



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 8, 2020 – 01:28 PM BST

PDB ID : 6YZ7  
Title : H11-D4, SARS-CoV-2 RBD, CR3022 ternary complex  
Authors : Naismith, J.H.; Ren, J.; Zhou, D.; Zhao, Y.; Stuart, D.I.  
Deposited on : 2020-05-06  
Resolution : 3.30 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 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 : 2.13.1  
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) : 2.13.1

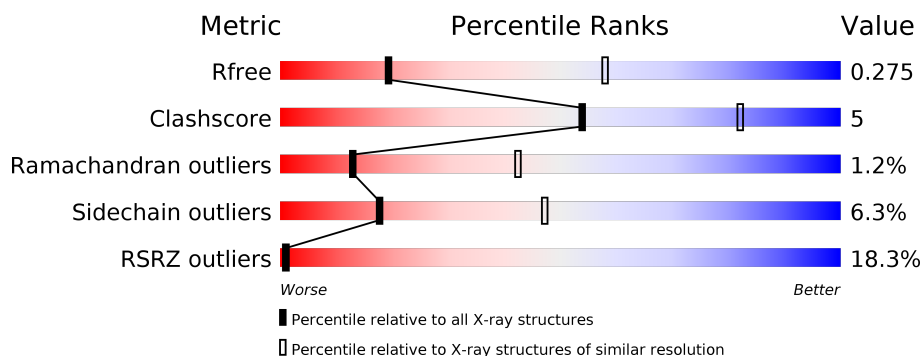
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.30 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	1149 (3.34-3.26)
Clashscore	141614	1205 (3.34-3.26)
Ramachandran outliers	138981	1183 (3.34-3.26)
Sidechain outliers	138945	1182 (3.34-3.26)
RSRZ outliers	127900	1115 (3.34-3.26)

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 on 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	AAA	210	<div> <div>19%</div> <div>79%</div> <div>12%</div> <div>7%</div> </div>
1	EEE	210	<div> <div>10%</div> <div>80%</div> <div>12%</div> <div>7%</div> </div>
2	BBB	229	<div> <div>14%</div> <div>82%</div> <div>11%</div> <div>6%</div> </div>
2	HHH	229	<div> <div>10%</div> <div>79%</div> <div>14%</div> <div>6%</div> </div>
3	CCC	220	<div> <div>11%</div> <div>82%</div> <div>16%</div> <div>.</div> </div>
3	LLL	220	<div> <div>12%</div> <div>82%</div> <div>16%</div> <div>.</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	DDD	134	<div> <div>70%</div> <div>75%</div> <div>18%</div> <div>5%</div> </div>
4	FFF	134	<div> <div>12%</div> <div>76%</div> <div>16%</div> <div>5%</div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 11718 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 glycoprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	195	Total	C	N	O	S	0	0	0
			1545	991	258	288	8			
1	EEE	195	Total	C	N	O	S	0	0	0
			1545	991	258	288	8			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	533	LYS	-	expression tag	UNP P0DTC2
AAA	534	HIS	-	expression tag	UNP P0DTC2
AAA	535	HIS	-	expression tag	UNP P0DTC2
AAA	536	HIS	-	expression tag	UNP P0DTC2
AAA	537	HIS	-	expression tag	UNP P0DTC2
AAA	538	HIS	-	expression tag	UNP P0DTC2
AAA	539	HIS	-	expression tag	UNP P0DTC2
EEE	533	LYS	-	expression tag	UNP P0DTC2
EEE	534	HIS	-	expression tag	UNP P0DTC2
EEE	535	HIS	-	expression tag	UNP P0DTC2
EEE	536	HIS	-	expression tag	UNP P0DTC2
EEE	537	HIS	-	expression tag	UNP P0DTC2
EEE	538	HIS	-	expression tag	UNP P0DTC2
EEE	539	HIS	-	expression tag	UNP P0DTC2

- Molecule 2 is a protein called Antibody Cr3022.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	BBB	216	Total	C	N	O	S	0	1	0
			1609	1023	261	317	8			
2	HHH	216	Total	C	N	O	S	0	1	0
			1609	1023	261	317	8			

