



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 12, 2024 – 08:52 PM EDT

PDB ID : 3ED4
Title : Crystal structure of putative arylsulfatase from escherichia coli
Authors : Patskovsky, Y.; Ozyurt, S.; Gilmore, M.; Chang, S.; Bain, K.; Wasserman, S.; Koss, J.; Sauder, J.M.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on : 2008-09-02
Resolution : 1.70 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 : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 1.20.1
EDS : 2.36.2
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) : 2.36.2

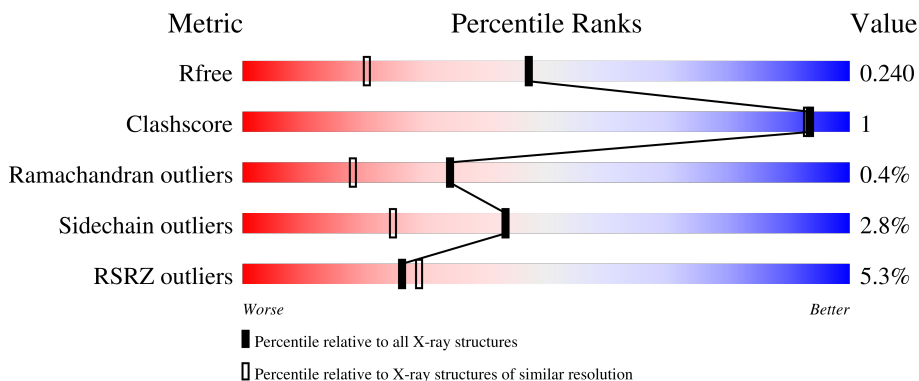
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.70 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	4298 (1.70-1.70)
Clashscore	141614	4695 (1.70-1.70)
Ramachandran outliers	138981	4610 (1.70-1.70)
Sidechain outliers	138945	4610 (1.70-1.70)
RSRZ outliers	127900	4222 (1.70-1.70)

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	502	 4% 88% 5% 7%
1	B	502	 2% 90% 6%
1	C	502	 8% 88% 7%
1	D	502	 6% 87% 7% 6%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard

residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	GOL	B	1	-	-	X	-
5	UNL	D	520	-	-	X	-

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 16828 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 ARYLSULFATASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	468	3800	2426	642	714	18	0	12	0
1	B	474	3807	2428	643	718	18	0	7	0
1	C	467	3747	2390	633	706	18	0	4	0
1	D	474	3883	2481	658	725	19	0	18	0

There are 44 discrepancies between the modelled and reference sequences:

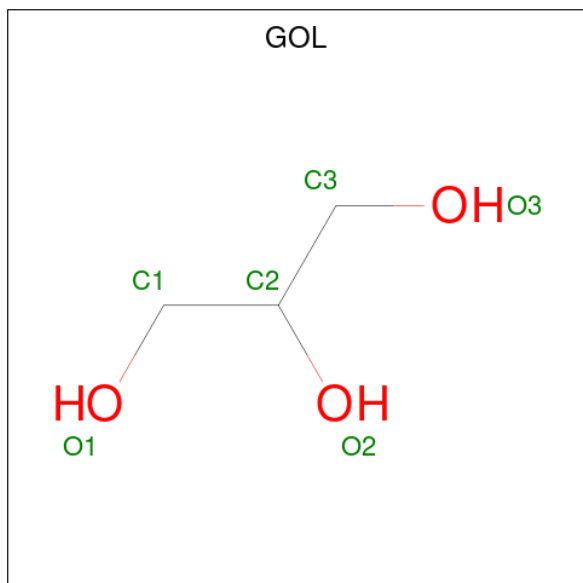
Chain	Residue	Modelled	Actual	Comment	Reference
A	16	MET	-	expression tag	UNP Q8FLC3
A	17	SER	-	expression tag	UNP Q8FLC3
A	18	LEU	-	expression tag	UNP Q8FLC3
A	510	GLU	-	expression tag	UNP Q8FLC3
A	511	GLY	-	expression tag	UNP Q8FLC3
A	512	HIS	-	expression tag	UNP Q8FLC3
A	513	HIS	-	expression tag	UNP Q8FLC3
A	514	HIS	-	expression tag	UNP Q8FLC3
A	515	HIS	-	expression tag	UNP Q8FLC3
A	516	HIS	-	expression tag	UNP Q8FLC3
A	517	HIS	-	expression tag	UNP Q8FLC3
B	16	MET	-	expression tag	UNP Q8FLC3
B	17	SER	-	expression tag	UNP Q8FLC3
B	18	LEU	-	expression tag	UNP Q8FLC3
B	510	GLU	-	expression tag	UNP Q8FLC3
B	511	GLY	-	expression tag	UNP Q8FLC3
B	512	HIS	-	expression tag	UNP Q8FLC3
B	513	HIS	-	expression tag	UNP Q8FLC3
B	514	HIS	-	expression tag	UNP Q8FLC3
B	515	HIS	-	expression tag	UNP Q8FLC3
B	516	HIS	-	expression tag	UNP Q8FLC3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	517	HIS	-	expression tag	UNP Q8FLC3
C	16	MET	-	expression tag	UNP Q8FLC3
C	17	SER	-	expression tag	UNP Q8FLC3
C	18	LEU	-	expression tag	UNP Q8FLC3
C	510	GLU	-	expression tag	UNP Q8FLC3
C	511	GLY	-	expression tag	UNP Q8FLC3
C	512	HIS	-	expression tag	UNP Q8FLC3
C	513	HIS	-	expression tag	UNP Q8FLC3
C	514	HIS	-	expression tag	UNP Q8FLC3
C	515	HIS	-	expression tag	UNP Q8FLC3
C	516	HIS	-	expression tag	UNP Q8FLC3
C	517	HIS	-	expression tag	UNP Q8FLC3
D	16	MET	-	expression tag	UNP Q8FLC3
D	17	SER	-	expression tag	UNP Q8FLC3
D	18	LEU	-	expression tag	UNP Q8FLC3
D	510	GLU	-	expression tag	UNP Q8FLC3
D	511	GLY	-	expression tag	UNP Q8FLC3
D	512	HIS	-	expression tag	UNP Q8FLC3
D	513	HIS	-	expression tag	UNP Q8FLC3
D	514	HIS	-	expression tag	UNP Q8FLC3
D	515	HIS	-	expression tag	UNP Q8FLC3
D	516	HIS	-	expression tag	UNP Q8FLC3
D	517	HIS	-	expression tag	UNP Q8FLC3

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

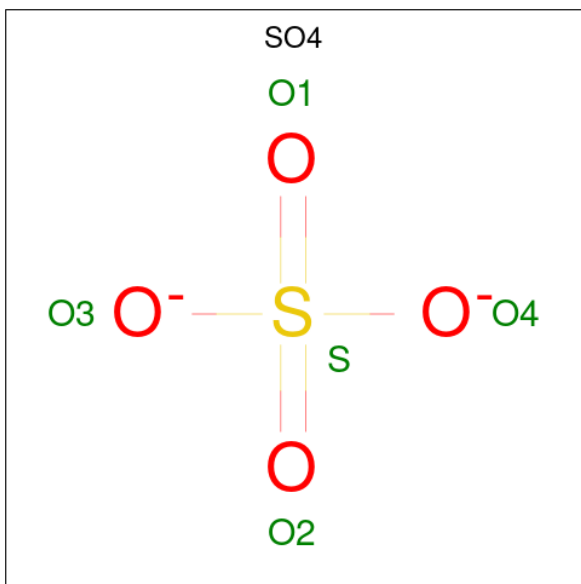


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			6	3	3		
2	B	1	Total	C	O	0	0
			6	3	3		
2	B	1	Total	C	O	0	0
			6	3	3		

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	2	Total	Na	0	0
			2	2		
3	B	1	Total	Na	0	0
			1	1		
3	C	2	Total	Na	0	0
			2	2		
3	D	1	Total	Na	0	0
			1	1		

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



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

Continued on next page...

Continued from previous page...

