Full wwPDB X-ray Structure Validation Report (i)

Feb 1, 2016 – 01:04 PM GMT

PDB ID	:	3SJR
Title	:	Crystal structure of conserved unkown function protein CV_1783 from Chromobacterium violaceum ATCC 12472
Authors	:	Chang, C.; Hatzos-Skintges, C.; Clancy, S.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2011-06-21
Resolution	:	2.94 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

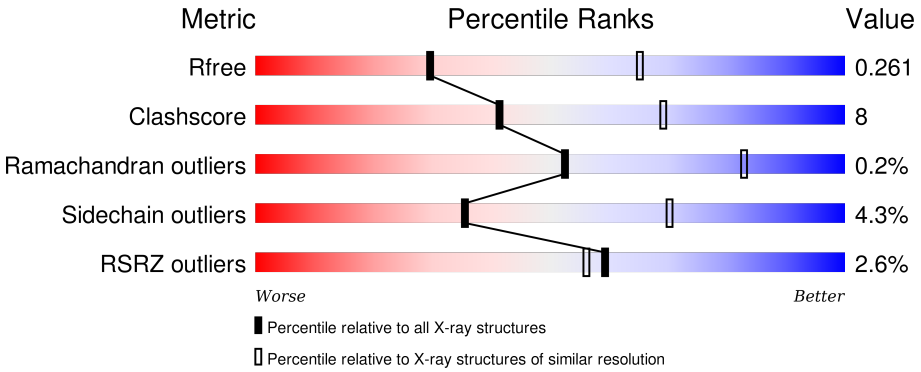
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 2.94 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	2057 (2.98-2.90)
Clashscore	102246	2308 (2.98-2.90)
Ramachandran outliers	100387	2245 (2.98-2.90)
Sidechain outliers	100360	2247 (2.98-2.90)
RSRZ outliers	91569	2065 (2.98-2.90)

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 >=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 <=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	175	2% 60% 11% • 28%
1	B	175	% 58% 7% • 33%
1	C	175	2% 43% 22% • 33%
1	D	175	2% 50% 13% • 37%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3652 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 protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	126	Total	C	N	O	S	Se	0	0	0
			974	615	175	180	1	3			
1	B	117	Total	C	N	O	S	Se	0	0	0
			904	570	163	167	1	3			
1	C	117	Total	C	N	O	S	Se	0	0	0
			909	572	163	170	1	3			
1	D	110	Total	C	N	O	S	Se	0	0	0
			861	543	156	158	1	3			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	EXPRESSION TAG	UNP Q7NX46
A	-1	ASN	-	EXPRESSION TAG	UNP Q7NX46
A	0	ALA	-	EXPRESSION TAG	UNP Q7NX46
B	-2	SER	-	EXPRESSION TAG	UNP Q7NX46
B	-1	ASN	-	EXPRESSION TAG	UNP Q7NX46
B	0	ALA	-	EXPRESSION TAG	UNP Q7NX46
C	-2	SER	-	EXPRESSION TAG	UNP Q7NX46
C	-1	ASN	-	EXPRESSION TAG	UNP Q7NX46
C	0	ALA	-	EXPRESSION TAG	UNP Q7NX46
D	-2	SER	-	EXPRESSION TAG	UNP Q7NX46
D	-1	ASN	-	EXPRESSION TAG	UNP Q7NX46
D	0	ALA	-	EXPRESSION TAG	UNP Q7NX46

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	O	0	0
			1	1		
2	B	1	Total	O	0	0
			1	1		

Continued on next page...

Continued from previous page...

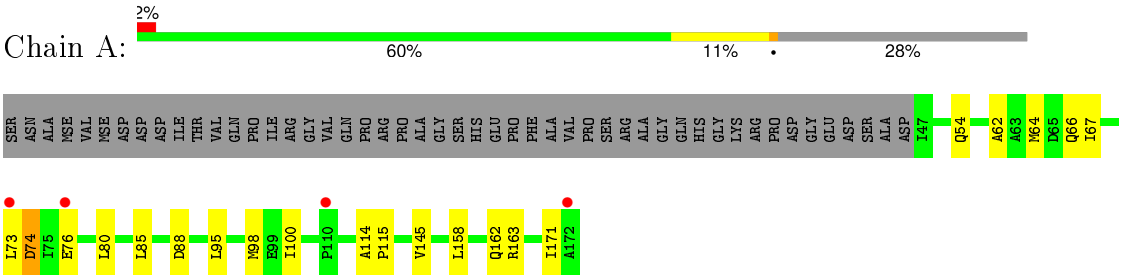
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	C	1	Total O 1 1	0	0
2	D	1	Total O 1 1	0	0

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.

