



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 15, 2023 – 09:56 AM EDT

PDB ID : 7RI1
Title : Crystal structure of anti-HIV llama VHH antibody J3 in complex with HIV-1 C1086 gp120
Authors : Zhou, T.; Kwong, P.D.
Deposited on : 2021-07-19
Resolution : 2.55 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 ⓘ 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 : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.36
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) : 2.36

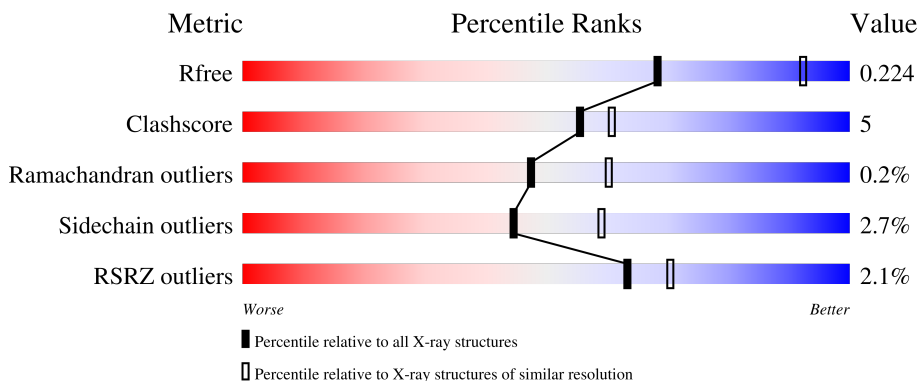
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.55 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	1284 (2.56-2.52)
Clashscore	141614	1332 (2.56-2.52)
Ramachandran outliers	138981	1315 (2.56-2.52)
Sidechain outliers	138945	1315 (2.56-2.52)
RSRZ outliers	127900	1272 (2.56-2.52)

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 $\leq 5\%$. 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	A	382	 3% 75% 15% 9%
2	B	130	 85% 8% 6%

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 3916 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 Glycoprotein 120.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	347	2708	1692	474	522	20	0	0	0

There are 78 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	20	ASP	-	expression tag	UNP R4GRV3
A	21	ILE	-	expression tag	UNP R4GRV3
A	22	ALA	-	expression tag	UNP R4GRV3
A	23	THR	-	expression tag	UNP R4GRV3
A	24	MET	-	expression tag	UNP R4GRV3
A	25	TYR	-	expression tag	UNP R4GRV3
A	26	SER	-	expression tag	UNP R4GRV3
A	27	MET	-	expression tag	UNP R4GRV3
A	28	GLN	-	expression tag	UNP R4GRV3
A	29	LEU	-	expression tag	UNP R4GRV3
A	30	ALA	-	expression tag	UNP R4GRV3
A	31	SER	-	expression tag	UNP R4GRV3
A	32	CYS	-	expression tag	UNP R4GRV3
A	33	VAL	-	expression tag	UNP R4GRV3
A	34	THR	-	expression tag	UNP R4GRV3
A	35	LEU	-	expression tag	UNP R4GRV3
A	36	THR	-	expression tag	UNP R4GRV3
A	37	LEU	-	expression tag	UNP R4GRV3
A	38	VAL	-	expression tag	UNP R4GRV3
A	39	LEU	-	expression tag	UNP R4GRV3
A	40	LEU	-	expression tag	UNP R4GRV3
A	41	VAL	-	expression tag	UNP R4GRV3
A	42	ASN	-	expression tag	UNP R4GRV3
A	43	SER	-	expression tag	UNP R4GRV3
A	63	LYS	ARG	conflict	UNP R4GRV3
A	85	VAL	ALA	conflict	UNP R4GRV3
A	87	ALA	GLY	conflict	UNP R4GRV3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	102	GLU	ASP	conflict	UNP R4GRV3
A	114	GLU	GLN	conflict	UNP R4GRV3
A	200	ALA	THR	conflict	UNP R4GRV3
A	201	ILE	VAL	conflict	UNP R4GRV3
A	215	LEU	ILE	conflict	UNP R4GRV3
A	223	PHE	TYR	conflict	UNP R4GRV3
A	231	LYS	GLU	conflict	UNP R4GRV3
A	234	ASN	GLU	conflict	UNP R4GRV3
A	240	ARG	ASN	conflict	UNP R4GRV3
A	267	GLU	LYS	conflict	UNP R4GRV3
A	271	ILE	VAL	conflict	UNP R4GRV3
A	279	ASN	ASP	conflict	UNP R4GRV3
A	289	ASN	LYS	conflict	UNP R4GRV3
A	293	ASN	GLU	conflict	UNP R4GRV3
A	300	ASN	GLY	conflict	UNP R4GRV3
A	325	ASN	ASP	conflict	UNP R4GRV3
A	326	ILE	THR	conflict	UNP R4GRV3
A	334	ASN	SER	conflict	UNP R4GRV3
A	336	SER	GLU	conflict	UNP R4GRV3
A	340	ASN	LYS	conflict	UNP R4GRV3
A	346	GLY	SER	conflict	UNP R4GRV3
A	347	GLU	LYS	conflict	UNP R4GRV3
A	348	GLU	ILE	conflict	UNP R4GRV3
A	350	ALA	GLN	conflict	UNP R4GRV3
A	351	LYS	GLU	conflict	UNP R4GRV3
A	356	SER	ASN	conflict	UNP R4GRV3
A	358	THR	ALA	conflict	UNP R4GRV3
A	364	SER	HIS	conflict	UNP R4GRV3
A	388	SER	THR	conflict	UNP R4GRV3
A	389	ASP	LYS	conflict	UNP R4GRV3
A	396	ARG	-	insertion	UNP R4GRV3
A	398	GLY	SER	conflict	UNP R4GRV3
A	403	TYR	ALA	conflict	UNP R4GRV3
A	405	HIS	SER	conflict	UNP R4GRV3
A	408	ARG	SER	conflict	UNP R4GRV3
A	409	SER	VAL	conflict	UNP R4GRV3
A	412	GLY	THR	conflict	UNP R4GRV3
A	417	GLN	PRO	conflict	UNP R4GRV3
A	419	LYS	ARG	conflict	UNP R4GRV3
A	424	ILE	VAL	conflict	UNP R4GRV3
A	429	GLU	GLY	conflict	UNP R4GRV3
A	434	ILE	MET	conflict	UNP R4GRV3

