wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 01:55 AM GMT

PDB ID	:	2ES7
Title	:	Crystal structure of Q8ZP25 from Salmonella typhimurium LT2. NESG TARGET STR70
Authors	:	Benach, J.; Su, M.; Forouhar, F.; Chen, Y.; Ho, C.K.; Janjua, H.; Cunningham, K.; Ma, L.-C.; Rong, X.; Liu, J.; Baran, M.; Acton, T.B.; Rost, B.; Montelione, G.T.; Tong, L.; Hunt, J.F.; Northeast Structural Genomics Consortium (NESG)
Deposited on	:	2005-10-25
Resolution	:	2.80 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
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:

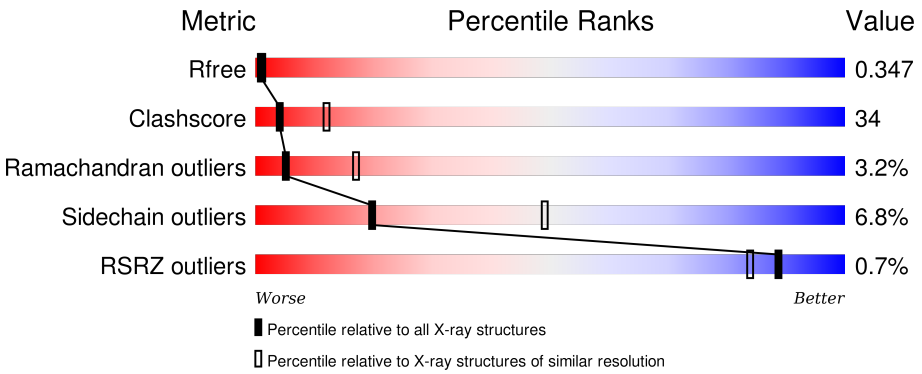
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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	142	% 34% 35% • • 27%
1	B	142	31% 37% • 27%
1	C	142	% 31% 38% • 27%
1	D	142	% 30% 39% • 27%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3840 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 putative thiol-disulfide isomerase and thioredoxin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
1	A	103	820	525	143	150	2	0	0	0
1	B	103	820	525	143	150	2	0	0	0
1	C	103	820	525	143	150	2	0	0	0
1	D	103	820	525	143	150	2	0	0	0

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	GB 16420320
A	57	MSE	MET	MODIFIED RESIDUE	GB 16420320
A	120	MSE	MET	MODIFIED RESIDUE	GB 16420320
A	135	LEU	-	CLONING ARTIFACT	GB 16420320
A	136	GLU	-	CLONING ARTIFACT	GB 16420320
A	137	HIS	-	EXPRESSION TAG	GB 16420320
A	138	HIS	-	EXPRESSION TAG	GB 16420320
A	139	HIS	-	EXPRESSION TAG	GB 16420320
A	140	HIS	-	EXPRESSION TAG	GB 16420320
A	141	HIS	-	EXPRESSION TAG	GB 16420320
A	142	HIS	-	EXPRESSION TAG	GB 16420320
B	1	MSE	MET	MODIFIED RESIDUE	GB 16420320
B	57	MSE	MET	MODIFIED RESIDUE	GB 16420320
B	120	MSE	MET	MODIFIED RESIDUE	GB 16420320
B	135	LEU	-	CLONING ARTIFACT	GB 16420320
B	136	GLU	-	CLONING ARTIFACT	GB 16420320
B	137	HIS	-	EXPRESSION TAG	GB 16420320
B	138	HIS	-	EXPRESSION TAG	GB 16420320
B	139	HIS	-	EXPRESSION TAG	GB 16420320
B	140	HIS	-	EXPRESSION TAG	GB 16420320
B	141	HIS	-	EXPRESSION TAG	GB 16420320