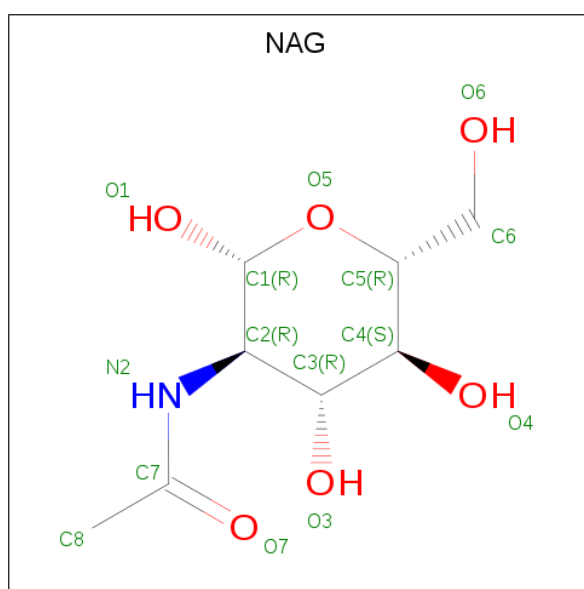
- Molecule 3 is a protein called Antibody light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	CCC	219	Total	C	N	O	S	0	0	0
			1703	1070	282	347	4			
3	LLL	219	Total	C	N	O	S	0	0	0
			1703	1070	282	347	4			

- Molecule 4 is a protein called Nanobody.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	DDD	127	Total	C	N	O	S	0	0	0
			988	621	173	189	5			
4	FFF	127	Total	C	N	O	S	0	0	0
			988	621	173	189	5			

- Molecule 5 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).

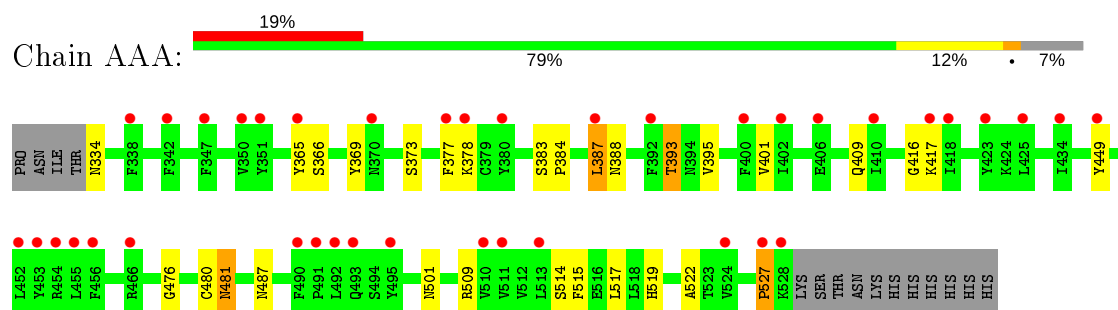


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	AAA	1	Total	C	N	O	0	0
			14	8	1	5		
5	EEE	1	Total	C	N	O	0	0
			14	8	1	5		

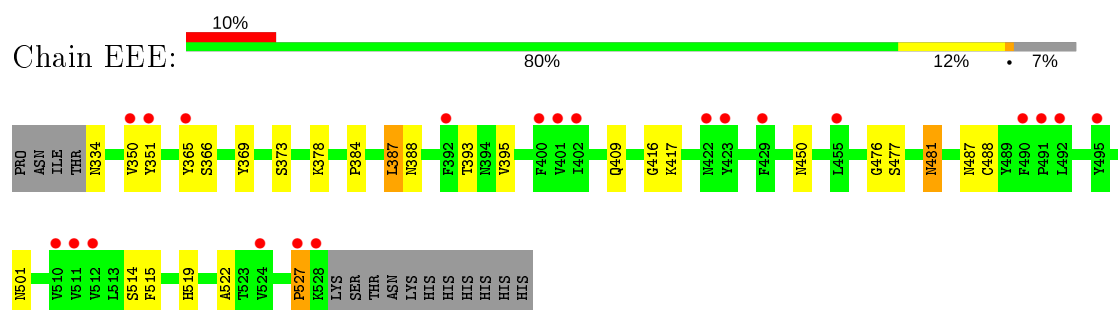
### 3 Residue-property plots [i](#)

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.

