



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 08:24 am BST

PDB ID : 5UHA  
Title : Crystal structure of Mycobacterium tuberculosis transcription initiation complex  
Authors : Lin, W.; Das, K.; Feng, Y.; Ebright, R.H.  
Deposited on : 2017-01-11  
Resolution : 3.91 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

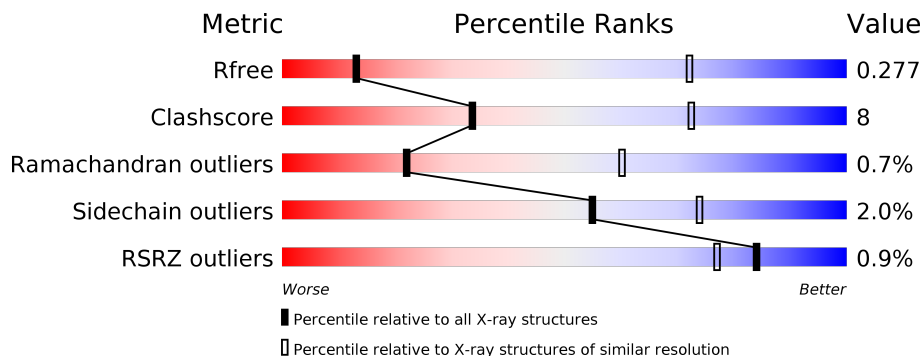
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.91 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	1002 (4.14-3.66)
Clashscore	141614	1004 (4.12-3.68)
Ramachandran outliers	138981	1021 (4.14-3.66)
Sidechain outliers	138945	1014 (4.14-3.66)
RSRZ outliers	127900	1275 (4.20-3.60)

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	A	347	
1	B	347	
2	C	1178	
3	D	1316	
4	E	110	
5	F	528	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
6	H	23	 39% 61%
7	G	16	 13% 50% 44% 6%

## 2 Entry composition [i](#)

There are 9 unique types of molecules in this entry. The entry contains 25997 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 DNA-directed RNA polymerase subunit alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	224	Total	C	N	O	S	0	0	0
			1704	1072	295	335	2			
1	B	227	Total	C	N	O	S	0	0	0
			1715	1080	291	342	2			

- Molecule 2 is a protein called DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	1126	Total	C	N	O	S	0	0	0
			8714	5454	1528	1693	39			

- Molecule 3 is a protein called DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	D	1265	Total	C	N	O	S	0	0	0
			9887	6188	1793	1866	40			

- Molecule 4 is a protein called DNA-directed RNA polymerase subunit omega.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
4	E	81	Total	C	N	O	0	0	0
			637	408	106	123			

- Molecule 5 is a protein called RNA polymerase sigma factor SigA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	F	322	Total	C	N	O	S	0	0	0
			2555	1589	461	496	9			

- Molecule 6 is a DNA chain called DNA (5'-D(\*TP\*AP\*TP\*AP\*AP\*TP\*GP\*GP\*GP\*AP\*GP\*CP\*TP\*GP\*TP\*CP\*AP\*CP\*GP\*GP\*AP\*TP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
6	H	23	476	227	91	136	22	0	0	0

- Molecule 7 is a DNA chain called DNA (5'-D(\*CP\*AP\*TP\*CP\*CP\*GP\*TP\*GP\*AP\*GP\*TP\*CP\*GP\*AP\*GP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
7	G	15	306	146	58	88	14	0	0	0

- Molecule 8 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	D	2	Total	Zn	0	0
			2	2		

