wwPDB X-ray Structure Validation Summary Report (i)

Oct 9, 2023 – 08:40 AM EDT

PDB ID	:	7UXZ
Title	:	Crystal structure of SARS-CoV-2 nucleocapsid protein C-terminal domain complexed with Chicoric acid
Authors	:	Bezerra, E.H.S.; Tonoli, C.C.C.; Soprano, A.S.; Franchini, K.G.; Trivella, D.B.B.; Benedetti, C.E.
Deposited on	:	2022-05-06
Resolution	:	1.73 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:

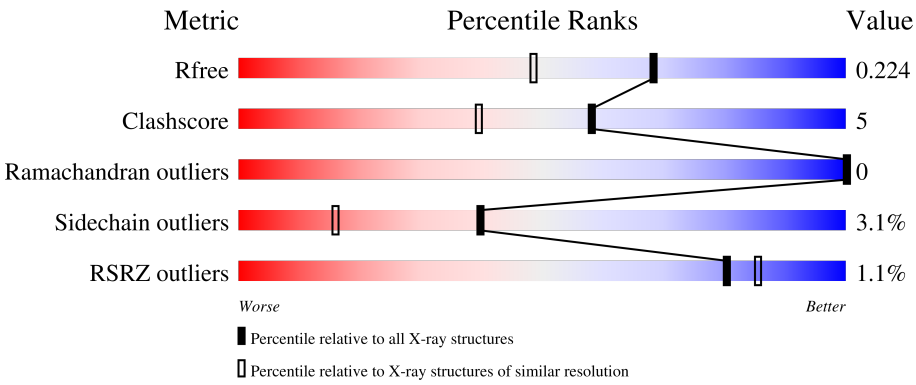
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh &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

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.73 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	3764 (1.76-1.72)
Clashscore	141614	3923 (1.76-1.72)
Ramachandran outliers	138981	3878 (1.76-1.72)
Sidechain outliers	138945	3878 (1.76-1.72)
RSRZ outliers	127900	3705 (1.76-1.72)

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% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	AAA	114	% 85% 10% • •
1	BBB	114	85% 9% • •
1	CCC	114	4% 88% 6% • •
1	DDD	114	86% 10% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	EEE	114	% 87% 9% •
1	FFF	114	82% 14% •

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	PEG	FFF	401	-	-	X	-

2 Entry composition (i)

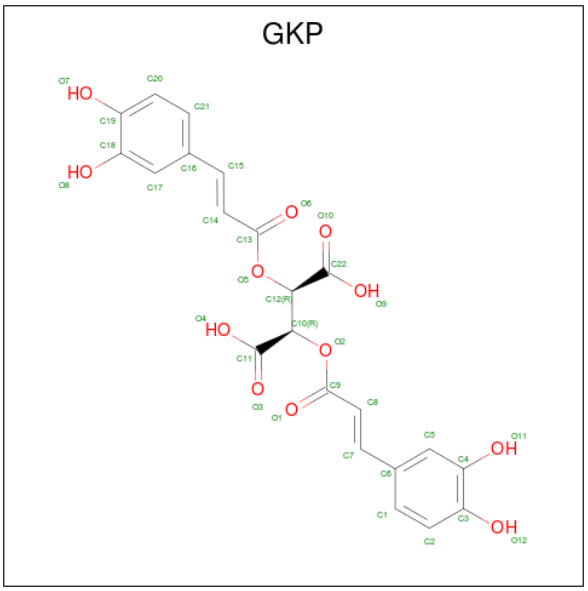
There are 7 unique types of molecules in this entry. The entry contains 6210 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Nucleoprotein.

