W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

May 26, 2022 – 04:16 pm BST

PDB ID	:	7ZF9
Title	:	SARS-CoV-2 Omicron BA.2 RBD in complex with COVOX-150 Fab (P21)
Authors	:	Zhou, D.; Huo, J.; Ren, J.; Stuart, D.I.
Deposited on	:	2022-04-01
Resolution	:	3.25 Å(reported)

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

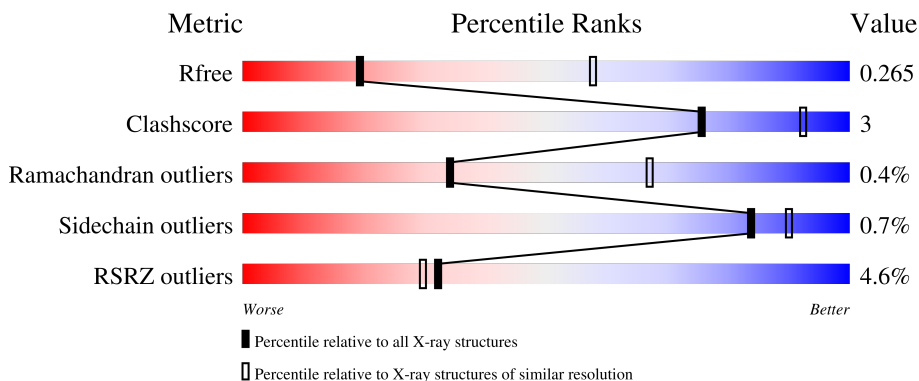|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.28.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.28.1

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 3.25 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.
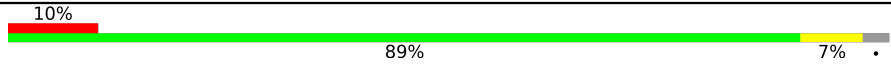
Metric	Percentile Ranks	Value
Rfree		0.265
Clashscore		3
Ramachandran outliers		0.4%
Sidechain outliers		0.7%
RSRZ outliers		4.6%

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	1191 (3.30-3.22)
Clashscore	141614	1251 (3.30-3.22)
Ramachandran outliers	138981	1229 (3.30-3.22)
Sidechain outliers	138945	1228 (3.30-3.22)
RSRZ outliers	127900	1154 (3.30-3.22)

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	222	4% / 86% / 11% / ·
1	H	222	5% / 89% / 10% / ·
2	B	216	89% / 10%
2	L	216	% / 93% / 6%
3	C	202	8% / 89% / 8% / ·

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	E	202	

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 9639 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 COVOX-150 heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	H	220	Total	C	N	O	S	0	0	0
			1620	1017	270	324	9			
1	A	216	Total	C	N	O	S	0	0	0
			1597	1005	266	317	9			

- Molecule 2 is a protein called COVOX-150 light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	L	215	Total	C	N	O	S	0	0	0
			1642	1030	272	335	5			
2	B	215	Total	C	N	O	S	0	0	0
			1642	1030	272	335	5			

