wwPDB X-ray Structure Validation Summary Report (i)

Jan 13, 2024 – 11:12 pm GMT

PDB ID	:	6Y5Z
Title	:	Structure of apo Human Polyomavirus 12 VP1
Authors	:	Stroh, L.J.; Rustmeier, N.H.; Stehle, T.
Deposited on	:	2020-02-26
Resolution	:	1.55 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*
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:

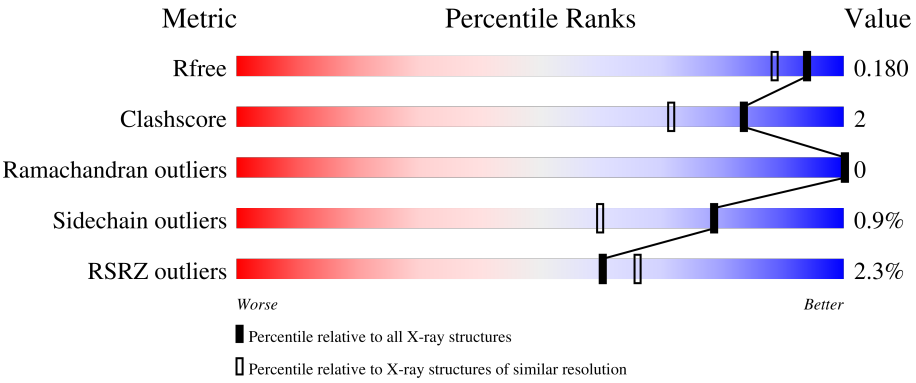
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 1.55 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	2556 (1.56-1.52)
Clashscore	141614	2634 (1.56-1.52)
Ramachandran outliers	138981	2580 (1.56-1.52)
Sidechain outliers	138945	2577 (1.56-1.52)
RSRZ outliers	127900	2524 (1.56-1.52)

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	277	2% 91% • 5%
1	BBB	277	5% 92% • 5%
1	CCC	277	% 91% • 5%
1	DDD	277	2% 92% 6% •
1	EEE	277	% 90% 5% •

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 11949 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 Capsid protein VP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	264	Total	C	N	O	S	0	9	0
			2078	1323	348	393	14			
1	BBB	262	Total	C	N	O	S	0	4	0
			2028	1295	336	383	14			
1	CCC	262	Total	C	N	O	S	0	9	0
			2052	1314	338	386	14			
1	DDD	272	Total	C	N	O	S	0	6	0
			2109	1342	355	396	16			
1	EEE	265	Total	C	N	O	S	0	7	0
			2051	1311	340	386	14			

There are 20 discrepancies between the modelled and reference sequences:

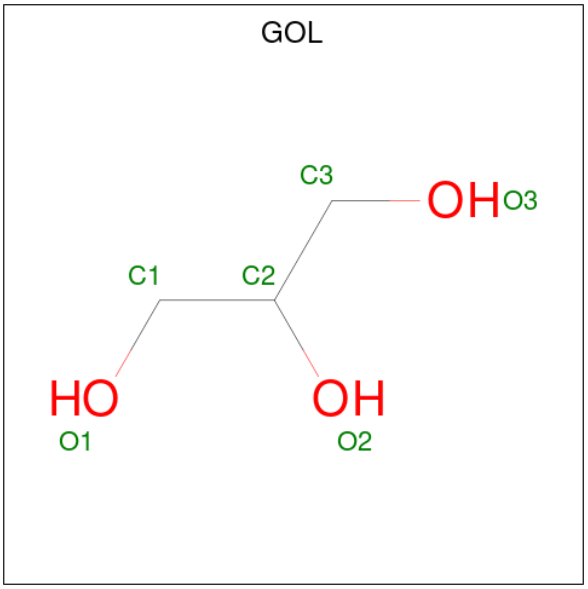
Chain	Residue	Modelled	Actual	Comment	Reference
AAA	23	GLY	-	expression tag	UNP M4T5D3
AAA	24	SER	-	expression tag	UNP M4T5D3
AAA	25	HIS	-	expression tag	UNP M4T5D3
AAA	26	MET	-	expression tag	UNP M4T5D3
BBB	23	GLY	-	expression tag	UNP M4T5D3
BBB	24	SER	-	expression tag	UNP M4T5D3
BBB	25	HIS	-	expression tag	UNP M4T5D3
BBB	26	MET	-	expression tag	UNP M4T5D3
CCC	23	GLY	-	expression tag	UNP M4T5D3
CCC	24	SER	-	expression tag	UNP M4T5D3
CCC	25	HIS	-	expression tag	UNP M4T5D3
CCC	26	MET	-	expression tag	UNP M4T5D3
DDD	23	GLY	-	expression tag	UNP M4T5D3
DDD	24	SER	-	expression tag	UNP M4T5D3
DDD	25	HIS	-	expression tag	UNP M4T5D3
DDD	26	MET	-	expression tag	UNP M4T5D3
EEE	23	GLY	-	expression tag	UNP M4T5D3
EEE	24	SER	-	expression tag	UNP M4T5D3
EEE	25	HIS	-	expression tag	UNP M4T5D3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
EEE	26	MET	-	expression tag	UNP M4T5D3