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	142	HIS	-	EXPRESSION TAG	GB 16420320
C	1	MSE	MET	MODIFIED RESIDUE	GB 16420320
C	57	MSE	MET	MODIFIED RESIDUE	GB 16420320
C	120	MSE	MET	MODIFIED RESIDUE	GB 16420320
C	135	LEU	-	CLONING ARTIFACT	GB 16420320
C	136	GLU	-	CLONING ARTIFACT	GB 16420320
C	137	HIS	-	EXPRESSION TAG	GB 16420320
C	138	HIS	-	EXPRESSION TAG	GB 16420320
C	139	HIS	-	EXPRESSION TAG	GB 16420320
C	140	HIS	-	EXPRESSION TAG	GB 16420320
C	141	HIS	-	EXPRESSION TAG	GB 16420320
C	142	HIS	-	EXPRESSION TAG	GB 16420320
D	1	MSE	MET	MODIFIED RESIDUE	GB 16420320
D	57	MSE	MET	MODIFIED RESIDUE	GB 16420320
D	120	MSE	MET	MODIFIED RESIDUE	GB 16420320
D	135	LEU	-	CLONING ARTIFACT	GB 16420320
D	136	GLU	-	CLONING ARTIFACT	GB 16420320
D	137	HIS	-	EXPRESSION TAG	GB 16420320
D	138	HIS	-	EXPRESSION TAG	GB 16420320
D	139	HIS	-	EXPRESSION TAG	GB 16420320
D	140	HIS	-	EXPRESSION TAG	GB 16420320
D	141	HIS	-	EXPRESSION TAG	GB 16420320
D	142	HIS	-	EXPRESSION TAG	GB 16420320

- Molecule 2 is water.

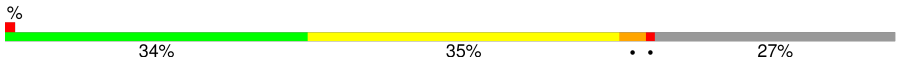
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	139	139	139	0	0
			Total	O		
2	B	149	149	149	0	0
			Total	O		
2	C	133	133	133	0	0
			Total	O		
2	D	139	139	139	0	0

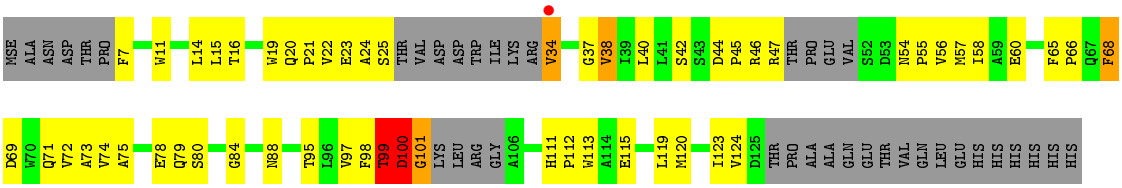
3 Residue-property plots (i)

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: putative thiol-disulfide isomerase and thioredoxin

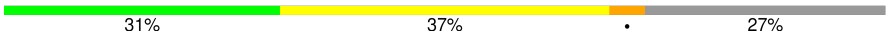
Chain A:

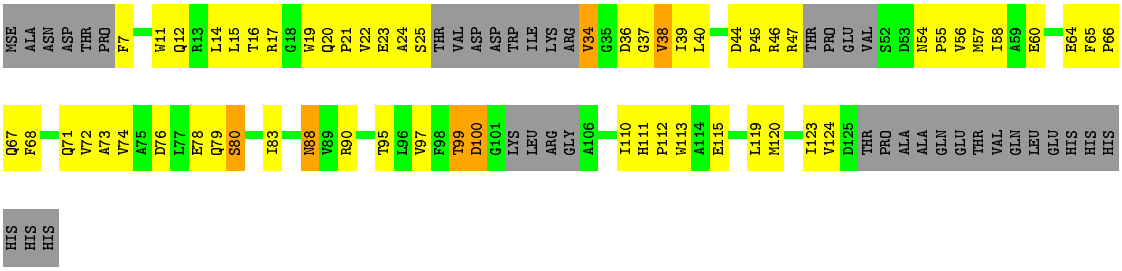




- Molecule 1: putative thiol-disulfide isomerase and thioredoxin

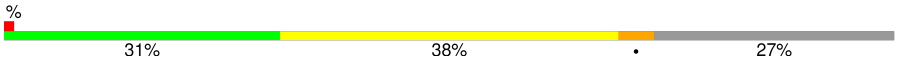
Chain B:

