W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jun 28, 2016 – 03:51 PM EDT

PDB ID	:	5AWH
Title	:	Rhodobacter sphaeroides Argonaute in complex with guide RNA/target DNA heteroduplex
Authors	:	Miyoshi, T.; Ito, K.; Murakami, R.; Uchiumi, T.
Deposited on	:	2015-07-03
Resolution	:	2.00 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 (i)) were used in the production of this report:

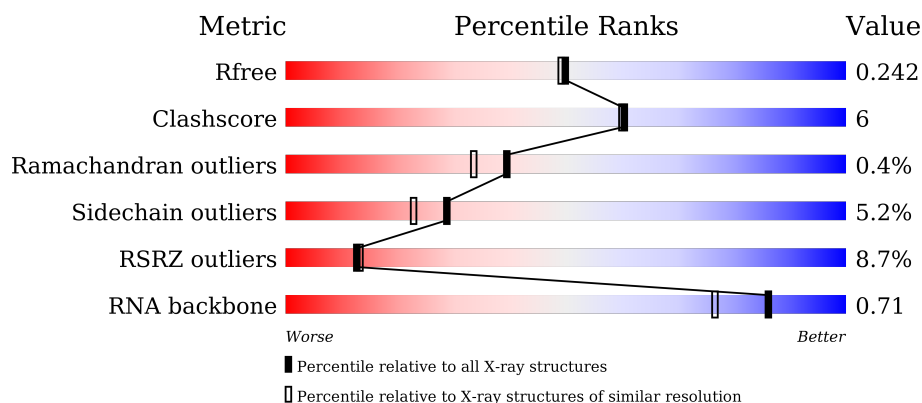|		
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027790
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027790

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 2.00 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	Percentile Ranks	Value
Rfree		0.242
Clashscore		6
Ramachandran outliers		0.4%
Sidechain outliers		5.2%
RSRZ outliers		8.7%
RNA backbone		0.71

Worse ... *Better*

■ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	91344	6249 (2.00-2.00)
Clashscore	102246	7340 (2.00-2.00)
Ramachandran outliers	100387	7248 (2.00-2.00)
Sidechain outliers	100360	7247 (2.00-2.00)
RSRZ outliers	91569	6262 (2.00-2.00)
RNA backbone	2183	1002 (2.72-1.30)

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	778	6% · 81% · 14% · ·
1	B	778	11% · 81% · 10% · 8%
2	C	18	44% · 56%
2	E	18	67% · 22% · 6% · 6%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	D	18	6% 83% 17%
3	F	18	6% 78% 17% 6%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 13805 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 Uncharacterized protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	759	Total 6005	C 3808	N 1084	O 1096	S 17	0	0	0
1	B	715	Total 5664	C 3600	N 1027	O 1022	S 15	0	0	0

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	expression tag	UNP A4WYU7
A	1	GLY	-	expression tag	UNP A4WYU7
A	2	SER	-	expression tag	UNP A4WYU7
A	3	SER	-	expression tag	UNP A4WYU7
A	4	HIS	-	expression tag	UNP A4WYU7
A	5	HIS	-	expression tag	UNP A4WYU7
A	6	HIS	-	expression tag	UNP A4WYU7
A	7	HIS	-	expression tag	UNP A4WYU7
A	8	HIS	-	expression tag	UNP A4WYU7
A	9	HIS	-	expression tag	UNP A4WYU7
A	10	SER	-	expression tag	UNP A4WYU7
A	11	SER	-	expression tag	UNP A4WYU7
A	12	GLY	-	expression tag	UNP A4WYU7
A	13	LEU	-	expression tag	UNP A4WYU7
A	14	VAL	-	expression tag	UNP A4WYU7
A	15	PRO	-	expression tag	UNP A4WYU7
A	16	ALA	-	expression tag	UNP A4WYU7
A	17	GLY	-	expression tag	UNP A4WYU7
A	18	SER	-	expression tag	UNP A4WYU7
A	19	HIS	-	expression tag	UNP A4WYU7
A	20	MET	-	expression tag	UNP A4WYU7
B	0	MET	-	expression tag	UNP A4WYU7
B	1	GLY	-	expression tag	UNP A4WYU7
B	2	SER	-	expression tag	UNP A4WYU7
B	3	SER	-	expression tag	UNP A4WYU7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	4	HIS	-	expression tag	UNP A4WYU7
B	5	HIS	-	expression tag	UNP A4WYU7
B	6	HIS	-	expression tag	UNP A4WYU7
B	7	HIS	-	expression tag	UNP A4WYU7
B	8	HIS	-	expression tag	UNP A4WYU7
B	9	HIS	-	expression tag	UNP A4WYU7
B	10	SER	-	expression tag	UNP A4WYU7
B	11	SER	-	expression tag	UNP A4WYU7
B	12	GLY	-	expression tag	UNP A4WYU7
B	13	LEU	-	expression tag	UNP A4WYU7
B	14	VAL	-	expression tag	UNP A4WYU7
B	15	PRO	-	expression tag	UNP A4WYU7
B	16	ALA	-	expression tag	UNP A4WYU7
B	17	GLY	-	expression tag	UNP A4WYU7
B	18	SER	-	expression tag	UNP A4WYU7
B	19	HIS	-	expression tag	UNP A4WYU7
B	20	MET	-	expression tag	UNP A4WYU7

