wwPDB X-ray Structure Validation Summary Report (i)

May 25, 2020 – 06:01 am BST

PDB ID	:	6OI7
Title	:	Se-Met structure of apo- Escherichia coli dGTPase
Authors	:	Calero, G.; Barnes, C.O.; Wu, Y.
Deposited on	:	2019-04-08
Resolution	:	2.90 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 following versions of software and data (see references (1)) were used in the production of this report:

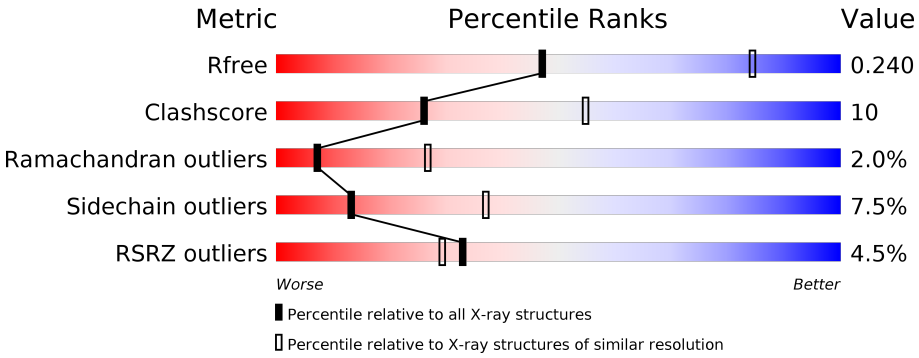
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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 on 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	505	71% 23%
1	B	505	7% 70% 22%
1	C	505	3% 73% 22%
1	D	505	10% 68% 25% 6%
1	E	505	3% 76% 21%
1	F	505	2% 75% 21%

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
3	SO4	A	602	-	-	X	X
3	SO4	C	603	-	-	X	-

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 24823 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.

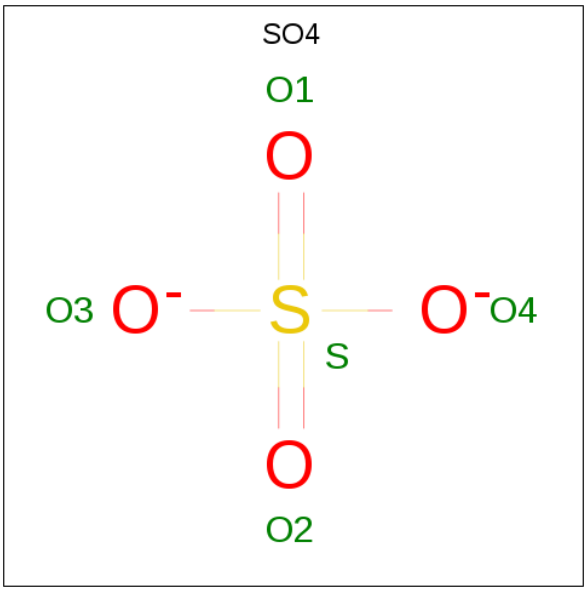
- Molecule 1 is a protein called Deoxyguanosinetriphosphate triphosphohydrolase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	498	Total	C	N	O	S	Se	0	0	0
			4135	2639	737	743	6	10			
1	B	488	Total	C	N	O	S	Se	0	0	0
			4050	2591	721	722	6	10			
1	C	497	Total	C	N	O	S	Se	0	0	0
			4125	2633	736	740	6	10			
1	D	497	Total	C	N	O	S	Se	0	0	0
			4132	2638	737	741	6	10			
1	E	498	Total	C	N	O	S	Se	0	0	0
			4139	2641	738	744	6	10			
1	F	503	Total	C	N	O	S	Se	0	0	0
			4178	2663	747	752	6	10			

