



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 17, 2024 – 07:38 PM EDT

PDB ID : 5XB6
Title : Crystal structure of YcjY from E. Coli
Authors : Chen, C.F.; Gao, Z.Q.; Liu, Q.S.
Deposited on : 2017-03-16
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

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
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)	:	2.37.1

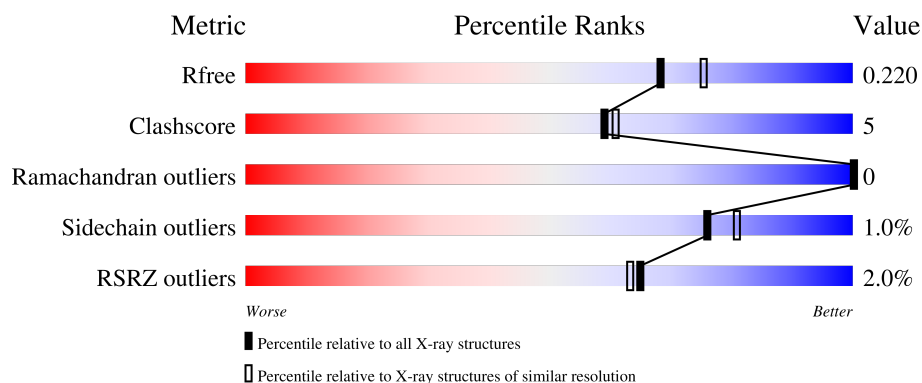
1 Overall quality at a glance

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.



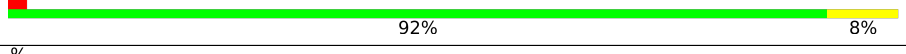
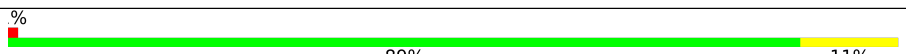
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 of 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 $\leq 5\%$. 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	306	<div> <div>2%</div> <div>90%</div> <div>10%</div> </div>
1	B	306	<div> <div>2%</div> <div>90%</div> <div>10%</div> </div>
1	C	306	<div> <div>2%</div> <div>85%</div> <div>15%</div> </div>
1	D	306	<div> <div>2%</div> <div>90%</div> <div>10%</div> </div>
1	E	306	<div> <div>%</div> <div>88%</div> <div>12%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	306	 92% 8%
1	G	306	 90% 10%
1	H	306	 92% 8%
1	I	306	 90% 10%
1	J	306	 90% 9%
1	K	306	 89% 11%
1	L	306	 89% 11%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 32624 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 YcjY.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	306	Total	C	N	O	S	0	0	0
			2371	1492	399	471	9			
1	B	306	Total	C	N	O	S	0	1	0
			2380	1498	401	472	9			
1	C	306	Total	C	N	O	S	0	0	0
			2371	1492	399	471	9			
1	D	306	Total	C	N	O	S	0	1	0
			2380	1498	401	472	9			
1	E	306	Total	C	N	O	S	0	1	0
			2377	1495	400	473	9			
1	F	306	Total	C	N	O	S	0	1	0
			2380	1498	401	472	9			
1	G	306	Total	C	N	O	S	0	1	0
			2380	1498	401	472	9			
1	H	306	Total	C	N	O	S	0	1	0
			2380	1498	401	472	9			
1	I	306	Total	C	N	O	S	0	1	0
			2380	1498	401	472	9			
1	J	305	Total	C	N	O	S	0	1	0
			2372	1493	400	471	8			
1	K	306	Total	C	N	O	S	0	1	0
			2380	1498	401	472	9			
1	L	306	Total	C	N	O	S	0	0	0
			2371	1492	399	471	9			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	317	Total	O	0	0
			317	317		
2	B	358	Total	O	0	0
			358	358		

Continued on next page...