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	AAA	1	Total	C	O	0	0
			6	3	3		
2	AAA	1	Total	C	O	0	0
			6	3	3		
2	AAA	1	Total	C	O	0	0
			6	3	3		
2	BBB	1	Total	C	O	0	0
			6	3	3		
2	BBB	1	Total	C	O	0	0
			6	3	3		
2	BBB	1	Total	C	O	0	0
			6	3	3		
2	BBB	1	Total	C	O	0	0
			6	3	3		
2	CCC	1	Total	C	O	0	0
			6	3	3		
2	CCC	1	Total	C	O	0	0
			6	3	3		
2	CCC	1	Total	C	O	0	0
			6	3	3		
2	DDD	1	Total	C	O	0	0
			6	3	3		

Continued on next page...

Continued from previous page...

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

- Molecule 3 is water.

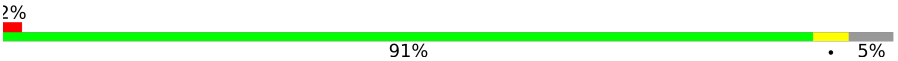
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	AAA	327	Total O 330 330	0	3
3	BBB	273	Total O 274 274	0	1
3	CCC	294	Total O 294 294	0	0
3	DDD	310	Total O 311 311	0	1
3	EEE	335	Total O 338 338	0	3

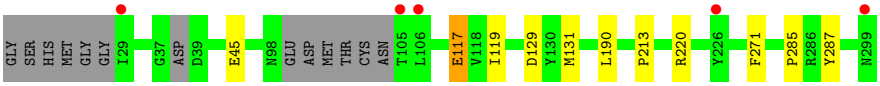
3 Residue-property plots (i)

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: Capsid protein VP1

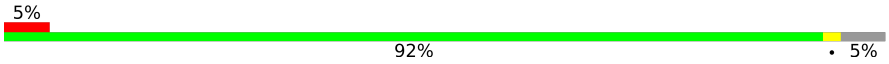
Chain AAA:

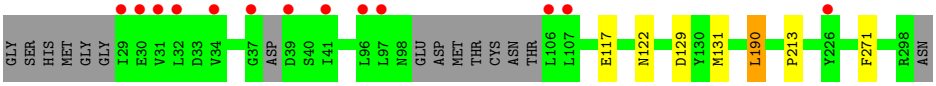




• Molecule 1: Capsid protein VP1

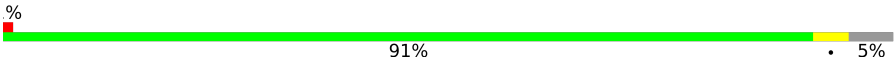
Chain BBB:

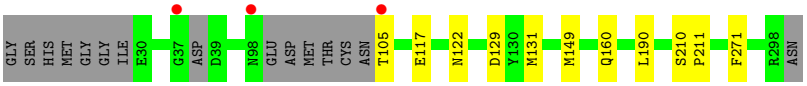




• Molecule 1: Capsid protein VP1

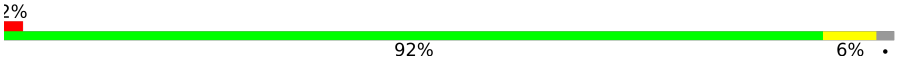
Chain CCC:

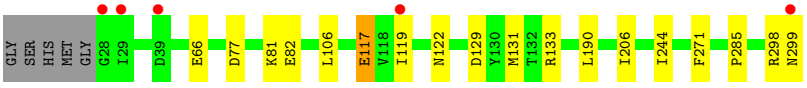




• Molecule 1: Capsid protein VP1

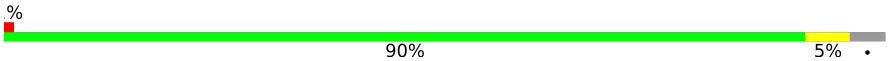
Chain DDD:

