WORLDWIDE PDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 10, 2023 – 04:08 AM EDT

PDB ID	:	7MLJ
Title	:	Crystal structure of Thermus thermophilus reiterative transcription complex with 4nt oligo-G RNA
Authors	:	Liu, Y.; Ebright, R.H.
Deposited on	:	2021-04-28
Resolution	:	3.75 Å(reported)

The following versions of software and data (see references ⓘ) were used in the production of this report:

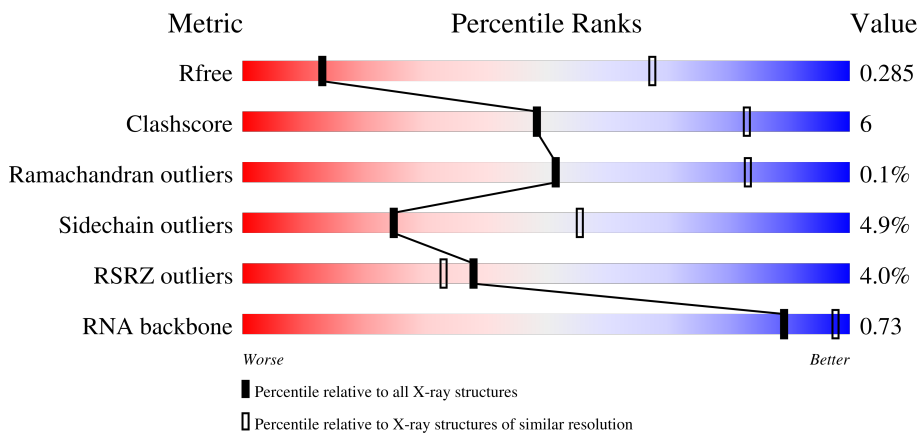|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 3.75 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.
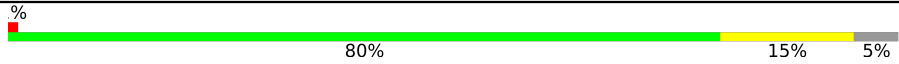
Metric	Percentile Ranks	Value
Rfree		0.285
Clashscore		6
Ramachandran outliers		0.1%
Sidechain outliers		4.9%
RSRZ outliers		4.0%
RNA backbone		0.73

Worse — *Better*

■ Percentile relative to all X-ray structures

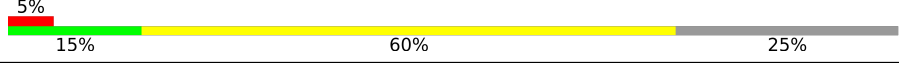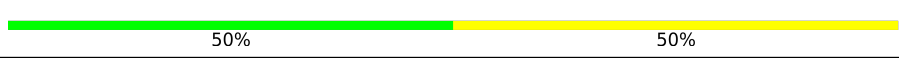▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1039 (3.94-3.58)
Clashscore	141614	1051 (3.92-3.60)
Ramachandran outliers	138981	1015 (3.92-3.60)
Sidechain outliers	138945	1011 (3.92-3.60)
RSRZ outliers	127900	1050 (3.96-3.56)
RNA backbone	3102	1035 (4.52-3.00)

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 <=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	315	2% — 59% / 13% / 28%
1	B	315	57% / 12% / • / 30%
2	C	1119	4% — 81% / 17% / ••
3	D	1524	5% — 78% / 18% / ••

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	99	% ▇ 80% 15% 5%
5	F	443	5% 64% 14% 22%
6	G	20	5% 15% 60% 25%
7	I	4	50% 50%
8	H	27	41% 44% 15%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
9	MG	B	2002	-	-	-	X

