



# wwPDB X-ray Structure Validation Summary Report

Sep 6, 2023 – 06:24 PM EDT

PDB ID : 4GHA  
Title : Crystal structure of Marburg virus VP35 RNA binding domain bound to 12-bp dsRNA  
Authors : Bale, S.; Jean-Philippe, J.; Bornholdt, Z.A.; Kimberlin, C.K.; Halfmann, P.; Zandonatti, M.A.; Kunert, J.; Kroon, G.J.A.; Kawaoka, Y.; MacRae, I.J.; Wilson, I.A.; Saphire, E.O.  
Deposited on : 2012-08-07  
Resolution : 2.50 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the  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](#) ) were used in the production of this report:

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.35  
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) : 2.35

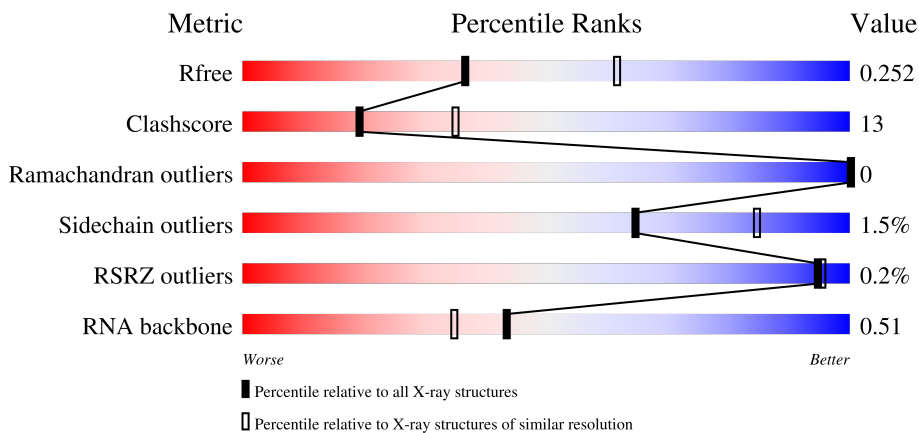
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)
RNA backbone	3102	1008 (2.84-2.16)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	146	 62% 21% 16%
1	C	146	 62% 22% 16%
1	E	146	 64% 20% 16%
1	G	146	 68% 14% 16%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	I	12	 58% 25% 17%
2	J	12	 75% 17% 8%

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4432 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 Polymerase cofactor VP35.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	122	943	607	166	168	2	0	0	0
1	C	122	943	607	166	168	2	0	0	0
1	E	122	943	607	166	168	2	0	0	0
1	G	122	943	607	166	168	2	0	0	0

There are 84 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	184	MET	-	expression tag	UNP P35259
A	185	ALA	-	expression tag	UNP P35259
A	186	HIS	-	expression tag	UNP P35259
A	187	HIS	-	expression tag	UNP P35259
A	188	HIS	-	expression tag	UNP P35259
A	189	HIS	-	expression tag	UNP P35259
A	190	HIS	-	expression tag	UNP P35259
A	191	HIS	-	expression tag	UNP P35259
A	192	VAL	-	expression tag	UNP P35259
A	193	ASP	-	expression tag	UNP P35259
A	194	ASP	-	expression tag	UNP P35259
A	195	ASP	-	expression tag	UNP P35259
A	196	ASP	-	expression tag	UNP P35259
A	197	LYS	-	expression tag	UNP P35259
A	198	GLU	-	expression tag	UNP P35259
A	199	ASN	-	expression tag	UNP P35259
A	200	LEU	-	expression tag	UNP P35259
A	201	TYR	-	expression tag	UNP P35259
A	202	PHE	-	expression tag	UNP P35259
A	203	GLN	-	expression tag	UNP P35259
A	296	CYS	SER	variant	UNP P35259

