wwPDB X-ray Structure Validation Summary Report (i)

Dec 13, 2022 – 12:07 pm GMT

PDB ID	:	7ZIP
Title	:	JC Polyomavirus VP1 in complex with 3'-Sialyllactose glycomacromolecules (aliphatic linker)
Authors	:	Freytag, J.; Mueller, J.C.; Stehle, T.
Deposited on	:	2022-04-08
Resolution	:	1.90 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 (i)) were used in the production of this report:

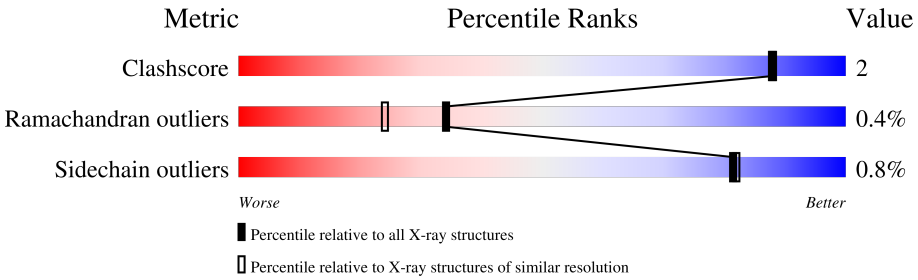
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	FAILED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.3

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 1.90 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))
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)

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%

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	AAA	275	87% 7% 6%
1	BBB	275	89% . 7%
1	CCC	275	83% 7% 10%
1	DDD	275	93% . .
1	EEE	275	86% 8% 5%
2	DaD	3	33% 67%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 10755 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 Major capsid protein VP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	258	Total	C	N	O	S	0	7	0
			1996	1258	339	386	13			
1	BBB	255	Total	C	N	O	S	0	3	0
			1951	1228	333	377	13			
1	CCC	247	Total	C	N	O	S	0	3	0
			1895	1194	328	360	13			
1	DDD	267	Total	C	N	O	S	0	1	0
			2037	1281	352	391	13			
1	EEE	260	Total	C	N	O	S	0	7	0
			2028	1277	349	389	13			

There are 35 discrepancies between the modelled and reference sequences:

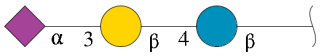
Chain	Residue	Modelled	Actual	Comment	Reference
AAA	26	GLY	-	expression tag	UNP P03088
AAA	27	SER	-	expression tag	UNP P03088
AAA	28	HIS	-	expression tag	UNP P03088
AAA	29	MET	-	expression tag	UNP P03088
AAA	157	ASP	GLU	variant	UNP P03088
AAA	170	THR	SER	variant	UNP P03088
AAA	218	THR	ALA	variant	UNP P03088
BBB	26	GLY	-	expression tag	UNP P03088
BBB	27	SER	-	expression tag	UNP P03088
BBB	28	HIS	-	expression tag	UNP P03088
BBB	29	MET	-	expression tag	UNP P03088
BBB	157	ASP	GLU	variant	UNP P03088
BBB	170	THR	SER	variant	UNP P03088
BBB	218	THR	ALA	variant	UNP P03088
CCC	26	GLY	-	expression tag	UNP P03088
CCC	27	SER	-	expression tag	UNP P03088
CCC	28	HIS	-	expression tag	UNP P03088
CCC	29	MET	-	expression tag	UNP P03088
CCC	157	ASP	GLU	variant	UNP P03088

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
CCC	170	THR	SER	variant	UNP P03088
CCC	218	THR	ALA	variant	UNP P03088
DDD	26	GLY	-	expression tag	UNP P03088
DDD	27	SER	-	expression tag	UNP P03088
DDD	28	HIS	-	expression tag	UNP P03088
DDD	29	MET	-	expression tag	UNP P03088
DDD	157	ASP	GLU	variant	UNP P03088
DDD	170	THR	SER	variant	UNP P03088
DDD	218	THR	ALA	variant	UNP P03088
EEE	26	GLY	-	expression tag	UNP P03088
EEE	27	SER	-	expression tag	UNP P03088
EEE	28	HIS	-	expression tag	UNP P03088
EEE	29	MET	-	expression tag	UNP P03088
EEE	157	ASP	GLU	variant	UNP P03088
EEE	170	THR	SER	variant	UNP P03088
EEE	218	THR	ALA	variant	UNP P03088

- Molecule 2 is an oligosaccharide called N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galacto pyranose-(1-4)-beta-D-glucopyranose.



