wwPDB X-ray Structure Validation Summary Report (i)

Jun 18, 2024 – 01:35 AM EDT

PDB ID	:	2YOC
Title	:	Crystal structure of PulA from Klebsiella oxytoca
Authors	:	Francetic, O.; Mechaly, A.E.; Tello-Manigne, D.; Buschiazzo, A.; Bernarde, C.; Nadeau, N.; Pugsley, A.P.; Alzari, P.M.
Deposited on	:	2012-10-23
Resolution	:	2.88 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

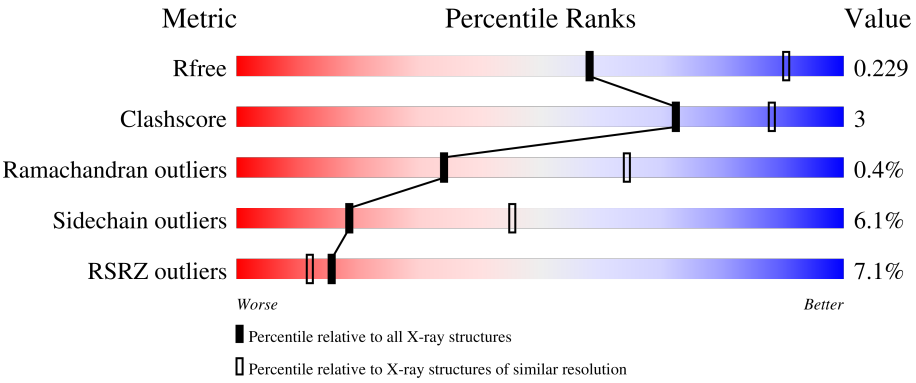
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.88 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	2691 (2.90-2.86)
Clashscore	141614	2947 (2.90-2.86)
Ramachandran outliers	138981	2868 (2.90-2.86)
Sidechain outliers	138945	2871 (2.90-2.86)
RSRZ outliers	127900	2629 (2.90-2.86)

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 <=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	1078	3% 85% 12%
1	B	1078	10% 85% 11%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 16490 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 PULLULANASE.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	1051	Total	C	N	O	S	Se	0	0	0
			8030	5013	1368	1621	6	22			
1	B	1055	Total	C	N	O	S	Se	0	0	0
			8074	5043	1376	1627	6	22			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	expression tag	UNP P07206
A	1071	TYR	-	expression tag	UNP P07206
A	1072	TYR	-	expression tag	UNP P07206
A	1073	HIS	-	expression tag	UNP P07206
A	1074	HIS	-	expression tag	UNP P07206
A	1075	HIS	-	expression tag	UNP P07206
A	1076	HIS	-	expression tag	UNP P07206
A	1077	HIS	-	expression tag	UNP P07206
A	1078	HIS	-	expression tag	UNP P07206
B	1	MSE	-	expression tag	UNP P07206
B	1071	TYR	-	expression tag	UNP P07206
B	1072	TYR	-	expression tag	UNP P07206
B	1073	HIS	-	expression tag	UNP P07206
B	1074	HIS	-	expression tag	UNP P07206
B	1075	HIS	-	expression tag	UNP P07206
B	1076	HIS	-	expression tag	UNP P07206
B	1077	HIS	-	expression tag	UNP P07206
B	1078	HIS	-	expression tag	UNP P07206

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

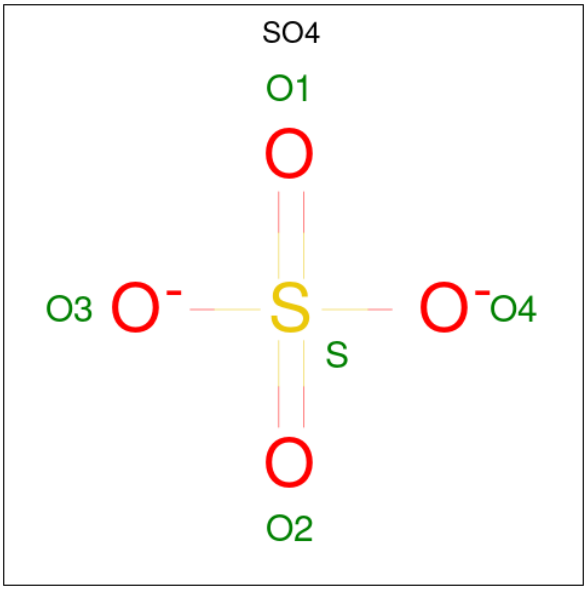
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	4	Total	Ca	0	0
			4	4		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	4	Total Ca 4 4	0	0

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



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

Continued on next page...

Continued from previous page...

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

- Molecule 4 is water.