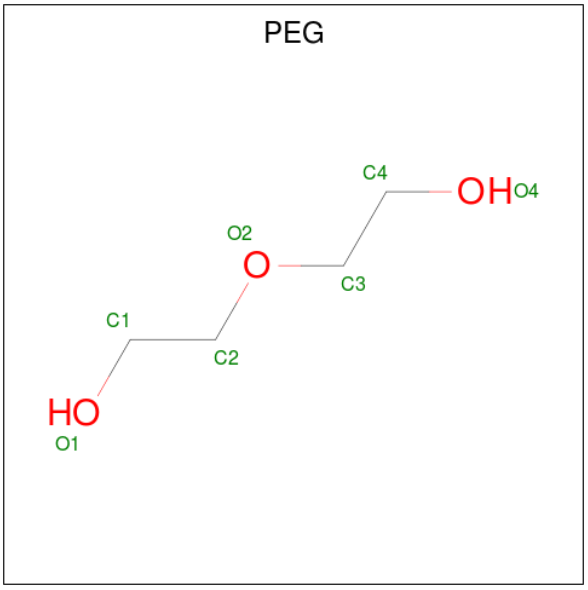
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	AAA	109	897	570	161	164	2	0	3	0
			Total	C	N	O	S			
1	BBB	109	890	565	158	164	3	0	2	0
			Total	C	N	O	S			
1	CCC	109	892	567	159	163	3	0	2	0
			Total	C	N	O	S			
1	DDD	109	901	571	162	165	3	0	3	0
			Total	C	N	O	S			
1	EEE	109	890	565	158	164	3	0	2	0
			Total	C	N	O	S			
1	FFF	109	898	570	159	165	4	0	3	0

- Molecule 2 is (2R,3R)-2,3-bis{[(2E)-3-(3,4-dihydroxyphenyl)prop-2-enoyl]oxy}butanedioic acid (three-letter code: GKP) (formula: $C_{22}H_{18}O_{12}$) (labeled as "Ligand of Interest" by depositor).



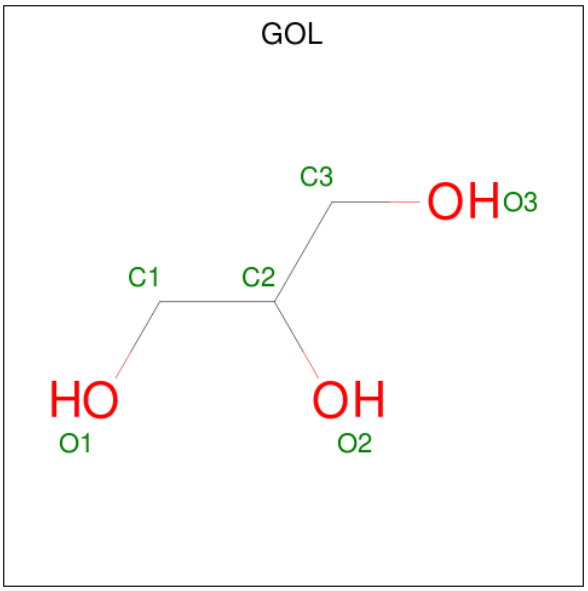
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	AAA	1	34	22	12	0	0

- Molecule 3 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
3	AAA	1	7	4	3	0	0
3	FFF	1	7	4	3	0	0

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



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
4	AAA	1	6	3	3	0	0
			Total	C	O		
4	DDD	1	6	3	3	0	0
			Total	C	O		
4	FFF	1	6	3	3	0	0

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Na		
5	AAA	1	1	1	0	0

- Molecule 6 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cl		
6	EEE	1	1	1	0	0

- Molecule 7 is water.

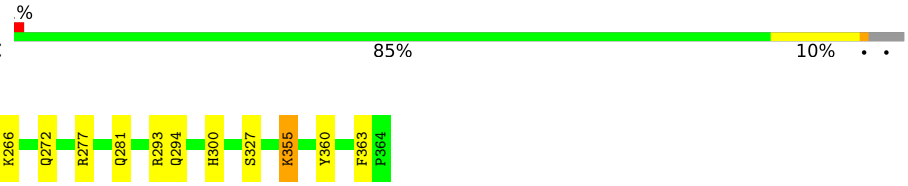
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
7	AAA	112	112	112	0	0
			Total	O		
7	BBB	128	128	128	0	0
			Total	O		
7	CCC	135	135	135	0	0
			Total	O		
7	DDD	142	142	142	0	0
			Total	O		
7	EEE	135	135	135	0	0
			Total	O		
7	FFF	122	122	122	0	0

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 and electron density. 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. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). 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.

