wwPDB X-ray Structure Validation Summary Report (i)

May 21, 2020 – 02:43 am BST

PDB ID : 1R4Q
Title : Shiga toxin
Authors : Fraser, M.E.; Fujinaga, M.; Cherney, M.M.; Melton-Celsa, A.R.; Twiddy, E.M.; O'Brien, A.D.; James, M.N.G.
Deposited on : 2003-10-07
Resolution : 2.50 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
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)
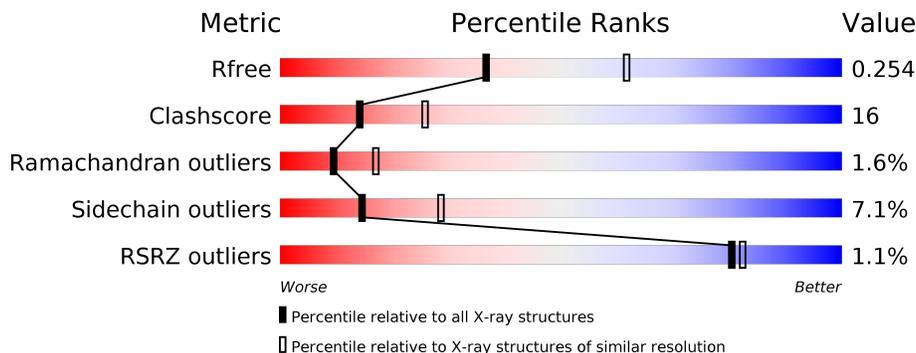Validation Pipeline (wwPDB-VP) : 2.11

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 2.50 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		0.254
Clashscore		16
Ramachandran outliers		1.6%
Sidechain outliers		7.1%
RSRZ outliers		1.1%

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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	293	66% 24% ·· 6%
1	L	293	68% 22% · 5%
2	B	69	67% 28% 6%
2	C	69	61% 38% ·
2	D	69	75% 23% ·
2	E	69	77% 22% ·

Continued on next page...

WORLDWIDE PDB PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	69	3% ... 64% ... 35%
2	G	69	80% ... 16%
2	H	69	78% ... 19%
2	I	69	67% ... 29%
2	J	69	% ... 62% ... 35%
2	K	69	4% ... 57% ... 36%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 9725 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 SHT cytotoxin A subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	276	Total 2136	C 1334	N 386	O 407	S 9	0	0	0
1	L	277	Total 2143	C 1338	N 387	O 409	S 9	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	168	ALA	PRO	see remark 999	UNP Q7BQ99
L	168	ALA	PRO	see remark 999	UNP Q7BQ99

- Molecule 2 is a protein called Shigella toxin chain B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	69	Total 540	C 339	N 90	O 108	S 3	0	0	0
2	C	69	Total 540	C 339	N 90	O 108	S 3	0	0	0
2	D	69	Total 540	C 339	N 90	O 108	S 3	0	0	0
2	E	69	Total 540	C 339	N 90	O 108	S 3	0	0	0
2	F	69	Total 540	C 339	N 90	O 108	S 3	0	0	0
2	G	69	Total 540	C 339	N 90	O 108	S 3	0	0	0
2	H	69	Total 540	C 339	N 90	O 108	S 3	0	0	0
2	I	69	Total 540	C 339	N 90	O 108	S 3	0	0	0
2	J	69	Total 540	C 339	N 90	O 108	S 3	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	K	69	540	339	90	108	3	0	0	0

- Molecule 3 is water.