- Molecule 2 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	D	1	Total	Mn	0	0
			1	1		
2	E	1	Total	Mn	0	0
			1	1		
2	B	1	Total	Mn	0	0
			1	1		
2	C	1	Total	Mn	0	0
			1	1		
2	A	1	Total	Mn	0	0
			1	1		
2	F	1	Total	Mn	0	0
			1	1		

- 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	B	1	Total O S 5 4 1	0	0
3	C	1	Total O S 5 4 1	0	0
3	C	1	Total O S 5 4 1	0	0
3	C	1	Total O S 5 4 1	0	0
3	C	1	Total O S 5 4 1	0	0
3	D	1	Total O S 5 4 1	0	0
3	E	1	Total O S 5 4 1	0	0
3	F	1	Total O S 5 4 1	0	0

- Molecule 4 is water.

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

Continued on next page...

Continued from previous page...

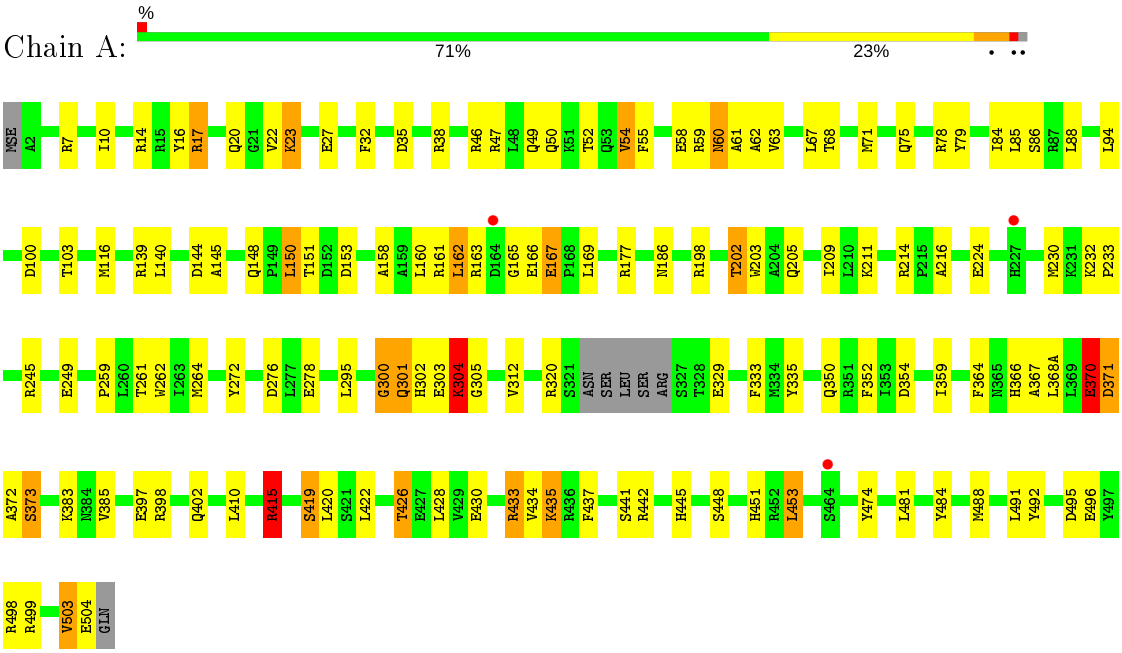
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	E	2	Total O 2 2	0	0
4	F	2	Total O 2 2	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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.

