



Full wwPDB X-ray Structure Validation Report ⓘ

Jan 23, 2025 – 12:08 PM JST

PDB ID : 8X1H
Title : Crystal structure of N-terminal domain of Nucleocapsid protein of SARS-CoV-2
Authors : Kumari, S.; Gupta, G.D.
Deposited on : 2023-11-07
Resolution : 2.00 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

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)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

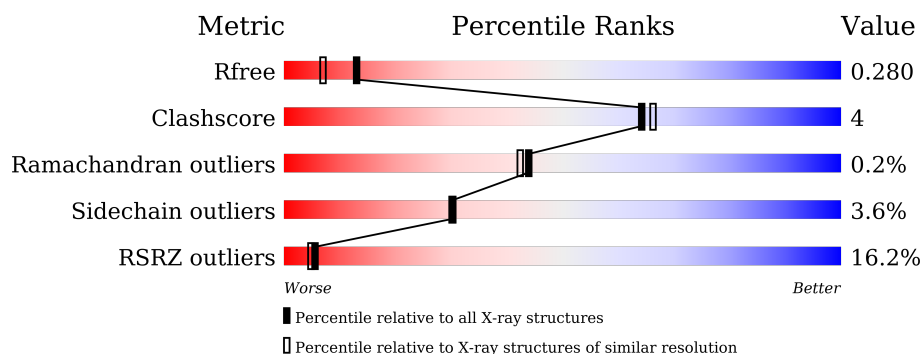
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.00 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}	164625	9409 (2.00-2.00)
Clashscore	180529	10737 (2.00-2.00)
Ramachandran outliers	177936	10628 (2.00-2.00)
Sidechain outliers	177891	10627 (2.00-2.00)
RSRZ outliers	164620	9409 (2.00-2.00)

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 $\leq 5\%$. 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	134	<div> <div>4%</div> <div> <div></div> <div>86%</div> <div>10%</div> <div></div> </div> </div>
1	B	134	<div> <div>8%</div> <div> <div></div> <div>87%</div> <div>10%</div> <div></div> </div> </div>
1	C	134	<div> <div>10%</div> <div> <div></div> <div>85%</div> <div>9%</div> <div>5%</div> </div> </div>
1	D	134	<div> <div>39%</div> <div> <div></div> <div>76%</div> <div>14%</div> <div>9%</div> </div> </div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4134 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 Nucleoprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	128	Total	C	N	O	S	0	0	0
			993	627	179	186	1			
1	B	130	Total	C	N	O	S	0	0	0
			1004	634	181	188	1			
1	C	127	Total	C	N	O	S	0	0	0
			986	625	178	182	1			
1	D	122	Total	C	N	O	S	0	0	0
			929	593	162	173	1			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	42	SER	-	expression tag	UNP P0DTC9
A	43	MET	-	expression tag	UNP P0DTC9
B	42	SER	-	expression tag	UNP P0DTC9
B	43	MET	-	expression tag	UNP P0DTC9
C	42	SER	-	expression tag	UNP P0DTC9
C	43	MET	-	expression tag	UNP P0DTC9
D	42	SER	-	expression tag	UNP P0DTC9
D	43	MET	-	expression tag	UNP P0DTC9

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



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

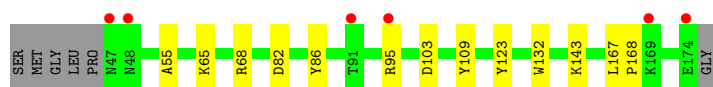
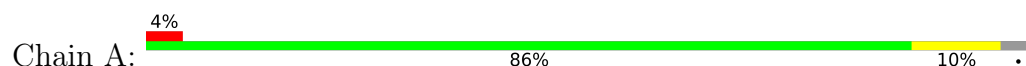
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	70	Total	O	0	0
			70	70		
3	B	50	Total	O	0	0
			50	50		
3	C	48	Total	O	0	0
			48	48		
3	D	42	Total	O	0	0
			42	42		

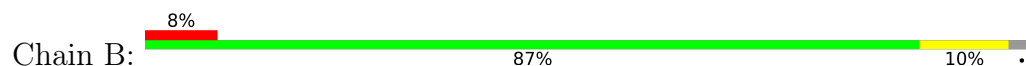
3 Residue-property plots [i](#)

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.

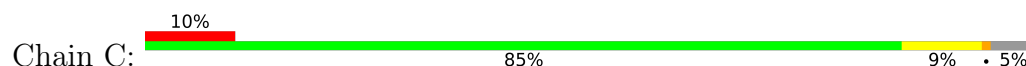
• Molecule 1: Nucleoprotein



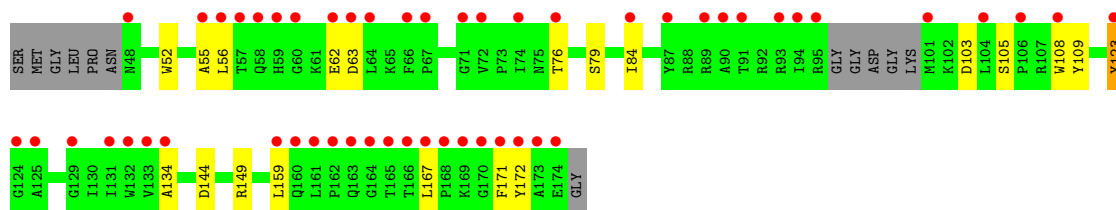
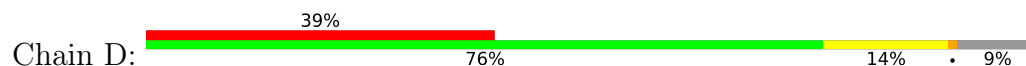
• Molecule 1: Nucleoprotein



