WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Oct 4, 2023 – 05:24 PM EDT

PDB ID	:	6P6Q
Title	:	HCV NS3/4A protease domain of genotype 1a3a chimera in complex with grazoprevir
Authors	:	Timm, J.; Schiffer, C.A.
Deposited on	:	2019-06-04
Resolution	:	3.50 Å(reported)

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

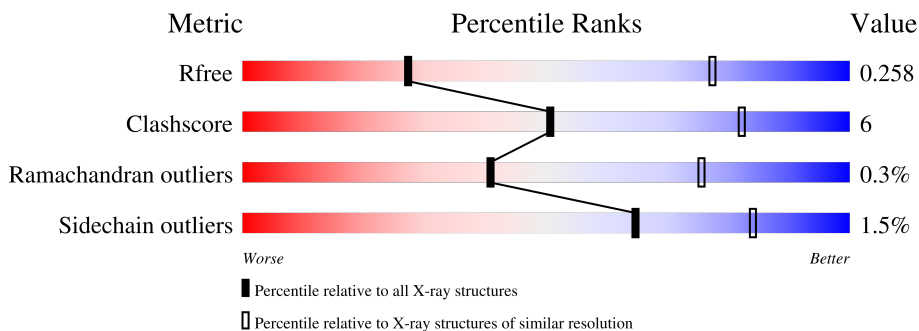|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 3.50 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		0.258
Clashscore		6
Ramachandran outliers		0.3%
Sidechain outliers		1.5%

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	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)

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	203	87% 9% •
1	B	203	85% 11% •

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
2	SUE	A	1301	X	-	-	-
2	SUE	B	1301	X	-	-	-

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 5399 atoms, of which 2461 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 4A,Serine protease NS3.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	197	Total	C	H	N	O	S	0	0	0
			2607	862	1224	245	269	7			
1	B	196	Total	C	H	N	O	S	0	0	0
			2445	836	1111	239	254	5			

There are 64 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	980	GLY	-	expression tag	UNP P26664
A	981	SER	-	expression tag	UNP P26664
A	982	HIS	-	expression tag	UNP P26664
A	983	MET	-	expression tag	UNP P26664
A	984	ALA	-	expression tag	UNP P26664
A	985	SER	-	expression tag	UNP P26664
A	986	MET	-	expression tag	UNP P26664
A	987	LYS	-	expression tag	UNP P26664
A	988	LYS	-	expression tag	UNP P26664
A	989	LYS	-	expression tag	UNP P26664
A	991	SER	CYS	engineered mutation	UNP P26664
A	998	ILE	VAL	engineered mutation	UNP P26664
A	999	ASN	VAL	engineered mutation	UNP P26664
A	1001	SER	-	linker	UNP P26664
A	1002	GLY	-	linker	UNP P26664
A	1003	ASP	-	linker	UNP P26664
A	1013	GLU	LEU	engineered mutation	UNP P26664
A	1014	GLU	LEU	engineered mutation	UNP P26664
A	1017	GLN	ILE	engineered mutation	UNP P26664
A	1018	GLU	ILE	engineered mutation	UNP P26664
A	1021	GLN	LEU	engineered mutation	UNP P26664
A	1040	THR	ALA	engineered mutation	UNP P26664
A	1047	SER	CYS	engineered mutation	UNP P26664
A	1052	LEU	CYS	engineered mutation	UNP P26664
A	1072	THR	ILE	engineered mutation	UNP P26664

