wwPDB X-ray Structure Validation Summary Report (i)

Jun 16, 2024 – 11:16 AM EDT

PDB ID	:	4WSA
Title	:	Crystal structure of Influenza B polymerase bound to the vRNA promoter (FluB1 form)
Authors	:	Reich, S.; Guilligay, D.; Pflug, A.; Cusack, S.
Deposited on	:	2014-10-26
Resolution	:	3.40 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 (1)) were used in the production of this report:

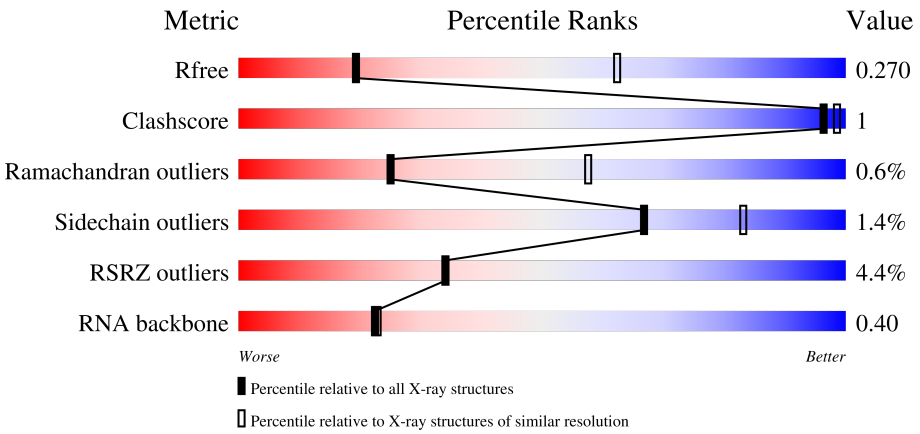
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 3.40 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}	130704	1026 (3.48-3.32)
Clashscore	141614	1055 (3.48-3.32)
Ramachandran outliers	138981	1038 (3.48-3.32)
Sidechain outliers	138945	1038 (3.48-3.32)
RSRZ outliers	127900	2173 (3.50-3.30)
RNA backbone	3102	1006 (3.84-2.96)

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% 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	R	18	50% 28% 22%
2	V	14	79% 21%
3	A	751	4% 91% · 5%
4	B	772	% 92% · 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	C	798	7% 86% 6% 9%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 17940 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 RNA chain called Influenza B vRNA promoter 3' end.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	R	14	282	128	42	99	13	0	0	0

- Molecule 2 is a RNA chain called Influenza B vRNA promoter 5' end.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	V	14	307	137	62	94	14	0	0	0

- Molecule 3 is a protein called PA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	716	5747	3651	962	1094	40	0	0	0

There are 25 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-13	GLY	-	expression tag	UNP Q5V8Z9
A	-12	SER	-	expression tag	UNP Q5V8Z9
A	-11	HIS	-	expression tag	UNP Q5V8Z9
A	-10	HIS	-	expression tag	UNP Q5V8Z9
A	-9	HIS	-	expression tag	UNP Q5V8Z9
A	-8	HIS	-	expression tag	UNP Q5V8Z9
A	-7	HIS	-	expression tag	UNP Q5V8Z9
A	-6	HIS	-	expression tag	UNP Q5V8Z9
A	-5	HIS	-	expression tag	UNP Q5V8Z9
A	-4	HIS	-	expression tag	UNP Q5V8Z9
A	-3	GLY	-	expression tag	UNP Q5V8Z9
A	-2	SER	-	expression tag	UNP Q5V8Z9
A	-1	GLY	-	expression tag	UNP Q5V8Z9
A	0	SER	-	expression tag	UNP Q5V8Z9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	727	GLY	-	expression tag	UNP Q5V8Z9
A	728	SER	-	expression tag	UNP Q5V8Z9
A	729	GLY	-	expression tag	UNP Q5V8Z9
A	730	SER	-	expression tag	UNP Q5V8Z9
A	731	GLY	-	expression tag	UNP Q5V8Z9
A	732	GLU	-	expression tag	UNP Q5V8Z9
A	733	ASN	-	expression tag	UNP Q5V8Z9
A	734	LEU	-	expression tag	UNP Q5V8Z9
A	735	TYR	-	expression tag	UNP Q5V8Z9
A	736	PHE	-	expression tag	UNP Q5V8Z9
A	737	GLN	-	expression tag	UNP Q5V8Z9

