



wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 10, 2023 – 05:12 pm GMT

PDB ID : 2J5G
Title : The Native structure of a beta-Diketone Hydrolase from the Cyanobacterium Anabaena sp. PCC 7120
Authors : Bennett, J.P.; Whittingham, J.L.; Brzozowski, A.M.; Leonard, P.M.; Grogan, G.
Deposited on : 2006-09-18
Resolution : 1.46 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 : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.36
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

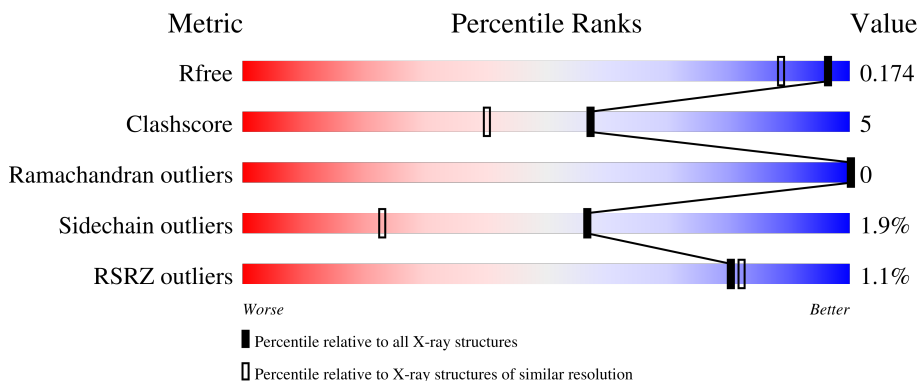
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.46 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	1156 (1.46-1.46)
Clashscore	141614	1202 (1.46-1.46)
Ramachandran outliers	138981	1178 (1.46-1.46)
Sidechain outliers	138945	1178 (1.46-1.46)
RSRZ outliers	127900	1139 (1.46-1.46)

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	263	 89% 5% • 5%
1	B	263	 86% 8% • 5%
1	C	263	 87% 7% • 5%
1	E	263	 87% 6% • 5%
1	F	263	 87% 6% • 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	263	<p>85% 8% • 5%</p>
1	H	263	<p>83% 11% • 5%</p>
1	I	263	<p>84% 10% • 5%</p>
1	J	263	<p>83% 10% • 5%</p>
1	K	263	<p>84% 9% • 6%</p>
1	L	263	<p>85% 8% • 5%</p>
2	D	263	<p>86% 8% • 5%</p>

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 28705 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 ALR4455 PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	249	2018	1290	351	373	4	0	4	1
1	B	250	2032	1298	354	376	4	0	4	1
1	C	249	2029	1297	353	375	4	0	5	1
1	E	249	2028	1295	352	377	4	0	4	0
1	F	249	2021	1291	352	374	4	0	4	1
1	G	249	2021	1291	352	374	4	0	4	1
1	H	249	2021	1291	352	374	4	0	4	1
1	I	249	2021	1291	352	374	4	0	4	1
1	J	249	2021	1291	352	374	4	0	4	1
1	K	248	2020	1291	351	374	4	0	4	0
1	L	249	2021	1291	352	374	4	0	4	1

- Molecule 2 is a protein called ALR4455 PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	D	249	2026	1294	352	376	4	0	4	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	253	ASN	THR	conflict	UNP Q8YNV6

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	S	0	0
			5	4	1		
3	D	1	Total	O	S	0	0
			5	4	1		
3	I	1	Total	O	S	0	0
			5	4	1		
3	J	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	412	Total	O	0	0
			412	412		
4	B	393	Total	O	0	0
			393	393		
4	C	366	Total	O	0	0
			366	366		
4	D	366	Total	O	0	0
			366	366		
4	E	384	Total	O	0	0
			384	384		
4	F	358	Total	O	0	0
			358	358		
4	G	375	Total	O	0	0
			375	375		

Continued on next page...

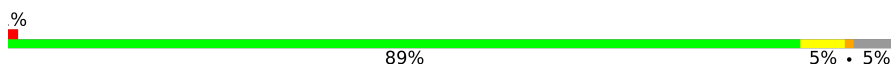
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	H	334	Total 334	O 334	0	0
4	I	371	Total 371	O 371	0	0
4	J	363	Total 363	O 363	0	0
4	K	331	Total 331	O 331	0	0
4	L	353	Total 353	O 353	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.