2 Entry composition ⓘ

There are 10 unique types of molecules in this entry. The entry contains 28489 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	226	Total 1782	C 1138	N 310	O 332	S 2	0	0	0
1	B	222	Total 1750	C 1118	N 304	O 326	S 2	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1111	Total 8770	C 5548	N 1564	O 1634	S 24	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1486	Total 11738	C 7441	N 2067	O 2195	S 35	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	94	Total 761	C 486	N 132	O 139	S 4	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	346	Total 2807	C 1770	N 509	O 524	S 4	0	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-19	MET	-	initiating methionine	UNP Q5SKW1
F	-18	GLY	-	expression tag	UNP Q5SKW1
F	-17	SER	-	expression tag	UNP Q5SKW1
F	-16	SER	-	expression tag	UNP Q5SKW1
F	-15	HIS	-	expression tag	UNP Q5SKW1
F	-14	HIS	-	expression tag	UNP Q5SKW1
F	-13	HIS	-	expression tag	UNP Q5SKW1
F	-12	HIS	-	expression tag	UNP Q5SKW1
F	-11	HIS	-	expression tag	UNP Q5SKW1
F	-10	HIS	-	expression tag	UNP Q5SKW1
F	-9	SER	-	expression tag	UNP Q5SKW1
F	-8	SER	-	expression tag	UNP Q5SKW1
F	-7	GLY	-	expression tag	UNP Q5SKW1
F	-6	LEU	-	expression tag	UNP Q5SKW1
F	-5	VAL	-	expression tag	UNP Q5SKW1
F	-4	PRO	-	expression tag	UNP Q5SKW1
F	-3	ARG	-	expression tag	UNP Q5SKW1
F	-2	GLY	-	expression tag	UNP Q5SKW1
F	-1	SER	-	expression tag	UNP Q5SKW1
F	0	HIS	-	expression tag	UNP Q5SKW1

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	15	Total 304	C 144	N 54	O 91	P 15	0	0	0

- Molecule 7 is a RNA chain called RNA (5'-R(P*GP*GP*GP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	I	4	Total 92	C 40	N 20	O 28	P 4	0	0	0

- Molecule 8 is a DNA chain called DNA (5'-D(*TP*AP*TP*AP*AP*TP*GP*GP*GP*AP*GP*CP*TP*GP*GP*CP*AP*CP*GP*GP*AP*TP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	H	23	Total 478	C 227	N 94	O 135	P 22	0	0	0

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

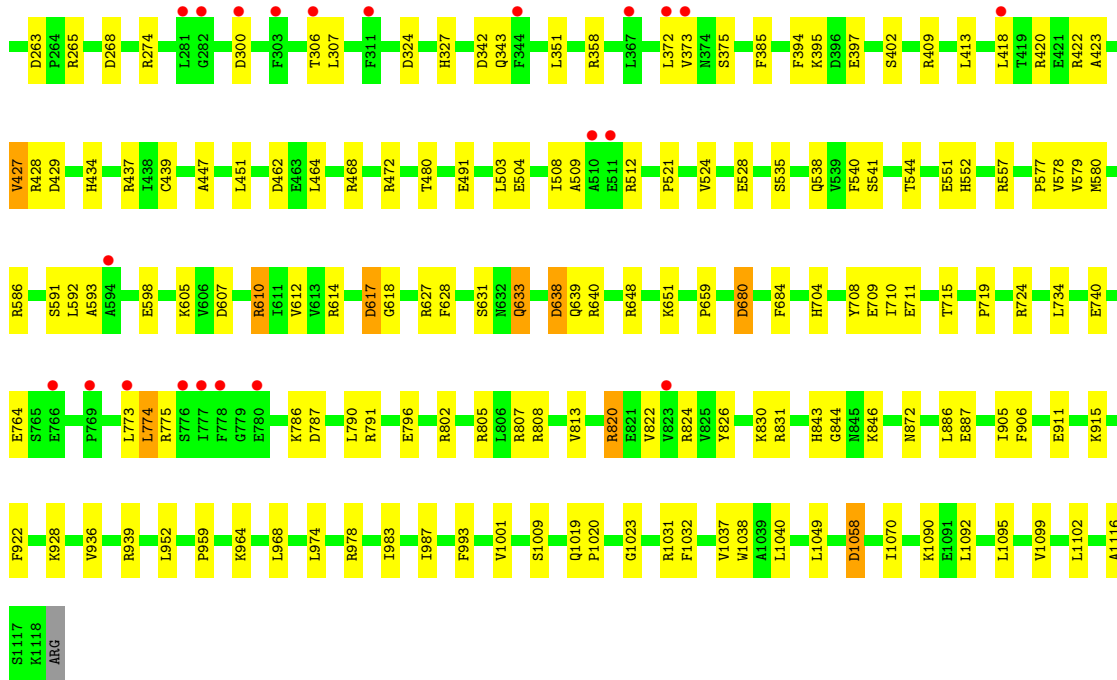| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	B	2	Total 2	Mg 2	0	0
9	D	2	Total 2	Mg 2	0	0
9	F	1	Total 1	Mg 1	0	0

