



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 26, 2020 – 04:56 pm BST

PDB ID : 6TYE  
Title : Crystal structure of MTB sigma L transcription initiation complex with 5 nt long RNA primer  
Authors : Molodtsov, V.; Ebright, R.H.  
Deposited on : 2019-08-08  
Resolution : 3.79 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

<https://www.wwpdb.org/validation/2017/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
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

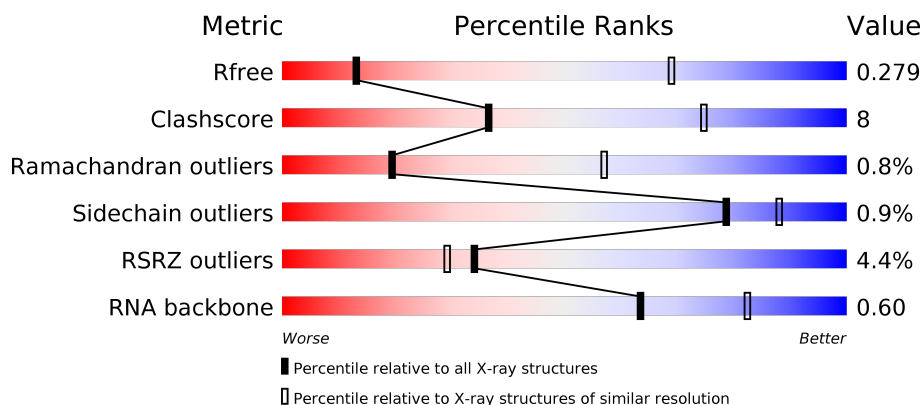
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.79 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	1212 (4.00-3.60)
Clashscore	141614	1288 (4.00-3.60)
Ramachandran outliers	138981	1243 (4.00-3.60)
Sidechain outliers	138945	1237 (4.00-3.60)
RSRZ outliers	127900	1121 (4.00-3.60)
RNA backbone	3102	1036 (4.60-3.00)

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 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	F	177	<div> <div>4%</div> <div> <div></div> <div>71%</div> <div>20%</div> <div>• 8%</div> </div> </div>
2	G	17	<div> <div>41%</div> <div>59%</div> </div>
3	H	27	<div> <div>11%</div> <div> <div></div> <div>48%</div> <div>37%</div> <div>15%</div> </div> </div>
4	I	5	<div> <div>20%</div> <div> <div></div> <div>80%</div> <div>20%</div> </div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	A	347	<div><div>%</div><div><div></div><div>48%</div><div>16%</div><div>35%</div></div></div>
5	B	347	<div><div>%</div><div><div></div><div>49%</div><div>18%</div><div>33%</div></div></div>
6	C	1178	<div><div>%</div><div><div></div><div>77%</div><div>17%</div><div>5%</div></div></div>
7	D	1316	<div><div>8%</div><div><div></div><div>80%</div><div>15%</div></div><div></div></div>
8	E	110	<div><div>5%</div><div><div></div><div>65%</div><div>8%</div><div>26%</div></div></div>

## 2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 49194 atoms, of which 24415 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 RNA polymerase sigma factor.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	F	163	Total	C	H	N	O	S	0	0	0
			2516	780	1259	238	237	2			

- Molecule 2 is a DNA chain called DNA (5'-D(\*GP\*CP\*AP\*TP\*CP\*CP\*GP\*TP\*GP\*AP\*AP\*TP\*CP\*GP\*AP\*GP\*G)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	G	17	Total	C	H	N	O	P	0	0	0
			541	166	192	68	99	16			

- Molecule 3 is a DNA chain called DNA (5'-D(P\*GP\*TP\*GP\*TP\*CP\*AP\*GP\*TP\*AP\*GP\*CP\*TP\*GP\*TP\*CP\*AP\*CP\*GP\*GP\*AP\*TP\*GP\*C)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	H	23	Total	C	H	N	O	P	0	0	0
			734	225	259	87	140	23			

- Molecule 4 is a RNA chain called RNA (5'-R(P\*CP\*UP\*CP\*GP\*A)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
4	I	5	Total	C	H	N	O	P	0	0	0
			160	47	55	18	35	5			