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	184	MET	-	expression tag	UNP P35259
C	185	ALA	-	expression tag	UNP P35259
C	186	HIS	-	expression tag	UNP P35259
C	187	HIS	-	expression tag	UNP P35259
C	188	HIS	-	expression tag	UNP P35259
C	189	HIS	-	expression tag	UNP P35259
C	190	HIS	-	expression tag	UNP P35259
C	191	HIS	-	expression tag	UNP P35259
C	192	VAL	-	expression tag	UNP P35259
C	193	ASP	-	expression tag	UNP P35259
C	194	ASP	-	expression tag	UNP P35259
C	195	ASP	-	expression tag	UNP P35259
C	196	ASP	-	expression tag	UNP P35259
C	197	LYS	-	expression tag	UNP P35259
C	198	GLU	-	expression tag	UNP P35259
C	199	ASN	-	expression tag	UNP P35259
C	200	LEU	-	expression tag	UNP P35259
C	201	TYR	-	expression tag	UNP P35259
C	202	PHE	-	expression tag	UNP P35259
C	203	GLN	-	expression tag	UNP P35259
C	296	CYS	SER	variant	UNP P35259
E	184	MET	-	expression tag	UNP P35259
E	185	ALA	-	expression tag	UNP P35259
E	186	HIS	-	expression tag	UNP P35259
E	187	HIS	-	expression tag	UNP P35259
E	188	HIS	-	expression tag	UNP P35259
E	189	HIS	-	expression tag	UNP P35259
E	190	HIS	-	expression tag	UNP P35259
E	191	HIS	-	expression tag	UNP P35259
E	192	VAL	-	expression tag	UNP P35259
E	193	ASP	-	expression tag	UNP P35259
E	194	ASP	-	expression tag	UNP P35259
E	195	ASP	-	expression tag	UNP P35259
E	196	ASP	-	expression tag	UNP P35259
E	197	LYS	-	expression tag	UNP P35259
E	198	GLU	-	expression tag	UNP P35259
E	199	ASN	-	expression tag	UNP P35259
E	200	LEU	-	expression tag	UNP P35259
E	201	TYR	-	expression tag	UNP P35259
E	202	PHE	-	expression tag	UNP P35259
E	203	GLN	-	expression tag	UNP P35259
E	296	CYS	SER	variant	UNP P35259

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	184	MET	-	expression tag	UNP P35259
G	185	ALA	-	expression tag	UNP P35259
G	186	HIS	-	expression tag	UNP P35259
G	187	HIS	-	expression tag	UNP P35259
G	188	HIS	-	expression tag	UNP P35259
G	189	HIS	-	expression tag	UNP P35259
G	190	HIS	-	expression tag	UNP P35259
G	191	HIS	-	expression tag	UNP P35259
G	192	VAL	-	expression tag	UNP P35259
G	193	ASP	-	expression tag	UNP P35259
G	194	ASP	-	expression tag	UNP P35259
G	195	ASP	-	expression tag	UNP P35259
G	196	ASP	-	expression tag	UNP P35259
G	197	LYS	-	expression tag	UNP P35259
G	198	GLU	-	expression tag	UNP P35259
G	199	ASN	-	expression tag	UNP P35259
G	200	LEU	-	expression tag	UNP P35259
G	201	TYR	-	expression tag	UNP P35259
G	202	PHE	-	expression tag	UNP P35259
G	203	GLN	-	expression tag	UNP P35259
G	296	CYS	SER	variant	UNP P35259

- Molecule 2 is a RNA chain called RNA (5'-R(\*CP\*UP\*AP\*GP\*AP\*CP\*GP\*UP\*CP\*UP\*AP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	I	12	Total	C	N	O	P	0	0	0
			252	114	45	82	11			
2	J	12	Total	C	N	O	P	0	0	0
			252	114	45	82	11			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	35	Total	O	0	0
			35	35		
3	C	31	Total	O	0	0
			31	31		
3	E	27	Total	O	0	0
			27	27		
3	G	32	Total	O	0	0
			32	32		

Continued on next page...