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	DaD	3	Total	C	N	O	0	0	1
			32	17	1	14			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	AAA	159	Total	O	0	0
			159	159		
3	BBB	159	Total	O	0	0
			159	159		
3	CCC	143	Total	O	0	0
			143	143		
3	DDD	174	Total	O	0	0
			174	174		
3	EEE	181	Total	O	0	0
			181	181		

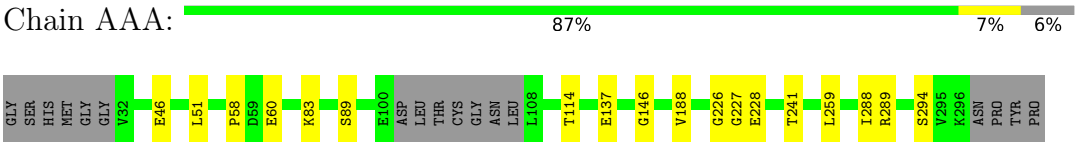
3 Residue-property plots (i)

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. 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. 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.

Note EDS failed to run properly.

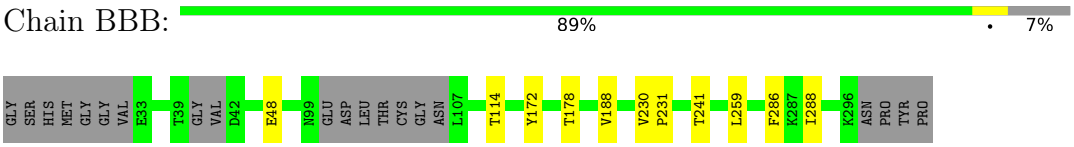
• Molecule 1: Major capsid protein VP1

Chain AAA:



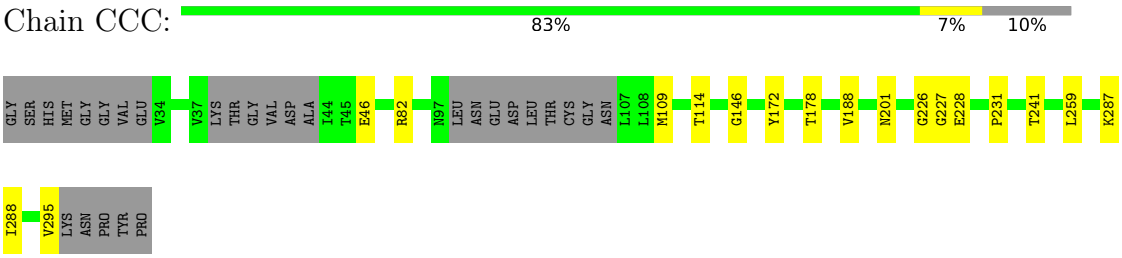
• Molecule 1: Major capsid protein VP1

Chain BBB:



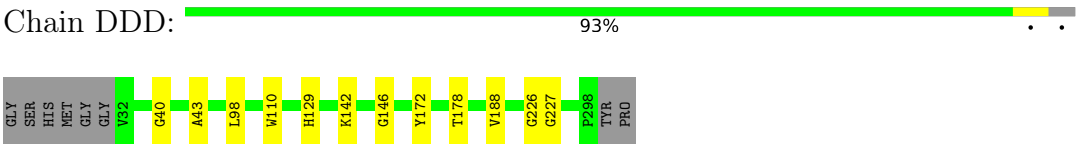
• Molecule 1: Major capsid protein VP1

Chain CCC:



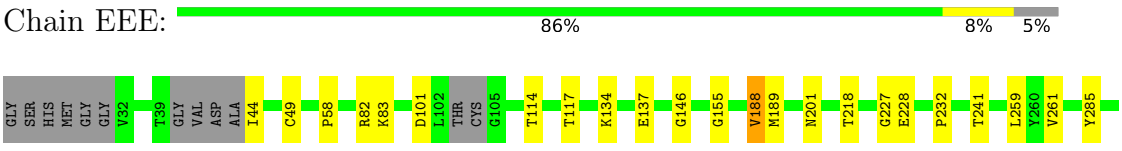
• Molecule 1: Major capsid protein VP1

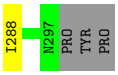
Chain DDD:



• Molecule 1: Major capsid protein VP1

Chain EEE:





• Molecule 2: N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose-(1-4)-beta-D-glucopyranose

Chain DaD:

4 Data and refinement statistics (i)

EDS failed to run properly - this section is therefore incomplete.

