



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 11:50 AM GMT

PDB ID : 3Q63  
Title : X-ray crystal structure of protein MLL2253 from Mesorhizobium loti, Northeast Structural Genomics Consortium Target MIR404.  
Authors : Forouhar, F.; Lew, S.; Seetharaman, J.; Wang, D.; Ciccocanti, C.; Sahdev, S.; Nair, R.; Rost, B.; Acton, T.B.; Xiao, R.; Everett, J.K.; Montelione, G.T.; Hunt, J.F.; Tong, L.; Northeast Structural Genomics Consortium (NESG)  
Deposited on : 2010-12-30  
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](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

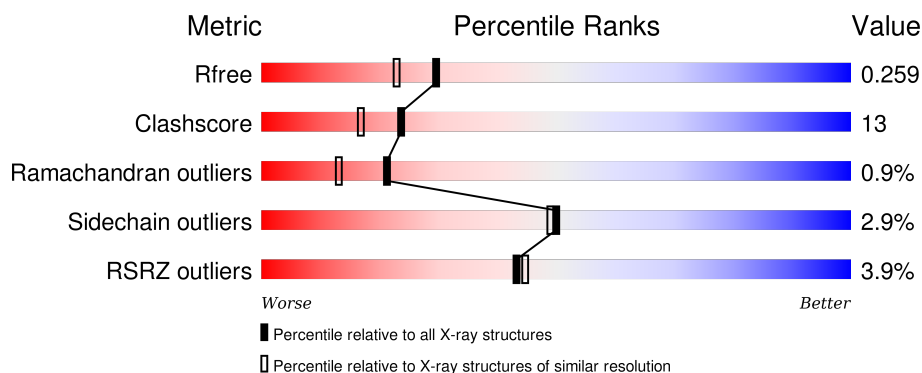
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 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	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)

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	151	<div> <div>68%</div> <div>17%</div> <div>•</div> <div>13%</div> </div>
1	B	151	<div>6%</div> <div>65%</div> <div>24%</div> <div>•</div> <div>9%</div>
1	C	151	<div>5%</div> <div>62%</div> <div>28%</div> <div>•</div> <div>9%</div>
1	D	151	<div>3%</div> <div>70%</div> <div>20%</div> <div>•</div> <div>10%</div>
1	E	151	<div>3%</div> <div>62%</div> <div>26%</div> <div>•</div> <div>11%</div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	151	<div><div><div>%</div><div><div></div><div>63%</div><div>27%</div><div>• 8%</div></div></div></div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 7357 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 Mll2253 protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	132	Total	C	N	O	S	Se	0	0	0
			1077	690	191	192	1	3			
1	B	137	Total	C	N	O	S	Se	0	0	0
			1119	714	200	201	1	3			
1	C	137	Total	C	N	O	S	Se	0	0	0
			1119	714	200	201	1	3			
1	D	136	Total	C	N	O	S	Se	0	0	0
			1111	710	198	199	1	3			
1	E	135	Total	C	N	O	S	Se	0	0	0
			1105	705	198	198	1	3			
1	F	139	Total	C	N	O	S	Se	0	0	0
			1135	723	202	206	1	3			

There are 54 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	EXPRESSION TAG	UNP Q98IT8
A	144	LEU	-	EXPRESSION TAG	UNP Q98IT8
A	145	GLU	-	EXPRESSION TAG	UNP Q98IT8
A	146	HIS	-	EXPRESSION TAG	UNP Q98IT8
A	147	HIS	-	EXPRESSION TAG	UNP Q98IT8
A	148	HIS	-	EXPRESSION TAG	UNP Q98IT8
A	149	HIS	-	EXPRESSION TAG	UNP Q98IT8
A	150	HIS	-	EXPRESSION TAG	UNP Q98IT8
A	151	HIS	-	EXPRESSION TAG	UNP Q98IT8
B	1	MSE	-	EXPRESSION TAG	UNP Q98IT8
B	144	LEU	-	EXPRESSION TAG	UNP Q98IT8
B	145	GLU	-	EXPRESSION TAG	UNP Q98IT8
B	146	HIS	-	EXPRESSION TAG	UNP Q98IT8
B	147	HIS	-	EXPRESSION TAG	UNP Q98IT8
B	148	HIS	-	EXPRESSION TAG	UNP Q98IT8
B	149	HIS	-	EXPRESSION TAG	UNP Q98IT8
B	150	HIS	-	EXPRESSION TAG	UNP Q98IT8

