wwPDB X-ray Structure Validation Summary Report (i)

Sep 28, 2016 – 01:15 PM EDT

PDB ID	:	5JRK
Title	:	Crystal Structure of the Sphingopyxin I Lasso Peptide Isopeptidase SpI-IsoP (SeMet-derived)
Authors	:	Fage, C.D.; Hegemann, J.D.; Bange, G.; Marahiel, M.A.
Deposited on	:	2016-05-06
Resolution	:	3.00 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

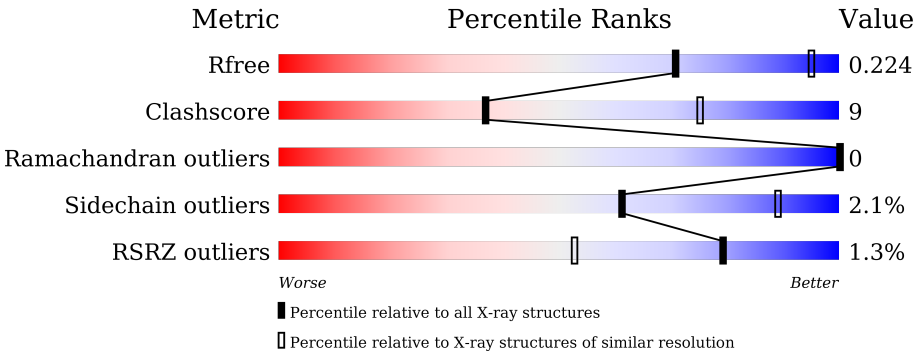
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

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 3.00 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}	91344	1578 (3.00-3.00)
Clashscore	102246	1912 (3.00-3.00)
Ramachandran outliers	100387	1853 (3.00-3.00)
Sidechain outliers	100360	1856 (3.00-3.00)
RSRZ outliers	91569	1592 (3.00-3.00)

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 >=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 <=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	756	71% 21% • 8%
1	B	756	2% 68% 21% • 10%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 10681 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 Dipeptidyl aminopeptidases/acylaminoacyl-peptidases-like protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	695	Total	C	N	O	S	Se	0	0	0
			5393	3373	989	1017	4	10			
1	B	677	Total	C	N	O	S	Se	0	0	0
			5264	3295	967	989	4	9			

There are 60 discrepancies between the modelled and reference sequences:

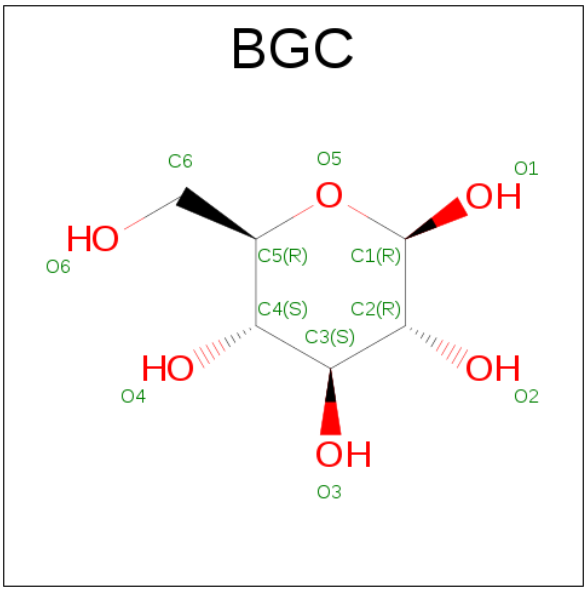
Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MSE	-	initiating methionine	UNP Q1GQ33
A	-18	GLY	-	expression tag	UNP Q1GQ33
A	-17	SER	-	expression tag	UNP Q1GQ33
A	-16	SER	-	expression tag	UNP Q1GQ33
A	-15	HIS	-	expression tag	UNP Q1GQ33
A	-14	HIS	-	expression tag	UNP Q1GQ33
A	-13	HIS	-	expression tag	UNP Q1GQ33
A	-12	HIS	-	expression tag	UNP Q1GQ33
A	-11	HIS	-	expression tag	UNP Q1GQ33
A	-10	HIS	-	expression tag	UNP Q1GQ33
A	-9	SER	-	expression tag	UNP Q1GQ33
A	-8	SER	-	expression tag	UNP Q1GQ33
A	-7	GLY	-	expression tag	UNP Q1GQ33
A	-6	LEU	-	expression tag	UNP Q1GQ33
A	-5	VAL	-	expression tag	UNP Q1GQ33
A	-4	PRO	-	expression tag	UNP Q1GQ33
A	-3	ARG	-	expression tag	UNP Q1GQ33
A	-2	GLY	-	expression tag	UNP Q1GQ33
A	-1	SER	-	expression tag	UNP Q1GQ33
A	0	HIS	-	expression tag	UNP Q1GQ33
A	727	SER	-	expression tag	UNP Q1GQ33
A	728	ALA	-	expression tag	UNP Q1GQ33
A	729	TRP	-	expression tag	UNP Q1GQ33
A	730	SER	-	expression tag	UNP Q1GQ33