- Molecule 10 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

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'

L930 F939 T943 I947 I951 D952 E959 K960 L971 M980 L983 T984 Y989 L993 W996 T1000 V1003 R1036 L1041 R1042 Q1046 K1047 T1052 F1053 E1054 V1055 P1056 V1057 R1062 E1063 L1068 I1072 T1084 T1088 V1099 D1100

V1101 T1102 R1108 C1112 E1127 V1128 T1129 R1130 R1137 E1141 R1147 E1152 V1155 E1161 E1162 V1188 Q1195 T1196 R1197 Y1198 D1208 V1215 E1219 G1222 L1236 T1237 MET ARG THR PHE HIS THR GLY GLY VAL ALA GLY ALA ALA ASP I1252 T1253 Q1254

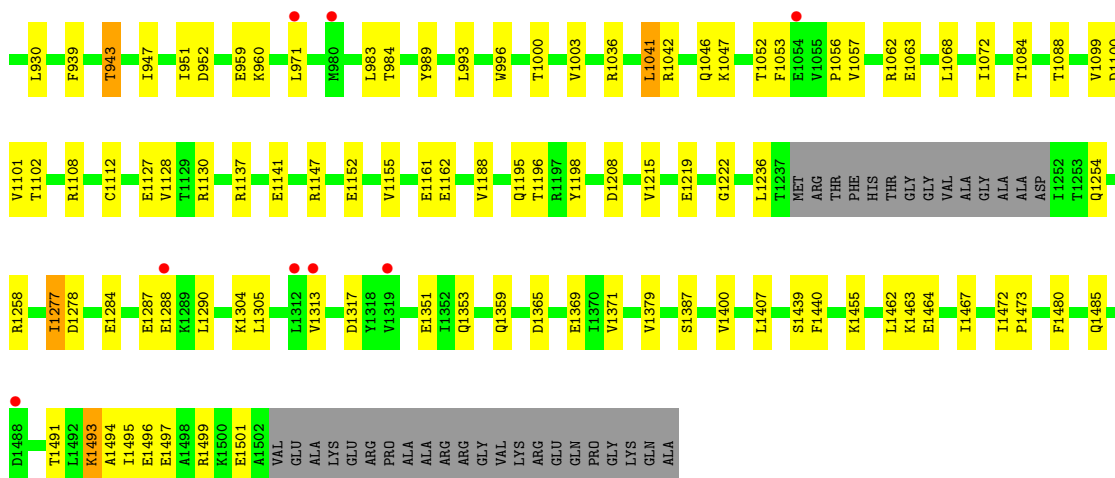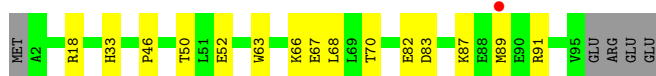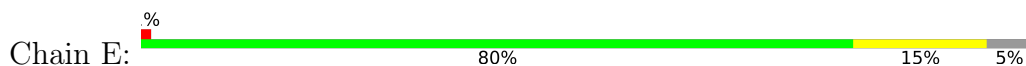R1258 I1277 D1278 E1284 E1287 K1288 K1289 L1290 K1304 L1305 L1312 V1313 D1317 Y1318 V1319 E1351 I1352 Q1353 Q1359 D1365 E1369 I1370 V1371 V1379 S1387 V1400 L1407 S1439 F1440 K1455 L1462 K1463 E1464 I1467 I1472 P1473 F1480 Q1485

D1488 T1491 L1492 K1493 A1494 I1495 E1496 E1497 A1498 R1499 K1500 E1501 A1502 VAL GLU ALA LYS GLU ARG PRO ALA ALA ARG GLY VAL LYS ARG GLU GLN PRO GLY LYS GLN ALA

