wwPDB X-ray Structure Validation Summary Report (i)

Jan 3, 2023 – 06:52 pm GMT

PDB ID	:	8ASY
Title	:	SARS-CoV-2 Omicron BA.2.75 RBD in complex with ACE2
Authors	:	Zhou, D.; Huo, J.; Ren, J.; Stuart, D.I.
Deposited on	:	2022-08-22
Resolution	:	2.85 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 (1)) were used in the production of this report:

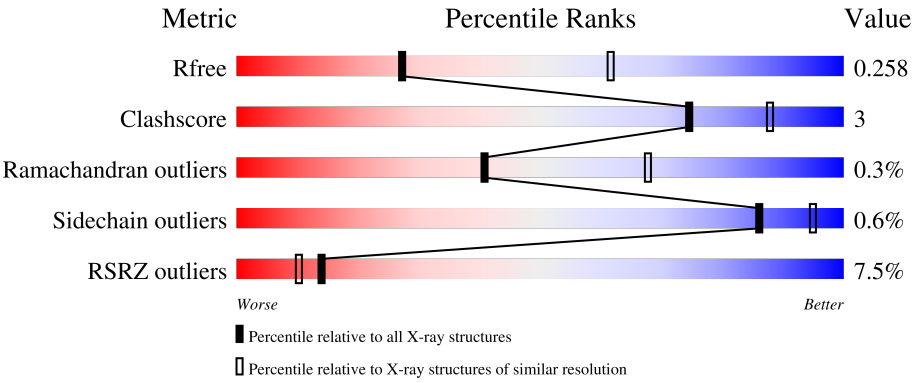
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.31.3
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.3

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 2.85 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	3168 (2.90-2.82)
Clashscore	141614	3438 (2.90-2.82)
Ramachandran outliers	138981	3348 (2.90-2.82)
Sidechain outliers	138945	3351 (2.90-2.82)
RSRZ outliers	127900	3103 (2.90-2.82)

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	A	604	4% 89% 9%
2	B	219	16% 81% 11% 8%

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	GOL	A	714	-	-	-	X
3	GOL	B	601	-	-	-	X
6	PGE	A	716	-	-	-	X

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 6631 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 Processed angiotensin-converting enzyme 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	596	Total	C	N	O	S	0	0	0
			4862	3111	805	917	29			

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	616	ARG	-	expression tag	UNP Q9BYF1
A	617	HIS	-	expression tag	UNP Q9BYF1
A	618	HIS	-	expression tag	UNP Q9BYF1
A	619	HIS	-	expression tag	UNP Q9BYF1
A	620	HIS	-	expression tag	UNP Q9BYF1
A	621	HIS	-	expression tag	UNP Q9BYF1
A	622	HIS	-	expression tag	UNP Q9BYF1

- Molecule 2 is a protein called Spike protein S1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	201	Total	C	N	O	S	0	0	0
			1602	1036	269	289	8			

There are 42 discrepancies between the modelled and reference sequences:

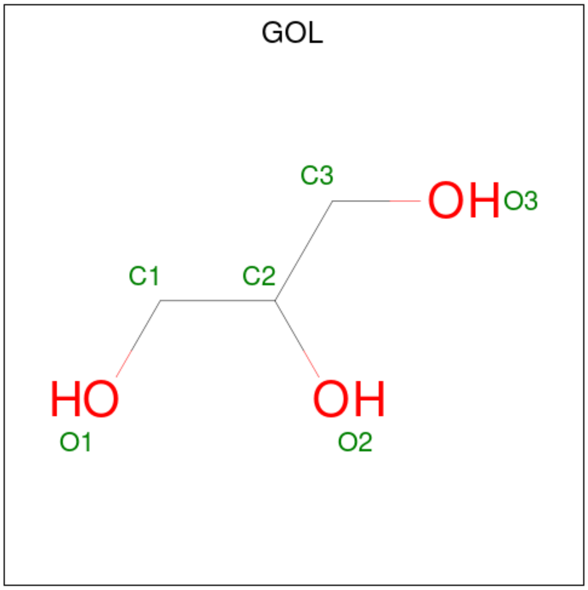
Chain	Residue	Modelled	Actual	Comment	Reference
B	339	HIS	GLY	variant	UNP P0DTC2
B	371	PHE	SER	variant	UNP P0DTC2
B	373	PRO	SER	variant	UNP P0DTC2
B	375	PHE	SER	variant	UNP P0DTC2
B	376	ALA	THR	variant	UNP P0DTC2
B	405	ASN	ASP	variant	UNP P0DTC2
B	408	SER	ARG	variant	UNP P0DTC2
B	417	ASN	LYS	variant	UNP P0DTC2
B	440	LYS	ASN	variant	UNP P0DTC2

Continued on next page...

