



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 02:09 AM GMT

PDB ID : 2FS3  
Title : Bacteriophage HK97 K169Y Head I  
Authors : Gan, L.; Speir, J.A.; Conway, J.F.; Lander, G.; Cheng, N.; Firek, B.A.; Hendrix, R.W.; Duda, R.L.; Liljas, L.; Johnson, J.E.  
Deposited on : 2006-01-20  
Resolution : 4.20 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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
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) : trunk26865

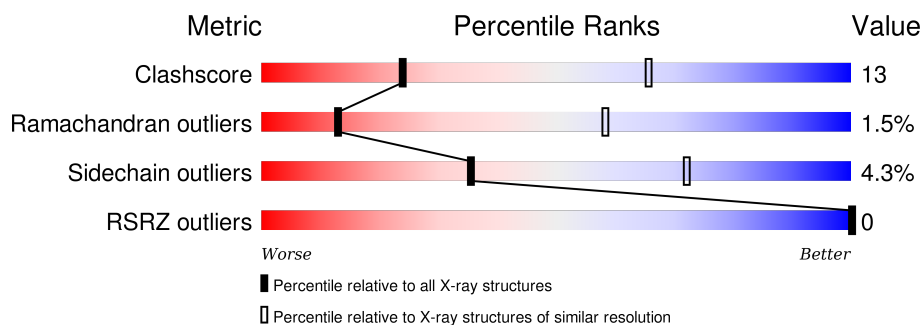
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 4.20 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))
Clashscore	102246	1140 (4.80-3.60)
Ramachandran outliers	100387	1083 (4.80-3.60)
Sidechain outliers	100360	1067 (4.80-3.60)
RSRZ outliers	91569	1042 (4.80-3.60)

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	282	 73% 24% ..
1	B	282	 74% 23% ..
1	C	282	 67% 28% . .
1	D	282	 68% 29% ..
1	E	282	 74% 23% ..
1	F	282	 64% 25% • 9%
1	G	282	 58% 24% • 14%

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 14664 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 Major capsid protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	280	Total	C	N	O	S	0	0	0
			2154	1347	374	423	10			
1	B	280	Total	C	N	O	S	0	0	0
			2154	1347	374	423	10			
1	C	280	Total	C	N	O	S	0	0	0
			2154	1347	374	423	10			
1	D	280	Total	C	N	O	S	0	0	0
			2154	1347	374	423	10			
1	E	280	Total	C	N	O	S	0	0	0
			2154	1347	374	423	10			
1	F	258	Total	C	N	O	S	0	0	0
			2004	1254	350	391	9			
1	G	243	Total	C	N	O	S	0	0	0
			1890	1181	333	368	8			

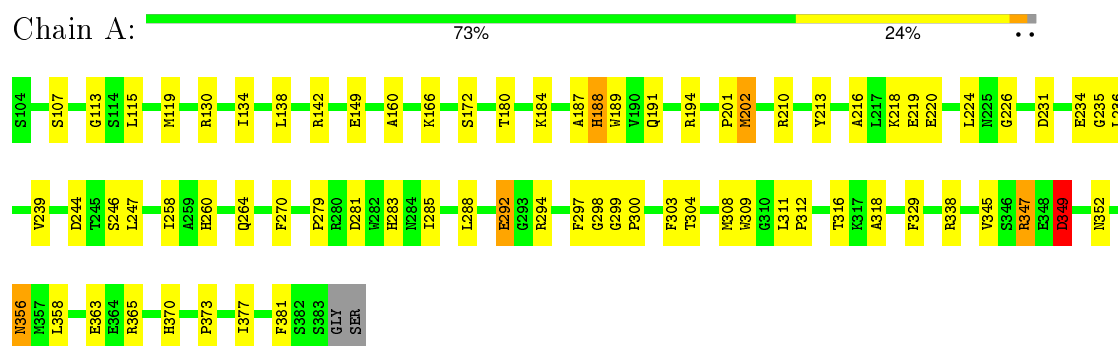
There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	169	TYR	LYS	ENGINEERED	UNP P49861
B	169	TYR	LYS	ENGINEERED	UNP P49861
C	169	TYR	LYS	ENGINEERED	UNP P49861
D	169	TYR	LYS	ENGINEERED	UNP P49861
E	169	TYR	LYS	ENGINEERED	UNP P49861
F	169	TYR	LYS	ENGINEERED	UNP P49861
G	169	TYR	LYS	ENGINEERED	UNP P49861

