



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 4, 2022 – 04:20 PM JST

PDB ID : 8GX9  
Title : Crystal structure of SARS-CoV-2 RBD with P2C-1F11 and P2B-1G5  
Authors : Wang, X.; Zhang, L.; Ge, J.  
Deposited on : 2022-09-19  
Resolution : 4.01 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](mailto: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 : 2.31.2  
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) : 2.31.2

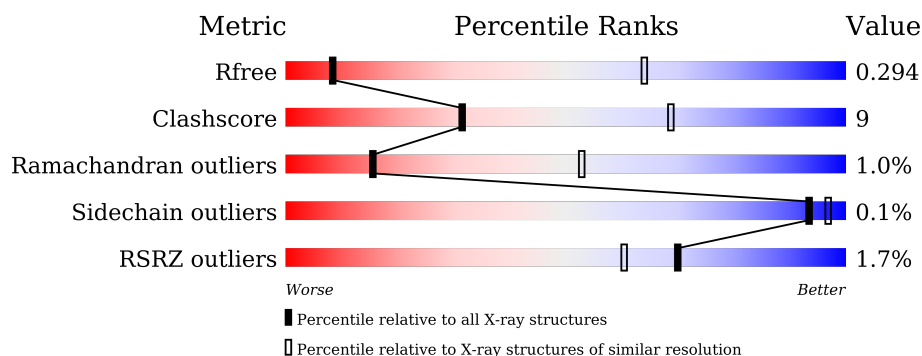
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 4.01 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	1098 (4.34-3.70)
Clashscore	141614	1159 (4.34-3.70)
Ramachandran outliers	138981	1118 (4.34-3.70)
Sidechain outliers	138945	1108 (4.34-3.70)
RSRZ outliers	127900	1034 (4.38-3.66)

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  $\geq 3$ , 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	211	<div> <div>2%</div> <div> <div></div> <div>71%</div> <div>21%</div> <div>8%</div> </div> </div>
1	E	211	<div> <div>72%</div> <div>18%</div> <div>9%</div> </div>
2	F	214	<div> <div>86%</div> <div>14%</div> </div>
2	L	214	<div> <div>83%</div> <div>16%</div> <div>.</div> </div>
3	H	216	<div> <div>3%</div> <div> <div></div> <div>74%</div> <div>22%</div> <div>.</div> </div> </div>
3	I	216	<div> <div>74%</div> <div>22%</div> <div>..</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
4	M	214	 81% 17% 2%
4	N	214	 79% 19% 2%
5	O	218	 74% 24% 8%
5	P	218	 80% 19% 1%

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 15761 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 Spike protein S1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	194	Total	C	N	O	S	0	0	0
			1536	984	256	288	8			
1	E	191	Total	C	N	O	S	0	0	0
			1510	966	251	285	8			

- Molecule 2 is a protein called light chain of P2C-1F11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	214	Total	C	N	O	S	0	0	0
			1632	1019	276	332	5			
2	L	214	Total	C	N	O	S	0	0	0
			1633	1020	276	332	5			

- Molecule 3 is a protein called heavy chain of P2C-1F11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	H	209	Total	C	N	O	S	0	0	0
			1550	975	262	306	7			
3	I	209	Total	C	N	O	S	0	0	0
			1550	975	262	306	7			

