



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jul 31, 2017 – 06:00 PM EDT

PDB ID : 5HSO  
Title : Crystal structure of MYCOBACTERIUM TUBERCULOSIS MARR FAMILY  
PROTEIN Rv2887 complex with DNA  
Authors : Gao, Y.R.; Li, D.F.; Wang, D.C.; Bi, L.J.  
Deposited on : unknown  
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

<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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20029824
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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-20029824

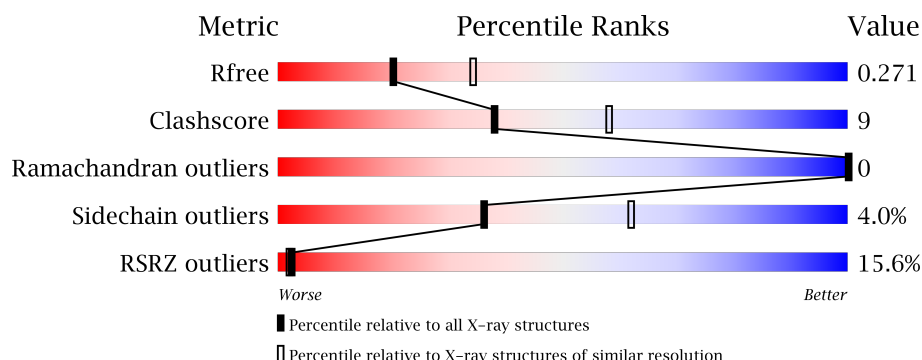
# 1 Overall quality at a glance

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))
$R_{free}$	100719	3846 (2.50-2.50)
Clashscore	112137	4554 (2.50-2.50)
Ramachandran outliers	110173	4463 (2.50-2.50)
Sidechain outliers	110143	4465 (2.50-2.50)
RSRZ outliers	101464	3876 (2.50-2.50)

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	160	<div> <div>16%</div> <div>62%</div> <div>12%</div> <div>•</div> <div>24%</div> </div>
1	B	160	<div> <div>12%</div> <div>59%</div> <div>16%</div> <div>•</div> <div>24%</div> </div>
1	C	160	<div> <div>13%</div> <div>61%</div> <div>16%</div> <div>•</div> <div>22%</div> </div>
1	D	160	<div> <div>12%</div> <div>64%</div> <div>11%</div> <div>•</div> <div>24%</div> </div>
2	E	30	<div> <div>63%</div> <div>37%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	F	30	<div><div></div><div>3%</div><div>57%</div><div>43%</div></div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4896 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 HTH-type transcriptional regulator Rv2887.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	122	Total	C	N	O	S	0	0	0
			910	571	172	163	4			
1	B	122	Total	C	N	O	S	0	0	0
			910	571	172	163	4			
1	C	125	Total	C	N	O	S	0	0	0
			936	588	178	166	4			
1	D	122	Total	C	N	O	S	0	0	0
			910	571	172	163	4			

There are 84 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-20	MET	-	initiating methionine	UNP P9WME9
A	-19	GLY	-	expression tag	UNP P9WME9
A	-18	SER	-	expression tag	UNP P9WME9
A	-17	SER	-	expression tag	UNP P9WME9
A	-16	HIS	-	expression tag	UNP P9WME9
A	-15	HIS	-	expression tag	UNP P9WME9
A	-14	HIS	-	expression tag	UNP P9WME9
A	-13	HIS	-	expression tag	UNP P9WME9
A	-12	HIS	-	expression tag	UNP P9WME9
A	-11	HIS	-	expression tag	UNP P9WME9
A	-10	SER	-	expression tag	UNP P9WME9
A	-9	SER	-	expression tag	UNP P9WME9
A	-8	GLY	-	expression tag	UNP P9WME9
A	-7	LEU	-	expression tag	UNP P9WME9
A	-6	VAL	-	expression tag	UNP P9WME9
A	-5	PRO	-	expression tag	UNP P9WME9
A	-4	ARG	-	expression tag	UNP P9WME9
A	-3	GLY	-	expression tag	UNP P9WME9
A	-2	SER	-	expression tag	UNP P9WME9
A	-1	HIS	-	expression tag	UNP P9WME9
A	0	MET	-	expression tag	UNP P9WME9