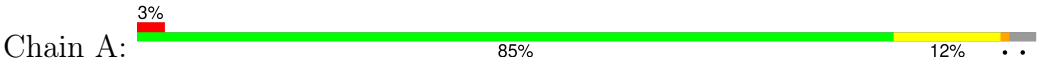
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	145	Total O 145 145	0	0
4	B	163	Total O 163 163	0	0

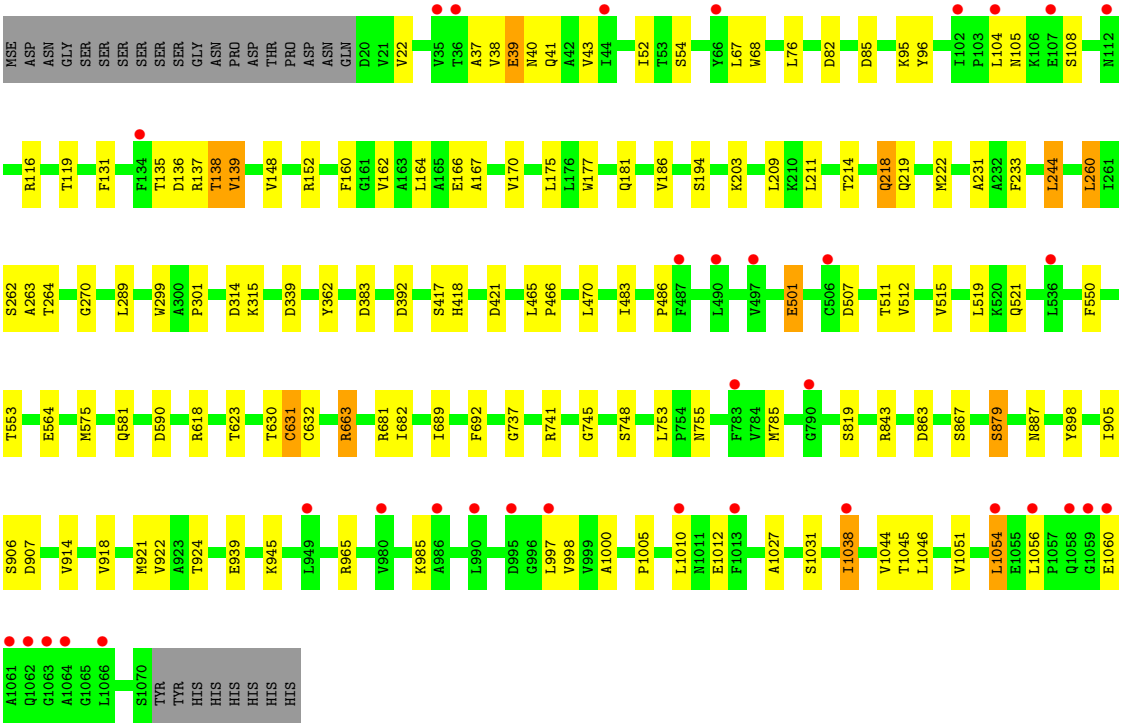
3 Residue-property plots (i)

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: PULLULANASE

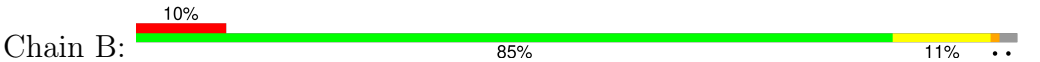
Chain A:

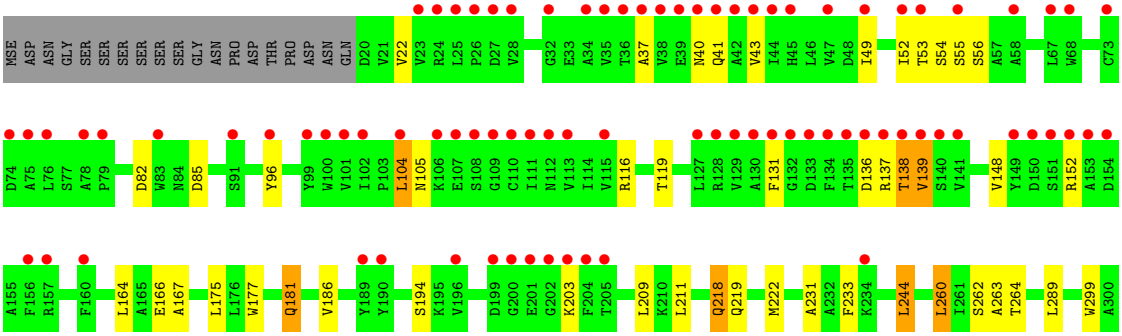


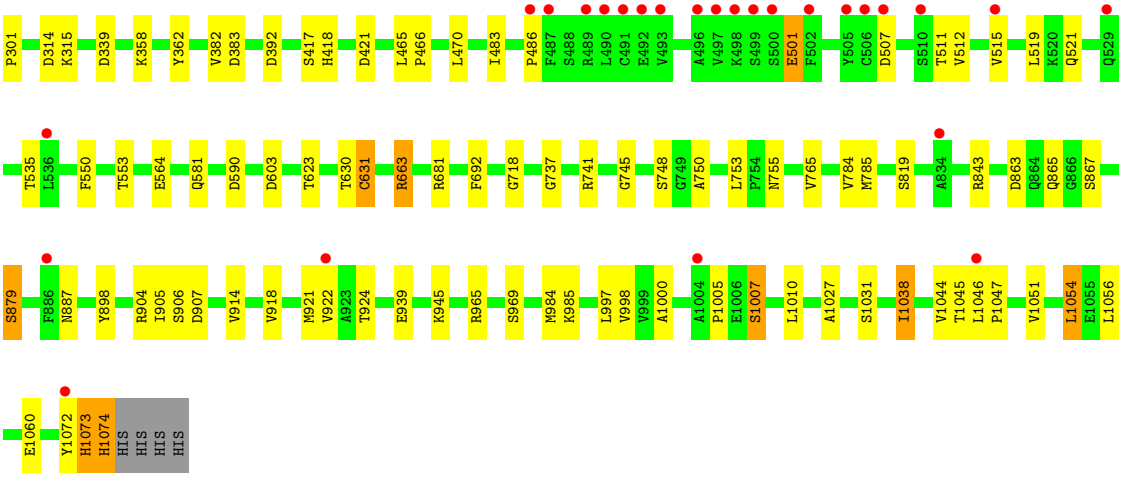


• Molecule 1: PULLULANASE

Chain B:






P301 D314 K315 D339 K358 Y362 V382 D383 D392 S417 H418 D421 L465 P466 L470 I483 P486 F487 S488 R489 L490 C491 E492 V493 A496 V497 K498 S499 S500 E501 F502 Y505 C506 D507 S510 T511 V512 V515 L519 K520 Q521 Q529
T535 L536 F550 T553 E564 Q581 D590 D603 T623 T630 C631 R663 R681 F692 G718 G737 R741 G745 S748 G749 A750 L753 P754 N755 V765 V784 M785 S819 A834 R843 D863 Q864 Q865 G866 S867
S879 F886 N887 Y898 R904 I905 S906 D907 V914 V918 M921 V922 A923 T924 E939 K945 R965 S969 M984 K985 L997 V998 Y999 A1000 A1004 P1005 E1006 S1007 L1010 A1027 S1031 I1038 V1044 T1045 L1046 P1047 V1051 L1054 E1055 L1056
E1060 Y1072 H1073 H1074 HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 63 2 2	Depositor
Cell constants a, b, c, α, β, γ	179.15Å 179.15Å 334.69Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	42.17 – 2.88 58.64 – 2.88	Depositor EDS
% Data completeness (in resolution range)	99.5 (42.17-2.88) 99.8 (58.64-2.88)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.15 (at 2.86Å)	Xtriage
Refinement program	BUSTER 2.11.1	Depositor
R, R_{free}	0.171 , 0.213 0.187 , 0.229	Depositor DCC
R_{free} test set	3635 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	69.0	Xtriage
Anisotropy	0.132	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 73.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	16490	wwPDB-VP
Average B, all atoms (Å^2)	81.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.37% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CA, 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	0.50	0/8171	0.74	2/11086 (0.0%)
1	B	0.50	0/8219	0.75	4/11152 (0.0%)
All	All	0.50	0/16390	0.74	6/22238 (0.0%)

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	631	CYS	N-CA-C	-5.71	95.59	111.00
1	A	631	CYS	N-CA-C	-5.64	95.76	111.00
1	B	53	THR	C-N-CA	5.59	135.68	121.70
1	B	1073	HIS	C-N-CA	5.31	134.97	121.70
1	A	630	THR	C-N-CA	5.21	134.72	121.70

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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	8030	0	7714	56	0
1	B	8074	0	7746	52	0
2	A	4	0	0	0	0
2	B	4	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	20	0	0	0	0
3	B	50	0	0	0	0
4	A	145	0	0	0	0
4	B	163	0	0	0	0
All	All	16490	0	15460	106	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 3.

