



wwPDB X-ray Structure Validation Summary Report ⓘ

Apr 25, 2026 – 02:10 PM EDT

PDB ID : 6KOO / pdb_00006koo
Title : Mycobacterium tuberculosis initial transcription complex comprising sigma H
and 5'-OH RNA of 7 nt
Authors : Li, L.; Zhang, Y.
Deposited on : 2019-08-12
Resolution : 2.80 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	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

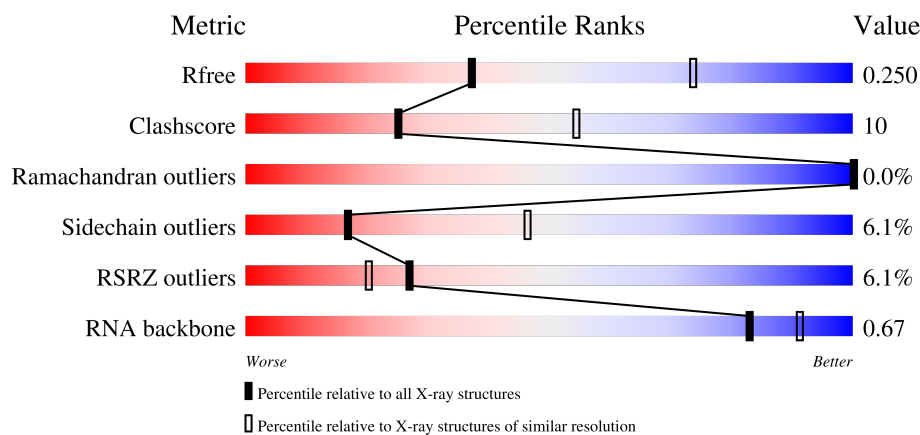
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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.



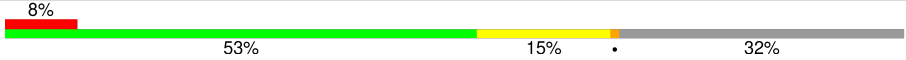
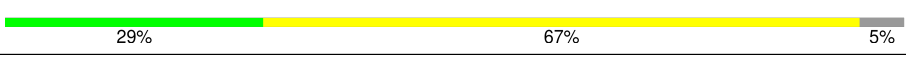
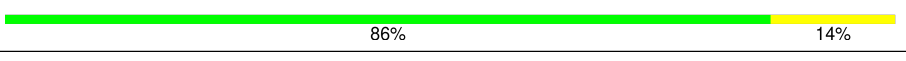
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	180053	3866 (2.80-2.80)
Clashscore	190562	4276 (2.80-2.80)
Ramachandran outliers	187476	4196 (2.80-2.80)
Sidechain outliers	187428	4198 (2.80-2.80)
RSRZ outliers	180081	3869 (2.80-2.80)
RNA backbone	3983	1114 (3.00-2.60)

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 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	368	48% 10% 41%
1	B	368	7% 43% 17% 38%
2	C	1174	6% 71% 24%
3	D	1317	4% 73% 21%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	110	
5	F	218	
6	G	23	
7	H	21	
8	I	7	

2 Entry composition

There are 11 unique types of molecules in this entry. The entry contains 24837 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	217	Total	C	N	O	S	0	0	0
			1646	1035	285	324	2			
1	B	230	Total	C	N	O	S	0	0	0
			1670	1057	283	328	2			