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	144.73Å 152.50Å 62.95Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.96 – 1.90	Depositor
% Data completeness (in resolution range)	99.9 (47.96-1.90)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.14 (at 1.90Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.172 , 0.215	Depositor
Wilson B-factor (Å^2)	26.0	Xtriage
Anisotropy	0.041	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.009 for k,h,-l	Xtriage
Total number of atoms	10755	wwPDB-VP
Average B, all atoms (Å^2)	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.32% 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: BGC, GAL, SIA

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	0.69	0/2062	0.86	0/2808
1	BBB	0.69	0/2001	0.87	0/2726
1	CCC	0.69	0/1948	0.85	0/2650
1	DDD	0.70	0/2087	0.85	0/2845
1	EEE	0.72	0/2093	0.85	0/2847
All	All	0.70	0/10191	0.85	0/13876

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

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	1996	0	1928	9	0
1	BBB	1951	0	1852	5	0
1	CCC	1895	0	1817	10	0
1	DDD	2037	0	1968	7	0
1	EEE	2028	0	1970	12	0
2	DaD	32	0	26	0	0
3	AAA	159	0	0	1	0
3	BBB	159	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	CCC	143	0	0	1	0
3	DDD	174	0	0	0	0
3	EEE	181	0	0	2	0
All	All	10755	0	9561	39	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 39 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:EEE:137:GLU:HG2	3:EEE:560:HOH:O	1.59	0.99
1:CCC:287:LYS:NZ	3:CCC:401:HOH:O	2.27	0.67
1:EEE:82[B]:ARG:NH1	1:EEE:201:ASN:O	2.29	0.65
1:CCC:82[B]:ARG:NH2	1:CCC:201:ASN:OD1	2.32	0.62
1:EEE:58:PRO:HA	1:EEE:83:LYS:HD3	1.85	0.59

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	261/275 (95%)	248 (95%)	12 (5%)	1 (0%)	34	24
1	BBB	252/275 (92%)	240 (95%)	11 (4%)	1 (0%)	34	24
1	CCC	244/275 (89%)	235 (96%)	8 (3%)	1 (0%)	34	24
1	DDD	266/275 (97%)	253 (95%)	12 (4%)	1 (0%)	34	24
1	EEE	261/275 (95%)	252 (97%)	8 (3%)	1 (0%)	34	24
All	All	1284/1375 (93%)	1228 (96%)	51 (4%)	5 (0%)	34	24

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	AAA	188	VAL
1	BBB	188	VAL
1	DDD	188	VAL
1	EEE	188	VAL
1	CCC	188	VAL

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	220/236 (93%)	217 (99%)	3 (1%)	67 65
1	BBB	210/236 (89%)	209 (100%)	1 (0%)	88 89
1	CCC	205/236 (87%)	204 (100%)	1 (0%)	88 89
1	DDD	223/236 (94%)	223 (100%)	0	100 100
1	EEE	225/236 (95%)	222 (99%)	3 (1%)	69 68
All	All	1083/1180 (92%)	1075 (99%)	8 (1%)	81 84

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	EEE	134	LYS
1	EEE	101	ASP
1	CCC	228	GLU
1	BBB	230	VAL
1	EEE	44	ILE

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)

Of 3 monosaccharides modelled in this entry, 2 were used for Mogul analysis.

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	GAL	DaD	2	2	11,11,12	0.90	1 (9%)	15,15,17	2.05	8 (53%)
2	SIA	DaD	3	2	20,20,21	1.38	3 (15%)	24,28,31	1.53	6 (25%)

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	GAL	DaD	2	2	-	0/2/19/22	0/1/1/1
2	SIA	DaD	3	2	-	1/18/34/38	0/1/1/1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	DaD	3	SIA	C2-C1	3.81	1.55	1.52
2	DaD	3	SIA	O1A-C1	2.34	1.29	1.22
2	DaD	2	GAL	C4-C5	2.14	1.57	1.53
2	DaD	3	SIA	C3-C2	2.08	1.55	1.52

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	DaD	2	GAL	C1-O5-C5	3.89	117.46	112.19
2	DaD	2	GAL	O3-C3-C2	3.04	115.82	109.99
2	DaD	3	SIA	O6-C2-C3	-2.89	106.48	110.46

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	DaD	2	GAL	C1-C2-C3	-2.76	106.27	109.67
2	DaD	3	SIA	C6-O6-C2	2.72	117.15	111.34

There are no chirality outliers.