- Molecule 5 is a protein called DNA-directed RNA polymerase subunit alpha.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
5	A	225	Total	C	H	N	O	S	0	0	0
			3472	1080	1756	296	338	2			
5	B	232	Total	C	H	N	O	S	0	0	0
			3486	1093	1754	296	341	2			

- Molecule 6 is a protein called DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
6	C	1114	Total	C	H	N	O	S	0	0	0
			17219	5411	8576	1512	1681	39			

- Molecule 7 is a protein called DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
7	D	1262	Total	C	H	N	O	S	0	0	0
			19814	6182	9942	1790	1860	40			

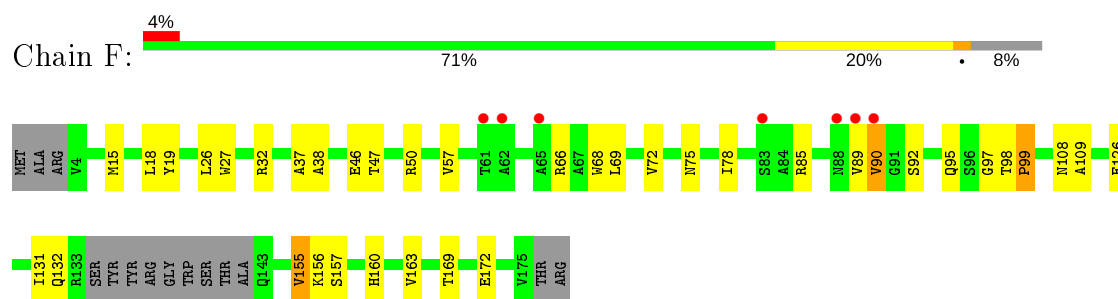
- Molecule 8 is a protein called DNA-directed RNA polymerase subunit omega.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	E	81	Total	C	H	N	O	0	0	0
			1252	403	622	106	121			

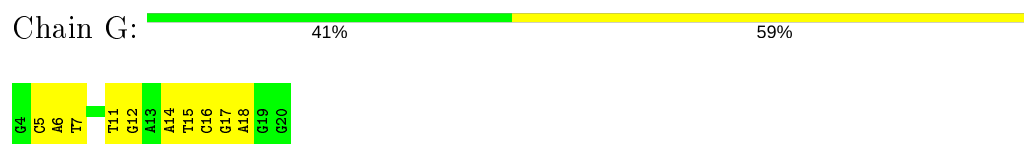
### 3 Residue-property plots [i](#)

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.

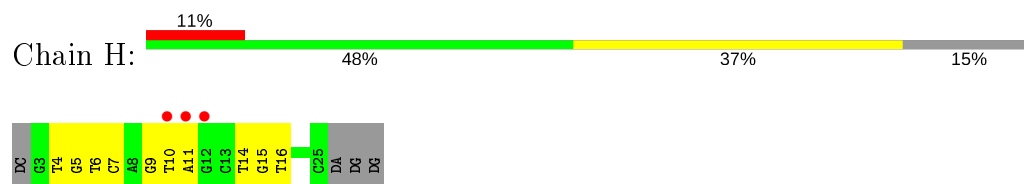
- Molecule 1: RNA polymerase sigma factor



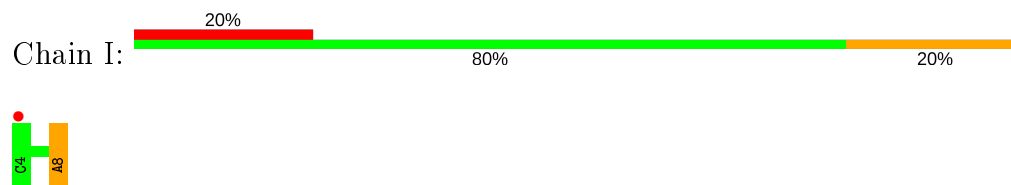
- Molecule 2: DNA (5'-D(\*GP\*CP\*AP\*TP\*CP\*GP\*TP\*GP\*AP\*AP\*TP\*CP\*GP\*AP\*GP\*G)-3')



- Molecule 3: DNA (5'-D(P\*GP\*TP\*GP\*TP\*CP\*AP\*GP\*TP\*AP\*GP\*CP\*TP\*GP\*TP\*CP\*AP\*CP\*GP\*GP\*AP\*TP\*GP\*C)-3')

