wwPDB X-ray Structure Validation Summary Report (i)

Sep 27, 2023 – 09:13 PM EDT

PDB ID	:	1OHG
Title	:	STRUCTURE OF THE DSDNA BACTERIOPHAGE HK97 MATURE EMPTY CAPSID
Authors	:	Helgstrand, C.; Wikoff, W.R.; Duda, R.L.; Hendrix, R.W.; Johnson, J.E.; Liljas, L.
Deposited on	:	2003-05-26
Resolution	:	3.45 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:

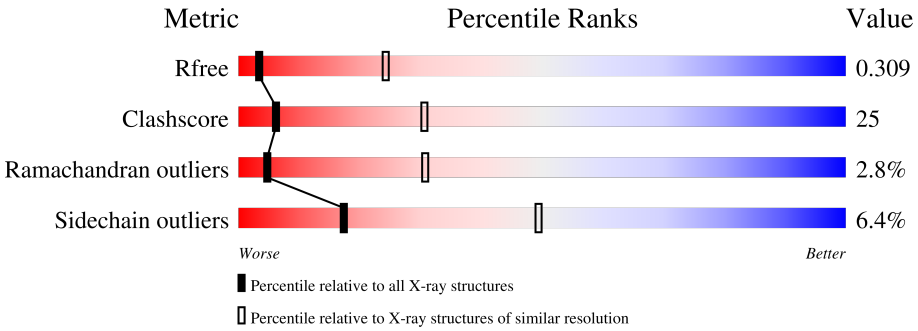
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 3.45 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}	130704	1291 (3.52-3.40)
Clashscore	141614	1372 (3.52-3.40)
Ramachandran outliers	138981	1337 (3.52-3.40)
Sidechain outliers	138945	1338 (3.52-3.40)

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%

Mol	Chain	Length	Quality of chain
1	A	282	60% 35% 5%
1	B	282	59% 37% 5%
1	C	282	57% 37% 6%
1	D	282	54% 41% .
1	E	282	58% 37% 5%
1	F	282	57% 38% 5%
1	G	282	58% 38% .

2 Entry composition (i)

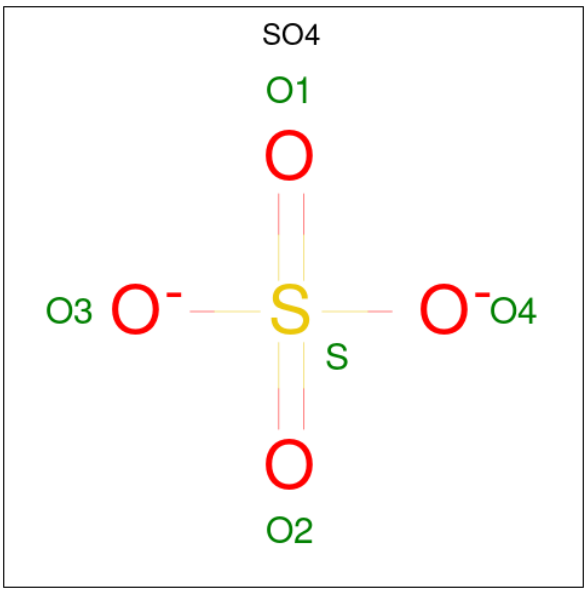
There are 3 unique types of molecules in this entry. The entry contains 15070 atoms, of which 0 are hydrogens and 0 are deuteriums.

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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	281	Total	C	N	O	S	0	0	1
			2152	1344	376	422	10			
1	B	281	Total	C	N	O	S	0	0	1
			2152	1344	376	422	10			
1	C	281	Total	C	N	O	S	0	0	1
			2152	1344	376	422	10			
1	D	281	Total	C	N	O	S	0	0	1
			2152	1344	376	422	10			
1	E	281	Total	C	N	O	S	0	0	1
			2152	1344	376	422	10			
1	F	281	Total	C	N	O	S	0	0	1
			2152	1344	376	422	10			
1	G	281	Total	C	N	O	S	0	0	1
			2152	1344	376	422	10			