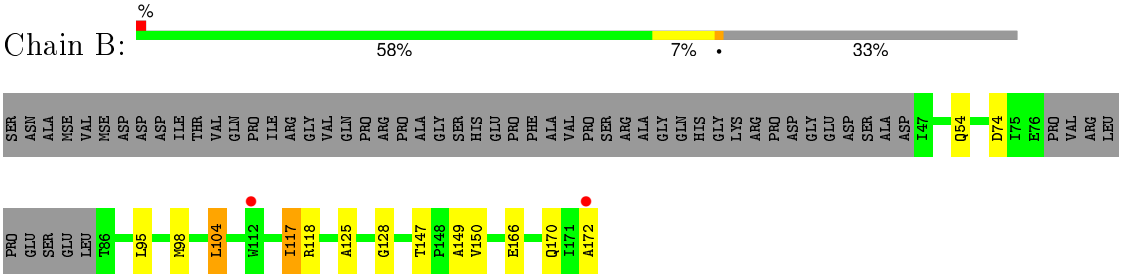
- Molecule 1: Uncharacterized protein

Chain A:



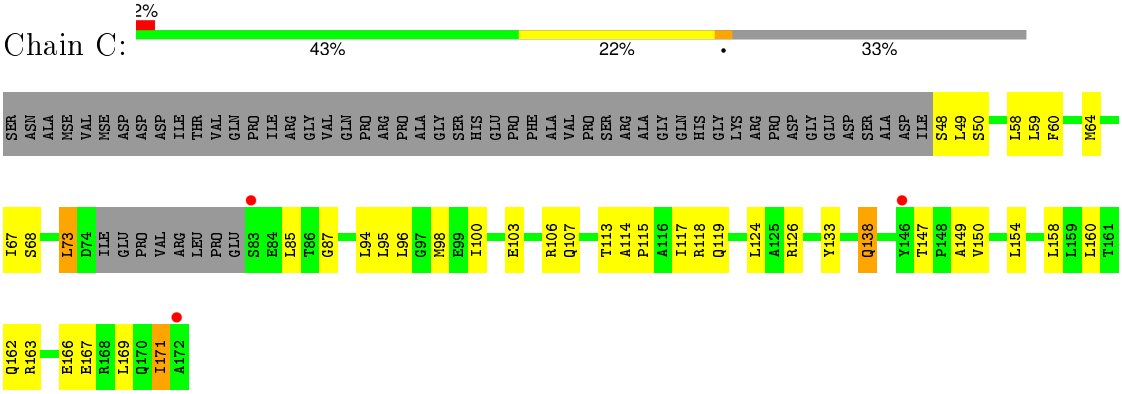
- Molecule 1: Uncharacterized protein

Chain B:



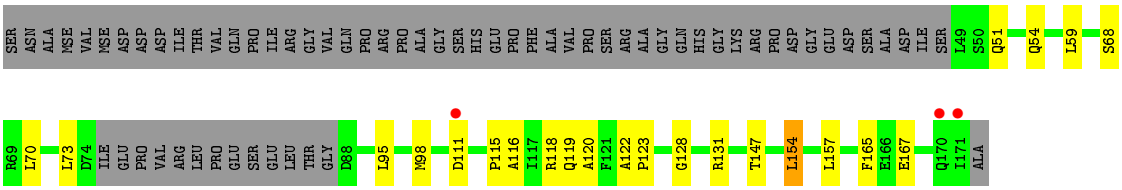
- Molecule 1: Uncharacterized protein

Chain C:



- Molecule 1: Uncharacterized protein

Chain D:


SER ASN ALA MSE VAL MSE ASP ASP ASP ILE THR VAL GLN PRO ILE ARG GLY VAL GLN PRO ARG PRO ALA GLY SER HIS GLU PRO PHE ALA VAL PRO SER ARG ALA GLY GLN HIS GLY LYS ARG PRO ASP GLY GLU ASP SER ALA ASP ILE SER L49 S50 Q51 Q54 L59 S68
R69 L70 L73 D74 ILE GLU PRO VAL ARG LEU PRO GLU SER GLU LEU THR GLY D88 L95 M98 D111 P115 A116 I117 R118 Q119 A120 F121 A122 P123 G128 R131 T147 L154 L157 F165 E166 E167 Q170 I171 ALA