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	-20	MET	-	initiating methionine	UNP P9WME9
B	-19	GLY	-	expression tag	UNP P9WME9
B	-18	SER	-	expression tag	UNP P9WME9
B	-17	SER	-	expression tag	UNP P9WME9
B	-16	HIS	-	expression tag	UNP P9WME9
B	-15	HIS	-	expression tag	UNP P9WME9
B	-14	HIS	-	expression tag	UNP P9WME9
B	-13	HIS	-	expression tag	UNP P9WME9
B	-12	HIS	-	expression tag	UNP P9WME9
B	-11	HIS	-	expression tag	UNP P9WME9
B	-10	SER	-	expression tag	UNP P9WME9
B	-9	SER	-	expression tag	UNP P9WME9
B	-8	GLY	-	expression tag	UNP P9WME9
B	-7	LEU	-	expression tag	UNP P9WME9
B	-6	VAL	-	expression tag	UNP P9WME9
B	-5	PRO	-	expression tag	UNP P9WME9
B	-4	ARG	-	expression tag	UNP P9WME9
B	-3	GLY	-	expression tag	UNP P9WME9
B	-2	SER	-	expression tag	UNP P9WME9
B	-1	HIS	-	expression tag	UNP P9WME9
B	0	MET	-	expression tag	UNP P9WME9
C	-20	MET	-	initiating methionine	UNP P9WME9
C	-19	GLY	-	expression tag	UNP P9WME9
C	-18	SER	-	expression tag	UNP P9WME9
C	-17	SER	-	expression tag	UNP P9WME9
C	-16	HIS	-	expression tag	UNP P9WME9
C	-15	HIS	-	expression tag	UNP P9WME9
C	-14	HIS	-	expression tag	UNP P9WME9
C	-13	HIS	-	expression tag	UNP P9WME9
C	-12	HIS	-	expression tag	UNP P9WME9
C	-11	HIS	-	expression tag	UNP P9WME9
C	-10	SER	-	expression tag	UNP P9WME9
C	-9	SER	-	expression tag	UNP P9WME9
C	-8	GLY	-	expression tag	UNP P9WME9
C	-7	LEU	-	expression tag	UNP P9WME9
C	-6	VAL	-	expression tag	UNP P9WME9
C	-5	PRO	-	expression tag	UNP P9WME9
C	-4	ARG	-	expression tag	UNP P9WME9
C	-3	GLY	-	expression tag	UNP P9WME9
C	-2	SER	-	expression tag	UNP P9WME9
C	-1	HIS	-	expression tag	UNP P9WME9
C	0	MET	-	expression tag	UNP P9WME9

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	-20	MET	-	initiating methionine	UNP P9WME9
D	-19	GLY	-	expression tag	UNP P9WME9
D	-18	SER	-	expression tag	UNP P9WME9
D	-17	SER	-	expression tag	UNP P9WME9
D	-16	HIS	-	expression tag	UNP P9WME9
D	-15	HIS	-	expression tag	UNP P9WME9
D	-14	HIS	-	expression tag	UNP P9WME9
D	-13	HIS	-	expression tag	UNP P9WME9
D	-12	HIS	-	expression tag	UNP P9WME9
D	-11	HIS	-	expression tag	UNP P9WME9
D	-10	SER	-	expression tag	UNP P9WME9
D	-9	SER	-	expression tag	UNP P9WME9
D	-8	GLY	-	expression tag	UNP P9WME9
D	-7	LEU	-	expression tag	UNP P9WME9
D	-6	VAL	-	expression tag	UNP P9WME9
D	-5	PRO	-	expression tag	UNP P9WME9
D	-4	ARG	-	expression tag	UNP P9WME9
D	-3	GLY	-	expression tag	UNP P9WME9
D	-2	SER	-	expression tag	UNP P9WME9
D	-1	HIS	-	expression tag	UNP P9WME9
D	0	MET	-	expression tag	UNP P9WME9

- Molecule 2 is a DNA chain called DNA (30-MER), the upstream sequence of Rv0560c.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	30	Total	C	N	O	P	0	0	0
			618	291	117	180	30			