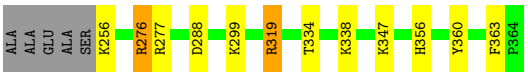
• Molecule 1: Nucleoprotein

Chain AAA:



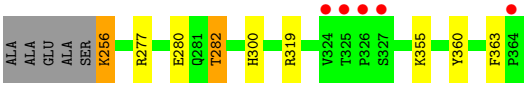
• Molecule 1: Nucleoprotein

Chain BBB:



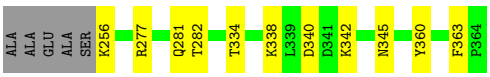
• Molecule 1: Nucleoprotein

Chain CCC:



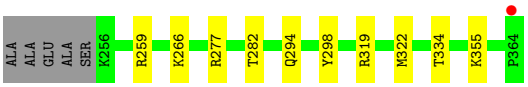
• Molecule 1: Nucleoprotein

Chain DDD:



• Molecule 1: Nucleoprotein

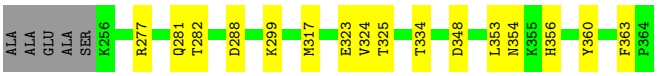
Chain EEE:



• Molecule 1: Nucleoprotein

Chain FFF:

82%
14%
•


ALA
ALA
GLU
ALA
SER
K256
R277
Q281
T282
D288
K299
M317
E323
V324
T325
T334
D348
L353
N354
K355
H356
Y360
F363
P364

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	43.67Å 120.66Å 128.50Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.98 – 1.73 43.98 – 1.73	Depositor EDS
% Data completeness (in resolution range)	99.9 (43.98-1.73) 99.9 (43.98-1.73)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.37 (at 1.73Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.174 , 0.215 0.185 , 0.224	Depositor DCC
R_{free} test set	3571 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	18.2	Xtriage
Anisotropy	0.060	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 43.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	6210	wwPDB-VP
Average B, all atoms (Å^2)	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 15.30% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: PEG, GKP, NA, CL, 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	AAA	0.69	0/926	0.79	0/1247
1	BBB	0.70	0/913	0.80	2/1231 (0.2%)
1	CCC	0.70	0/915	0.84	2/1232 (0.2%)
1	DDD	0.69	0/924	0.79	0/1245
1	EEE	0.68	0/913	0.79	0/1231
1	FFF	0.69	0/921	0.79	0/1241
All	All	0.69	0/5512	0.80	4/7427 (0.1%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	BBB	319	ARG	NE-CZ-NH1	6.71	123.66	120.30
1	CCC	319	ARG	NE-CZ-NH1	6.29	123.44	120.30
1	BBB	319	ARG	NE-CZ-NH2	-5.82	117.39	120.30
1	CCC	319	ARG	NE-CZ-NH2	-5.07	117.76	120.30

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	AAA	897	0	888	9	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	BBB	890	0	873	10	0
1	CCC	892	0	878	6	0
1	DDD	901	0	885	6	0
1	EEE	890	0	873	11	0
1	FFF	898	0	881	16	0
2	AAA	34	0	0	0	0
3	AAA	7	0	10	0	0
3	FFF	7	0	10	4	0
4	AAA	6	0	8	0	0
4	DDD	6	0	8	0	0
4	FFF	6	0	8	1	0
5	AAA	1	0	0	0	0
6	EEE	1	0	0	1	0
7	AAA	112	0	0	5	0
7	BBB	128	0	0	1	0
7	CCC	135	0	0	3	0
7	DDD	142	0	0	4	0
7	EEE	135	0	0	6	0
7	FFF	122	0	0	1	0
All	All	6210	0	5322	52	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 5.

The worst 5 of 52 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:BBB:319:ARG:HB2	1:BBB:334[B]:THR:HG22	1.63	0.79
1:FFF:288:ASP:H	1:FFF:356:HIS:HD2	1.32	0.77
1:AAA:293[B]:ARG:NH2	7:AAA:601:HOH:O	2.21	0.73
1:EEE:322:MET:CE	1:FFF:353:LEU:HB2	2.19	0.72
7:EEE:504:HOH:O	1:FFF:334[B]:THR:HG21	1.89	0.71

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	AAA	110/114 (96%)	108 (98%)	2 (2%)	0	100	100
1	BBB	109/114 (96%)	108 (99%)	1 (1%)	0	100	100
1	CCC	108/114 (95%)	106 (98%)	2 (2%)	0	100	100
1	DDD	110/114 (96%)	109 (99%)	1 (1%)	0	100	100
1	EEE	109/114 (96%)	107 (98%)	2 (2%)	0	100	100
1	FFF	110/114 (96%)	108 (98%)	2 (2%)	0	100	100
All	All	656/684 (96%)	646 (98%)	10 (2%)	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	AAA	94/93 (101%)	91 (97%)	3 (3%)	39	15
1	BBB	93/93 (100%)	89 (96%)	4 (4%)	29	8
1	CCC	93/93 (100%)	89 (96%)	4 (4%)	29	8
1	DDD	94/93 (101%)	90 (96%)	4 (4%)	29	8
1	EEE	93/93 (100%)	91 (98%)	2 (2%)	52	29
1	FFF	94/93 (101%)	93 (99%)	1 (1%)	73	59
All	All	561/558 (100%)	543 (97%)	18 (3%)	40	15