- Molecule 4 is a protein called RNA-directed RNA polymerase catalytic subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	B	733	5761	3636	999	1074	52	0	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-8	GLY	-	expression tag	UNP Q5V8Y6
B	-7	SER	-	expression tag	UNP Q5V8Y6
B	-6	GLY	-	expression tag	UNP Q5V8Y6
B	-5	SER	-	expression tag	UNP Q5V8Y6
B	-4	GLY	-	expression tag	UNP Q5V8Y6
B	-3	SER	-	expression tag	UNP Q5V8Y6
B	-2	GLY	-	expression tag	UNP Q5V8Y6
B	-1	SER	-	expression tag	UNP Q5V8Y6
B	0	GLY	-	expression tag	UNP Q5V8Y6
B	753	GLY	-	expression tag	UNP Q5V8Y6
B	754	SER	-	expression tag	UNP Q5V8Y6
B	755	GLY	-	expression tag	UNP Q5V8Y6
B	756	SER	-	expression tag	UNP Q5V8Y6
B	757	GLY	-	expression tag	UNP Q5V8Y6
B	758	GLU	-	expression tag	UNP Q5V8Y6
B	759	ASN	-	expression tag	UNP Q5V8Y6
B	760	LEU	-	expression tag	UNP Q5V8Y6
B	761	TYR	-	expression tag	UNP Q5V8Y6
B	762	PHE	-	expression tag	UNP Q5V8Y6
B	763	GLN	-	expression tag	UNP Q5V8Y6

- Molecule 5 is a protein called PB2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	C	730	5843	3715	1023	1065	40	0	0	0

There are 28 discrepancies between the modelled and reference sequences:

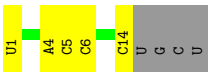
Chain	Residue	Modelled	Actual	Comment	Reference
C	-8	GLY	-	expression tag	UNP Q5V8X3
C	-7	SER	-	expression tag	UNP Q5V8X3
C	-6	GLY	-	expression tag	UNP Q5V8X3
C	-5	SER	-	expression tag	UNP Q5V8X3
C	-4	GLY	-	expression tag	UNP Q5V8X3
C	-3	SER	-	expression tag	UNP Q5V8X3
C	-2	GLY	-	expression tag	UNP Q5V8X3
C	-1	SER	-	expression tag	UNP Q5V8X3
C	0	GLY	-	expression tag	UNP Q5V8X3
C	771	GLY	-	expression tag	UNP Q5V8X3
C	772	TRP	-	expression tag	UNP Q5V8X3
C	773	SER	-	expression tag	UNP Q5V8X3
C	774	HIS	-	expression tag	UNP Q5V8X3
C	775	PRO	-	expression tag	UNP Q5V8X3
C	776	GLN	-	expression tag	UNP Q5V8X3
C	777	PHE	-	expression tag	UNP Q5V8X3
C	778	GLU	-	expression tag	UNP Q5V8X3
C	779	LYS	-	expression tag	UNP Q5V8X3
C	780	GLY	-	expression tag	UNP Q5V8X3
C	781	SER	-	expression tag	UNP Q5V8X3
C	782	GLY	-	expression tag	UNP Q5V8X3
C	783	SER	-	expression tag	UNP Q5V8X3
C	784	GLU	-	expression tag	UNP Q5V8X3
C	785	ASN	-	expression tag	UNP Q5V8X3
C	786	LEU	-	expression tag	UNP Q5V8X3
C	787	TYR	-	expression tag	UNP Q5V8X3
C	788	PHE	-	expression tag	UNP Q5V8X3
C	789	GLN	-	expression tag	UNP Q5V8X3

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 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: Influenza B vRNA promoter 3' end

Chain R:

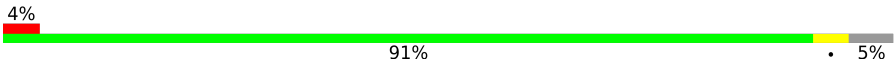


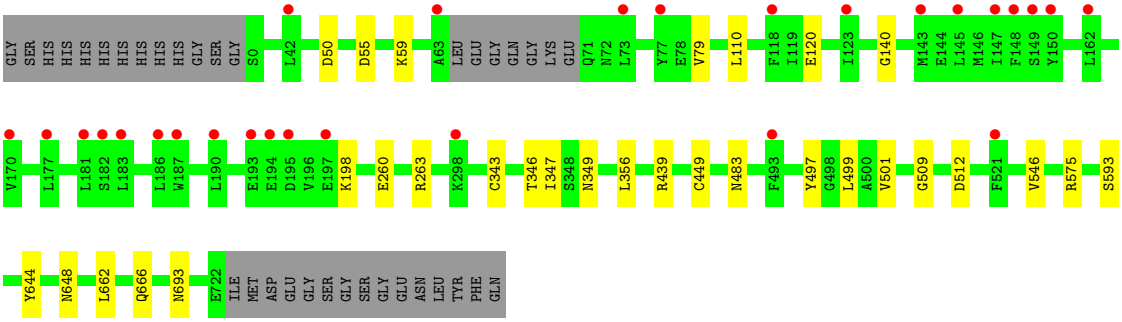
- Molecule 2: Influenza B vRNA promoter 5' end

Chain V:

- Molecule 3: PA

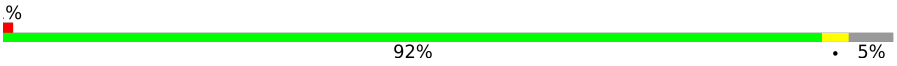
Chain A:

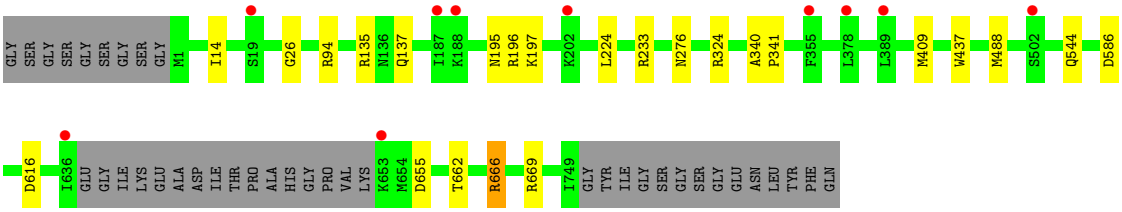




- Molecule 4: RNA-directed RNA polymerase catalytic subunit

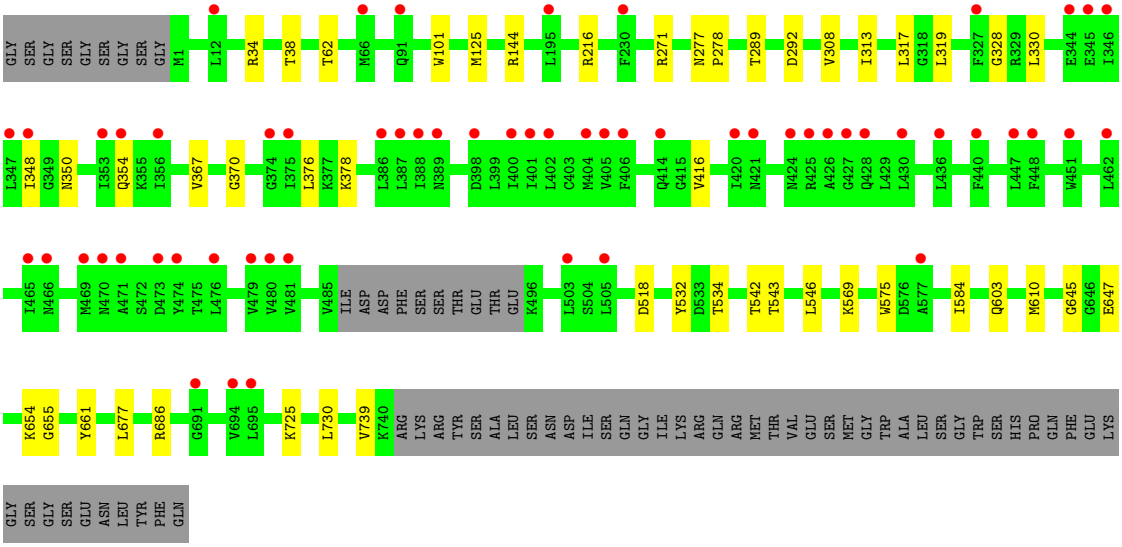
Chain B:





