



wwPDB X-ray Structure Validation Summary Report ⓘ

May 16, 2020 – 11:49 am BST

PDB ID : 6TYG
Title : Crystal structure of MTB sigma L transcription initiation complex with 9 nt long RNA primer
Authors : Molodtsov, V.; Ebright, R.H.
Deposited on : 2019-08-08
Resolution : 3.50 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 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

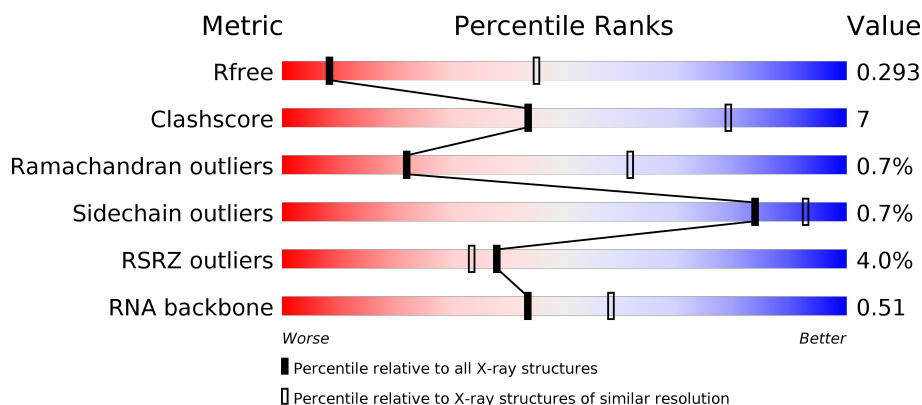
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.50 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	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)
RSRZ outliers	127900	1559 (3.60-3.40)
RNA backbone	3102	1002 (4.00-3.00)

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 ≥ 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>2%</div> <div> <div></div> <div>60%</div> <div>21%</div> <div>•</div> <div>17%</div> </div> </div>
2	G	20	<div> <div>60%</div> <div>40%</div> </div>
3	H	27	<div> <div>4%</div> <div> <div></div> <div>37%</div> <div>48%</div> <div>15%</div> </div> </div>
4	I	9	<div> <div>22%</div> <div>33%</div> <div>44%</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></div></div><div><div>%</div><div>54%</div><div>11%</div><div>•</div><div>35%</div></div></div>
5	B	347	<div><div><div></div><div></div><div></div></div><div><div>2%</div><div>51%</div><div>15%</div><div>•</div><div>33%</div></div></div>
6	C	1178	<div><div><div></div><div></div><div></div></div><div><div>%</div><div>82%</div><div>13%</div><div>5%</div></div></div>
7	D	1316	<div><div><div></div><div></div><div></div></div><div><div>7%</div><div>82%</div><div>14%</div><div>•</div></div></div>
8	E	110	<div><div><div></div><div></div><div></div></div><div><div>4%</div><div>64%</div><div>10%</div><div>26%</div></div></div>

2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 49186 atoms, of which 24383 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	147	Total	C	H	N	O	S	0	0	0
			2286	709	1149	216	210	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*GP*GP*TP*G)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	G	20	Total	C	H	N	O	P	0	0	0
			639	196	226	80	118	19			

- 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*AP*CP*CP*CP*UP*CP*GP*A)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
4	I	9	Total	C	H	N	O	P	0	0	0
			284	84	99	32	60	9			

- 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			

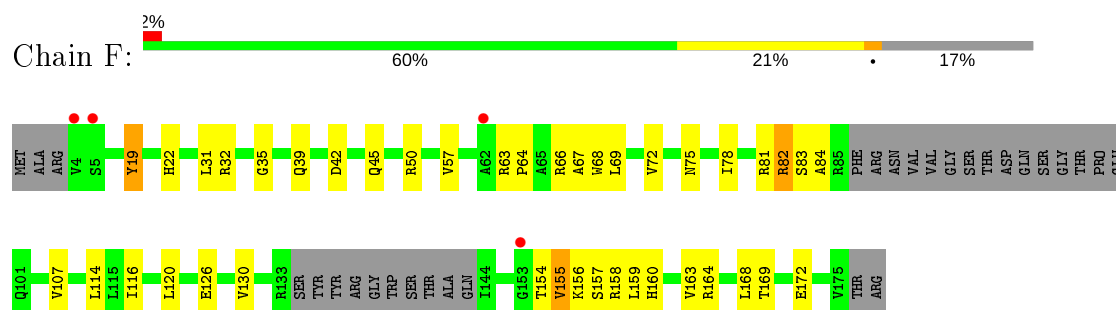
- Molecule 8 is a protein called DNA-directed RNA polymerase subunit omega.