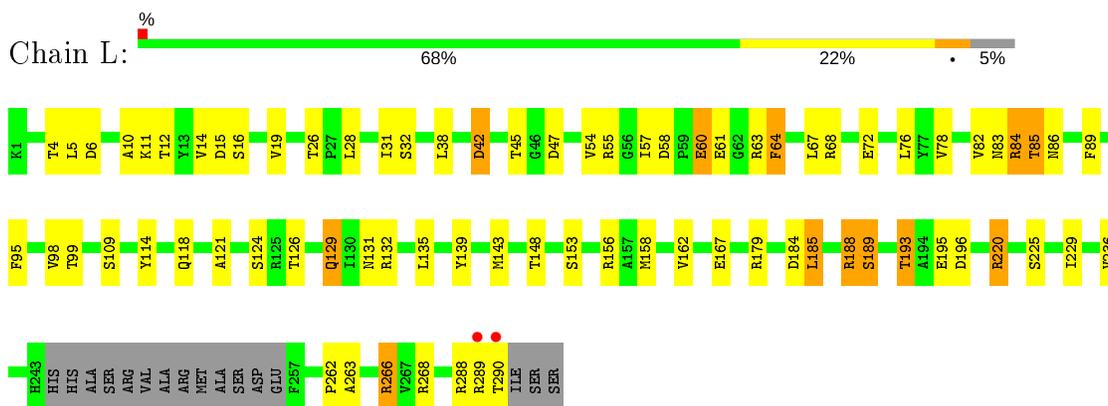| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	9	Total 9	O 9	0	0
3	L	6	Total 6	O 6	0	0
3	B	1	Total 1	O 1	0	0
3	C	1	Total 1	O 1	0	0
3	D	1	Total 1	O 1	0	0
3	E	4	Total 4	O 4	0	0
3	F	2	Total 2	O 2	0	0
3	G	10	Total 10	O 10	0	0
3	H	5	Total 5	O 5	0	0
3	I	1	Total 1	O 1	0	0
3	J	2	Total 2	O 2	0	0
3	K	4	Total 4	O 4	0	0

3 Residue-property plots ⓘ

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.
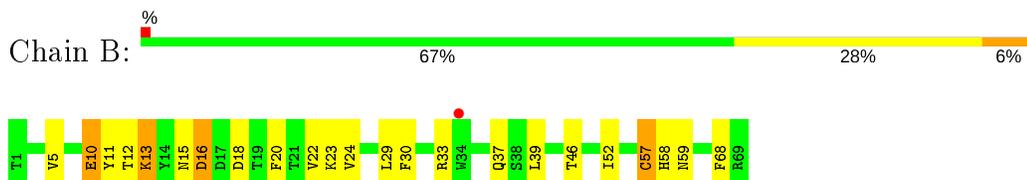
- Molecule 1: SHT cytotoxin A subunit
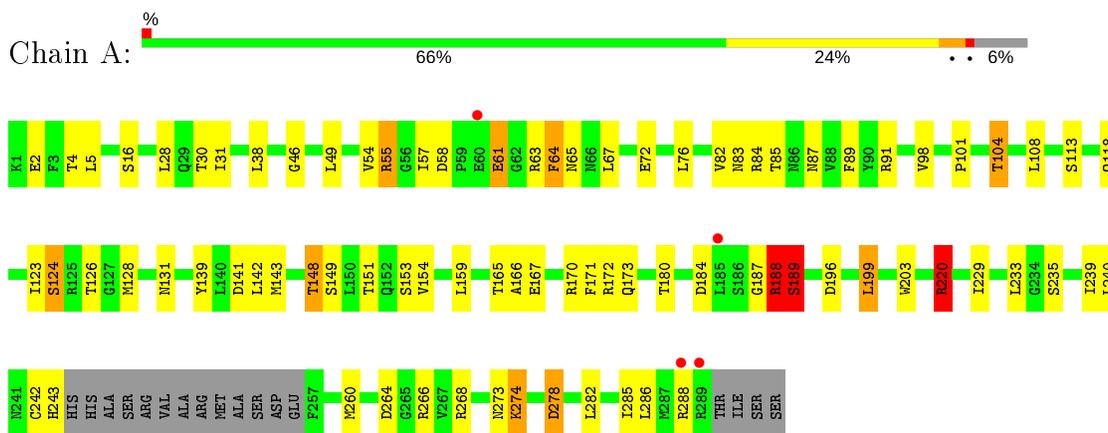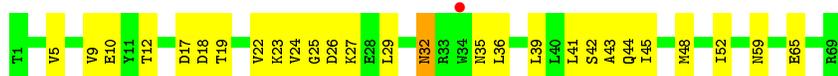

- Molecule 1: SHT cytotoxin A subunit

- Molecule 2: Shigella toxin chain B

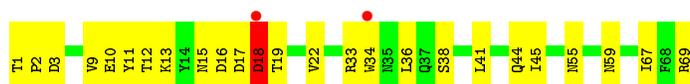
- Molecule 2: Shigella toxin chain B

T1 V5 V9 E10 Y11 T12 D17 D18 T19 V22 K23 V24 G25 D26 K27 E28 L29 N32 R33 W34 N35 L36 L39 L40 L41 S42 A43 Q44 I45 M48 I52 N59 E65 R69

- Molecule 2: Shigella toxin chain B

Chain D:

75%   23%

T1 P2 T12 K13 Y14 F20 T21 V22 D26 K27 Q37 L41 T46 G47 M48 K53 T54 R69

- Molecule 2: Shigella toxin chain B

Chain E:

77%   22%

T1 P2 T12 K13 Y14 N15 D16 D17 D18 T19 F20 T21 V22 N35 L39 L40 K53 T54 I67 F68 R69

- Molecule 2: Shigella toxin chain B

3%

Chain F:

64%   35%

T1 P2 D3 V9 E10 Y11 T12 K13 Y14 N15 D16 D17 D18 T19 V22 R33 W34 N35 L36 Q37 S38 L41 Q44 I45 N55 N59 I67 F68 R69

- Molecule 2: Shigella toxin chain B

Chain G:

80%   16%

T1 D18 T19 F20 T21 V22 K23 V24 N32 R33 W34 Q37 V50 K53 T54 V66 I67 F68 R69

- Molecule 2: Shigella toxin chain B

Chain H:

78%   19%

T1 P2 V5 Y11 T12 F20 T21 V22 K23 V24 E28 L29 F30 T31 N32 R33 Q37 V66 I67 F68 R69

- Molecule 2: Shigella toxin chain B

Chain I:

67%   29%

T1 P2 D3 C4 T12 K13 Y14 N15 D16 D17 D18 T19 V22 K23 V24 G25 D26 K27 E28 L29 F30 W34 S38 L39 L40 Q44 I52 N55 A56 C57 F63 R69

- Molecule 2: Shigella toxin chain B

%

Chain J:

62%   35%

• Molecule 2: Shigella toxin chain B

Chain K:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	133.00Å 147.18Å 82.85Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	98.60 – 2.50 98.68 – 2.50	Depositor EDS
% Data completeness (in resolution range)	(Not available) (98.60-2.50) 94.0 (98.68-2.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.56 (at 2.51Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.199 , 0.257 0.198 , 0.254	Depositor DCC
R_{free} test set	2747 reflections (5.13%)	wwPDB-VP
Wilson B-factor (Å²)	38.6	Xtriage
Anisotropy	0.214	Xtriage
Bulk solvent k_{sol}(e/Å³), B_{sol}(Å²)	0.27 , 37.4	EDS
L-test for twinning [2]	$<|L|> = 0.48$, $<L^2> = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	9725	wwPDB-VP
Average B, all atoms (Å²)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.39% 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 (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	0.73	0/2171	0.91	3/2941 (0.1%)
1	L	0.67	0/2178	0.86	0/2951
2	B	0.65	0/549	0.80	0/742
2	C	0.70	0/549	0.86	0/742
2	D	0.78	0/549	0.86	0/742
2	E	0.89	0/549	0.94	0/742
2	F	0.88	0/549	0.99	0/742
2	G	0.84	0/549	0.93	0/742
2	H	0.82	0/549	0.91	0/742
2	I	0.73	1/549 (0.2%)	0.79	0/742
2	J	0.72	0/549	0.82	0/742
2	K	0.98	1/549 (0.2%)	0.98	2/742 (0.3%)
All	All	0.76	2/9839 (0.0%)	0.89	5/13312 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	K	34	TRP	CB-CG	9.90	1.68	1.50
2	I	34	TRP	CB-CG	5.44	1.60	1.50

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	K	41	LEU	CA-CB-CG	-7.52	98.00	115.30
2	K	34	TRP	CA-CB-CG	6.21	125.50	113.70
1	A	220	ARG	NE-CZ-NH1	5.50	123.05	120.30
1	A	104	THR	N-CA-C	-5.28	96.73	111.00
1	A	278	ASP	N-CA-C	-5.10	97.23	111.00

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	2136	0	2129	71	0
1	L	2143	0	2136	72	0
2	B	540	0	527	20	0
2	C	540	0	527	21	0
2	D	540	0	527	14	0
2	E	540	0	527	13	0
2	F	540	0	527	23	0
2	G	540	0	527	11	0
2	H	540	0	527	12	0
2	I	540	0	527	17	0
2	J	540	0	527	23	0
2	K	540	0	527	42	0
3	A	9	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
3	E	4	0	0	0	0
3	F	2	0	0	0	0
3	G	10	0	0	0	0
3	H	5	0	0	0	0
3	I	1	0	0	0	0
3	J	2	0	0	0	0
3	K	4	0	0	1	0
3	L	6	0	0	0	0
All	All	9725	0	9535	300	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 16.