Continued from previous page...

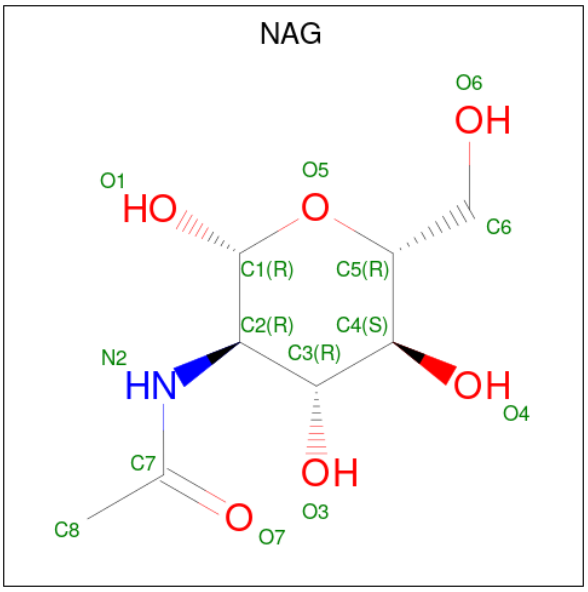
Chain	Residue	Modelled	Actual	Comment	Reference
B	446	SER	GLY	variant	UNP P0DTC2
B	460	LYS	ASN	variant	UNP P0DTC2
B	477	ASN	SER	variant	UNP P0DTC2
B	478	LYS	THR	variant	UNP P0DTC2
B	484	ALA	GLU	variant	UNP P0DTC2
B	498	ARG	GLN	variant	UNP P0DTC2
B	501	TYR	ASN	variant	UNP P0DTC2
B	505	HIS	TYR	variant	UNP P0DTC2
B	527	LYS	-	expression tag	UNP P0DTC2
B	528	LYS	-	expression tag	UNP P0DTC2
B	529	SER	-	expression tag	UNP P0DTC2
B	530	LEU	-	expression tag	UNP P0DTC2
B	531	LEU	-	expression tag	UNP P0DTC2
B	532	ASN	-	expression tag	UNP P0DTC2
B	533	ASP	-	expression tag	UNP P0DTC2
B	534	ILE	-	expression tag	UNP P0DTC2
B	535	PHE	-	expression tag	UNP P0DTC2
B	536	GLU	-	expression tag	UNP P0DTC2
B	537	ALA	-	expression tag	UNP P0DTC2
B	538	GLN	-	expression tag	UNP P0DTC2
B	539	LYS	-	expression tag	UNP P0DTC2
B	540	ILE	-	expression tag	UNP P0DTC2
B	541	GLU	-	expression tag	UNP P0DTC2
B	542	TRP	-	expression tag	UNP P0DTC2
B	543	HIS	-	expression tag	UNP P0DTC2
B	544	GLU	-	expression tag	UNP P0DTC2
B	545	LYS	-	expression tag	UNP P0DTC2
B	546	HIS	-	expression tag	UNP P0DTC2
B	547	HIS	-	expression tag	UNP P0DTC2
B	548	HIS	-	expression tag	UNP P0DTC2
B	549	HIS	-	expression tag	UNP P0DTC2
B	550	HIS	-	expression tag	UNP P0DTC2
B	551	HIS	-	expression tag	UNP P0DTC2

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

- Molecule 4 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).