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	95.37Å 109.47Å 151.60Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.94 37.13 – 2.94	Depositor EDS
% Data completeness (in resolution range)	99.7 (50.00-2.94) 99.4 (37.13-2.94)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.92 (at 2.95Å)	Xtriage
Refinement program	REFMAC 5.5.0109, CNS	Depositor
R, R_{free}	0.226 , 0.267 0.224 , 0.261	Depositor DCC
R_{free} test set	867 reflections (5.35%)	DCC
Wilson B-factor (Å^2)	81.6	Xtriage
Anisotropy	0.348	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 38.1	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Outliers	0 of 17059 reflections	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	3652	wwPDB-VP
Average B, all atoms (Å^2)	90.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.02% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.51	0/987	0.62	0/1334
1	B	0.50	0/914	0.62	0/1232
1	C	0.48	0/919	0.61	0/1238
1	D	0.46	0/871	0.57	0/1173
All	All	0.49	0/3691	0.61	0/4977

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	974	0	975	14	0
1	B	904	0	904	11	0
1	C	909	0	911	29	0
1	D	861	0	860	16	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
All	All	3652	0	3650	62	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 8.

All (62) 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:67:ILE:HG12	1:D:128:GLY:HA3	1.78	0.66
1:D:116:ALA:HA	1:D:119:GLN:HG3	1.81	0.62
1:D:115:PRO:O	1:D:119:GLN:HG2	1.99	0.61
1:C:126:ARG:HH21	1:C:166:GLU:HG3	1.65	0.61
1:D:122:ALA:HB3	1:D:123:PRO:HD3	1.84	0.59
1:C:95:LEU:HA	1:C:98:MSE:CE	2.34	0.58
1:C:114:ALA:HB1	1:C:118:ARG:HH21	1.70	0.56
1:C:113:THR:HB	1:C:115:PRO:HD2	1.87	0.56
1:B:98:MSE:HE2	1:B:125:ALA:HB1	1.88	0.55
1:C:96:LEU:O	1:C:100:ILE:HG12	2.05	0.55
1:A:158:LEU:O	1:A:162:GLN:HG3	2.07	0.55
1:C:126:ARG:HA	1:C:162:GLN:OE1	2.07	0.54
1:C:95:LEU:HA	1:C:98:MSE:HE3	1.90	0.53
1:C:87:GLY:HA2	1:C:160:LEU:HD21	1.89	0.53
1:C:115:PRO:O	1:C:119:GLN:HG2	2.09	0.53
1:C:133:TYR:CD2	1:C:158:LEU:HD22	2.44	0.53
1:A:67:ILE:HG12	1:B:128:GLY:HA3	1.91	0.52
1:B:147:THR:HB	1:B:150:VAL:H	1.74	0.52
1:C:60:PHE:O	1:C:64:MSE:HG2	2.09	0.52
1:A:114:ALA:N	1:A:115:PRO:HD2	2.25	0.51
1:A:62:ALA:O	1:A:66:GLN:HG2	2.12	0.50
1:C:48:SER:C	1:C:50:SER:H	2.14	0.50
1:C:147:THR:HG22	1:C:149:ALA:H	1.77	0.49
1:C:103:GLU:O	1:C:107:GLN:HG2	2.11	0.49
1:D:131:ARG:HA	1:D:131:ARG:NE	2.27	0.49
1:D:98:MSE:HE3	1:D:165:PHE:CD1	2.48	0.48
1:A:95:LEU:HA	1:A:98:MSE:CE	2.44	0.48
1:C:106:ARG:NH2	1:C:117:ILE:HD11	2.28	0.48
1:D:120:ALA:O	1:D:123:PRO:HD2	2.15	0.47
1:A:145:VAL:O	1:A:145:VAL:HG12	2.14	0.47
1:C:138:GLN:HA	1:C:138:GLN:OE1	2.14	0.47
1:C:94:LEU:O	1:C:98:MSE:HG3	2.14	0.47
1:A:100:ILE:HG23	1:B:104:LEU:HD22	1.97	0.46
1:D:115:PRO:HA	1:D:118:ARG:NH1	2.31	0.46
1:C:68:SER:HA	1:C:73:LEU:HD22	1.98	0.46
1:D:51:GLN:O	1:D:54:GLN:HG2	2.16	0.45
1:D:68:SER:HA	1:D:73:LEU:HD12	1.98	0.45
1:A:85:LEU:HD11	1:A:163:ARG:HG3	1.97	0.45
1:A:73:LEU:HD22	1:B:104:LEU:HD23	1.99	0.45
1:B:117:ILE:HG13	1:B:118:ARG:N	2.32	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:95:LEU:HA	1:A:98:MSE:HE2	1.97	0.45
1:C:147:THR:HB	1:C:150:VAL:H	1.81	0.45
1:A:95:LEU:HD23	1:A:98:MSE:CE	2.47	0.45
1:A:80:LEU:HD23	1:A:171:ILE:CD1	2.47	0.45
1:C:48:SER:C	1:C:50:SER:N	2.70	0.44
1:B:166:GLU:O	1:B:170:GLN:HG3	2.17	0.44
1:D:154:LEU:HD12	1:D:154:LEU:HA	1.74	0.44
1:C:59:LEU:HD23	1:D:157:LEU:HD23	2.01	0.43
1:C:154:LEU:HD23	1:C:154:LEU:HA	1.89	0.43
1:C:85:LEU:HD11	1:C:163:ARG:NE	2.34	0.43
1:C:154:LEU:HD22	1:D:59:LEU:HD22	2.02	0.42
1:B:118:ARG:HH22	1:B:172:ALA:C	2.23	0.41
1:C:58:LEU:HA	1:C:58:LEU:HD23	1.93	0.41
1:D:115:PRO:O	1:D:119:GLN:CG	2.67	0.41
1:C:167:GLU:O	1:C:171:ILE:HD12	2.20	0.41
1:C:124:LEU:HD22	1:D:70:LEU:HB3	2.02	0.41
1:B:95:LEU:HD23	1:B:95:LEU:HA	1.79	0.41
1:D:95:LEU:HA	1:D:98:MSE:HE2	2.02	0.41
1:A:54:GLN:NE2	1:B:54:GLN:OE1	2.48	0.41
1:B:147:THR:HG22	1:B:149:ALA:H	1.85	0.40
1:C:133:TYR:CG	1:C:158:LEU:HD22	2.56	0.40
1:A:80:LEU:HD23	1:A:171:ILE:HD13	2.03	0.40