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.

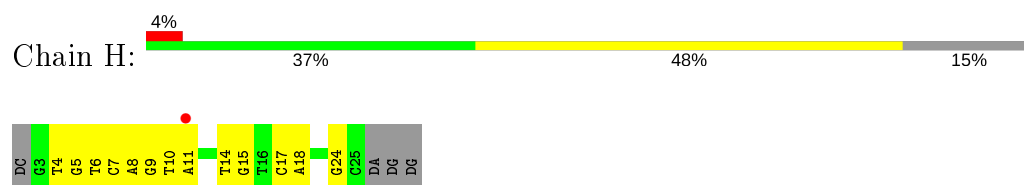
- Molecule 1: RNA polymerase sigma factor



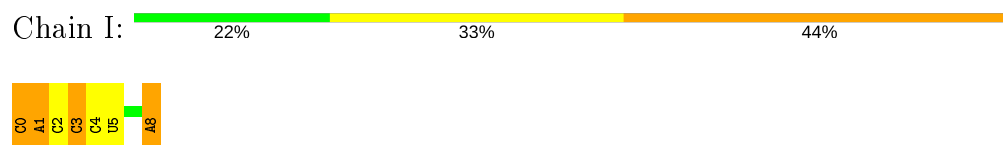
- Molecule 2: DNA (5'-D(*GP*CP*AP*TP*CP*CP*GP*TP*GP*AP*AP*TP*CP*GP*AP*GP*GP*GP*TP*G)-3')



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



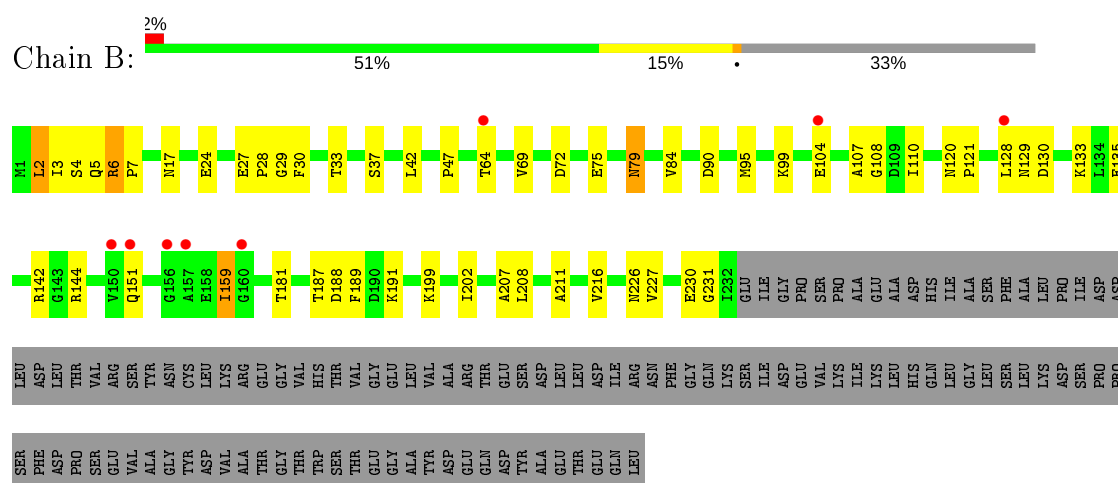
- Molecule 4: RNA (5'-R(P*CP*AP*CP*CP*CP*UP*CP*GP*A)-3')



