



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 28, 2023 – 08:45 AM EDT

PDB ID : 3KCM
Title : The crystal structure of thioredoxin protein from *Geobacter metallireducens*
Authors : Zhang, Z.; Burley, S.K.; Swaminathan, S.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on : 2009-10-21
Resolution : 2.45 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
Mogul : 1.8.5 (274361), CSD as541be (2020)
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

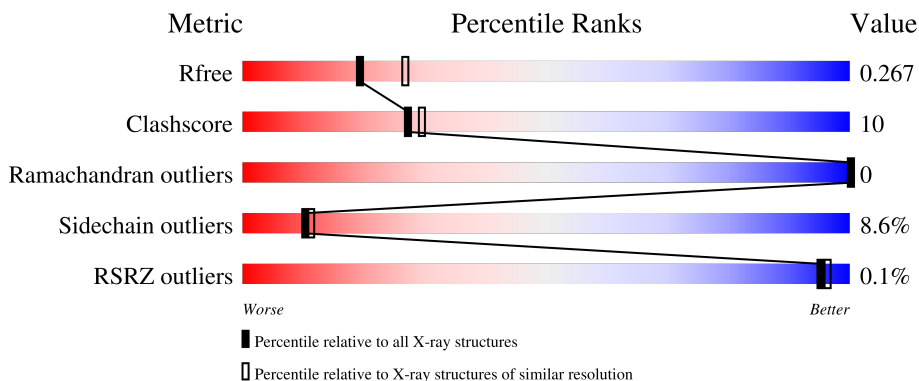
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.45 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	1544 (2.48-2.44)
Clashscore	141614	1613 (2.48-2.44)
Ramachandran outliers	138981	1598 (2.48-2.44)
Sidechain outliers	138945	1598 (2.48-2.44)
RSRZ outliers	127900	1523 (2.48-2.44)

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	154	 66% 21% 10%
1	B	154	 68% 20% 8%
1	C	154	 64% 21% 6% 10%
1	D	154	 75% 16% 8%
1	E	154	 75% 14% 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	154	 63% 25% 10%

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 6720 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 Thioredoxin family protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	139	1081	698	179	196	3	5	0	0	0
1	B	141	1088	704	180	196	3	5	0	0	0
1	C	139	1075	695	177	195	3	5	0	0	0
1	D	142	1091	705	181	197	3	5	0	0	0
1	E	141	1094	706	184	196	3	5	0	0	0
1	F	138	1063	687	176	192	3	5	0	0	0

There are 66 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	25	MSE	-	expression tag	UNP Q39SU7
A	26	SER	-	expression tag	UNP Q39SU7
A	27	LEU	-	expression tag	UNP Q39SU7
A	171	GLU	-	expression tag	UNP Q39SU7
A	172	GLY	-	expression tag	UNP Q39SU7
A	173	HIS	-	expression tag	UNP Q39SU7
A	174	HIS	-	expression tag	UNP Q39SU7
A	175	HIS	-	expression tag	UNP Q39SU7
A	176	HIS	-	expression tag	UNP Q39SU7
A	177	HIS	-	expression tag	UNP Q39SU7
A	178	HIS	-	expression tag	UNP Q39SU7
B	25	MSE	-	expression tag	UNP Q39SU7
B	26	SER	-	expression tag	UNP Q39SU7
B	27	LEU	-	expression tag	UNP Q39SU7
B	171	GLU	-	expression tag	UNP Q39SU7
B	172	GLY	-	expression tag	UNP Q39SU7
B	173	HIS	-	expression tag	UNP Q39SU7

Continued on next page...

Continued from previous page...