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-20	MET	-	initiating methionine	UNP P9WGZ1
A	-19	GLY	-	expression tag	UNP P9WGZ1
A	-18	HIS	-	expression tag	UNP P9WGZ1
A	-17	HIS	-	expression tag	UNP P9WGZ1
A	-16	HIS	-	expression tag	UNP P9WGZ1
A	-15	HIS	-	expression tag	UNP P9WGZ1
A	-14	HIS	-	expression tag	UNP P9WGZ1
A	-13	HIS	-	expression tag	UNP P9WGZ1
A	-12	HIS	-	expression tag	UNP P9WGZ1
A	-11	HIS	-	expression tag	UNP P9WGZ1
A	-10	HIS	-	expression tag	UNP P9WGZ1
A	-9	HIS	-	expression tag	UNP P9WGZ1
A	-8	SER	-	expression tag	UNP P9WGZ1
A	-7	SER	-	expression tag	UNP P9WGZ1
A	-6	GLY	-	expression tag	UNP P9WGZ1
A	-5	HIS	-	expression tag	UNP P9WGZ1
A	-4	ILE	-	expression tag	UNP P9WGZ1
A	-3	GLU	-	expression tag	UNP P9WGZ1
A	-2	GLY	-	expression tag	UNP P9WGZ1
A	-1	ARG	-	expression tag	UNP P9WGZ1
A	0	HIS	-	expression tag	UNP P9WGZ1
B	-20	MET	-	initiating methionine	UNP P9WGZ1
B	-19	GLY	-	expression tag	UNP P9WGZ1
B	-18	HIS	-	expression tag	UNP P9WGZ1
B	-17	HIS	-	expression tag	UNP P9WGZ1

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-16	HIS	-	expression tag	UNP P9WGZ1
B	-15	HIS	-	expression tag	UNP P9WGZ1
B	-14	HIS	-	expression tag	UNP P9WGZ1
B	-13	HIS	-	expression tag	UNP P9WGZ1
B	-12	HIS	-	expression tag	UNP P9WGZ1
B	-11	HIS	-	expression tag	UNP P9WGZ1
B	-10	HIS	-	expression tag	UNP P9WGZ1
B	-9	HIS	-	expression tag	UNP P9WGZ1
B	-8	SER	-	expression tag	UNP P9WGZ1
B	-7	SER	-	expression tag	UNP P9WGZ1
B	-6	GLY	-	expression tag	UNP P9WGZ1
B	-5	HIS	-	expression tag	UNP P9WGZ1
B	-4	ILE	-	expression tag	UNP P9WGZ1
B	-3	GLU	-	expression tag	UNP P9WGZ1
B	-2	GLY	-	expression tag	UNP P9WGZ1
B	-1	ARG	-	expression tag	UNP P9WGZ1
B	0	HIS	-	expression tag	UNP P9WGZ1

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1135	Total	C	N	O	S	0	0	0
			8673	5428	1519	1687	39			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-1	MET	-	initiating methionine	UNP P9WGY9
C	0	VAL	-	expression tag	UNP P9WGY9

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1258	Total	C	N	O	S	0	0	0
			9763	6118	1766	1838	41			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	0	MET	-	initiating methionine	UNP P9WGY7
D	1	VAL	-	expression tag	UNP P9WGY7

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	E	75	Total	C	N	O	0	0	0
			592	378	99	115			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	182	Total	C	N	O	S	0	0	0
			1432	896	251	280	5			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-1	GLY	-	expression tag	UNP P9WGH9
F	0	ALA	-	expression tag	UNP P9WGH9

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	23	Total	C	N	O	P	0	0	0
			475	226	89	138	22			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	H	20	Total	C	N	O	P	0	0	0
			412	196	77	120	19			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	I	7	Total	C	N	O	P	0	0	0
			142	65	24	47	6			

- Molecule 9 is ZINC ION (CCD ID: ZN) (formula: Zn).