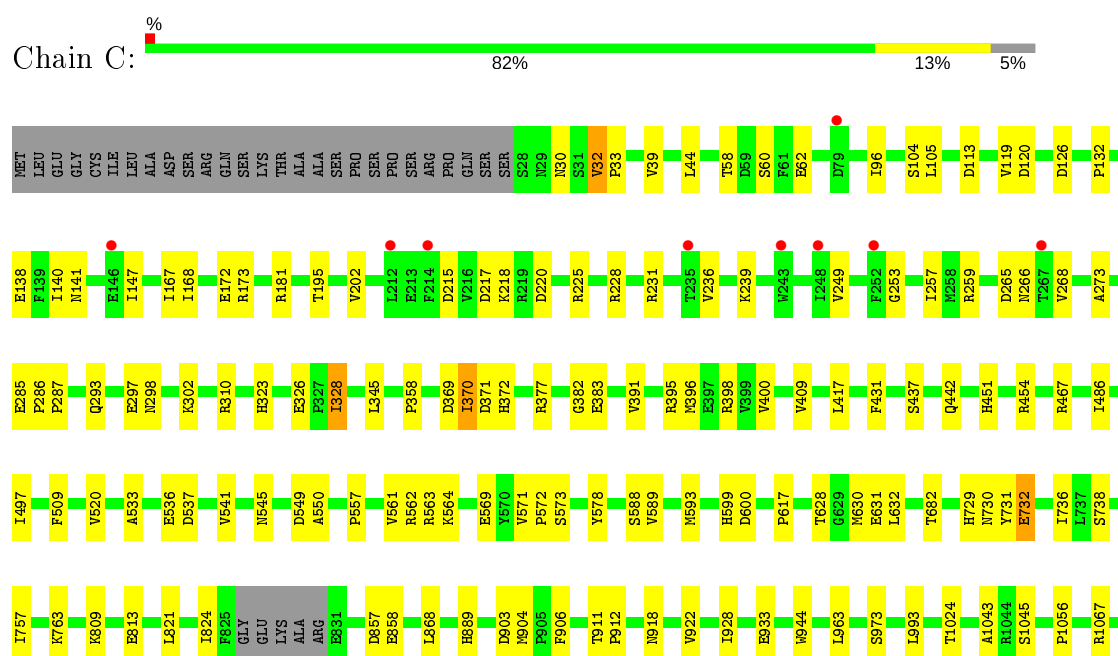
- Molecule 5: DNA-directed RNA polymerase subunit alpha