- Molecule 5: PB2

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	199.70Å 199.70Å 252.68Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.98 – 3.40 48.98 – 3.40	Depositor EDS
% Data completeness (in resolution range)	98.5 (48.98-3.40) 98.6 (48.98-3.40)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.71 (at 3.40Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1760)	Depositor
R, R_{free}	0.229 , 0.265 0.234 , 0.270	Depositor DCC
R_{free} test set	3821 reflections (4.82%)	wwPDB-VP
Wilson B-factor (Å^2)	129.1	Xtriage
Anisotropy	0.639	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 97.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.047 for -h,-k,l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	17940	wwPDB-VP
Average B, all atoms (Å^2)	158.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.50% 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)

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	R	0.26	0/312	0.81	0/482
2	V	0.62	1/345 (0.3%)	0.70	0/535
3	A	0.23	0/5862	0.38	0/7904
4	B	0.23	0/5873	0.39	0/7915
5	C	0.23	0/5943	0.40	0/7987
All	All	0.24	1/18335 (0.0%)	0.41	0/24823

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	V	1	A	OP3-P	-10.62	1.48	1.61

There are no bond angle outliers.

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	R	282	0	146	0	0
2	V	307	0	153	0	0
3	A	5747	0	5718	13	0
4	B	5761	0	5783	14	0
5	C	5843	0	6015	19	0
All	All	17940	0	17815	37	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 1.

The worst 5 of 37 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)
4:B:195:ASN:O	4:B:197:LYS:N	2.32	0.62
3:A:648:ASN:OD1	3:A:693:ASN:ND2	2.40	0.55
3:A:346:THR:O	3:A:349:ASN:ND2	2.40	0.54
3:A:509:GLY:N	3:A:512:ASP:OD2	2.42	0.52
4:B:669:ARG:NH1	5:C:38:THR:OG1	2.44	0.52

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	
3	A	710/751 (94%)	666 (94%)	41 (6%)	3 (0%)	34	67
4	B	729/772 (94%)	688 (94%)	40 (6%)	1 (0%)	51	82
5	C	726/798 (91%)	676 (93%)	42 (6%)	8 (1%)	14	44
All	All	2165/2321 (93%)	2030 (94%)	123 (6%)	12 (1%)	25	57

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	B	196	ARG
5	C	292	ASP
5	C	569	LYS
5	C	730	LEU
3	A	50	ASP

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	
3	A	637/664 (96%)	629 (99%)	8 (1%)	69	84
4	B	631/657 (96%)	624 (99%)	7 (1%)	73	86
5	C	638/694 (92%)	627 (98%)	11 (2%)	60	80
All	All	1906/2015 (95%)	1880 (99%)	26 (1%)	67	83

5 of 26 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	B	666	ARG
5	C	271	ARG
5	C	610	MET
5	C	125	MET
5	C	319	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
3	A	666	GLN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	R	13/18 (72%)	4 (30%)	1 (7%)
2	V	13/14 (92%)	2 (15%)	0
All	All	26/32 (81%)	6 (23%)	1 (3%)

5 of 6 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	R	4	A
1	R	5	C
1	R	6	C

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	R	14	C
2	V	7	A

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	R	1	U

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)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
3	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	164:GLU	C	165:GLU	N	3.00

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(Å^2)	Q<0.9
1	R	14/18 (77%)	-0.18	0	100	100	110, 129, 172, 209	0
2	V	14/14 (100%)	-0.28	0	100	100	99, 113, 127, 127	0
3	A	716/751 (95%)	0.24	28 (3%)	39	38	96, 151, 238, 271	0
4	B	733/772 (94%)	0.16	10 (1%)	75	74	101, 148, 178, 209	0
5	C	730/798 (91%)	0.42	59 (8%)	12	13	101, 157, 234, 262	0
All	All	2207/2353 (93%)	0.27	97 (4%)	34	34	96, 151, 230, 271	0

The worst 5 of 97 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	C	425	ARG	5.4
3	A	73	LEU	5.4
3	A	148	PHE	5.1
5	C	402	LEU	4.8
5	C	426	ALA	4.8

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 monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.