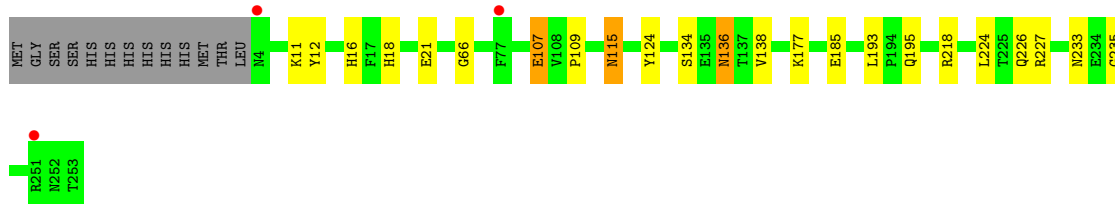
- Molecule 1: ALR4455 PROTEIN

Chain A: 




- Molecule 1: ALR4455 PROTEIN

Chain B: 




- Molecule 1: ALR4455 PROTEIN

Chain C: 




- Molecule 1: ALR4455 PROTEIN

Chain E: 

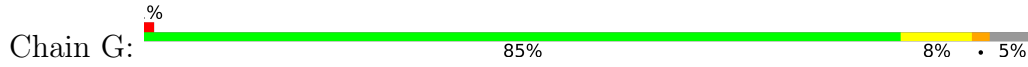


- Molecule 1: ALR4455 PROTEIN

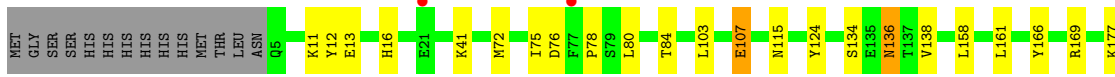
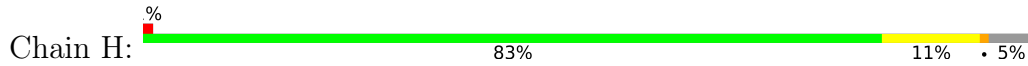
Chain F: 



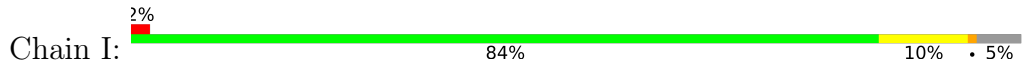
• Molecule 1: ALR4455 PROTEIN



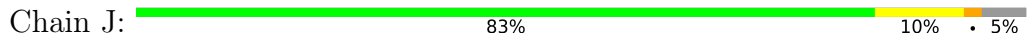
• Molecule 1: ALR4455 PROTEIN



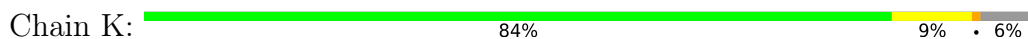
• Molecule 1: ALR4455 PROTEIN

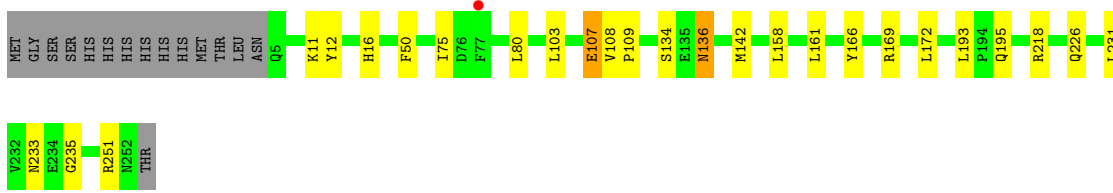


• Molecule 1: ALR4455 PROTEIN

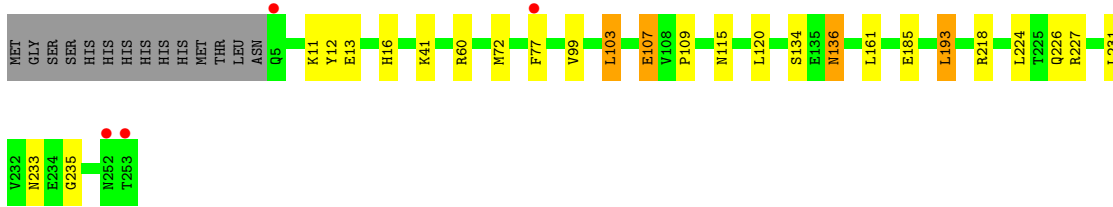
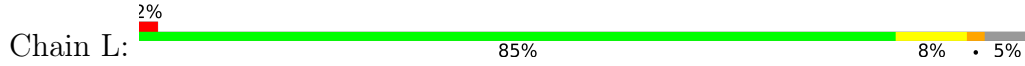