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	1080	LYS	GLN	engineered mutation	UNP P26664
A	1086	GLN	PRO	engineered mutation	UNP P26664
A	1123	THR	ARG	engineered mutation	UNP P26664
A	1132	LEU	ILE	engineered mutation	UNP P26664
A	1159	SER	CYS	engineered mutation	UNP P26664
A	1168	GLN	ASP	engineered mutation	UNP P26664
A	1174	SER	ASN	engineered mutation	UNP P26664
B	981	GLY	-	expression tag	UNP P26664
B	982	SER	-	expression tag	UNP P26664
B	983	HIS	-	expression tag	UNP P26664
B	984	MET	-	expression tag	UNP P26664
B	985	ALA	-	expression tag	UNP P26664
B	985A	SER	-	expression tag	UNP P26664
B	985B	MET	-	expression tag	UNP P26664
B	987	LYS	-	expression tag	UNP P26664
B	988	LYS	-	expression tag	UNP P26664
B	989	LYS	-	expression tag	UNP P26664
B	991	SER	CYS	engineered mutation	UNP P26664
B	998	ILE	VAL	engineered mutation	UNP P26664
B	999	ASN	VAL	engineered mutation	UNP P26664
B	1001	SER	-	linker	UNP P26664
B	1002	GLY	-	linker	UNP P26664
B	1003	ASP	-	linker	UNP P26664
B	1013	GLU	LEU	engineered mutation	UNP P26664
B	1014	GLU	LEU	engineered mutation	UNP P26664
B	1017	GLN	ILE	engineered mutation	UNP P26664
B	1018	GLU	ILE	engineered mutation	UNP P26664
B	1021	GLN	LEU	engineered mutation	UNP P26664
B	1040	THR	ALA	engineered mutation	UNP P26664
B	1047	SER	CYS	engineered mutation	UNP P26664
B	1052	LEU	CYS	engineered mutation	UNP P26664
B	1072	THR	ILE	engineered mutation	UNP P26664
B	1080	LYS	GLN	engineered mutation	UNP P26664
B	1086	GLN	PRO	engineered mutation	UNP P26664
B	1123	THR	ARG	engineered mutation	UNP P26664
B	1132	LEU	ILE	engineered mutation	UNP P26664
B	1159	SER	CYS	engineered mutation	UNP P26664
B	1168	GLN	ASP	engineered mutation	UNP P26664
B	1174	SER	ASN	engineered mutation	UNP P26664

- Molecule 2 is (1aR,5S,8S,10R,22aR)-5-tert-butyl-N-{(1R,2S)-1-[(cyclopropylsulfonyl)carba moyl]-2-ethenylcyclopropyl}-14-methoxy-3,6-di oxo-1,1a,3,4,5,6,9,10,18,19,20,21,22,22a-tetr adecahydro-8H-7,10-methanocyclopropa[18,19][1,10,3,6]dioxadiazacyclononadec ino[11,12-

b]quinoxaline-8-carboxamide (three-letter code: SUE) (formula: $C_{38}H_{50}N_6O_9S$) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
2	A	1	Total	C	H	N	O	S	0	0
			104	38	50	6	9	1		
2	B	1	Total	C	H	N	O	S	0	0
			104	38	50	6	9	1		

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Zn	0	0
			1	1		
3	B	1	Total	Zn	0	0
			1	1		

- Molecule 4 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: $C_6H_{13}NO_4S$).

MES

O1

C2　　　　C6

C3　　H　　C5
　　　N+
　　　N4

　　　C7
C8

S
O2S　O＝S＝O O1S

O⁻
O3S

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
4	A	1	Total	C	H	N	O	S	0	0
			25	6	13	1	4	1		
4	A	1	Total	C	H	N	O	S	0	0
			25	6	13	1	4	1		

• Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O_4S).

SO4

O1

O

O3 O⁻ — S — O⁻ O4
　　　　　S

O

O2

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	O	S	0	0
			5	4	1		
5	A	1	Total	O	S	0	0
			5	4	1		

Continued on next page...