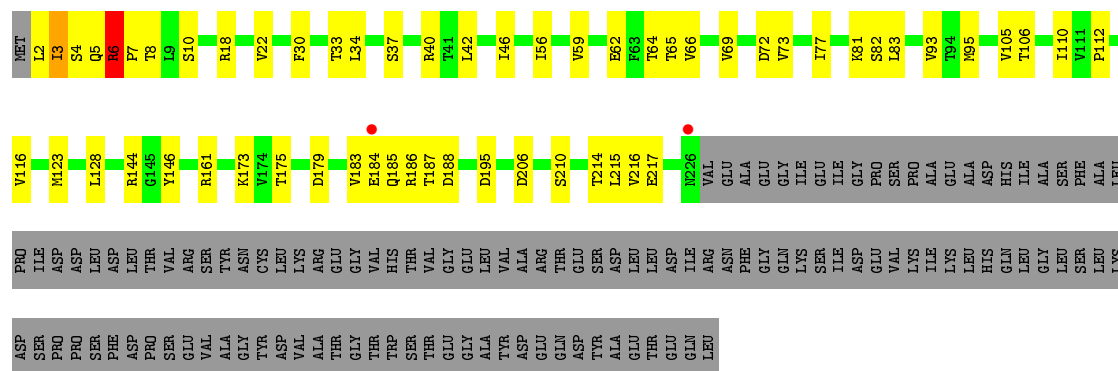


- Molecule 4: RNA (5'-R(P\*CP\*UP\*CP\*GP\*A)-3')