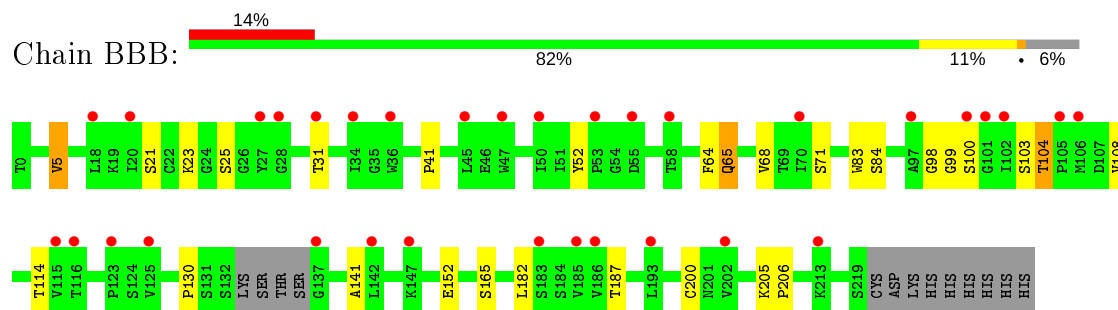
- Molecule 1: Spike glycoprotein



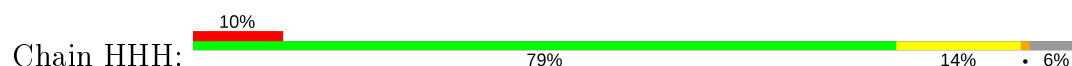
- Molecule 1: Spike glycoprotein

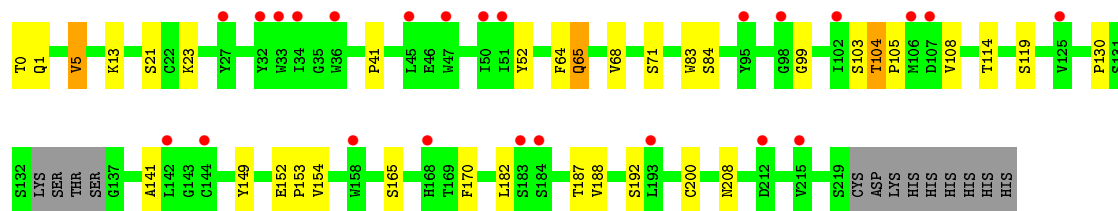


- Molecule 2: Antibody Cr3022

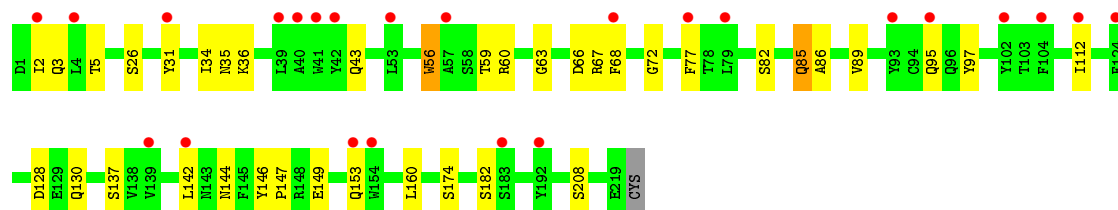
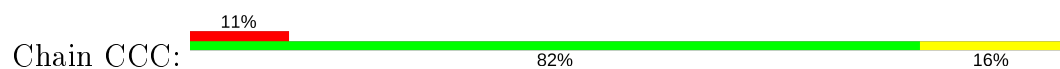


- Molecule 2: Antibody Cr3022

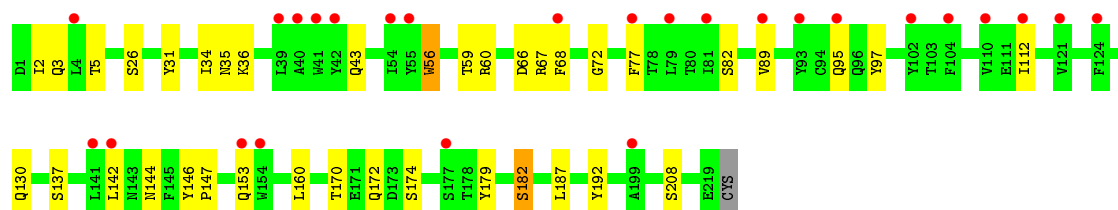
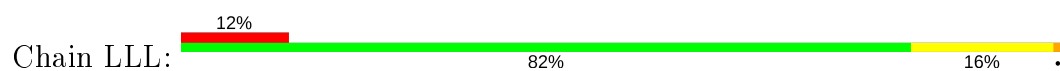