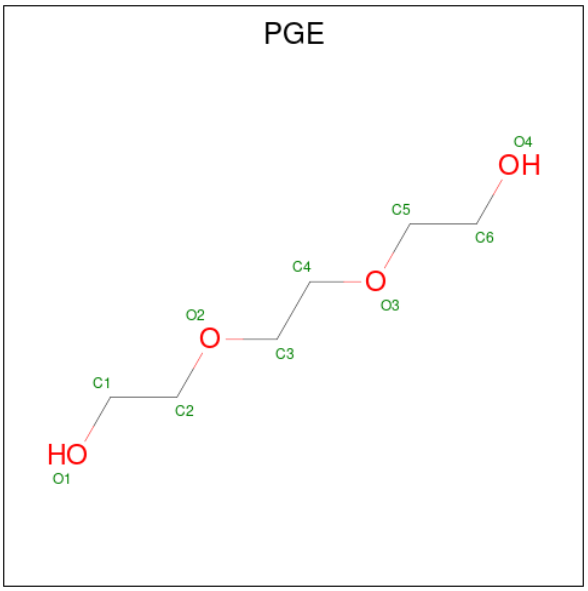


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C N O 14 8 1 5	0	0
4	A	1	Total C N O 14 8 1 5	0	0
4	A	1	Total C N O 14 8 1 5	0	0
4	A	1	Total C N O 14 8 1 5	0	0
4	A	1	Total C N O 14 8 1 5	0	0
4	A	1	Total C N O 14 8 1 5	0	0

- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total Cl 1 1	0	0

- Molecule 6 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).

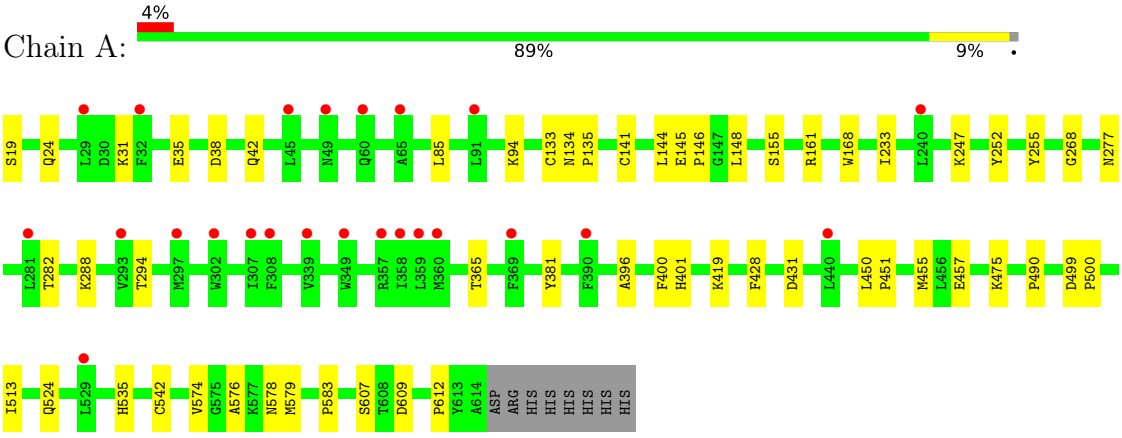


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
6	A	1	10	6	4	0	0

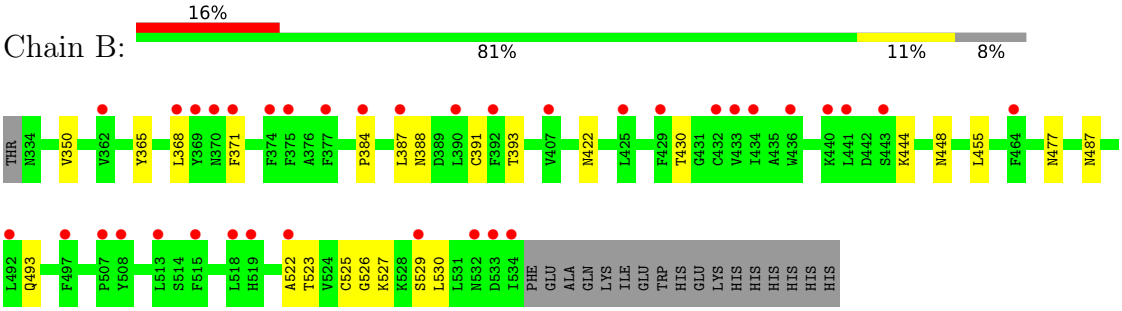
3 Residue-property plots (i)

