



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 02:13 PM GMT

PDB ID : 3WFP  
Title : tRNA processing enzyme (apo form 2)  
Authors : Yamashita, S.; Takeshita, D.; Tomita, K.  
Deposited on : 2013-07-23  
Resolution : 4.00 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865



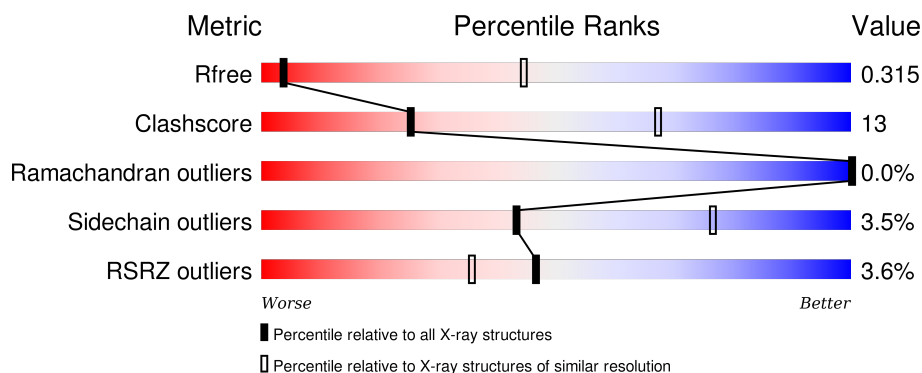
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 4.00 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))
$R_{free}$	91344	1010 (4.42-3.56)
Clashscore	102246	1052 (4.40-3.60)
Ramachandran outliers	100387	1005 (4.40-3.60)
Sidechain outliers	100360	1013 (4.42-3.58)
RSRZ outliers	91569	1013 (4.42-3.56)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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	497	<div> <div> <div></div> <div>65%</div> <div>24%</div> <div>8%</div> </div> </div>
1	B	497	<div> <div> <div></div> <div>67%</div> <div>22%</div> <div>9%</div> </div> </div>
1	C	497	<div> <div> <div></div> <div>63%</div> <div>20%</div> <div>16%</div> </div> </div>
1	D	497	<div> <div> <div></div> <div>64%</div> <div>19%</div> <div>16%</div> </div> </div>
1	E	497	<div> <div> <div></div> <div>61%</div> <div>21%</div> <div>17%</div> </div> </div>

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	497	
1	G	497	
1	H	497	

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	SO4	B	1001	-	-	-	X
2	SO4	C	1001	-	-	X	X
2	SO4	G	1001	-	-	-	X
2	SO4	H	1001	-	-	X	-



## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 27564 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 Poly A polymerase.

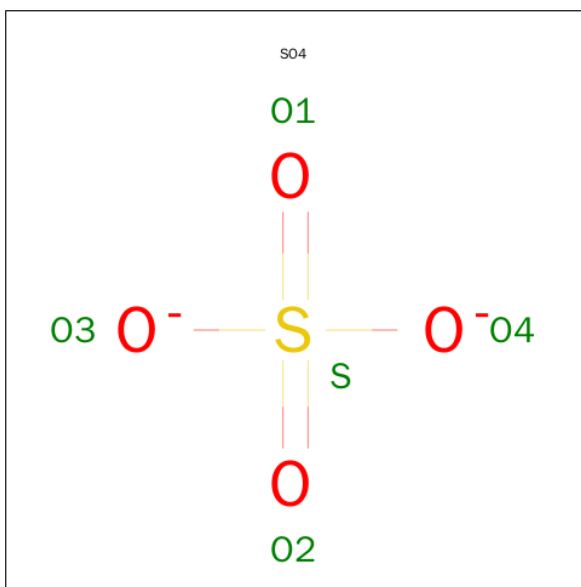
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	455	Total	C	N	O	S	0	0	0
			3624	2334	624	659	7			
1	B	450	Total	C	N	O	S	0	0	0
			3568	2296	611	654	7			
1	C	419	Total	C	N	O	S	0	0	0
			3436	2222	588	619	7			
1	D	419	Total	C	N	O	S	0	0	0
			3436	2222	588	619	7			
1	E	411	Total	C	N	O	S	0	0	0
			3365	2175	572	611	7			
1	F	411	Total	C	N	O	S	0	0	0
			3365	2175	572	611	7			
1	G	411	Total	C	N	O	S	0	0	0
			3365	2175	572	611	7			
1	H	411	Total	C	N	O	S	0	0	0
			3365	2175	572	611	7			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	16	MET	-	EXPRESSION TAG	UNP O67911
B	16	MET	-	EXPRESSION TAG	UNP O67911
C	16	MET	-	EXPRESSION TAG	UNP O67911
D	16	MET	-	EXPRESSION TAG	UNP O67911
E	16	MET	-	EXPRESSION TAG	UNP O67911
F	16	MET	-	EXPRESSION TAG	UNP O67911
G	16	MET	-	EXPRESSION TAG	UNP O67911
H	16	MET	-	EXPRESSION TAG	UNP O67911