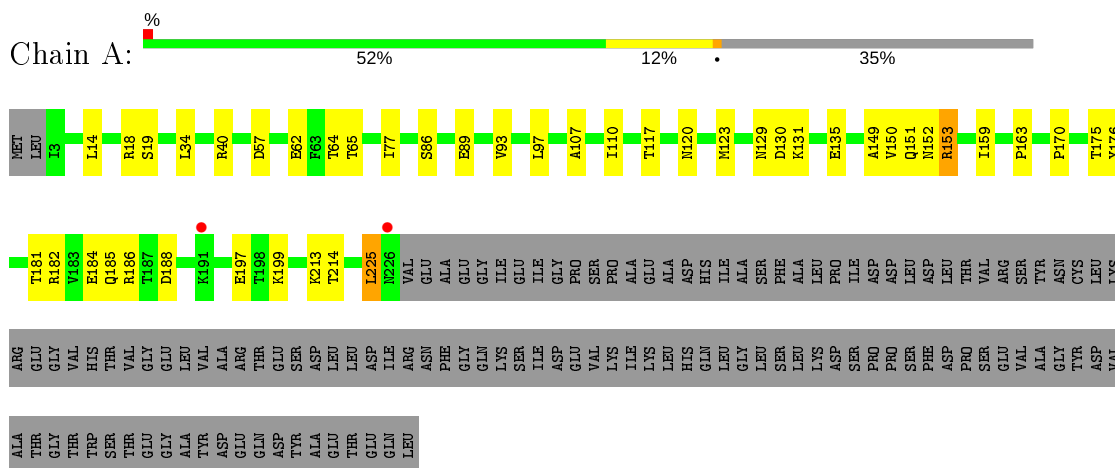
- Molecule 9 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	D	1	Total	Mg	0	0
			1	1		

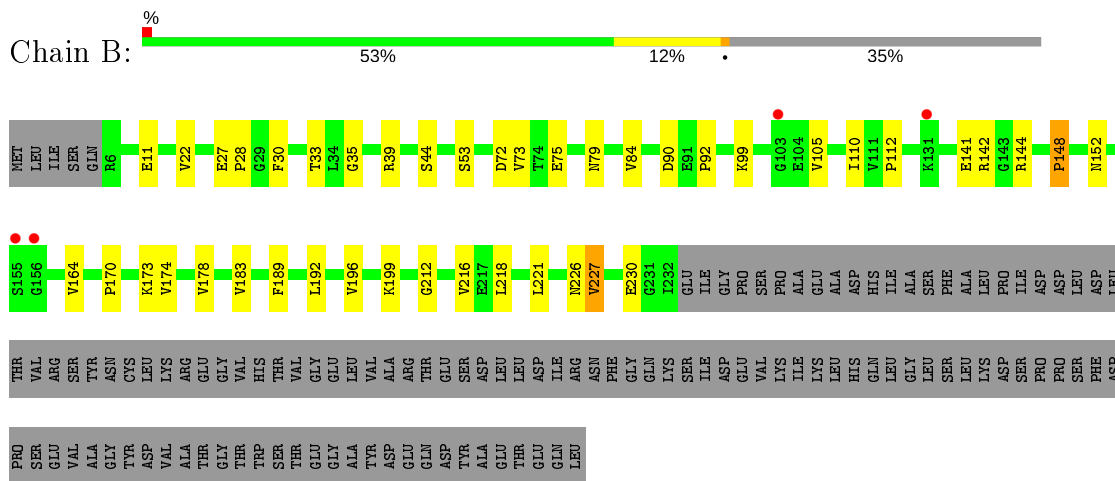
### 3 Residue-property plots [i](#)

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.

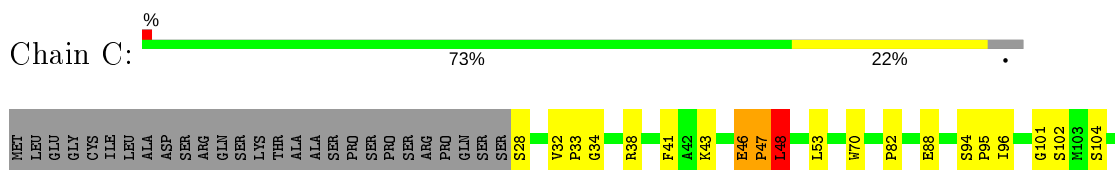
#### ● Molecule 1: DNA-directed RNA polymerase subunit alpha