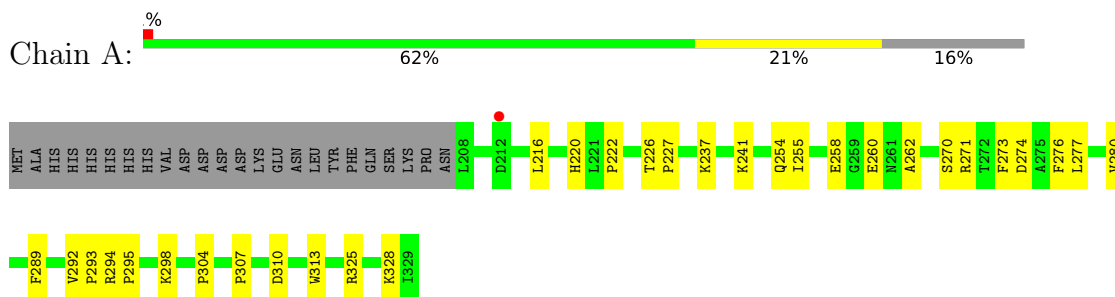
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
3	I	17	Total	O	0	0
			17	17		
3	J	14	Total	O	0	0
			14	14		

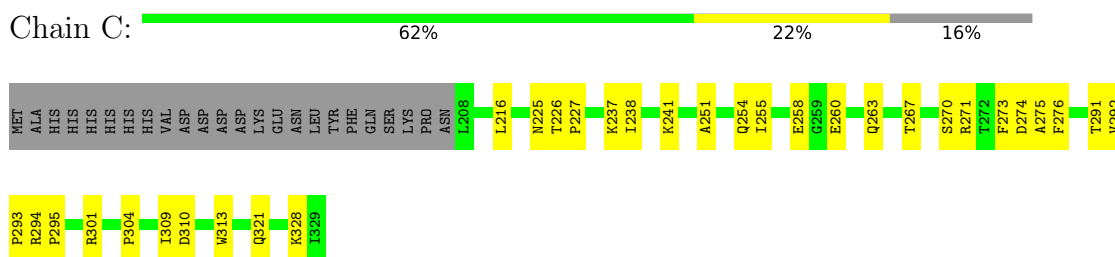
### 3 Residue-property plots [i](#)

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.

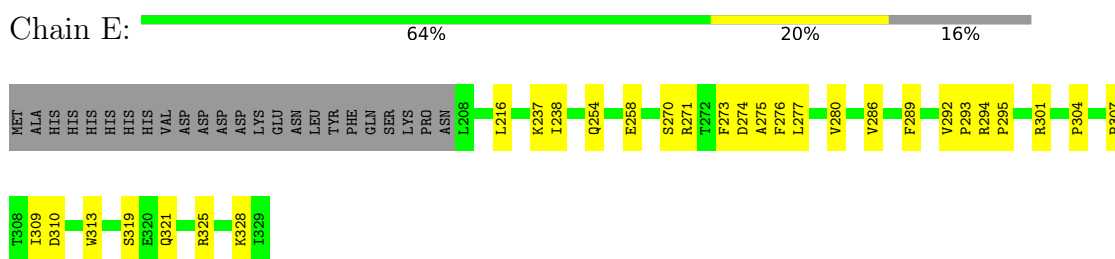
- Molecule 1: Polymerase cofactor VP35



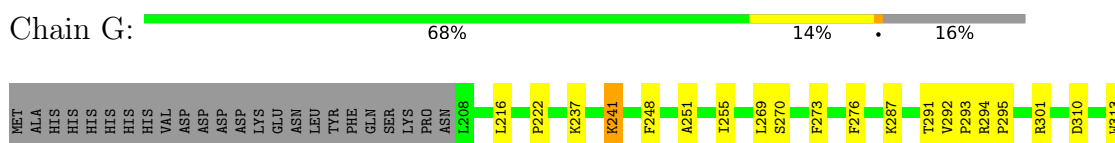
- Molecule 1: Polymerase cofactor VP35



- Molecule 1: Polymerase cofactor VP35



- Molecule 1: Polymerase cofactor VP35



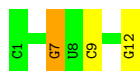




- Molecule 2: RNA (5'-R(\*CP\*UP\*AP\*GP\*AP\*CP\*GP\*UP\*CP\*UP\*AP\*G)-3')



