



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 28, 2023 – 07:13 AM EDT

PDB ID : 3KWQ  
Title : Structural characterization of H3K56Q nucleosomes and nucleosomal arrays  
Authors : Lilyestrom, W.G.; Clark, N.  
Deposited on : 2009-12-01  
Resolution : 3.50 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](mailto: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 : 2.35  
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) : 2.35

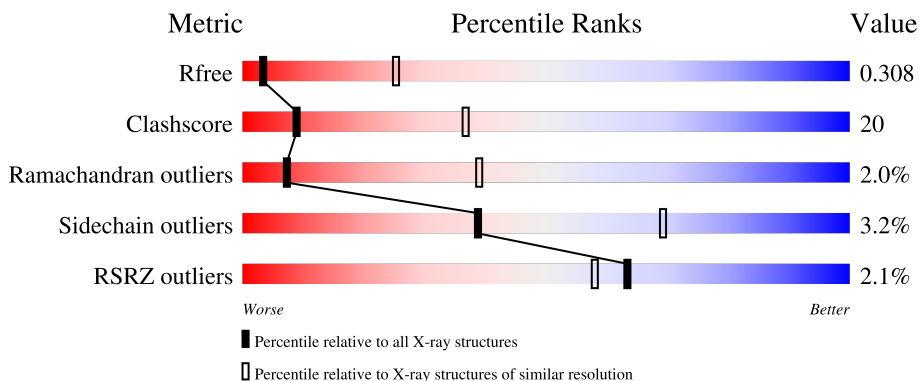
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.50 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.



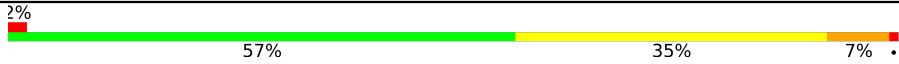

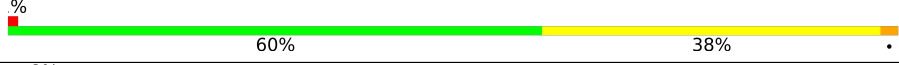


Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)
RSRZ outliers	127900	1559 (3.60-3.40)

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  $\geq 3$ , 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	98	 2% 57% 42%
1	E	98	 2% 50% 49%
2	B	83	 49% 43% 5%
2	F	83	 1% 58% 34% 8%
3	C	107	 2% 71% 25%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	G	107	 2% 57% 35% 7% ..
4	D	93	 % 57% 42% .
4	H	93	 % 60% 38% .
5	I	146	 8% 55% 45% .
5	J	146	 47% 53% .

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 12010 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 Histone H3.2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	98	808	508	155	142	3	0	0	0
1	E	98	808	508	155	142	3	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	56	GLU	LYS	engineered mutation	UNP P84233
A	102	ALA	GLY	SEE REMARK 999	UNP P84233
E	56	GLU	LYS	engineered mutation	UNP P84233
E	102	ALA	GLY	SEE REMARK 999	UNP P84233

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	79	627	395	121	110	1	0	0	0
2	F	83	662	418	129	114	1	0	0	0

- Molecule 3 is a protein called Histone H2A.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	C	107	826	520	161	145	0	0	0
3	G	106	818	516	160	142	0	0	0

- Molecule 4 is a protein called Histone H2B 1.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	93	Total	C	N	O	S	0	0	0
			729	459	131	137	2			
4	H	93	Total	C	N	O	S	0	0	0
			729	459	131	137	2			

- Molecule 5 is a DNA chain called DNA (146-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	I	146	Total	C	N	O	P	0	0	0
			2990	1430	541	874	145			
5	J	146	Total	C	N	O	P	0	0	0
			2990	1430	541	874	145			