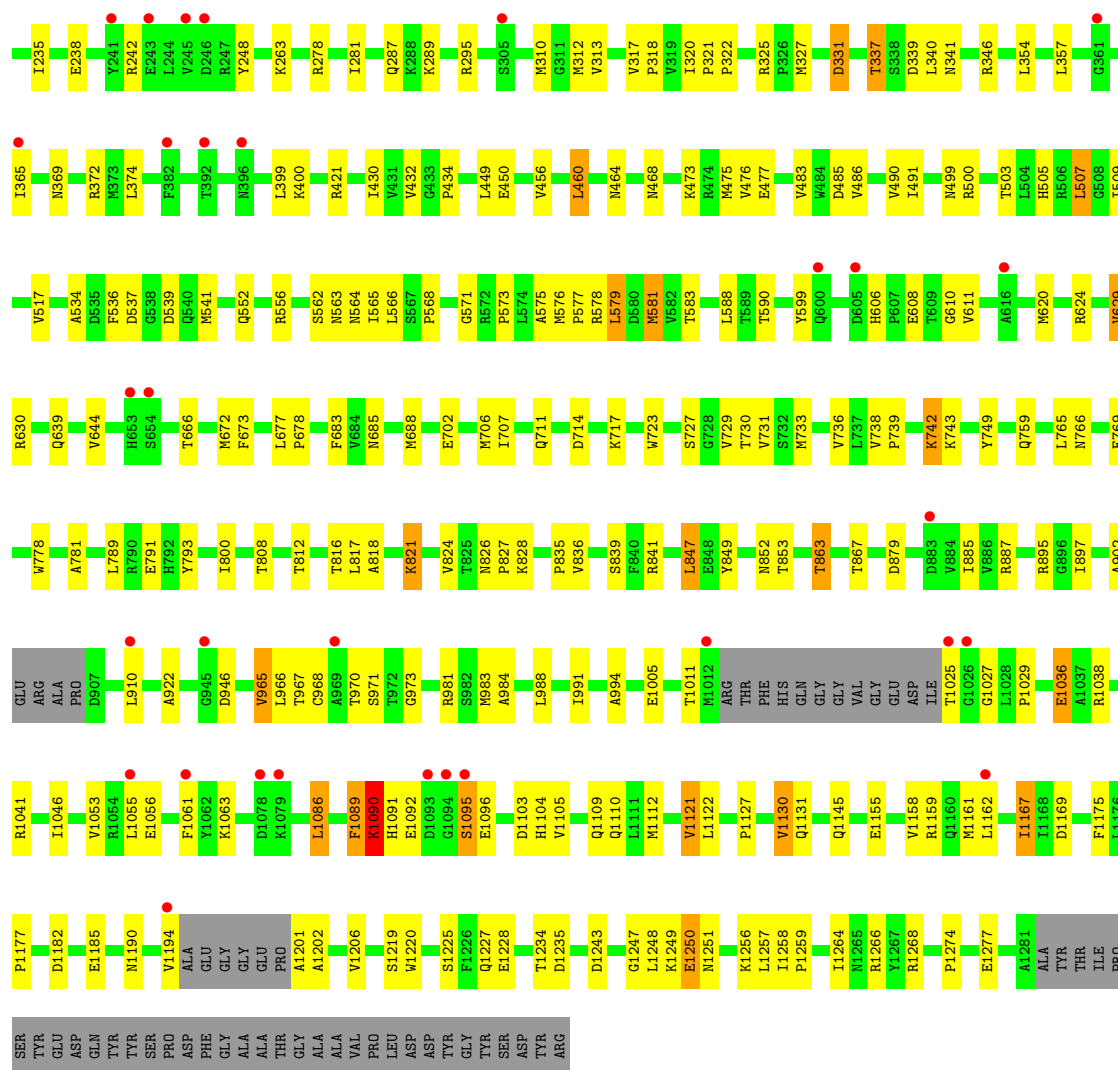
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	D	2	Total	Zn	0	0
			2	2		

- Molecule 10 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

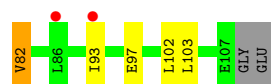
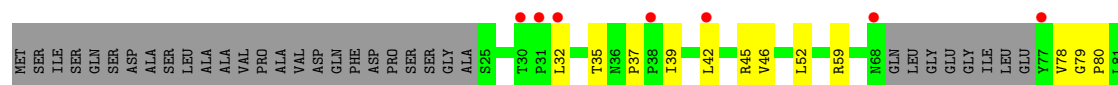
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	D	1	Total 1	Mg 1	0	0

- Molecule 11 is water.