- Molecule 2: RNA (5'-R(\*CP\*UP\*AP\*GP\*AP\*CP\*GP\*UP\*CP\*UP\*AP\*G)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	36.59Å 99.19Å 81.80Å 90.00° 89.92° 90.00°	Depositor
Resolution (Å)	81.80 – 2.50 49.60 – 2.50	Depositor EDS
% Data completeness (in resolution range)	98.7 (81.80-2.50) 98.6 (49.60-2.50)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	0.13	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.72 (at 2.51Å)	Xtrriage
Refinement program	REFMAC 5.5.0102	Depositor
R, $R_{free}$	0.209 , 0.250 0.213 , 0.252	Depositor DCC
$R_{free}$ test set	1017 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	29.3	Xtrriage
Anisotropy	0.363	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 1.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtrriage
Estimated twinning fraction	0.470 for h,-k,-l	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	4432	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	17.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.29% 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 [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	A	0.21	0/965	0.35	0/1310
1	C	0.22	0/965	0.36	0/1310
1	E	0.22	0/965	0.36	0/1310
1	G	0.22	0/965	0.36	0/1310
2	I	0.37	0/281	0.82	0/436
2	J	0.39	0/281	0.85	0/436
All	All	0.24	0/4422	0.46	0/6112

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 [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	943	0	971	23	0
1	C	943	0	971	26	0
1	E	943	0	971	24	0
1	G	943	0	971	30	0
2	I	252	0	131	10	0
2	J	252	0	131	4	0
3	A	35	0	0	1	0
3	C	31	0	0	4	0
3	E	27	0	0	6	0
3	G	32	0	0	6	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	I	17	0	0	0	0
3	J	14	0	0	0	0
All	All	4432	0	4146	107	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 13.

The worst 5 of 107 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:271:ARG:HD3	3:C:416:HOH:O	1.74	0.85
1:C:301:ARG:NE	3:C:402:HOH:O	2.22	0.73
1:A:237:LYS:O	1:A:241:LYS:HG3	1.89	0.72
2:I:2:U:H6	2:I:2:U:H5''	1.53	0.71
1:C:237:LYS:O	1:C:241:LYS:HG3	1.90	0.71

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	120/146 (82%)	119 (99%)	1 (1%)	0	100	100
1	C	120/146 (82%)	119 (99%)	1 (1%)	0	100	100
1	E	120/146 (82%)	119 (99%)	1 (1%)	0	100	100
1	G	120/146 (82%)	120 (100%)	0	0	100	100
All	All	480/584 (82%)	477 (99%)	3 (1%)	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	103/126 (82%)	102 (99%)	1 (1%)	76	90
1	C	103/126 (82%)	102 (99%)	1 (1%)	76	90
1	E	103/126 (82%)	101 (98%)	2 (2%)	57	80
1	G	103/126 (82%)	101 (98%)	2 (2%)	57	80
All	All	412/504 (82%)	406 (98%)	6 (2%)	65	85

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	280	VAL
1	G	241	LYS
1	G	321	GLN
1	C	321	GLN
1	A	280	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	321	GLN
1	C	224	ASN
1	C	321	GLN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	I	11/12 (91%)	3 (27%)	0
2	J	11/12 (91%)	2 (18%)	0
All	All	22/24 (91%)	5 (22%)	0

All (5) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	I	2	U
2	I	6	C
2	I	12	G
2	J	7	G
2	J	9	C

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 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](#)

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	122/146 (83%)	-0.24	1 (0%) 86   87	8, 15, 29, 37	0
1	C	122/146 (83%)	-0.28	0 100   100	9, 16, 28, 34	0
1	E	122/146 (83%)	-0.26	0 100   100	9, 15, 28, 34	0
1	G	122/146 (83%)	-0.26	0 100   100	9, 15, 27, 29	0
2	I	12/12 (100%)	-0.90	0 100   100	12, 16, 18, 18	0
2	J	12/12 (100%)	-0.92	0 100   100	14, 16, 19, 19	0
All	All	512/608 (84%)	-0.29	1 (0%) 95   95	8, 15, 28, 37	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	212	ASP	2.2

### 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.