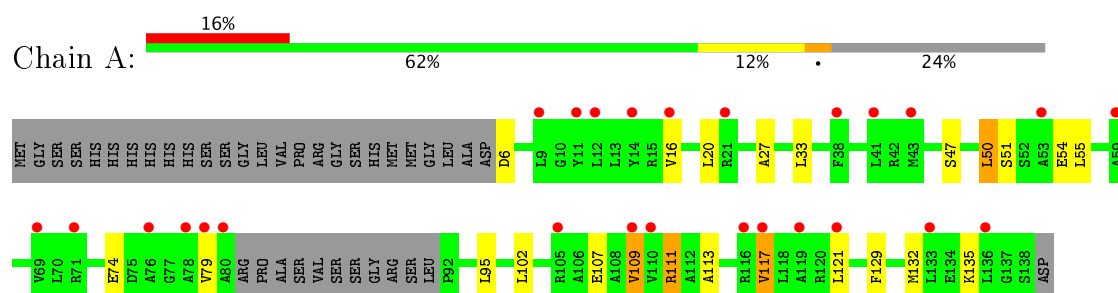
- Molecule 3 is a DNA chain called DNA (30-MER), the upstream sequence of Rv0560c.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	30	Total	C	N	O	P	0	0	0
			612	289	113	180	30			

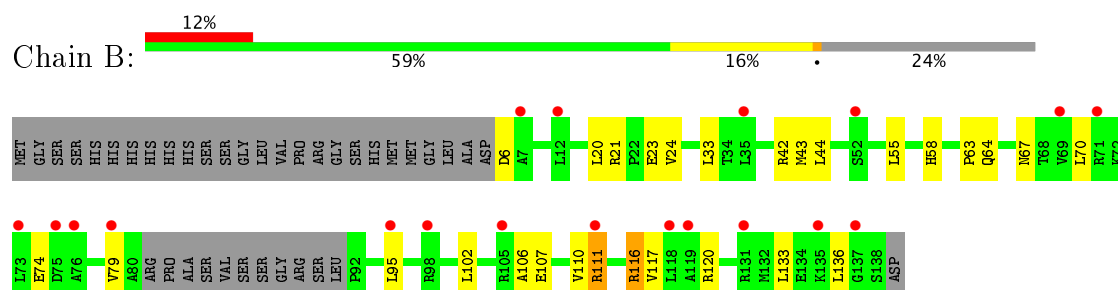
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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.

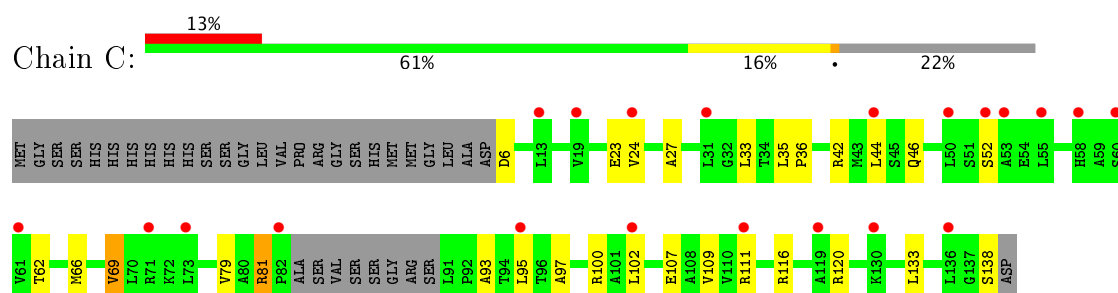
- Molecule 1: Uncharacterized HTH-type transcriptional regulator Rv2887



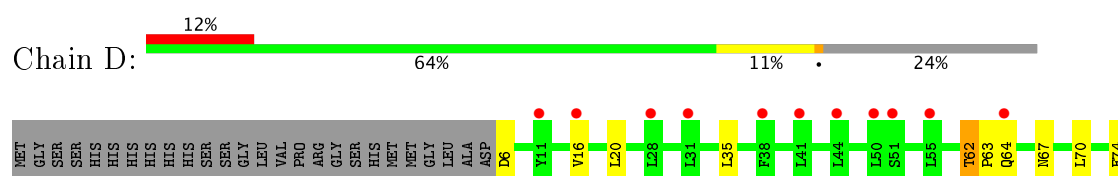
- Molecule 1: Uncharacterized HTH-type transcriptional regulator Rv2887

