



# Full wwPDB X-ray Structure Validation Report ⓘ

May 18, 2020 – 10:04 AM EDT

PDB ID : 6WZQ  
Title : Structure of SARS-CoV-2 Nucleocapsid dimerization domain, P21 form  
Authors : Ye, Q.; Corbett, K.D.  
Deposited on : 2020-05-14  
Resolution : 1.45 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 : 2.10.1  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
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) : 2.10.1

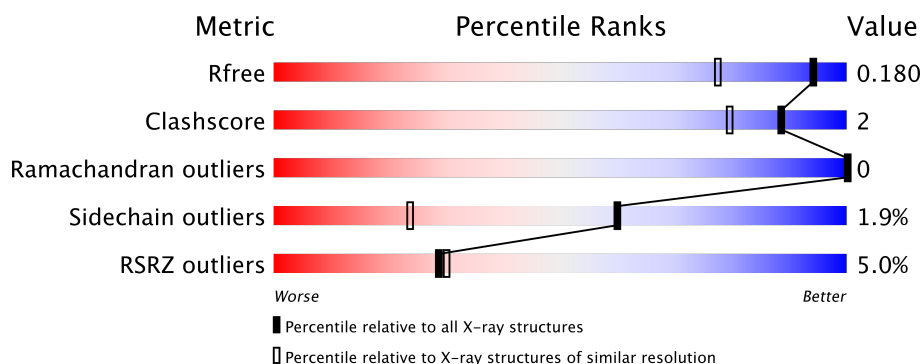
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.45 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))
$R_{free}$	111664	1761 (1.48-1.44)
Clashscore	122126	1816 (1.48-1.44)
Ramachandran outliers	120053	1793 (1.48-1.44)
Sidechain outliers	120020	1793 (1.48-1.44)
RSRZ outliers	108989	1733 (1.48-1.44)

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	A	137	<div> <div>4%</div> <div> <div></div> <div>80%</div> <div>• •</div> <div>17%</div> </div> </div>
1	B	137	<div> <div>4%</div> <div> <div></div> <div>78%</div> <div>5%</div> <div>17%</div> </div> </div>
1	C	137	<div> <div>3%</div> <div> <div></div> <div>81%</div> <div>•</div> <div>15%</div> </div> </div>
1	D	137	<div> <div>7%</div> <div> <div></div> <div>77%</div> <div>5%</div> <div>•</div> <div>16%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 7768 atoms, of which 3580 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 Nucleoprotein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	114	Total	C	H	N	O	S	0	1	0
			1804	577	892	162	171	2			
1	B	114	Total	C	H	N	O	S	0	1	0
			1804	577	892	162	171	2			
1	C	116	Total	C	H	N	O	S	0	1	0
			1822	583	899	164	174	2			
1	D	115	Total	C	H	N	O	S	0	1	0
			1815	580	897	163	173	2			

There are 76 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	228	MET	-	expression tag	UNP P0DTC9
A	229	LYS	-	expression tag	UNP P0DTC9
A	230	SER	-	expression tag	UNP P0DTC9
A	231	SER	-	expression tag	UNP P0DTC9
A	232	HIS	-	expression tag	UNP P0DTC9
A	233	HIS	-	expression tag	UNP P0DTC9
A	234	HIS	-	expression tag	UNP P0DTC9
A	235	HIS	-	expression tag	UNP P0DTC9
A	236	HIS	-	expression tag	UNP P0DTC9
A	237	HIS	-	expression tag	UNP P0DTC9
A	238	GLU	-	expression tag	UNP P0DTC9
A	239	ASN	-	expression tag	UNP P0DTC9
A	240	LEU	-	expression tag	UNP P0DTC9
A	241	TYR	-	expression tag	UNP P0DTC9
A	242	PHE	-	expression tag	UNP P0DTC9
A	243	GLN	-	expression tag	UNP P0DTC9
A	244	SER	-	expression tag	UNP P0DTC9
A	245	ASN	-	expression tag	UNP P0DTC9
A	246	ALA	-	expression tag	UNP P0DTC9
B	228	MET	-	expression tag	UNP P0DTC9
B	229	LYS	-	expression tag	UNP P0DTC9