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	DDD	342	LYS
1	FFF	277	ARG
1	EEE	355	LYS
1	CCC	256[B]	LYS
1	DDD	282	THR

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

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 8 ligands modelled in this entry, 2 are monoatomic - leaving 6 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	PEG	AAA	502	-	6,6,6	0.16	0	5,5,5	0.21	0
4	GOL	AAA	503	-	5,5,5	0.14	0	5,5,5	0.37	0
4	GOL	DDD	501	-	5,5,5	0.23	0	5,5,5	0.61	0
3	PEG	FFF	401	-	6,6,6	0.22	0	5,5,5	0.17	0
2	GKP	AAA	501	-	35,35,35	1.47	9 (25%)	48,48,48	3.05	21 (43%)
4	GOL	FFF	402	-	5,5,5	0.29	0	5,5,5	0.68	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	PEG	AAA	502	-	-	3/4/4/4	-
4	GOL	AAA	503	-	-	2/4/4/4	-
4	GOL	DDD	501	-	-	2/4/4/4	-
3	PEG	FFF	401	-	-	1/4/4/4	-
2	GKP	AAA	501	-	-	13/30/30/30	0/2/2/2
4	GOL	FFF	402	-	-	2/4/4/4	-

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	AAA	501	GKP	O2-C9	3.06	1.41	1.34
2	AAA	501	GKP	O8-C18	2.86	1.42	1.36
2	AAA	501	GKP	C8-C9	2.60	1.54	1.48
2	AAA	501	GKP	O5-C13	2.59	1.40	1.34
2	AAA	501	GKP	O12-C3	2.27	1.41	1.36

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	AAA	501	GKP	C12-C10-C11	10.22	131.06	110.65
2	AAA	501	GKP	C16-C15-C14	7.69	144.51	126.91
2	AAA	501	GKP	C12-O5-C13	7.60	126.91	116.94
2	AAA	501	GKP	O5-C12-C10	-7.45	94.90	107.17
2	AAA	501	GKP	O5-C12-C22	-4.17	102.54	111.27

There are no chirality outliers.

5 of 23 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	AAA	501	GKP	C11-C10-C12-O5
2	AAA	501	GKP	C12-C10-O2-C9
2	AAA	501	GKP	O1-C9-O2-C10
2	AAA	501	GKP	C7-C8-C9-O2
3	AAA	502	PEG	O1-C1-C2-O2

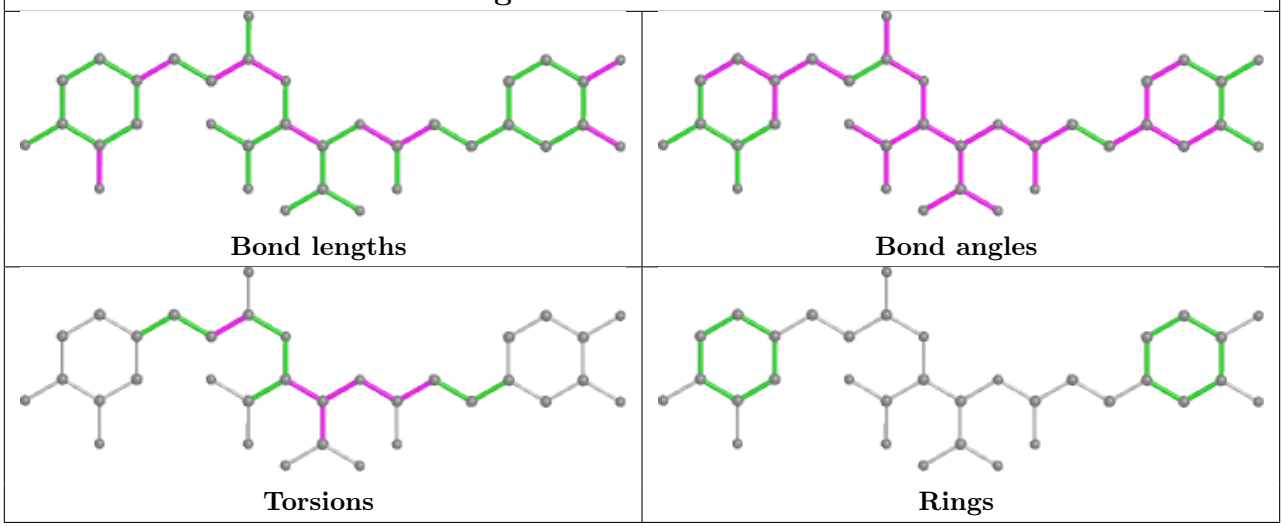
There are no ring outliers.

