



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 9, 2024 – 09:22 AM EST

PDB ID : 5UH9
Title : Crystal structure of Mycobacterium tuberculosis transcription initiation complex containing 2nt RNA
Authors : Lin, W.; Das, K.; Feng, Y.; Ebright, R.H.
Deposited on : 2017-01-11
Resolution : 4.40 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
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

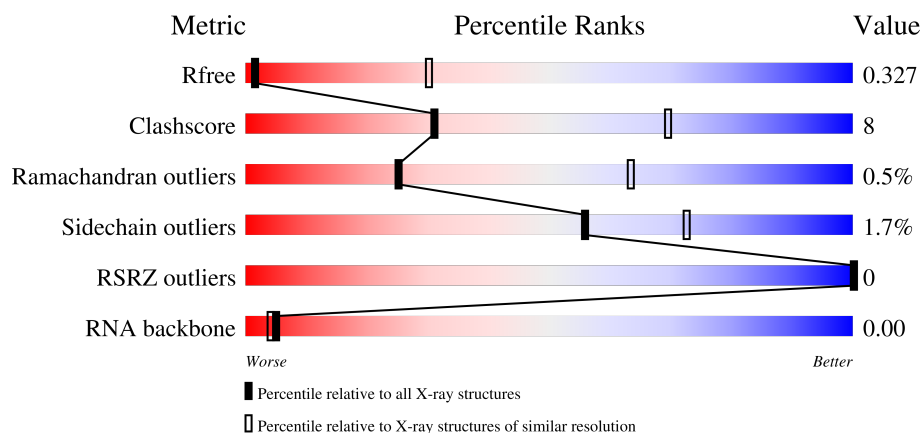
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 4.40 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}	164625	1013 (4.90-3.90)
Clashscore	180529	1066 (4.90-3.90)
Ramachandran outliers	177936	1019 (4.92-3.86)
Sidechain outliers	177891	1003 (4.92-3.86)
RSRZ outliers	164620	1010 (4.90-3.90)
RNA backbone	3690	1156 (5.60-3.00)

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 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	347	
1	B	347	
2	C	1178	
3	D	1316	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	110	<div><div></div><div>65%</div><div>8%</div><div>26%</div></div>
5	F	528	<div><div></div><div>49%</div><div>11%</div><div>39%</div></div>
6	H	23	<div><div></div><div>48%</div><div>43%</div><div>9%</div></div>
7	G	16	<div><div></div><div>62%</div><div>25%</div><div>6%</div><div>6%</div></div>
8	I	2	<div><div></div><div>100%</div></div>

2 Entry composition [i](#)

There are 10 unique types of molecules in this entry. The entry contains 26030 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
1	A	224	Total	C	N	O	S	0	0	0
			1704	1072	295	335	2			
1	B	226	Total	C	N	O	S	0	0	0
			1709	1077	290	340	2			

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

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

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	G	15	Total	C	N	O	P	0	0	0
			303	145	56	88	14			

- Molecule 8 is a RNA chain called RNA (5'-R(*GP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	I	2	Total	C	N	O	P	0	0	0
			42	20	10	11	1			

- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

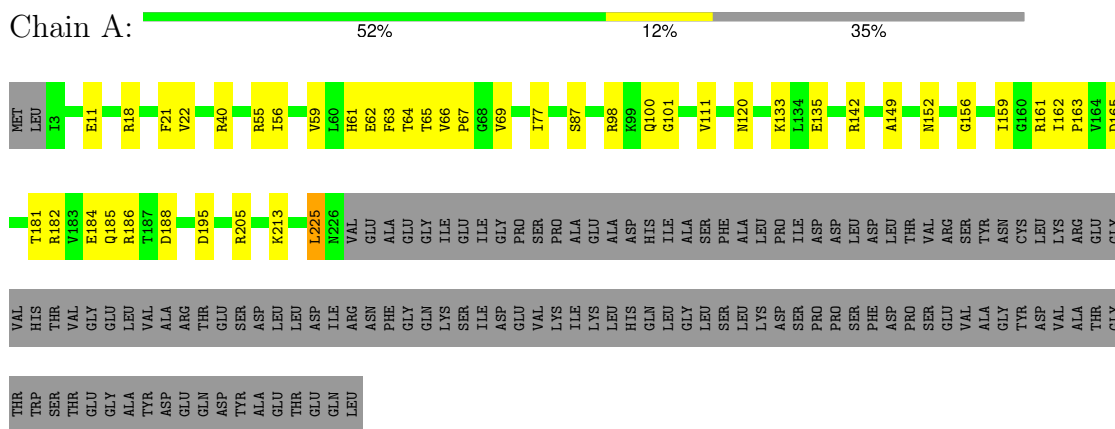
- Molecule 10 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