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	151	HIS	-	EXPRESSION TAG	UNP Q98IT8
C	1	MSE	-	EXPRESSION TAG	UNP Q98IT8
C	144	LEU	-	EXPRESSION TAG	UNP Q98IT8
C	145	GLU	-	EXPRESSION TAG	UNP Q98IT8
C	146	HIS	-	EXPRESSION TAG	UNP Q98IT8
C	147	HIS	-	EXPRESSION TAG	UNP Q98IT8
C	148	HIS	-	EXPRESSION TAG	UNP Q98IT8
C	149	HIS	-	EXPRESSION TAG	UNP Q98IT8
C	150	HIS	-	EXPRESSION TAG	UNP Q98IT8
C	151	HIS	-	EXPRESSION TAG	UNP Q98IT8
D	1	MSE	-	EXPRESSION TAG	UNP Q98IT8
D	144	LEU	-	EXPRESSION TAG	UNP Q98IT8
D	145	GLU	-	EXPRESSION TAG	UNP Q98IT8
D	146	HIS	-	EXPRESSION TAG	UNP Q98IT8
D	147	HIS	-	EXPRESSION TAG	UNP Q98IT8
D	148	HIS	-	EXPRESSION TAG	UNP Q98IT8
D	149	HIS	-	EXPRESSION TAG	UNP Q98IT8
D	150	HIS	-	EXPRESSION TAG	UNP Q98IT8
D	151	HIS	-	EXPRESSION TAG	UNP Q98IT8
E	1	MSE	-	EXPRESSION TAG	UNP Q98IT8
E	144	LEU	-	EXPRESSION TAG	UNP Q98IT8
E	145	GLU	-	EXPRESSION TAG	UNP Q98IT8
E	146	HIS	-	EXPRESSION TAG	UNP Q98IT8
E	147	HIS	-	EXPRESSION TAG	UNP Q98IT8
E	148	HIS	-	EXPRESSION TAG	UNP Q98IT8
E	149	HIS	-	EXPRESSION TAG	UNP Q98IT8
E	150	HIS	-	EXPRESSION TAG	UNP Q98IT8
E	151	HIS	-	EXPRESSION TAG	UNP Q98IT8
F	1	MSE	-	EXPRESSION TAG	UNP Q98IT8
F	144	LEU	-	EXPRESSION TAG	UNP Q98IT8
F	145	GLU	-	EXPRESSION TAG	UNP Q98IT8
F	146	HIS	-	EXPRESSION TAG	UNP Q98IT8
F	147	HIS	-	EXPRESSION TAG	UNP Q98IT8
F	148	HIS	-	EXPRESSION TAG	UNP Q98IT8
F	149	HIS	-	EXPRESSION TAG	UNP Q98IT8
F	150	HIS	-	EXPRESSION TAG	UNP Q98IT8
F	151	HIS	-	EXPRESSION TAG	UNP Q98IT8

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	118	Total O 118 118	0	0

*Continued on next page...*

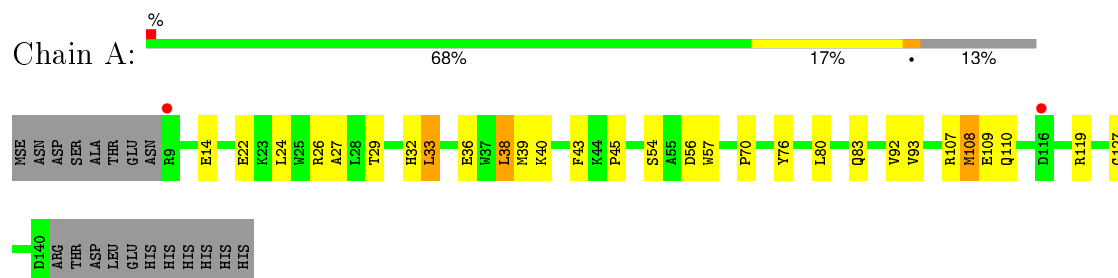
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	110	Total 110	O 110	0	0
2	C	108	Total 108	O 108	0	0
2	D	121	Total 121	O 121	0	0
2	E	103	Total 103	O 103	0	0
2	F	131	Total 131	O 131	0	0