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).





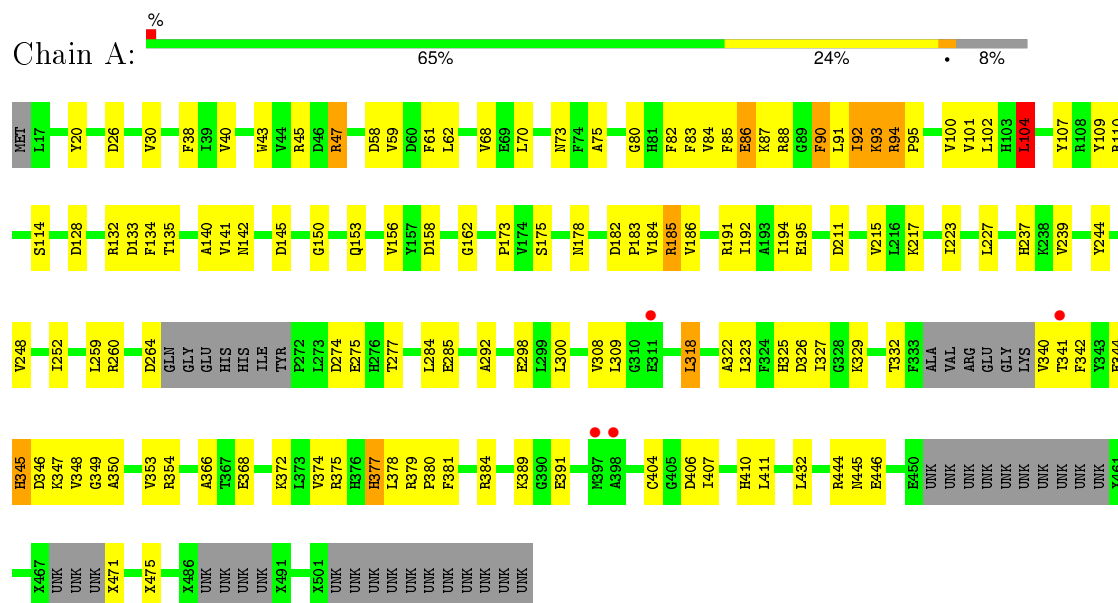
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		
2	C	1	Total	O	S	0	0
			5	4	1		
2	D	1	Total	O	S	0	0
			5	4	1		
2	E	1	Total	O	S	0	0
			5	4	1		
2	F	1	Total	O	S	0	0
			5	4	1		
2	G	1	Total	O	S	0	0
			5	4	1		
2	H	1	Total	O	S	0	0
			5	4	1		