2 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	FFF	401	PEG	4	0
4	FFF	402	GOL	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 GKP AAA 501



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 (i)

6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	AAA	109/114 (95%)	-0.15	1 (0%) 84 88	11, 21, 40, 65	0
1	BBB	109/114 (95%)	-0.29	0 100 100	12, 19, 38, 53	0
1	CCC	109/114 (95%)	-0.18	5 (4%) 32 38	11, 16, 44, 53	0
1	DDD	109/114 (95%)	-0.36	0 100 100	11, 15, 29, 42	0
1	EEE	109/114 (95%)	-0.35	1 (0%) 84 88	11, 16, 33, 47	0
1	FFF	109/114 (95%)	-0.34	0 100 100	10, 17, 32, 57	0
All	All	654/684 (95%)	-0.28	7 (1%) 80 85	10, 17, 38, 65	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	CCC	326	PRO	3.4
1	CCC	325	THR	3.3
1	CCC	327	SER	3.1
1	CCC	364	PRO	2.8
1	EEE	364	PRO	2.6

6.2 Non-standard residues in protein, DNA, RNA chains (i)

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

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

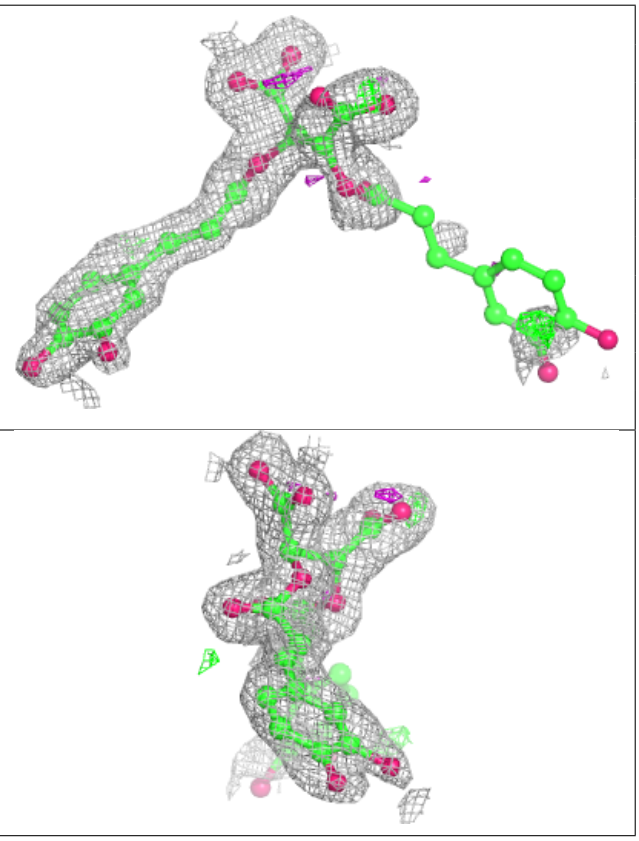
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	GOL	AAA	503	6/6	0.74	0.19	46,50,51,52	0
4	GOL	DDD	501	6/6	0.79	0.15	26,31,35,39	0
3	PEG	FFF	401	7/7	0.80	0.26	38,45,54,58	0
3	PEG	AAA	502	7/7	0.82	0.15	53,54,59,64	0
4	GOL	FFF	402	6/6	0.83	0.19	29,36,38,38	0
2	GKP	AAA	501	34/34	0.90	0.20	28,43,100,106	0
5	NA	AAA	504	1/1	0.94	0.08	36,36,36,36	0
6	CL	EEE	401	1/1	0.94	0.05	60,60,60,60	0

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 GKP AAA 501:

$2mF_o$-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 (i)

There are no such residues in this entry.