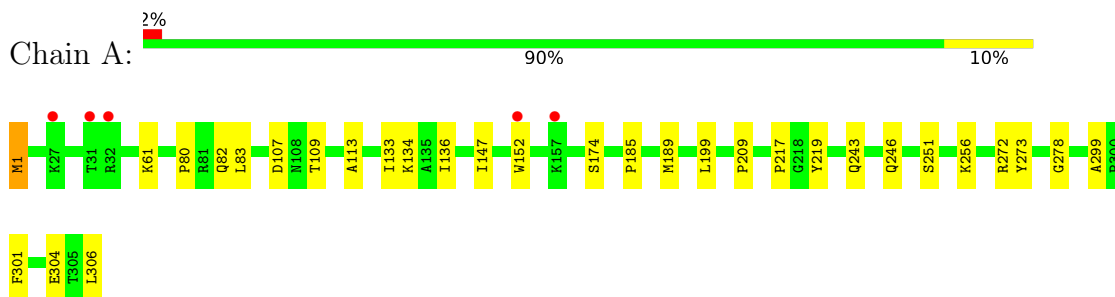
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	C	358	Total 358	O 358	0	0
2	D	389	Total 389	O 389	0	0
2	E	325	Total 325	O 325	0	0
2	F	387	Total 387	O 387	0	0
2	G	395	Total 395	O 395	0	0
2	H	324	Total 324	O 324	0	0
2	I	347	Total 347	O 347	0	0
2	J	331	Total 331	O 331	0	0
2	K	374	Total 374	O 374	0	0
2	L	197	Total 197	O 197	0	0

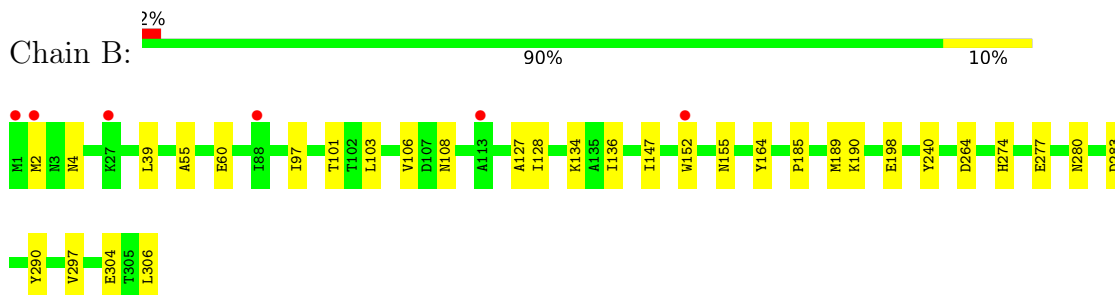
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

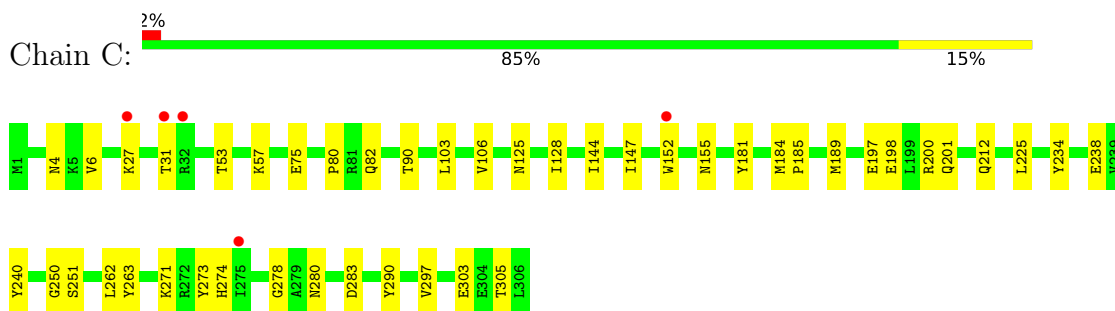
- Molecule 1: Uncharacterized protein YcjY



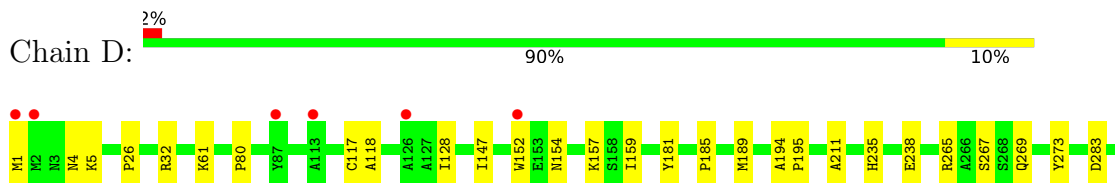
- Molecule 1: Uncharacterized protein YcjY