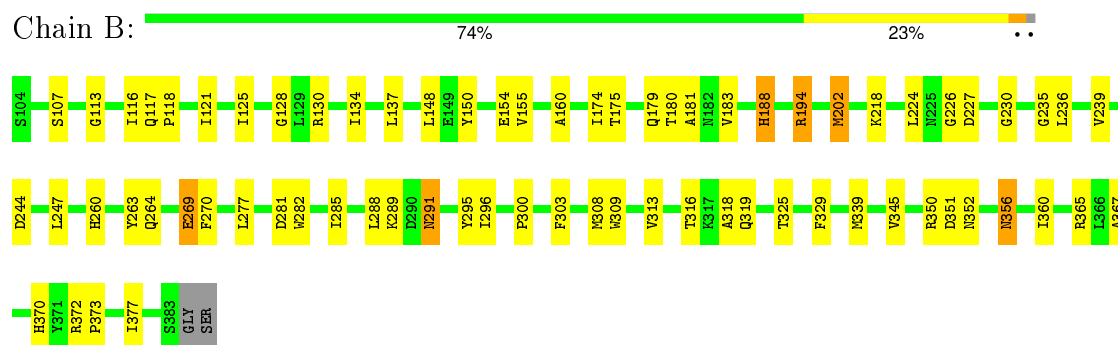
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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 ( $\text{RSRZ} > 2$ ). 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.