Continued on next page...

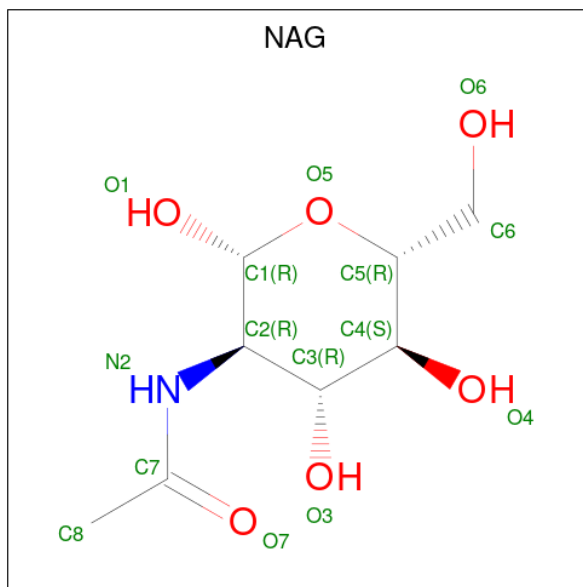
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	440	GLU	LYS	conflict	UNP R4GRV3
A	442	GLU	ASN	conflict	UNP R4GRV3
A	446	ASN	LYS	conflict	UNP R4GRV3
A	448	ASN	ILE	conflict	UNP R4GRV3
A	455	LEU	THR	conflict	UNP R4GRV3
A	459A	GLN	-	insertion	UNP R4GRV3
A	459B	SER	-	insertion	UNP R4GRV3
A	459D	GLU	ASP	conflict	UNP R4GRV3
A	459E	THR	ASP	conflict	UNP R4GRV3

- Molecule 2 is a protein called Lamma VHH antibody J3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	122	951	602	165	179	5	2	0	0

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



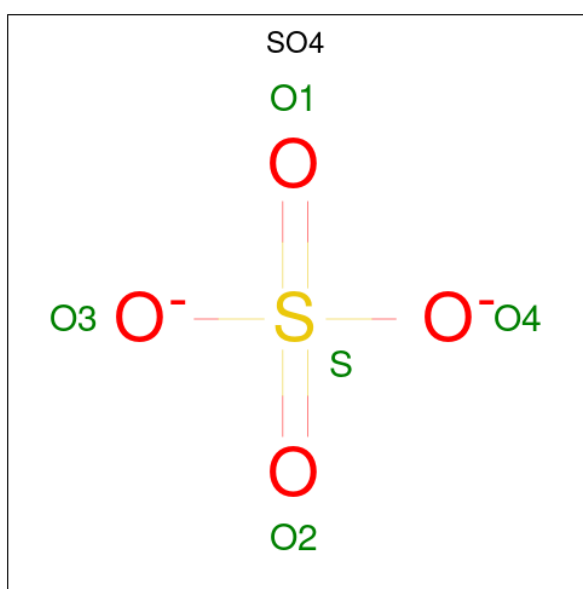
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
3	A	1	14	8	1	5	0	0
3	A	1	14	8	1	5	0	0
3	A	1	14	8	1	5	0	0

Continued on next page...