All (1) torsion outliers are listed below:

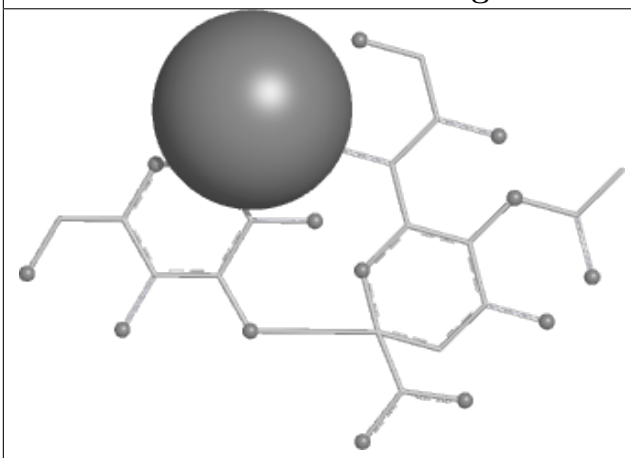
Mol	Chain	Res	Type	Atoms
2	DaD	3	SIA	O1A-C1-C2-O6

There are no ring outliers.

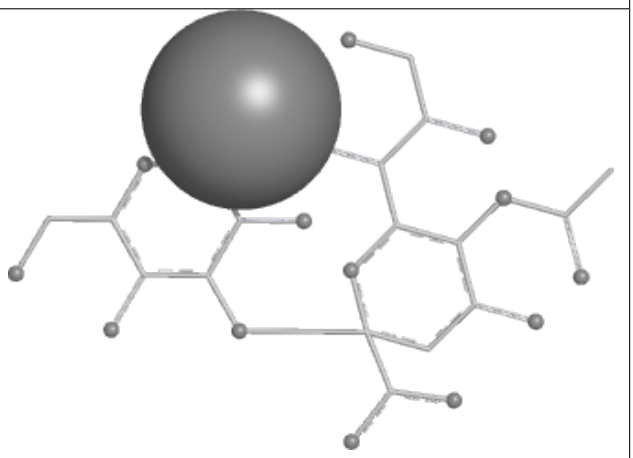
No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.

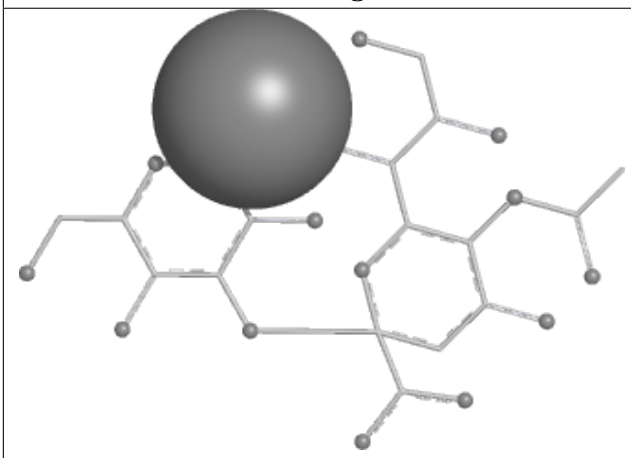
Oligosaccharide Chain DaD



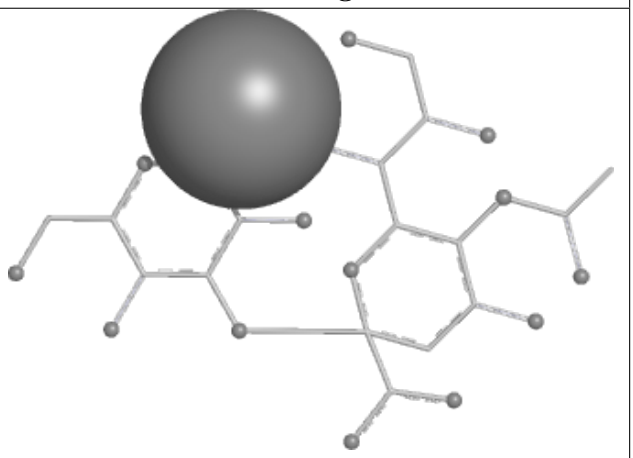
Bond lengths



Bond angles



Torsions



Rings

5.6 Ligand geometry (i)

There are no ligands in this entry.

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

EDS failed to run properly - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

EDS failed to run properly - this section is therefore empty.

6.3 Carbohydrates

EDS failed to run properly - this section is therefore empty.

6.4 Ligands

EDS failed to run properly - this section is therefore empty.

6.5 Other polymers

EDS failed to run properly - this section is therefore empty.