- Molecule 4: DNA-directed RNA polymerase subunit omega

Chain E:

MET A2 R18 H33 P46 T50 L51 E52 W63 K66 E67 L68 L69 T70 E82 D83 K87 E88 M89 E90 R91 V95 GLU ARG GLU GLU

- Molecule 5: RNA polymerase sigma factor SigA

Chain F:

MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS MET LYS LYS SER LYS ARG LYS ASN ALA GLN ALA GLN GLU ALA GLN GLU THR GLU VAL LEU VAL GLN GLU GLU ALA GLU GLU LEU PRO GLU PHE PRO GLU GLY GLU PRO ASP PRO ASP LEU

GLU ASP PRO ASP LEU THR LEU GLU ASP ASP LEU LEU ASP PRO GLU GLY GLY GLY LEU LEU GLU ASP GLU GLU GLU ASP ALA LEU PRO ILE LYS ILE SER THR S78 V81 R82 Q83 E87 I88 V131 R140 V141 T150 P153 V156 L163 K164 S165

L166 P167 H170 K171 R172 Y173 L174 H175 I176 A177 R178 H186 L187 I188 L192 R193 K201 R205 F209 L210 I213 Q214 Q218 I221 V224 Y229 K230 F235 S236 T237 R244 K270 L278 E281 T287 M295 V315 L350 L354

L361 R364 E365 R372 K373 G374 L375 I376 D377 G378 H381 V386 F389 F390 G391 V392 T393 R396 I397 I400 K403 A404 K407 L408 K417 L418 F421 L422 D423

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

Chain G:

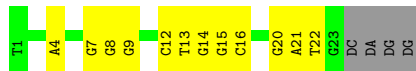DC DC DT DG C6 A7 T8 C9 C10 G11 T12 G13 C14 C15 C16 T17 G18 A19 G20 DG

- Molecule 7: RNA (5'-R(P*GP*GP*GP*G)-3')

Chain I:

G1 G4

• Molecule 8: DNA (5'-D(*TP*AP*TP*AP*AP*TP*GP*GP*GP*AP*GP*CP*TP*GP*GP*CP*AP*CP*GP*GP*AP*TP*G)-3')

Chain H:

41%　　　　44%　　　　15%

T1 A4 G7 G8 G9 C12 T13 G14 G15 C16 G20 A21 T22 G23 DC DA DG DG

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	186.31Å 104.66Å 298.41Å 90.00° 98.12° 90.00°	Depositor
Resolution (Å)	46.21 – 3.75 49.32 – 3.75	Depositor EDS
% Data completeness (in resolution range)	97.1 (46.21-3.75) 97.5 (49.32-3.75)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	0.10	Depositor
$<I/\sigma(I)>$ [1]	2.65 (at 3.77Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.265 , 0.283 0.265 , 0.285	Depositor DCC
R_{free} test set	2009 reflections (3.51%)	wwPDB-VP
Wilson B-factor (Å^2)	106.3	Xtriage
Anisotropy	0.807	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 80.9	EDS
L-test for twinning [2]	$<|L|> = 0.46, <L^2> = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	28489	wwPDB-VP
Average B, all atoms (Å^2)	152.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.79% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $<|L|>$, $<L^2>$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5   Model quality ⓘ

5.1   Standard geometry ⓘ

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	0.25	0/1814	0.51	0/2466
1	B	0.25	0/1782	0.52	0/2424
2	C	0.25	0/8937	0.52	0/12087
3	D	0.25	0/11944	0.50	0/16149
4	E	0.24	0/775	0.47	0/1045
5	F	0.24	0/2852	0.49	0/3837
6	G	0.58	0/339	0.93	0/520
7	I	0.36	0/103	1.04	0/160
8	H	0.53	0/538	0.90	0/831
All	All	0.26	0/29084	0.53	0/39519

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 ⓘ

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	1782	0	1834	23	0
1	B	1750	0	1797	24	0
2	C	8770	0	8874	113	0
3	D	11738	0	11971	154	0

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	E	761	0	778	12	0
5	F	2807	0	2882	40	0
6	G	304	0	169	9	0
7	I	92	0	44	2	0
8	H	478	0	260	18	0
9	B	2	0	0	0	0
9	D	2	0	0	0	0
9	F	1	0	0	0	0
10	D	2	0	0	0	0
All	All	28489	0	28609	344	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 6.