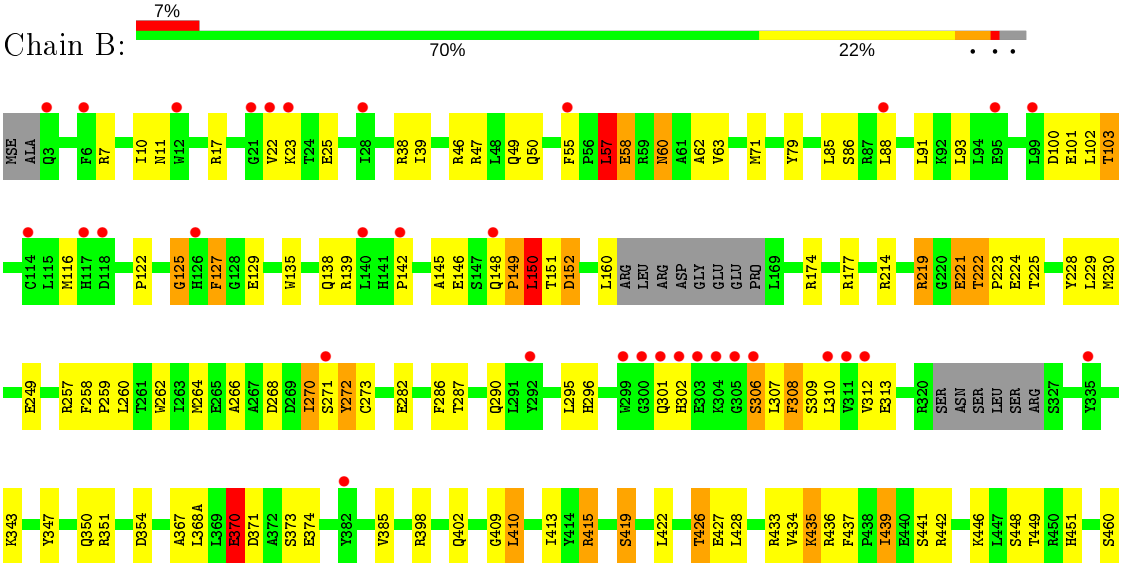
• Molecule 1: Deoxyguanosinetriphosphate triphosphohydrolase

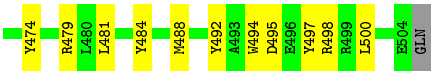
Chain A:



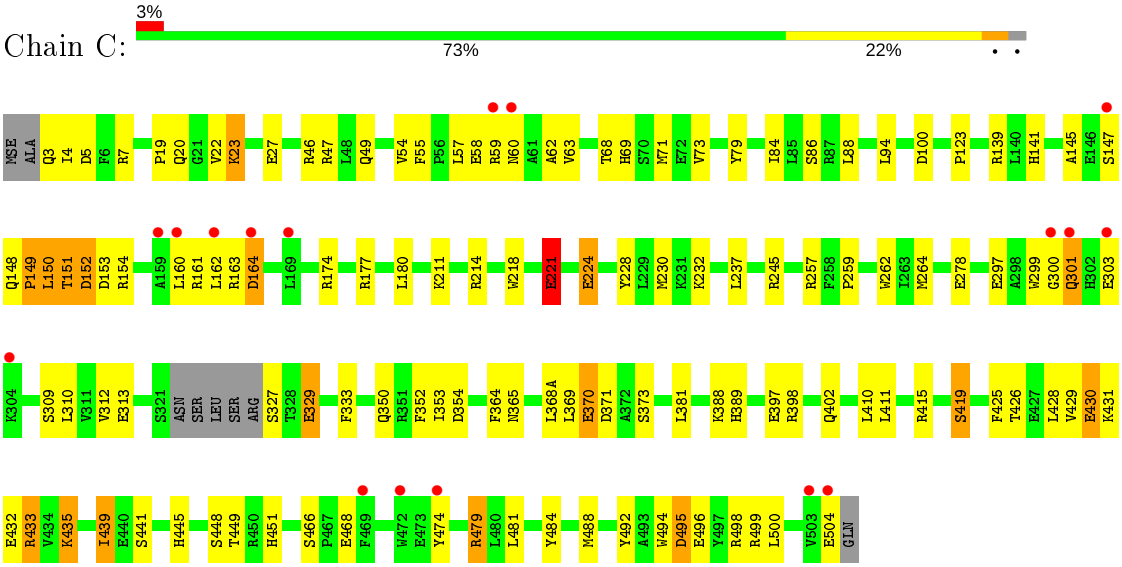
• Molecule 1: Deoxyguanosinetriphosphate triphosphohydrolase