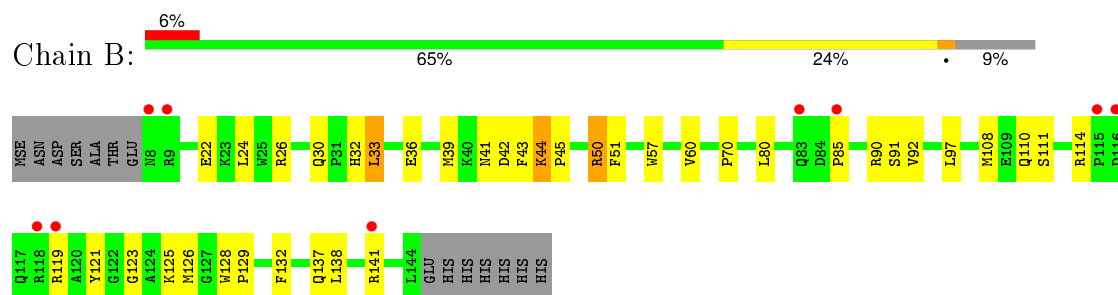
### 3 Residue-property plots [i](#)

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.

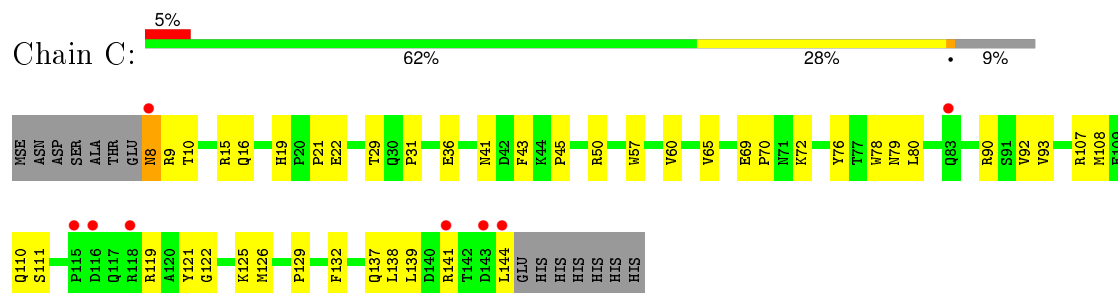
#### • Molecule 1: Mll2253 protein



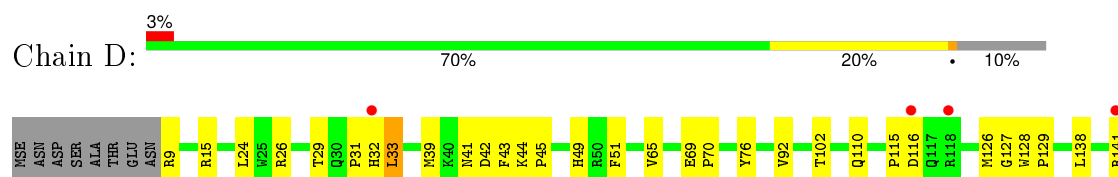
#### • Molecule 1: Mll2253 protein

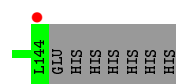


#### • Molecule 1: Mll2253 protein

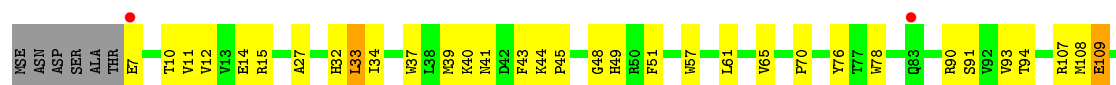


#### • Molecule 1: Mll2253 protein





- Molecule 1: Mll2253 protein



- Molecule 1: Mll2253 protein





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	40.12Å 148.12Å 87.39Å 90.00° 98.33° 90.00°	Depositor
Resolution (Å)	19.99 – 2.00 29.07 – 1.95	Depositor EDS
% Data completeness (in resolution range)	82.1 (19.99-2.00) 93.4 (29.07-1.95)	Depositor EDS
$R_{merge}$	0.15	Depositor
$R_{sym}$	0.14	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.58 (at 1.95Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, $R_{free}$	0.202 , 0.244 0.217 , 0.259	Depositor DCC
$R_{free}$ test set	6470 reflections (11.27%)	DCC
Wilson B-factor (Å <sup>2</sup> )	21.6	Xtriage
Anisotropy	0.821	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 49.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	1 of 135983 reflections (0.001%)	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	7357	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.95% 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	0.34	0/1107	0.62	0/1502
1	B	0.34	0/1149	0.58	0/1559
1	C	0.33	0/1149	0.58	0/1559
1	D	0.33	0/1141	0.61	0/1548
1	E	0.33	0/1135	0.57	0/1539
1	F	0.34	0/1165	0.58	0/1581
All	All	0.33	0/6846	0.59	0/9288

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	1077	0	1051	29	0
1	B	1119	0	1092	35	0
1	C	1119	0	1092	39	0
1	D	1111	0	1086	19	0
1	E	1105	0	1076	35	0
1	F	1135	0	1105	34	0
2	A	118	0	0	5	0
2	B	110	0	0	2	0
2	C	108	0	0	3	0
2	D	121	0	0	5	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	103	0	0	4	0
2	F	131	0	0	6	0
