



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 01:58 am GMT

PDB ID : 5C3M  
Title : Crystal structure of Gan4C, a GH4 6-phospho-glucosidase from *Geobacillus stearothermophilus*  
Authors : Cohen, T.; Lansky, S.; Zehavi, A.; Shoham, Y.; Shoham, G.  
Deposited on : 2015-06-17  
Resolution : 3.06 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
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

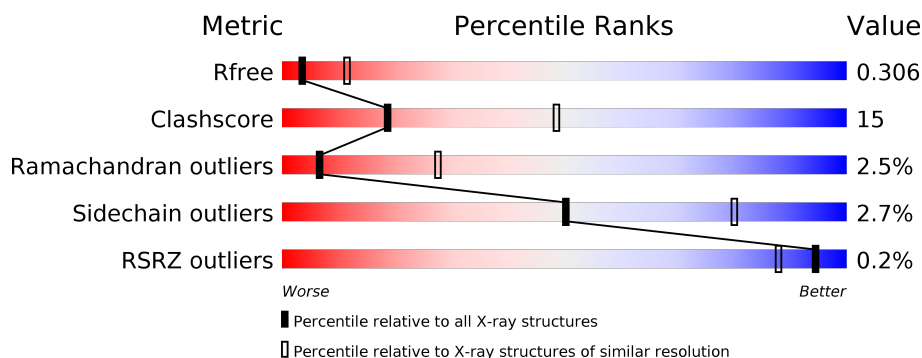
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.06 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	1348 (3.10-3.02)
Clashscore	112137	1462 (3.10-3.02)
Ramachandran outliers	110173	1410 (3.10-3.02)
Sidechain outliers	110143	1410 (3.10-3.02)
RSRZ outliers	101464	1355 (3.10-3.02)

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	447	
1	B	447	
1	C	447	
1	D	447	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	MN	A	501	-	-	X	-
2	MN	B	501	-	-	X	-
2	MN	C	501	-	-	X	-
2	MN	D	501	-	-	X	-

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 12494 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 Putative 6-phospho-beta-glucosidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	410	Total	C	N	O	S	0	0	0
			3080	1966	514	584	16			
1	B	411	Total	C	N	O	S	0	0	0
			3069	1964	516	573	16			
1	C	410	Total	C	N	O	S	0	0	0
			3114	1996	522	580	16			
1	D	409	Total	C	N	O	S	0	0	1
			3136	2006	528	586	16			

- Molecule 2 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Mn	0	0
			1	1		
2	A	1	Total	Mn	0	0
			1	1		
2	D	1	Total	Mn	0	0
			1	1		
2	C	1	Total	Mn	0	0
			1	1		

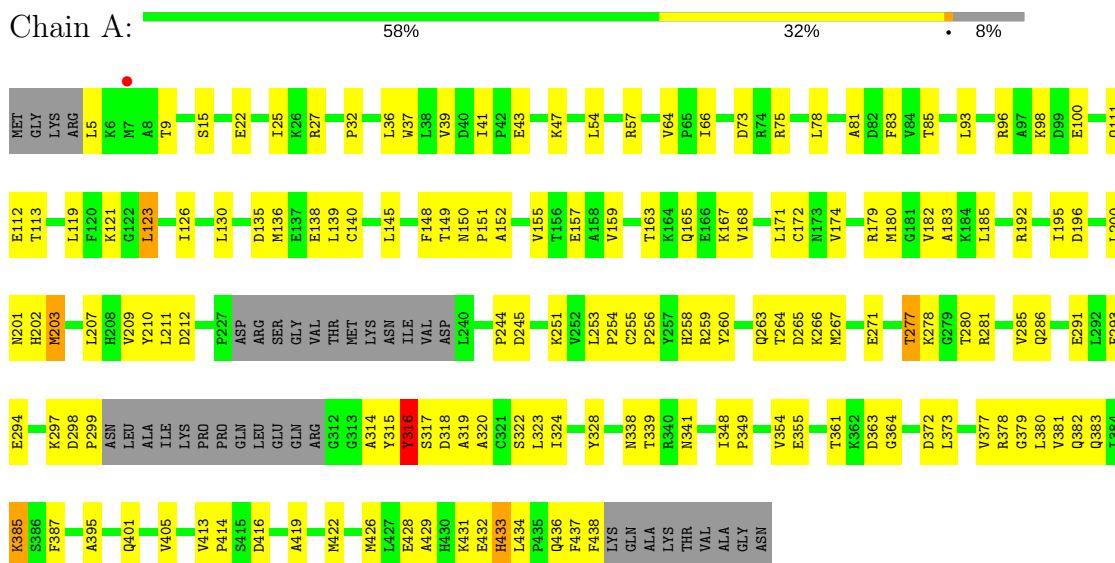
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	21	Total	O	0	0
			21	21		
3	B	27	Total	O	0	0
			27	27		
3	C	19	Total	O	0	0
			19	19		
3	D	24	Total	O	0	0
			24	24		