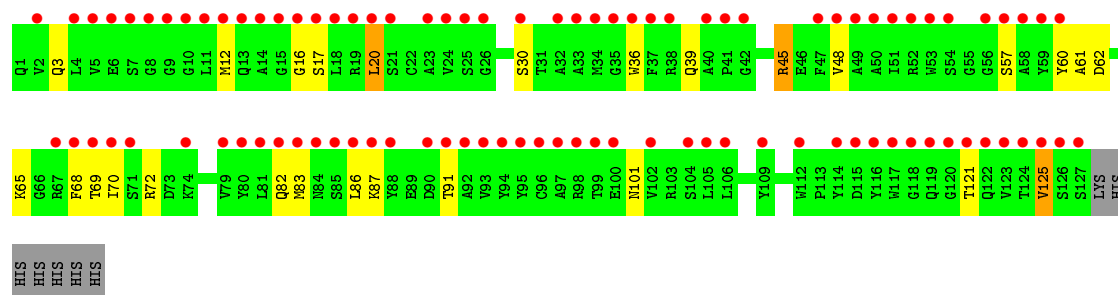
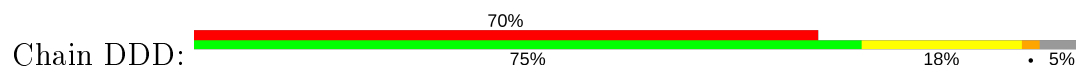
• Molecule 3: Antibody light chain



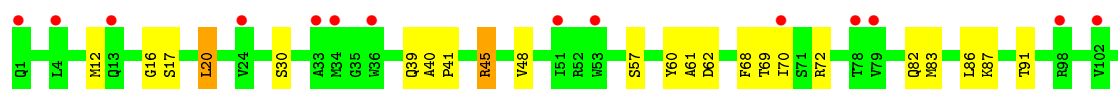
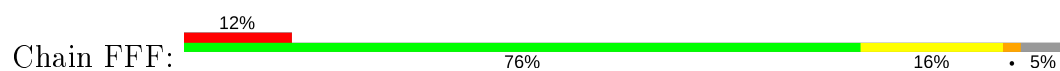
• Molecule 3: Antibody light chain



• Molecule 4: Nanobody



• Molecule 4: Nanobody







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	154.60 Å 154.60 Å 229.31 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	128.19 – 3.30 128.19 – 3.30	Depositor EDS
% Data completeness (in resolution range)	96.3 (128.19-3.30) 96.3 (128.19-3.30)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.95 (at 3.33 Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, $R_{free}$	0.238 , 0.269 0.242 , 0.275	Depositor DCC
$R_{free}$ test set	1945 reflections (4.74%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	128.1	Xtriage
Anisotropy	0.005	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 129.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.40$ , $\langle L^2 \rangle = 0.23$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	11718	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	157.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 32.63 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 9.1168e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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	AAA	0.66	0/1589	0.79	0/2162
1	EEE	0.65	0/1589	0.80	0/2162
2	BBB	0.67	0/1654	0.79	0/2253
2	HHH	0.66	0/1654	0.81	0/2253
3	CCC	0.66	0/1741	0.80	0/2367
3	LLL	0.66	0/1741	0.80	0/2367
4	DDD	0.68	0/1010	0.81	0/1366
4	FFF	0.65	0/1010	0.82	0/1366
All	All	0.66	0/11988	0.80	0/16296

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
2	BBB	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	BBB	98	GLY	Peptide

## 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	AAA	1545	0	1465	15	0
1	EEE	1545	0	1465	13	0
2	BBB	1609	0	1585	10	0
2	HHH	1609	0	1585	16	0
3	CCC	1703	0	1649	14	0
3	LLL	1703	0	1649	16	0
4	DDD	988	0	951	24	0
4	FFF	988	0	951	24	0
5	AAA	14	0	13	0	0
5	EEE	14	0	13	0	0
All	All	11718	0	11326	122	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 122 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)
4:DDD:60:TYR:OH	4:DDD:70:ILE:N	1.68	1.26
4:FFF:60:TYR:OH	4:FFF:70:ILE:N	1.69	1.23
4:DDD:60:TYR:OH	4:DDD:69:THR:CA	1.94	1.15
4:FFF:60:TYR:OH	4:FFF:69:THR:CA	1.94	1.14
4:FFF:60:TYR:OH	4:FFF:69:THR:HA	1.47	1.13

