wwPDB X-ray Structure Validation Summary Report (i)

Aug 22, 2022 – 11:09 pm BST

PDB ID	:	7B6C
Title	:	BK Polyomavirus VP1 pentamer fusion with long C-terminal extended arm
Authors	:	Osipov, E.M.; Beelen, S.; Strelkov, S.V.
Deposited on	:	2020-12-07
Resolution	:	2.48 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:

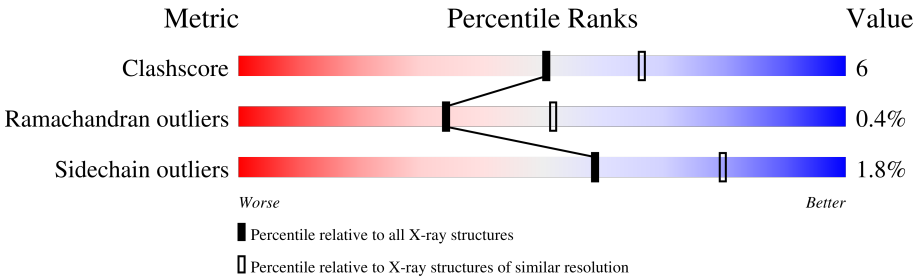
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	FAILED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

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 2.48 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	6594 (2.50-2.46)
Ramachandran outliers	138981	6469 (2.50-2.46)
Sidechain outliers	138945	6471 (2.50-2.46)

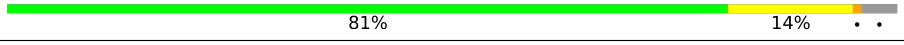
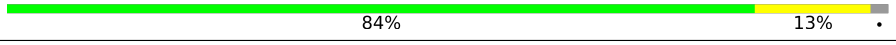
The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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	283	87% 12% .
1	BBB	283	84% 11% .
1	CCC	283	83% 15% ..
1	DDD	283	81% 17% .
1	EEE	283	86% 12% ..
1	FFF	283	87% 11% ..
1	GGG	283	83% 14% ..
1	HHH	283	82% 14% .

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	III	283	
1	JJJ	283	

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 22259 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,Major capsid protein VP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	279	Total	C	N	O	S	0	0	0
			2154	1347	371	421	15			
1	BBB	271	Total	C	N	O	S	0	0	0
			2092	1313	360	405	14			
1	CCC	279	Total	C	N	O	S	0	0	0
			2157	1347	373	422	15			
1	DDD	280	Total	C	N	O	S	0	0	0
			2155	1347	371	422	15			
1	EEE	279	Total	C	N	O	S	0	0	0
			2153	1347	373	418	15			
1	FFF	280	Total	C	N	O	S	0	0	0
			2174	1361	373	425	15			
1	GGG	279	Total	C	N	O	S	0	0	0
			2164	1353	375	421	15			
1	HHH	271	Total	C	N	O	S	0	0	0
			2102	1316	364	408	14			
1	III	272	Total	C	N	O	S	0	0	0
			2108	1321	362	411	14			
1	JJJ	276	Total	C	N	O	S	0	0	0
			2126	1331	367	414	14			

There are 10 discrepancies between the modelled and reference sequences:

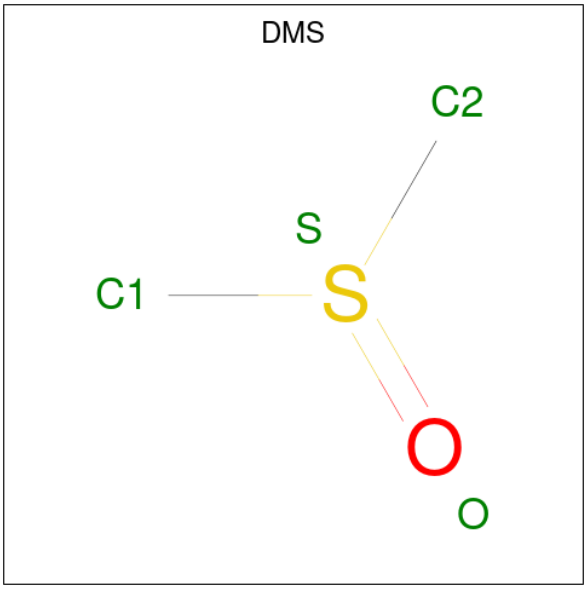
Chain	Residue	Modelled	Actual	Comment	Reference
AAA	17	GLY	-	expression tag	UNP P03088
BBB	17	GLY	-	expression tag	UNP P03088
CCC	17	GLY	-	expression tag	UNP P03088
DDD	17	GLY	-	expression tag	UNP P03088
EEE	17	GLY	-	expression tag	UNP P03088
FFF	17	GLY	-	expression tag	UNP P03088
GGG	17	GLY	-	expression tag	UNP P03088
HHH	17	GLY	-	expression tag	UNP P03088
III	17	GLY	-	expression tag	UNP P03088