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



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

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

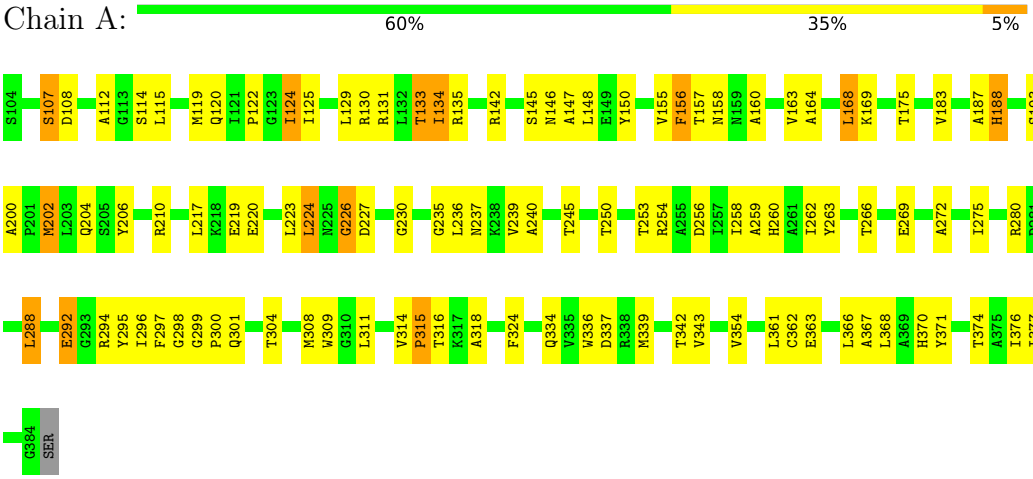
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	G	1	Total Cl 1 1	0	0

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. 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.

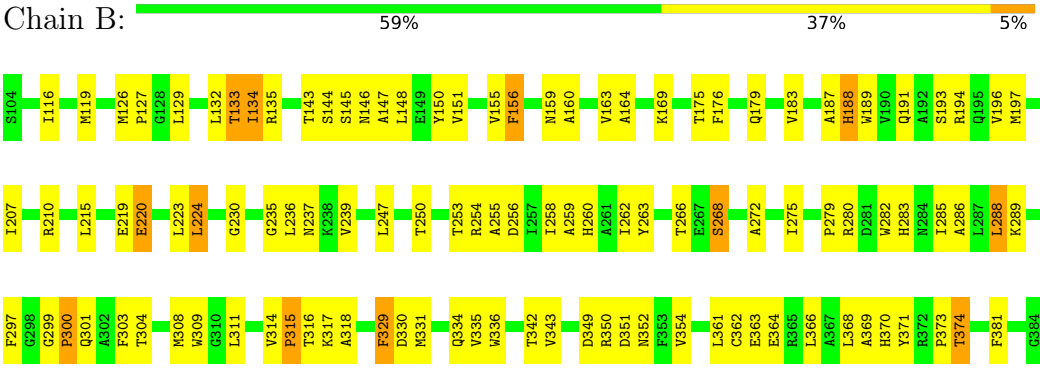
• Molecule 1: MAJOR CAPSID PROTEIN

Chain A:



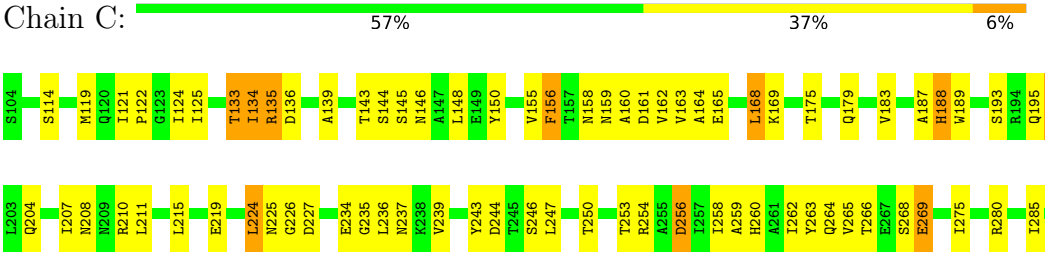
• Molecule 1: MAJOR CAPSID PROTEIN

Chain B:



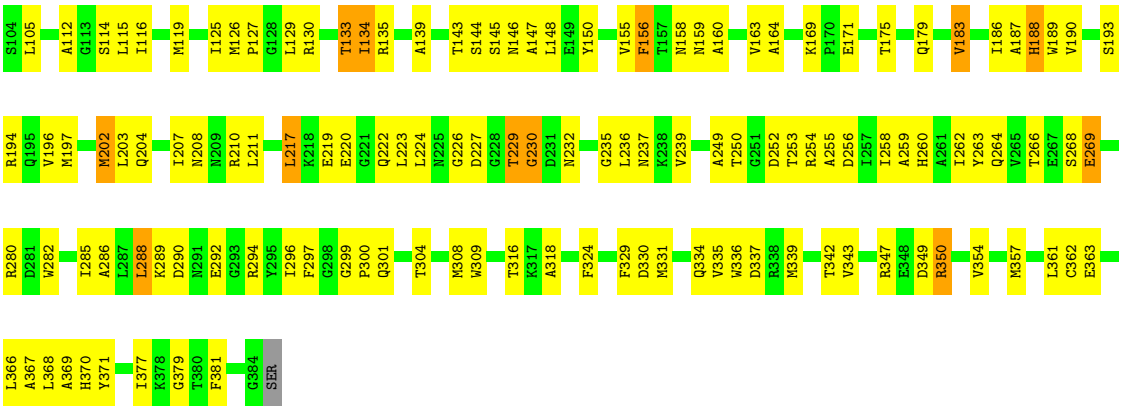
• Molecule 1: MAJOR CAPSID PROTEIN

Chain C:



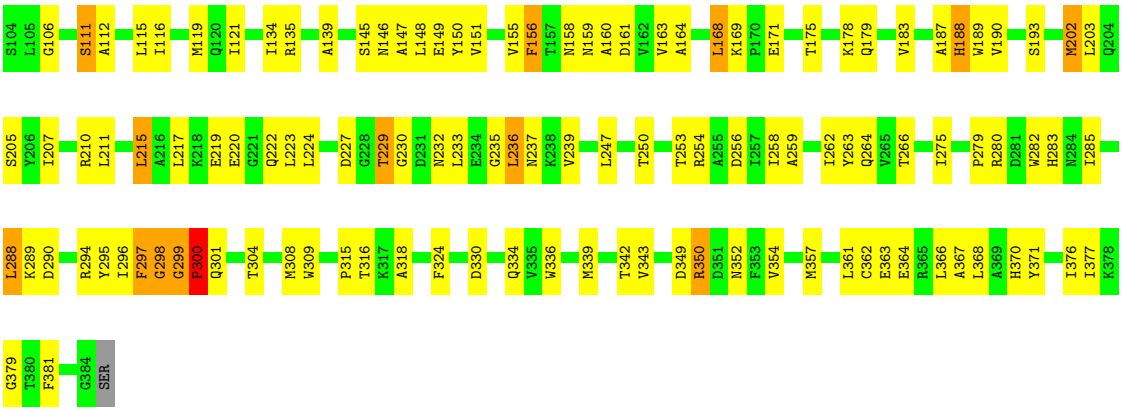
• Molecule 1: MAJOR CAPSID PROTEIN

Chain D:



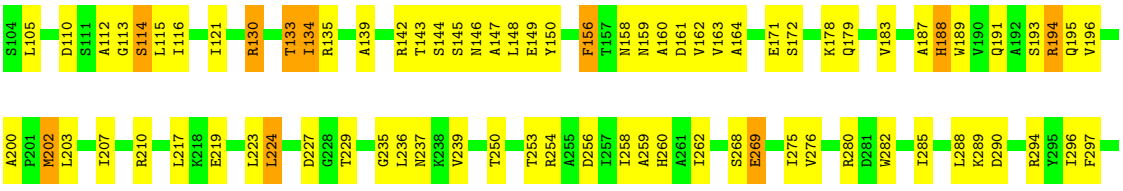
• Molecule 1: MAJOR CAPSID PROTEIN

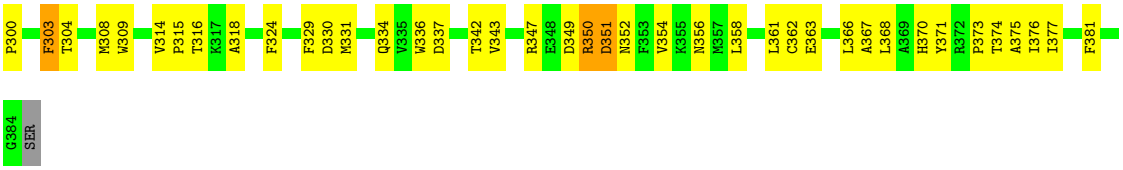
Chain E:



• Molecule 1: MAJOR CAPSID PROTEIN

Chain F:





• Molecule 1: MAJOR CAPSID PROTEIN

Chain G:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	579.70Å 626.65Å 787.20Å 90.00° 89.90° 90.00°	Depositor
Resolution (Å)	190.13 – 3.45 190.13 – 3.45	Depositor EDS
% Data completeness (in resolution range)	65.2 (190.13-3.45) 65.3 (190.13-3.45)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.90 (at 3.41Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.373 , 0.374 0.307 , 0.309	Depositor DCC
R_{free} test set	4792 reflections (0.10%)	wwPDB-VP
Wilson B-factor (Å^2)	97.5	Xtriage
Anisotropy	0.326	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.23 , 55.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.003 for h,-k,-l	Xtriage
F_o,F_c correlation	0.87	EDS
Total number of atoms	15070	wwPDB-VP
Average B, all atoms (Å^2)	75.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.49% 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: SO4, CL

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	0.44	0/2189	0.68	1/2971 (0.0%)
1	B	0.44	0/2189	0.67	0/2971
1	C	0.45	0/2189	0.68	0/2971
1	D	0.48	0/2189	0.67	0/2971
1	E	0.46	0/2189	0.69	2/2971 (0.1%)
1	F	0.44	0/2189	0.68	0/2971
1	G	0.46	0/2189	0.69	1/2971 (0.0%)
All	All	0.45	0/15323	0.68	4/20797 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	298	GLY	N-CA-C	5.27	126.28	113.10
1	E	298	GLY	N-CA-C	5.18	126.05	113.10
1	E	299	GLY	C-N-CD	-5.09	109.40	120.60
1	G	132	LEU	CA-CB-CG	-5.01	103.77	115.30

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	A	2152	0	2119	107	0
1	B	2152	0	2119	119	0
1	C	2152	0	2119	121	0
1	D	2152	0	2119	138	0
1	E	2152	0	2119	118	0
1	F	2152	0	2119	109	0
1	G	2152	0	2119	99	0
2	A	5	0	0	0	0
3	G	1	0	0	0	0
All	All	15070	0	14833	733	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 25.

The worst 5 of 733 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:B:316:THR:HG22	1:B:318:ALA:H	1.11	1.09
1:D:316:THR:HG22	1:D:318:ALA:H	1.13	1.09
1:E:316:THR:HG22	1:E:318:ALA:H	1.13	1.09
1:C:316:THR:HG22	1:C:318:ALA:H	1.17	1.06
1:G:316:THR:HG22	1:G:318:ALA:H	1.21	1.05

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	A	279/282 (99%)	237 (85%)	36 (13%)	6 (2%)	6 35
1	B	279/282 (99%)	229 (82%)	43 (15%)	7 (2%)	5 32
1	C	279/282 (99%)	236 (85%)	34 (12%)	9 (3%)	4 28

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	279/282 (99%)	232 (83%)	39 (14%)	8 (3%)	4	30
1	E	279/282 (99%)	240 (86%)	32 (12%)	7 (2%)	5	32
1	F	279/282 (99%)	231 (83%)	40 (14%)	8 (3%)	4	30
1	G	279/282 (99%)	232 (83%)	38 (14%)	9 (3%)	4	28
All	All	1953/1974 (99%)	1637 (84%)	262 (13%)	54 (3%)	5	31

5 of 54 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	300	PRO
1	B	352	ASN
1	C	134	ILE
1	C	300	PRO
1	D	229	THR

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	A	230/231 (100%)	211 (92%)	19 (8%)	11	38
1	B	230/231 (100%)	216 (94%)	14 (6%)	18	51
1	C	230/231 (100%)	211 (92%)	19 (8%)	11	38
1	D	230/231 (100%)	220 (96%)	10 (4%)	29	61
1	E	230/231 (100%)	214 (93%)	16 (7%)	15	46
1	F	230/231 (100%)	214 (93%)	16 (7%)	15	46
1	G	230/231 (100%)	221 (96%)	9 (4%)	32	64
All	All	1610/1617 (100%)	1507 (94%)	103 (6%)	17	49

5 of 103 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	202	MET

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	E	229	THR
1	G	256	ASP
1	D	217	LEU
1	E	168	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 29 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	291	ASN
1	G	232	ASN
1	E	208	ASN
1	F	237	ASN
1	E	191	GLN

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)

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 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	SO4	A	1384	-	4,4,4	1.89	2 (50%)	6,6,6	0.18	0

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1384	SO4	O2-S	2.39	1.59	1.46
2	A	1384	SO4	O1-S	2.39	1.59	1.46

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

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.