wwPDB X-ray Structure Validation Summary Report (i)

Aug 6, 2020 – 01:48 PM BST

PDB ID	:	6S3Y
Title	:	CBDP35 SeMet structure
Authors	:	Hermoso, J.A.; Bartual, S.G.
Deposited on	:	2019-06-26
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 (i)) were used in the production of this report:

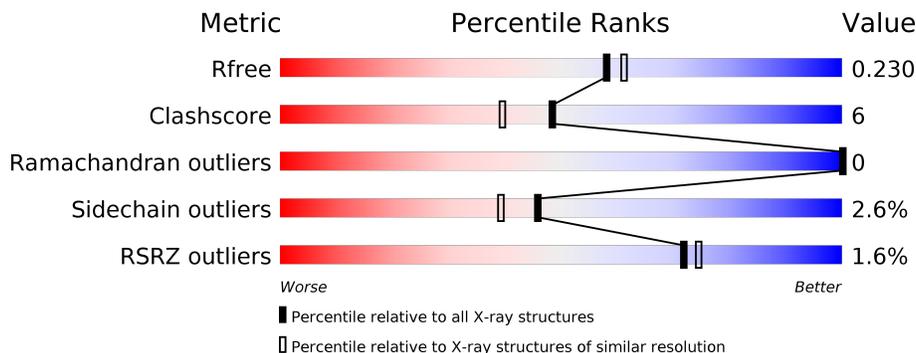|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

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.
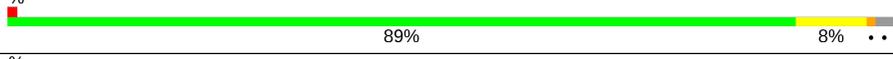
Metric	Percentile Ranks	Value
Rfree		0.230
Clashscore		6
Ramachandran outliers		0
Sidechain outliers		2.6%
RSRZ outliers		1.6%

Worse　　　　　　　　　　　　　　　　　　Better

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

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	140	86%　　13%　••
1	B	140	%　89%　　9%　•
1	C	140	2%　84%　　14%　••
1	D	140	%　87%　　11%　••
1	E	140	%　90%　　8%　••
1	F	140	%　89%　　9%　••

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	140	89% 9%
1	H	140	86% 13%
1	I	140	86% 10%
1	J	140	89% 8%
1	K	140	86% 12%
1	L	140	87% 11%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 15137 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 PlyP35.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	139	Total 1144	C 737	N 204	O 198	S 1	Se 4	0	1	0
1	B	140	Total 1145	C 738	N 205	O 197	S 1	Se 4	0	0	0
1	C	139	Total 1136	C 732	N 203	O 196	S 1	Se 4	0	0	0
1	D	139	Total 1135	C 732	N 202	O 196	S 1	Se 4	0	0	0
1	E	138	Total 1127	C 727	N 201	O 195	S 1	Se 3	0	0	0
1	F	139	Total 1135	C 732	N 202	O 196	S 1	Se 4	0	0	0
1	G	139	Total 1144	C 737	N 204	O 198	S 1	Se 4	0	1	0
1	H	139	Total 1135	C 732	N 202	O 196	S 1	Se 4	0	0	0
1	I	138	Total 1127	C 727	N 201	O 195	S 1	Se 3	0	0	0
1	J	137	Total 1127	C 726	N 201	O 196	S 1	Se 3	0	1	0
1	L	139	Total 1135	C 732	N 202	O 196	S 1	Se 4	0	0	0
1	K	139	Total 1135	C 732	N 202	O 196	S 1	Se 4	0	0	0

