



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 09:48 pm GMT

PDB ID : 5L7I  
Title : Structure of human Smoothened in complex with Vismodegib  
Authors : Byrne, E.X.B.; Sircar, R.; Miller, P.S.; Hedger, G.; Luchetti, G.; Nachtergaele, S.; Tully, M.D.; Mydock-McGrane, L.; Covey, D.F.; Rambo, R.F.; Sansom, M.S.P.; Newstead, S.; Rohatgi, R.; Siebold, C.  
Deposited on : 2016-06-03  
Resolution : 3.30 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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

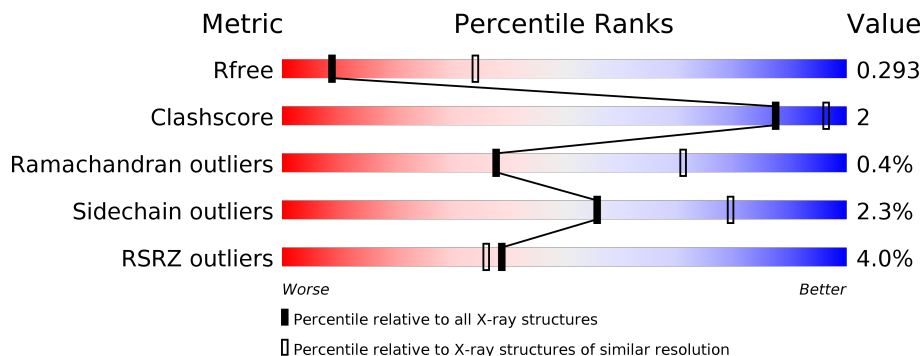
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.30 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))
$R_{free}$	100719	1034 (3.36-3.24)
Clashscore	112137	1100 (3.36-3.24)
Ramachandran outliers	110173	1081 (3.36-3.24)
Sidechain outliers	110143	1080 (3.36-3.24)
RSRZ outliers	101464	1039 (3.36-3.24)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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  $\leq 5\%$ . 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	A	638	<div> <div>3%</div> <div> <div></div> <div>80%</div> <div>8%</div> <div>12%</div> </div> </div>
1	B	638	<div> <div>4%</div> <div> <div></div> <div>82%</div> <div>7%</div> <div>10%</div> </div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 9107 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 Smoothened homolog,Soluble cytochrome b562,Smoothened homolog.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	564	Total	C	N	O	S	0	0	0
			4464	2870	764	798	32			
1	B	572	Total	C	N	O	S	0	0	0
			4524	2909	773	810	32			

There are 52 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	329	PHE	VAL	conflict	UNP Q99835
A	1011	ALA	-	linker	UNP Q99835
A	1012	ARG	-	linker	UNP Q99835
A	1013	ARG	-	linker	UNP Q99835
A	1014	GLN	-	linker	UNP Q99835
A	1015	LEU	-	linker	UNP Q99835
A	1022	TRP	MET	conflict	UNP P0ABE7
A	1117	ILE	HIS	conflict	UNP P0ABE7
A	1121	LEU	-	linker	UNP P0ABE7
A	1122	GLU	-	linker	UNP P0ABE7
A	1123	ARG	-	linker	UNP P0ABE7
A	1124	ALA	-	linker	UNP P0ABE7
A	1125	ARG	-	linker	UNP P0ABE7
A	1126	SER	-	linker	UNP P0ABE7
A	1127	THR	-	linker	UNP P0ABE7
A	1128	LEU	-	linker	UNP P0ABE7
A	556	GLY	-	expression tag	UNP Q99835
A	557	THR	-	expression tag	UNP Q99835
A	558	GLU	-	expression tag	UNP Q99835
A	559	THR	-	expression tag	UNP Q99835
A	560	SER	-	expression tag	UNP Q99835
A	561	GLN	-	expression tag	UNP Q99835
A	562	VAL	-	expression tag	UNP Q99835
A	563	ALA	-	expression tag	UNP Q99835

