Full wwPDB X-ray Structure Validation Report (i)

Jan 20, 2024 – 09:45 pm GMT

PDB ID	:	7OAU
Title	:	Nanobody C5 bound to Kent variant RBD (N501Y)
Authors	:	Naismith, J.H.; Mikolajek, H.
Deposited on	:	2021-04-20
Resolution	:	1.65 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 (i) 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 (1)) were used in the production of this report:

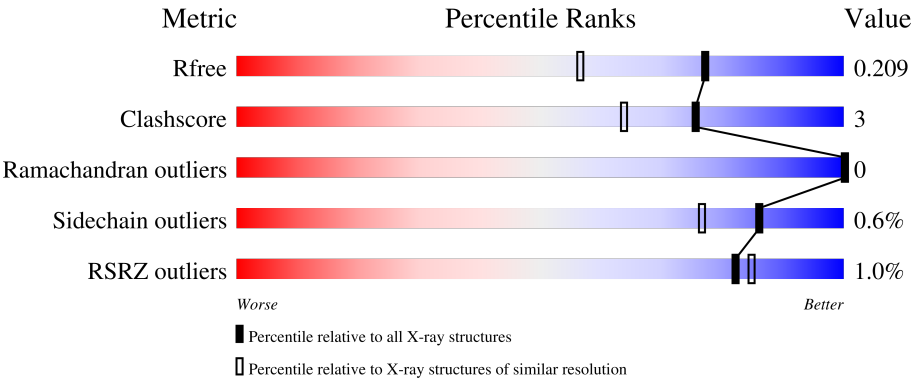
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.65 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	1827 (1.66-1.66)
Clashscore	141614	1931 (1.66-1.66)
Ramachandran outliers	138981	1891 (1.66-1.66)
Sidechain outliers	138945	1891 (1.66-1.66)
RSRZ outliers	127900	1791 (1.66-1.66)

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 >=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	AAA	210	86% 6% 8%
1	EEE	210	88% . 8%
2	BBB	129	% 87% 8% 5%
2	FFF	129	3% 82% 12% 6%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 5494 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	AAA	194	Total	C	N	O	S	0	2	0
			1547	995	255	289	8			
1	EEE	194	Total	C	N	O	S	0	3	0
			1554	999	256	291	8			

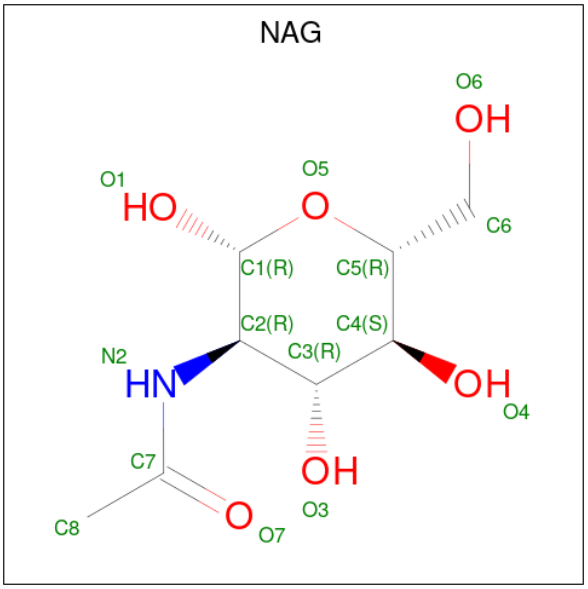
There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	501	TYR	ASN	engineered mutation	UNP P0DTC2
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	501	TYR	ASN	engineered mutation	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 C5.

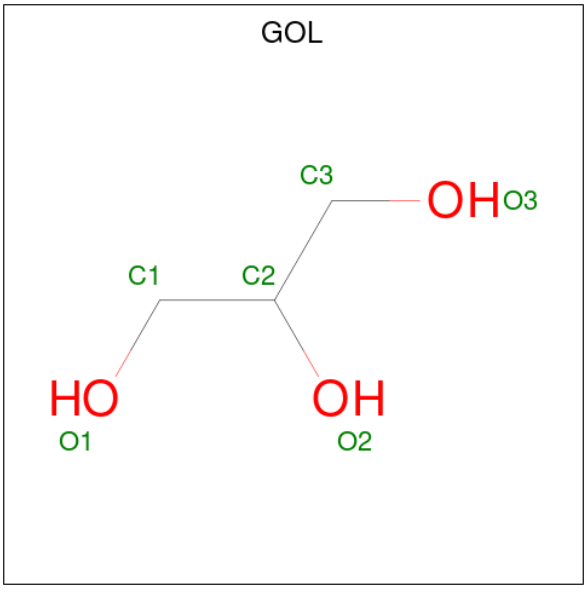
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	BBB	122	Total	C	N	O	S	0	7	0
			948	594	163	183	8			
2	FFF	121	Total	C	N	O	S	0	7	0
			941	592	160	181	8			

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



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

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	BBB	1	Total C O 6 3 3	0	0

- Molecule 5 is water.