- Molecule 2 is a RNA chain called RNA (5'-D(P*UP*UP*AP*CP*AP*AP*CP*CP*UP*AP *CP*UP*AP*CP*CP*UP*CP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	18	Total 374	C 168	N 61	O 127	P 18	0	0	0
2	E	18	Total 374	C 168	N 61	O 127	P 18	0	0	0

- Molecule 3 is a DNA chain called DNA (5'-D(*CP*GP*AP*GP*GP*TP*AP*GP*TP*AP* GP*GP*TP*TP*GP*TP*AP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	18	Total 375	C 179	N 73	O 106	P 17	0	0	0
3	F	18	Total 375	C 179	N 73	O 106	P 17	0	0	0

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	1	Total 1	Mg 1	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	E	1	Total 1	Mg 1	0	0

- Molecule 5 is water.

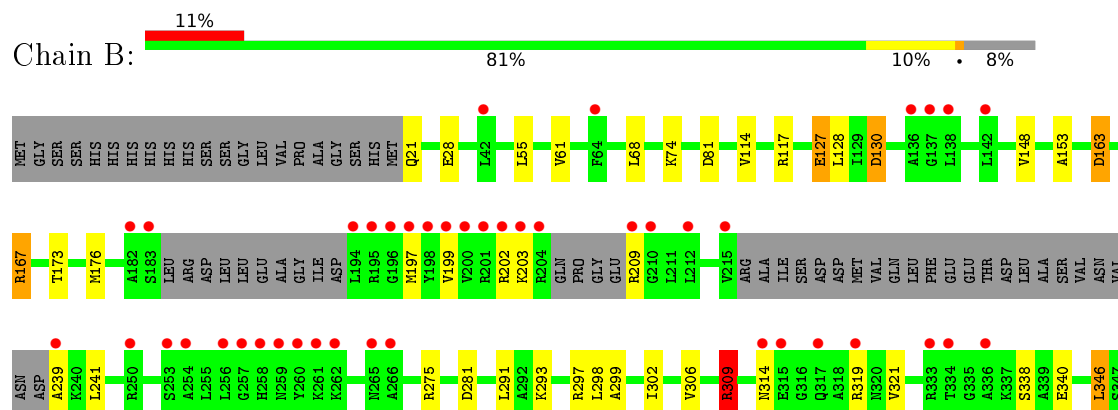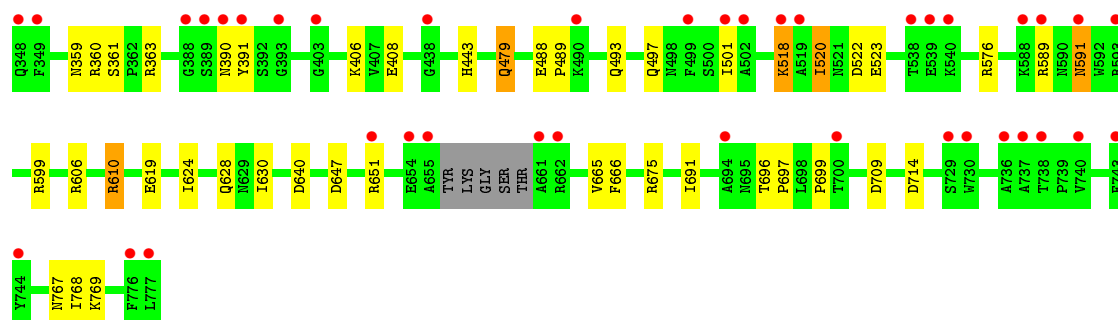| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	292	Total 292	O 292	0	0
5	B	182	Total 182	O 182	0	0
5	C	50	Total 50	O 50	0	0
5	D	57	Total 57	O 57	0	0
5	E	30	Total 30	O 30	0	0
5	F	25	Total 25	O 25	0	0