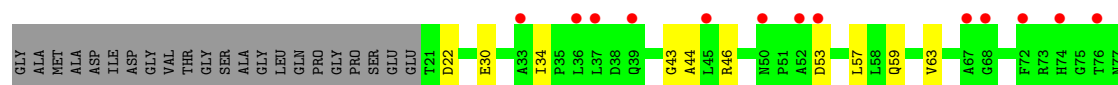
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
11	C	8	Total 8	O 8	0	0
11	D	12	Total 12	O 12	0	0
11	F	4	Total 4	O 4	0	0
11	H	3	Total 3	O 3	0	0
11	I	2	Total 2	O 2	0	0

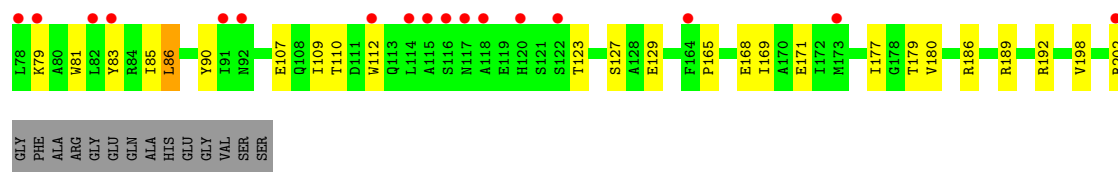


• Molecule 4: DNA-directed RNA polymerase subunit omega

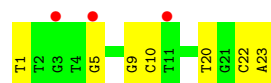


• Molecule 5: ECF RNA polymerase sigma factor SigH

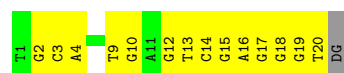
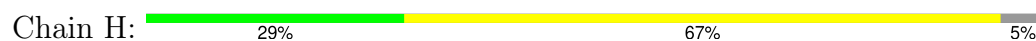




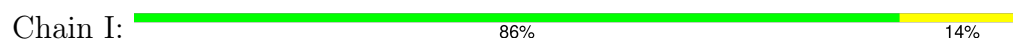
- Molecule 6: DNA (5'-D(*TP*TP*GP*TP*GP*GP*GP*AP*GP*CP*TP*GP*TP*CP*AP*CP*GP*GP*AP*TP*GP*CP*A)-3')



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



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



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	125.03Å 162.31Å 128.48Å 90.00° 117.03° 90.00°	Depositor
Resolution (Å)	31.38 – 2.80 31.38 – 2.80	Depositor EDS
% Data completeness (in resolution range)	96.9 (31.38-2.80) 97.4 (31.38-2.80)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.77 (at 2.81Å)	Xtriage
Refinement program	PHENIX 1.14_3260	Depositor
R, R_{free}	0.206 , 0.248 0.208 , 0.250	Depositor DCC
R_{free} test set	2398 reflections (2.18%)	wwPDB-VP
Wilson B-factor (Å ²)	77.6	Xtriage
Anisotropy	0.424	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 55.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.016 for l,-k,h	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	24837	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 2.35% 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	0.24	0/1671	0.48	0/2273
1	B	0.33	0/1695	0.49	1/2314 (0.0%)
2	C	0.33	1/8830 (0.0%)	0.48	0/11988
3	D	0.28	1/9923 (0.0%)	0.46	1/13425 (0.0%)
4	E	0.20	0/604	0.40	0/822
5	F	0.21	0/1460	0.38	0/1983
6	G	0.27	0/533	0.47	0/823
7	H	0.43	0/462	0.50	0/713
8	I	0.28	0/157	0.50	0/242
All	All	0.30	2/25335 (0.0%)	0.47	2/34583 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	77	VAL	C-O	-5.61	1.18	1.23
3	D	1250	GLU	C-O	-5.52	1.17	1.24

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	228	GLU	N-CA-C	8.92	120.42	108.38
3	D	1090	LYS	N-CA-C	5.59	117.01	108.52

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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	1646	0	1676	28	0
1	B	1670	0	1651	46	0
2	C	8673	0	8508	187	0
3	D	9763	0	9772	205	0
4	E	592	0	587	11	0
5	F	1432	0	1359	22	0
6	G	475	0	261	6	0
7	H	412	0	227	22	0
8	I	142	0	78	1	0
9	D	2	0	0	0	0
10	D	1	0	0	0	0
11	C	8	0	0	0	0
11	D	12	0	0	0	0
11	F	4	0	0	0	0
11	H	3	0	0	0	0
11	I	2	0	0	0	0
All	All	24837	0	24119	476	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 10.