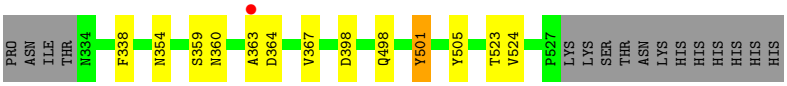
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	AAA	149	Total O 149 149	0	0
5	BBB	93	Total O 93 93	0	0
5	EEE	161	Total O 161 161	0	0
5	FFF	67	Total O 67 67	0	0

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.

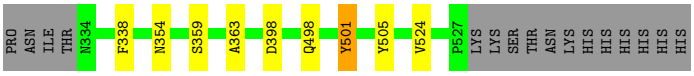
- Molecule 1: Spike protein S1

Chain AAA:



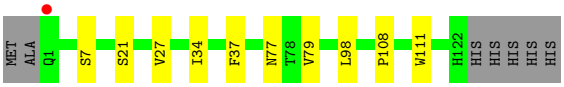
- Molecule 1: Spike protein S1

Chain EEE:



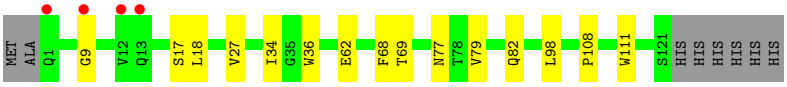
- Molecule 2: C5

Chain BBB:



- Molecule 2: C5

Chain FFF:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	28.81Å 153.73Å 75.85Å 90.00° 100.33° 90.00°	Depositor
Resolution (Å)	38.46 – 1.65 38.43 – 1.65	Depositor EDS
% Data completeness (in resolution range)	100.0 (38.46-1.65) 100.0 (38.43-1.65)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.60 (at 1.65Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.175 , 0.199 0.185 , 0.209	Depositor DCC
R_{free} test set	3776 reflections (4.86%)	wwPDB-VP
Wilson B-factor (Å^2)	24.3	Xtriage
Anisotropy	0.163	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 31.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.087 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	5494	wwPDB-VP
Average B, all atoms (Å^2)	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.40% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, 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	0.62	0/1598	0.74	1/2177 (0.0%)
1	EEE	0.64	0/1605	0.77	1/2187 (0.0%)
2	BBB	0.64	0/988	0.73	0/1339
2	FFF	0.66	0/977	0.76	0/1326
All	All	0.64	0/5168	0.75	2/7029 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	EEE	501	TYR	CB-CG-CD1	7.41	125.44	121.00
1	AAA	501	TYR	CB-CG-CD1	5.47	124.28	121.00

There are no chirality outliers.

There are no planarity outliers.

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	1547	0	1467	8	0
1	EEE	1554	0	1473	6	0
2	BBB	948	0	933	6	0
2	FFF	941	0	932	11	0
3	AAA	14	0	13	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	EEE	14	0	13	0	0
4	BBB	6	0	8	0	0
5	AAA	149	0	0	0	0
5	BBB	93	0	0	0	0
5	EEE	161	0	0	0	0
5	FFF	67	0	0	1	0
All	All	5494	0	4839	31	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.