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

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total 5	O 4	S 1	0	0
5	B	1	Total 5	O 4	S 1	0	0
5	B	1	Total 5	O 4	S 1	0	0
5	B	1	Total 5	O 4	S 1	0	0
5	B	1	Total 5	O 4	S 1	0	0

• Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	28	Total 28	O 28	0	0
6	B	24	Total 24	O 24	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: Non-structural protein 4A,Serine protease NS3

Chain A:

• Molecule 1: Non-structural protein 4A,Serine protease NS3

Chain B:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	49.75Å 58.97Å 65.89Å 90.00° 100.27° 90.00°	Depositor
Resolution (Å)	48.95 – 3.50 48.95 – 2.16	Depositor EDS
% Data completeness (in resolution range)	99.8 (48.95-3.50) 75.0 (48.95-2.16)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.07 (at 2.16Å)	Xtriage
Refinement program	PHENIX (1.14_3260: ???)	Depositor
R, R_{free}	0.220 , 0.258 0.221 , 0.258	Depositor DCC
R_{free} test set	1013 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	24.8	Xtriage
Anisotropy	0.167	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 46.3	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.86	EDS
Total number of atoms	5399	wwPDB-VP
Average B, all atoms (Å^2)	34.0	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 26.10 % 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 2.8029e-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: SUE, ZN, MES, SO4

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	0.25	0/1406	0.46	0/1920
1	B	0.25	0/1357	0.46	0/1862
All	All	0.25	0/2763	0.46	0/3782

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	1383	1224	1327	13	0
1	B	1334	1111	1250	17	0
2	A	54	50	48	2	0
2	B	54	50	48	4	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	24	26	24	2	0
5	A	15	0	0	0	0
5	B	20	0	0	2	0
6	A	28	0	0	2	0
6	B	24	0	0	5	0
All	All	2938	2461	2697	32	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 6.

The worst 5 of 32 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:990:GLY:N	4:A:1303:MES:O3S	2.06	0.88
1:B:1062:ARG:NH2	5:B:1306:SO4:O2	2.09	0.84
4:A:1304:MES:O3S	6:A:1401:HOH:O	1.96	0.82
1:A:1155:ARG:NH2	1:A:1168:GLN:OE1	2.19	0.76
1:A:1042:THR:OG1	1:A:1109:ARG:NH1	2.24	0.70

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	195/203 (96%)	183 (94%)	11 (6%)	1 (0%)	29	68
1	B	192/203 (95%)	181 (94%)	11 (6%)	0	100	100
All	All	387/406 (95%)	364 (94%)	22 (6%)	1 (0%)	41	75

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	990	GLY

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	138/165 (84%)	136 (99%)	2 (1%)	67	85
1	B	126/165 (76%)	124 (98%)	2 (2%)	62	83
All	All	264/330 (80%)	260 (98%)	4 (2%)	65	84

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	983	MET
1	A	1174	SER
1	B	983	HIS
1	B	1132	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 13 ligands modelled in this entry, 2 are monoatomic - leaving 11 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
4	MES	A	1303	-	12,12,12	2.17	1 (8%)	14,16,16	1.69	3 (21%)
5	SO4	A	1305	-	4,4,4	0.14	0	6,6,6	0.05	0
5	SO4	B	1303	-	4,4,4	0.14	0	6,6,6	0.05	0
5	SO4	B	1304	-	4,4,4	0.14	0	6,6,6	0.05	0
5	SO4	A	1306	-	4,4,4	0.14	0	6,6,6	0.05	0
5	SO4	B	1306	-	4,4,4	0.17	0	6,6,6	0.10	0
2	SUE	B	1301	-	54,60,60	4.22	23 (42%)	73,92,92	3.41	13 (17%)
5	SO4	A	1307	-	4,4,4	0.17	0	6,6,6	0.09	0
5	SO4	B	1305	-	4,4,4	0.14	0	6,6,6	0.05	0
4	MES	A	1304	-	12,12,12	2.19	1 (8%)	14,16,16	1.42	2 (14%)
2	SUE	A	1301	-	54,60,60	4.25	24 (44%)	73,92,92	3.47	11 (15%)

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	MES	A	1304	-	-	4/6/14/14	0/1/1/1
2	SUE	B	1301	-	1/1/19/19	5/61/91/91	0/6/7/7
4	MES	A	1303	-	-	5/6/14/14	0/1/1/1
2	SUE	A	1301	-	1/1/19/19	6/61/91/91	0/6/7/7