• Molecule 1: Nucleoprotein



• Molecule 1: Nucleoprotein



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	58.95Å 92.34Å 96.01Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.24 – 2.00 30.24 – 2.00	Depositor EDS
% Data completeness (in resolution range)	100.0 (30.24-2.00) 99.9 (30.24-2.00)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.49 (at 2.00Å)	Xtriage
Refinement program	PHENIX (1.19.1_4122: ???)	Depositor
R, R_{free}	0.237 , 0.278 0.238 , 0.280	Depositor DCC
R_{free} test set	1838 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å ²)	33.9	Xtriage
Anisotropy	0.221	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 55.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.015 for -h,l,k	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	4134	wwPDB-VP
Average B, all atoms (Å ²)	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.15% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: GOL

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	0.28	0/1021	0.53	0/1390
1	B	0.26	0/1033	0.51	0/1406
1	C	0.27	0/1014	0.51	0/1379
1	D	0.27	0/956	0.55	0/1307
All	All	0.27	0/4024	0.52	0/5482

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	993	0	954	8	0
1	B	1004	0	965	7	0
1	C	986	0	957	5	0
1	D	929	0	874	9	0
2	A	6	0	8	1	0
2	C	6	0	8	0	0
3	A	70	0	0	0	0
3	B	50	0	0	1	0
3	C	48	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	D	42	0	0	0	0
All	All	4134	0	3766	29	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 4.

All (29) 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:D:62:GLU:HG2	1:D:63:ASP:H	1.59	0.67
1:A:68:ARG:H	1:A:68:ARG:HD2	1.71	0.56
1:A:95:ARG:HG2	1:A:95:ARG:HH11	1.70	0.56
1:D:167:LEU:HD13	1:D:172:TYR:HA	1.87	0.55
1:A:55:ALA:HB2	1:A:109:TYR:CE2	2.44	0.52
1:B:82:ASP:HA	1:B:143:LYS:HE3	1.91	0.52
1:C:65:LYS:HD2	1:C:66:PHE:H	1.74	0.51
1:A:82:ASP:HA	1:A:143:LYS:HE3	1.92	0.51
1:A:167:LEU:HD12	1:A:168:PRO:HD2	1.92	0.51
1:D:84:ILE:HD11	1:D:123:TYR:CD1	2.46	0.51
1:B:100:LYS:HD2	1:B:101:MET:H	1.76	0.50
1:B:87:TYR:HB2	1:B:131:ILE:HG13	1.94	0.50
1:D:84:ILE:HD11	1:D:123:TYR:HD1	1.76	0.49
1:C:80:PRO:HA	1:C:83:GLN:HG2	1.93	0.49
1:A:109:TYR:CD2	2:A:201:GOL:H31	2.49	0.47
1:C:95:ARG:NH1	1:C:99:GLY:O	2.46	0.47
1:C:55:ALA:HB2	1:C:109:TYR:CE2	2.50	0.46
1:A:86:TYR:HB3	1:A:132:TRP:CE3	2.51	0.46
1:B:55:ALA:HB2	1:B:109:TYR:CE1	2.51	0.45
1:D:55:ALA:HB2	1:D:109:TYR:CE2	2.52	0.45
1:B:118:GLU:HG3	1:B:130:ILE:HD11	1.98	0.45
1:D:52:TRP:CD2	1:D:149:ARG:HB2	2.53	0.43
1:A:123:TYR:CE1	1:A:132:TRP:HB3	2.54	0.43
1:B:174:GLU:HG2	3:B:211:HOH:O	2.19	0.42
1:D:144:ASP:OD1	1:D:144:ASP:N	2.50	0.42
1:D:123:TYR:HE1	1:D:134:ALA:HB2	1.85	0.41
1:B:47:ASN:OD1	1:B:47:ASN:N	2.51	0.41
1:D:108:TRP:CD1	1:D:171:PHE:HE2	2.39	0.41
1:C:86:TYR:HB3	1:C:132:TRP:CE3	2.56	0.41

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	126/134 (94%)	123 (98%)	3 (2%)	0	100	100
1	B	128/134 (96%)	121 (94%)	6 (5%)	1 (1%)	16	12
1	C	125/134 (93%)	121 (97%)	4 (3%)	0	100	100
1	D	118/134 (88%)	113 (96%)	5 (4%)	0	100	100
All	All	497/536 (93%)	478 (96%)	18 (4%)	1 (0%)	44	42