*Continued on next page...*

*Continued from previous page...*

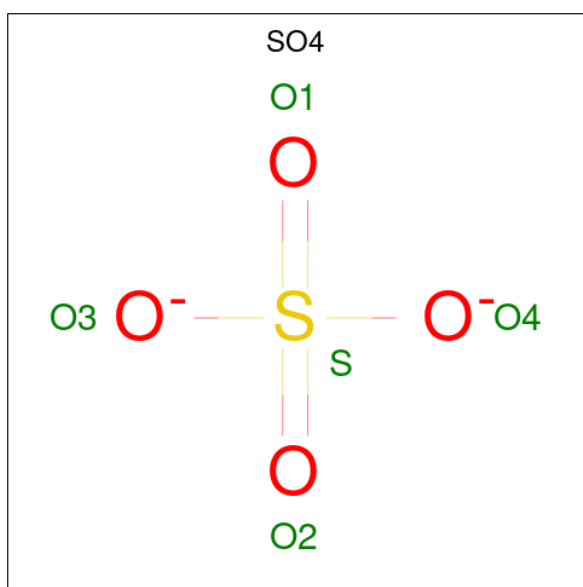
Chain	Residue	Modelled	Actual	Comment	Reference
B	230	SER	-	expression tag	UNP P0DTC9
B	231	SER	-	expression tag	UNP P0DTC9
B	232	HIS	-	expression tag	UNP P0DTC9
B	233	HIS	-	expression tag	UNP P0DTC9
B	234	HIS	-	expression tag	UNP P0DTC9
B	235	HIS	-	expression tag	UNP P0DTC9
B	236	HIS	-	expression tag	UNP P0DTC9
B	237	HIS	-	expression tag	UNP P0DTC9
B	238	GLU	-	expression tag	UNP P0DTC9
B	239	ASN	-	expression tag	UNP P0DTC9
B	240	LEU	-	expression tag	UNP P0DTC9
B	241	TYR	-	expression tag	UNP P0DTC9
B	242	PHE	-	expression tag	UNP P0DTC9
B	243	GLN	-	expression tag	UNP P0DTC9
B	244	SER	-	expression tag	UNP P0DTC9
B	245	ASN	-	expression tag	UNP P0DTC9
B	246	ALA	-	expression tag	UNP P0DTC9
C	228	MET	-	expression tag	UNP P0DTC9
C	229	LYS	-	expression tag	UNP P0DTC9
C	230	SER	-	expression tag	UNP P0DTC9
C	231	SER	-	expression tag	UNP P0DTC9
C	232	HIS	-	expression tag	UNP P0DTC9
C	233	HIS	-	expression tag	UNP P0DTC9
C	234	HIS	-	expression tag	UNP P0DTC9
C	235	HIS	-	expression tag	UNP P0DTC9
C	236	HIS	-	expression tag	UNP P0DTC9
C	237	HIS	-	expression tag	UNP P0DTC9
C	238	GLU	-	expression tag	UNP P0DTC9
C	239	ASN	-	expression tag	UNP P0DTC9
C	240	LEU	-	expression tag	UNP P0DTC9
C	241	TYR	-	expression tag	UNP P0DTC9
C	242	PHE	-	expression tag	UNP P0DTC9
C	243	GLN	-	expression tag	UNP P0DTC9
C	244	SER	-	expression tag	UNP P0DTC9
C	245	ASN	-	expression tag	UNP P0DTC9
C	246	ALA	-	expression tag	UNP P0DTC9
D	228	MET	-	expression tag	UNP P0DTC9
D	229	LYS	-	expression tag	UNP P0DTC9
D	230	SER	-	expression tag	UNP P0DTC9
D	231	SER	-	expression tag	UNP P0DTC9
D	232	HIS	-	expression tag	UNP P0DTC9
D	233	HIS	-	expression tag	UNP P0DTC9

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	234	HIS	-	expression tag	UNP P0DTC9
D	235	HIS	-	expression tag	UNP P0DTC9
D	236	HIS	-	expression tag	UNP P0DTC9
D	237	HIS	-	expression tag	UNP P0DTC9
D	238	GLU	-	expression tag	UNP P0DTC9
D	239	ASN	-	expression tag	UNP P0DTC9
D	240	LEU	-	expression tag	UNP P0DTC9
D	241	TYR	-	expression tag	UNP P0DTC9
D	242	PHE	-	expression tag	UNP P0DTC9
D	243	GLN	-	expression tag	UNP P0DTC9
D	244	SER	-	expression tag	UNP P0DTC9
D	245	ASN	-	expression tag	UNP P0DTC9
D	246	ALA	-	expression tag	UNP P0DTC9

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



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

Continued on next page...