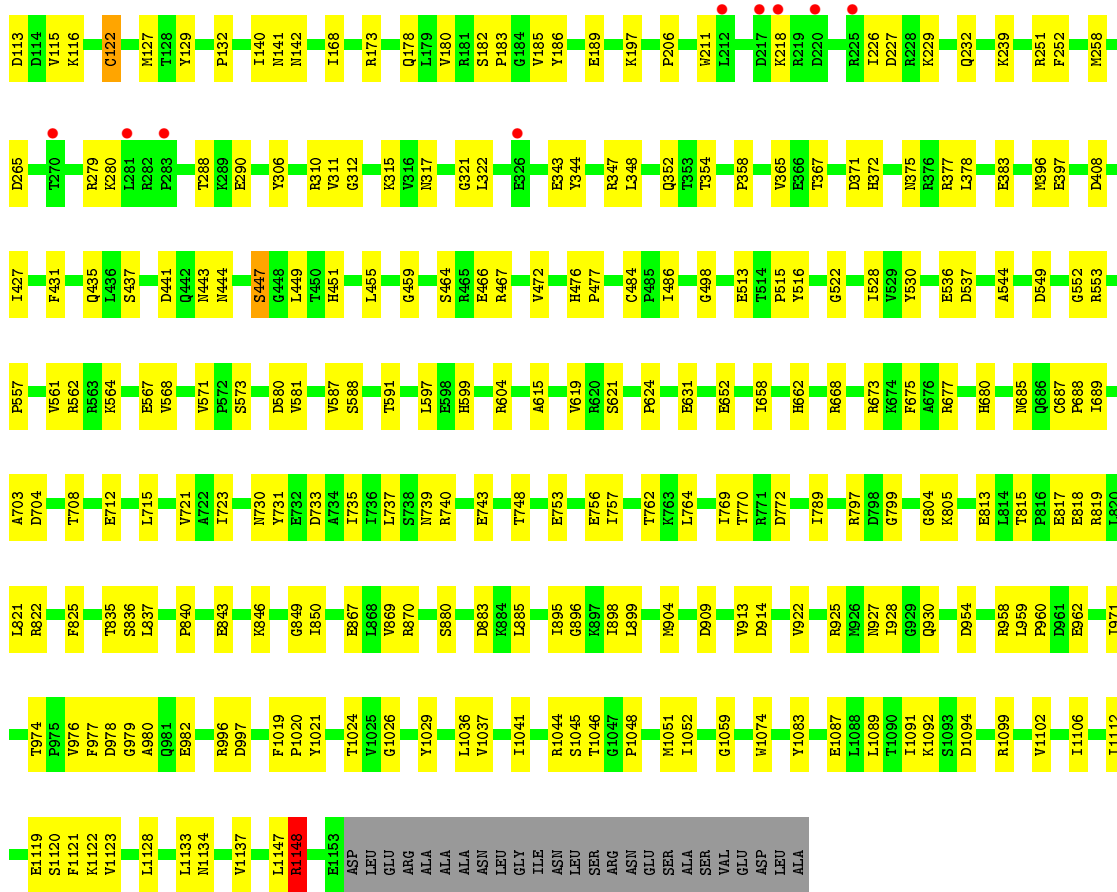


#### ● Molecule 1: DNA-directed RNA polymerase subunit alpha

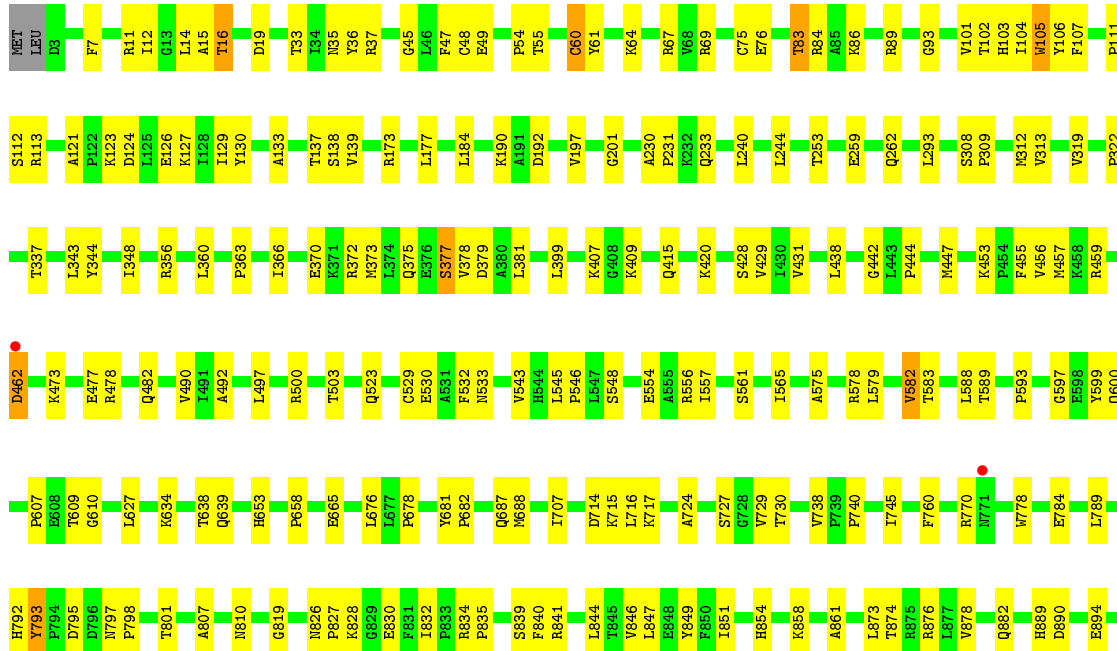


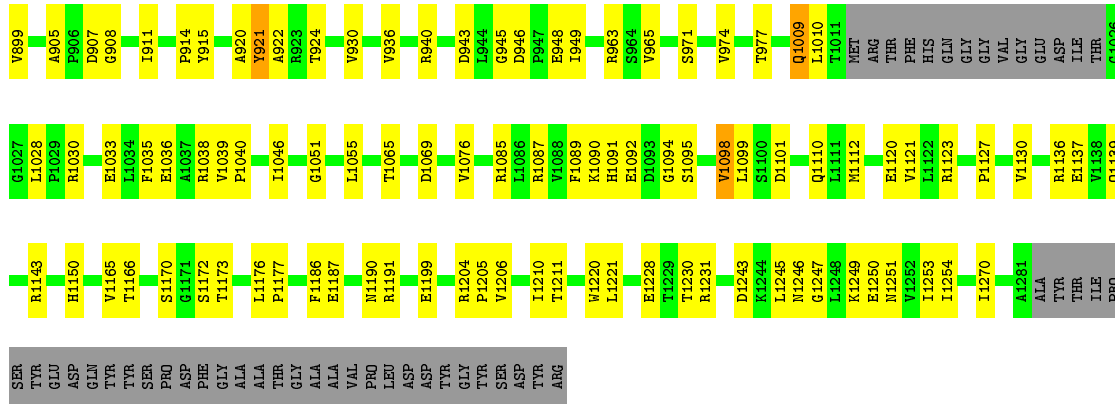
#### ● Molecule 2: DNA-directed RNA polymerase subunit beta



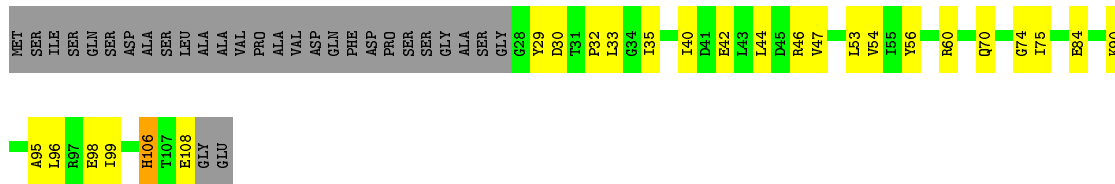


• Molecule 3: DNA-directed RNA polymerase subunit beta'