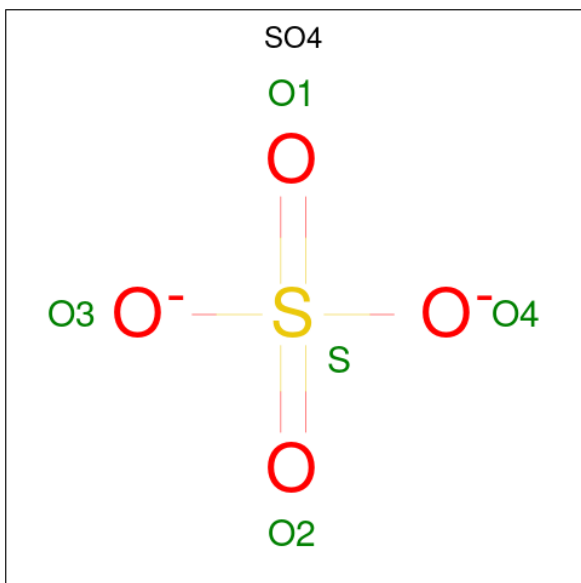
Chain	Residue	Modelled	Actual	Comment	Reference
B	174	HIS	-	expression tag	UNP Q39SU7
B	175	HIS	-	expression tag	UNP Q39SU7
B	176	HIS	-	expression tag	UNP Q39SU7
B	177	HIS	-	expression tag	UNP Q39SU7
B	178	HIS	-	expression tag	UNP Q39SU7
C	25	MSE	-	expression tag	UNP Q39SU7
C	26	SER	-	expression tag	UNP Q39SU7
C	27	LEU	-	expression tag	UNP Q39SU7
C	171	GLU	-	expression tag	UNP Q39SU7
C	172	GLY	-	expression tag	UNP Q39SU7
C	173	HIS	-	expression tag	UNP Q39SU7
C	174	HIS	-	expression tag	UNP Q39SU7
C	175	HIS	-	expression tag	UNP Q39SU7
C	176	HIS	-	expression tag	UNP Q39SU7
C	177	HIS	-	expression tag	UNP Q39SU7
C	178	HIS	-	expression tag	UNP Q39SU7
D	25	MSE	-	expression tag	UNP Q39SU7
D	26	SER	-	expression tag	UNP Q39SU7
D	27	LEU	-	expression tag	UNP Q39SU7
D	171	GLU	-	expression tag	UNP Q39SU7
D	172	GLY	-	expression tag	UNP Q39SU7
D	173	HIS	-	expression tag	UNP Q39SU7
D	174	HIS	-	expression tag	UNP Q39SU7
D	175	HIS	-	expression tag	UNP Q39SU7
D	176	HIS	-	expression tag	UNP Q39SU7
D	177	HIS	-	expression tag	UNP Q39SU7
D	178	HIS	-	expression tag	UNP Q39SU7
E	25	MSE	-	expression tag	UNP Q39SU7
E	26	SER	-	expression tag	UNP Q39SU7
E	27	LEU	-	expression tag	UNP Q39SU7
E	171	GLU	-	expression tag	UNP Q39SU7
E	172	GLY	-	expression tag	UNP Q39SU7
E	173	HIS	-	expression tag	UNP Q39SU7
E	174	HIS	-	expression tag	UNP Q39SU7
E	175	HIS	-	expression tag	UNP Q39SU7
E	176	HIS	-	expression tag	UNP Q39SU7
E	177	HIS	-	expression tag	UNP Q39SU7
E	178	HIS	-	expression tag	UNP Q39SU7
F	25	MSE	-	expression tag	UNP Q39SU7
F	26	SER	-	expression tag	UNP Q39SU7
F	27	LEU	-	expression tag	UNP Q39SU7
F	171	GLU	-	expression tag	UNP Q39SU7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	172	GLY	-	expression tag	UNP Q39SU7
F	173	HIS	-	expression tag	UNP Q39SU7
F	174	HIS	-	expression tag	UNP Q39SU7
F	175	HIS	-	expression tag	UNP Q39SU7
F	176	HIS	-	expression tag	UNP Q39SU7
F	177	HIS	-	expression tag	UNP Q39SU7
F	178	HIS	-	expression tag	UNP Q39SU7

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		
2	C	1	Total	O	S	0	0
			5	4	1		
2	D	1	Total	O	S	0	0
			5	4	1		
2	E	1	Total	O	S	0	0
			5	4	1		
2	F	1	Total	O	S	0	0
			5	4	1		

- Molecule 3 is water.