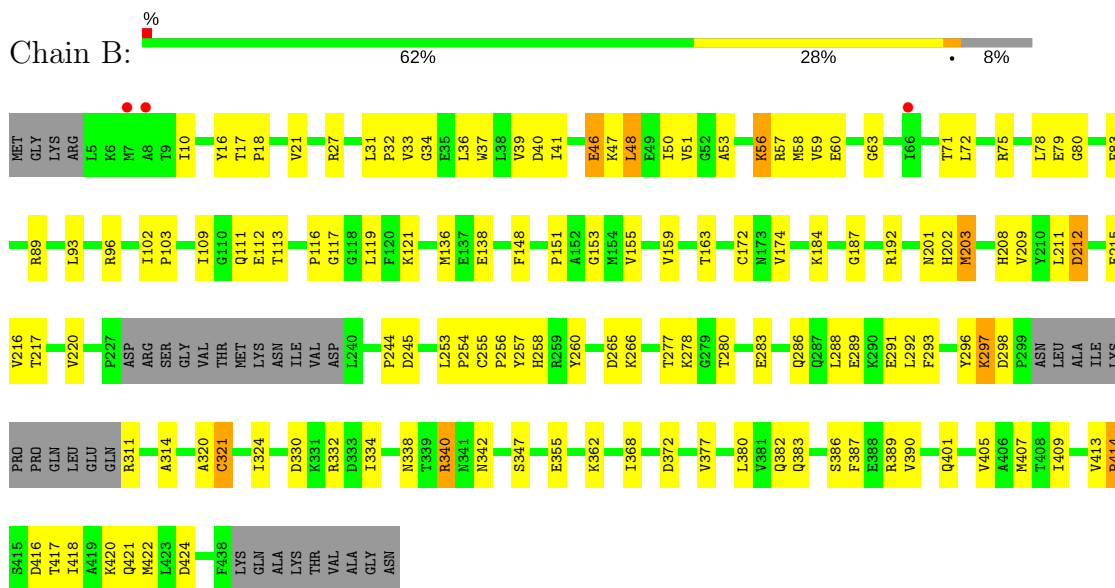
### 3 Residue-property plots

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: Putative 6-phospho-beta-glucosidase