- Molecule 1: Uncharacterized protein YcjY

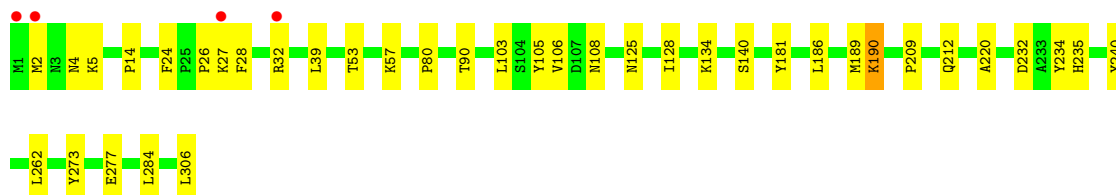
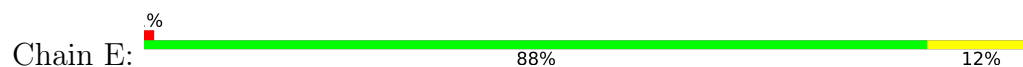


- Molecule 1: Uncharacterized protein YcjY





- Molecule 1: Uncharacterized protein YcjY



- Molecule 1: Uncharacterized protein YcjY



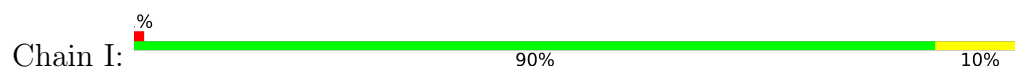
- Molecule 1: Uncharacterized protein YcjY



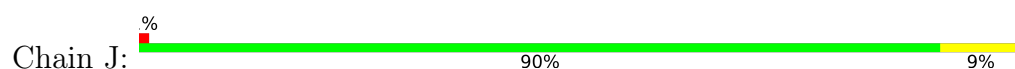
- Molecule 1: Uncharacterized protein YcjY



- Molecule 1: Uncharacterized protein YcjY



- Molecule 1: Uncharacterized protein YcjY



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	77.58Å 216.41Å 258.84Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.36 – 2.00 38.36 – 2.00	Depositor EDS
% Data completeness (in resolution range)	91.7 (38.36-2.00) 91.7 (38.36-2.00)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.71 (at 2.00Å)	Xtriage
Refinement program	PHENIX 1.8.1_1168	Depositor
R, R_{free}	0.182 , 0.219 0.184 , 0.220	Depositor DCC
R_{free} test set	13583 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å ²)	22.0	Xtriage
Anisotropy	0.646	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 51.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	32624	wwPDB-VP
Average B, all atoms (Å ²)	19.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.61% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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 [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	0.44	0/2425	0.57	0/3299
1	B	0.40	0/2434	0.54	0/3310
1	C	0.38	0/2425	0.55	0/3299
1	D	0.40	0/2434	0.56	1/3310 (0.0%)
1	E	0.40	0/2431	0.57	0/3307
1	F	0.39	0/2434	0.56	0/3310
1	G	0.42	0/2434	0.58	0/3310
1	H	0.38	0/2434	0.52	0/3310
1	I	0.39	0/2434	0.58	0/3310
1	J	0.41	0/2426	0.56	0/3300
1	K	0.39	0/2434	0.55	0/3310
1	L	0.36	0/2425	0.53	0/3299
All	All	0.40	0/29170	0.56	1/39674 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	211	ALA	N-CA-CB	-5.17	102.87	110.10

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	2371	0	2281	22	0
1	B	2380	0	2293	22	0
1	C	2371	0	2281	32	0
1	D	2380	0	2293	19	0
1	E	2377	0	2285	28	0
1	F	2380	0	2293	21	0
1	G	2380	0	2293	20	0
1	H	2380	0	2293	16	0
1	I	2380	0	2293	24	0
1	J	2372	0	2281	23	0
1	K	2380	0	2293	25	0
1	L	2371	0	2281	23	0
2	A	317	0	0	2	0
2	B	358	0	0	9	0
2	C	358	0	0	11	0
2	D	389	0	0	8	1
2	E	325	0	0	10	1
2	F	387	0	0	6	5
2	G	395	0	0	6	5
2	H	324	0	0	6	0
2	I	347	0	0	7	8
2	J	331	0	0	5	1
2	K	374	0	0	7	7
2	L	197	0	0	2	0
All	All	32624	0	27460	264	14

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 5.