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	152	HIS	-	expression tag	UNP A8ATR6
A	153	MSE	-	expression tag	UNP A8ATR6
A	154	MSE	-	expression tag	UNP A8ATR6
B	152	HIS	-	expression tag	UNP A8ATR6
B	153	MSE	-	expression tag	UNP A8ATR6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	154	MSE	-	expression tag	UNP A8ATR6
C	152	HIS	-	expression tag	UNP A8ATR6
C	153	MSE	-	expression tag	UNP A8ATR6
C	154	MSE	-	expression tag	UNP A8ATR6
D	152	HIS	-	expression tag	UNP A8ATR6
D	153	MSE	-	expression tag	UNP A8ATR6
D	154	MSE	-	expression tag	UNP A8ATR6
E	152	HIS	-	expression tag	UNP A8ATR6
E	153	MSE	-	expression tag	UNP A8ATR6
E	154	MSE	-	expression tag	UNP A8ATR6
F	152	HIS	-	expression tag	UNP A8ATR6
F	153	MSE	-	expression tag	UNP A8ATR6
F	154	MSE	-	expression tag	UNP A8ATR6
G	152	HIS	-	expression tag	UNP A8ATR6
G	153	MSE	-	expression tag	UNP A8ATR6
G	154	MSE	-	expression tag	UNP A8ATR6
H	152	HIS	-	expression tag	UNP A8ATR6
H	153	MSE	-	expression tag	UNP A8ATR6
H	154	MSE	-	expression tag	UNP A8ATR6
I	152	HIS	-	expression tag	UNP A8ATR6
I	153	MSE	-	expression tag	UNP A8ATR6
I	154	MSE	-	expression tag	UNP A8ATR6
J	152	HIS	-	expression tag	UNP A8ATR6
J	153	MSE	-	expression tag	UNP A8ATR6
J	154	MSE	-	expression tag	UNP A8ATR6
L	152	HIS	-	expression tag	UNP A8ATR6
L	153	MSE	-	expression tag	UNP A8ATR6
L	154	MSE	-	expression tag	UNP A8ATR6
K	152	HIS	-	expression tag	UNP A8ATR6
K	153	MSE	-	expression tag	UNP A8ATR6
K	154	MSE	-	expression tag	UNP A8ATR6

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	148	Total 148	O 148	0	0
2	B	114	Total 114	O 114	0	0
2	C	114	Total 114	O 114	0	0
2	D	175	Total 175	O 175	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	E	114	Total 114	O 114	0	0
2	F	102	Total 102	O 102	0	0
2	G	147	Total 147	O 147	0	0
2	H	165	Total 165	O 165	0	0
2	I	108	Total 108	O 108	0	0
2	J	79	Total 79	O 79	0	0
2	L	103	Total 103	O 103	0	0
2	K	143	Total 143	O 143	0	0

3 Residue-property plots ⓘ

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.

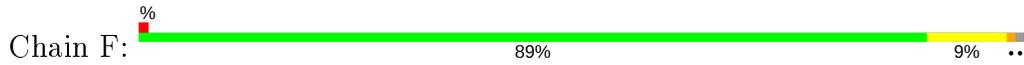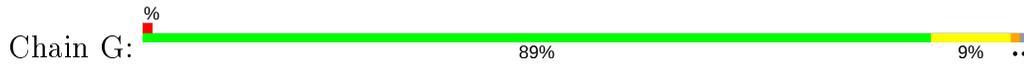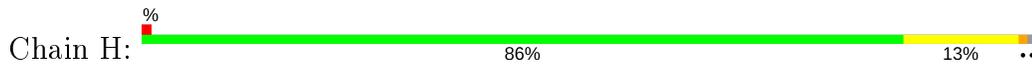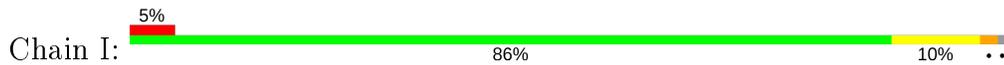- Molecule 1: PlyP35

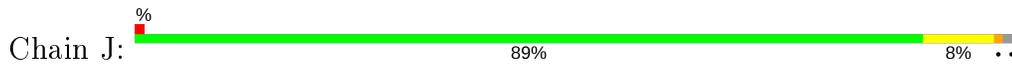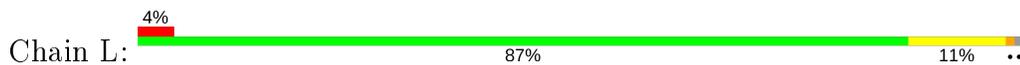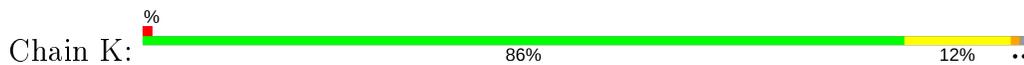- Molecule 1: PlyP35

- Molecule 1: PlyP35

- Molecule 1: PlyP35

- Molecule 1: PlyP35

- Molecule 1: PlyP35

%

Chain F:

89% 9%

HIS M153 K183 H193 K207 S216 K223 N224 S225 N230 H236 A237 W240 L245 R276 K277 K291

• Molecule 1: PlyP35

%

Chain G:

89% 9%

HIS M153 N156 H184 H193 K219 N224 A235 K254 I260 S265 Y272 F273 A274 S275 R276 T279 K290 K291

• Molecule 1: PlyP35

%

Chain H:

86% 13%

HIS M153 N156 N175 H184 H193 K203 N206 H215 S216 A235 H236 A237 H262 P263 A264 Y272 F273 A274 S275 R276 K277 K278 T279 K290 K291

• Molecule 1: PlyP35

5%

Chain I:

86% 10%

HIS MSE M154 E155 N156 K183 T191 D204 K207 R210 S216 K219 N224 S225 D226 G227 S228 T232 H236 A237 N242 I260 S265 Y272 F273 A274 K291

• Molecule 1: PlyP35

%

Chain J:

89% 8%

HIS MSE M154 E155 N156 S176 K177 H184 H193 A235 K254 I260 T261 H262 Y272 F273 A274 K277 K290 LYS

• Molecule 1: PlyP35

4%

Chain L:

87% 11%

HIS M153 M154 E155 N156 H184 T191 K192 H193 N194 H215 V220 N224 S225 S228 V229 N230 A235 K254 A264 Y272 F273 A274 S275 R276 T279 K290 K291

• Molecule 1: PlyP35

%

Chain K:

86% 12%

HIS M153 N156 H184 H193 N194 K203 N206 K207 A218 A235 W240 L245 K254 F257 K258 G259 I260 T261 H262 Y272 K277 K278 T279 K290 K291

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	335.88Å 95.56Å 84.41Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.84 – 2.04 47.78 – 2.04	Depositor EDS
% Data completeness (in resolution range)	99.8 (19.84-2.04) 99.8 (47.78-2.04)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.89 (at 2.05Å)	Xtriage
Refinement program	REFMAC 5.8.0238	Depositor
R, R_{free}	0.203 , 0.226 0.209 , 0.230	Depositor DCC
R_{free} test set	8475 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	22.1	Xtriage
Anisotropy	0.638	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 50.4	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.017 for -h,-k,l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	15137	wwPDB-VP
Average B, all atoms (Å^2)	30.0	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 43.74 % 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 1.6884e-04. 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, 0.333 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	0.70	0/1173	0.89	1/1570 (0.1%)
1	B	0.67	0/1174	0.96	3/1570 (0.2%)
1	C	0.67	0/1165	0.87	0/1559
1	D	0.72	0/1164	0.90	2/1558 (0.1%)
1	E	0.70	0/1157	0.91	1/1551 (0.1%)
1	F	0.71	0/1164	0.88	1/1558 (0.1%)
1	G	0.69	0/1173	0.86	0/1570
1	H	0.71	0/1164	0.90	0/1558
1	I	0.67	0/1157	0.88	2/1551 (0.1%)
1	J	0.66	0/1157	0.85	0/1552
1	K	0.70	0/1164	0.85	0/1558
1	L	0.67	0/1164	0.89	2/1558 (0.1%)
All	All	0.69	0/13976	0.89	12/18713 (0.1%)

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	153	MSE	CG-SE-CE	13.39	128.37	98.90
1	I	154	MSE	CG-SE-CE	7.21	114.75	98.90
1	L	154	MSE	CG-SE-CE	6.38	112.93	98.90
1	F	153	MSE	CG-SE-CE	6.28	112.72	98.90
1	E	276	ARG	CA-C-N	6.21	130.87	117.20

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	1144	0	1137	19	0
1	B	1145	0	1137	12	0
1	C	1136	0	1124	20	0
1	D	1135	0	1130	13	0
1	E	1127	0	1121	11	0
1	F	1135	0	1130	8	0
1	G	1144	0	1137	12	0
1	H	1135	0	1130	24	0
1	I	1127	0	1121	14	0
1	J	1127	0	1115	16	0
1	K	1135	0	1130	25	0
1	L	1135	0	1130	16	0
2	A	148	0	0	8	0
2	B	114	0	0	1	0
2	C	114	0	0	3	0
2	D	175	0	0	7	0
2	E	114	0	0	7	0
2	F	102	0	0	1	0
2	G	147	0	0	6	0
2	H	165	0	0	10	0
2	I	108	0	0	2	0
2	J	79	0	0	5	0
2	K	143	0	0	6	0
2	L	103	0	0	5	0
All	All	15137	0	13542	163	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 6.