Chain B:

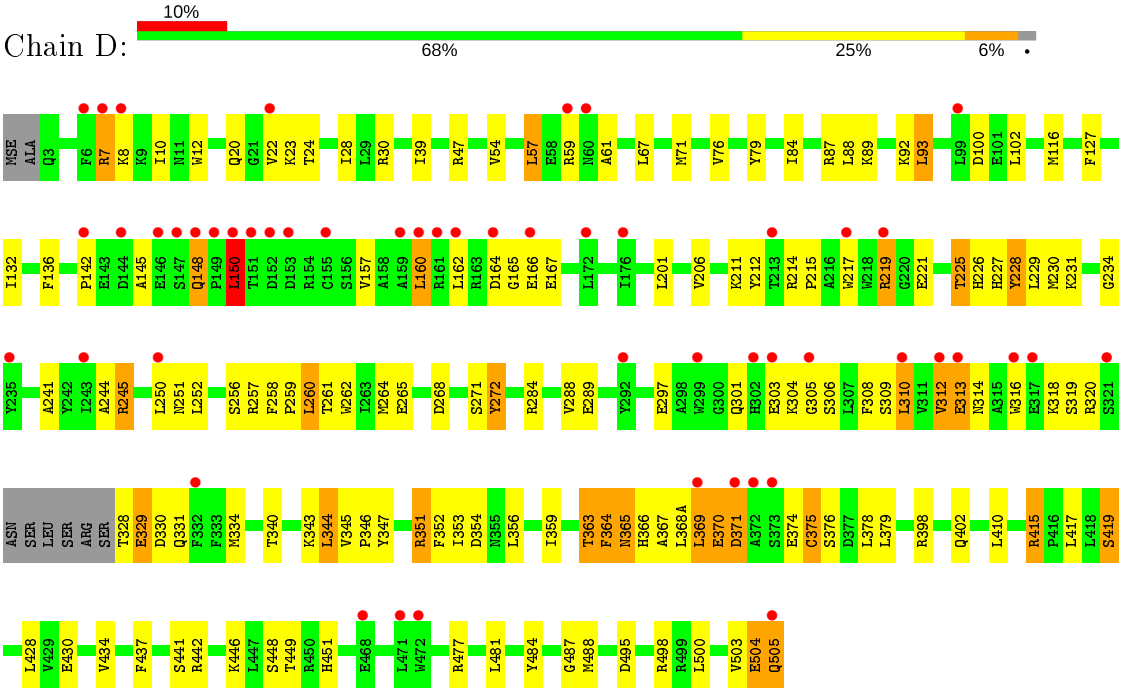




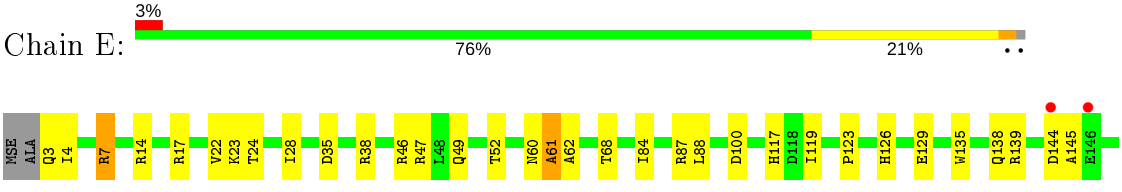
• Molecule 1: Deoxyguanosinetriphosphate triphosphohydrolase