*Continued on next page...*

*Continued from previous page...*

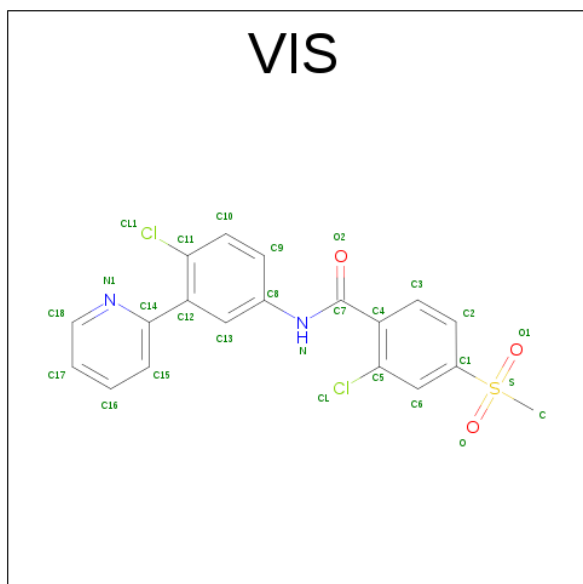
Chain	Residue	Modelled	Actual	Comment	Reference
A	564	PRO	-	expression tag	UNP Q99835
A	565	ALA	-	expression tag	UNP Q99835
B	329	PHE	VAL	conflict	UNP Q99835
B	1011	ALA	-	linker	UNP Q99835
B	1012	ARG	-	linker	UNP Q99835
B	1013	ARG	-	linker	UNP Q99835
B	1014	GLN	-	linker	UNP Q99835
B	1015	LEU	-	linker	UNP Q99835
B	1022	TRP	MET	conflict	UNP P0ABE7
B	1117	ILE	HIS	conflict	UNP P0ABE7
B	1121	LEU	-	linker	UNP P0ABE7
B	1122	GLU	-	linker	UNP P0ABE7
B	1123	ARG	-	linker	UNP P0ABE7
B	1124	ALA	-	linker	UNP P0ABE7
B	1125	ARG	-	linker	UNP P0ABE7
B	1126	SER	-	linker	UNP P0ABE7
B	1127	THR	-	linker	UNP P0ABE7
B	1128	LEU	-	linker	UNP P0ABE7
B	556	GLY	-	expression tag	UNP Q99835
B	557	THR	-	expression tag	UNP Q99835
B	558	GLU	-	expression tag	UNP Q99835
B	559	THR	-	expression tag	UNP Q99835
B	560	SER	-	expression tag	UNP Q99835
B	561	GLN	-	expression tag	UNP Q99835
B	562	VAL	-	expression tag	UNP Q99835
B	563	ALA	-	expression tag	UNP Q99835
B	564	PRO	-	expression tag	UNP Q99835
B	565	ALA	-	expression tag	UNP Q99835

- Molecule 2 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 3 is 2-chloranyl- {N}-(4-chloranyl-3-pyridin-2-yl-phenyl)-4-methylsulfonyl-benzamide (three-letter code: VIS) (formula:  $C_{19}H_{14}Cl_2N_2O_3S$ ).

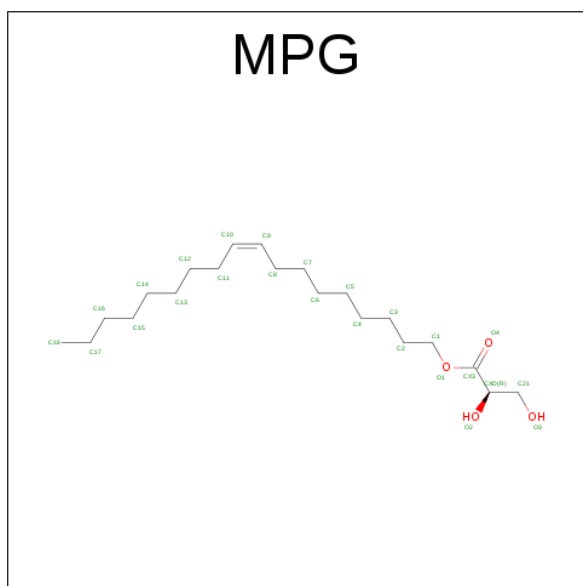


Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	A	1	Total	C	Cl	N	O	S	0	0
			27	19	2	2	3	1		
3	B	1	Total	C	Cl	N	O	S	0	0
			27	19	2	2	3	1		

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Na	0	0
			1	1		

- Molecule 5 is [(Z)-octadec-9-enyl] (2R)-2,3-bis(oxidanyl)propanoate (three-letter code: MPG) (formula: C<sub>21</sub>H<sub>40</sub>O<sub>4</sub>).