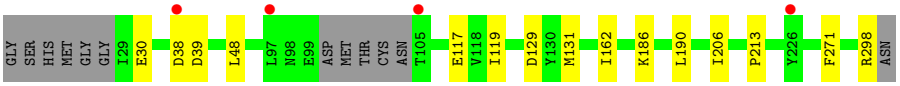




• Molecule 1: Capsid protein VP1

Chain EEE:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	61.22Å 136.42Å 85.38Å 90.00° 109.56° 90.00°	Depositor
Resolution (Å)	44.05 – 1.55 44.05 – 1.55	Depositor EDS
% Data completeness (in resolution range)	95.1 (44.05-1.55) 95.1 (44.05-1.55)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.08 (at 1.55Å)	Xtriage
Refinement program	REFMAC 5.8.0253	Depositor
R, R_{free}	0.143 , 0.169 0.154 , 0.180	Depositor DCC
R_{free} test set	3634 reflections (2.00%)	wwPDB-VP
Wilson B-factor (Å^2)	16.4	Xtriage
Anisotropy	0.133	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 46.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.018 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	11949	wwPDB-VP
Average B, all atoms (Å^2)	19.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.58% 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

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	0.63	0/2147	0.81	1/2924 (0.0%)
1	BBB	0.64	0/2088	0.82	0/2848
1	CCC	0.63	0/2127	0.82	0/2899
1	DDD	0.64	0/2176	0.82	1/2966 (0.0%)
1	EEE	0.64	0/2121	0.80	0/2896
All	All	0.64	0/10659	0.81	2/14533 (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	AAA	220	ARG	NE-CZ-NH2	-5.62	117.49	120.30
1	DDD	133	ARG	N-CA-CB	-5.24	101.17	110.60

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	2078	0	2031	12	0
1	BBB	2028	0	1965	7	0
1	CCC	2052	0	2003	8	0
1	DDD	2109	0	2065	20	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	EEE	2051	0	1988	14	0
2	AAA	18	0	24	0	0
2	BBB	24	0	32	0	0
2	CCC	18	0	24	0	0
2	DDD	18	0	24	0	0
2	EEE	6	0	8	0	0
3	AAA	330	0	0	3	0
3	BBB	274	0	0	0	0
3	CCC	294	0	0	0	0
3	DDD	311	0	0	6	0
3	EEE	338	0	0	4	0
All	All	11949	0	10164	49	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 2.

The worst 5 of 49 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:AAA:119[A]:ILE:HD11	1:BBB:213:PRO:HG3	1.58	0.86
1:AAA:213:PRO:HG3	1:EEE:119[A]:ILE:HD11	1.58	0.84
1:DDD:119[A]:ILE:HD11	1:EEE:213:PRO:HG3	1.60	0.83
1:DDD:244:ILE:HD11	3:DDD:709:HOH:O	1.90	0.71
1:DDD:117:GLU:HG2	1:DDD:119[A]:ILE:CD1	2.29	0.63

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	267/277 (96%)	258 (97%)	9 (3%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	BBB	260/277 (94%)	250 (96%)	10 (4%)	0	100	100
1	CCC	265/277 (96%)	256 (97%)	9 (3%)	0	100	100
1	DDD	276/277 (100%)	268 (97%)	8 (3%)	0	100	100
1	EEE	268/277 (97%)	258 (96%)	10 (4%)	0	100	100
All	All	1336/1385 (96%)	1290 (97%)	46 (3%)	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	228/240 (95%)	226 (99%)	2 (1%)	78	60
1	BBB	219/240 (91%)	217 (99%)	2 (1%)	78	60
1	CCC	222/240 (92%)	220 (99%)	2 (1%)	78	60
1	DDD	229/240 (95%)	227 (99%)	2 (1%)	78	60
1	EEE	220/240 (92%)	218 (99%)	2 (1%)	78	60
All	All	1118/1200 (93%)	1108 (99%)	10 (1%)	78	60

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	DDD	190	LEU
1	EEE	117	GLU
1	EEE	190	LEU
1	BBB	190	LEU
1	CCC	105	THR

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)