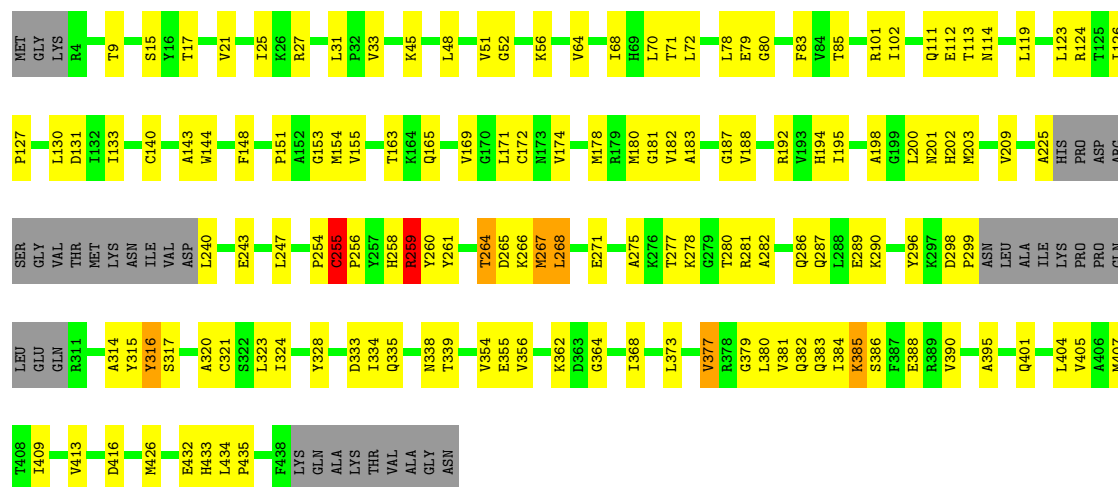


#### ● Molecule 1: Putative 6-phospho-beta-glucosidase



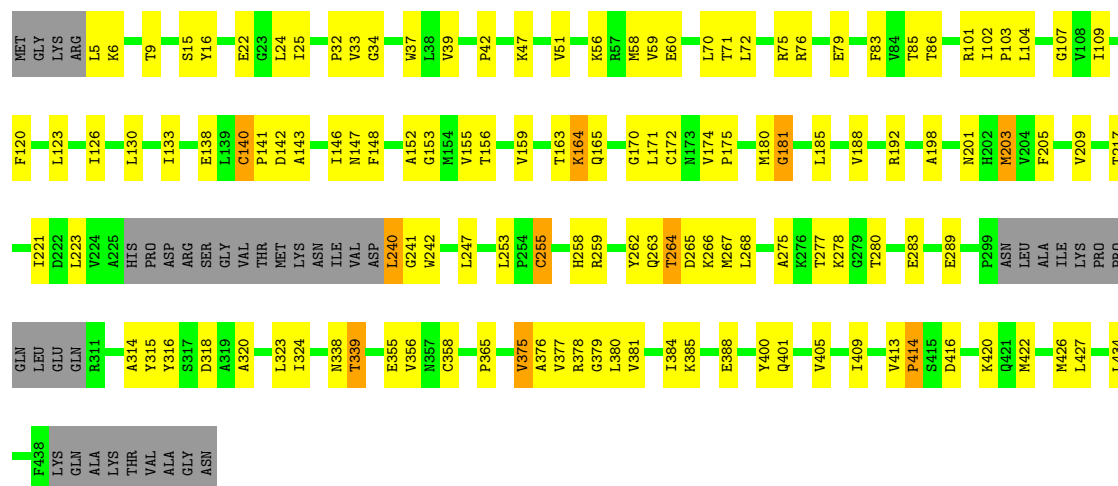
#### ● Molecule 1: Putative 6-phospho-beta-glucosidase

Chain C: 



• Molecule 1: Putative 6-phospho-beta-glucosidase

Chain D: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	185.32Å 83.20Å 151.37Å 90.00° 113.61° 90.00°	Depositor
Resolution (Å)	46.60 – 3.06 46.60 – 3.06	Depositor EDS
% Data completeness (in resolution range)	98.9 (46.60-3.06) 99.0 (46.60-3.06)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.36 (at 3.06Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.230 , 0.306 0.230 , 0.306	Depositor DCC
$R_{free}$ test set	1963 reflections (4.94%)	DCC
Wilson B-factor (Å <sup>2</sup> )	61.0	Xtriage
Anisotropy	0.743	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 35.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	12494	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 44.26 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 1.5664e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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: MN

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	0.50	0/3141	0.60	0/4275
1	B	0.51	0/3130	0.61	0/4260
1	C	0.56	1/3173 (0.0%)	0.63	4/4314 (0.1%)
1	D	0.59	1/3195 (0.0%)	0.61	1/4338 (0.0%)
All	All	0.54	2/12639 (0.0%)	0.61	5/17187 (0.0%)