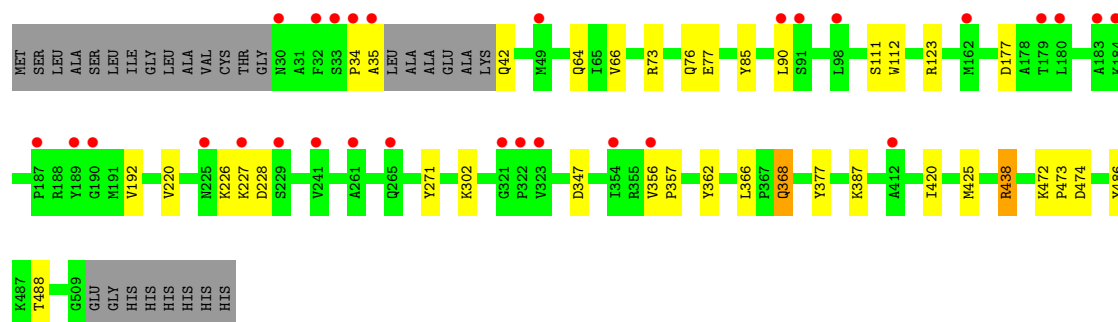
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	D	1	Total O S 5 4 1	0	0

- Molecule 5 is UNKNOWN LIGAND (three-letter code: UNL) (formula:).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C O 4 1 3	0	0
5	B	1	Total C O 4 1 3	0	0
5	C	1	Total C O 4 1 3	0	0
5	D	3	Total C O 12 3 9	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	387	Total O 387 387	0	0
6	B	438	Total O 438 438	0	0
6	C	315	Total O 315 315	0	0
6	D	382	Total O 383 383	0	1



4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	71.93Å 178.86Å 76.28Å 90.00° 92.17° 90.00°	Depositor
Resolution (Å)	20.00 – 1.70 39.74 – 1.70	Depositor EDS
% Data completeness (in resolution range)	87.3 (20.00-1.70) 87.2 (39.74-1.70)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	0.06	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.40 (at 1.70Å)	Xtrriage
Refinement program	REFMAC 5.3.0034	Depositor
R, R_{free}	0.196 , 0.242 0.194 , 0.240	Depositor DCC
R_{free} test set	5541 reflections (3.00%)	wwPDB-VP
Wilson B-factor (Å ²)	29.8	Xtrriage
Anisotropy	0.235	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 45.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.027 for h,-k,-l	Xtrriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	16828	wwPDB-VP
Average B, all atoms (Å ²)	34.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.14% 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: NA, SO4, UNL, GOL

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	0.49	0/3921	0.60	0/5308
1	B	0.48	0/3923	0.60	0/5312
1	C	0.43	0/3851	0.59	1/5216 (0.0%)
1	D	0.47	0/4032	0.64	1/5455 (0.0%)
All	All	0.47	0/15727	0.61	2/21291 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	90	LEU	CA-CB-CG	5.49	127.93	115.30
1	C	90	LEU	CA-CB-CG	5.06	126.94	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	3800	0	3773	9	0
1	B	3807	0	3760	11	0
1	C	3747	0	3683	9	0
1	D	3883	0	3876	14	0
2	A	6	0	8	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	12	0	14	6	0
3	A	2	0	0	0	0
3	B	1	0	0	0	0
3	C	2	0	0	0	0
3	D	1	0	0	0	0
4	A	5	0	0	0	0
4	B	5	0	0	1	0
4	C	5	0	0	0	0
4	D	5	0	0	0	0
5	A	4	0	0	0	0
5	B	4	0	0	0	0
5	C	4	0	0	0	0
5	D	12	0	0	4	0
6	A	387	0	0	1	0
6	B	438	0	0	0	0
6	C	315	0	0	0	0
6	D	383	0	0	1	0
All	All	16828	0	15114	43	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 1.

The worst 5 of 43 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:A:462:ARG:HH21	2:A:1:GOL:H32	1.34	0.88
1:A:239[B]:THR:HG23	6:A:802:HOH:O	1.93	0.67
1:B:375:PRO:HB2	2:B:1:GOL:H31	1.77	0.66
1:B:375:PRO:CB	2:B:1:GOL:H31	2.29	0.63
1:D:34:PRO:HA	1:D:35:ALA:HB3	1.82	0.62

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	478/502 (95%)	462 (97%)	14 (3%)	2 (0%)	34	18
1	B	477/502 (95%)	465 (98%)	10 (2%)	2 (0%)	34	18
1	C	467/502 (93%)	447 (96%)	18 (4%)	2 (0%)	34	18
1	D	489/502 (97%)	476 (97%)	11 (2%)	2 (0%)	34	18
All	All	1911/2008 (95%)	1850 (97%)	53 (3%)	8 (0%)	34	18

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	347	ASP
1	C	347	ASP
1	D	111	SER
1	D	347	ASP
1	A	111	SER