*Continued from previous page...*

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	D	1	Total	O	S	0	0
			5	4	1		

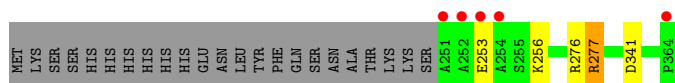
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	113	Total	O	0	0
			113	113		
3	B	140	Total	O	0	0
			140	140		
3	C	152	Total	O	0	0
			152	152		
3	D	88	Total	O	0	0
			88	88		

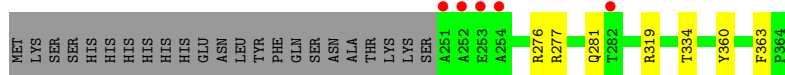
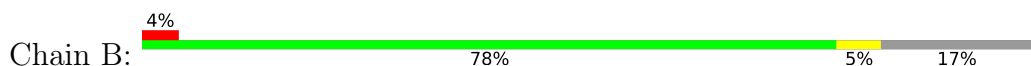
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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.

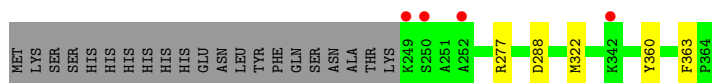
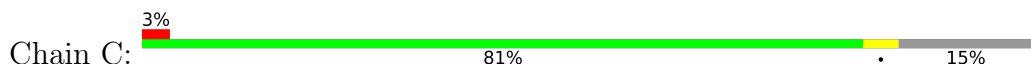
- Molecule 1: Nucleoprotein



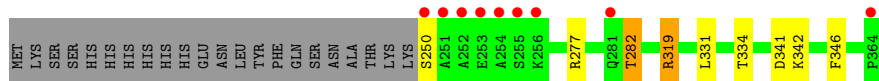
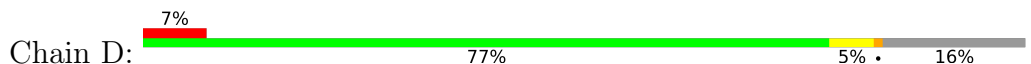
- Molecule 1: Nucleoprotein



- Molecule 1: Nucleoprotein



- Molecule 1: Nucleoprotein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	71.15Å 43.55Å 74.23Å 90.00° 94.87° 90.00°	Depositor
Resolution (Å)	73.97 – 1.45 73.97 – 1.45	Depositor EDS
% Data completeness (in resolution range)	98.7 (73.97-1.45) 98.7 (73.97-1.45)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.22 (at 1.45Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, $R_{free}$	0.167 , 0.180 0.167 , 0.180	Depositor DCC
$R_{free}$ test set	3920 reflections (4.89%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	21.3	Xtriage
Anisotropy	0.329	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 45.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.021 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	7768	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.87% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: SO4

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	0.47	0/935	0.63	0/1262
1	B	0.50	0/935	0.63	0/1262
1	C	0.53	0/946	0.66	1/1277 (0.1%)
1	D	0.45	0/941	0.65	1/1270 (0.1%)
All	All	0.49	0/3757	0.64	2/5071 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	319	ARG	NE-CZ-NH2	7.10	123.85	120.30
1	C	288	ASP	CB-CG-OD1	5.24	123.01	118.30

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	912	892	891	2	0
1	B	912	892	891	5	0
1	C	923	899	898	2	0
1	D	918	897	896	6	0
2	A	10	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	5	0	0	0	0
2	C	10	0	0	0	0
2	D	5	0	0	0	0
3	A	113	0	0	1	1
3	B	140	0	0	2	0
3	C	152	0	0	1	1
3	D	88	0	0	2	0
All	All	4188	3580	3576	15	1

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 2.