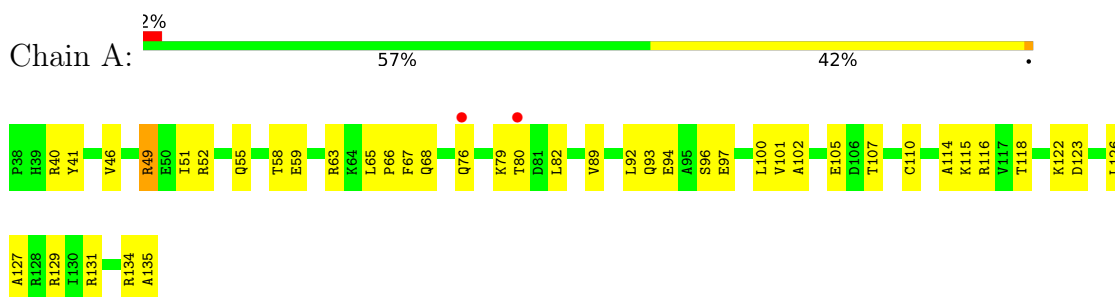
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	3	Total	O	0	0
			3	3		
6	B	1	Total	O	0	0
			1	1		
6	C	3	Total	O	0	0
			3	3		
6	D	1	Total	O	0	0
			1	1		
6	E	1	Total	O	0	0
			1	1		
6	F	2	Total	O	0	0
			2	2		
6	G	2	Total	O	0	0
			2	2		
6	H	2	Total	O	0	0
			2	2		
6	I	2	Total	O	0	0
			2	2		
6	J	6	Total	O	0	0
			6	6		

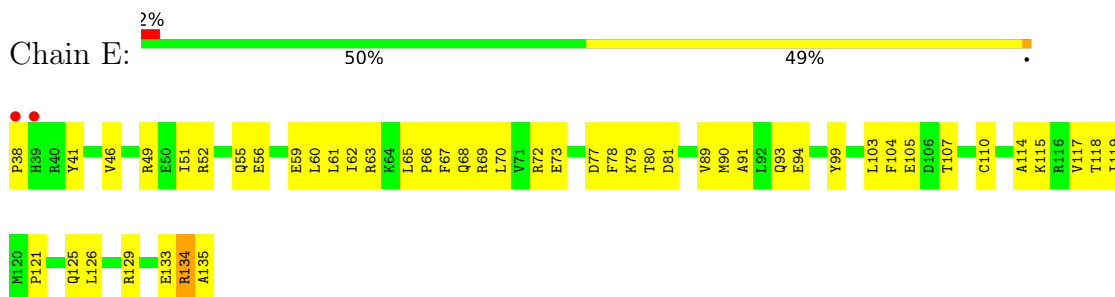
### 3 Residue-property plots [i](#)

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.