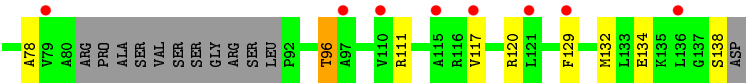


- Molecule 1: Uncharacterized HTH-type transcriptional regulator Rv2887

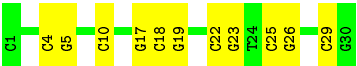


- Molecule 1: Uncharacterized HTH-type transcriptional regulator Rv2887

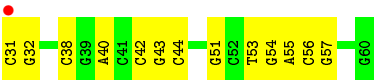




- Molecule 2: DNA (30-MER), the upstream sequence of Rv0560c



- Molecule 3: DNA (30-MER), the upstream sequence of Rv0560c





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	48.22Å 48.23Å 101.11Å 89.95° 90.04° 107.13°	Depositor
Resolution (Å)	36.21 – 2.50 46.09 – 2.50	Depositor EDS
% Data completeness (in resolution range)	98.1 (36.21-2.50) 92.1 (46.09-2.50)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	0.04	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.67 (at 2.51Å)	Xtriage
Refinement program	PHENIX dev_1980	Depositor
R, $R_{free}$	0.239 , 0.278 0.231 , 0.271	Depositor DCC
$R_{free}$ test set	1333 reflections (4.80%)	DCC
Wilson B-factor (Å <sup>2</sup> )	67.9	Xtriage
Anisotropy	0.507	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	(Not available) , (Not available)	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.456 for -h,-k,l 0.459 for -k,-h,-l 0.466 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	4896	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	89.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.59% 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

### 5.1 Standard geometry

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	0/921	0.49	0/1245
1	B	0.25	0/921	0.47	0/1245
1	C	0.28	0/948	0.52	0/1283
1	D	0.25	0/921	0.46	0/1245
2	E	0.52	0/693	0.86	0/1068
3	F	0.50	0/685	0.86	0/1054
All	All	0.34	0/5089	0.62	0/7140

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	910	0	966	18	0
1	B	910	0	966	18	0
1	C	936	0	996	21	0
1	D	910	0	966	13	0
2	E	618	0	336	8	0
3	F	612	0	336	11	0
All	All	4896	0	4566	76	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 76 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:107:GLU:OE2	1:A:111:ARG:NH1	2.22	0.73
1:C:52:SER:HB2	1:C:81:ARG:NH2	2.04	0.72
2:E:17:DG:H1	3:F:44:DC:H42	1.37	0.71
1:C:138:SER:OG	1:D:120:ARG:NH1	2.26	0.68
1:B:63:PRO:O	1:B:67:ASN:ND2	2.28	0.67

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	118/160 (74%)	118 (100%)	0	0	100	100
1	B	118/160 (74%)	117 (99%)	1 (1%)	0	100	100
1	C	121/160 (76%)	118 (98%)	3 (2%)	0	100	100
1	D	118/160 (74%)	117 (99%)	1 (1%)	0	100	100
All	All	475/640 (74%)	470 (99%)	5 (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	93/124 (75%)	89 (96%)	4 (4%)	33	58
1	B	93/124 (75%)	90 (97%)	3 (3%)	44	71
1	C	96/124 (77%)	93 (97%)	3 (3%)	45	73
1	D	93/124 (75%)	88 (95%)	5 (5%)	26	47
All	All	375/496 (76%)	360 (96%)	15 (4%)	36	62

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	116	ARG
1	C	24	VAL
1	D	96	THR
1	B	111	ARG
1	D	62	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

There are no ligands in this entry.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues ⓘ

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	122/160 (76%)	1.20	26 (21%) <b>1</b> <b>1</b>	57, 78, 101, 119	0
1	B	122/160 (76%)	0.94	19 (15%) <b>2</b> <b>2</b>	52, 80, 99, 118	0
1	C	125/160 (78%)	1.06	21 (16%) <b>2</b> <b>1</b>	56, 77, 110, 129	0
1	D	122/160 (76%)	0.92	19 (15%) <b>2</b> <b>2</b>	57, 78, 99, 114	0
2	E	30/30 (100%)	0.23	0 <b>100</b> <b>100</b>	77, 124, 144, 145	0
3	F	30/30 (100%)	0.11	1 (3%) 47 50	76, 117, 136, 152	0
All	All	551/700 (78%)	0.93	86 (15%) <b>2</b> <b>2</b>	52, 81, 122, 152	0

The worst 5 of 86 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	50	LEU	7.6
1	C	52	SER	7.0
1	A	78	ALA	6.2
1	D	11	TYR	6.0
1	B	105	ARG	5.0

### 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.