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: Processed angiotensin-converting enzyme 2



• Molecule 2: Spike protein S1



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	105.26Å 105.26Å 220.75Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	76.18 – 2.85 76.18 – 2.85	Depositor EDS
% Data completeness (in resolution range)	99.0 (76.18-2.85) 99.1 (76.18-2.85)	Depositor EDS
R_{merge}	0.44	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.99 (at 2.86Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.217 , 0.265 0.218 , 0.258	Depositor DCC
R_{free} test set	1444 reflections (4.88%)	wwPDB-VP
Wilson B-factor (Å^2)	76.3	Xtriage
Anisotropy	0.352	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 50.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6631	wwPDB-VP
Average B, all atoms (Å^2)	86.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.58% 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: PGE, NAG, GOL, CL

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	0.24	0/4999	0.42	0/6792
2	B	0.26	0/1650	0.47	1/2245 (0.0%)
All	All	0.24	0/6649	0.43	1/9037 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	371	PHE	C-N-CA	-6.00	106.69	121.70

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	4862	0	4633	31	0
2	B	1602	0	1525	12	0
3	A	60	0	80	1	0
3	B	12	0	16	0	0
4	A	84	0	78	0	0
5	A	1	0	0	1	0
6	A	10	0	14	0	0
All	All	6631	0	6346	41	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 3.

The worst 5 of 41 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)
2:B:384:PRO:HA	2:B:387:LEU:HG	1.80	0.64
1:A:288:LYS:NZ	1:A:431:ASP:OD2	2.28	0.64
1:A:574:VAL:HG23	1:A:576:ALA:H	1.63	0.64
2:B:365:TYR:HD2	2:B:388:ASN:HB3	1.63	0.63
1:A:31:LYS:NZ	1:A:35:GLU:OE2	2.33	0.61

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	594/604 (98%)	580 (98%)	14 (2%)	0	100 100
2	B	199/219 (91%)	185 (93%)	12 (6%)	2 (1%)	15 40
All	All	793/823 (96%)	765 (96%)	26 (3%)	2 (0%)	41 68

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	368	LEU
2	B	529	SER

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	526/534 (98%)	523 (99%)	3 (1%)	86 95
2	B	172/191 (90%)	171 (99%)	1 (1%)	86 95
All	All	698/725 (96%)	694 (99%)	4 (1%)	86 95