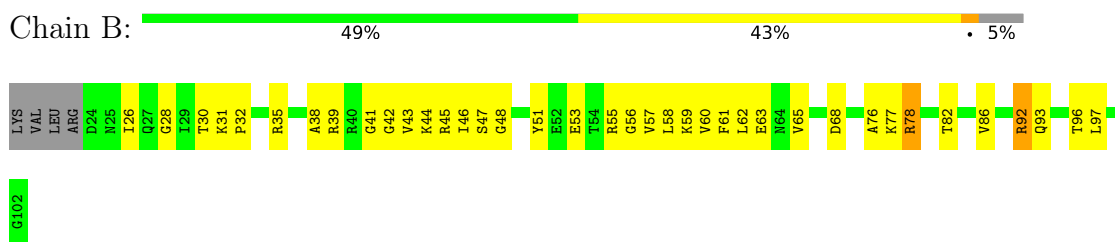
- Molecule 1: Histone H3.2



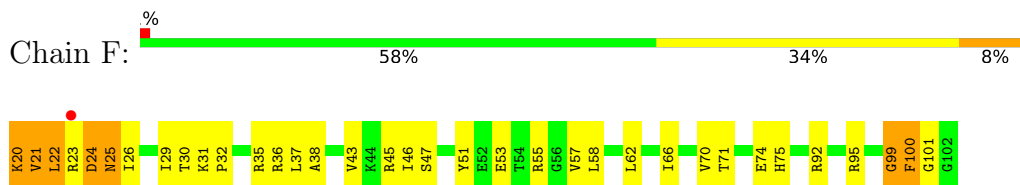
- Molecule 1: Histone H3.2



- Molecule 2: Histone H4



- Molecule 2: Histone H4



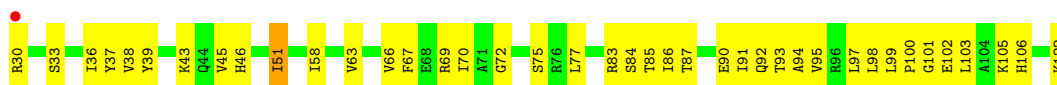
- Molecule 3: Histone H2A



• Molecule 3: Histone H2A



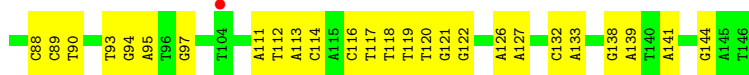
• Molecule 4: Histone H2B 1.1



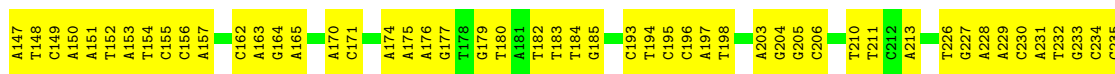
• Molecule 4: Histone H2B 1.1



• Molecule 5: DNA (146-MER)



• Molecule 5: DNA (146-MER)



T236	T237	G240	G246	C247	A248	G249	T258	A259	C260	A261	C262	T266	G267	G268	T269	A270	G271	A272	A273	T274	C275	T276	A279	T286	A287	T288	T289	G290	A291	T292
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	105.56Å 109.54Å 180.65Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 3.50 39.69 – 3.50	Depositor EDS
% Data completeness (in resolution range)	100.0 (20.00-3.50) 99.5 (39.69-3.50)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	0.11	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.28 (at 3.48Å)	Xtrriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.268 , 0.315 0.274 , 0.308	Depositor DCC
$R_{free}$ test set	1342 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	80.7	Xtrriage
Anisotropy	0.531	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.23 , 62.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.40$ , $\langle L^2 \rangle = 0.22$	Xtrriage
Estimated twinning fraction	0.076 for k,h,-l	Xtrriage
$F_o, F_c$ correlation	0.87	EDS
Total number of atoms	12010	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	117.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.74% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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	0.47	0/820	0.67	0/1100
1	E	0.56	0/820	0.78	1/1100 (0.1%)
2	B	0.51	0/634	0.75	0/848
2	F	0.80	1/669 (0.1%)	1.05	7/894 (0.8%)
3	C	0.56	0/836	0.67	0/1127
3	G	0.60	0/828	0.86	3/1117 (0.3%)
4	D	0.66	0/740	0.73	1/994 (0.1%)
4	H	0.51	0/740	0.68	0/994
5	I	0.41	0/3354	0.70	0/5175
5	J	0.40	0/3354	0.71	0/5175
All	All	0.50	1/12795 (0.0%)	0.74	12/18524 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	F	24	ASP	N-CA	5.03	1.56	1.46

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

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	F	99	GLY	N-CA-C	9.43	136.68	113.10
3	G	16	THR	N-CA-C	8.68	134.44	111.00
1	E	77	ASP	CB-CG-OD1	7.83	125.35	118.30
3	G	32	ARG	NE-CZ-NH1	-6.87	116.87	120.30
2	F	25	ASN	N-CA-C	6.73	129.17	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	808	0	839	43	0
1	E	808	0	839	58	0
2	B	627	0	663	38	0
2	F	662	0	709	60	0
3	C	826	0	884	41	0
3	G	818	0	877	52	0
4	D	729	0	753	44	0
4	H	729	0	753	45	0
5	I	2990	0	1651	81	0
5	J	2990	0	1651	89	0
6	A	3	0	0	0	0
6	B	1	0	0	0	0
6	C	3	0	0	0	0
6	D	1	0	0	0	0
6	E	1	0	0	0	0
6	F	2	0	0	1	0
6	G	2	0	0	0	0
6	H	2	0	0	0	0
6	I	2	0	0	0	0
6	J	6	0	0	0	0
All	All	12010	0	9619	438	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 20.