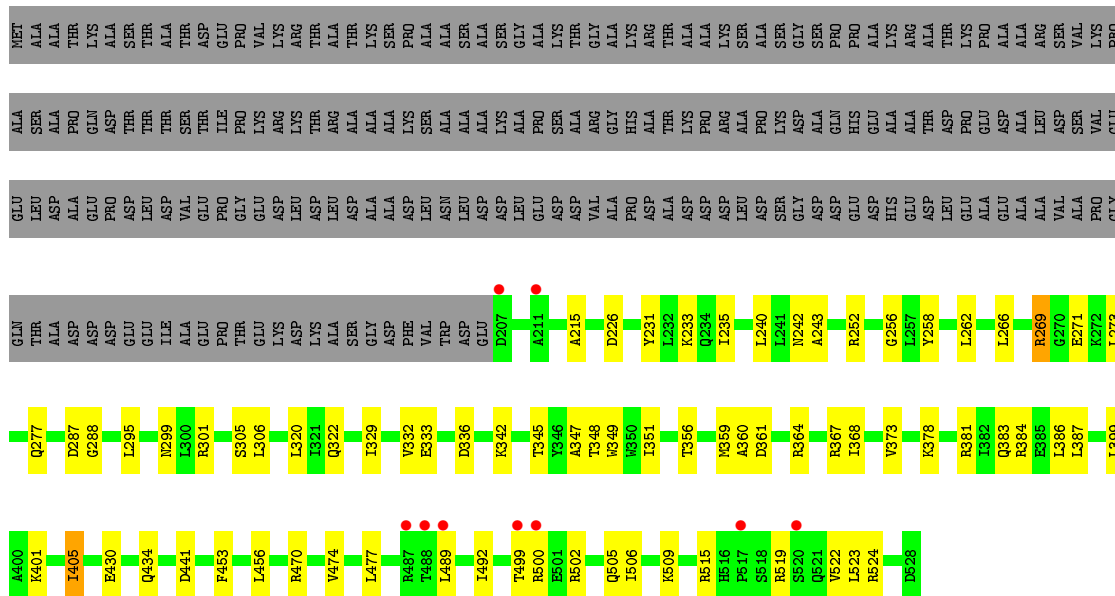




• Molecule 4: DNA-directed RNA polymerase subunit omega



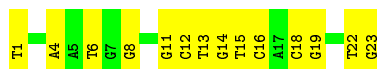
• Molecule 5: RNA polymerase sigma factor SigA



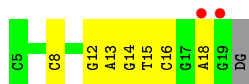
• Molecule 6: DNA (5'-D(\*TP\*AP\*TP\*AP\*AP\*TP\*GP\*GP\*GP\*AP\*GP\*CP\*TP\*GP\*TP\*CP\*AP\*CP\*GP\*GP\*AP\*TP\*G)-3')







- Molecule 7: DNA (5'-D(\*CP\*AP\*TP\*CP\*CP\*GP\*TP\*GP\*AP\*GP\*TP\*CP\*GP\*AP\*GP\*G)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	150.20Å 163.48Å 195.65Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.91 – 3.91 48.91 – 3.91	Depositor EDS
% Data completeness (in resolution range)	77.9 (48.91-3.91) 61.2 (48.91-3.91)	Depositor EDS
$R_{merge}$	0.22	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.60 (at 3.88Å)	Xtrriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.211 , 0.279 0.215 , 0.277	Depositor DCC
$R_{free}$ test set	1988 reflections (5.75%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	26.4	Xtrriage
Anisotropy	0.591	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , -28.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.37$ , $\langle L^2 \rangle = 0.20$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.83	EDS
Total number of atoms	25997	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.97% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, MG