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	B	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	358	CYS	CB-SG	-5.65	1.72	1.81
1	C	255	CYS	CB-SG	-5.37	1.73	1.81

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	240	LEU	CB-CG-CD2	-6.92	99.24	111.00
1	C	267	MET	N-CA-C	6.57	128.74	111.00
1	C	259	ARG	NE-CZ-NH1	-5.81	117.39	120.30
1	C	267	MET	C-N-CA	5.34	135.05	121.70
1	D	181	GLY	N-CA-C	-5.26	99.94	113.10

There are no chirality outliers.



All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	136	MET	Peptide

## 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	3080	0	2981	106	0
1	B	3069	0	2968	88	0
1	C	3114	0	3094	98	1
1	D	3136	0	3145	81	1
2	A	1	0	0	2	0
2	B	1	0	0	2	0
2	C	1	0	0	3	0
2	D	1	0	0	2	0
3	A	21	0	0	2	0
3	B	27	0	0	2	0
3	C	19	0	0	3	0
3	D	24	0	0	0	0
All	All	12494	0	12188	364	1

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 15.

The worst 5 of 364 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:C:52:GLY:HA3	1:C:70:LEU:HD21	1.52	0.92
1:C:27:ARG:HH21	1:C:321:CYS:HB2	1.36	0.90
1:D:140:CYS:O	1:D:142:ASP:N	2.03	0.90
1:C:172:CYS:SG	2:C:501:MN:MN	1.83	0.77
1:A:265:ASP:OD1	1:A:266:LYS:N	2.17	0.76

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:131:ASP:OD1	1:D:76:ARG:NH2[3_556]	2.13	0.07

## 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	404/447 (90%)	361 (89%)	34 (8%)	9 (2%)	8	31
1	B	405/447 (91%)	353 (87%)	37 (9%)	15 (4%)	4	20
1	C	404/447 (90%)	360 (89%)	39 (10%)	5 (1%)	15	47
1	D	403/447 (90%)	356 (88%)	35 (9%)	12 (3%)	5	24
All	All	1616/1788 (90%)	1430 (88%)	145 (9%)	41 (2%)	6	27

5 of 41 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	73	ASP
1	A	149	THR
1	A	277	THR
1	B	46	GLU
1	B	57	ARG

### 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	308/371 (83%)	297 (96%)	11 (4%)	40 74

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	302/371 (81%)	294 (97%)	8 (3%)	51	81
1	C	319/371 (86%)	313 (98%)	6 (2%)	62	85
1	D	329/371 (89%)	320 (97%)	9 (3%)	50	80
All	All	1258/1484 (85%)	1224 (97%)	34 (3%)	50	80

5 of 34 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	311	ARG
1	C	255	CYS
1	D	264	THR
1	B	332	ARG
1	A	323	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	165	GLN
1	A	173	ASN

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

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 [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	410/447 (91%)	-0.34	1 (0%) 94 88	33, 61, 82, 97	0
1	B	411/447 (91%)	-0.41	3 (0%) 87 72	36, 56, 77, 87	0
1	C	410/447 (91%)	-0.54	0 100 100	27, 50, 66, 79	0
1	D	409/447 (91%)	-0.52	0 100 100	28, 49, 65, 91	0
All	All	1640/1788 (91%)	-0.45	4 (0%) 94 88	27, 52, 77, 97	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	8	ALA	2.4
1	B	66	ILE	2.4
1	A	7	MET	2.3
1	B	7	MET	2.3

### 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
2	MN	B	501	1/1	0.99	0.09	-2.26	50,50,50,50	0
2	MN	D	501	1/1	0.98	0.10	-2.33	54,54,54,54	0
2	MN	A	501	1/1	0.98	0.08	-3.06	58,58,58,58	0
2	MN	C	501	1/1	0.98	0.10	-3.34	51,51,51,51	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.