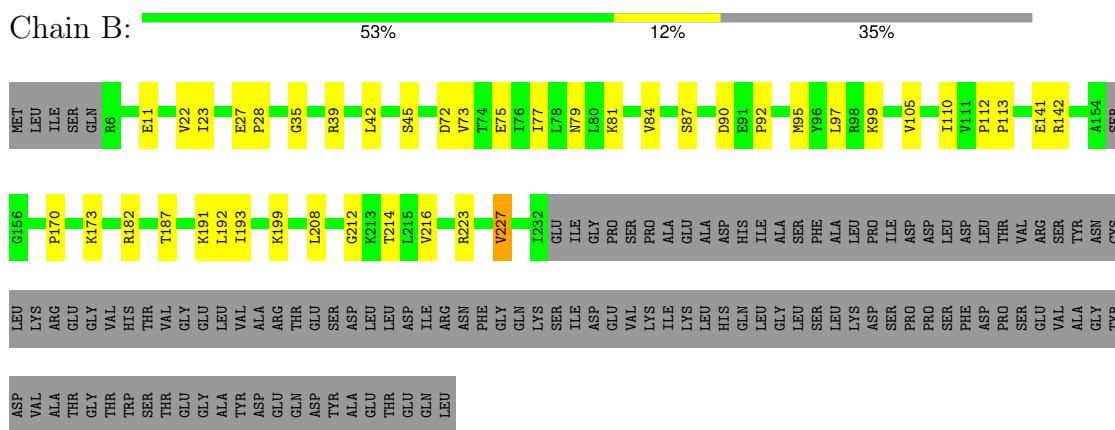
3 Residue-property plots

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.

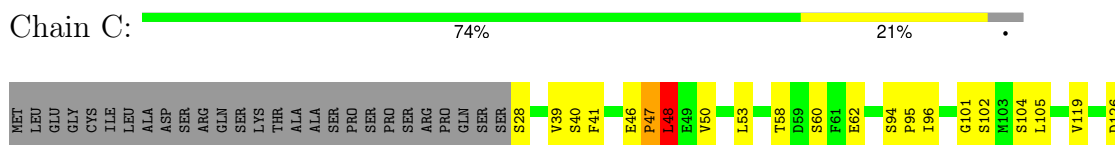
• Molecule 1: DNA-directed RNA polymerase subunit alpha

