wwPDB X-ray Structure Validation Summary Report (i)

May 22, 2020 – 01:15 am BST

PDB ID	:	4IRV
Title	:	Structure of the Helicobacter pylori CagA Oncogene Bound to the Human Tumor Suppressor Apoptosis-stimulating Protein of p53-2
Authors	:	Stebbins, C.E.; Nesic, D.
Deposited on	:	2013-01-15
Resolution	:	2.04 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
https://www.wwpdb.org/validation/2017/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:

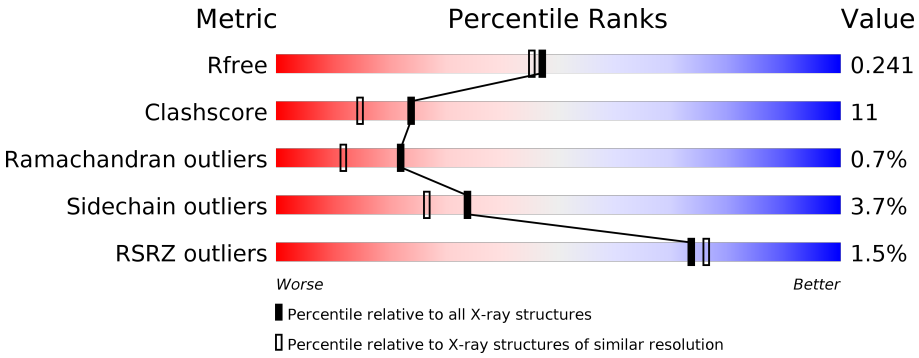
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.04 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.



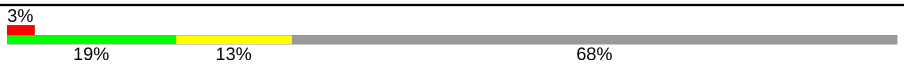
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1692 (2.04-2.04)
Clashscore	141614	1773 (2.04-2.04)
Ramachandran outliers	138981	1752 (2.04-2.04)
Sidechain outliers	138945	1752 (2.04-2.04)
RSRZ outliers	127900	1672 (2.04-2.04)

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 respectively. 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	221	68% 17% • 11%
1	B	221	% 68% 17% • 12%
1	C	221	2% 68% 18% • 13%
1	D	221	75% 12% • 10%
2	E	62	23% 10% 68%
2	F	62	3% 24% 8% 68%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	62	
2	H	62	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 7530 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 Cytotoxicity-associated immunodominant antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	197	Total	C	N	O	Se	0	0	0
			1609	1018	276	313	2			
1	B	194	Total	C	N	O	Se	0	0	0
			1592	1008	272	310	2			
1	C	192	Total	C	N	O	Se	0	0	0
			1572	994	270	306	2			
1	D	198	Total	C	N	O	Se	0	0	0
			1616	1023	277	314	2			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	15	GLY	-	EXPRESSION TAG	UNP P55980
A	16	PRO	-	EXPRESSION TAG	UNP P55980
A	17	VAL	-	EXPRESSION TAG	UNP P55980
A	18	ASP	-	EXPRESSION TAG	UNP P55980
B	15	GLY	-	EXPRESSION TAG	UNP P55980
B	16	PRO	-	EXPRESSION TAG	UNP P55980
B	17	VAL	-	EXPRESSION TAG	UNP P55980
B	18	ASP	-	EXPRESSION TAG	UNP P55980
C	15	GLY	-	EXPRESSION TAG	UNP P55980
C	16	PRO	-	EXPRESSION TAG	UNP P55980
C	17	VAL	-	EXPRESSION TAG	UNP P55980
C	18	ASP	-	EXPRESSION TAG	UNP P55980
D	15	GLY	-	EXPRESSION TAG	UNP P55980
D	16	PRO	-	EXPRESSION TAG	UNP P55980
D	17	VAL	-	EXPRESSION TAG	UNP P55980
D	18	ASP	-	EXPRESSION TAG	UNP P55980