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	731	HIS	-	expression tag	UNP Q1GQ33
A	732	PRO	-	expression tag	UNP Q1GQ33
A	733	GLN	-	expression tag	UNP Q1GQ33
A	734	PHE	-	expression tag	UNP Q1GQ33
A	735	GLU	-	expression tag	UNP Q1GQ33
A	736	LYS	-	expression tag	UNP Q1GQ33
B	-19	MSE	-	initiating methionine	UNP Q1GQ33
B	-18	GLY	-	expression tag	UNP Q1GQ33
B	-17	SER	-	expression tag	UNP Q1GQ33
B	-16	SER	-	expression tag	UNP Q1GQ33
B	-15	HIS	-	expression tag	UNP Q1GQ33
B	-14	HIS	-	expression tag	UNP Q1GQ33
B	-13	HIS	-	expression tag	UNP Q1GQ33
B	-12	HIS	-	expression tag	UNP Q1GQ33
B	-11	HIS	-	expression tag	UNP Q1GQ33
B	-10	HIS	-	expression tag	UNP Q1GQ33
B	-9	SER	-	expression tag	UNP Q1GQ33
B	-8	SER	-	expression tag	UNP Q1GQ33
B	-7	GLY	-	expression tag	UNP Q1GQ33
B	-6	LEU	-	expression tag	UNP Q1GQ33
B	-5	VAL	-	expression tag	UNP Q1GQ33
B	-4	PRO	-	expression tag	UNP Q1GQ33
B	-3	ARG	-	expression tag	UNP Q1GQ33
B	-2	GLY	-	expression tag	UNP Q1GQ33
B	-1	SER	-	expression tag	UNP Q1GQ33
B	0	HIS	-	expression tag	UNP Q1GQ33
B	727	SER	-	expression tag	UNP Q1GQ33
B	728	ALA	-	expression tag	UNP Q1GQ33
B	729	TRP	-	expression tag	UNP Q1GQ33
B	730	SER	-	expression tag	UNP Q1GQ33
B	731	HIS	-	expression tag	UNP Q1GQ33
B	732	PRO	-	expression tag	UNP Q1GQ33
B	733	GLN	-	expression tag	UNP Q1GQ33
B	734	PHE	-	expression tag	UNP Q1GQ33
B	735	GLU	-	expression tag	UNP Q1GQ33
B	736	LYS	-	expression tag	UNP Q1GQ33

- Molecule 2 is BETA-D-GLUCOSE (three-letter code: BGC) (formula: $C_6H_{12}O_6$).



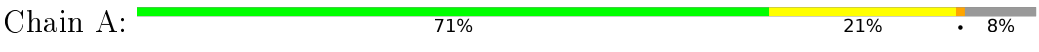
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			12	6	6		
2	B	1	Total	C	O	0	0
			12	6	6		