All	All	7357	0	6502	174	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 174 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:80:LEU:HD21	1:B:60:VAL:HG12	1.53	0.91
1:F:11:VAL:HG13	1:F:125:LYS:HD2	1.57	0.87
1:E:10:THR:HG21	2:E:536:HOH:O	1.81	0.81
1:C:41:ASN:HD22	1:C:43:PHE:H	1.28	0.80
1:C:126:MSE:HB2	2:C:527:HOH:O	1.83	0.78

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	130/151 (86%)	122 (94%)	7 (5%)	1 (1%)	24	15
1	B	135/151 (89%)	124 (92%)	10 (7%)	1 (1%)	26	19
1	C	135/151 (89%)	124 (92%)	10 (7%)	1 (1%)	26	19
1	D	134/151 (89%)	125 (93%)	8 (6%)	1 (1%)	26	19
1	E	133/151 (88%)	122 (92%)	9 (7%)	2 (2%)	13	5
1	F	137/151 (91%)	130 (95%)	6 (4%)	1 (1%)	26	19
All	All	804/906 (89%)	747 (93%)	50 (6%)	7 (1%)	21	13

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	70	PRO
1	E	70	PRO
1	E	109	GLU
1	A	70	PRO
1	F	70	PRO

### 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	116/130 (89%)	111 (96%)	5 (4%)	35	30
1	B	121/130 (93%)	117 (97%)	4 (3%)	45	43
1	C	121/130 (93%)	118 (98%)	3 (2%)	55	55
1	D	120/130 (92%)	115 (96%)	5 (4%)	36	31
1	E	119/130 (92%)	118 (99%)	1 (1%)	86	89
1	F	123/130 (95%)	120 (98%)	3 (2%)	57	58
All	All	720/780 (92%)	699 (97%)	21 (3%)	50	49

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	8	ASN
1	C	119	ARG
1	E	33	LEU
1	B	50	ARG
1	F	10	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 29 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	137	GLN
1	D	71	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	F	41	ASN
1	D	41	ASN
1	E	8	ASN

### 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	129/151 (85%)	0.07	2 (1%) 74 75	13, 21, 31, 40	0
1	B	134/151 (88%)	0.27	9 (6%) 21 22	14, 22, 43, 58	0
1	C	134/151 (88%)	0.31	8 (5%) 25 27	14, 23, 37, 44	0
1	D	133/151 (88%)	0.16	5 (3%) 44 45	14, 21, 34, 41	0
1	E	132/151 (87%)	0.21	5 (3%) 44 45	15, 24, 36, 47	0
1	F	136/151 (90%)	0.12	2 (1%) 76 77	13, 21, 35, 45	0
All	All	798/906 (88%)	0.19	31 (3%) 43 45	13, 22, 36, 58	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	6	THR	5.4
1	B	118	ARG	4.9
1	E	119	ARG	4.7
1	C	141	ARG	4.1
1	B	115	PRO	3.5

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.