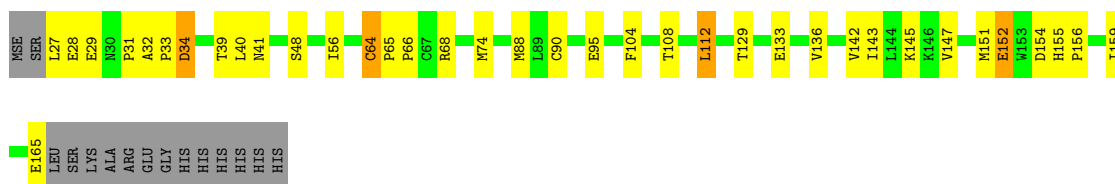
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	39	Total O 39 39	0	0
3	B	28	Total O 28 28	0	0
3	C	27	Total O 27 27	0	0
3	D	37	Total O 37 37	0	0
3	E	39	Total O 39 39	0	0
3	F	28	Total O 28 28	0	0

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.

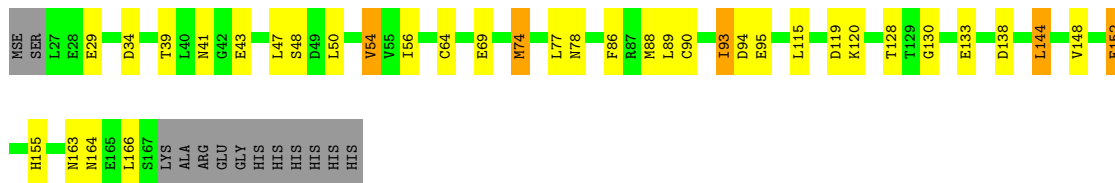
- Molecule 1: Thioredoxin family protein

Chain A: 



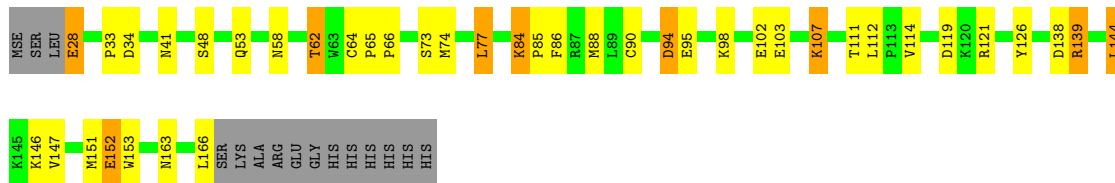
- Molecule 1: Thioredoxin family protein

Chain B: 



- Molecule 1: Thioredoxin family protein

Chain C: 



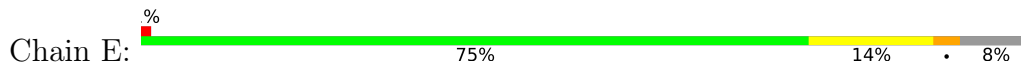
- Molecule 1: Thioredoxin family protein

Chain D: 



HIS
HIS
HIS
HIS
HIS

• Molecule 1: Thioredoxin family protein



HIS
HIS
HIS

• Molecule 1: Thioredoxin family protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	51.99Å 74.56Å 82.41Å 100.66° 109.05° 99.07°	Depositor
Resolution (Å)	75.36 – 2.45 75.36 – 2.45	Depositor EDS
% Data completeness (in resolution range)	94.4 (75.36-2.45) 94.4 (75.36-2.45)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.77 (at 2.45Å)	Xtrriage
Refinement program	REFMAC 5.5.0072	Depositor
R, R_{free}	0.214 , 0.270 0.214 , 0.267	Depositor DCC
R_{free} test set	1953 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å ²)	37.0	Xtrriage
Anisotropy	0.050	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 40.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	6720	wwPDB-VP
Average B, all atoms (Å ²)	23.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.82% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: SO4

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	0.90	0/1101	0.85	2/1487 (0.1%)
1	B	0.91	0/1108	0.89	0/1498
1	C	0.81	0/1095	0.85	0/1480
1	D	0.93	0/1111	0.91	1/1501 (0.1%)
1	E	0.97	0/1114	0.88	1/1503 (0.1%)
1	F	0.88	1/1083 (0.1%)	0.85	0/1464
All	All	0.90	1/6612 (0.0%)	0.87	4/8933 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	101	VAL	CB-CG2	-5.37	1.41	1.52

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	112	LEU	CA-CB-CG	5.67	128.34	115.30
1	A	34	ASP	CB-CG-OD1	5.47	123.22	118.30
1	E	154	ASP	CB-CG-OD2	5.36	123.12	118.30
1	A	112	LEU	CA-CB-CG	5.15	127.14	115.30