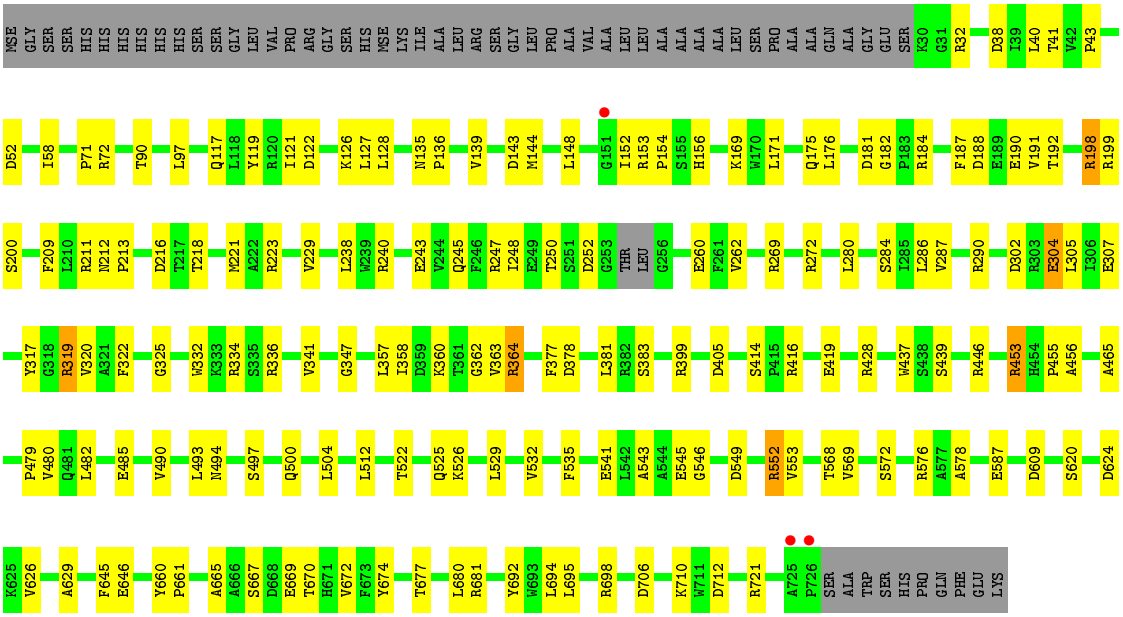
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: Dipeptidyl aminopeptidases/acylaminoacyl-peptidases-like protein

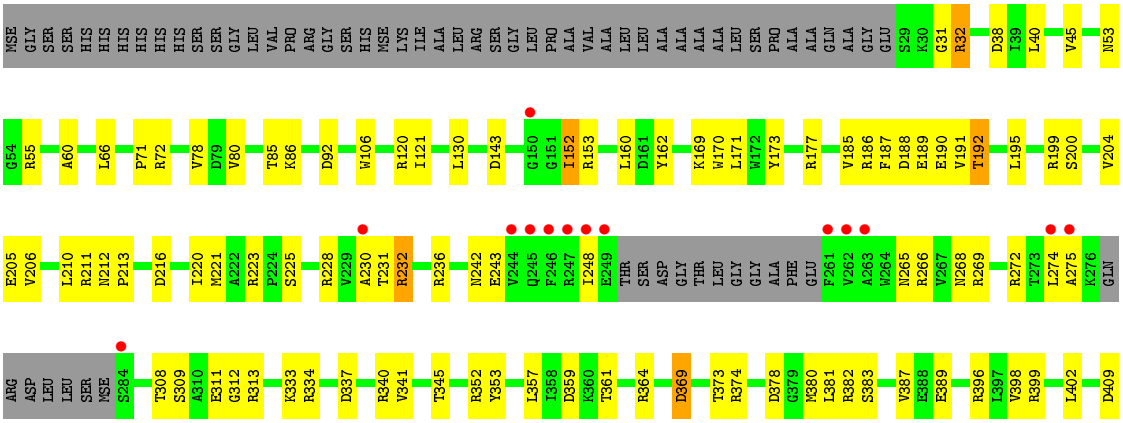
Chain A:

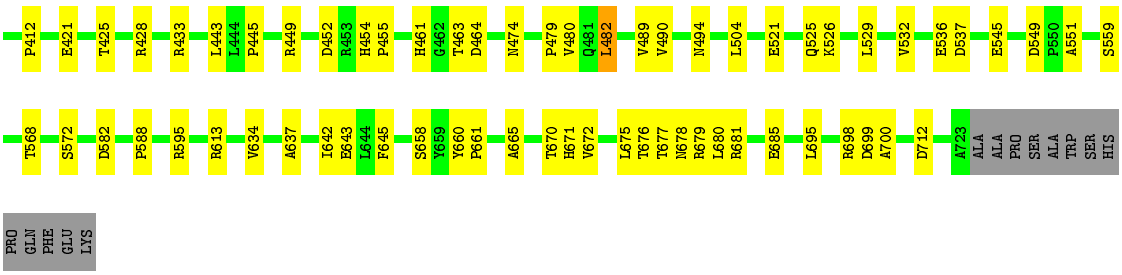




- Molecule 1: Dipeptidyl aminopeptidases/acylaminoacyl-peptidases-like protein

Chain B:




P412
E421
T425
R428
R433
L443
L444
P445
R449
D452
R453
H454
P455
H461
G462
T463
D464
N474
P479
V480
Q481
L482
V489
V490
N494
L504
E521
Q525
K526
L529
V532
E536
D537
E545
D549
P550
A551
S559
T568
S572
D582
P588
R595
R613
V634
A637
I642
E643
L644
F645
S658
Y659
Y660
P661
A665
T670
H671
V672
L675
T676
T677
N678
R679
L680
R681
E685
L695
R698
D699
A700
D712
A723
ALA
ALA
PRO
SER
ALA
TRP
SER
HIS
PRO
GLN
PHE
GLU
LYS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	111.11Å 133.50Å 169.07Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.08 – 3.00 49.08 – 3.00	Depositor EDS
% Data completeness (in resolution range)	100.0 (49.08-3.00) 100.0 (49.08-3.00)	Depositor EDS
R_{merge}	0.22	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.76 (at 3.01Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.237 , 0.259 0.223 , 0.224	Depositor DCC
R_{free} test set	2607 reflections (5.11%)	DCC
Wilson B-factor (Å^2)	88.6	Xtriage
Anisotropy	0.456	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 36.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	10681	wwPDB-VP
Average B, all atoms (Å^2)	79.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.54% 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

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	0.26	0/5501	0.45	0/7453
1	B	0.26	0/5370	0.45	0/7276
All	All	0.26	0/10871	0.45	0/14729

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	5393	0	5316	108	0
1	B	5264	0	5193	106	0
2	A	12	0	12	2	0
2	B	12	0	12	0	0
All	All	10681	0	10533	200	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 9.

The worst 5 of 200 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:A:710:LYS:NZ	1:B:188:ASP:OD2	1.81	1.12
1:A:446:ARG:NH1	1:A:485:GLU:OE2	1.93	1.00
1:A:32:ARG:NH1	1:A:38:ASP:OD2	2.08	0.86
1:B:243:GLU:OE2	1:B:272:ARG:NH1	2.08	0.86
1:A:428:ARG:NH1	1:A:545:GLU:OE2	2.09	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	A	691/756 (91%)	663 (96%)	28 (4%)	0	100	100
1	B	671/756 (89%)	646 (96%)	25 (4%)	0	100	100
All	All	1362/1512 (90%)	1309 (96%)	53 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	565/598 (94%)	552 (98%)	13 (2%)	58	87
1	B	552/598 (92%)	542 (98%)	10 (2%)	66	91
All	All	1117/1196 (93%)	1094 (98%)	23 (2%)	61	89

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	512	LEU
1	A	669	GLU
1	B	382	ARG
1	A	552	ARG
1	B	32	ARG

Some 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 carbohydrates in this entry.

5.6 Ligand geometry (i)

2 ligands are modelled in this entry.

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	BGC	A	801	-	12,12,12	0.56	0	17,17,17	0.51	0
2	BGC	B	801	-	12,12,12	0.55	0	17,17,17	0.46	0

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	BGC	A	801	-	-	0/2/22/22	0/1/1/1
2	BGC	B	801	-	-	0/2/22/22	0/1/1/1

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.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	801	BGC	2	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

6.1 Protein, DNA and RNA chains

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^{th} 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^2)	Q<0.9
1	A	685/756 (90%)	-0.09	3 (0%)	93	80	55, 77, 104, 137	0
1	B	668/756 (88%)	0.05	14 (2%)	67	36	52, 75, 113, 165	0
All	All	1353/1512 (89%)	-0.02	17 (1%)	79	53	52, 76, 110, 165	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	274	LEU	3.3
1	A	151	GLY	3.1
1	B	230	ALA	2.9
1	B	275	ALA	2.8
1	B	150	GLY	2.8

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

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

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

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^{th} 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(Å^2)	Q<0.9
2	BGC	B	801	12/12	0.87	0.35	1.94	93,97,103,104	0
2	BGC	A	801	12/12	0.89	0.17	-0.69	88,93,99,102	0

6.5 Other polymers (i)

There are no such residues in this entry.