Continued from previous page...

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

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



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

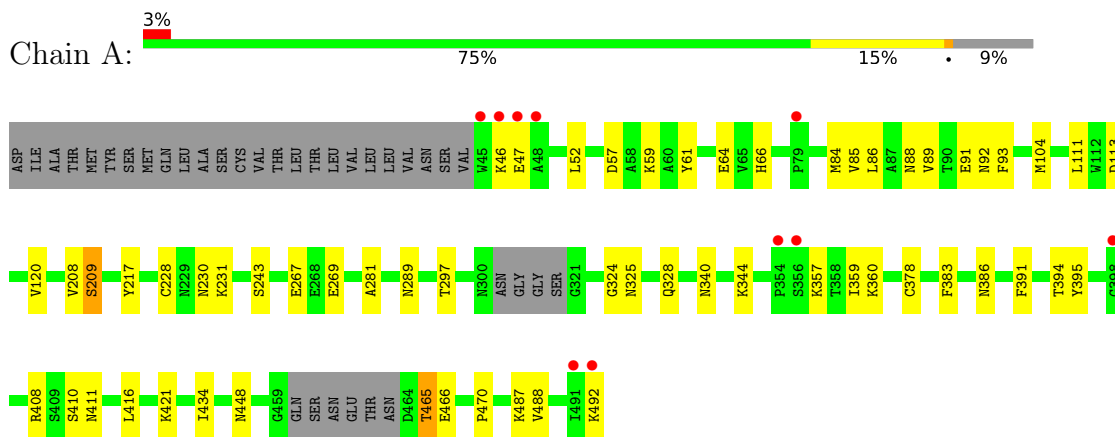
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	94	Total O 94 94	0	0
5	B	40	Total O 40 40	0	0

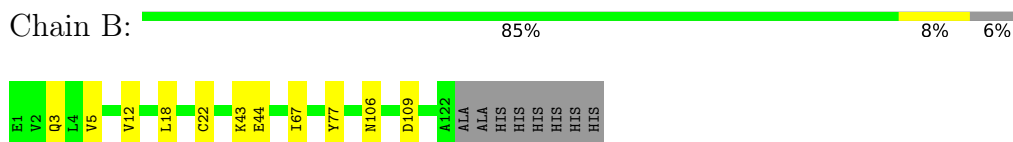
3 Residue-property plots [i](#)

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: Glycoprotein 120



- Molecule 2: Lamma VHH antibody J3



4 Data and refinement statistics

Property	Value	Source
Space group	P 42 21 2	Depositor
Cell constants a, b, c, α , β , γ	119.09Å 119.09Å 110.52Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.20 – 2.55 46.20 – 2.55	Depositor EDS
% Data completeness (in resolution range)	97.3 (46.20-2.55) 97.3 (46.20-2.55)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.86 (at 2.54Å)	Xtrriage
Refinement program	PHENIX 1.19-4092	Depositor
R, R_{free}	0.176 , 0.225 0.175 , 0.224	Depositor DCC
R_{free} test set	1316 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å ²)	40.1	Xtrriage
Anisotropy	0.300	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 39.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	3916	wwPDB-VP
Average B, all atoms (Å ²)	50.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.52% 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: SO4, NAG

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	0.43	0/2765	0.61	0/3748
2	B	0.48	0/973	0.64	0/1311
All	All	0.44	0/3738	0.62	0/5059

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	2708	0	2619	31	0
2	B	951	0	916	7	0
3	A	98	0	91	0	0
4	A	20	0	0	2	0
4	B	5	0	0	0	0
5	A	94	0	0	5	0
5	B	40	0	0	1	0
All	All	3916	0	3626	38	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 38 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:410:SER:OG	1:A:411:ASN:N	2.13	0.81
1:A:231:LYS:HD2	1:A:267:GLU:HG2	1.70	0.74
1:A:46:LYS:HD2	1:A:492:LYS:HD2	1.69	0.73
2:B:12:VAL:HG22	2:B:18:LEU:HD13	1.73	0.69
1:A:391:PHE:CD2	1:A:470:PRO:HG3	2.29	0.66