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
JJJ	17	GLY	-	expression tag	UNP P03088

- Molecule 2 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula: C_2H_6OS).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	CCC	1	Total C O S 4 2 1 1	0	0
2	HHH	1	Total C O S 4 2 1 1	0	0

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	DDD	1	Total Ca 1 1	0	0
3	FFF	1	Total Ca 1 1	0	0
3	III	1	Total Ca 1 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	AAA	93	Total O 93 93	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	BBB	89	Total O 89 89	0	0
4	CCC	92	Total O 92 92	0	0
4	DDD	88	Total O 88 88	0	0
4	EEE	84	Total O 84 84	0	0
4	FFF	98	Total O 98 98	0	0
4	GGG	80	Total O 80 80	0	0
4	HHH	64	Total O 64 64	0	0
4	III	84	Total O 84 84	0	0
4	JJJ	91	Total O 91 91	0	0

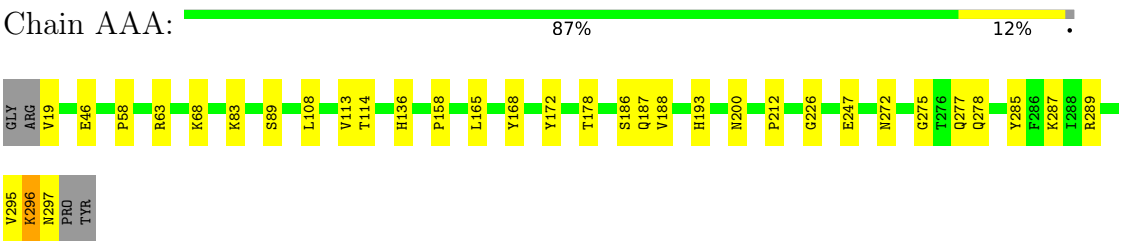
3 Residue-property plots (i)

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.

Note EDS failed to run properly.

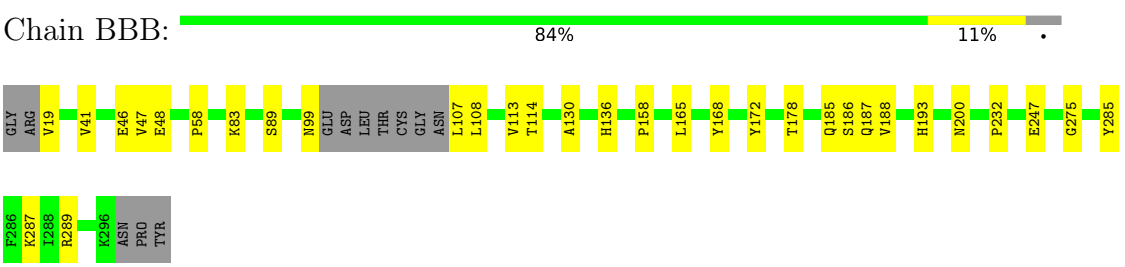
• Molecule 1: Major capsid protein VP1,Major capsid protein VP1

Chain AAA:



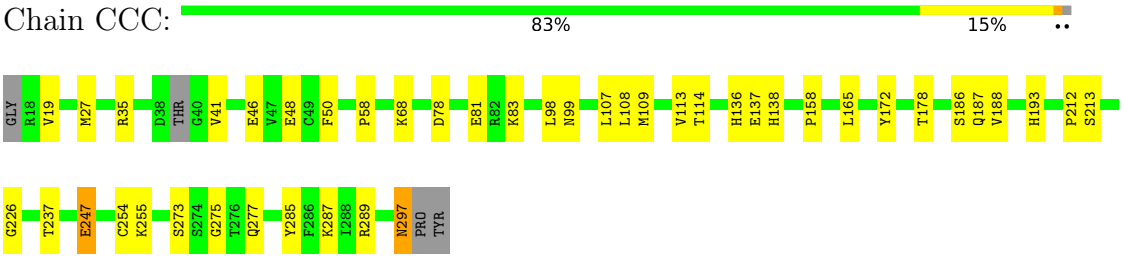
• Molecule 1: Major capsid protein VP1,Major capsid protein VP1

Chain BBB:



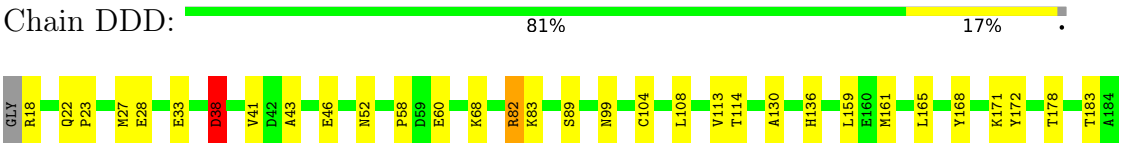
• Molecule 1: Major capsid protein VP1,Major capsid protein VP1

Chain CCC:



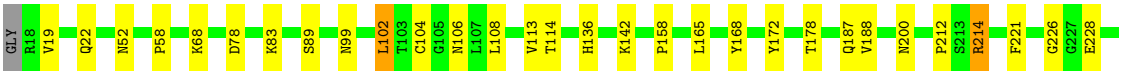
• Molecule 1: Major capsid protein VP1,Major capsid protein VP1

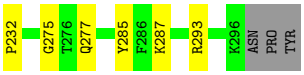
Chain DDD:



• Molecule 1: Major capsid protein VP1,Major capsid protein VP1

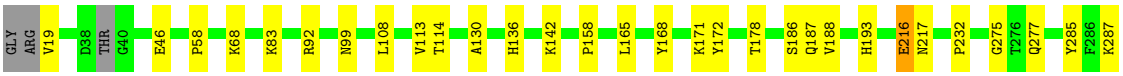
Chain EEE:

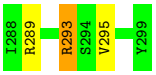




• Molecule 1: Major capsid protein VP1,Major capsid protein VP1

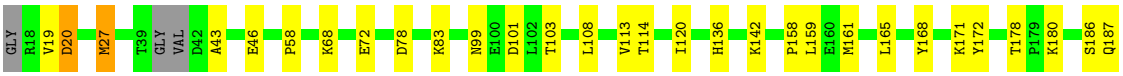
Chain FFF:





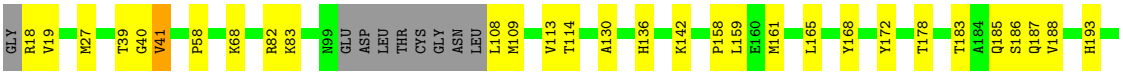
• Molecule 1: Major capsid protein VP1,Major capsid protein VP1

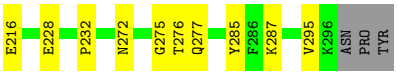
Chain GGG:



• Molecule 1: Major capsid protein VP1,Major capsid protein VP1

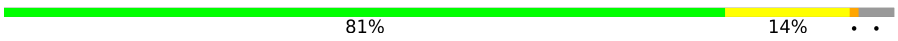
Chain HHH:

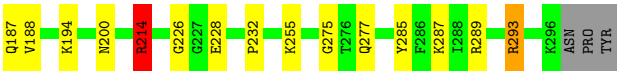




• Molecule 1: Major capsid protein VP1,Major capsid protein VP1

Chain III:





• Molecule 1: Major capsid protein VP1,Major capsid protein VP1

Chain JJJ:

4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	234.30Å 97.20Å 146.17Å 90.00° 98.38° 90.00°	Depositor
Resolution (Å)	90.00 – 2.48	Depositor
% Data completeness (in resolution range)	99.1 (90.00-2.48)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.86 (at 2.48Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.205 , 0.233	Depositor
Wilson B-factor (Å^2)	39.2	Xtriage
Anisotropy	0.394	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	22259	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.31% 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: DMS, CA

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	0/2203	0.88	0/2994
1	BBB	0.69	0/2140	0.84	0/2907
1	CCC	0.75	2/2205 (0.1%)	0.88	0/2995
1	DDD	0.74	1/2204 (0.0%)	0.94	5/2997 (0.2%)
1	EEE	0.73	1/2202 (0.0%)	0.88	3/2992 (0.1%)
1	FFF	0.71	0/2224	0.90	2/3019 (0.1%)
1	GGG	0.69	0/2213	0.87	3/3005 (0.1%)
1	HHH	0.71	0/2150	0.86	1/2919 (0.0%)
1	III	0.72	0/2156	0.91	5/2928 (0.2%)
1	JJJ	0.73	1/2174 (0.0%)	0.90	2/2953 (0.1%)
All	All	0.72	5/21871 (0.0%)	0.89	21/29709 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	HHH	0	1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	EEE	228	GLU	CD-OE2	7.66	1.34	1.25
1	JJJ	206	GLU	CD-OE1	7.16	1.33	1.25
1	DDD	206	GLU	CD-OE1	5.67	1.31	1.25
1	CCC	254	CYS	C-O	5.64	1.34	1.23
1	CCC	81	GLU	CD-OE2	5.02	1.31	1.25