The worst 5 of 49 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1301	SUE	CBF-NBH	11.77	1.47	1.32
2	A	1301	SUE	CBF-NBH	11.68	1.47	1.32
2	A	1301	SUE	CAE-CBJ	9.80	1.54	1.37
2	A	1301	SUE	CAE-CBG	9.63	1.56	1.41
2	B	1301	SUE	CAE-CBJ	9.59	1.54	1.37

The worst 5 of 29 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	1301	SUE	OBK-SAN-OAM	-24.30	101.01	119.24
2	B	1301	SUE	OBK-SAN-OAM	-23.74	101.43	119.24
2	B	1301	SUE	OBT-CAB-N	7.85	120.50	110.32
2	A	1301	SUE	OBT-CAB-N	7.83	120.47	110.32
2	B	1301	SUE	CAU-CAG-CAH	7.80	141.48	122.75

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	A	1301	SUE	NBZ
2	B	1301	SUE	NBZ

5 of 20 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1301	SUE	CAF-CAH-OBT-CAB
2	A	1301	SUE	CBO-NAV-SAN-CBL
2	B	1301	SUE	CBO-NAV-SAN-CBL
2	B	1301	SUE	CBQ-CAS-CAW-CBN
4	A	1303	MES	C8-C7-N4-C5

There are no ring outliers.

6 monomers are involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1303	MES	1	0
5	B	1304	SO4	1	0
5	B	1306	SO4	1	0
2	B	1301	SUE	4	0
4	A	1304	MES	1	0
2	A	1301	SUE	2	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 SUE B 1301