- Molecule 3 is a protein called Spike protein S1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	195	Total	C	N	O	S	0	0	0
			1555	1004	262	281	8			
3	C	195	Total	C	N	O	S	0	0	0
			1555	1004	262	281	8			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	327	HIS	-	expression tag	UNP P0DTC2
E	328	HIS	-	expression tag	UNP P0DTC2
E	329	HIS	-	expression tag	UNP P0DTC2
E	330	HIS	-	expression tag	UNP P0DTC2
E	331	HIS	-	expression tag	UNP P0DTC2
E	332	HIS	-	expression tag	UNP P0DTC2
E	339	ASP	GLY	variant	UNP P0DTC2
E	371	PHE	SER	variant	UNP P0DTC2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	373	PRO	SER	variant	UNP P0DTC2
E	375	PHE	SER	variant	UNP P0DTC2
E	376	ALA	THR	variant	UNP P0DTC2
E	405	ASN	ASP	variant	UNP P0DTC2
E	408	SER	ARG	variant	UNP P0DTC2
E	417	ASN	LYS	variant	UNP P0DTC2
E	440	LYS	ASN	variant	UNP P0DTC2
E	477	ASN	SER	variant	UNP P0DTC2
E	478	LYS	THR	variant	UNP P0DTC2
E	484	ALA	GLU	variant	UNP P0DTC2
E	493	ARG	GLN	variant	UNP P0DTC2
E	498	ARG	GLN	variant	UNP P0DTC2
E	501	TYR	ASN	variant	UNP P0DTC2
E	505	HIS	TYR	variant	UNP P0DTC2
E	527	LYS	-	expression tag	UNP P0DTC2
E	528	LYS	-	expression tag	UNP P0DTC2
C	327	HIS	-	expression tag	UNP P0DTC2
C	328	HIS	-	expression tag	UNP P0DTC2
C	329	HIS	-	expression tag	UNP P0DTC2
C	330	HIS	-	expression tag	UNP P0DTC2
C	331	HIS	-	expression tag	UNP P0DTC2
C	332	HIS	-	expression tag	UNP P0DTC2
C	339	ASP	GLY	variant	UNP P0DTC2
C	371	PHE	SER	variant	UNP P0DTC2
C	373	PRO	SER	variant	UNP P0DTC2
C	375	PHE	SER	variant	UNP P0DTC2
C	376	ALA	THR	variant	UNP P0DTC2
C	405	ASN	ASP	variant	UNP P0DTC2
C	408	SER	ARG	variant	UNP P0DTC2
C	417	ASN	LYS	variant	UNP P0DTC2
C	440	LYS	ASN	variant	UNP P0DTC2
C	477	ASN	SER	variant	UNP P0DTC2
C	478	LYS	THR	variant	UNP P0DTC2
C	484	ALA	GLU	variant	UNP P0DTC2
C	493	ARG	GLN	variant	UNP P0DTC2
C	498	ARG	GLN	variant	UNP P0DTC2
C	501	TYR	ASN	variant	UNP P0DTC2
C	505	HIS	TYR	variant	UNP P0DTC2
C	527	LYS	-	expression tag	UNP P0DTC2
C	528	LYS	-	expression tag	UNP P0DTC2

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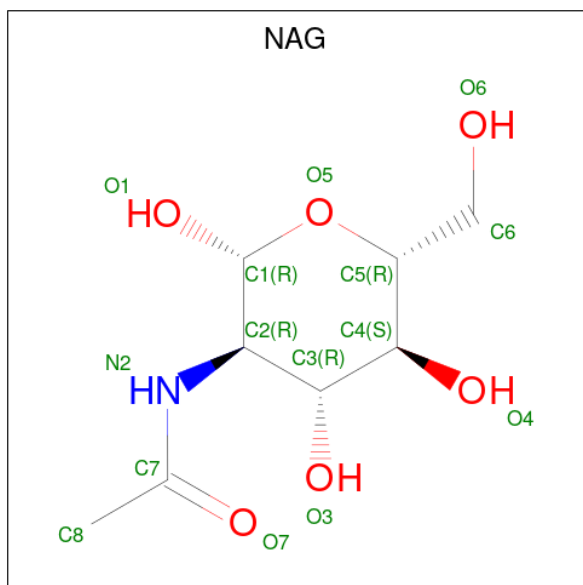
NAG

O6
OH

O1
HO

O5
O

C6

C1(R)   C5(R)

C2(R)   C4(S)
C3(R)