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: Uncharacterized protein

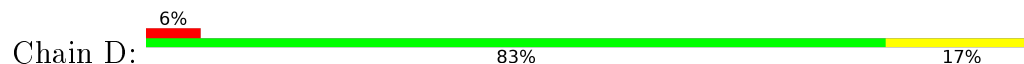- Molecule 1: Uncharacterized protein

Q348 F349 N359 R360 S361 P362 R363 G388 S389 N390 Y391 S392 G393 G403 K406 V407 E408 G438 H443 Q479 E488 P489 K490 Q493 Q497 N498 F499 S500 I501 A502 K518 A519 I520 N521 D522 E523 T538 E539 K540 R576 K588 K589 N590 N591 W592 R593

R599 R606 R610 E619 I624 Q628 N629 I630 D640 D647 R651 E654 A655 TYR LYS GLY SER THR A661 R662 V665 F666 R675 I691 A694 N695 T696 P697 L698 P699 T700 D709 D714 S729 W730 A736 A737 T738 P739 V740 F743

Y744 N767 I768 K769 F776 L777

• Molecule 2: RNA (5'-D(P*UP*UP*AP*CP*AP*AP*CP*CP*UP*AP*CP*UP*AP*CP*CP*U P*CP*G)-3')

Chain C: [bar: 44% green, 56% yellow]

U1 U2 A3 C4 A5 A6 C7 C8 C11 U12 A13 G18

• Molecule 2: RNA (5'-D(P*UP*UP*AP*CP*AP*AP*CP*CP*UP*AP*CP*UP*AP*CP*CP*U P*CP*G)-3')

Chain E: [bar: 67% green, 22% yellow, 6% orange, 6% red]

U1 C8 U9 A10 C11 U12 A13 G18

• Molecule 3: DNA (5'-D(*CP*GP*AP*GP*GP*TP*AP*GP*TP*AP*GP*GP*TP*TP*GP*TP *AP*A)-3')

Chain D: [bar: 6% red, 83% green, 17% yellow]

C1 G4 A7 G8 T9 A18

• Molecule 3: DNA (5'-D(*CP*GP*AP*GP*GP*TP*AP*GP*TP*AP*GP*GP*TP*TP*GP*TP *AP*A)-3')

Chain F: [bar: 6% red, 78% green, 17% yellow, 6% orange]

C1 A7 G8 T9 A10 A18

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	68.19Å 118.30Å 118.48Å 90.00° 95.83° 90.00°	Depositor
Resolution (Å)	35.63 – 2.00 35.60 – 1.99	Depositor EDS
% Data completeness (in resolution range)	97.1 (35.63-2.00) 97.1 (35.60-1.99)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.56 (at 2.00Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.184 , 0.233 0.194 , 0.242	Depositor DCC
R_{free} test set	6190 reflections (5.29%)	DCC
Wilson B-factor (Å^2)	39.9	Xtriage
Anisotropy	0.074	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.40 , 65.6	EDS
L-test for twinning[2]	$< |L| > = 0.48, < L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	13805	wwPDB-VP
Average B, all atoms (Å^2)	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.22% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: MG

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	0.89	2/6136 (0.0%)	0.95	14/8305 (0.2%)
1	B	0.75	0/5787	0.89	9/7826 (0.1%)
2	C	0.98	1/415 (0.2%)	1.02	1/640 (0.2%)
2	E	0.88	1/415 (0.2%)	1.03	2/640 (0.3%)
3	D	0.80	1/422 (0.2%)	1.00	1/652 (0.2%)
3	F	0.62	0/422	0.94	1/652 (0.2%)
All	All	0.82	5/13597 (0.0%)	0.93	28/18715 (0.1%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	1	U	OP3-P	-11.11	1.47	1.61
2	C	1	U	OP3-P	-7.72	1.51	1.61
1	A	615	SER	CB-OG	-7.14	1.32	1.42
3	D	9	DT	O3'-P	-5.79	1.54	1.61
1	A	567	GLU	CG-CD	5.56	1.60	1.51