The worst 5 of 264 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)
2:E:712:HOH:O	1:J:2:MET:SD	2.15	1.02
1:G:283:ASP:OD1	2:G:401:HOH:O	1.80	0.99
1:K:1:MET:SD	2:K:560:HOH:O	2.21	0.97
1:C:201:GLN:NE2	2:C:402:HOH:O	1.99	0.96
1:B:4:ASN:ND2	2:B:404:HOH:O	2.00	0.92

The worst 5 of 14 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)
2:F:590:HOH:O	2:I:407:HOH:O[4_455]	1.78	0.42
2:F:586:HOH:O	2:G:686:HOH:O[3_455]	1.89	0.31
2:E:455:HOH:O	2:J:489:HOH:O[4_455]	1.91	0.29
2:F:543:HOH:O	2:I:696:HOH:O[4_455]	1.91	0.29
2:F:749:HOH:O	2:I:729:HOH:O[4_455]	1.94	0.26

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	304/306 (99%)	295 (97%)	9 (3%)	0	100	100
1	B	305/306 (100%)	292 (96%)	13 (4%)	0	100	100
1	C	304/306 (99%)	292 (96%)	12 (4%)	0	100	100
1	D	305/306 (100%)	293 (96%)	12 (4%)	0	100	100
1	E	305/306 (100%)	296 (97%)	9 (3%)	0	100	100
1	F	305/306 (100%)	296 (97%)	9 (3%)	0	100	100
1	G	305/306 (100%)	295 (97%)	10 (3%)	0	100	100
1	H	305/306 (100%)	295 (97%)	10 (3%)	0	100	100
1	I	305/306 (100%)	298 (98%)	7 (2%)	0	100	100
1	J	304/306 (99%)	294 (97%)	10 (3%)	0	100	100
1	K	305/306 (100%)	298 (98%)	7 (2%)	0	100	100
1	L	304/306 (99%)	291 (96%)	13 (4%)	0	100	100
All	All	3656/3672 (100%)	3535 (97%)	121 (3%)	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	248/248 (100%)	245 (99%)	3 (1%)	71	76
1	B	249/248 (100%)	249 (100%)	0	100	100
1	C	248/248 (100%)	247 (100%)	1 (0%)	91	93
1	D	249/248 (100%)	248 (100%)	1 (0%)	91	93
1	E	249/248 (100%)	245 (98%)	4 (2%)	62	67
1	F	249/248 (100%)	246 (99%)	3 (1%)	71	76
1	G	249/248 (100%)	246 (99%)	3 (1%)	71	76
1	H	249/248 (100%)	246 (99%)	3 (1%)	71	76
1	I	249/248 (100%)	248 (100%)	1 (0%)	91	93
1	J	248/248 (100%)	245 (99%)	3 (1%)	71	76
1	K	249/248 (100%)	246 (99%)	3 (1%)	71	76
1	L	248/248 (100%)	244 (98%)	4 (2%)	62	67
All	All	2984/2976 (100%)	2955 (99%)	29 (1%)	76	81

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

Mol	Chain	Res	Type
1	G	288	LYS
1	L	198	GLU
1	H	273	TYR
1	K	288	LYS
1	H	271	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	J	4	ASN
1	K	35	GLN
1	L	4	ASN
1	D	154	ASN
1	D	269	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 monosaccharides 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	306/306 (100%)	-0.08	5 (1%) 72 70	8, 19, 33, 46	0
1	B	306/306 (100%)	-0.10	6 (1%) 65 63	7, 16, 32, 57	0
1	C	306/306 (100%)	-0.09	5 (1%) 72 70	8, 18, 32, 45	0
1	D	306/306 (100%)	-0.16	6 (1%) 65 63	6, 15, 30, 67	0
1	E	306/306 (100%)	-0.22	4 (1%) 77 76	7, 17, 28, 41	0
1	F	306/306 (100%)	-0.24	3 (0%) 82 81	6, 15, 27, 45	0
1	G	306/306 (100%)	-0.27	0 100 100	5, 14, 31, 46	0
1	H	306/306 (100%)	0.05	7 (2%) 60 59	8, 20, 37, 58	0
1	I	306/306 (100%)	-0.30	2 (0%) 87 87	6, 14, 29, 39	0
1	J	305/306 (99%)	-0.19	4 (1%) 77 76	7, 17, 30, 41	0
1	K	306/306 (100%)	-0.32	2 (0%) 87 87	6, 14, 28, 41	0
1	L	306/306 (100%)	0.42	30 (9%) 7 7	11, 28, 47, 60	0
All	All	3671/3672 (99%)	-0.12	74 (2%) 65 63	5, 17, 35, 67	0

The worst 5 of 74 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	L	1	MET	9.4
1	B	1	MET	7.9
1	D	1	MET	5.5
1	H	2	MET	5.0
1	L	279	ALA	4.6

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 monosaccharides 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.