The worst 5 of 106 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:52:ILE:HD11	1:A:160:PHE:CZ	2.02	0.93
1:A:52:ILE:HD11	1:A:160:PHE:HZ	1.38	0.84
1:A:54:SER:O	1:A:95:LYS:NZ	2.24	0.70
1:B:483:ILE:HA	1:B:512:VAL:HB	1.80	0.63
1:A:483:ILE:HA	1:A:512:VAL:HB	1.81	0.63

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	1049/1078 (97%)	987 (94%)	58 (6%)	4 (0%)	34	64
1	B	1053/1078 (98%)	986 (94%)	63 (6%)	4 (0%)	34	64
All	All	2102/2156 (98%)	1973 (94%)	121 (6%)	8 (0%)	34	64

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	631	CYS
1	B	631	CYS
1	A	466	PRO
1	B	466	PRO
1	A	501	GLU

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	861/863 (100%)	811 (94%)	50 (6%)	20	48
1	B	865/863 (100%)	809 (94%)	56 (6%)	17	42
All	All	1726/1726 (100%)	1620 (94%)	106 (6%)	18	45

5 of 106 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	139	VAL
1	B	392	ASP
1	B	1038	ILE
1	B	181	GLN
1	B	262	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 22 ligands modelled in this entry, 8 are monoatomic - leaving 14 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	$\#\|Z\| > 2$
3	SO4	B	2082	-	4,4,4	0.40	0	6,6,6	0.25	0
3	SO4	B	2081	-	4,4,4	0.56	0	6,6,6	0.16	0
3	SO4	B	2083	-	4,4,4	0.44	0	6,6,6	0.11	0
3	SO4	B	2079	-	4,4,4	0.42	0	6,6,6	0.17	0
3	SO4	B	2086	-	4,4,4	0.36	0	6,6,6	0.19	0
3	SO4	B	2087	-	4,4,4	0.44	0	6,6,6	0.22	0
3	SO4	B	2088	-	4,4,4	0.29	0	6,6,6	0.11	0
3	SO4	B	2085	-	4,4,4	0.39	0	6,6,6	0.24	0
3	SO4	B	2080	-	4,4,4	0.42	0	6,6,6	0.20	0
3	SO4	B	2084	-	4,4,4	0.48	0	6,6,6	0.16	0
3	SO4	A	2076	-	4,4,4	0.41	0	6,6,6	0.35	0
3	SO4	A	2077	-	4,4,4	0.34	0	6,6,6	0.19	0
3	SO4	A	2075	-	4,4,4	0.57	0	6,6,6	0.34	0
3	SO4	A	2078	-	4,4,4	0.49	0	6,6,6	0.15	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.

No monomer is involved in short contacts.

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

6.1 Protein, DNA and RNA chains

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^{th} 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	1029/1078 (95%)	0.17	35 (3%) 45 40	43, 77, 128, 162	0
1	B	1033/1078 (95%)	0.52	111 (10%) 6 4	41, 74, 141, 170	0
All	All	2062/2156 (95%)	0.35	146 (7%) 16 12	41, 76, 136, 170	0

The worst 5 of 146 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	129	VAL	9.5
1	B	131	PHE	6.9
1	B	43	VAL	6.4
1	B	101	VAL	6.4
1	B	108	SER	6.2

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

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, 95^{th} 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(Å^2)	Q<0.9
2	CA	A	2074	1/1	0.62	0.26	106,106,106,106	0
3	SO4	B	2086	5/5	0.73	0.29	156,161,162,162	0
3	SO4	B	2083	5/5	0.80	0.31	156,159,161,162	0
3	SO4	A	2078	5/5	0.83	0.51	154,158,159,159	0
3	SO4	B	2081	5/5	0.83	0.32	145,149,151,151	0
2	CA	B	2078	1/1	0.84	0.43	106,106,106,106	0
3	SO4	B	2084	5/5	0.86	0.20	147,152,152,154	0
3	SO4	B	2082	5/5	0.86	0.37	141,145,146,147	0
3	SO4	B	2088	5/5	0.88	0.28	142,146,147,147	0
3	SO4	A	2076	5/5	0.89	0.26	130,133,136,137	0
3	SO4	B	2087	5/5	0.89	0.25	131,135,136,137	0
2	CA	A	2073	1/1	0.89	0.15	123,123,123,123	0
3	SO4	B	2080	5/5	0.90	0.18	130,133,135,137	0
3	SO4	A	2077	5/5	0.93	0.17	135,139,141,142	0
2	CA	A	2072	1/1	0.94	0.47	140,140,140,140	0
2	CA	B	2075	1/1	0.95	0.17	58,58,58,58	0
3	SO4	A	2075	5/5	0.95	0.15	78,82,83,85	0
3	SO4	B	2085	5/5	0.95	0.11	124,128,130,130	0
2	CA	B	2076	1/1	0.96	0.12	106,106,106,106	0
3	SO4	B	2079	5/5	0.98	0.10	106,111,111,112	0
2	CA	A	2071	1/1	0.98	0.10	51,51,51,51	0
2	CA	B	2077	1/1	0.99	0.13	78,78,78,78	0

6.5 Other polymers (i)

There are no such residues in this entry.