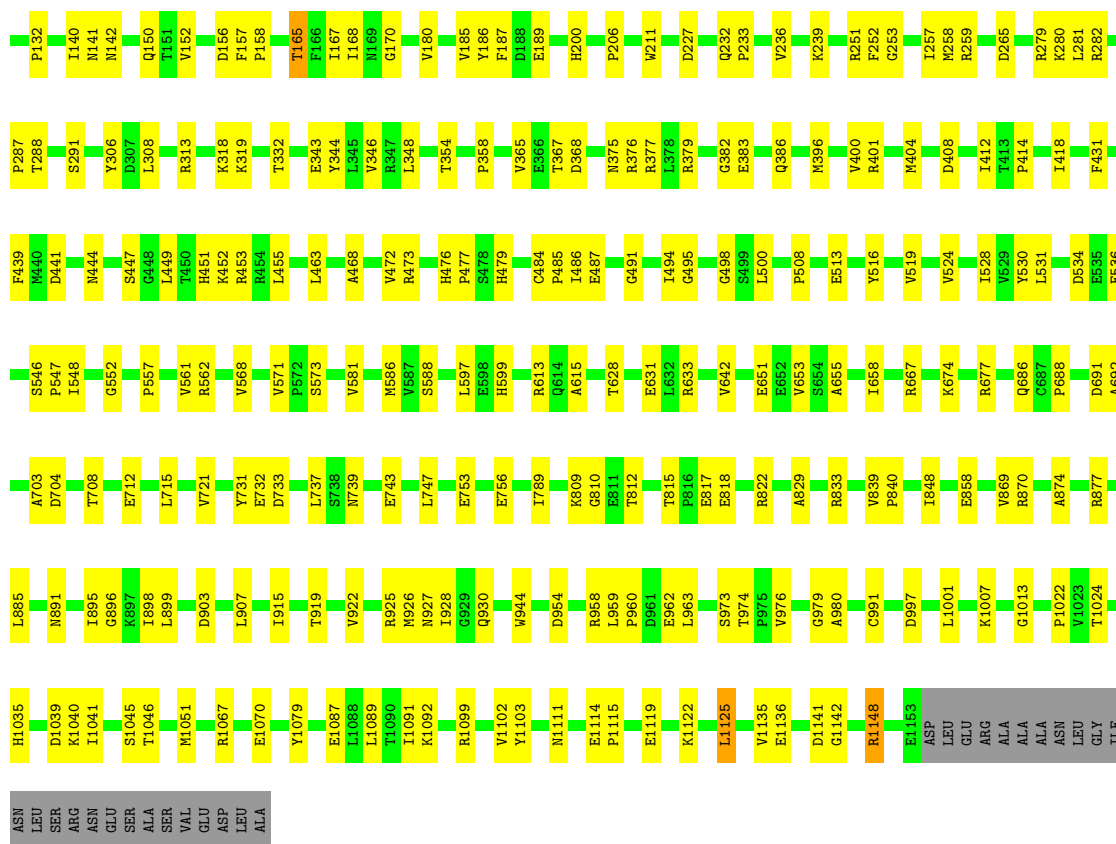


• Molecule 1: DNA-directed RNA polymerase subunit alpha



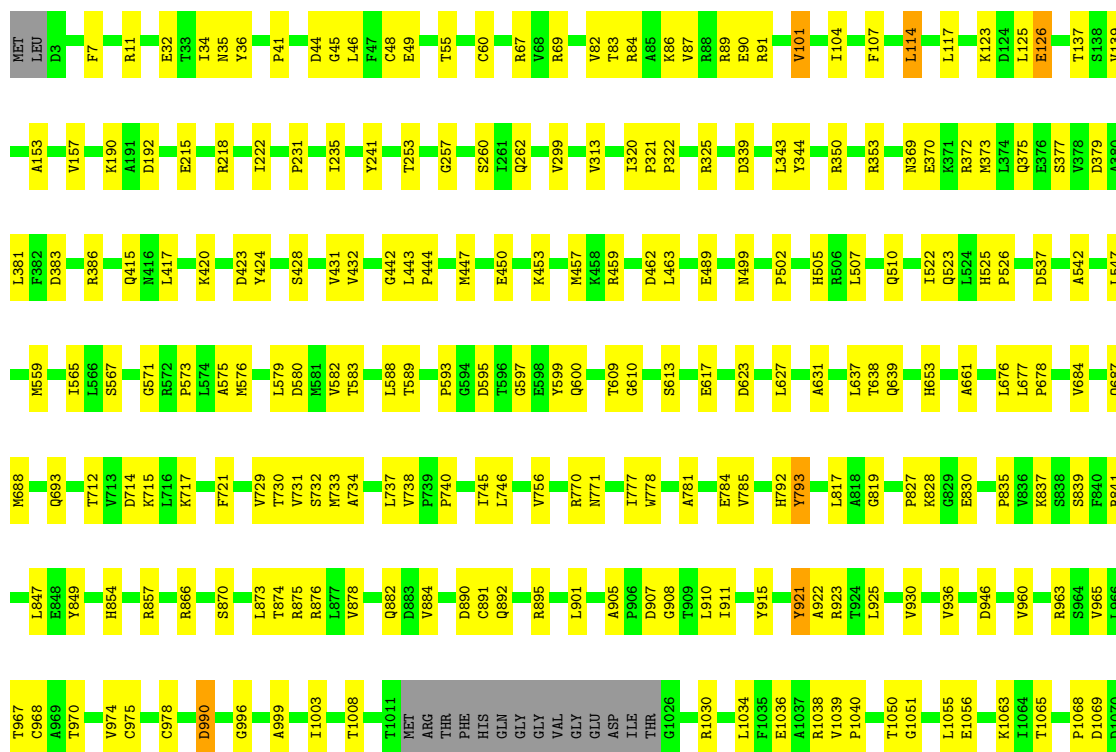
• Molecule 2: DNA-directed RNA polymerase subunit beta