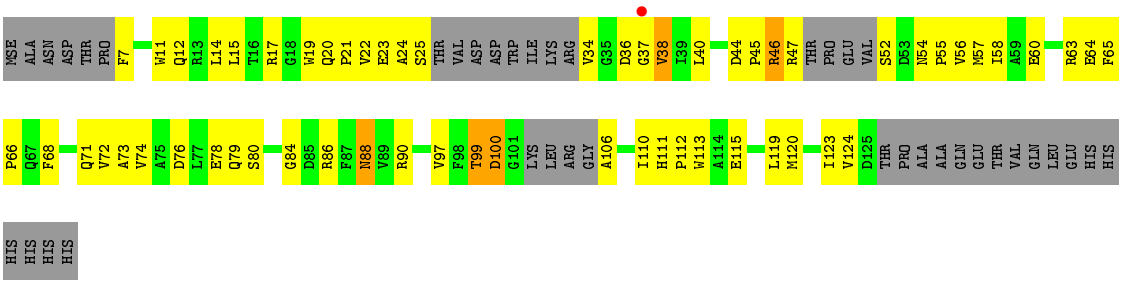




- Molecule 1: putative thiol-disulfide isomerase and thioredoxin

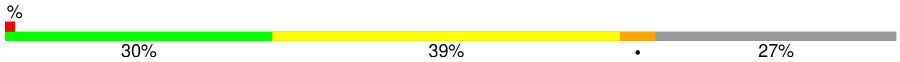
Chain C:

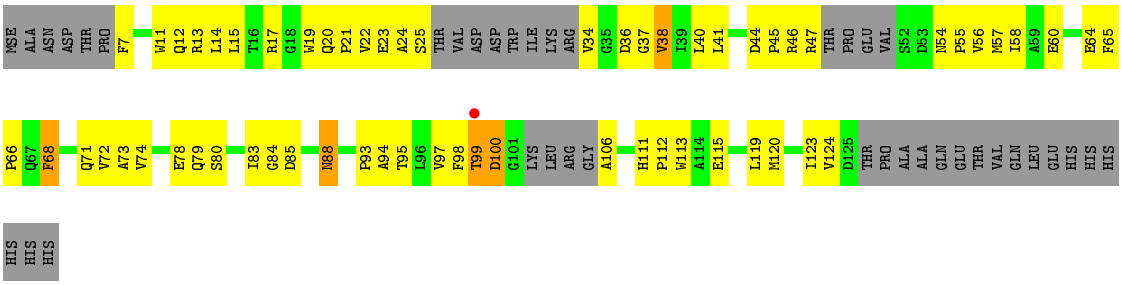




- Molecule 1: putative thiol-disulfide isomerase and thioredoxin

Chain D:




MSE
ALA
ASN
ASP
THR
PRO
F7
W11
Q12
R13
L14
L15
T16
R17
G18
W19
Q20
P21
V22
E23
A24
S25
THR
VAL
ASP
ASP
TRP
ILE
LYS
ARG
V34
G35
D36
G37
V38
I39
L40
L41
D44
P45
R46
R47
THR
PRO
GLU
VAL
S52
D53
N54
P55
V56
M57
I58
A59
E60
E64
F65
P66
Q67
F68
Q71
V72
A73
V74
E78
Q79
S80
I83
G84
D85
N88
P93
A94
T95
L96
V97
F98
T99
D100
G101
LYS
LEU
ARG
GLY
A106
H111
P112
W113
A114
E115
L119
M120
I123
V124
D125
THR
PRO
ALA
ALA
GLN
GLU
THR
VAL
GLN
LEU
GLU
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	39.55Å 71.16Å 95.00Å 90.00° 90.11° 90.00°	Depositor
Resolution (Å)	19.90 – 2.80 19.89 – 2.40	Depositor EDS
% Data completeness (in resolution range)	96.9 (19.90-2.80) 94.0 (19.89-2.40)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.92 (at 2.41Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.285 , 0.338 0.295 , 0.347	Depositor DCC
R_{free} test set	1239 reflections (9.59%)	DCC
Wilson B-factor (Å^2)	42.2	Xtriage
Anisotropy	0.234	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.21 , 84.7	EDS
Estimated twinning fraction	0.467 for h,-k,-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.26$	Xtriage
Outliers	1 of 38063 reflections (0.003%)	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	3840	wwPDB-VP
Average B, all atoms (Å^2)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 74.03 % of the origin peak, indicating pseudo translational symmetry.The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 1.6473e-06.The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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 (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	1.14	4/835 (0.5%)	2.30	6/1130 (0.5%)
1	B	0.55	0/835	0.78	2/1130 (0.2%)
1	C	0.56	0/835	0.78	2/1130 (0.2%)
1	D	0.59	1/835 (0.1%)	0.78	2/1130 (0.2%)
All	All	0.75	5/3340 (0.1%)	1.33	12/4520 (0.3%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	99	THR	C-N	-21.49	0.84	1.34
1	A	98	PHE	C-N	-18.43	0.91	1.34
1	A	99	THR	N-CA	-6.38	1.33	1.46
1	D	98	PHE	C-N	6.30	1.48	1.34
1	A	101	GLY	N-CA	5.05	1.53	1.46