The worst 5 of 28 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	13	A	O5'-P-OP2	-11.31	95.52	105.70
1	B	610	ARG	NE-CZ-NH1	8.08	124.34	120.30
1	B	610	ARG	NE-CZ-NH2	-7.70	116.45	120.30
1	A	599	ARG	NE-CZ-NH2	-7.46	116.57	120.30
1	A	567	GLU	OE1-CD-OE2	-7.36	114.46	123.30

There are no chirality outliers.

There are no planarity outliers.

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	A	6005	0	6039	83	0
1	B	5664	0	5713	53	0
2	C	374	0	194	5	0
2	E	374	0	194	3	0
3	D	375	0	205	1	0
3	F	375	0	205	3	0
4	C	1	0	0	0	0
4	E	1	0	0	0	0
5	A	292	0	0	9	0
5	B	182	0	0	4	0
5	C	50	0	0	0	0
5	D	57	0	0	0	0
5	E	30	0	0	0	0
5	F	25	0	0	0	0
All	All	13805	0	12550	144	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 144 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:148:VAL:HG12	1:B:176:MET:CE	1.99	0.93
1:B:606:ARG:HG2	5:B:966:HOH:O	1.78	0.84
1:B:363:ARG:H	1:B:443:HIS:HD2	1.24	0.83
1:B:363:ARG:H	1:B:443:HIS:CD2	1.99	0.80
1:A:610:ARG:CZ	5:A:802:HOH:O	2.33	0.77

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	757/778 (97%)	728 (96%)	25 (3%)	4 (0%)	34	26
1	B	705/778 (91%)	686 (97%)	17 (2%)	2 (0%)	46	41
All	All	1462/1556 (94%)	1414 (97%)	42 (3%)	6 (0%)	39	33

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	518	LYS
1	A	76	GLY
1	A	518	LYS
1	A	522	ASP
1	B	768	ILE

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	643/658 (98%)	605 (94%)	38 (6%)	24	18
1	B	605/658 (92%)	578 (96%)	27 (4%)	34	29
All	All	1248/1316 (95%)	1183 (95%)	65 (5%)	29	23

5 of 65 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	495	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	747	ARG
1	B	520	ILE
1	A	518	LYS
1	A	564	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	498	ASN
1	A	516	HIS
1	B	479	GLN
1	A	443	HIS
1	A	479	GLN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	C	17/18 (94%)	0	0
2	E	17/18 (94%)	2 (11%)	0
All	All	34/36 (94%)	2 (5%)	0

All (2) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	E	12	U
2	E	13	A

There are no RNA pucker outliers to report.

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 2 ligands modelled in this entry, 2 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, 95th 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(Å²)	Q<0.9
1	A	759/778 (97%)	0.18	50 (6%)	22	22	29, 51, 85, 110	0
1	B	715/778 (91%)	0.51	82 (11%)	6	7	34, 62, 99, 130	0
2	C	18/18 (100%)	-0.80	0	100	100	34, 41, 62, 80	0
2	E	18/18 (100%)	-0.72	0	100	100	43, 53, 70, 86	0
3	D	18/18 (100%)	-0.39	1 (5%)	28	29	39, 47, 74, 99	0
3	F	18/18 (100%)	-0.27	1 (5%)	28	29	41, 64, 74, 106	0
All	All	1546/1628 (94%)	0.30	134 (8%)	13	13	29, 56, 92, 130	0

The worst 5 of 134 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	389	SER	7.1
1	B	196	GLY	6.9
1	A	317	GLN	6.7
1	B	215	VAL	6.7
1	B	204	ARG	6.5

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, 95th 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
4	MG	E	101	1/1	0.98	0.16	-0.68	46,46,46,46	0
4	MG	C	101	1/1	0.98	0.15	-0.82	33,33,33,33	0

6.5 Other polymers (i)

There are no such residues in this entry.