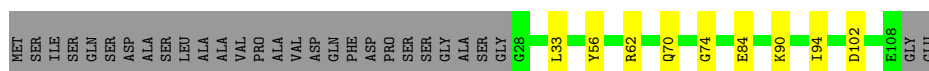
• Molecule 3: DNA-directed RNA polymerase subunit beta'

Chain D: 74% 21%



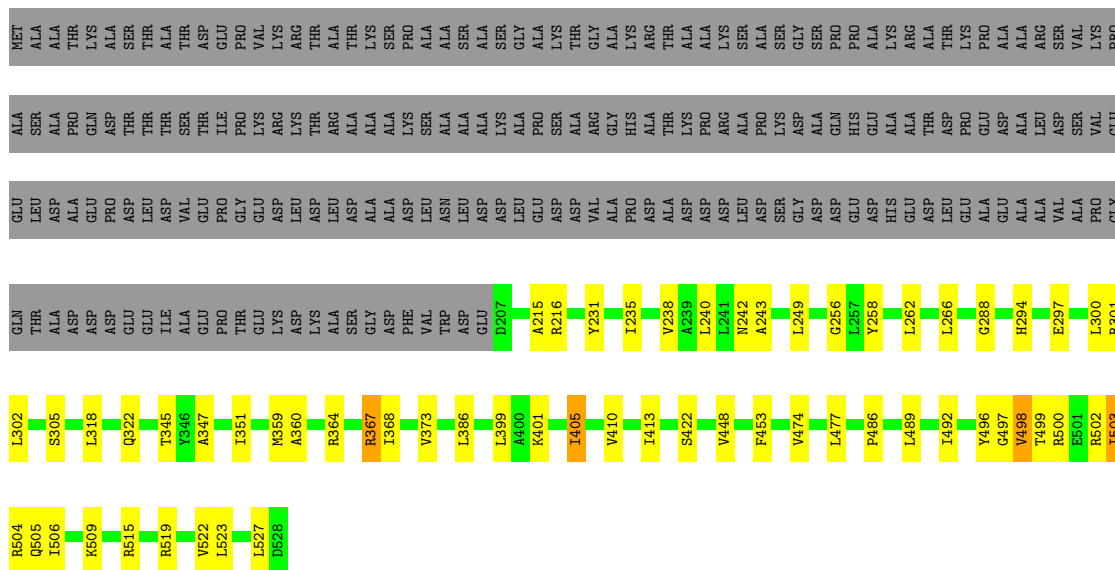
- Molecule 4: DNA-directed RNA polymerase subunit omega

Chain E:  65% 8% 26%



- Molecule 5: RNA polymerase sigma factor SigA

Chain F: 49% 11% 39%



- 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')

Chain H: 48% 43% 9%



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

Chain G: 62% 25% 6% 6%



- Molecule 8: RNA (5'-R(*GP*A)-3')

Chain I:  100%

There are no outlier residues recorded for this chain.

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	154.46Å 164.63Å 201.85Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.89 – 4.40 49.89 – 4.40	Depositor EDS
% Data completeness (in resolution range)	85.8 (49.89-4.40) 71.6 (49.89-4.40)	Depositor EDS
R_{merge}	0.19	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.13 (at 4.45Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.281 , 0.332 0.286 , 0.327	Depositor DCC
R_{free} test set	26563 reflections (6.98%)	wwPDB-VP
Wilson B-factor (Å ²)	49.6	Xtriage
Anisotropy	0.691	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 522.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.35$, $\langle L^2 \rangle = 0.18$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.80	EDS
Total number of atoms	26030	wwPDB-VP
Average B, all atoms (Å ²)	88.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.33% of the height of the origin peak. No significant pseudotranslation is detected.*

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: 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	0.23	0/1730	0.45	0/2354
1	B	0.24	0/1734	0.44	0/2360
2	C	0.24	0/8873	0.42	1/12031 (0.0%)
3	D	0.24	0/10052	0.41	0/13591
4	E	0.23	0/650	0.42	0/886
5	F	0.24	0/2585	0.41	0/3485
6	H	0.68	2/535 (0.4%)	1.07	4/826 (0.5%)
7	G	0.82	1/339 (0.3%)	0.95	2/521 (0.4%)
8	I	0.30	0/47	0.60	0/72
All	All	0.27	3/26545 (0.0%)	0.46	7/36126 (0.0%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	G	12	DG	O3'-P	10.61	1.73	1.61
6	H	13	DT	O3'-P	6.39	1.68	1.61
6	H	16	DC	O3'-P	5.84	1.68	1.61

