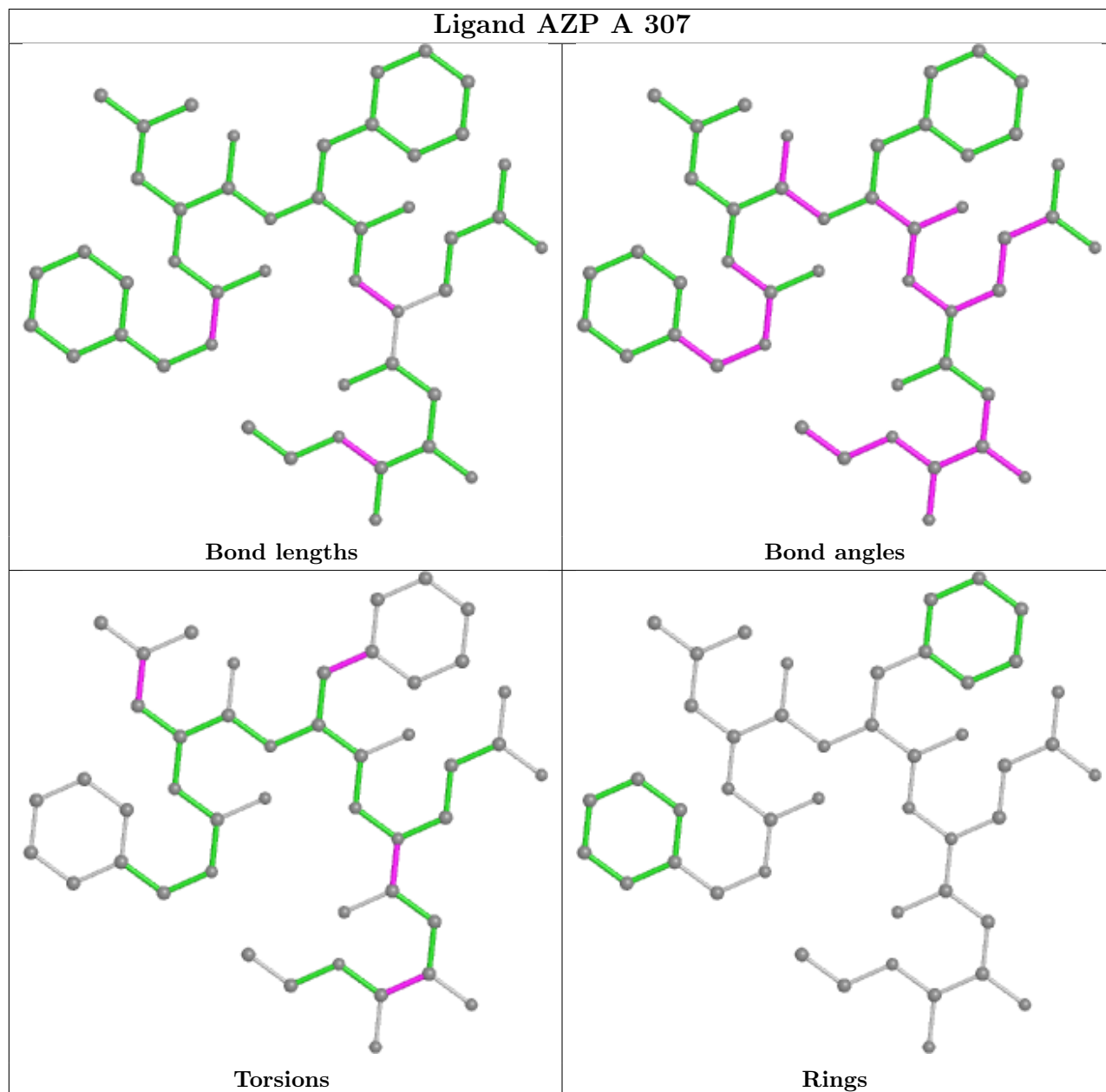
5.7 Other polymers ⓘ

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 ⓘ

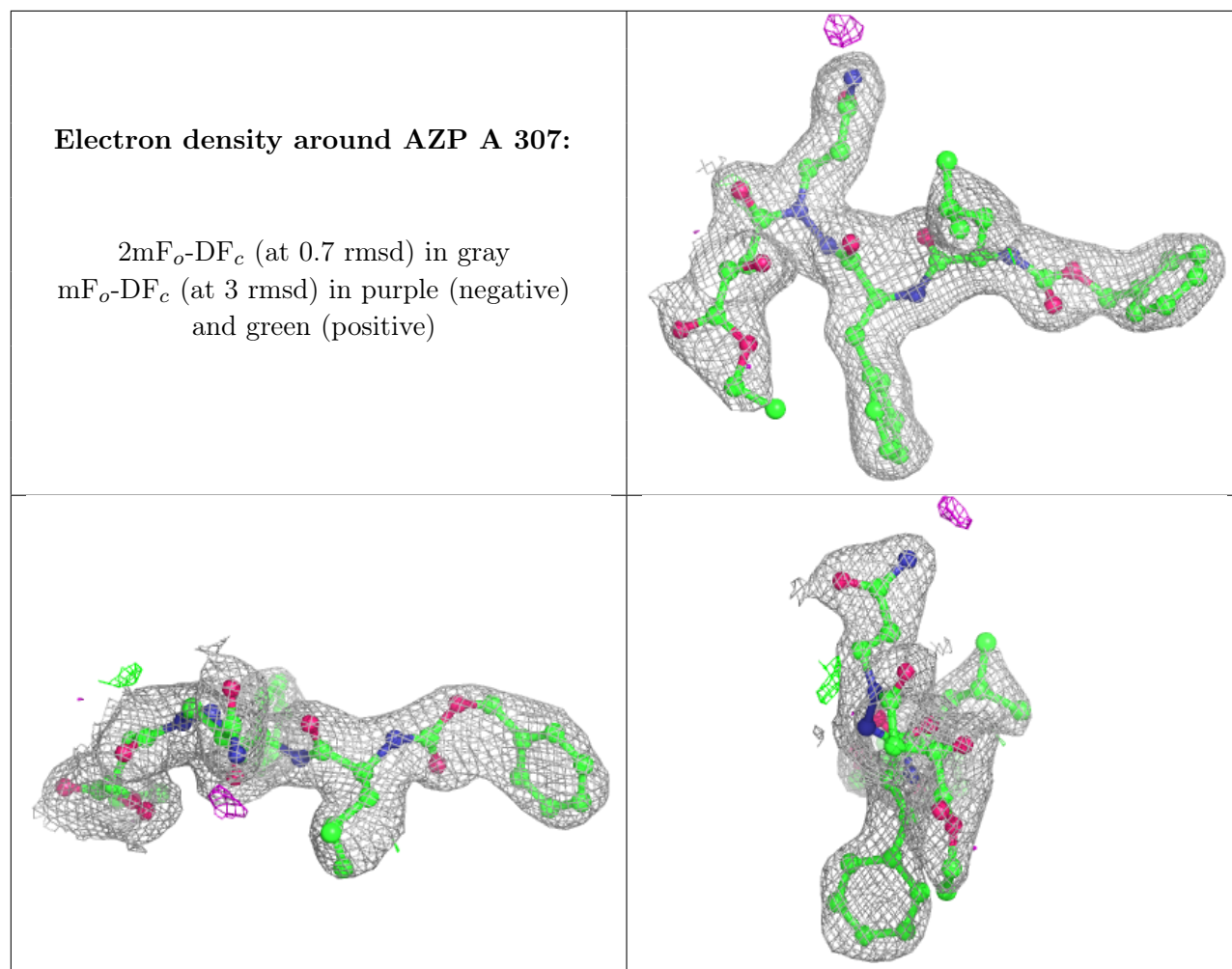Unable to reproduce the depositors R factor - this section is therefore empty.

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 AZP A 307:

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

6.5 Other polymers ⓘ

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