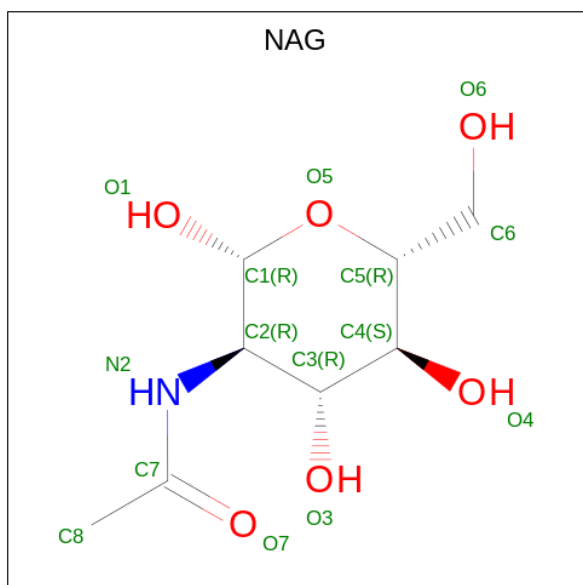
- Molecule 4 is a protein called light chain of P2B-1G5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	M	210	Total	C	N	O	S	0	4	0
			1570	987	260	319	4			
4	N	210	Total	C	N	O	S	0	4	0
			1579	991	264	320	4			

- Molecule 5 is a protein called heavy chain of P2B-1G5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	O	214	Total	C	N	O	S	0	0	0
			1574	993	257	316	8			
5	P	217	Total	C	N	O	S	0	0	0
			1599	1009	261	321	8			

- Molecule 6 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).

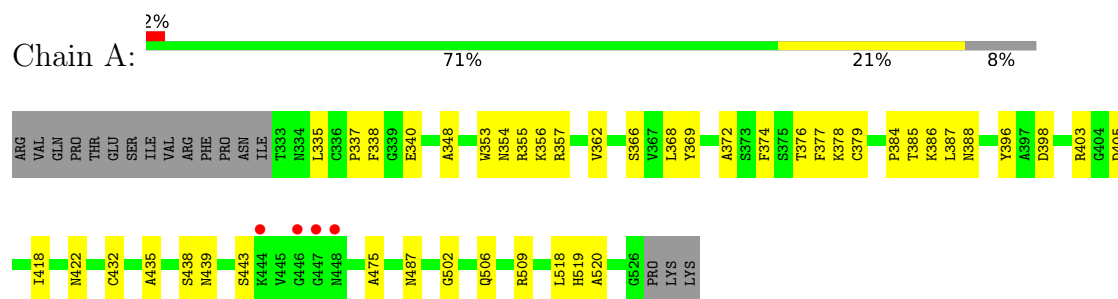


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	A	1	Total	C	N	O	0	0
			14	8	1	5		
6	E	1	Total	C	N	O	0	0
			14	8	1	5		

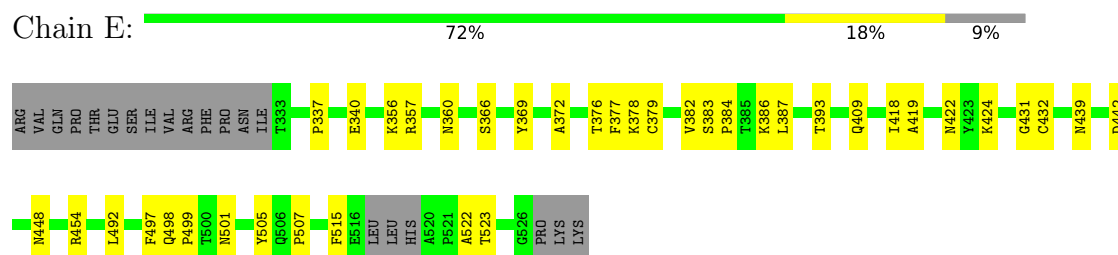
### 3 Residue-property plots

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.