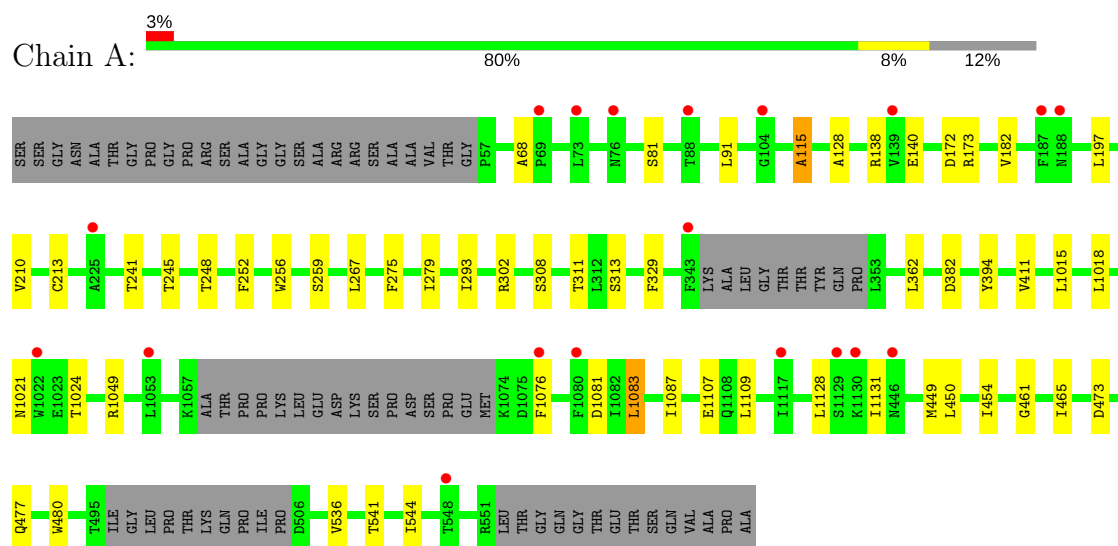


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	B	1	Total	C	O	0	0
			25	21	4		
5	B	1	Total	C	O	0	0
			25	21	4		

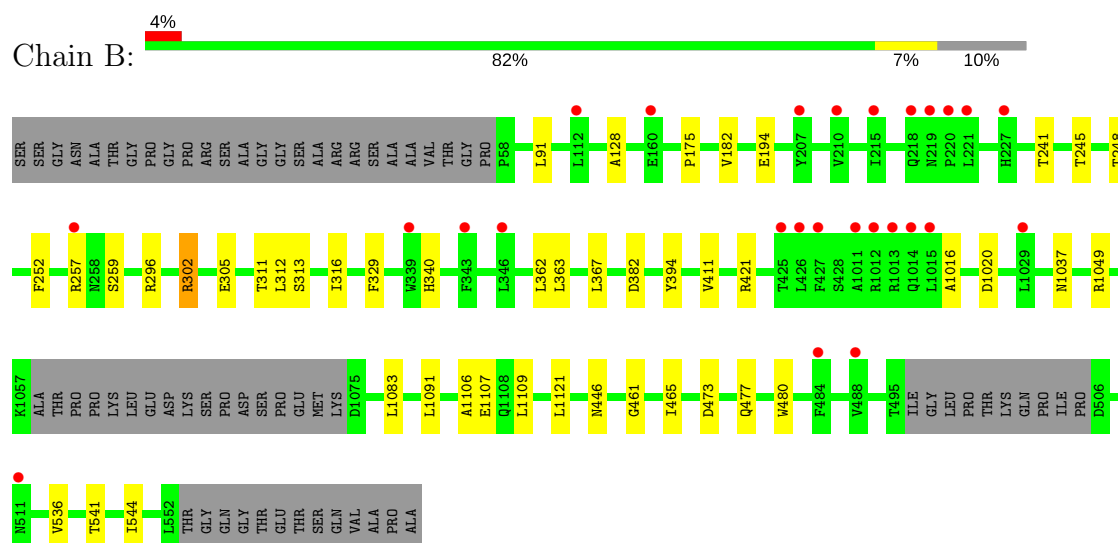
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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: Smoothened homolog,Soluble cytochrome b562,Smoothened homolog



- Molecule 1: Smoothened homolog,Soluble cytochrome b562,Smoothened homolog



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	52.82Å 71.29Å 107.02Å 91.53° 98.20° 105.92°	Depositor
Resolution (Å)	59.32 – 3.30 59.32 – 3.30	Depositor EDS
% Data completeness (in resolution range)	94.5 (59.32-3.30) 92.4 (59.32-3.30)	Depositor EDS
$R_{merge}$	0.15	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.24 (at 3.33Å)	Xtriage
Refinement program	BUSTER 2.10.2	Depositor
R, $R_{free}$	0.245 , 0.257 0.270 , 0.293	Depositor DCC
$R_{free}$ test set	1012 reflections (4.92%)	DCC
Wilson B-factor (Å <sup>2</sup> )	94.1	Xtriage
Anisotropy	0.398	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 81.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	9107	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	123.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.03% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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: MPG, NAG, NA, VIS