N2
HN

OH
O4

C7

OH
O3

C8

O
O7

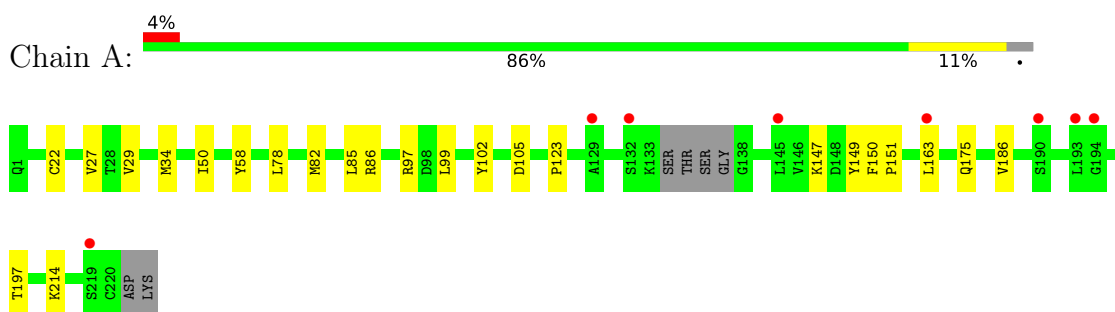| Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	E	1	Total	C	N	O	0	0
			14	8	1	5		
4	C	1	Total	C	N	O	0	0
			14	8	1	5		

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 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.
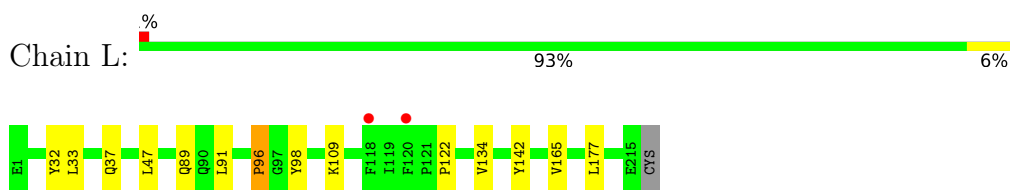
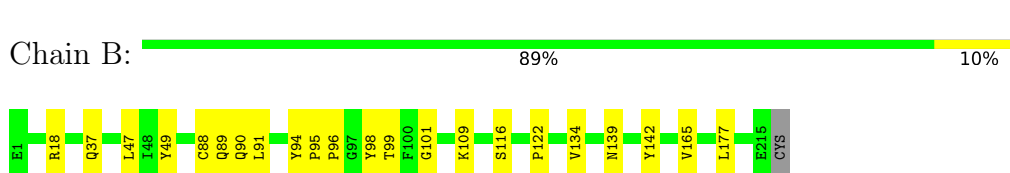• Molecule 1: COVOX-150 heavy chain
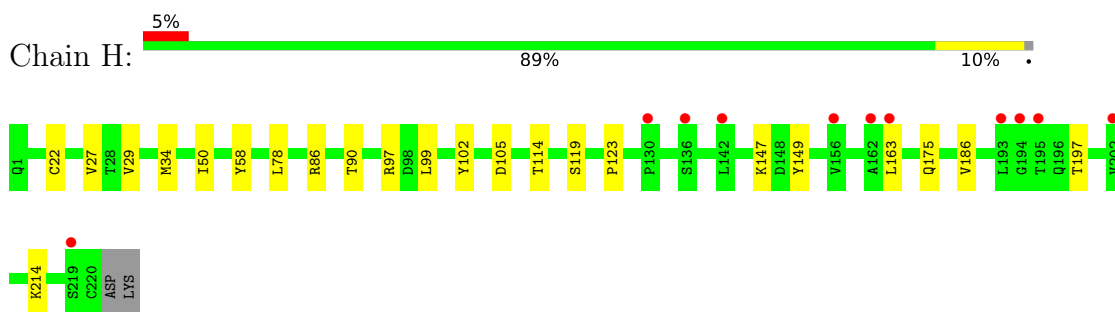
Chain H:

• Molecule 1: COVOX-150 heavy chain

Chain A:

• Molecule 2: COVOX-150 light chain

Chain L:

• Molecule 2: COVOX-150 light chain

Chain B:

• Molecule 3: Spike protein S1

Chain E:

10%

89%
7%

HIS HIS HIS HIS HIS T333 N334 L335 F338 V350 I358 Y365 S366 V367 L368 Y369 N370 F371 A372 P373 F374 F377 T385 K386 L387 F392 F400 R403 N422 T430 N439 S443 F497 G502 Q506 P507 F515 H519 A520 P521

V524 K527 LYS

• Molecule 3: Spike protein S1

Chain C:

8%

89%
8%

HIS HIS HIS HIS HIS T333 N334 V350 I358 S366 V367 L368 A372 P373 F374 F377 P384 L387 N388 D389 V401 I402 R403 G404 N405 N422 G431 N439 S443 K444 V445 R509 F515 H519 A520 P521 A522 T523 V524 K527 LYS
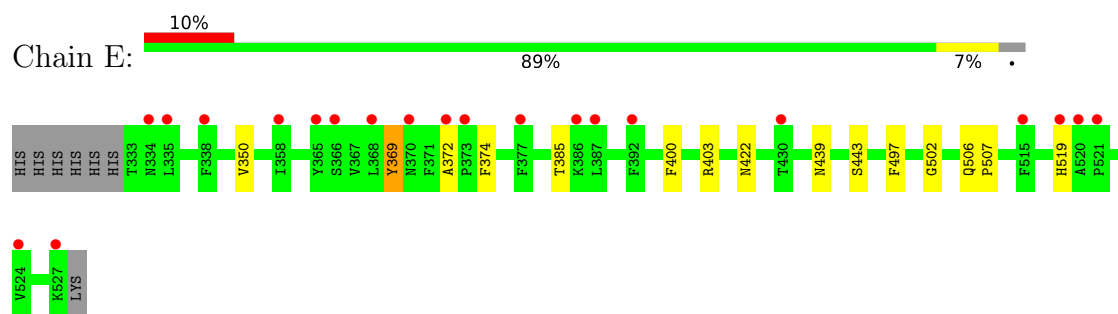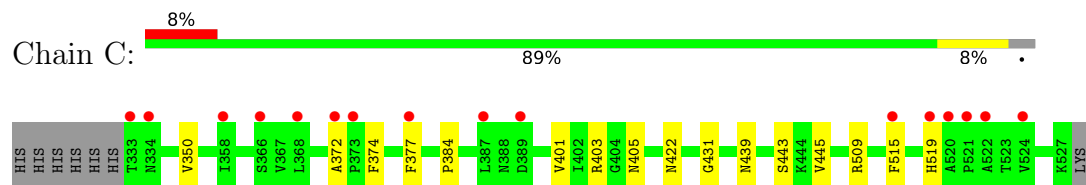
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	89.97Å 83.89Å 110.70Å 90.00° 101.97° 90.00°	Depositor
Resolution (Å)	62.27 – 3.25 66.32 – 3.25	Depositor EDS
% Data completeness (in resolution range)	99.8 (62.27-3.25) 99.8 (66.32-3.25)	Depositor EDS
R_{merge}	0.63	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.02 (at 3.26Å)	Xtriage
Refinement program	PHENIX 1.19_4092	Depositor
R, R_{free}	0.222 , 0.265 0.223 , 0.265	Depositor DCC
R_{free} test set	1205 reflections (4.69%)	wwPDB-VP
Wilson B-factor (Å^2)	87.6	Xtriage
Anisotropy	0.241	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< |L| > = 0.41, < L^2 > = 0.24$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	9639	wwPDB-VP
Average B, all atoms (Å^2)	94.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.10% 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: 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	0.25	0/1631	0.49	0/2223
1	H	0.25	0/1655	0.48	0/2257
2	B	0.25	0/1678	0.48	0/2280
2	L	0.25	0/1678	0.48	0/2280
3	C	0.25	0/1602	0.47	0/2180
3	E	0.25	0/1602	0.48	0/2180
All	All	0.25	0/9846	0.48	0/13400