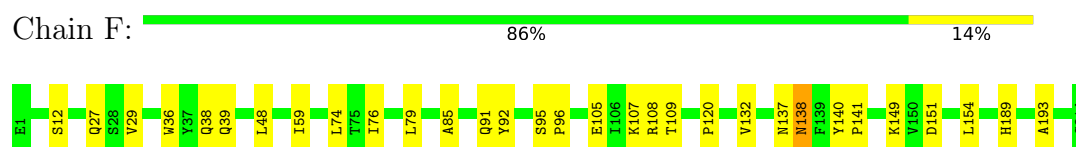
- Molecule 1: Spike protein S1



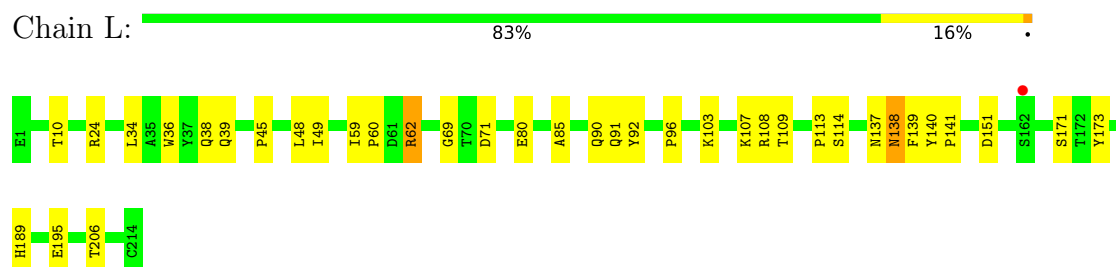
- Molecule 1: Spike protein S1



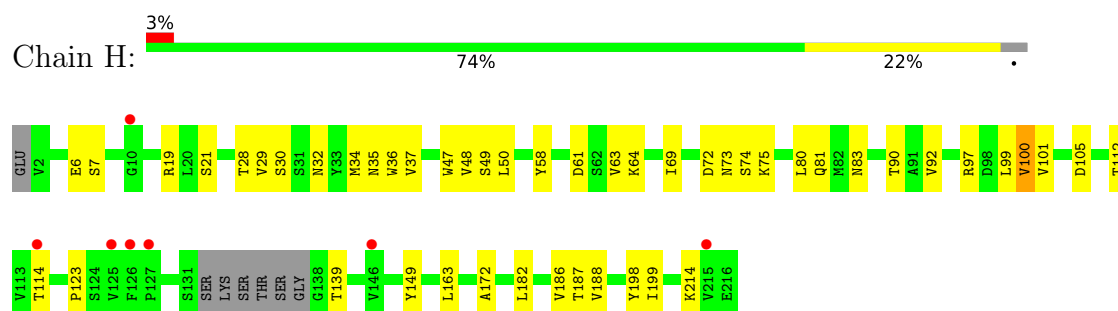
- Molecule 2: light chain of P2C-1F11



- Molecule 2: light chain of P2C-1F11



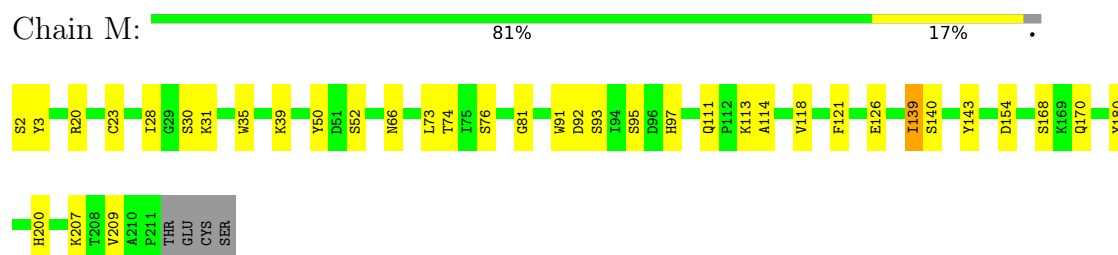
- Molecule 3: heavy chain of P2C-1F11

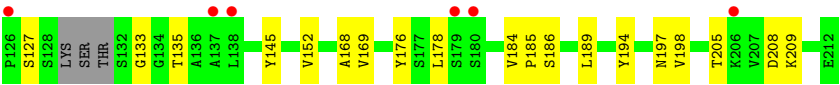


- Molecule 3: heavy chain of P2C-1F11

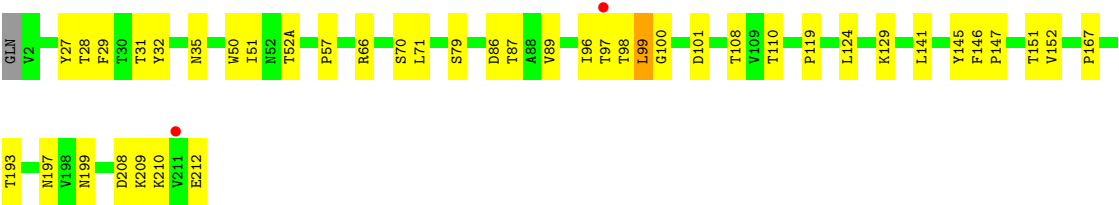
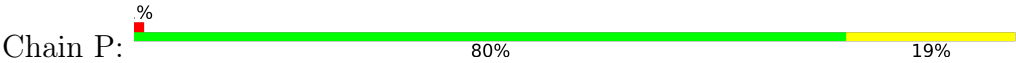


- Molecule 4: light chain of P2B-1G5





● Molecule 5: heavy chain of P2B-1G5





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	110.51Å 154.86Å 158.42Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.08 – 4.01 49.98 – 4.01	Depositor EDS
% Data completeness (in resolution range)	97.3 (49.08-4.01) 98.3 (49.98-4.01)	Depositor EDS
$R_{merge}$	0.29	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.69 (at 4.00Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, $R_{free}$	0.238 , 0.297 0.242 , 0.294	Depositor DCC
$R_{free}$ test set	1244 reflections (5.40%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	105.2	Xtriage
Anisotropy	1.152	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 87.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.26$	Xtriage
Estimated twinning fraction	0.048 for -h,l,k	Xtriage
$F_o, F_c$ correlation	0.87	EDS