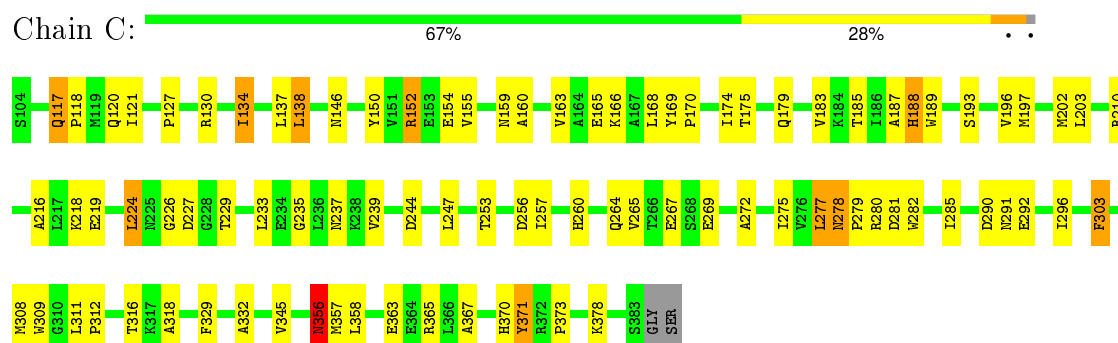
- Molecule 1: Major capsid protein



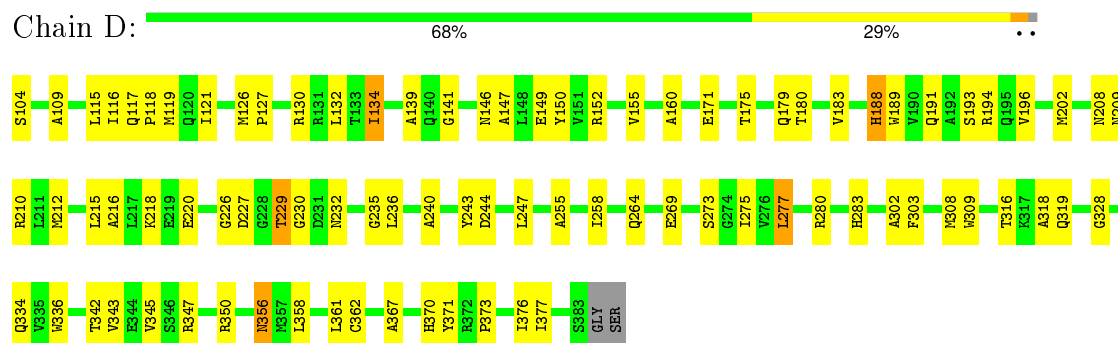
- Molecule 1: Major capsid protein



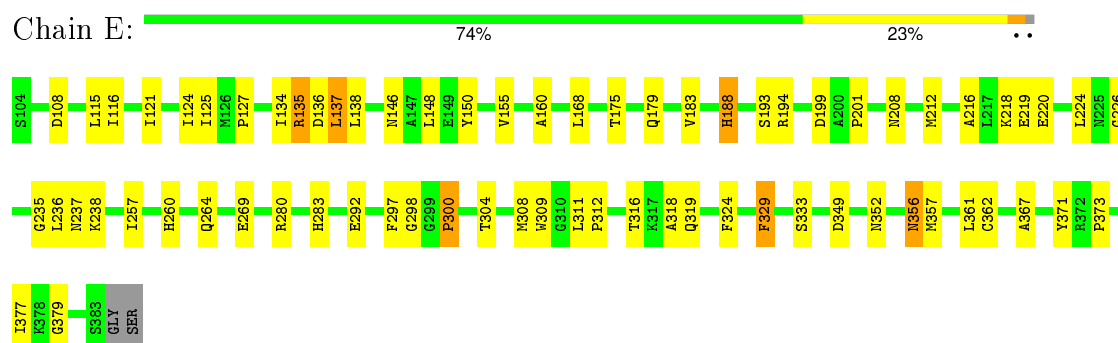
- Molecule 1: Major capsid protein



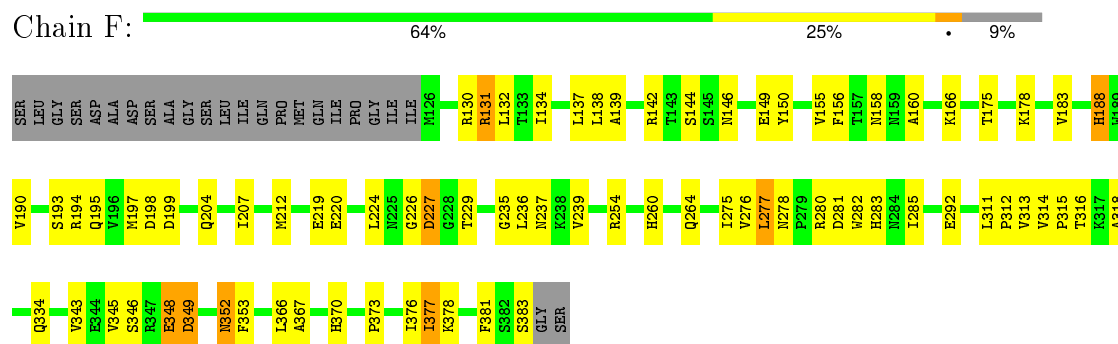
- Molecule 1: Major capsid protein



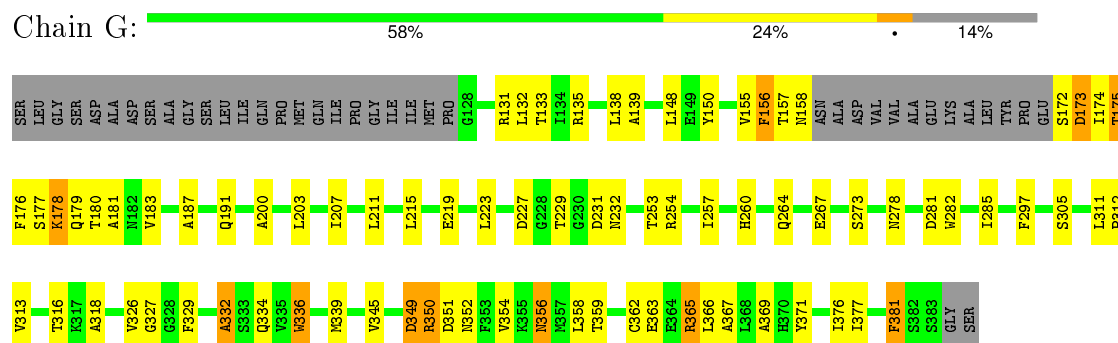
- Molecule 1: Major capsid protein



- Molecule 1: Major capsid protein



- Molecule 1: Major capsid protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	1009.64Å 1009.64Å 729.81Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 4.20 29.98 – 4.20	Depositor EDS
% Data completeness (in resolution range)	63.1 (30.00-4.20) 63.1 (29.98-4.20)	Depositor EDS
$R_{merge}$	0.16	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.95 (at 4.26Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.384 , (Not available) 0.328 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	20.0	Xtriage
Anisotropy	2.024	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.20 , -4.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.42$ , $\langle L^2 \rangle = 0.24$	Xtriage
Outliers	0 of 1678427 reflections	Xtriage
$F_o, F_c$ correlation	0.16	EDS
Total number of atoms	14664	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	80.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.10% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.25	0/2192	0.48	0/2976
1	B	0.25	0/2192	0.49	0/2976
1	C	0.25	0/2192	0.50	1/2976 (0.0%)
1	D	0.25	0/2192	0.48	0/2976
1	E	0.25	0/2192	0.50	0/2976
1	F	0.25	0/2040	0.49	0/2769
1	G	0.28	0/1922	0.51	0/2605
All	All	0.25	0/14922	0.49	1/20254 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	356	ASN	CB-CA-C	6.04	122.49	110.40

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	2154	0	2115	56	0
1	B	2154	0	2115	54	0
1	C	2154	0	2115	69	0
1	D	2154	0	2115	70	0
1	E	2154	0	2115	53	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	F	2004	0	1963	61	0
1	G	1890	0	1851	66	0
All	All	14664	0	14389	380	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 380 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:F:130:ARG:HD3	1:F:131:ARG:H	1.33	0.93
1:D:264:GLN:HE21	1:D:377:ILE:HD12	1.40	0.86