The worst 5 of 300 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:J:39:LEU:HD22	2:K:41:LEU:HD22	1.41	1.01
2:F:16:ASP:C	2:F:18:ASP:H	1.64	0.98
2:D:69:ARG:HH11	2:D:69:ARG:HG3	1.28	0.98

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:L:10:ALA:HB1	1:L:179:ARG:HG2	1.47	0.97
2:K:12:THR:HG22	2:K:22:VAL:HG12	1.45	0.94

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	272/293 (93%)	254 (93%)	12 (4%)	6 (2%)	6	10
1	L	273/293 (93%)	245 (90%)	23 (8%)	5 (2%)	8	14
2	B	67/69 (97%)	62 (92%)	4 (6%)	1 (2%)	10	18
2	C	67/69 (97%)	62 (92%)	5 (8%)	0	100	100
2	D	67/69 (97%)	60 (90%)	7 (10%)	0	100	100
2	E	67/69 (97%)	65 (97%)	2 (3%)	0	100	100
2	F	67/69 (97%)	61 (91%)	4 (6%)	2 (3%)	4	6
2	G	67/69 (97%)	64 (96%)	3 (4%)	0	100	100
2	H	67/69 (97%)	66 (98%)	1 (2%)	0	100	100
2	I	67/69 (97%)	63 (94%)	3 (4%)	1 (2%)	10	18
2	J	67/69 (97%)	62 (92%)	3 (4%)	2 (3%)	4	6
2	K	67/69 (97%)	60 (90%)	5 (8%)	2 (3%)	4	6
All	All	1215/1276 (95%)	1124 (92%)	72 (6%)	19 (2%)	9	17

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	64	PHE
1	A	188	ARG
1	A	189	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	F	18	ASP
2	J	57	CYS

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	238/252 (94%)	219 (92%)	19 (8%)	12	23
1	L	239/252 (95%)	219 (92%)	20 (8%)	11	21
2	B	61/61 (100%)	57 (93%)	4 (7%)	16	32
2	C	61/61 (100%)	58 (95%)	3 (5%)	25	47
2	D	61/61 (100%)	59 (97%)	2 (3%)	38	64
2	E	61/61 (100%)	59 (97%)	2 (3%)	38	64
2	F	61/61 (100%)	57 (93%)	4 (7%)	16	32
2	G	61/61 (100%)	56 (92%)	5 (8%)	11	22
2	H	61/61 (100%)	59 (97%)	2 (3%)	38	64
2	I	61/61 (100%)	55 (90%)	6 (10%)	8	15
2	J	61/61 (100%)	55 (90%)	6 (10%)	8	15
2	K	61/61 (100%)	57 (93%)	4 (7%)	16	32
All	All	1087/1114 (98%)	1010 (93%)	77 (7%)	14	28

5 of 77 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	L	188	ARG
2	C	10	GLU
2	J	64	SER
1	L	193	THR
2	B	5	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	32	ASN
2	C	37	GLN
2	F	59	ASN
2	B	55	ASN
2	B	59	ASN

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	276/293 (94%)	-0.37	4 (1%)	75	77	18, 37, 71, 86	0
1	L	277/293 (94%)	-0.38	2 (0%)	87	89	30, 49, 82, 93	0
2	B	69/69 (100%)	-0.47	1 (1%)	75	77	28, 45, 56, 66	0
2	C	69/69 (100%)	-0.46	1 (1%)	75	77	26, 36, 49, 63	0
2	D	69/69 (100%)	-0.52	0	100	100	20, 37, 52, 60	0
2	E	69/69 (100%)	-0.55	0	100	100	15, 24, 40, 56	0
2	F	69/69 (100%)	-0.24	2 (2%)	51	55	26, 38, 67, 81	0
2	G	69/69 (100%)	-0.63	0	100	100	17, 22, 35, 43	0
2	H	69/69 (100%)	-0.54	0	100	100	19, 31, 43, 52	0
2	I	69/69 (100%)	-0.47	0	100	100	33, 50, 64, 72	0
2	J	69/69 (100%)	-0.34	1 (1%)	75	77	33, 47, 62, 66	0
2	K	69/69 (100%)	-0.31	3 (4%)	35	38	23, 42, 75, 87	0
All	All	1243/1276 (97%)	-0.42	14 (1%)	80	82	15, 40, 68, 93	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	L	290	THR	6.6
2	F	34	TRP	5.2
1	A	289	ARG	4.6
2	K	34	TRP	4.2
2	F	18	ASP	3.5

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.