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	0.26	0/1730	0.47	0/2354
1	B	0.25	0/1741	0.45	0/2371
2	C	0.28	0/8873	0.46	1/12031 (0.0%)
3	D	0.29	0/10052	0.45	0/13591
4	E	0.27	0/650	0.44	0/886
5	F	0.27	0/2585	0.43	0/3485
6	H	0.57	0/535	0.89	0/826
7	G	0.63	0/343	0.88	0/528
All	All	0.30	0/26509	0.48	1/36072 (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	C	48	LEU	CA-CB-CG	5.39	127.70	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	1704	0	1741	26	0
1	B	1715	0	1739	25	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	8714	0	8636	171	0
3	D	9887	0	9943	185	0
4	E	637	0	635	19	0
5	F	2555	0	2579	48	0
6	H	476	0	261	13	0
7	G	306	0	170	7	0
8	D	2	0	0	0	0
9	D	1	0	0	0	0
All	All	25997	0	25704	439	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 439 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)
2:C:279:ARG:HD3	5:F:215:ALA:HB1	1.68	0.76
3:D:1165:VAL:HG12	3:D:1205:PRO:HA	1.67	0.75
6:H:16:DC:O2	7:G:12:DG:N2	2.15	0.75
3:D:137:THR:OG1	3:D:253:THR:O	2.01	0.74
3:D:1090:LYS:HB3	3:D:1092:GLU:HG2	1.69	0.74

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	222/347 (64%)	208 (94%)	12 (5%)	2 (1%)	17 54
1	B	225/347 (65%)	206 (92%)	15 (7%)	4 (2%)	8 42
2	C	1124/1178 (95%)	1039 (92%)	74 (7%)	11 (1%)	15 52

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	D	1261/1316 (96%)	1184 (94%)	72 (6%)	5 (0%)	34	71
4	E	79/110 (72%)	74 (94%)	5 (6%)	0	100	100
5	F	320/528 (61%)	303 (95%)	16 (5%)	1 (0%)	41	75
All	All	3231/3826 (84%)	3014 (93%)	194 (6%)	23 (1%)	22	60

5 of 23 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	678	PRO
2	C	1148	ARG
5	F	405	ILE
1	A	184	GLU
2	C	33	PRO

### 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	192/297 (65%)	188 (98%)	4 (2%)	53	73
1	B	192/297 (65%)	192 (100%)	0	100	100
2	C	948/998 (95%)	933 (98%)	15 (2%)	62	79
3	D	1048/1095 (96%)	1023 (98%)	25 (2%)	49	69
4	E	68/90 (76%)	66 (97%)	2 (3%)	42	65
5	F	271/427 (64%)	262 (97%)	9 (3%)	38	63
All	All	2719/3204 (85%)	2664 (98%)	55 (2%)	55	74

5 of 55 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	83	THR
3	D	503	THR
5	F	336	ASP
3	D	105	TRP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	D	337	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
2	C	662	HIS
4	E	69	ASN

### 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](#)

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

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

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	A	224/347 (64%)	-0.27	2 (0%) 84 77	3, 28, 99, 153	0
1	B	227/347 (65%)	0.13	4 (1%) 68 59	22, 70, 133, 162	0
2	C	1126/1178 (95%)	-0.33	9 (0%) 86 79	1, 17, 107, 132	0
3	D	1265/1316 (96%)	-0.44	2 (0%) 95 93	1, 16, 88, 125	0
4	E	81/110 (73%)	-0.26	0 100 100	6, 16, 53, 91	0
5	F	322/528 (60%)	-0.16	9 (2%) 53 41	2, 39, 140, 174	0
6	H	23/23 (100%)	-0.03	0 100 100	7, 86, 132, 162	0
7	G	15/16 (93%)	0.81	2 (13%) 3 3	80, 103, 134, 134	0
All	All	3283/3865 (84%)	-0.31	28 (0%) 84 77	1, 22, 110, 174	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	155	SER	4.4
5	F	500	ARG	3.9
5	F	487	ARG	3.7
5	F	488	THR	3.7
7	G	19	DG	3.5

### 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 [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<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	B-factors( $\text{\AA}^2$ )	Q<0.9
9	MG	D	1403	1/1	0.98	0.12	2,2,2,2	0
8	ZN	D	1401	1/1	0.98	0.09	55,55,55,55	0
8	ZN	D	1402	1/1	1.00	0.07	25,25,25,25	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.