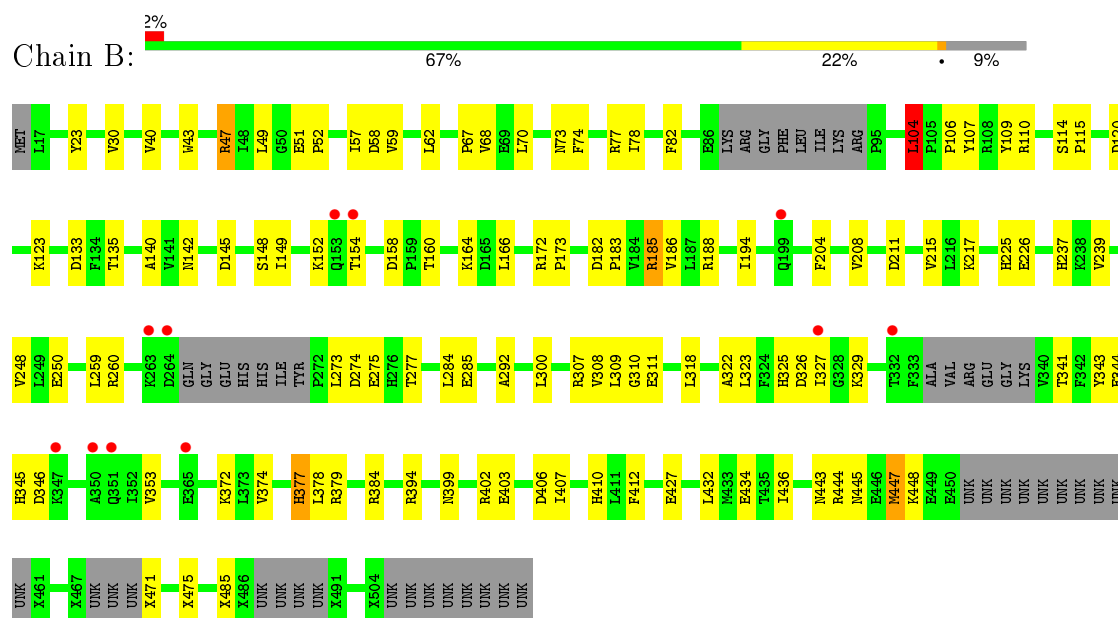
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: Poly A polymerase



#### • Molecule 1: Poly A polymerase





Chain C:

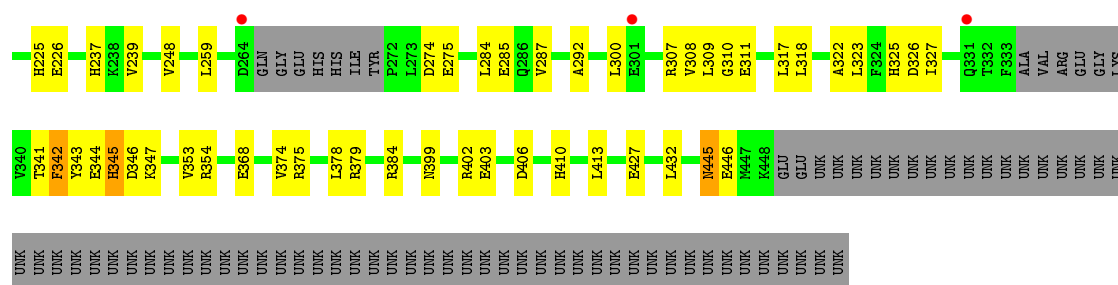
3% 63% 20% 16%

Chain C: MET L17 Y20 V30 L31 P32 R33 E34 H35 Y36 C37 V40 R47 I46 L49 G50 E51 D58 L62 T63 P67 V68 E69 L70 A71 K72 N73 F74 H81 F82 F83 V84 F85 E86 K87 F90 L91 I92 K93 R94 V101 L104 F105 Y107 R110 F111 P112 F113 S114 P115 D120 K123 D133 F134 T135 A140 V141 N142 D145 S148 I149 G150 A151 K152 Q153 T154 I155 V156 Y157 L158 F159 T160 L166 D182 P183 V184 R185 V186 L187 R188 R191 D211 T214 V215 L216 K217 S218 H225 H237 V248 L249 E250 R258 L259 R260 E261 D264 G1N G1Y G1U H1S H1S I1E T1Y P272 L273 D274 E275 L284 E285 Q286 V287 A292 L300 R307 V308 L309 G310 E311 L318 A322 R325 D326 G331 T332 F333 A1A V1A A1G G1U G1Y L1S V340 T341 F342 E344 H345 D346 K347 V348 V259