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	341/382 (89%)	323 (95%)	17 (5%)	1 (0%)	41	51
2	B	120/130 (92%)	118 (98%)	2 (2%)	0	100	100
All	All	461/512 (90%)	441 (96%)	19 (4%)	1 (0%)	47	60

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	324	GLY

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	303/334 (91%)	294 (97%)	9 (3%)	41	55
2	B	98/104 (94%)	96 (98%)	2 (2%)	55	70
All	All	401/438 (92%)	390 (97%)	11 (3%)	44	59

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	408	ARG
1	A	465	THR
2	B	109	ASP
2	B	43	LYS
1	A	209	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	203	GLN
1	A	280	ASN
1	A	328	GLN
2	B	3	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](#)

12 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	SO4	A	510	-	4,4,4	0.44	0	6,6,6	0.32	0
3	NAG	A	504	1	14,14,15	0.45	0	17,19,21	0.59	0
4	SO4	A	508	-	4,4,4	0.17	0	6,6,6	0.36	0
4	SO4	A	509	-	4,4,4	0.39	0	6,6,6	0.24	0
3	NAG	A	505	1	14,14,15	0.19	0	17,19,21	0.50	0
3	NAG	A	506	1	14,14,15	1.12	1 (7%)	17,19,21	1.06	1 (5%)
3	NAG	A	502	1	14,14,15	0.31	0	17,19,21	0.47	0
3	NAG	A	503	1	14,14,15	0.50	0	17,19,21	0.56	0
3	NAG	A	507	1	14,14,15	0.67	1 (7%)	17,19,21	0.79	1 (5%)
4	SO4	A	511	-	4,4,4	0.21	0	6,6,6	0.14	0
3	NAG	A	501	1	14,14,15	0.38	0	17,19,21	0.30	0
4	SO4	B	201	-	4,4,4	0.22	0	6,6,6	0.24	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	NAG	A	504	1	-	2/6/23/26	0/1/1/1
3	NAG	A	505	1	-	4/6/23/26	0/1/1/1
3	NAG	A	506	1	-	0/6/23/26	0/1/1/1
3	NAG	A	502	1	-	2/6/23/26	0/1/1/1
3	NAG	A	503	1	-	2/6/23/26	0/1/1/1
3	NAG	A	507	1	-	4/6/23/26	0/1/1/1
3	NAG	A	501	1	-	2/6/23/26	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	506	NAG	O5-C1	3.97	1.50	1.43
3	A	507	NAG	O5-C1	2.08	1.47	1.43

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	506	NAG	C1-O5-C5	3.77	117.30	112.19
3	A	507	NAG	C1-O5-C5	2.79	115.98	112.19

There are no chirality outliers.

5 of 16 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	501	NAG	O5-C5-C6-O6
3	A	502	NAG	C4-C5-C6-O6
3	A	502	NAG	O5-C5-C6-O6
3	A	501	NAG	C4-C5-C6-O6
3	A	505	NAG	C8-C7-N2-C2

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	508	SO4	1	0
4	A	509	SO4	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 [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	A	347/382 (90%)	0.06	10 (2%) 51 59	25, 49, 88, 120	0
2	B	122/130 (93%)	-0.15	0 100 100	27, 41, 65, 84	1 (0%)
All	All	469/512 (91%)	0.01	10 (2%) 63 70	25, 47, 85, 120	1 (0%)

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	45	TRP	5.6
1	A	492	LYS	4.8
1	A	47	GLU	4.6
1	A	46	LYS	4.2
1	A	356	SER	3.7

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(\AA^2)	Q<0.9
3	NAG	A	502	14/15	0.87	0.33	72,78,85,86	0
3	NAG	A	505	14/15	0.91	0.17	63,67,73,76	0
3	NAG	A	504	14/15	0.92	0.23	63,67,71,72	0
4	SO4	A	510	5/5	0.93	0.20	42,48,55,77	0
3	NAG	A	501	14/15	0.94	0.18	60,66,71,74	0
4	SO4	B	201	5/5	0.94	0.22	68,70,84,87	0
3	NAG	A	507	14/15	0.95	0.13	48,55,58,59	0
4	SO4	A	509	5/5	0.95	0.44	47,47,55,64	0
4	SO4	A	511	5/5	0.96	0.28	76,78,81,91	0
3	NAG	A	503	14/15	0.96	0.12	38,45,47,52	0
3	NAG	A	506	14/15	0.97	0.12	29,36,44,46	0
4	SO4	A	508	5/5	1.00	0.16	30,31,34,34	0

6.5 Other polymers [\(i\)](#)

There are no such residues in this entry.