Bond lengths

Bond angles

Torsions

Rings

Ligand SUE A 1301

Bond lengths

Bond angles

Torsions

Rings
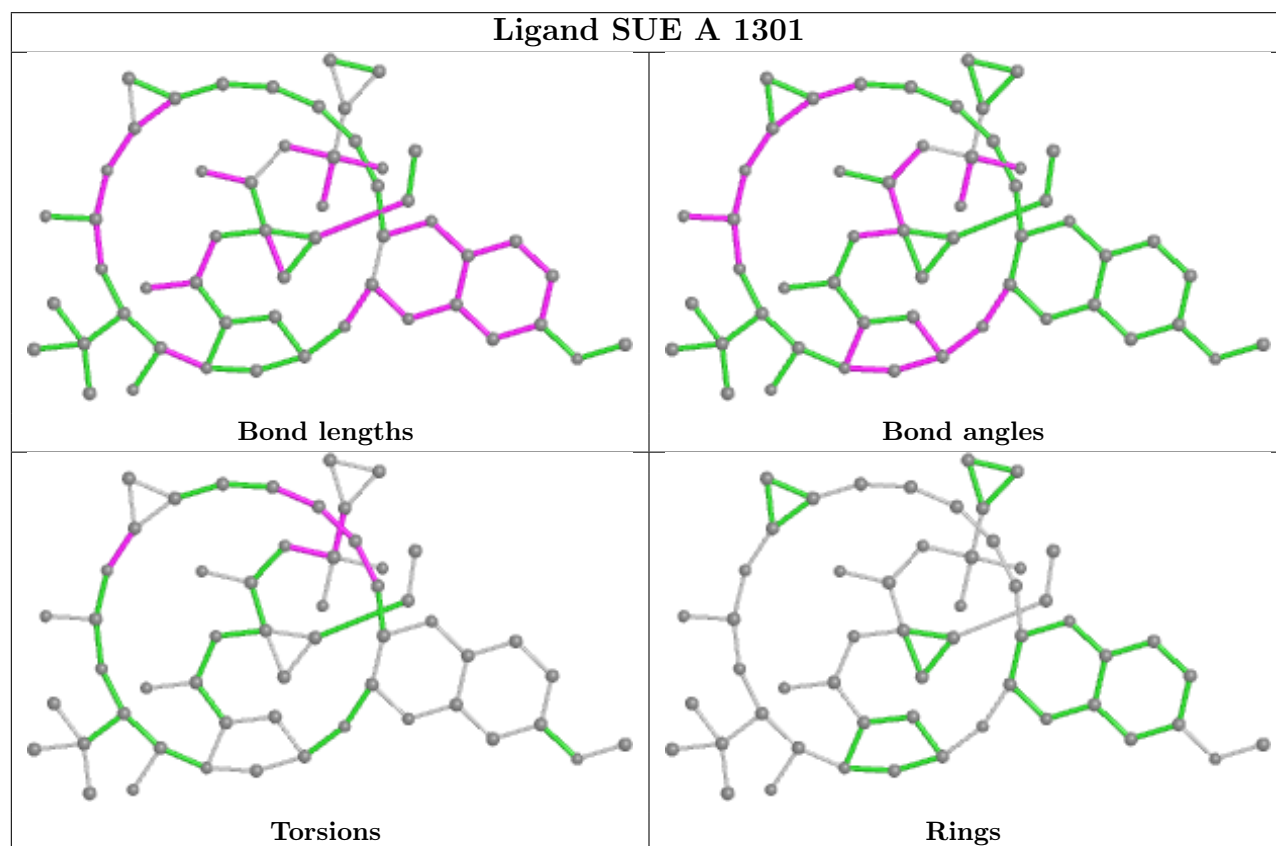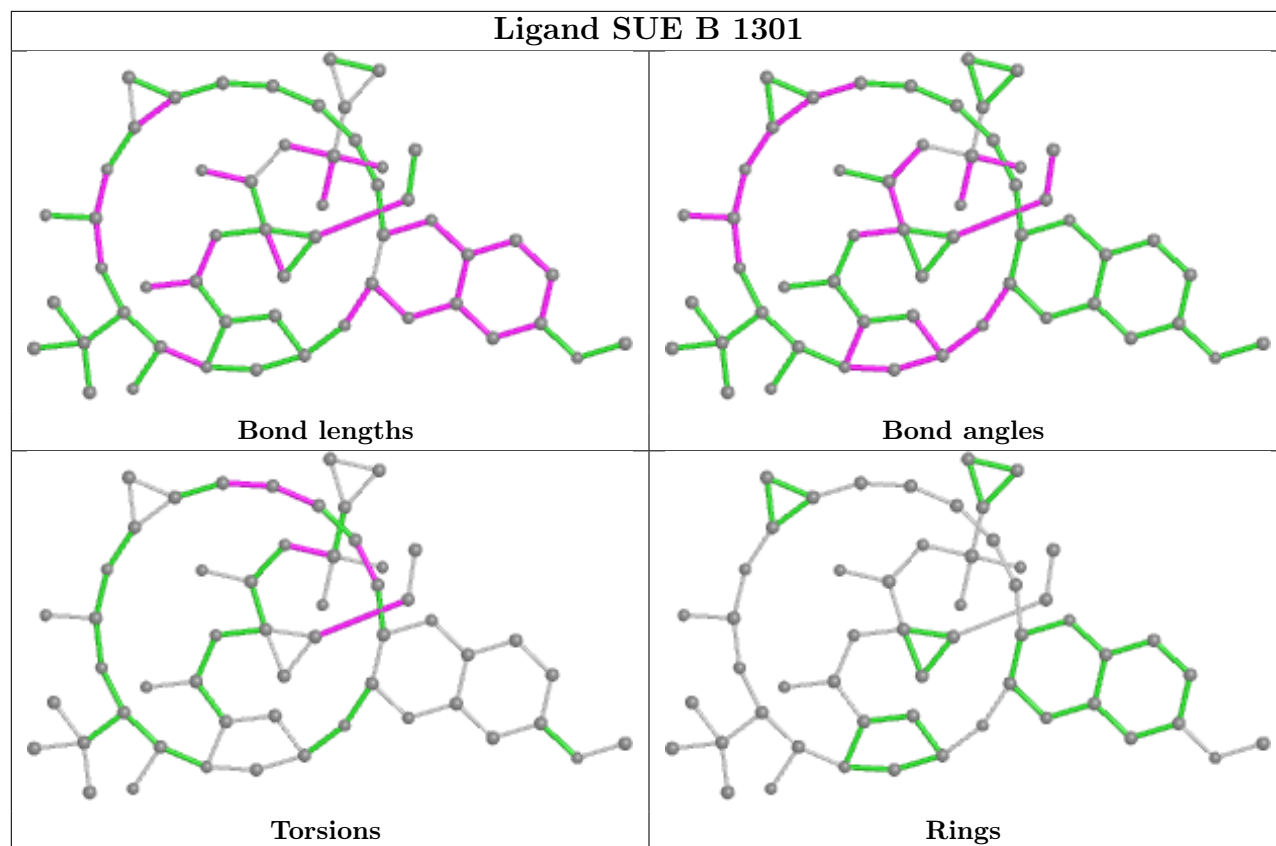
5.7 Other polymers ⓘ