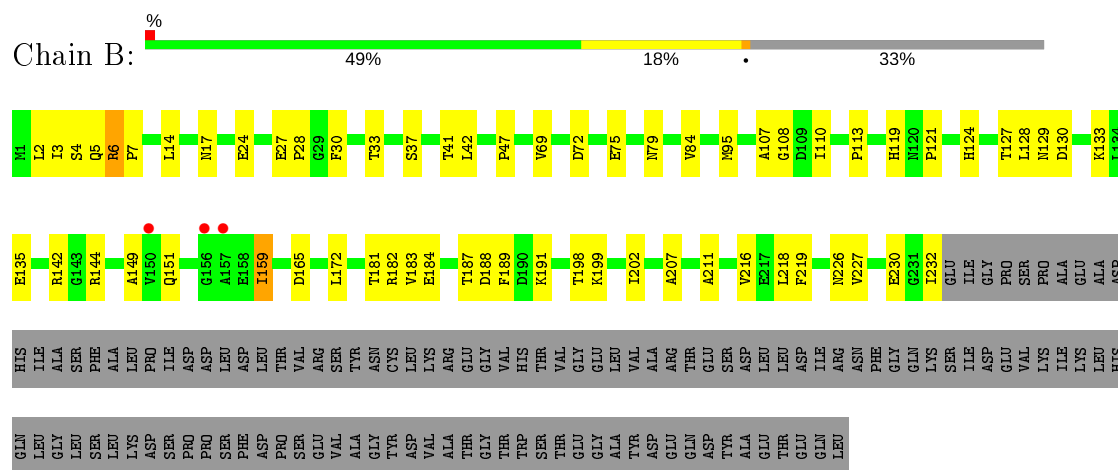


- Molecule 5: DNA-directed RNA polymerase subunit alpha

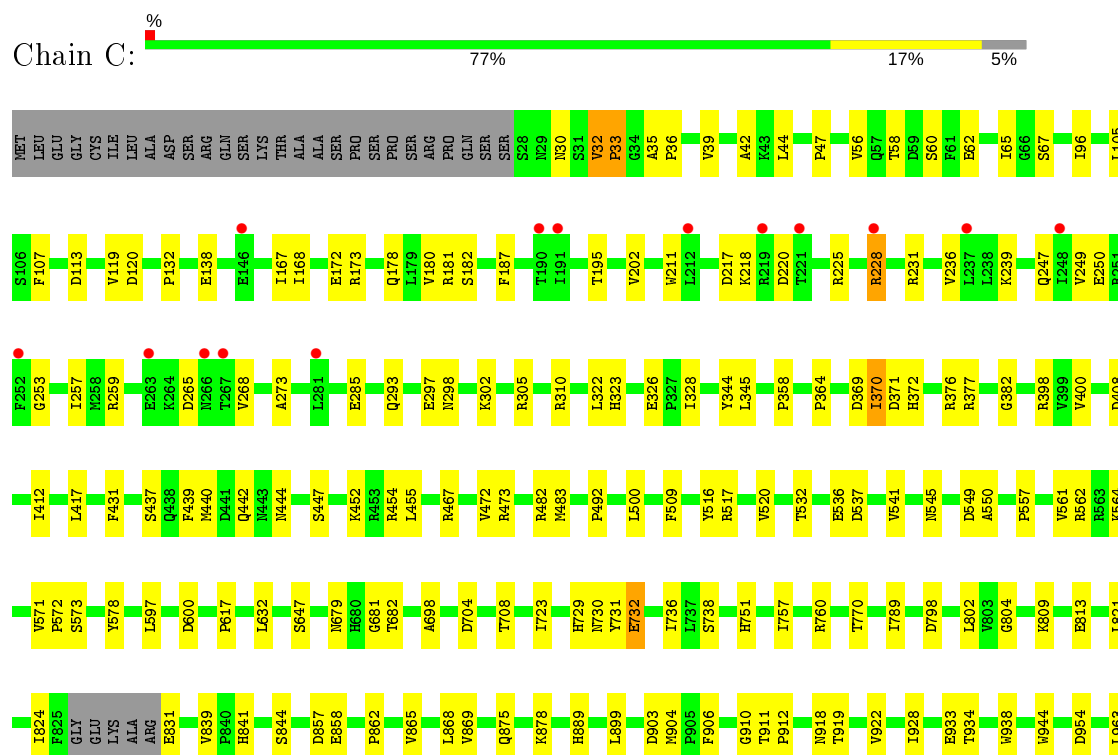


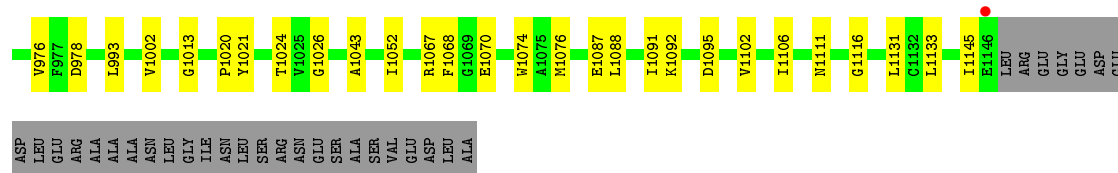


• Molecule 5: DNA-directed RNA polymerase subunit alpha

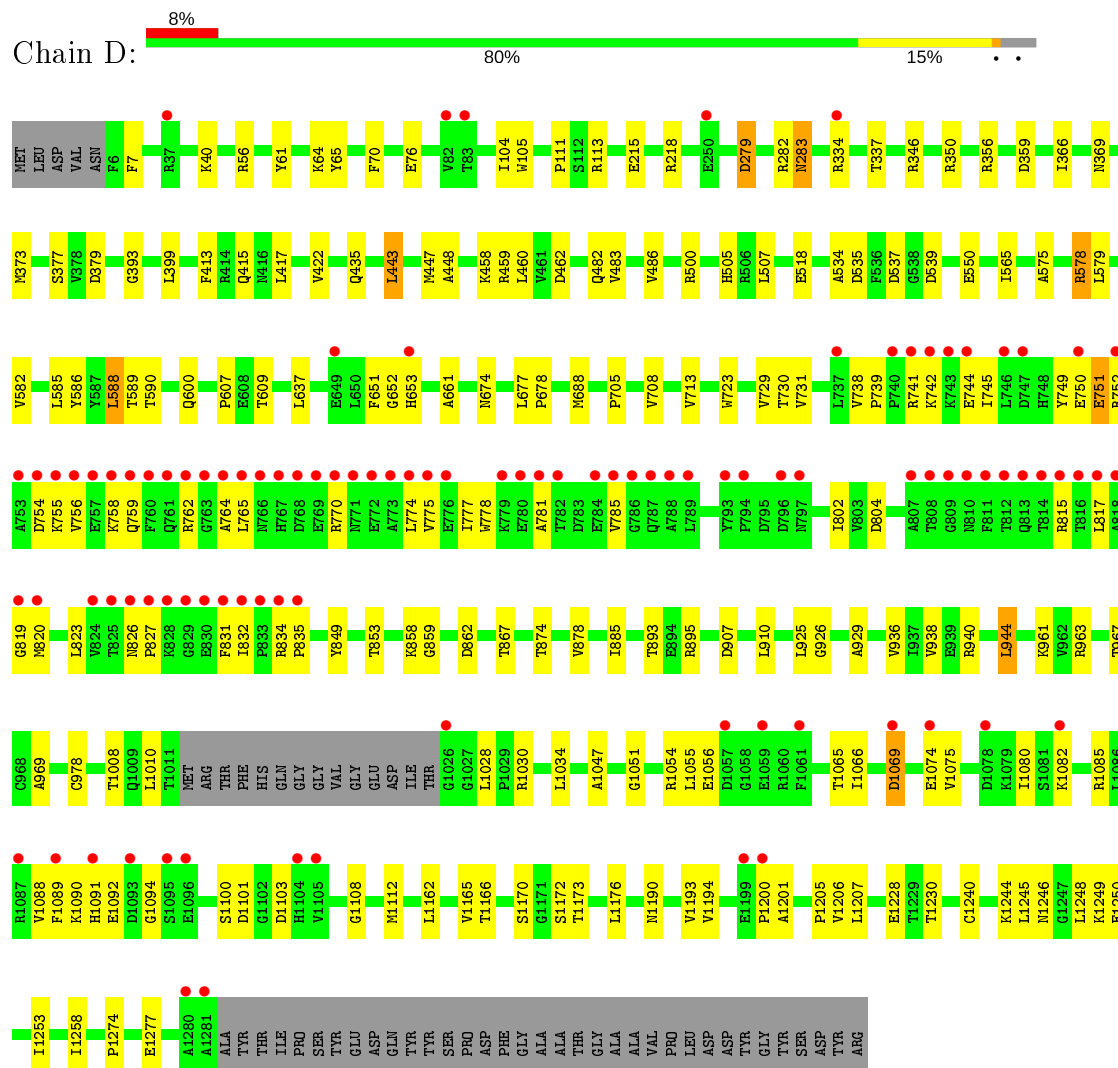


• Molecule 6: DNA-directed RNA polymerase subunit beta

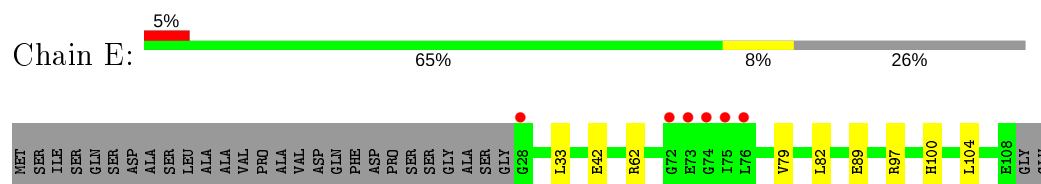




• Molecule 7: DNA-directed RNA polymerase subunit beta'



• Molecule 8: DNA-directed RNA polymerase subunit omega





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	129.53Å 158.52Å 214.66Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.02 – 3.79 48.02 – 3.79	Depositor EDS
% Data completeness (in resolution range)	96.1 (48.02-3.79) 96.1 (48.02-3.79)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.75 (at 3.77Å)	Xtriage
Refinement program	PHENIX 1.16 _3549	Depositor
R, $R_{free}$	0.235 , 0.279 0.235 , 0.279	Depositor DCC
$R_{free}$ test set	1904 reflections (4.45%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	60.2	Xtriage
Anisotropy	0.406	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 29.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.87	EDS
Total number of atoms	49194	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	91.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.06% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

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	F	0.26	0/1274	0.46	0/1731
2	G	0.59	0/392	0.90	0/604
3	H	0.59	0/532	0.96	0/820
4	I	0.29	0/116	1.00	0/178
5	A	0.32	0/1742	0.57	0/2370
5	B	0.27	0/1758	0.52	0/2397
6	C	0.26	0/8801	0.45	0/11933
7	D	0.27	0/10038	0.45	0/13568
8	E	0.25	0/643	0.41	0/877
All	All	0.29	0/25296	0.50	0/34478

There are no bond length outliers.