All (15) 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:C:322:MET:HE2	3:D:501:HOH:O	2.00	0.61
1:B:281:GLN:O	3:B:501:HOH:O	2.17	0.56
1:D:342:LYS:NZ	3:D:505:HOH:O	2.40	0.54
3:C:502:HOH:O	1:D:334[A]:THR:HG23	2.10	0.52
1:C:360:TYR:HA	1:C:363:PHE:CZ	2.46	0.51
1:A:253:GLU:HA	1:A:256:LYS:HD2	1.93	0.50
1:B:319:ARG:HB2	1:B:334[A]:THR:HG22	1.94	0.50
1:D:319:ARG:HB2	1:D:334[A]:THR:HG22	1.94	0.50
1:B:360:TYR:HA	1:B:363:PHE:CZ	2.48	0.49
1:D:331:LEU:C	1:D:331:LEU:HD23	2.33	0.48
1:D:282:THR:O	1:D:282:THR:CG2	2.63	0.46
3:A:570:HOH:O	1:B:334[A]:THR:HG23	2.17	0.45
1:D:341:ASP:HA	1:D:346:PHE:CD1	2.55	0.41
1:A:276:ARG:HG2	1:A:277:ARG:O	2.21	0.41
1:B:276:ARG:NH2	3:B:511:HOH:O	2.54	0.41

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:586:HOH:O	3:C:525:HOH:O[1_655]	2.04	0.16

## 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	113/137 (82%)	112 (99%)	1 (1%)	0	100	100
1	B	113/137 (82%)	112 (99%)	1 (1%)	0	100	100
1	C	115/137 (84%)	114 (99%)	1 (1%)	0	100	100
1	D	114/137 (83%)	113 (99%)	1 (1%)	0	100	100
All	All	455/548 (83%)	451 (99%)	4 (1%)	0	100	100

There are no Ramachandran outliers to report.

### 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	94/115 (82%)	92 (98%)	2 (2%)	56	20
1	B	94/115 (82%)	93 (99%)	1 (1%)	76	47
1	C	95/115 (83%)	94 (99%)	1 (1%)	76	47
1	D	95/115 (83%)	92 (97%)	3 (3%)	42	9
All	All	378/460 (82%)	371 (98%)	7 (2%)	60	24

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	277	ARG
1	A	341	ASP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	277	ARG
1	C	277	ARG
1	D	250	SER
1	D	277	ARG
1	D	282	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

6 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	$\# Z  > 2$
2	SO4	B	401	-	4,4,4	0.13	0	6,6,6	0.05	0
2	SO4	A	401	-	4,4,4	0.15	0	6,6,6	0.05	0
2	SO4	D	401	-	4,4,4	0.15	0	6,6,6	0.05	0
2	SO4	C	402	-	4,4,4	0.14	0	6,6,6	0.06	0
2	SO4	A	402	-	4,4,4	0.15	0	6,6,6	0.05	0
2	SO4	C	401	-	4,4,4	0.15	0	6,6,6	0.06	0

There are no bond length outliers.

There are no bond angle outliers.

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 [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<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	114/137 (83%)	-0.19	5 (4%) 34 36	17, 28, 55, 81	0
1	B	114/137 (83%)	-0.28	5 (4%) 34 36	17, 25, 50, 73	0
1	C	116/137 (84%)	-0.33	4 (3%) 45 46	16, 25, 43, 65	0
1	D	115/137 (83%)	0.29	9 (7%) 13 15	17, 38, 73, 89	0
All	All	459/548 (83%)	-0.13	23 (5%) 29 30	16, 28, 60, 89	0

All (23) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	254	ALA	7.0
1	D	252	ALA	5.9
1	D	251	ALA	5.6
1	A	251	ALA	5.0
1	A	252	ALA	5.0
1	D	364	PRO	4.7
1	D	250	SER	4.6
1	D	253	GLU	4.5
1	B	252	ALA	4.3
1	D	281	GLN	4.1
1	A	253	GLU	3.9
1	B	253	GLU	3.4
1	A	254	ALA	3.2
1	B	251	ALA	3.2
1	C	250	SER	3.0
1	B	254	ALA	2.9
1	C	342	LYS	2.7
1	A	364	PRO	2.4
1	D	256	LYS	2.4
1	D	255	SER	2.4
1	B	282	THR	2.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	C	252	ALA	2.2
1	C	249	LYS	2.1

## 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 carbohydrates 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
2	SO4	C	402	5/5	0.64	0.17	110,111,116,119	0
2	SO4	D	401	5/5	0.74	0.14	111,115,116,119	0
2	SO4	B	401	5/5	0.87	0.16	105,110,113,117	0
2	SO4	A	401	5/5	0.87	0.17	100,106,110,110	0
2	SO4	A	402	5/5	0.89	0.09	122,123,130,134	0
2	SO4	C	401	5/5	0.92	0.12	77,79,86,86	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.