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	124/175 (71%)	120 (97%)	3 (2%)	1 (1%)	24	58
1	B	113/175 (65%)	108 (96%)	5 (4%)	0	100	100
1	C	113/175 (65%)	111 (98%)	2 (2%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	D	106/175 (61%)	100 (94%)	6 (6%)	0	100 100
All	All	456/700 (65%)	439 (96%)	16 (4%)	1 (0%)	52 83

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	74	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
1	A	99/136 (73%)	95 (96%)	4 (4%)	38 73
1	B	91/136 (67%)	88 (97%)	3 (3%)	45 79
1	C	93/136 (68%)	88 (95%)	5 (5%)	27 61
1	D	87/136 (64%)	83 (95%)	4 (5%)	33 68
All	All	370/544 (68%)	354 (96%)	16 (4%)	35 70

All (16) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	64	MSE
1	A	74	ASP
1	A	76	GLU
1	A	88	ASP
1	B	74	ASP
1	B	104	LEU
1	B	117	ILE
1	C	49	LEU
1	C	73	LEU
1	C	138	GLN
1	C	169	LEU
1	C	171	ILE
1	D	111	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	147	THR
1	D	154	LEU
1	D	167	GLU

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

Mol	Chain	Res	Type
1	D	142	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

6.1 Protein, DNA and RNA chains

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^{th} 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	A	123/175 (70%)	-0.02	4 (3%)	50	45	62, 85, 112, 132	0
1	B	114/175 (65%)	-0.04	2 (1%)	71	70	63, 85, 118, 142	0
1	C	114/175 (65%)	0.03	3 (2%)	59	56	66, 88, 120, 142	0
1	D	107/175 (61%)	-0.20	3 (2%)	56	53	66, 93, 127, 143	0
All	All	458/700 (65%)	-0.05	12 (2%)	59	56	62, 87, 122, 143	0

All (12) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	172	ALA	5.1
1	A	172	ALA	4.4
1	C	83	SER	3.3
1	A	110	PRO	2.4
1	C	146	TYR	2.4
1	B	112	TRP	2.4
1	A	73	LEU	2.3
1	D	111	ASP	2.3
1	A	76	GLU	2.1
1	D	171	ILE	2.1
1	D	170	GLN	2.0
1	C	172	ALA	2.0

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.