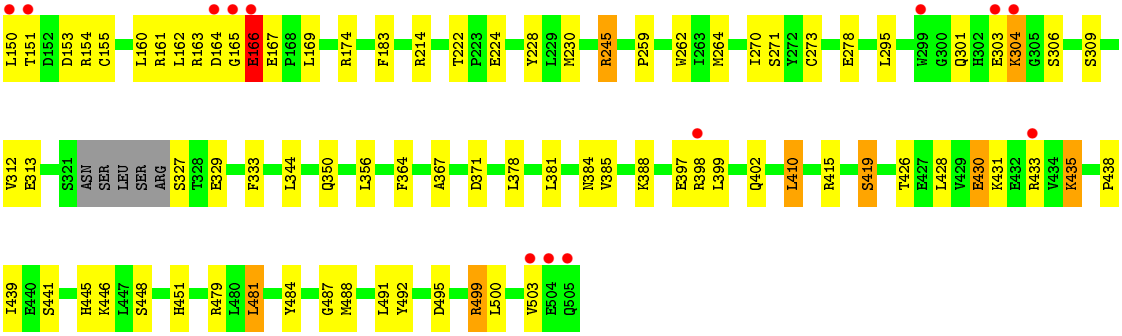


• Molecule 1: Deoxyguanosinetriphosphate triphosphohydrolase

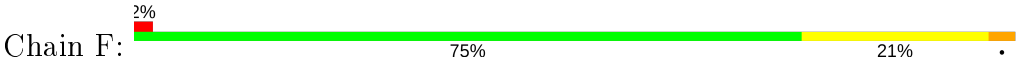


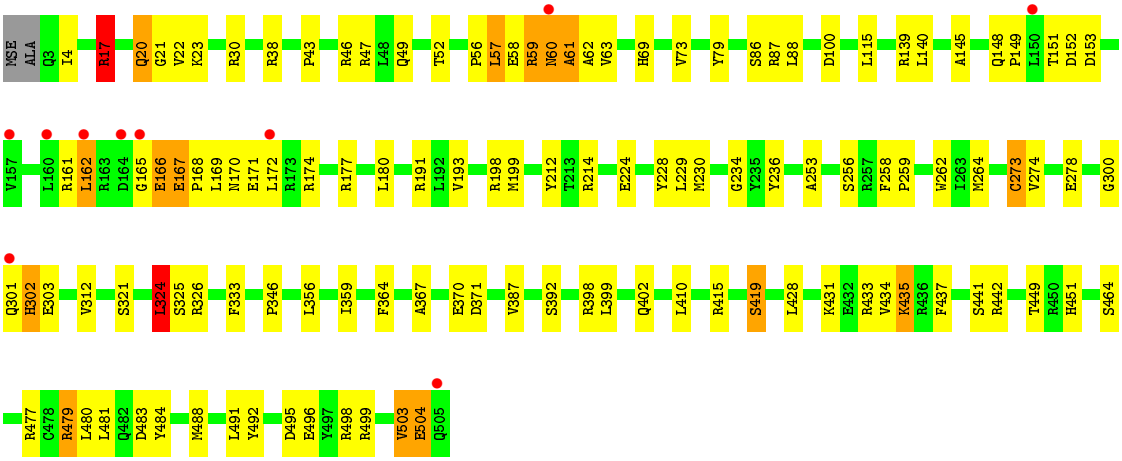
• Molecule 1: Deoxyguanosinetriphosphate triphosphohydrolase