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	III	293	ARG	NE-CZ-NH1	12.11	126.36	120.30
1	FFF	293	ARG	NE-CZ-NH1	-11.03	114.79	120.30
1	DDD	38	ASP	CB-CA-C	8.10	126.60	110.40
1	EEE	214	ARG	CB-CG-CD	7.74	131.74	111.60
1	JJJ	63	ARG	NE-CZ-NH2	-7.20	116.70	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	HHH	40	GLY	Peptide

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	AAA	2154	0	2072	23	0
1	BBB	2092	0	2017	20	0
1	CCC	2157	0	2068	31	0
1	DDD	2155	0	2063	42	0
1	EEE	2153	0	2075	25	0
1	FFF	2174	0	2093	28	0
1	GGG	2164	0	2088	30	0
1	HHH	2102	0	2027	30	0
1	III	2108	0	2031	31	0
1	JJJ	2126	0	2040	24	0
2	CCC	4	0	6	0	0
2	HHH	4	0	6	0	0
3	DDD	1	0	0	0	0
3	FFF	1	0	0	0	0
3	III	1	0	0	0	0
4	AAA	93	0	0	1	0
4	BBB	89	0	0	0	0
4	CCC	92	0	0	2	0
4	DDD	88	0	0	1	0
4	EEE	84	0	0	3	0
4	FFF	98	0	0	3	0
4	GGG	80	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	HHH	64	0	0	1	0
4	III	84	0	0	0	0
4	JJJ	91	0	0	2	0
All	All	22259	0	20586	238	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 6.

The worst 5 of 238 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:CCC:109:MET:CE	1:CCC:255:LYS:HA	1.79	1.10
1:BBB:41:VAL:HG12	1:HHH:295:VAL:HG12	1.39	1.03
1:CCC:109:MET:HE3	1:CCC:255:LYS:HA	1.50	0.94
1:III:157:GLU:OE2	1:III:255:LYS:HE3	1.73	0.87
1:HHH:68:LYS:HB2	1:HHH:276:THR:HG23	1.55	0.85

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	277/283 (98%)	265 (96%)	11 (4%)	1 (0%)	34	52
1	BBB	267/283 (94%)	257 (96%)	9 (3%)	1 (0%)	34	52
1	CCC	275/283 (97%)	265 (96%)	9 (3%)	1 (0%)	34	52
1	DDD	278/283 (98%)	265 (95%)	12 (4%)	1 (0%)	34	52
1	EEE	277/283 (98%)	266 (96%)	10 (4%)	1 (0%)	34	52
1	FFF	276/283 (98%)	268 (97%)	7 (2%)	1 (0%)	34	52
1	GGG	275/283 (97%)	265 (96%)	8 (3%)	2 (1%)	22	36

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	HHH	267/283 (94%)	255 (96%)	11 (4%)	1 (0%)	34	52
1	III	268/283 (95%)	257 (96%)	10 (4%)	1 (0%)	34	52
1	JJJ	272/283 (96%)	262 (96%)	9 (3%)	1 (0%)	34	52
All	All	2732/2830 (96%)	2625 (96%)	96 (4%)	11 (0%)	34	52

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	GGG	43	ALA
1	CCC	188	VAL
1	EEE	188	VAL
1	GGG	188	VAL
1	HHH	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	238/243 (98%)	236 (99%)	2 (1%)	81	92
1	BBB	230/243 (95%)	227 (99%)	3 (1%)	69	86
1	CCC	238/243 (98%)	234 (98%)	4 (2%)	60	81
1	DDD	237/243 (98%)	230 (97%)	7 (3%)	41	65
1	EEE	237/243 (98%)	234 (99%)	3 (1%)	69	86
1	FFF	241/243 (99%)	239 (99%)	2 (1%)	81	92
1	GGG	240/243 (99%)	234 (98%)	6 (2%)	47	71
1	HHH	232/243 (96%)	225 (97%)	7 (3%)	41	65
1	III	233/243 (96%)	229 (98%)	4 (2%)	60	81
1	JJJ	233/243 (96%)	229 (98%)	4 (2%)	60	81
All	All	2359/2430 (97%)	2317 (98%)	42 (2%)	59	80

5 of 42 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	HHH	27	MET
1	III	100	GLU
1	HHH	39	THR
1	HHH	108	LEU
1	III	214	ARG

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)

Of 5 ligands modelled in this entry, 3 are monoatomic - leaving 2 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	DMS	HHH	301	-	3,3,3	0.42	0	3,3,3	0.36	0
2	DMS	CCC	301	-	3,3,3	0.51	0	3,3,3	0.31	0

There are no bond length outliers.

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

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.