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	H	16	DC	P-O3'-C3'	8.31	129.67	119.70
6	H	19	DG	C5'-C4'-O4'	7.34	123.24	109.30
6	H	19	DG	O4'-C4'-C3'	-6.85	101.76	104.50
7	G	12	DG	P-O3'-C3'	-6.21	112.25	119.70
6	H	16	DC	OP1-P-O3'	5.77	117.89	105.20

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	25	0
1	B	1709	0	1733	25	0
2	C	8714	0	8636	164	0
3	D	9887	0	9943	186	0
4	E	637	0	635	6	0
5	F	2555	0	2579	42	0
6	H	476	0	261	15	0
7	G	303	0	170	4	0
8	I	42	0	22	0	0
9	D	2	0	0	0	0
10	D	1	0	0	0	0
All	All	26030	0	25720	414	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 414 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)
6:H:16:DC:O2	7:G:12:DG:N2	1.92	1.03
3:D:891:CYS:SG	3:D:970:THR:OG1	2.23	0.96
2:C:211:TRP:NE1	6:H:13:DT:O2	2.00	0.94
6:H:16:DC:N3	7:G:12:DG:N1	2.23	0.83
6:H:15:DT:H2''	6:H:16:DC:H5'	1.68	0.73

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%)	209 (94%)	11 (5%)	2 (1%)	14	50
1	B	222/347 (64%)	202 (91%)	18 (8%)	2 (1%)	14	50
2	C	1124/1178 (95%)	1055 (94%)	62 (6%)	7 (1%)	22	59
3	D	1261/1316 (96%)	1192 (94%)	67 (5%)	2 (0%)	44	78
4	E	79/110 (72%)	77 (98%)	2 (2%)	0	100	100
5	F	320/528 (61%)	304 (95%)	13 (4%)	3 (1%)	14	50
All	All	3228/3826 (84%)	3039 (94%)	173 (5%)	16 (0%)	25	63

5 of 16 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	678	PRO
5	F	497	GLY
1	A	184	GLU
2	C	1148	ARG
3	D	593	PRO

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	192/297 (65%)	188 (98%)	4 (2%)	48	67
1	B	191/297 (64%)	190 (100%)	1 (0%)	86	90
2	C	948/998 (95%)	935 (99%)	13 (1%)	62	76
3	D	1048/1095 (96%)	1026 (98%)	22 (2%)	48	67
4	E	68/90 (76%)	67 (98%)	1 (2%)	60	75
5	F	271/427 (64%)	265 (98%)	6 (2%)	47	66
All	All	2718/3204 (85%)	2671 (98%)	47 (2%)	56	72

5 of 47 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	684	VAL
3	D	968	CYS
3	D	714	ASP
3	D	901	LEU
3	D	978	CYS

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

Mol	Chain	Res	Type
2	C	142	ASN
2	C	930	GLN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	I	1/2 (50%)	0	0

There are no RNA backbone outliers to report.

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 oligosaccharides 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 [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	224/347 (64%)	-0.97	0 100 100	61, 73, 133, 185	0
1	B	226/347 (65%)	-0.92	0 100 100	66, 116, 184, 220	0
2	C	1126/1178 (95%)	-0.98	0 100 100	59, 72, 145, 193	0
3	D	1265/1316 (96%)	-0.98	0 100 100	59, 73, 131, 183	0
4	E	81/110 (73%)	-1.02	0 100 100	63, 79, 125, 199	0
5	F	322/528 (60%)	-0.90	0 100 100	62, 84, 204, 218	0
6	H	23/23 (100%)	-0.62	0 100 100	69, 90, 119, 133	0
7	G	15/16 (93%)	-0.53	0 100 100	65, 91, 114, 121	0
8	I	2/2 (100%)	-1.24	0 100 100	89, 89, 89, 91	0
All	All	3284/3867 (84%)	-0.96	0 100 100	59, 75, 157, 220	0

There are no RSRZ outliers to report.

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
9	ZN	D	1401	1/1	1.00	0.01	116,116,116,116	0
9	ZN	D	1402	1/1	1.00	0.01	116,116,116,116	0
10	MG	D	1403	1/1	1.00	0.03	70,70,70,70	0

6.5 Other polymers [i](#)

There are no such residues in this entry.