The worst 5 of 476 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)
7:H:14:DC:H2'	7:H:15:DG:H5''	1.20	1.11
7:H:14:DC:C2'	7:H:15:DG:H5''	1.96	0.95
3:D:1090:LYS:HG3	3:D:1095:SER:C	1.96	0.90
3:D:1090:LYS:HG3	3:D:1095:SER:O	1.75	0.87
7:H:19:DG:C2'	7:H:20:DT:H5'	2.06	0.86

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	213/368 (58%)	208 (98%)	5 (2%)	0	100	100
1	B	224/368 (61%)	211 (94%)	13 (6%)	0	100	100
2	C	1131/1174 (96%)	1082 (96%)	48 (4%)	1 (0%)	48	77
3	D	1250/1317 (95%)	1211 (97%)	39 (3%)	0	100	100
4	E	71/110 (64%)	68 (96%)	3 (4%)	0	100	100
5	F	180/218 (83%)	177 (98%)	3 (2%)	0	100	100
All	All	3069/3555 (86%)	2957 (96%)	111 (4%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	76	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	185/315 (59%)	177 (96%)	8 (4%)	26	60
1	B	177/315 (56%)	159 (90%)	18 (10%)	7	23
2	C	929/995 (93%)	866 (93%)	63 (7%)	14	42
3	D	1026/1096 (94%)	972 (95%)	54 (5%)	20	52
4	E	64/90 (71%)	62 (97%)	2 (3%)	35	70
5	F	143/175 (82%)	135 (94%)	8 (6%)	19	50

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	2524/2986 (84%)	2371 (94%)	153 (6%)	17 46

5 of 153 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	590	THR
4	E	45	ARG
3	D	644	VAL
3	D	971	SER
5	F	179	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	854	HIS
3	D	935	ASN
5	F	120	HIS
5	F	59	GLN
5	F	100	GLN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	I	6/7 (85%)	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 ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry

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

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, 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	217/368 (58%)	0.27	2 (0%) 81 74	53, 75, 108, 128	0
1	B	230/368 (62%)	0.88	25 (10%) 10 8	75, 112, 147, 156	0
2	C	1135/1174 (96%)	0.41	67 (5%) 28 21	47, 75, 147, 169	0
3	D	1258/1317 (95%)	0.37	56 (4%) 38 30	48, 78, 122, 153	0
4	E	75/110 (68%)	0.76	9 (12%) 9 6	73, 99, 138, 175	0
5	F	182/218 (83%)	1.00	30 (16%) 4 3	32, 103, 132, 147	12 (6%)
6	G	23/23 (100%)	0.58	3 (13%) 7 6	87, 105, 148, 161	3 (13%)
7	H	20/21 (95%)	-0.15	0 100 100	54, 78, 136, 137	0
8	I	7/7 (100%)	-0.68	0 100 100	53, 54, 81, 87	0
All	All	3147/3606 (87%)	0.46	192 (6%) 27 20	32, 81, 138, 175	15 (0%)

The worst 5 of 192 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	F	114	LEU	6.4
5	F	68	GLY	5.7
3	D	201	GLY	5.4
5	F	45	LEU	5.2
1	B	236	PRO	5.1

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 oligosaccharides 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($\text{\AA}^2$)	Q<0.9
10	MG	D	2003	1/1	0.88	0.11	54,54,54,54	0
9	ZN	D	2002	1/1	0.99	0.03	105,105,105,105	0
9	ZN	D	2001	1/1	0.99	0.02	87,87,87,87	0

6.5 Other polymers [i](#)

There are no such residues in this entry.