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	AAA	193/210 (92%)	169 (88%)	23 (12%)	1 (0%)	29	61
1	EEE	193/210 (92%)	170 (88%)	22 (11%)	1 (0%)	29	61
2	BBB	213/229 (93%)	189 (89%)	21 (10%)	3 (1%)	11	38
2	HHH	213/229 (93%)	189 (89%)	21 (10%)	3 (1%)	11	38
3	CCC	217/220 (99%)	197 (91%)	14 (6%)	6 (3%)	5	25
3	LLL	217/220 (99%)	198 (91%)	15 (7%)	4 (2%)	8	35
4	DDD	125/134 (93%)	116 (93%)	9 (7%)	0	100	100
4	FFF	125/134 (93%)	116 (93%)	9 (7%)	0	100	100
All	All	1496/1586 (94%)	1344 (90%)	134 (9%)	18 (1%)	13	42

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	BBB	165	SER
3	CCC	35	ASN
3	LLL	35	ASN
1	AAA	527	PRO
2	BBB	65	GLN

### 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	AAA	168/183 (92%)	157 (94%)	11 (6%)	17	46
1	EEE	168/183 (92%)	158 (94%)	10 (6%)	19	49
2	BBB	182/194 (94%)	170 (93%)	12 (7%)	16	46
2	HHH	182/194 (94%)	171 (94%)	11 (6%)	19	49
3	CCC	194/195 (100%)	183 (94%)	11 (6%)	20	51
3	LLL	194/195 (100%)	185 (95%)	9 (5%)	27	58
4	DDD	101/108 (94%)	91 (90%)	10 (10%)	8	28

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	FFF	101/108 (94%)	94 (93%)	7 (7%)	15	44
All	All	1290/1360 (95%)	1209 (94%)	81 (6%)	18	47

5 of 81 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	DDD	30	SER
1	EEE	373	SER
3	LLL	31	TYR
4	DDD	45	ARG
4	DDD	72	ARG

Some 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
5	NAG	EEE	601	1	14,14,15	0.32	0	17,19,21	1.03	1 (5%)
5	NAG	AAA	601	1	14,14,15	0.36	0	17,19,21	1.08	2 (11%)

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
5	NAG	EEE	601	1	-	2/6/23/26	0/1/1/1
5	NAG	AAA	601	1	-	2/6/23/26	0/1/1/1

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	EEE	601	NAG	C4-C3-C2	-2.98	106.65	111.02
5	AAA	601	NAG	O5-C1-C2	-2.59	107.21	111.29
5	AAA	601	NAG	O5-C5-C6	2.10	110.50	107.20

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	AAA	601	NAG	O5-C5-C6-O6
5	AAA	601	NAG	C4-C5-C6-O6
5	EEE	601	NAG	O5-C5-C6-O6
5	EEE	601	NAG	C4-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 ⓘ

### 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<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	AAA	195/210 (92%)	1.08	39 (20%) 1 1	130, 167, 212, 245	0
1	EEE	195/210 (92%)	0.90	21 (10%) 5 5	114, 135, 182, 226	0
2	BBB	216/229 (94%)	0.90	33 (15%) 2 2	111, 150, 183, 206	0
2	HHH	216/229 (94%)	0.91	24 (11%) 5 5	99, 134, 193, 216	0
3	CCC	219/220 (99%)	0.90	24 (10%) 5 5	113, 148, 174, 200	0
3	LLL	219/220 (99%)	0.91	26 (11%) 4 4	104, 143, 172, 197	0
4	DDD	127/134 (94%)	3.96	94 (74%) 0 0	214, 271, 304, 315	0
4	FFF	127/134 (94%)	1.04	16 (12%) 3 3	117, 140, 171, 207	0
All	All	1514/1586 (95%)	1.19	277 (18%) 1 1	99, 148, 267, 315	0

The worst 5 of 277 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	DDD	35	GLY	15.8
4	DDD	33	ALA	14.0
4	DDD	34	MET	13.9
4	DDD	81	LEU	12.3
4	DDD	97	ALA	12.3

### 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 [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(Å <sup>2</sup> )	Q<0.9
5	NAG	AAA	601	14/15	0.85	0.28	162,172,176,179	0
5	NAG	EEE	601	14/15	0.89	0.30	140,150,156,157	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.