Total number of atoms	15761	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	139.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.85% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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

### 5.1 Standard geometry

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	0.29	0/1579	0.49	0/2149
1	E	0.27	0/1551	0.45	0/2109
2	F	0.28	0/1667	0.51	0/2263
2	L	0.27	0/1668	0.50	0/2264
3	H	0.27	0/1583	0.50	0/2158
3	I	0.27	0/1583	0.51	0/2158
4	M	0.29	0/1627	0.52	1/2229 (0.0%)
4	N	0.27	0/1636	0.47	0/2238
5	O	0.29	0/1610	0.52	0/2198
5	P	0.27	0/1636	0.52	0/2234
All	All	0.28	0/16140	0.50	1/22000 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
4	M	0	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	M	139	ILE	C-N-CA	-7.17	103.77	121.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	M	139	ILE	Mainchain

## 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	1536	0	1452	44	0
1	E	1510	0	1422	36	0
2	F	1632	0	1583	23	0
2	L	1633	0	1585	25	0
3	H	1550	0	1525	34	0
3	I	1550	0	1525	37	0
4	M	1570	0	1497	33	0
4	N	1579	0	1515	27	0
5	O	1574	0	1534	33	0
5	P	1599	0	1569	32	0
6	A	14	0	13	0	0
6	E	14	0	13	0	0
All	All	15761	0	15233	279	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 9.

The worst 5 of 279 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)
3:I:138:GLY:O	3:I:139:THR:OG1	1.78	0.99
5:O:66:ARG:NH2	5:O:86:ASP:OD2	2.07	0.85
1:A:379:CYS:HB2	1:E:378:LYS:HE3	1.58	0.84
1:A:378:LYS:HE3	1:E:379:CYS:HB2	1.59	0.84
2:F:92:TYR:HE2	3:H:100:VAL:HG13	1.45	0.82