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	1597	0	1573	14	0
1	H	1620	0	1594	13	0
2	B	1642	0	1598	10	0
2	L	1642	0	1598	8	0
3	C	1555	0	1474	10	0
3	E	1555	0	1474	7	0
4	C	14	0	13	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	E	14	0	13	0	0
All	All	9639	0	9337	59	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 59 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:89:GLN:HE21	2:B:98:TYR:HB3	1.57	0.69
1:H:119:SER:H	3:C:445:VAL:HG13	1.56	0.69
2:L:122:PRO:HD3	2:L:134:VAL:HG22	1.75	0.69
1:H:123:PRO:HB3	1:H:149:TYR:HB3	1.75	0.68
1:H:29:VAL:HG13	1:H:34:MET:HG3	1.79	0.65

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	212/222 (96%)	206 (97%)	6 (3%)	0	100	100
1	H	218/222 (98%)	208 (95%)	10 (5%)	0	100	100
2	B	213/216 (99%)	202 (95%)	10 (5%)	1 (0%)	29	62
2	L	213/216 (99%)	203 (95%)	9 (4%)	1 (0%)	29	62
3	C	193/202 (96%)	178 (92%)	14 (7%)	1 (0%)	29	62
3	E	193/202 (96%)	178 (92%)	13 (7%)	2 (1%)	15	47
All	All	1242/1280 (97%)	1175 (95%)	62 (5%)	5 (0%)	34	67

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	L	96	PRO
2	B	96	PRO
3	E	519	HIS
3	C	519	HIS
3	E	372	ALA

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	181/187 (97%)	180 (99%)	1 (1%)	86	91
1	H	184/187 (98%)	183 (100%)	1 (0%)	88	93
2	B	187/188 (100%)	185 (99%)	2 (1%)	73	84
2	L	187/188 (100%)	186 (100%)	1 (0%)	88	93
3	C	166/174 (95%)	166 (100%)	0	100	100
3	E	166/174 (95%)	164 (99%)	2 (1%)	71	83
All	All	1071/1098 (98%)	1064 (99%)	7 (1%)	84	90

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	E	374	PHE
1	A	86	ARG
2	B	91	LEU
2	B	18	ARG
3	E	369	TYR

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 ⓘ

2 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	NAG	C	601	3	14,14,15	0.18	0	17,19,21	0.42	0
4	NAG	E	601	3	14,14,15	0.25	0	17,19,21	0.38	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	NAG	C	601	3	-	1/6/23/26	0/1/1/1
4	NAG	E	601	3	-	1/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	C	601	NAG	O5-C5-C6-O6
4	E	601	NAG	O5-C5-C6-O6

There are no ring outliers.

No monomer is involved in short contacts.

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 (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(Å²)	Q<0.9
1	A	216/222 (97%)	0.29	8 (3%)	41	38	59, 85, 138, 263	0
1	H	220/222 (99%)	0.38	11 (5%)	28	26	67, 85, 141, 235	0
2	B	215/216 (99%)	0.18	0	100	100	65, 85, 108, 136	0
2	L	215/216 (99%)	0.18	2 (0%)	84	84	68, 92, 114, 131	0
3	C	195/202 (96%)	0.68	16 (8%)	11	11	69, 89, 173, 227	0
3	E	195/202 (96%)	0.56	21 (10%)	5	6	71, 97, 158, 196	0
All	All	1256/1280 (98%)	0.37	58 (4%)	32	30	59, 89, 147, 263	0

The worst 5 of 58 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	521	PRO	9.6
3	C	333	THR	7.4
3	C	368	LEU	7.0
3	C	520	ALA	6.7
3	E	519	HIS	6.5

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	NAG	C	601	14/15	0.71	0.31	133,141,146,148	0
4	NAG	E	601	14/15	0.82	0.26	112,128,134,135	0

6.5 Other polymers (i)

There are no such residues in this entry.