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	408/420 (97%)	394 (97%)	14 (3%)	37	18
1	B	408/420 (97%)	398 (98%)	10 (2%)	47	29
1	C	399/420 (95%)	389 (98%)	10 (2%)	47	29
1	D	419/420 (100%)	404 (96%)	15 (4%)	35	16
All	All	1634/1680 (97%)	1585 (97%)	49 (3%)	43	22

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	142	MET
1	D	66	VAL
1	C	224	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	377	TYR
1	D	226	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	64	GLN
1	C	135	GLN
1	D	492	ASN
1	D	64	GLN
1	D	478	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](#)

Of 19 ligands modelled in this entry, 6 are monoatomic and 6 are unknown - leaving 7 for Mogul analysis.

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	GOL	A	1	-	5,5,5	0.24	0	5,5,5	0.80	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
4	SO4	B	2	-	4,4,4	0.22	0	6,6,6	0.17	0
4	SO4	C	3	-	4,4,4	0.22	0	6,6,6	0.08	0
4	SO4	A	520	-	4,4,4	0.25	0	6,6,6	0.06	0
2	GOL	B	519	-	5,5,5	0.34	0	5,5,5	0.32	0
4	SO4	D	4	-	4,4,4	0.61	0	6,6,6	0.60	0
2	GOL	B	1	-	5,5,5	1.66	2 (40%)	5,5,5	0.96	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GOL	A	1	-	-	1/4/4/4	-
2	GOL	B	1	-	-	2/4/4/4	-
2	GOL	B	519	-	-	4/4/4/4	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1	GOL	O2-C2	-2.63	1.35	1.43
2	B	1	GOL	O3-C3	-2.07	1.33	1.42

There are no bond angle outliers.

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	1	GOL	O2-C2-C3-O3
2	B	1	GOL	C1-C2-C3-O3
2	B	519	GOL	O1-C1-C2-C3
2	B	519	GOL	C1-C2-C3-O3
2	B	519	GOL	O2-C2-C3-O3

There are no ring outliers.

3 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1	GOL	2	0
4	B	2	SO4	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	1	GOL	6	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	468/502 (93%)	0.21	20 (4%) 35 39	18, 30, 51, 94	0
1	B	474/502 (94%)	0.05	12 (2%) 57 61	17, 28, 51, 88	0
1	C	467/502 (93%)	0.47	39 (8%) 11 12	21, 37, 66, 87	0
1	D	474/502 (94%)	0.31	29 (6%) 21 23	19, 30, 52, 84	0
All	All	1883/2008 (93%)	0.26	100 (5%) 26 29	17, 31, 56, 94	0

The worst 5 of 100 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	35	ALA	8.9
1	C	412	ALA	6.5
1	B	34	PRO	5.5
1	C	303	ILE	5.5
1	A	227	LYS	5.2

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
5	UNL	A	521	4/-	0.78	0.15	30,36,44,45	0
5	UNL	C	519	4/-	0.80	0.33	34,52,57,58	0
2	GOL	B	519	6/6	0.85	0.22	48,57,62,70	0
2	GOL	B	1	6/6	0.86	0.19	24,36,59,59	0
5	UNL	D	518	4/-	0.87	0.14	31,42,50,56	0
5	UNL	D	519	4/-	0.87	0.21	52,56,58,59	0
5	UNL	B	520	4/-	0.88	0.19	25,37,44,55	0
2	GOL	A	1	6/6	0.89	0.15	24,46,52,52	0
5	UNL	D	520	4/-	0.93	0.22	54,57,61,62	0
4	SO4	D	4	5/5	0.95	0.52	25,34,47,55	5
4	SO4	A	520	5/5	0.96	0.42	23,26,37,46	5
3	NA	C	518	1/1	0.96	0.20	30,30,30,30	0
3	NA	D	1	1/1	0.97	0.25	27,27,27,27	0
4	SO4	C	3	5/5	0.98	0.20	23,31,33,50	5
3	NA	C	1	1/1	0.98	0.20	19,19,19,19	0
3	NA	B	518	1/1	0.98	0.13	21,21,21,21	0
4	SO4	B	2	5/5	0.98	0.31	24,26,40,41	5
3	NA	A	518	1/1	0.99	0.18	19,19,19,19	0
3	NA	A	519	1/1	0.99	0.11	24,24,24,24	0

6.5 Other polymers [i](#)

There are no such residues in this entry.