• Molecule 1: ALR4455 PROTEIN

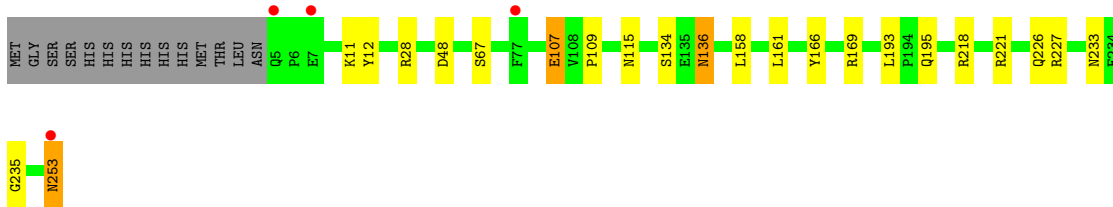
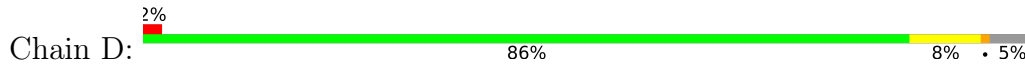




• Molecule 1: ALR4455 PROTEIN



• Molecule 2: ALR4455 PROTEIN



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	118.00Å 83.19Å 154.01Å 90.00° 90.08° 90.00°	Depositor
Resolution (Å)	154.30 – 1.46 46.86 – 1.46	Depositor EDS
% Data completeness (in resolution range)	98.8 (154.30-1.46) 83.1 (46.86-1.46)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.82 (at 1.46Å)	Xtrriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.150 , 0.177 0.148 , 0.174	Depositor DCC
R_{free} test set	21462 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å ²)	11.0	Xtrriage
Anisotropy	0.223	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 56.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.010 for h,-k,-l	Xtrriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	28705	wwPDB-VP
Average B, all atoms (Å ²)	13.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 68.93 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 4.0269e-06. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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	0.67	0/2077	0.81	3/2826 (0.1%)
1	B	0.70	0/2089	0.82	3/2843 (0.1%)
1	C	0.65	0/2085	0.82	1/2837 (0.0%)
1	E	0.68	0/2084	0.81	2/2834 (0.1%)
1	F	0.68	0/2077	0.81	1/2826 (0.0%)
1	G	0.67	0/2077	0.81	2/2826 (0.1%)
1	H	0.64	0/2077	0.76	3/2826 (0.1%)
1	I	0.68	0/2077	0.81	3/2826 (0.1%)
1	J	0.69	0/2077	0.80	2/2826 (0.1%)
1	K	0.64	0/2076	0.77	2/2824 (0.1%)
1	L	0.66	0/2077	0.80	3/2826 (0.1%)
2	D	0.69	0/2085	0.82	5/2835 (0.2%)
All	All	0.67	0/24958	0.80	30/33955 (0.1%)

There are no bond length outliers.

The worst 5 of 30 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	J	227	ARG	NE-CZ-NH2	-6.84	116.88	120.30
1	J	28	ARG	NE-CZ-NH2	-6.75	116.93	120.30
1	L	60	ARG	NE-CZ-NH2	-6.38	117.11	120.30
1	L	227	ARG	NE-CZ-NH2	-6.28	117.16	120.30
2	D	227	ARG	NE-CZ-NH2	-5.95	117.33	120.30

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	2018	0	2004	17	0
1	B	2032	0	2008	21	0
1	C	2029	0	2013	19	0
1	E	2028	0	2010	27	0
1	F	2021	0	2003	22	0
1	G	2021	0	2003	20	0
1	H	2021	0	2003	23	0
1	I	2021	0	2003	25	0
1	J	2021	0	2003	27	0
1	K	2020	0	2003	21	0
1	L	2021	0	2003	24	0
2	D	2026	0	2010	20	0
3	A	5	0	0	0	0
3	D	5	0	0	0	0
3	I	5	0	0	0	0
3	J	5	0	0	0	0
4	A	412	0	0	3	0
4	B	393	0	0	9	0
4	C	366	0	0	8	0
4	D	366	0	0	6	1
4	E	384	0	0	4	0
4	F	358	0	0	8	0
4	G	375	0	0	5	0
4	H	334	0	0	7	0
4	I	371	0	0	10	1
4	J	363	0	0	10	0
4	K	331	0	0	8	0
4	L	353	0	0	7	0
All	All	28705	0	24066	239	1

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 5.