- Molecule 2 is a protein called Apoptosis-stimulating of p53 protein 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	20	Total	C	N	O	Se	0	0	0
			157	100	27	29	1			
2	F	20	Total	C	N	O	Se	0	0	0
			157	100	27	29	1			
2	G	20	Total	C	N	O	Se	0	0	0
			157	100	27	29	1			
2	H	20	Total	C	N	O	Se	0	0	0
			157	100	27	29	1			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	721	GLY	-	EXPRESSION TAG	UNP Q13625
E	722	PRO	-	EXPRESSION TAG	UNP Q13625
E	723	LYS	-	EXPRESSION TAG	UNP Q13625
E	724	LEU	-	EXPRESSION TAG	UNP Q13625
E	725	ALA	-	EXPRESSION TAG	UNP Q13625
F	721	GLY	-	EXPRESSION TAG	UNP Q13625
F	722	PRO	-	EXPRESSION TAG	UNP Q13625
F	723	LYS	-	EXPRESSION TAG	UNP Q13625
F	724	LEU	-	EXPRESSION TAG	UNP Q13625
F	725	ALA	-	EXPRESSION TAG	UNP Q13625
G	721	GLY	-	EXPRESSION TAG	UNP Q13625
G	722	PRO	-	EXPRESSION TAG	UNP Q13625
G	723	LYS	-	EXPRESSION TAG	UNP Q13625
G	724	LEU	-	EXPRESSION TAG	UNP Q13625
G	725	ALA	-	EXPRESSION TAG	UNP Q13625
H	721	GLY	-	EXPRESSION TAG	UNP Q13625
H	722	PRO	-	EXPRESSION TAG	UNP Q13625
H	723	LYS	-	EXPRESSION TAG	UNP Q13625
H	724	LEU	-	EXPRESSION TAG	UNP Q13625
H	725	ALA	-	EXPRESSION TAG	UNP Q13625

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	124	Total	O	0	0
			124	124		
3	B	148	Total	O	0	0
			148	148		
3	C	92	Total	O	0	0
			92	92		
3	D	120	Total	O	0	0
			120	120		

Continued on next page...

Continued from previous page...

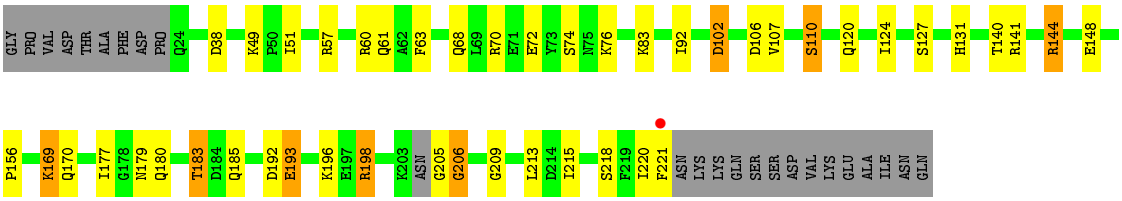
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	E	9	Total O 9 9	0	0
3	F	10	Total O 10 10	0	0
3	G	6	Total O 6 6	0	0
3	H	4	Total O 4 4	0	0

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 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.

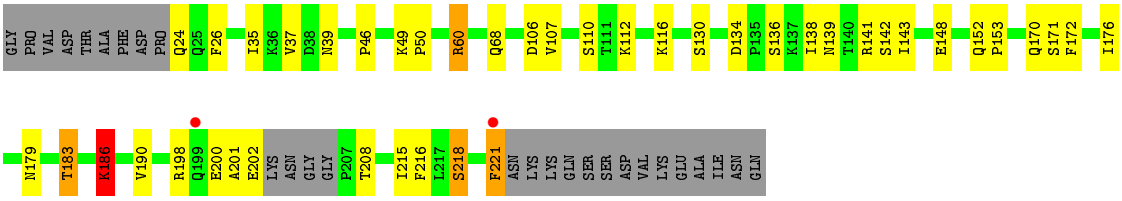
- Molecule 1: Cytotoxicity-associated immunodominant antigen

Chain A:



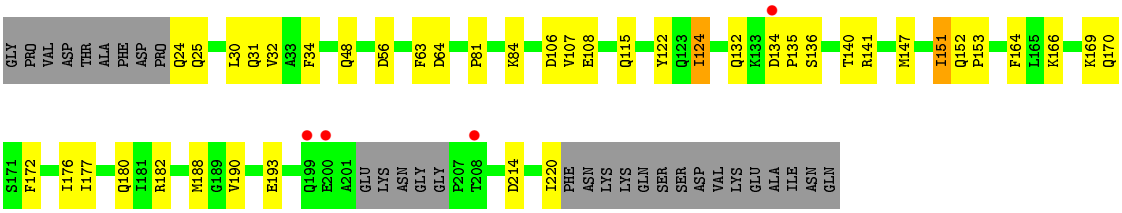
- Molecule 1: Cytotoxicity-associated immunodominant antigen