- Molecule 5: DNA-directed RNA polymerase subunit alpha

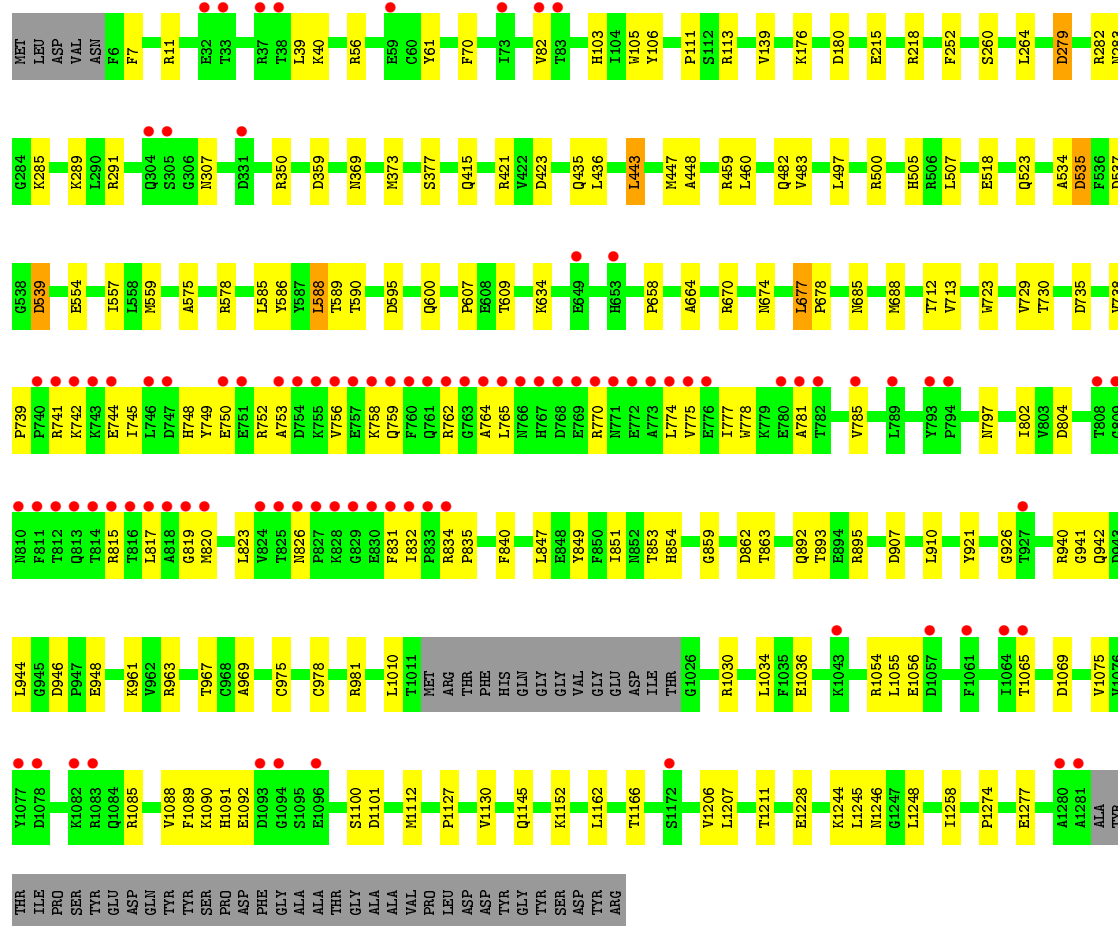
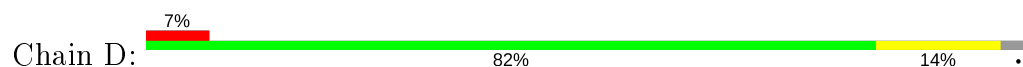


- Molecule 6: DNA-directed RNA polymerase subunit beta

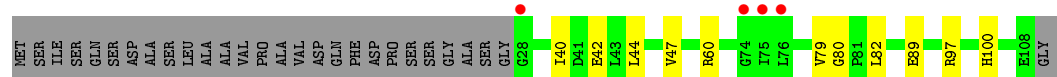




• Molecule 7: DNA-directed RNA polymerase subunit beta'



• 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, α , β , γ	128.65Å 158.23Å 215.25Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.64 – 3.50 49.64 – 3.50	Depositor EDS
% Data completeness (in resolution range)	99.5 (49.64-3.50) 99.5 (49.64-3.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	6.76 (at 3.48Å)	Xtriage
Refinement program	PHENIX 1.16_3549, PHENIX 1.16_3549	Depositor
R, R_{free}	0.246 , 0.293 0.246 , 0.293	Depositor DCC
R_{free} test set	1982 reflections (3.55%)	wwPDB-VP
Wilson B-factor (Å ²)	84.6	Xtriage
Anisotropy	0.711	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 41.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.53$, $\langle L^2 \rangle = 0.38$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	49186	wwPDB-VP
Average B, all atoms (Å ²)	120.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.40% 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](#)

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	0/1151	0.47	0/1562
2	G	0.75	0/464	0.99	0/716
3	H	0.60	0/532	0.94	0/820
4	I	0.51	0/205	1.21	1/314 (0.3%)
5	A	0.26	0/1742	0.51	0/2370
5	B	0.27	0/1758	0.50	0/2397
6	C	0.25	0/8801	0.44	0/11933
7	D	0.25	0/10038	0.44	0/13568
8	E	0.24	0/643	0.40	0/877
All	All	0.29	0/25334	0.50	1/34557 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
5	A	0	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	I	4	C	C5-C4-N4	-5.28	116.51	120.20

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
5	A	4	SER	Peptide

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	F	1137	1149	1149	30	0
2	G	413	226	226	5	0
3	H	475	259	260	18	0
4	I	185	99	99	8	0
5	A	1716	1756	1756	30	0
5	B	1732	1754	1754	38	0
6	C	8643	8576	8575	111	0
7	D	9872	9942	9942	134	1
8	E	630	622	622	9	0
All	All	24803	24383	24383	328	1

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 7.