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	0.41	0/4574	0.55	0/6212
1	B	0.41	0/4636	0.55	0/6299
All	All	0.41	0/9210	0.55	0/12511

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	A	4464	0	4391	24	0
1	B	4524	0	4454	20	0
2	A	14	0	13	0	0
3	A	27	0	0	0	0
3	B	27	0	0	0	0
4	B	1	0	0	0	0
5	B	50	0	80	1	0
All	All	9107	0	8938	43	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 43 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:1049:ARG:HH11	1:B:1091:LEU:HD21	1.52	0.73
1:B:296:ARG:HG3	1:B:302:ARG:HG3	1.78	0.65
1:B:382:ASP:HB2	1:B:394:TYR:HB2	1.81	0.63
1:A:382:ASP:HB2	1:A:394:TYR:HB2	1.81	0.63
1:B:363:LEU:HD23	1:B:367:LEU:HD11	1.89	0.55

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	556/638 (87%)	540 (97%)	14 (2%)	2 (0%)	38	71
1	B	566/638 (89%)	545 (96%)	19 (3%)	2 (0%)	38	71
All	All	1122/1276 (88%)	1085 (97%)	33 (3%)	4 (0%)	38	71

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	313	SER
1	B	446	ASN
1	A	313	SER
1	A	115	ALA

### 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	474/529 (90%)	464 (98%)	10 (2%)	59	81
1	B	480/529 (91%)	468 (98%)	12 (2%)	53	78
All	All	954/1058 (90%)	932 (98%)	22 (2%)	56	79

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	477	GLN
1	B	194	GLU
1	B	1121	LEU
1	B	91	LEU
1	B	175	PRO

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	340	HIS
1	B	477	GLN
1	B	380	GLN
1	B	123	GLN
1	B	1021	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry

Of 6 ligands modelled in this entry, 1 is monoatomic - leaving 5 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	NAG	A	1201	1	14,14,15	0.30	0	15,19,21	0.45	0
3	VIS	A	1202	-	29,29,29	0.19	0	42,42,42	0.34	0
3	VIS	B	1202	-	29,29,29	0.20	0	42,42,42	0.33	0
5	MPG	B	1203	-	24,24,24	1.35	1 (4%)	23,25,25	0.91	2 (8%)
5	MPG	B	1204	-	24,24,24	1.35	1 (4%)	23,25,25	0.93	1 (4%)

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	NAG	A	1201	1	-	0/6/23/26	0/1/1/1
3	VIS	A	1202	-	-	0/18/18/18	0/3/3/3
3	VIS	B	1202	-	-	0/18/18/18	0/3/3/3
5	MPG	B	1203	-	-	0/25/25/25	0/0/0/0
5	MPG	B	1204	-	-	0/25/25/25	0/0/0/0

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	B	1203	MPG	O1-CX3	6.22	1.46	1.33
5	B	1204	MPG	O1-CX3	6.37	1.46	1.33

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	B	1204	MPG	O1-CX3-O4	-2.93	118.41	124.11
5	B	1203	MPG	O1-CX3-O4	-2.92	118.44	124.11
5	B	1203	MPG	C1-O1-CX3	2.37	121.68	116.64

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	B	1203	MPG	1	0

## 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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	564/638 (88%)	0.11	19 (3%) 46 42	68, 127, 170, 206	0
1	B	572/638 (89%)	0.18	26 (4%) 34 32	62, 116, 173, 254	0
All	All	1136/1276 (89%)	0.14	45 (3%) 39 36	62, 120, 171, 254	0

The worst 5 of 45 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	343	PHE	5.1
1	A	1129	SER	4.7
1	B	220	PRO	4.5
1	A	1130	LYS	3.9
1	A	104	GLY	3.8

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [i](#)

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
5	MPG	B	1204	25/25	0.77	0.29	1.29	81,93,121,121	0
2	NAG	A	1201	14/15	0.66	0.31	0.88	163,164,165,165	0
3	VIS	B	1202	27/27	0.82	0.39	0.87	135,142,155,156	0
3	VIS	A	1202	27/27	0.91	0.28	0.57	109,113,120,120	0
5	MPG	B	1203	25/25	0.85	0.22	0.54	74,85,105,105	0
4	NA	B	1201	1/1	0.97	0.15	-1.33	49,49,49,49	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.