The worst 5 of 344 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)
3:D:1491:THR:HG21	4:E:89:MET:HG2	1.63	0.79
2:C:773:LEU:HB2	5:F:375:LEU:HD11	1.67	0.77
1:B:206:THR:HG22	1:B:209:GLU:H	1.52	0.74
2:C:462:ASP:HB3	2:C:468:ARG:HD2	1.69	0.74
2:C:422:ARG:HA	8:H:16:DC:H5'	1.68	0.74

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	224/315 (71%)	219 (98%)	5 (2%)	0	100	100
1	B	220/315 (70%)	215 (98%)	5 (2%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	C	1107/1119 (99%)	1082 (98%)	23 (2%)	2 (0%)	47	78
3	D	1482/1524 (97%)	1449 (98%)	31 (2%)	2 (0%)	51	83
4	E	92/99 (93%)	91 (99%)	1 (1%)	0	100	100
5	F	344/443 (78%)	338 (98%)	6 (2%)	0	100	100
All	All	3469/3815 (91%)	3394 (98%)	71 (2%)	4 (0%)	51	83

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	592	LEU
3	D	1440	PHE
2	C	215	GLY
3	D	530	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	A	199/273 (73%)	192 (96%)	7 (4%)	36	63
1	B	195/273 (71%)	190 (97%)	5 (3%)	46	70
2	C	936/941 (100%)	885 (95%)	51 (5%)	22	54
3	D	1253/1279 (98%)	1185 (95%)	68 (5%)	22	54
4	E	83/88 (94%)	82 (99%)	1 (1%)	71	84
5	F	301/388 (78%)	289 (96%)	12 (4%)	31	60
All	All	2967/3242 (92%)	2823 (95%)	144 (5%)	25	55

5 of 144 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	1188	VAL
5	F	422	LEU
3	D	1284	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	D	1501	GLU
2	C	774	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
3	D	350	HIS
3	D	1172	HIS
3	D	1195	GLN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
7	I	3/4 (75%)	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 7 ligands modelled in this entry, 7 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 ⓘ

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(Å²)	Q<0.9
1	A	226/315 (71%)	-0.07	5 (2%)	62	56	113, 141, 165, 179	0
1	B	222/315 (70%)	-0.18	0	100	100	105, 148, 179, 195	0
2	C	1111/1119 (99%)	0.23	40 (3%)	42	37	83, 151, 200, 219	0
3	D	1486/1524 (97%)	0.26	74 (4%)	28	26	90, 142, 197, 245	1 (0%)
4	E	94/99 (94%)	-0.04	1 (1%)	80	77	123, 166, 198, 211	0
5	F	346/443 (78%)	0.26	20 (5%)	23	19	112, 160, 213, 223	0
6	G	15/20 (75%)	-0.27	1 (6%)	17	14	130, 168, 224, 224	0
7	I	4/4 (100%)	1.32	0	100	100	155, 174, 199, 203	0
8	H	23/27 (85%)	0.10	0	100	100	168, 198, 281, 298	0
All	All	3527/3866 (91%)	0.19	141 (3%)	38	33	83, 149, 201, 298	1 (0%)

The worst 5 of 141 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	367	ILE	7.7
3	D	289	THR	6.9
3	D	305	ALA	6.8
3	D	322	VAL	6.5
3	D	355	VAL	6.5

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 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(Å²)	Q<0.9
9	MG	B	2002	1/1	0.74	0.48	92,92,92,92	0
9	MG	B	2001	1/1	0.83	0.38	139,139,139,139	0
10	ZN	D	2002	1/1	0.90	0.06	197,197,197,197	0
9	MG	F	2001	1/1	0.91	0.08	147,147,147,147	0
9	MG	D	2004	1/1	0.91	0.27	149,149,149,149	0
10	ZN	D	2001	1/1	0.95	0.22	123,123,123,123	0
9	MG	D	2003	1/1	0.98	0.40	98,98,98,98	0

6.5   Other polymers (i)

There are no such residues in this entry.