- Molecule 1: Deoxyguanosinetriphosphate triphosphohydrolase





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	192.18Å 192.18Å 287.19Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.97 – 2.90 49.97 – 2.90	Depositor EDS
% Data completeness (in resolution range)	99.9 (49.97-2.90) 99.9 (49.97-2.90)	Depositor EDS
R_{merge}	0.28	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.53 (at 2.91Å)	Xtriage
Refinement program	PHENIX (1.15.1_3469: ???)	Depositor
R, R_{free}	0.185 , 0.234 0.198 , 0.240	Depositor DCC
R_{free} test set	3558 reflections (3.00%)	wwPDB-VP
Wilson B-factor (Å^2)	94.1	Xtriage
Anisotropy	0.178	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 68.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	24823	wwPDB-VP
Average B, all atoms (Å^2)	110.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.42% 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: MN, 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	0.57	4/4227 (0.1%)	0.75	4/5699 (0.1%)
1	B	0.41	0/4140	0.62	2/5581 (0.0%)
1	C	0.53	1/4217 (0.0%)	0.72	3/5685 (0.1%)
1	D	0.43	1/4224 (0.0%)	0.65	3/5695 (0.1%)
1	E	0.53	0/4231	0.70	1/5704 (0.0%)
1	F	0.52	1/4271 (0.0%)	0.72	3/5759 (0.1%)
All	All	0.50	7/25310 (0.0%)	0.69	16/34123 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	C	0	2
1	F	0	1
All	All	0	4

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	54	VAL	C-N	6.86	1.49	1.34
1	A	54	VAL	C-N	6.25	1.48	1.34
1	F	273	CYS	CB-SG	-6.01	1.72	1.82
1	A	415	ARG	CG-CD	5.61	1.66	1.51
1	D	54	VAL	C-N	5.32	1.46	1.34

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	150	LEU	CA-CB-CG	7.99	133.67	115.30
1	F	17	ARG	NE-CZ-NH1	6.91	123.75	120.30
1	B	57	LEU	CA-CB-CG	6.70	130.70	115.30
1	F	57	LEU	CA-CB-CG	6.39	130.00	115.30
1	C	433	ARG	CG-CD-NE	-6.33	98.51	111.80

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	167	GLU	Peptide
1	C	221	GLU	Peptide
1	C	58	GLU	Peptide
1	F	504	GLU	Peptide

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	4135	0	4091	96	1
1	B	4050	0	4011	83	0
1	C	4125	0	4081	85	0
1	D	4132	0	4086	115	1
1	E	4139	0	4094	72	0
1	F	4178	0	4135	98	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	F	1	0	0	0	0
3	A	5	0	0	2	0
3	B	5	0	0	0	0
3	C	20	0	0	4	0
3	D	5	0	0	0	0
3	E	5	0	0	0	0
3	F	5	0	0	0	0
4	A	5	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	C	3	0	0	3	0
4	D	1	0	0	0	0
4	E	2	0	0	3	0
4	F	2	0	0	1	0
All	All	24823	0	24498	514	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 10.

The worst 5 of 514 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:347:TYR:CD1	1:D:375:CYS:SG	2.07	1.48
1:D:347:TYR:CG	1:D:375:CYS:SG	2.15	1.39
1:D:347:TYR:CE1	1:D:375:CYS:SG	2.44	1.10
1:D:57:LEU:HD12	1:D:57:LEU:H	1.19	1.06
1:F:498:ARG:HH21	1:F:503:VAL:HG12	1.24	1.01

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)
1:A:148:GLN:NE2	1:D:374:GLU:OE1[3_544]	2.17	0.03

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	494/505 (98%)	458 (93%)	24 (5%)	12 (2%)	6 22
1	B	482/505 (95%)	435 (90%)	33 (7%)	14 (3%)	4 18

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	C	493/505 (98%)	460 (93%)	24 (5%)	9 (2%)	8 29
1	D	493/505 (98%)	457 (93%)	29 (6%)	7 (1%)	11 36
1	E	494/505 (98%)	464 (94%)	22 (4%)	8 (2%)	9 32
1	F	501/505 (99%)	462 (92%)	29 (6%)	10 (2%)	7 27
All	All	2957/3030 (98%)	2736 (92%)	161 (5%)	60 (2%)	7 27