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	99	THR	O-C-N	-70.20	10.38	122.70
1	A	99	THR	C-N-CA	15.04	159.30	121.70
1	A	98	PHE	O-C-N	-11.22	104.74	122.70
1	B	99	THR	CA-C-N	-8.63	98.21	117.20
1	D	99	THR	CA-C-N	-8.63	98.22	117.20

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	99	THR	Mainchain

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	820	0	802	43	0
1	B	820	0	804	64	0
1	C	820	0	804	54	0
1	D	820	0	804	61	0
2	A	139	0	0	6	0
2	B	149	0	0	23	0
2	C	133	0	0	11	0
2	D	139	0	0	13	0
All	All	3840	0	3214	219	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 34.

The worst 5 of 219 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:97:VAL:HG11	1:C:120:MSE:HE1	1.43	0.99
1:D:97:VAL:HG11	1:D:120:MSE:HE1	1.52	0.88
1:B:15:LEU:HD21	1:B:21:PRO:HG3	1.58	0.85
1:C:15:LEU:HD21	1:C:21:PRO:HG3	1.60	0.84
1:B:97:VAL:HG11	1:B:120:MSE:HE1	1.59	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	95/142 (67%)	83 (87%)	9 (10%)	3 (3%)	5	17
1	B	95/142 (67%)	81 (85%)	12 (13%)	2 (2%)	9	29
1	C	95/142 (67%)	83 (87%)	8 (8%)	4 (4%)	3	11
1	D	95/142 (67%)	82 (86%)	10 (10%)	3 (3%)	5	17
All	All	380/568 (67%)	329 (87%)	39 (10%)	12 (3%)	5	17

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	100	ASP
1	A	24	ALA
1	B	24	ALA
1	C	24	ALA
1	D	24	ALA

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	88/120 (73%)	81 (92%)	7 (8%)	15	40
1	B	88/120 (73%)	83 (94%)	5 (6%)	25	58
1	C	88/120 (73%)	82 (93%)	6 (7%)	20	49
1	D	88/120 (73%)	82 (93%)	6 (7%)	20	49
All	All	352/480 (73%)	328 (93%)	24 (7%)	20	49

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	80	SER
1	C	38	VAL
1	D	80	SER
1	B	88	ASN
1	C	34	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	20	GLN
1	A	111	HIS
1	B	111	HIS
1	C	111	HIS
1	D	111	HIS

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, 95th 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	101/142 (71%)	-0.44	1 (0%) 84 77	20, 31, 49, 66	0
1	B	101/142 (71%)	-0.50	0 100 100	21, 31, 48, 66	0
1	C	101/142 (71%)	-0.20	1 (0%) 84 77	20, 32, 50, 66	0
1	D	101/142 (71%)	-0.26	1 (0%) 84 77	23, 34, 51, 67	0
All	All	404/568 (71%)	-0.35	3 (0%) 89 84	20, 32, 50, 67	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	37	GLY	3.7
1	D	99	THR	2.3
1	A	34	VAL	2.0

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.