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

Mol	Chain	Res	Type
1	A	381	TYR
1	A	401	HIS
1	A	455	MET
2	B	430	THR

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

Mol	Chain	Res	Type
1	A	134	ASN

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 20 ligands modelled in this entry, 1 is monoatomic - leaving 19 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	GOL	B	601	-	5,5,5	0.91	0	5,5,5	1.00	0
4	NAG	A	704	1	14,14,15	0.28	0	17,19,21	0.44	0
3	GOL	A	703	-	5,5,5	0.91	0	5,5,5	1.00	0
3	GOL	A	713	-	5,5,5	0.93	0	5,5,5	1.01	0
3	GOL	A	701	-	5,5,5	0.89	0	5,5,5	1.03	0
3	GOL	A	714	-	5,5,5	0.90	0	5,5,5	1.00	0
4	NAG	A	702	1	14,14,15	0.29	0	17,19,21	0.47	0
6	PGE	A	716	-	9,9,9	0.31	0	8,8,8	0.29	0
3	GOL	A	709	-	5,5,5	0.90	0	5,5,5	1.01	0
3	GOL	A	717	-	5,5,5	0.90	0	5,5,5	1.01	0
3	GOL	A	718	-	5,5,5	0.90	0	5,5,5	1.01	0
4	NAG	A	707	1	14,14,15	0.34	0	17,19,21	0.51	0
4	NAG	A	706	1	14,14,15	0.20	0	17,19,21	0.38	0
3	GOL	A	715	-	5,5,5	0.91	0	5,5,5	1.01	0
3	GOL	A	711	-	5,5,5	0.91	0	5,5,5	0.94	0
4	NAG	A	705	1	14,14,15	0.24	0	17,19,21	0.45	0
3	GOL	B	602	-	5,5,5	0.93	0	5,5,5	0.98	0
4	NAG	A	708	1	14,14,15	0.48	0	17,19,21	0.71	1 (5%)
3	GOL	A	712	-	5,5,5	0.91	0	5,5,5	0.96	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	GOL	B	601	-	-	2/4/4/4	-
4	NAG	A	704	1	-	0/6/23/26	0/1/1/1
3	GOL	A	703	-	-	0/4/4/4	-
3	GOL	A	713	-	-	2/4/4/4	-
3	GOL	A	701	-	-	0/4/4/4	-
3	GOL	A	714	-	-	4/4/4/4	-
4	NAG	A	702	1	-	0/6/23/26	0/1/1/1
6	PGE	A	716	-	-	4/7/7/7	-
3	GOL	A	709	-	-	0/4/4/4	-
3	GOL	A	717	-	-	2/4/4/4	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	GOL	A	718	-	-	2/4/4/4	-
4	NAG	A	707	1	-	0/6/23/26	0/1/1/1
4	NAG	A	706	1	-	2/6/23/26	0/1/1/1
3	GOL	A	715	-	-	0/4/4/4	-
3	GOL	A	711	-	-	0/4/4/4	-
4	NAG	A	705	1	-	1/6/23/26	0/1/1/1
3	GOL	B	602	-	-	0/4/4/4	-
4	NAG	A	708	1	-	1/6/23/26	0/1/1/1
3	GOL	A	712	-	-	0/4/4/4	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	708	NAG	C1-O5-C5	2.42	115.47	112.19

There are no chirality outliers.

5 of 20 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	717	GOL	O1-C1-C2-C3
3	B	601	GOL	C1-C2-C3-O3
6	A	716	PGE	O3-C5-C6-O4
4	A	706	NAG	C4-C5-C6-O6
3	B	601	GOL	O2-C2-C3-O3

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	717	GOL	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

6.1 Protein, DNA and RNA chains

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, 95^{th} 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	596/604 (98%)	0.59	24 (4%) 38 32	56, 77, 116, 168	0
2	B	201/219 (91%)	1.01	36 (17%) 1 1	73, 93, 145, 177	0
All	All	797/823 (96%)	0.69	60 (7%) 14 10	56, 82, 129, 177	0

The worst 5 of 60 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	374	PHE	7.0
2	B	532	ASN	6.2
2	B	368	LEU	4.5
2	B	375	PHE	3.9
2	B	370	ASN	3.7

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

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, 95^{th} 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	NAG	A	708	14/15	0.61	0.38	116,140,143,145	0
3	GOL	A	717	6/6	0.63	0.31	73,80,92,95	0
3	GOL	B	601	6/6	0.70	0.40	79,94,95,95	0
6	PGE	A	716	10/10	0.70	0.41	70,77,90,95	0
3	GOL	A	714	6/6	0.71	0.44	81,93,104,105	0
3	GOL	A	711	6/6	0.74	0.27	87,90,95,96	0
4	NAG	A	702	14/15	0.76	0.26	103,120,129,132	0
4	NAG	A	704	14/15	0.76	0.24	102,113,121,123	0
4	NAG	A	706	14/15	0.77	0.23	113,137,146,153	0
3	GOL	A	709	6/6	0.78	0.31	74,96,98,111	0
4	NAG	A	707	14/15	0.78	0.22	99,113,116,118	0
3	GOL	A	713	6/6	0.80	0.50	83,94,106,111	0
3	GOL	A	701	6/6	0.81	0.26	88,90,116,155	0
3	GOL	A	715	6/6	0.82	0.62	78,84,86,101	0
3	GOL	A	703	6/6	0.83	0.33	83,87,95,96	0
4	NAG	A	705	14/15	0.84	0.17	84,130,138,144	0
3	GOL	A	712	6/6	0.86	0.52	79,85,95,98	0
3	GOL	B	602	6/6	0.88	0.47	84,85,91,94	0
3	GOL	A	718	6/6	0.90	0.50	79,93,98,103	0
5	CL	A	710	1/1	0.94	0.14	67,67,67,67	0

6.5 Other polymers (i)

There are no such residues in this entry.