All (31) 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:BBB:7[A]:SER:OG	2:BBB:21:SER:OG	2.07	0.72
1:EEE:359:SER:HA	1:EEE:524:VAL:HG22	1.75	0.68
1:AAA:359:SER:HA	1:AAA:524:VAL:HG22	1.77	0.67
1:AAA:364:ASP:O	1:AAA:367:VAL:HG12	1.99	0.63
2:FFF:9:GLY:HA2	2:FFF:18:LEU:HD13	1.81	0.61
2:FFF:34:ILE:CG2	2:FFF:79[A]:VAL:HG11	2.34	0.58
2:BBB:34:ILE:CG2	2:BBB:79:VAL:HG11	2.34	0.57
1:EEE:359:SER:HA	1:EEE:524:VAL:CG2	2.34	0.56
2:FFF:36[B]:TRP:CH2	2:FFF:79[B]:VAL:HG13	2.40	0.56
1:AAA:359:SER:HA	1:AAA:524:VAL:CG2	2.36	0.55
2:FFF:108:PRO:HG2	2:FFF:111:TRP:CZ2	2.42	0.55
1:AAA:360:ASN:HA	1:AAA:523:THR:OG1	2.07	0.54
2:BBB:34:ILE:HG21	2:BBB:79:VAL:HG11	1.91	0.53
1:EEE:501:TYR:HB3	1:EEE:505:TYR:HB2	1.90	0.52
2:FFF:34:ILE:HG21	2:FFF:79[A]:VAL:HG11	1.92	0.51
2:FFF:36[A]:TRP:HZ3	5:FFF:206:HOH:O	1.95	0.50
2:BBB:108:PRO:HG2	2:BBB:111:TRP:CZ2	2.48	0.49
2:FFF:69[A]:THR:OG1	2:FFF:82:GLN:HB3	2.13	0.48
2:FFF:27[A]:VAL:HG11	2:FFF:98:LEU:HD21	1.96	0.47
1:EEE:354:ASN:O	1:EEE:398:ASP:HA	2.17	0.45
1:AAA:338:PHE:HE2	1:AAA:363:ALA:HB1	1.82	0.44
2:FFF:34:ILE:HG23	2:FFF:79[A]:VAL:HG11	1.99	0.43
1:AAA:354:ASN:O	1:AAA:398:ASP:HA	2.19	0.43
1:EEE:338:PHE:HE2	1:EEE:363:ALA:HB1	1.83	0.42
2:BBB:27:VAL:HG11	2:BBB:98:LEU:HD21	2.01	0.42
2:BBB:37:PHE:CE1	2:BBB:108:PRO:HG3	2.56	0.41
1:EEE:498:GLN:HB2	1:EEE:501:TYR:CD1	2.56	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:AAA:498:GLN:HB2	1:AAA:501:TYR:CD1	2.56	0.40
2:FFF:68:PHE:HA	2:FFF:82:GLN:O	2.21	0.40
1:AAA:501:TYR:HB3	1:AAA:505:TYR:HB2	2.01	0.40

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	AAA	194/210 (92%)	188 (97%)	6 (3%)	0	100	100
1	EEE	195/210 (93%)	189 (97%)	6 (3%)	0	100	100
2	BBB	127/129 (98%)	127 (100%)	0	0	100	100
2	FFF	126/129 (98%)	126 (100%)	0	0	100	100
All	All	642/678 (95%)	630 (98%)	12 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains (i)

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	169/183 (92%)	169 (100%)	0	100	100
1	EEE	170/183 (93%)	170 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	BBB	106/105 (101%)	105 (99%)	1 (1%)	78 66
2	FFF	105/105 (100%)	103 (98%)	2 (2%)	57 34
All	All	550/576 (96%)	547 (100%)	3 (0%)	86 81

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	BBB	77	ASN
2	FFF	62	GLU
2	FFF	77	ASN

Sometimes 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 monosaccharides in this entry.

5.6 Ligand geometry (i)

3 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	$\#\|Z\|>2$
3	NAG	AAA	601	1	14,14,15	0.38	0	17,19,21	1.03	1 (5%)
3	NAG	EEE	601	1	14,14,15	0.32	0	17,19,21	1.04	1 (5%)
4	GOL	BBB	201	-	5,5,5	0.10	0	5,5,5	0.27	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
3	NAG	AAA	601	1	-	0/6/23/26	0/1/1/1
3	NAG	EEE	601	1	-	0/6/23/26	0/1/1/1
4	GOL	BBB	201	-	-	3/4/4/4	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	AAA	601	NAG	C4-C3-C2	3.02	115.45	111.02
3	EEE	601	NAG	O5-C5-C6	2.11	110.52	107.20

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	BBB	201	GOL	O2-C2-C3-O3
4	BBB	201	GOL	C1-C2-C3-O3
4	BBB	201	GOL	O1-C1-C2-O2

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^{th} 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	AAA	194/210 (92%)	-0.31	1 (0%) 91 92	18, 31, 50, 62	0
1	EEE	194/210 (92%)	-0.44	0 100 100	17, 27, 42, 56	0
2	BBB	122/129 (94%)	-0.42	1 (0%) 86 88	18, 27, 40, 55	0
2	FFF	121/129 (93%)	-0.18	4 (3%) 46 47	18, 31, 57, 72	0
All	All	631/678 (93%)	-0.35	6 (0%) 82 85	17, 29, 49, 72	0

All (6) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	FFF	1	GLN	2.9
2	FFF	9	GLY	2.5
2	BBB	1	GLN	2.5
2	FFF	12	VAL	2.3
2	FFF	13	GLN	2.2
1	AAA	363	ALA	2.2

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

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^{th} 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
3	NAG	AAA	601	14/15	0.86	0.18	43,54,61,61	0
3	NAG	EEE	601	14/15	0.93	0.10	36,41,51,52	0
4	GOL	BBB	201	6/6	0.95	0.08	36,41,46,52	0

6.5 Other polymers (i)

There are no such residues in this entry.