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	1081	0	1081	27	0
1	B	1088	0	1083	19	0
1	C	1075	0	1072	28	0
1	D	1091	0	1091	18	0
1	E	1094	0	1107	16	0
1	F	1063	0	1057	27	0
2	A	5	0	0	0	0
2	B	5	0	0	0	0
2	C	5	0	0	1	0
2	D	5	0	0	0	0
2	E	5	0	0	0	0
2	F	5	0	0	0	0
3	A	39	0	0	3	0
3	B	28	0	0	1	0
3	C	27	0	0	2	0
3	D	37	0	0	0	0
3	E	39	0	0	3	0
3	F	28	0	0	1	0
All	All	6720	0	6491	132	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 10.

The worst 5 of 132 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:D:88:MSE:HE1	1:D:112:LEU:HD22	1.28	1.09
1:D:88:MSE:HE3	1:D:90:CYS:SG	2.03	0.98
1:D:88:MSE:CE	1:D:112:LEU:HD22	1.96	0.96
1:B:95:GLU:HG3	3:B:231:HOH:O	1.73	0.88
1:F:139:ARG:HE	1:F:139:ARG:H	1.22	0.88

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	137/154 (89%)	129 (94%)	8 (6%)	0	100	100
1	B	139/154 (90%)	132 (95%)	7 (5%)	0	100	100
1	C	137/154 (89%)	128 (93%)	9 (7%)	0	100	100
1	D	140/154 (91%)	134 (96%)	6 (4%)	0	100	100
1	E	139/154 (90%)	135 (97%)	4 (3%)	0	100	100
1	F	136/154 (88%)	130 (96%)	6 (4%)	0	100	100
All	All	828/924 (90%)	788 (95%)	40 (5%)	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	117/126 (93%)	111 (95%)	6 (5%)	24	31
1	B	116/126 (92%)	103 (89%)	13 (11%)	6	5
1	C	116/126 (92%)	103 (89%)	13 (11%)	6	5
1	D	117/126 (93%)	111 (95%)	6 (5%)	24	31
1	E	119/126 (94%)	109 (92%)	10 (8%)	11	12
1	F	114/126 (90%)	102 (90%)	12 (10%)	7	6
All	All	699/756 (92%)	639 (91%)	60 (9%)	10	11

5 of 60 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	107	LYS
1	F	69	GLU
1	D	92	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	F	64	CYS
1	F	144	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	53	GLN
1	C	58	ASN
1	F	58	ASN
1	E	41	ASN
1	F	53	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](#)

6 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
2	SO4	D	3	-	4,4,4	0.28	0	6,6,6	0.51	0
2	SO4	E	4	-	4,4,4	0.24	0	6,6,6	0.50	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	SO4	B	2	-	4,4,4	0.20	0	6,6,6	0.18	0
2	SO4	C	6	-	4,4,4	0.15	0	6,6,6	0.20	0
2	SO4	F	5	-	4,4,4	0.27	0	6,6,6	0.35	0
2	SO4	A	1	-	4,4,4	0.20	0	6,6,6	0.55	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	6	SO4	1	0

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	134/154 (87%)	-0.37	0 100 100	12, 21, 36, 49	0
1	B	136/154 (88%)	-0.33	0 100 100	12, 22, 40, 47	0
1	C	134/154 (87%)	-0.35	0 100 100	12, 21, 38, 51	0
1	D	137/154 (88%)	-0.41	0 100 100	12, 20, 33, 39	0
1	E	136/154 (88%)	-0.30	1 (0%) 87 88	13, 21, 39, 46	0
1	F	133/154 (86%)	-0.33	0 100 100	13, 23, 40, 47	0
All	All	810/924 (87%)	-0.35	1 (0%) 95 96	12, 21, 39, 51	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	107	LYS	2.0

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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
2	SO4	E	4	5/5	0.97	0.10	24,25,26,27	0
2	SO4	B	2	5/5	0.99	0.12	22,24,25,25	0
2	SO4	C	6	5/5	0.99	0.16	27,28,28,31	0
2	SO4	D	3	5/5	0.99	0.15	26,26,29,29	0
2	SO4	A	1	5/5	0.99	0.12	19,20,22,25	0
2	SO4	F	5	5/5	0.99	0.17	24,25,25,26	0

6.5 Other polymers [i](#)

There are no such residues in this entry.