14 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	CCC	301	-	5,5,5	0.08	0	5,5,5	0.28	0
2	GOL	DDD	302	-	5,5,5	0.07	0	5,5,5	0.20	0
2	GOL	BBB	301	-	5,5,5	0.13	0	5,5,5	0.36	0
2	GOL	BBB	302	-	5,5,5	0.12	0	5,5,5	0.38	0
2	GOL	AAA	303	-	5,5,5	0.12	0	5,5,5	0.29	0
2	GOL	AAA	302	-	5,5,5	0.13	0	5,5,5	0.28	0
2	GOL	AAA	301	-	5,5,5	0.05	0	5,5,5	0.12	0
2	GOL	CCC	302	-	5,5,5	0.11	0	5,5,5	0.24	0
2	GOL	DDD	303	-	5,5,5	0.09	0	5,5,5	0.35	0
2	GOL	EEE	301	-	5,5,5	0.12	0	5,5,5	0.25	0
2	GOL	BBB	303	-	5,5,5	0.10	0	5,5,5	0.36	0
2	GOL	DDD	301	-	5,5,5	0.11	0	5,5,5	0.33	0
2	GOL	CCC	303	-	5,5,5	0.17	0	5,5,5	0.36	0
2	GOL	BBB	304	-	5,5,5	0.08	0	5,5,5	0.26	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	CCC	301	-	-	0/4/4/4	-
2	GOL	DDD	302	-	-	1/4/4/4	-
2	GOL	BBB	301	-	-	0/4/4/4	-
2	GOL	BBB	302	-	-	2/4/4/4	-
2	GOL	AAA	303	-	-	4/4/4/4	-
2	GOL	AAA	302	-	-	2/4/4/4	-
2	GOL	AAA	301	-	-	0/4/4/4	-
2	GOL	CCC	302	-	-	1/4/4/4	-
2	GOL	DDD	303	-	-	0/4/4/4	-
2	GOL	EEE	301	-	-	0/4/4/4	-
2	GOL	BBB	303	-	-	4/4/4/4	-
2	GOL	DDD	301	-	-	0/4/4/4	-
2	GOL	CCC	303	-	-	4/4/4/4	-
2	GOL	BBB	304	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 20 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	AAA	302	GOL	O1-C1-C2-C3
2	AAA	303	GOL	O1-C1-C2-O2
2	AAA	303	GOL	O1-C1-C2-C3
2	BBB	302	GOL	C1-C2-C3-O3
2	BBB	303	GOL	C1-C2-C3-O3

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	264/277 (95%)	-0.32	5 (1%) 66 72	11, 15, 26, 40	0
1	BBB	262/277 (94%)	-0.09	13 (4%) 28 32	11, 16, 32, 52	0
1	CCC	262/277 (94%)	-0.34	3 (1%) 80 83	12, 16, 26, 52	0
1	DDD	272/277 (98%)	-0.29	5 (1%) 68 74	13, 17, 28, 48	0
1	EEE	265/277 (95%)	-0.32	4 (1%) 73 78	11, 15, 27, 53	0
All	All	1325/1385 (95%)	-0.27	30 (2%) 60 66	11, 16, 28, 53	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	BBB	106	LEU	4.3
1	EEE	105	THR	4.3
1	BBB	41	ILE	4.1
1	CCC	105	THR	3.6
1	BBB	34	VAL	3.3

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

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
2	GOL	BBB	302	6/6	0.77	0.16	33,40,42,43	0
2	GOL	AAA	303	6/6	0.78	0.18	45,47,48,53	0
2	GOL	CCC	303	6/6	0.81	0.17	41,46,48,49	0
2	GOL	EEE	301	6/6	0.84	0.16	38,38,40,42	0
2	GOL	DDD	301	6/6	0.85	0.25	37,40,42,46	0
2	GOL	BBB	303	6/6	0.86	0.24	24,40,46,48	0
2	GOL	BBB	301	6/6	0.87	0.25	43,44,44,45	0
2	GOL	CCC	302	6/6	0.88	0.18	41,46,47,47	0
2	GOL	AAA	301	6/6	0.89	0.12	23,26,28,30	0
2	GOL	BBB	304	6/6	0.89	0.18	29,33,38,43	0
2	GOL	CCC	301	6/6	0.90	0.17	31,35,38,40	0
2	GOL	DDD	303	6/6	0.91	0.15	34,38,39,40	0
2	GOL	AAA	302	6/6	0.92	0.20	26,40,46,48	0
2	GOL	DDD	302	6/6	0.96	0.23	36,39,40,40	0

6.5 Other polymers (i)

There are no such residues in this entry.