The worst 5 of 328 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	1.34	0.93
6:C:369:ASP:O	6:C:370:ILE:HG12	1.66	0.93
7:D:770:ARG:O	7:D:774:LEU:HD12	1.79	0.83
6:C:593:MET:HA	6:C:628:THR:HG21	1.61	0.81
5:B:75:GLU:O	5:B:79:ASN:ND2	2.15	0.80

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
7:D:748:HIS:NE2	7:D:892:GLN:OE1[4_545]	2.18	0.02

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	141/177 (80%)	120 (85%)	17 (12%)	4 (3%)	5	32
5	A	223/347 (64%)	213 (96%)	9 (4%)	1 (0%)	34	72
5	B	230/347 (66%)	212 (92%)	14 (6%)	4 (2%)	9	42
6	C	1110/1178 (94%)	1049 (94%)	53 (5%)	8 (1%)	22	61
7	D	1258/1316 (96%)	1204 (96%)	49 (4%)	5 (0%)	34	72
8	E	79/110 (72%)	74 (94%)	5 (6%)	0	100	100
All	All	3041/3475 (88%)	2872 (94%)	147 (5%)	22 (1%)	22	61

5 of 22 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	F	82	ARG
5	A	184	GLU
5	B	6	ARG
5	B	159	ILE
6	C	370	ILE

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	F	111/136 (82%)	110 (99%)	1 (1%)	78	90
5	A	194/297 (65%)	193 (100%)	1 (0%)	88	94
5	B	191/297 (64%)	189 (99%)	2 (1%)	76	88

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
6	C	945/998 (95%)	944 (100%)	1 (0%)	93	98
7	D	1047/1095 (96%)	1033 (99%)	14 (1%)	69	86
8	E	66/90 (73%)	66 (100%)	0	100	100
All	All	2554/2913 (88%)	2535 (99%)	19 (1%)	84	93

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
7	D	443	LEU
7	D	535	ASP
7	D	804	ASP
7	D	359	ASP
7	D	862	ASP

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

Mol	Chain	Res	Type
1	F	160	HIS
7	D	854	HIS

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
4	I	9/9 (100%)	3 (33%)	1 (11%)

All (3) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
4	I	1	A
4	I	3	C
4	I	8	A

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
4	I	0	C

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 ⓘ

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, 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	F	147/177 (83%)	0.20	4 (2%) 54 48	86, 148, 196, 225	0
2	G	20/20 (100%)	-0.20	0 100 100	59, 89, 145, 160	0
3	H	23/27 (85%)	0.29	1 (4%) 35 31	103, 143, 217, 219	0
4	I	9/9 (100%)	0.60	0 100 100	61, 70, 168, 172	0
5	A	225/347 (64%)	-0.13	2 (0%) 84 79	62, 89, 143, 198	0
5	B	232/347 (66%)	0.14	8 (3%) 45 40	75, 105, 158, 224	0
6	C	1114/1178 (94%)	-0.03	11 (0%) 82 77	48, 84, 163, 201	0
7	D	1262/1316 (95%)	0.32	93 (7%) 14 14	48, 99, 240, 330	0
8	E	81/110 (73%)	0.17	4 (4%) 29 26	79, 112, 177, 200	0
All	All	3113/3531 (88%)	0.14	123 (3%) 38 33	48, 97, 190, 330	0

The worst 5 of 123 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
7	D	763	GLY	11.6
7	D	771	ASN	11.1
7	D	828	LYS	10.9
7	D	814	THR	9.9
7	D	829	GLY	9.6

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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.