The worst 5 of 438 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:F:26:ILE:HG13	2:F:29:ILE:HD12	1.16	1.12
2:F:21:VAL:HG22	2:F:22:LEU:H	1.14	1.05
2:F:21:VAL:HG22	2:F:22:LEU:N	1.81	0.93
1:A:65:LEU:HB3	1:A:66:PRO:HD3	1.49	0.92
3:C:42:ARG:C	4:D:85:THR:HG23	1.90	0.92

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	96/98 (98%)	88 (92%)	8 (8%)	0	100	100
1	E	96/98 (98%)	84 (88%)	10 (10%)	2 (2%)	7	38
2	B	77/83 (93%)	66 (86%)	9 (12%)	2 (3%)	5	33
2	F	81/83 (98%)	72 (89%)	8 (10%)	1 (1%)	13	50
3	C	105/107 (98%)	92 (88%)	10 (10%)	3 (3%)	4	31
3	G	104/107 (97%)	95 (91%)	6 (6%)	3 (3%)	4	31
4	D	91/93 (98%)	81 (89%)	8 (9%)	2 (2%)	6	37
4	H	91/93 (98%)	76 (84%)	13 (14%)	2 (2%)	6	37
All	All	741/762 (97%)	654 (88%)	72 (10%)	15 (2%)	7	39

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	101	GLY
1	E	134	ARG
3	G	16	THR
3	C	118	LYS
2	F	21	VAL

### 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	85/85 (100%)	83 (98%)	2 (2%)	49	76

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	E	85/85 (100%)	84 (99%)	1 (1%)	71	87
2	B	64/68 (94%)	63 (98%)	1 (2%)	62	83
2	F	68/68 (100%)	65 (96%)	3 (4%)	28	62
3	C	85/85 (100%)	81 (95%)	4 (5%)	26	60
3	G	84/85 (99%)	79 (94%)	5 (6%)	19	52
4	D	79/79 (100%)	78 (99%)	1 (1%)	69	86
4	H	79/79 (100%)	76 (96%)	3 (4%)	33	65
All	All	629/634 (99%)	609 (97%)	20 (3%)	39	69

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	G	79	ILE
4	H	33	SER
4	H	84	SER
4	H	82	LYS
3	C	118	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
4	D	79	HIS
1	E	93	GLN
3	G	31	HIS
2	B	64	ASN
1	A	76	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	98/98 (100%)	-0.05	2 (2%) 65 60	0, 93, 123, 138	0
1	E	98/98 (100%)	-0.14	2 (2%) 65 60	0, 75, 118, 149	0
2	B	79/83 (95%)	-0.23	0 100 100	64, 85, 108, 128	0
2	F	83/83 (100%)	-0.24	1 (1%) 79 73	51, 72, 115, 165	0
3	C	107/107 (100%)	-0.26	2 (1%) 66 61	52, 75, 119, 164	0
3	G	106/107 (99%)	-0.11	2 (1%) 66 61	73, 92, 129, 161	0
4	D	93/93 (100%)	-0.22	1 (1%) 80 75	60, 80, 111, 141	0
4	H	93/93 (100%)	-0.11	1 (1%) 80 75	68, 89, 124, 156	0
5	I	146/146 (100%)	0.26	11 (7%) 14 14	94, 145, 188, 200	0
5	J	146/146 (100%)	-0.09	0 100 100	93, 146, 191, 207	0
All	All	1049/1054 (99%)	-0.10	22 (2%) 63 58	0, 93, 173, 207	0

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	I	55	DA	5.6
3	C	920	THR	5.1
1	E	38	PRO	3.8
1	E	39	HIS	3.5
1	A	76	GLN	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 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.