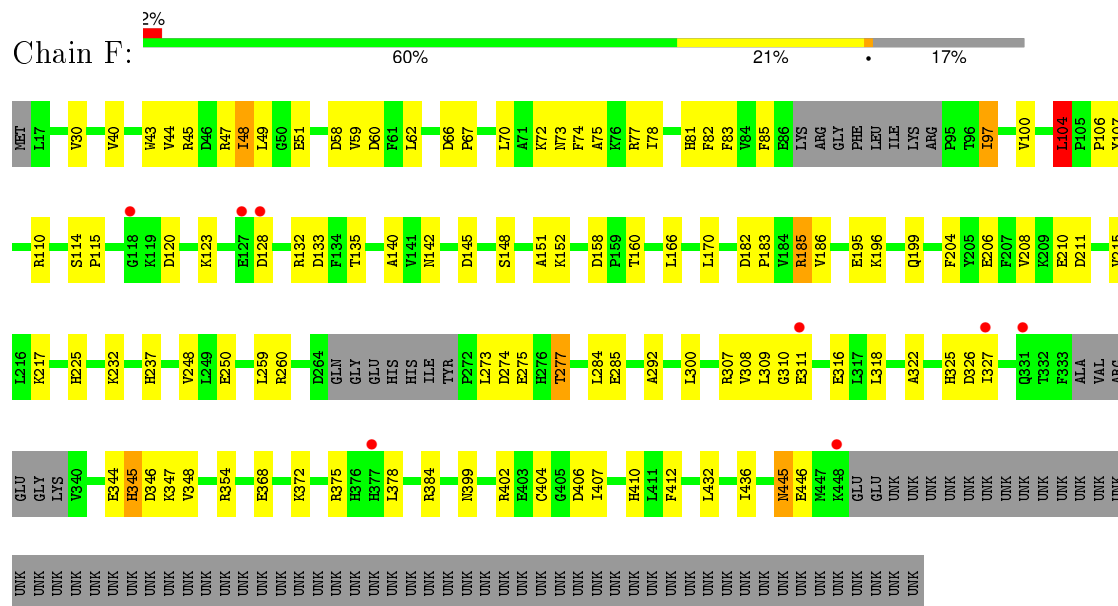
Chain D:

[illegible]