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	99	GLY

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	100/105 (95%)	98 (98%)	2 (2%)	50	55
1	B	101/105 (96%)	100 (99%)	1 (1%)	73	78
1	C	99/105 (94%)	95 (96%)	4 (4%)	27	26
1	D	91/105 (87%)	84 (92%)	7 (8%)	10	7
All	All	391/420 (93%)	377 (96%)	14 (4%)	30	30

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	65	LYS
1	A	103	ASP
1	B	163	GLN
1	C	61	LYS
1	C	62	GLU
1	C	65	LYS
1	C	93	ARG
1	D	56	LEU
1	D	76	THR
1	D	79	SER
1	D	103	ASP
1	D	105	SER
1	D	123	TYR
1	D	159	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	163	GLN
1	D	75	ASN
1	D	77	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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

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	$\# Z > 2$
2	GOL	A	201	-	5,5,5	0.83	0	5,5,5	0.99	0
2	GOL	C	201	-	5,5,5	0.88	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
2	GOL	A	201	-	-	4/4/4/4	-
2	GOL	C	201	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (6) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	201	GOL	O1-C1-C2-C3
2	C	201	GOL	O1-C1-C2-C3
2	A	201	GOL	O1-C1-C2-O2
2	A	201	GOL	C1-C2-C3-O3
2	A	201	GOL	O2-C2-C3-O3
2	C	201	GOL	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	201	GOL	1	0

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, 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(Å ²)	Q<0.9
1	A	128/134 (95%)	0.22	6 (4%) 37 35	23, 40, 70, 88	0
1	B	130/134 (97%)	0.64	11 (8%) 18 17	26, 47, 79, 116	0
1	C	127/134 (94%)	0.72	13 (10%) 13 12	28, 48, 82, 127	0
1	D	122/134 (91%)	1.82	52 (42%) 1 1	36, 63, 118, 128	0
All	All	507/536 (94%)	0.84	82 (16%) 5 5	23, 48, 89, 128	0

All (82) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	167	LEU	5.1
1	D	106	PRO	4.6
1	D	165	THR	4.5
1	C	79	SER	4.4
1	D	173	ALA	4.4
1	D	64	LEU	4.3
1	D	66	PHE	4.2
1	D	166	THR	4.1
1	D	162	PRO	4.0
1	D	94	ILE	4.0
1	D	56	LEU	3.8
1	D	108	TRP	3.6
1	D	57	THR	3.5
1	D	90	ALA	3.5
1	D	58	GLN	3.4
1	D	170	GLY	3.4
1	C	49	THR	3.3
1	D	161	LEU	3.2
1	D	91	THR	3.2
1	D	101	MET	3.2
1	D	74	ILE	3.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	91	THR	3.1
1	D	123	TYR	3.1
1	D	160	GLN	3.1
1	D	164	GLY	3.0
1	C	135	THR	3.0
1	C	140	ASN	3.0
1	B	175	GLY	3.0
1	D	59	HIS	2.9
1	C	139	LEU	2.9
1	C	80	PRO	2.8
1	D	63	ASP	2.8
1	B	97	GLY	2.7
1	D	169	LYS	2.7
1	D	168	PRO	2.7
1	A	95	ARG	2.7
1	D	174	GLU	2.7
1	D	60	GLY	2.7
1	D	132	TRP	2.6
1	D	71	GLY	2.6
1	D	48	ASN	2.6
1	D	172	TYR	2.5
1	B	131	ILE	2.5
1	A	48	ASN	2.5
1	D	62	GLU	2.4
1	B	64	LEU	2.4
1	D	84	ILE	2.4
1	A	47	ASN	2.4
1	C	173	ALA	2.4
1	D	72	VAL	2.4
1	D	133	VAL	2.4
1	A	174	GLU	2.3
1	C	141	THR	2.3
1	D	134	ALA	2.3
1	B	98	ASP	2.3
1	C	137	GLY	2.3
1	D	129	GLY	2.3
1	D	89	ARG	2.3
1	B	124	GLY	2.2
1	C	69	GLY	2.2
1	B	95	ARG	2.2
1	D	87	TYR	2.2
1	D	104	LEU	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	D	55	ALA	2.2
1	D	124	GLY	2.2
1	D	93	ARG	2.2
1	C	142	PRO	2.2
1	B	96	GLY	2.2
1	D	171	PHE	2.2
1	C	99	GLY	2.2
1	B	123	TYR	2.1
1	D	125	ALA	2.1
1	A	169	LYS	2.1
1	C	161	LEU	2.1
1	D	159	LEU	2.1
1	D	76	THR	2.1
1	D	95	ARG	2.1
1	D	131	ILE	2.1
1	D	163	GLN	2.1
1	B	46	PRO	2.0
1	D	67	PRO	2.0
1	B	129	GLY	2.0

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(\AA^2)	Q<0.9
2	GOL	A	201	6/6	0.79	0.13	32,44,53,56	0
2	GOL	C	201	6/6	0.82	0.15	42,46,49,55	0

6.5 Other polymers [i](#)

There are no such residues in this entry.