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 SUE A 1301:

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

Electron density around SUE B 1301:

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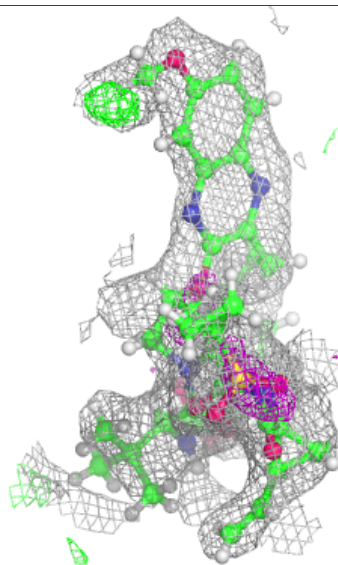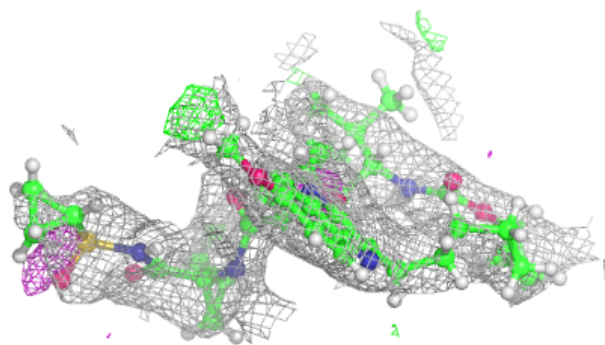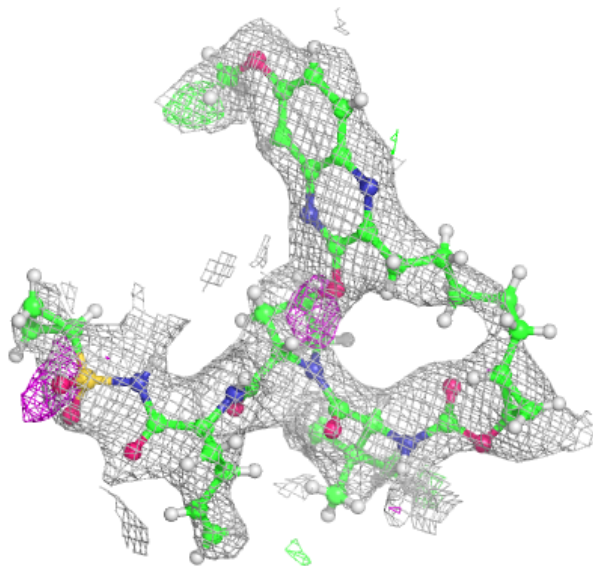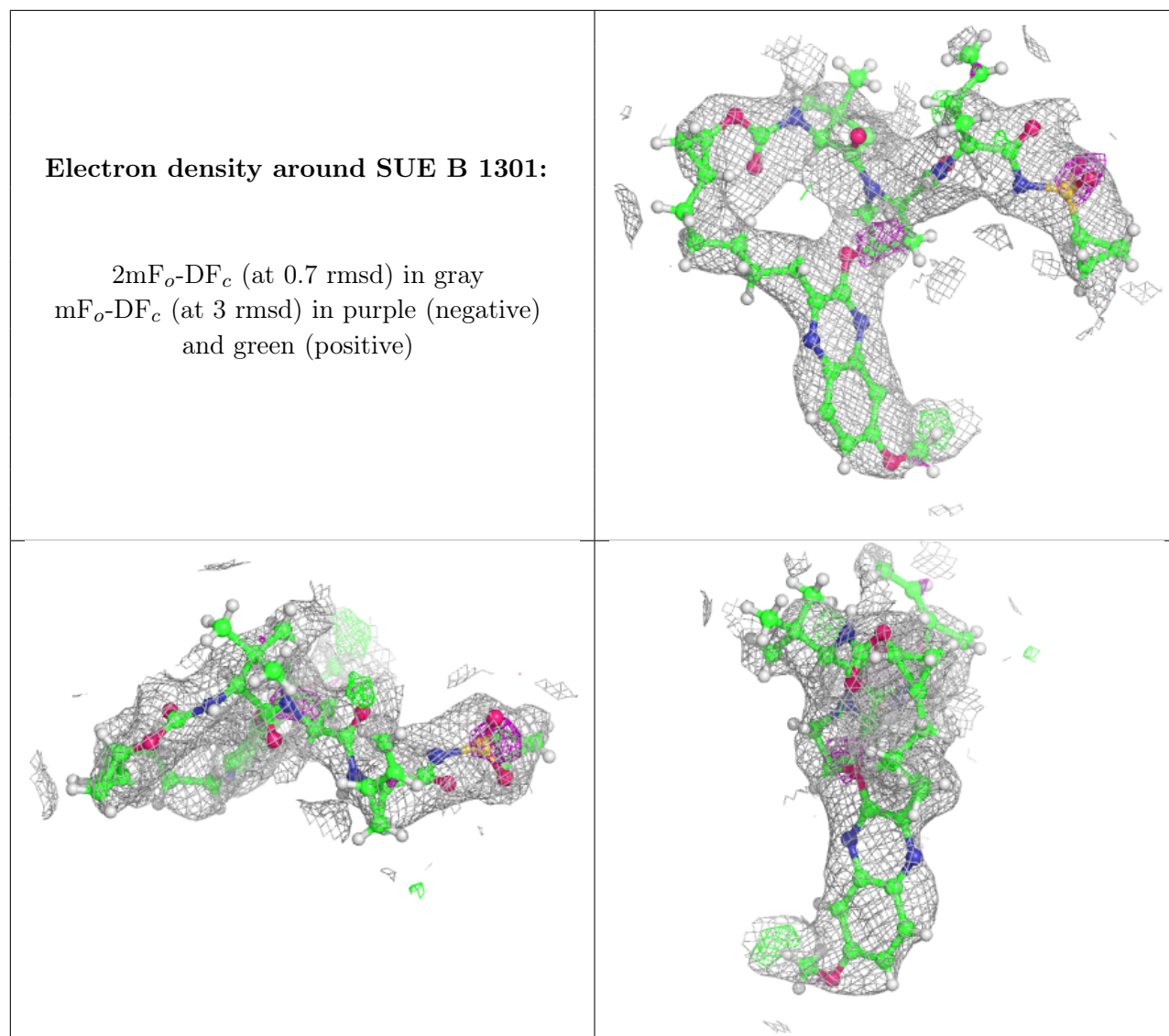
6.5   Other polymers (i)

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