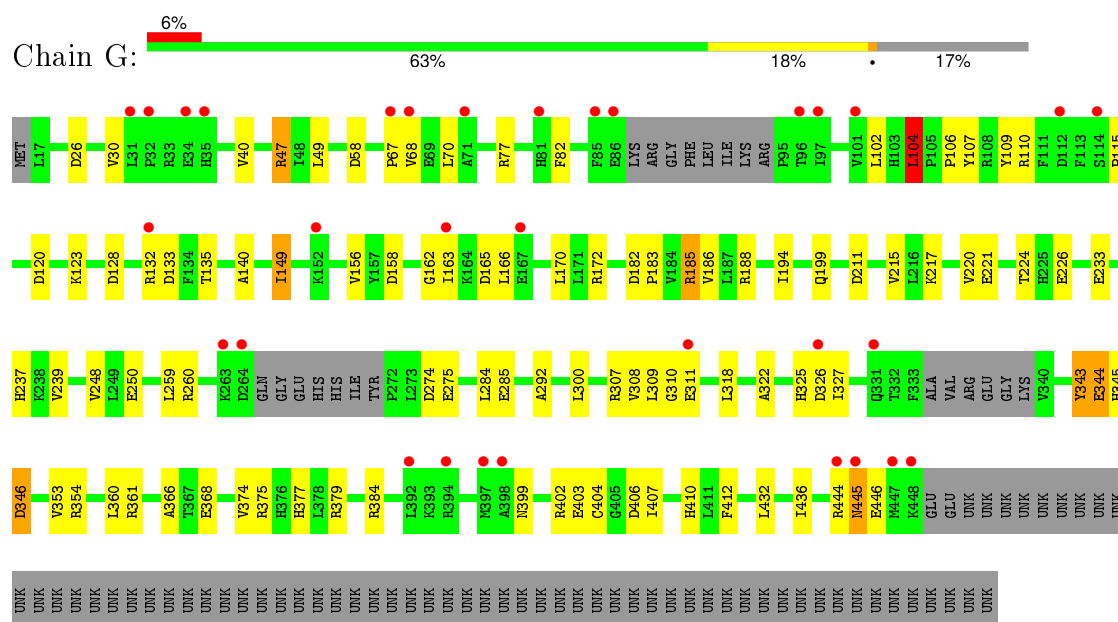




• Molecule 1: Poly A polymerase



• Molecule 1: Poly A polymerase



• Molecule 1: Poly A polymerase







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	126.91Å 138.36Å 147.62Å 90.00° 111.04° 90.00°	Depositor
Resolution (Å)	19.97 – 4.00 48.81 – 4.01	Depositor EDS
% Data completeness (in resolution range)	86.8 (19.97-4.00) 76.7 (48.81-4.01)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.68 (at 4.00Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, $R_{free}$	0.258 , 0.314 0.260 , 0.315	Depositor DCC
$R_{free}$ test set	1755 reflections (5.02%)	DCC
Wilson B-factor (Å <sup>2</sup> )	68.6	Xtriage
Anisotropy	0.055	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 107.0	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.38$ , $\langle L^2 \rangle = 0.21$	Xtriage
Outliers	2 of 34980 reflections (0.006%)	Xtriage
$F_o, F_c$ correlation	0.76	EDS
Total number of atoms	27564	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	142.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 12.68% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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	0.24	0/3527	0.42	1/4750 (0.0%)
1	B	0.23	0/3454	0.42	1/4653 (0.0%)
1	C	0.23	0/3509	0.42	1/4726 (0.0%)
1	D	0.26	0/3509	0.45	2/4726 (0.0%)
1	E	0.22	0/3436	0.40	2/4629 (0.0%)
1	F	0.27	0/3436	0.46	2/4629 (0.0%)
1	G	0.23	0/3436	0.42	2/4629 (0.0%)
1	H	0.22	0/3436	0.40	1/4629 (0.0%)
All	All	0.24	0/27743	0.42	12/37371 (0.0%)

There are no bond length outliers.

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	94	ARG	C-N-CD	5.83	140.64	128.40
1	B	104	LEU	CA-CB-CG	5.72	128.45	115.30
1	G	104	LEU	CA-CB-CG	5.71	128.43	115.30
1	F	104	LEU	CA-CB-CG	5.70	128.41	115.30
1	H	104	LEU	CA-CB-CG	5.70	128.41	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	3624	0	3522	107	0
1	B	3568	0	3439	80	0
1	C	3436	0	3470	104	0
1	D	3436	0	3470	107	0
1	E	3365	0	3384	76	0
1	F	3365	0	3384	96	0
1	G	3365	0	3384	84	0
1	H	3365	0	3384	59	0
2	A	5	0	0	0	0
2	B	5	0	0	1	0
2	C	5	0	0	2	0
2	D	5	0	0	0	0
2	E	5	0	0	1	0
2	F	5	0	0	1	0
2	G	5	0	0	0	0
2	H	5	0	0	2	0
All	All	27564	0	27437	690	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 13.

The worst 5 of 690 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:92:ILE:HD12	1:A:93:LYS:N	1.27	1.41
1:C:345:HIS:HD2	1:C:346:ASP:N	1.30	1.28
1:C:345:HIS:CD2	1:C:346:ASP:H	1.57	1.20
1:A:92:ILE:CD1	1:A:93:LYS:H	1.55	1.19
1:C:345:HIS:CD2	1:C:346:ASP:N	2.10	1.16

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	415/497 (84%)	409 (99%)	6 (1%)	0	100	100
1	B	405/497 (82%)	402 (99%)	2 (0%)	1 (0%)	52	86
1	C	413/497 (83%)	407 (98%)	6 (2%)	0	100	100
1	D	413/497 (83%)	407 (98%)	6 (2%)	0	100	100
1	E	403/497 (81%)	402 (100%)	1 (0%)	0	100	100
1	F	403/497 (81%)	400 (99%)	3 (1%)	0	100	100
1	G	403/497 (81%)	400 (99%)	3 (1%)	0	100	100
1	H	403/497 (81%)	399 (99%)	4 (1%)	0	100	100
All	All	3258/3976 (82%)	3226 (99%)	31 (1%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	345	HIS