There are no bond angle outliers.

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	F	1257	1259	1259	25	0
2	G	349	192	192	8	0
3	H	475	259	260	16	0
4	I	105	55	55	2	0
5	A	1716	1756	1756	47	0
5	B	1732	1754	1754	48	0
6	C	8643	8576	8575	145	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	D	9872	9942	9942	151	0
8	E	630	622	622	7	0
All	All	24779	24415	24415	389	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 8.

The worst 5 of 389 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)
5:A:40:ARG:HE	5:B:33:THR:HG22	1.31	0.94
7:D:745:ILE:HG21	7:D:785:VAL:HG22	1.55	0.86
6:C:1024:THR:H	7:D:730:THR:HG21	1.41	0.84
3:H:14:DT:OP1	6:C:467:ARG:NH1	2.10	0.84
7:D:832:ILE:HG22	7:D:834:ARG:H	1.42	0.84

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	F	159/177 (90%)	134 (84%)	21 (13%)	4 (2%)	5	36
5	A	223/347 (64%)	210 (94%)	10 (4%)	3 (1%)	12	48
5	B	230/347 (66%)	214 (93%)	12 (5%)	4 (2%)	9	43
6	C	1110/1178 (94%)	1041 (94%)	61 (6%)	8 (1%)	22	60
7	D	1258/1316 (96%)	1189 (94%)	63 (5%)	6 (0%)	29	66
8	E	79/110 (72%)	73 (92%)	6 (8%)	0	100	100
All	All	3059/3475 (88%)	2861 (94%)	173 (6%)	25 (1%)	19	57

5 of 25 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
5	A	184	GLU
5	B	6	ARG
5	B	159	ILE
6	C	370	ILE
1	F	90	VAL

### 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	F	125/136 (92%)	124 (99%)	1 (1%)	81	89
5	A	194/297 (65%)	189 (97%)	5 (3%)	46	69
5	B	191/297 (64%)	191 (100%)	0	100	100
6	C	945/998 (95%)	940 (100%)	5 (0%)	88	94
7	D	1047/1095 (96%)	1035 (99%)	12 (1%)	73	85
8	E	66/90 (73%)	66 (100%)	0	100	100
All	All	2568/2913 (88%)	2545 (99%)	23 (1%)	78	88

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
6	C	954	ASP
7	D	283	ASN
7	D	1240	CYS
7	D	279	ASP
7	D	359	ASP

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

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
4	I	4/5 (80%)	1 (25%)	0

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
4	I	8	A

There are no RNA pucker outliers to report.

## 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](#)

There are no ligands in this entry.

## 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, 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	F	163/177 (92%)	0.43	7 (4%) 35 30	60, 123, 162, 182	0
2	G	17/17 (100%)	-0.20	0 100 100	56, 75, 132, 142	0
3	H	23/27 (85%)	0.64	3 (13%) 3 4	83, 133, 182, 189	0
4	I	5/5 (100%)	0.65	1 (20%) 1 1	62, 67, 74, 115	0
5	A	225/347 (64%)	-0.12	2 (0%) 84 79	44, 69, 108, 136	0
5	B	232/347 (66%)	0.13	3 (1%) 77 70	49, 85, 122, 151	0
6	C	1114/1178 (94%)	-0.05	15 (1%) 77 70	27, 61, 137, 171	0
7	D	1262/1316 (95%)	0.25	101 (8%) 12 10	26, 73, 193, 275	0
8	E	81/110 (73%)	0.04	6 (7%) 14 11	53, 80, 135, 159	0
All	All	3122/3524 (88%)	0.11	138 (4%) 34 29	26, 73, 152, 275	0

The worst 5 of 138 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
7	D	814	THR	9.1
7	D	763	GLY	8.9
7	D	761	GLN	8.6
7	D	828	LYS	8.3
7	D	776	GLU	7.8

### 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 carbohydrates in this entry.

## 6.4 Ligands

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.