Chain B:



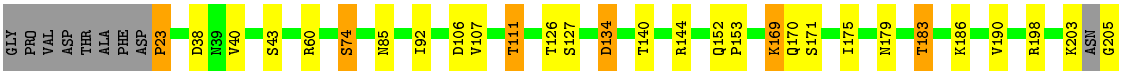
- Molecule 1: Cytotoxicity-associated immunodominant antigen

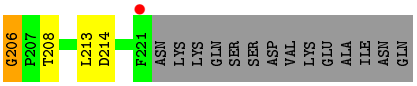
Chain C:



- Molecule 1: Cytotoxicity-associated immunodominant antigen

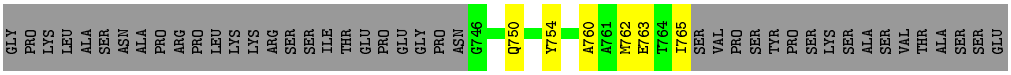
Chain D:





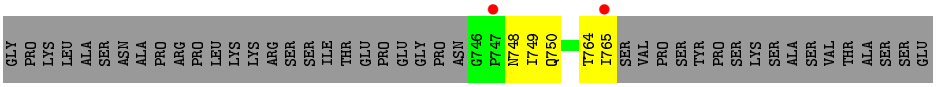
- Molecule 2: Apoptosis-stimulating of p53 protein 2

Chain E:



- Molecule 2: Apoptosis-stimulating of p53 protein 2

Chain F:

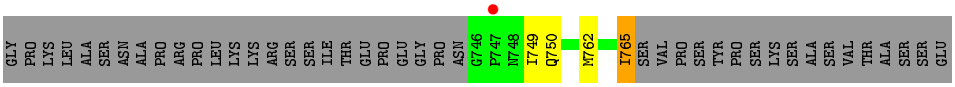


- Molecule 2: Apoptosis-stimulating of p53 protein 2

Chain G:

- Molecule 2: Apoptosis-stimulating of p53 protein 2

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	118.64Å 120.24Å 100.66Å 90.00° 115.64° 90.00°	Depositor
Resolution (Å)	90.75 – 2.04 90.75 – 2.04	Depositor EDS
% Data completeness (in resolution range)	96.7 (90.75-2.04) 96.7 (90.75-2.04)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.75 (at 2.05Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, R_{free}	0.189 , 0.234 0.195 , 0.241	Depositor DCC
R_{free} test set	4023 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	23.6	Xtriage
Anisotropy	0.629	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 47.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	7530	wwPDB-VP
Average B, all atoms (Å^2)	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.05% 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, 0.333 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	1.03	6/1638 (0.4%)	1.06	13/2197 (0.6%)
1	B	1.03	3/1621 (0.2%)	1.02	6/2175 (0.3%)
1	C	0.88	0/1600	0.91	3/2147 (0.1%)
1	D	0.98	1/1646 (0.1%)	0.98	5/2208 (0.2%)
2	E	0.87	0/157	1.01	1/210 (0.5%)
2	F	0.95	0/157	0.92	0/210
2	G	0.73	0/157	0.87	0/210
2	H	0.81	0/157	0.81	0/210
All	All	0.97	10/7133 (0.1%)	0.99	28/9567 (0.3%)

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	193	GLU	CD-OE1	9.42	1.36	1.25
1	B	171	SER	CB-OG	-8.95	1.30	1.42
1	A	193	GLU	CD-OE2	8.28	1.34	1.25
1	D	74	SER	CB-OG	-7.24	1.32	1.42
1	A	102	ASP	CG-OD2	7.22	1.42	1.25

The worst 5 of 28 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	60	ARG	NE-CZ-NH2	-10.54	115.03	120.30
1	A	102	ASP	CB-CG-OD2	10.04	127.34	118.30
1	A	102	ASP	CB-CG-OD1	-9.23	109.99	118.30
1	A	169	LYS	CD-CE-NZ	-8.53	92.09	111.70
1	D	60	ARG	NE-CZ-NH1	8.35	124.47	120.30