1:G:157:THR:HB	1:G:173:ASP:O	1.76	0.85
1:G:345:VAL:HG13	1:G:358:LEU:HD21	1.59	0.84
1:D:121:ILE:HD12	1:E:150:TYR:HB3	1.60	0.83

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	278/282 (99%)	241 (87%)	32 (12%)	5 (2%)	11	55
1	B	278/282 (99%)	238 (86%)	37 (13%)	3 (1%)	17	64
1	C	278/282 (99%)	248 (89%)	27 (10%)	3 (1%)	17	64
1	D	278/282 (99%)	247 (89%)	27 (10%)	4 (1%)	14	59
1	E	278/282 (99%)	244 (88%)	29 (10%)	5 (2%)	11	55
1	F	256/282 (91%)	228 (89%)	24 (9%)	4 (2%)	12	57
1	G	239/282 (85%)	204 (85%)	30 (13%)	5 (2%)	9	52
All	All	1885/1974 (96%)	1650 (88%)	206 (11%)	29 (2%)	13	58



5 of 29 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	201	PRO
1	E	300	PRO
1	F	352	ASN
1	A	202	MET
1	B	352	ASN

### 5.3.2 Protein sidechains ⓘ

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	230/231 (100%)	220 (96%)	10 (4%)	35	72
1	B	230/231 (100%)	221 (96%)	9 (4%)	39	74
1	C	230/231 (100%)	217 (94%)	13 (6%)	25	65
1	D	230/231 (100%)	224 (97%)	6 (3%)	54	81
1	E	230/231 (100%)	222 (96%)	8 (4%)	43	77
1	F	213/231 (92%)	202 (95%)	11 (5%)	29	68
1	G	201/231 (87%)	190 (94%)	11 (6%)	27	66
All	All	1564/1617 (97%)	1496 (96%)	68 (4%)	35	72

5 of 68 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	371	TYR
1	E	108	ASP
1	G	179	GLN
1	C	378	LYS
1	D	209	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 51 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	208	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	D	291	ASN
1	G	222	GLN
1	D	237	ASN
1	D	334	GLN

### 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](#)

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	280/282 (99%)	-0.45	0 100 100	8, 74, 127, 208	0
1	B	280/282 (99%)	-0.50	0 100 100	1, 68, 117, 152	0
1	C	280/282 (99%)	-0.49	0 100 100	1, 69, 114, 163	0
1	D	280/282 (99%)	-0.53	0 100 100	1, 69, 117, 187	0
1	E	280/282 (99%)	-0.48	0 100 100	8, 68, 119, 151	0
1	F	258/282 (91%)	-0.45	0 100 100	5, 72, 140, 226	0
1	G	243/282 (86%)	-0.26	0 100 100	41, 122, 182, 225	0
All	All	1901/1974 (96%)	-0.46	0 100 100	1, 74, 141, 226	0

There are no RSRZ outliers to report.

### 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](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.