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	192/211 (91%)	174 (91%)	17 (9%)	1 (0%)	29	67
1	E	187/211 (89%)	171 (91%)	16 (9%)	0	100	100
2	F	212/214 (99%)	200 (94%)	10 (5%)	2 (1%)	17	55
2	L	212/214 (99%)	198 (93%)	12 (6%)	2 (1%)	17	55
3	H	205/216 (95%)	191 (93%)	13 (6%)	1 (0%)	29	67
3	I	205/216 (95%)	186 (91%)	15 (7%)	4 (2%)	7	40
4	M	212/214 (99%)	201 (95%)	9 (4%)	2 (1%)	17	55
4	N	212/214 (99%)	194 (92%)	15 (7%)	3 (1%)	11	46
5	O	210/218 (96%)	189 (90%)	18 (9%)	3 (1%)	11	46
5	P	215/218 (99%)	198 (92%)	15 (7%)	2 (1%)	17	55
All	All	2062/2146 (96%)	1902 (92%)	140 (7%)	20 (1%)	15	53

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	519	HIS
2	F	109	THR
3	H	100	VAL
3	I	100	VAL
3	I	121	LYS

### 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	167/184 (91%)	167 (100%)	0	100	100
1	E	164/184 (89%)	164 (100%)	0	100	100
2	F	184/185 (100%)	184 (100%)	0	100	100
2	L	184/185 (100%)	183 (100%)	1 (0%)	88	93
3	H	175/181 (97%)	174 (99%)	1 (1%)	86	92
3	I	175/181 (97%)	175 (100%)	0	100	100
4	M	176/180 (98%)	176 (100%)	0	100	100
4	N	178/180 (99%)	178 (100%)	0	100	100
5	O	179/184 (97%)	179 (100%)	0	100	100
5	P	183/184 (100%)	183 (100%)	0	100	100
All	All	1765/1828 (97%)	1763 (100%)	2 (0%)	93	97

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	H	139	THR
2	L	62	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
3	H	203	ASN
5	O	53	ASN

### 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	$\# Z  > 2$
6	NAG	A	601	1	14,14,15	0.36	0	17,19,21	0.65	1 (5%)
6	NAG	E	601	1	14,14,15	0.36	0	17,19,21	0.65	1 (5%)

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
6	NAG	A	601	1	-	0/6/23/26	0/1/1/1
6	NAG	E	601	1	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	E	601	NAG	C1-O5-C5	2.32	115.33	112.19
6	A	601	NAG	C1-O5-C5	2.31	115.33	112.19

There are no chirality outliers.

There are no torsion outliers.

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	194/211 (91%)	-0.03	4 (2%) 63 54	100, 130, 172, 198	0
1	E	191/211 (90%)	-0.05	0 100 100	95, 128, 172, 188	0
2	F	214/214 (100%)	-0.10	0 100 100	98, 132, 157, 173	0
2	L	214/214 (100%)	-0.09	1 (0%) 91 85	106, 133, 161, 177	0
3	H	209/216 (96%)	0.13	7 (3%) 46 36	30, 126, 155, 171	0
3	I	209/216 (96%)	-0.16	0 100 100	30, 125, 157, 183	0
4	M	210/214 (98%)	-0.05	0 100 100	30, 148, 173, 184	0
4	N	210/214 (98%)	0.16	4 (1%) 66 58	112, 149, 167, 175	0
5	O	214/218 (98%)	0.31	17 (7%) 12 11	117, 145, 172, 188	0
5	P	217/218 (99%)	0.13	2 (0%) 84 77	139, 163, 192, 202	0
All	All	2082/2146 (97%)	0.03	35 (1%) 70 60	30, 140, 173, 202	0

The worst 5 of 35 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	O	100(A)	GLY	3.5
5	O	138	LEU	3.3
5	O	179	SER	3.2
3	H	127	PRO	3.1
5	O	137	ALA	3.1

### 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, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
6	NAG	A	601	14/15	0.76	0.33	133,151,158,159	0
6	NAG	E	601	14/15	0.77	0.39	127,142,166,175	0

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

There are no such residues in this entry.