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	1609	0	1570	38	0
1	B	1592	0	1552	34	0
1	C	1572	0	1537	39	0
1	D	1616	0	1578	32	0
2	E	157	0	167	4	0
2	F	157	0	167	5	0
2	G	157	0	167	12	0
2	H	157	0	167	6	0
3	A	124	0	0	12	0
3	B	148	0	0	20	0
3	C	92	0	0	11	0
3	D	120	0	0	12	2
3	E	9	0	0	1	0
3	F	10	0	0	0	0
3	G	6	0	0	2	0
3	H	4	0	0	1	0
All	All	7530	0	6905	155	2

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 11.

The worst 5 of 155 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:63:PHE:HD1	1:A:177:ILE:HD11	1.11	1.13
1:C:63:PHE:HD1	1:C:177:ILE:HD11	1.10	1.11
1:C:122:TYR:CE2	3:C:332:HOH:O	2.03	1.09
1:A:63:PHE:CD1	1:A:177:ILE:HD11	1.96	1.00
1:C:122:TYR:CD2	3:C:332:HOH:O	2.14	0.97

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:D:355:HOH:O	3:D:377:HOH:O[2_554]	1.98	0.22
3:D:355:HOH:O	3:D:404:HOH:O[2_554]	2.01	0.19

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	193/221 (87%)	191 (99%)	1 (0%)	1 (0%)	29 18
1	B	190/221 (86%)	188 (99%)	0	2 (1%)	14 5
1	C	188/221 (85%)	181 (96%)	5 (3%)	2 (1%)	14 5
1	D	194/221 (88%)	190 (98%)	3 (2%)	1 (0%)	29 18
2	E	18/62 (29%)	18 (100%)	0	0	100 100
2	F	18/62 (29%)	18 (100%)	0	0	100 100
2	G	18/62 (29%)	17 (94%)	1 (6%)	0	100 100
2	H	18/62 (29%)	18 (100%)	0	0	100 100
All	All	837/1132 (74%)	821 (98%)	10 (1%)	6 (1%)	22 12

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	206	GLY
1	D	206	GLY
1	C	140	THR
1	C	141	ARG
1	B	200	GLU

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	180/199 (90%)	175 (97%)	5 (3%)	43 37
1	B	179/199 (90%)	169 (94%)	10 (6%)	21 12
1	C	177/199 (89%)	173 (98%)	4 (2%)	50 44
1	D	181/199 (91%)	175 (97%)	6 (3%)	38 31
2	E	17/52 (33%)	17 (100%)	0	100 100
2	F	17/52 (33%)	15 (88%)	2 (12%)	5 1
2	G	17/52 (33%)	16 (94%)	1 (6%)	19 11
2	H	17/52 (33%)	16 (94%)	1 (6%)	19 11
All	All	785/1004 (78%)	756 (96%)	29 (4%)	34 27

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	218	SER
1	C	115	GLN
2	F	764	THR
1	B	221	PHE
1	C	124	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	132	GLN
1	C	39	ASN
1	D	132	GLN
1	B	170	GLN
1	C	24	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	195/221 (88%)	-0.34	1 (0%) 91 92	18, 29, 56, 84	0
1	B	192/221 (86%)	-0.34	2 (1%) 82 84	16, 26, 49, 85	0
1	C	190/221 (85%)	-0.20	4 (2%) 63 67	23, 35, 64, 94	0
1	D	196/221 (88%)	-0.35	1 (0%) 91 92	19, 29, 58, 82	0
2	E	19/62 (30%)	-0.23	0 100 100	20, 31, 59, 62	0
2	F	19/62 (30%)	0.19	2 (10%) 6 6	21, 31, 71, 82	0
2	G	19/62 (30%)	0.57	2 (10%) 6 6	29, 42, 90, 90	0
2	H	19/62 (30%)	-0.22	1 (5%) 26 28	25, 32, 60, 61	0
All	All	849/1132 (75%)	-0.27	13 (1%) 73 76	16, 30, 61, 94	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	G	746	GLY	6.1
2	F	747	PRO	4.7
1	B	221	PHE	4.4
2	G	747	PRO	3.1
1	C	199	GLN	2.9

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.