The worst 5 of 163 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:I:156:ASN:ND2	1:I:260:ILE:HG21	1.83	0.93
1:C:262:HIS:HE1	1:J:262:HIS:HE1	1.20	0.89
1:A:262:HIS:HE1	1:H:262:HIS:HE1	1.21	0.89
1:I:156:ASN:ND2	1:I:260:ILE:CG2	2.38	0.87
1:A:262:HIS:HE1	1:H:262:HIS:CE1	1.94	0.85

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	138/140 (99%)	135 (98%)	3 (2%)	0	100	100
1	B	138/140 (99%)	136 (99%)	2 (1%)	0	100	100
1	C	137/140 (98%)	135 (98%)	2 (2%)	0	100	100
1	D	137/140 (98%)	133 (97%)	4 (3%)	0	100	100
1	E	136/140 (97%)	133 (98%)	3 (2%)	0	100	100
1	F	137/140 (98%)	134 (98%)	3 (2%)	0	100	100
1	G	138/140 (99%)	134 (97%)	4 (3%)	0	100	100
1	H	137/140 (98%)	134 (98%)	3 (2%)	0	100	100
1	I	136/140 (97%)	134 (98%)	2 (2%)	0	100	100
1	J	136/140 (97%)	133 (98%)	3 (2%)	0	100	100
1	K	137/140 (98%)	134 (98%)	3 (2%)	0	100	100
1	L	137/140 (98%)	134 (98%)	3 (2%)	0	100	100
All	All	1644/1680 (98%)	1609 (98%)	35 (2%)	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	123/119 (103%)	120 (98%)	3 (2%)	49	42
1	B	123/119 (103%)	117 (95%)	6 (5%)	25	17

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	C	122/119 (102%)	117 (96%)	5 (4%)	30	23
1	D	122/119 (102%)	120 (98%)	2 (2%)	62	59
1	E	121/119 (102%)	120 (99%)	1 (1%)	81	82
1	F	122/119 (102%)	118 (97%)	4 (3%)	38	31
1	G	123/119 (103%)	119 (97%)	4 (3%)	38	31
1	H	122/119 (102%)	120 (98%)	2 (2%)	62	59
1	I	121/119 (102%)	118 (98%)	3 (2%)	47	40
1	J	121/119 (102%)	118 (98%)	3 (2%)	47	40
1	K	122/119 (102%)	119 (98%)	3 (2%)	47	40
1	L	122/119 (102%)	120 (98%)	2 (2%)	62	59
All	All	1464/1428 (102%)	1426 (97%)	38 (3%)	46	39

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	183	LYS
1	G	254	LYS
1	K	153	MSE
1	F	223	LYS
1	G	276	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 64 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	184	HIS
1	H	193	HIS
1	K	184	HIS
1	G	215	HIS
1	G	262	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 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	135/140 (96%)	-0.17	0	100	100	13, 21, 43, 56	0
1	B	136/140 (97%)	-0.13	1 (0%)	87	89	15, 26, 49, 86	0
1	C	135/140 (96%)	0.04	3 (2%)	62	66	15, 31, 56, 94	0
1	D	135/140 (96%)	-0.19	1 (0%)	87	89	13, 19, 46, 71	0
1	E	135/140 (96%)	-0.13	1 (0%)	87	89	16, 25, 45, 70	0
1	F	135/140 (96%)	-0.14	1 (0%)	87	89	17, 28, 50, 67	0
1	G	135/140 (96%)	-0.20	1 (0%)	87	89	14, 23, 46, 58	0
1	H	135/140 (96%)	-0.13	2 (1%)	73	76	13, 21, 46, 61	0
1	I	135/140 (96%)	0.04	7 (5%)	27	29	20, 34, 70, 109	0
1	J	134/140 (95%)	-0.02	2 (1%)	73	76	23, 34, 56, 77	0
1	K	135/140 (96%)	-0.15	2 (1%)	73	76	16, 23, 47, 71	0
1	L	135/140 (96%)	0.07	5 (3%)	41	45	17, 32, 64, 78	0
All	All	1620/1680 (96%)	-0.09	26 (1%)	72	74	13, 27, 55, 109	0

The worst 5 of 26 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	152	HIS	6.0
1	C	152	HIS	5.6
1	C	264	ALA	4.7
1	I	225	SER	3.8
1	I	274	ALA	3.3

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.