5 of 60 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	304	LYS
1	A	370	GLU
1	A	371	ASP
1	A	372	ALA
1	B	58	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	443/439 (101%)	412 (93%)	31 (7%)	15 41
1	B	433/439 (99%)	401 (93%)	32 (7%)	13 38
1	C	442/439 (101%)	413 (93%)	29 (7%)	16 44
1	D	442/439 (101%)	396 (90%)	46 (10%)	7 21
1	E	444/439 (101%)	417 (94%)	27 (6%)	18 48
1	F	449/439 (102%)	415 (92%)	34 (8%)	13 36
All	All	2653/2634 (101%)	2454 (92%)	199 (8%)	13 37

5 of 199 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	481	LEU
1	D	284	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	F	303	GLU
1	D	7	ARG
1	D	164	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	296	HIS
1	C	302	HIS
1	D	251	ASN
1	D	365	ASN
1	E	60	ASN

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

There are no carbohydrates in this entry.

5.6 Ligand geometry

Of 15 ligands modelled in this entry, 6 are monoatomic - leaving 9 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	C	602	-	4,4,4	0.15	0	6,6,6	0.58	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
3	SO4	C	605	-	4,4,4	0.21	0	6,6,6	0.27	0
3	SO4	E	602	-	4,4,4	0.14	0	6,6,6	0.22	0
3	SO4	F	602	-	4,4,4	0.13	0	6,6,6	0.09	0
3	SO4	D	602	-	4,4,4	0.16	0	6,6,6	0.13	0
3	SO4	B	602	-	4,4,4	0.14	0	6,6,6	0.17	0
3	SO4	C	603	-	4,4,4	0.17	0	6,6,6	0.45	0
3	SO4	C	604	-	4,4,4	0.21	0	6,6,6	0.42	0
3	SO4	A	602	-	4,4,4	0.15	0	6,6,6	0.10	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.

3 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	605	SO4	1	0
3	C	603	SO4	3	0
3	A	602	SO4	2	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

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	488/505 (96%)	-0.03	3 (0%) 89 89	48, 87, 154, 220	0
1	B	478/505 (94%)	0.39	33 (6%) 16 13	55, 132, 191, 225	0
1	C	487/505 (96%)	0.10	17 (3%) 44 38	50, 88, 141, 198	0
1	D	487/505 (96%)	0.47	52 (10%) 6 4	30, 136, 192, 225	0
1	E	488/505 (96%)	0.13	15 (3%) 49 44	30, 91, 144, 200	0
1	F	493/505 (97%)	0.06	10 (2%) 65 63	30, 93, 154, 231	0
All	All	2921/3030 (96%)	0.18	130 (4%) 33 29	30, 102, 177, 231	0

The worst 5 of 130 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	22	VAL	7.2
1	C	60	ASN	5.2
1	F	505	GLN	5.1
1	D	312	VAL	5.0
1	F	160	LEU	5.0

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 carbohydrates 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, 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
3	SO4	A	602	5/5	0.51	1.07	426,444,448,450	0
2	MN	E	601	1/1	0.70	0.29	123,123,123,123	1
3	SO4	D	602	5/5	0.77	0.34	169,176,178,181	0
2	MN	F	601	1/1	0.78	0.32	101,101,101,101	1
2	MN	C	601	1/1	0.81	0.25	102,102,102,102	1
2	MN	D	601	1/1	0.86	0.31	129,129,129,129	0
3	SO4	F	602	5/5	0.88	0.28	168,172,177,178	0
2	MN	B	601	1/1	0.88	0.26	101,101,101,101	1
3	SO4	E	602	5/5	0.89	0.32	142,143,146,158	0
2	MN	A	601	1/1	0.90	0.23	114,114,114,114	0
3	SO4	B	602	5/5	0.92	0.38	165,166,169,171	0
3	SO4	C	605	5/5	0.93	0.22	109,120,129,137	0
3	SO4	C	603	5/5	0.94	0.12	136,146,150,150	0
3	SO4	C	602	5/5	0.96	0.09	84,90,129,133	0
3	SO4	C	604	5/5	0.99	0.15	83,86,93,104	0

6.5 Other polymers (i)

There are no such residues in this entry.