### 5.3.2 Protein sidechains ⓘ

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	367/378 (97%)	344 (94%)	23 (6%)	22	61
1	B	360/378 (95%)	348 (97%)	12 (3%)	45	77
1	C	365/378 (97%)	350 (96%)	15 (4%)	37	73
1	D	365/378 (97%)	354 (97%)	11 (3%)	48	78
1	E	358/378 (95%)	346 (97%)	12 (3%)	44	77
1	F	358/378 (95%)	347 (97%)	11 (3%)	47	78
1	G	358/378 (95%)	349 (98%)	9 (2%)	55	82
1	H	358/378 (95%)	349 (98%)	9 (2%)	55	82
All	All	2889/3024 (96%)	2787 (96%)	102 (4%)	43	76

5 of 102 residues with a non-rotameric sidechain are listed below:



Mol	Chain	Res	Type
1	C	211	ASP
1	D	345	HIS
1	H	104	LEU
1	C	345	HIS
1	D	87	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	345	HIS
1	G	377	HIS
1	G	225	HIS
1	C	345	HIS
1	E	345	HIS

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

8 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	A	1001	-	4,4,4	0.25	0	6,6,6	0.08	0
2	SO4	B	1001	-	4,4,4	0.23	0	6,6,6	0.08	0
2	SO4	C	1001	-	4,4,4	0.24	0	6,6,6	0.09	0
2	SO4	D	1001	-	4,4,4	0.23	0	6,6,6	0.08	0
2	SO4	E	1001	-	4,4,4	0.23	0	6,6,6	0.08	0
2	SO4	F	1001	-	4,4,4	0.24	0	6,6,6	0.07	0
2	SO4	G	1001	-	4,4,4	1.48	0	6,6,6	1.71	1 (16%)
2	SO4	H	1001	-	4,4,4	0.24	0	6,6,6	0.08	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	SO4	A	1001	-	-	0/0/0/0	0/0/0/0
2	SO4	B	1001	-	-	0/0/0/0	0/0/0/0
2	SO4	C	1001	-	-	0/0/0/0	0/0/0/0
2	SO4	D	1001	-	-	0/0/0/0	0/0/0/0
2	SO4	E	1001	-	-	0/0/0/0	0/0/0/0
2	SO4	F	1001	-	-	0/0/0/0	0/0/0/0
2	SO4	G	1001	-	-	0/0/0/0	0/0/0/0
2	SO4	H	1001	-	-	0/0/0/0	0/0/0/0

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	G	1001	SO4	O4-S-O3	4.03	125.37	108.98

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	1001	SO4	1	0
2	C	1001	SO4	2	0
2	E	1001	SO4	1	0
2	F	1001	SO4	1	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	H	1001	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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	421/497 (84%)	-0.22	4 (0%) 84 77	51, 113, 166, 211	0
1	B	413/497 (83%)	-0.10	11 (2%) 58 46	62, 116, 177, 218	0
1	C	419/497 (84%)	0.03	15 (3%) 46 36	71, 139, 209, 259	0
1	D	419/497 (84%)	-0.03	7 (1%) 73 62	75, 133, 193, 243	0
1	E	411/497 (82%)	0.03	9 (2%) 65 54	83, 146, 198, 249	0
1	F	411/497 (82%)	-0.08	8 (1%) 70 59	65, 125, 176, 190	0
1	G	411/497 (82%)	0.33	32 (7%) 16 11	118, 167, 231, 259	0
1	H	411/497 (82%)	0.36	32 (7%) 16 11	102, 177, 223, 266	0
All	All	3316/3976 (83%)	0.04	118 (3%) 46 36	51, 140, 209, 266	0

The worst 5 of 118 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	264	ASP	6.4
1	H	264	ASP	5.6
1	C	264	ASP	5.5
1	G	31	LEU	5.3
1	E	180	LYS	4.7

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	SO4	B	1001	5/5	0.91	0.45	1.90	126,131,139,141	0
2	SO4	C	1001	5/5	0.73	0.49	1.48	162,165,166,172	0
2	SO4	H	1001	5/5	0.92	0.39	1.00	152,157,161,166	0
2	SO4	G	1001	5/5	0.92	0.43	0.42	172,174,178,186	0
2	SO4	A	1001	5/5	0.88	0.31	0.16	119,122,146,154	0
2	SO4	E	1001	5/5	0.95	0.24	-0.66	133,135,137,145	0
2	SO4	D	1001	5/5	0.94	0.29	-0.86	104,115,144,156	0
2	SO4	F	1001	5/5	0.94	0.16	-1.15	81,81,91,105	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.