The worst 5 of 239 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:C:88:GLU:HG2	4:C:2195:HOH:O	1.53	1.05
1:E:138[A]:VAL:CG1	1:E:177:LYS:HG2	1.91	1.01
1:I:21:GLU:HG3	4:I:2060:HOH:O	1.66	0.94
1:E:233[A]:ASN:ND2	4:E:2348:HOH:O	2.03	0.91
1:H:233[B]:ASN:ND2	4:H:2317:HOH:O	2.04	0.90

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:D:2078:HOH:O	4:I:2202:HOH:O[2_656]	2.04	0.16

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	251/263 (95%)	250 (100%)	1 (0%)	0	100	100
1	B	252/263 (96%)	251 (100%)	1 (0%)	0	100	100
1	C	252/263 (96%)	251 (100%)	1 (0%)	0	100	100
1	E	251/263 (95%)	249 (99%)	2 (1%)	0	100	100
1	F	251/263 (95%)	250 (100%)	1 (0%)	0	100	100
1	G	251/263 (95%)	250 (100%)	1 (0%)	0	100	100
1	H	251/263 (95%)	249 (99%)	2 (1%)	0	100	100
1	I	251/263 (95%)	249 (99%)	2 (1%)	0	100	100
1	J	251/263 (95%)	250 (100%)	1 (0%)	0	100	100
1	K	250/263 (95%)	247 (99%)	3 (1%)	0	100	100
1	L	251/263 (95%)	250 (100%)	1 (0%)	0	100	100
2	D	251/263 (95%)	250 (100%)	1 (0%)	0	100	100
All	All	3013/3156 (96%)	2996 (99%)	17 (1%)	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	218/228 (96%)	214 (98%)	4 (2%)	59	26
1	B	219/228 (96%)	216 (99%)	3 (1%)	67	37
1	C	219/228 (96%)	216 (99%)	3 (1%)	67	37
1	E	219/228 (96%)	215 (98%)	4 (2%)	59	26
1	F	218/228 (96%)	213 (98%)	5 (2%)	50	17
1	G	218/228 (96%)	213 (98%)	5 (2%)	50	17
1	H	218/228 (96%)	214 (98%)	4 (2%)	59	26
1	I	218/228 (96%)	214 (98%)	4 (2%)	59	26
1	J	218/228 (96%)	212 (97%)	6 (3%)	43	11
1	K	218/228 (96%)	216 (99%)	2 (1%)	78	57
1	L	218/228 (96%)	213 (98%)	5 (2%)	50	17
2	D	219/228 (96%)	216 (99%)	3 (1%)	67	37
All	All	2620/2736 (96%)	2572 (98%)	48 (2%)	57	26

5 of 48 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	H	107	GLU
1	J	32	ASN
1	H	136	ASN
1	I	107	GLU
1	J	107	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 76 such sidechains are listed below:

Mol	Chain	Res	Type
1	J	5	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	L	115	ASN
1	J	115	ASN
1	K	115	ASN
1	L	226	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](#)

4 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	$\# Z > 2$
3	SO4	I	1253	-	4,4,4	0.42	0	6,6,6	1.51	0
3	SO4	A	1253	-	4,4,4	0.31	0	6,6,6	1.37	0
3	SO4	J	1253	-	4,4,4	0.28	0	6,6,6	1.44	0
3	SO4	D	1254	-	4,4,4	0.25	0	6,6,6	1.34	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 [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	249/263 (94%)	-0.39	2 (0%) 86 87	6, 9, 18, 29	0
1	B	250/263 (95%)	-0.41	3 (1%) 79 80	5, 9, 18, 31	0
1	C	249/263 (94%)	-0.33	2 (0%) 86 87	5, 9, 18, 28	0
1	E	249/263 (94%)	-0.33	2 (0%) 86 87	6, 10, 19, 28	0
1	F	249/263 (94%)	-0.31	3 (1%) 79 80	5, 9, 17, 29	0
1	G	249/263 (94%)	-0.32	2 (0%) 86 87	6, 11, 19, 31	0
1	H	249/263 (94%)	-0.28	3 (1%) 79 80	7, 11, 21, 33	0
1	I	249/263 (94%)	-0.33	5 (2%) 65 67	5, 10, 20, 30	0
1	J	249/263 (94%)	-0.43	1 (0%) 92 94	5, 10, 18, 23	1 (0%)
1	K	248/263 (94%)	-0.21	1 (0%) 92 94	7, 13, 23, 31	0
1	L	249/263 (94%)	-0.29	4 (1%) 72 72	6, 10, 18, 30	0
2	D	249/263 (94%)	-0.43	4 (1%) 72 72	5, 9, 18, 26	0
All	All	2988/3156 (94%)	-0.34	32 (1%) 80 82	5, 10, 19, 33	1 (0%)

The worst 5 of 32 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	77	PHE	6.3
1	K	77	PHE	5.5
1	J	77	PHE	5.2
1	E	253	THR	5.2
1	A	77	PHE	4.8

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
3	SO4	D	1254	5/5	0.98	0.34	2,3,3,3	5
3	SO4	A	1253	5/5	0.99	0.32	2,2,3,3	5
3	SO4	I	1253	5/5	0.99	0.32	2,3,4,4	5
3	SO4	J	1253	5/5	0.99	0.33	2,3,4,4	5

6.5 Other polymers [i](#)

There are no such residues in this entry.