



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 21, 2020 – 03:00 AM BST

PDB ID : 6KQL  
Title : Thermus thermophilus initial transcription complex comprising sigma A and 5'-triphosphate RNA of 4 nt  
Authors : Zhang, Y.; Li, L.; Ebright, R.H.  
Deposited on : 2019-08-18  
Resolution : 2.89 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
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
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)	:	2.13.1

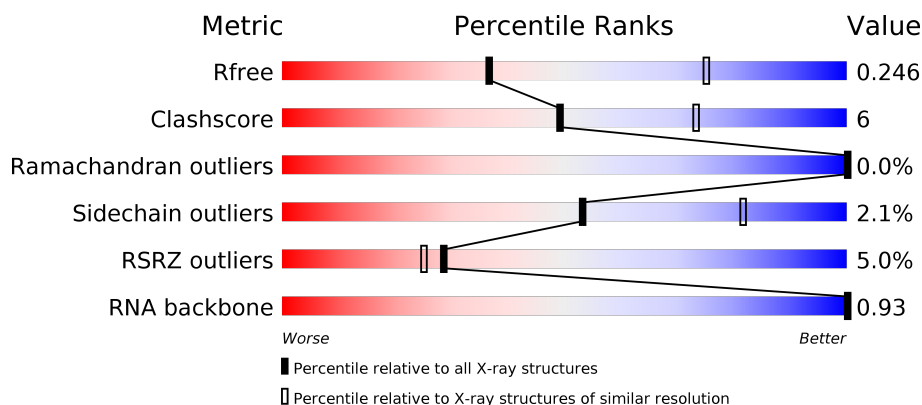
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.89 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)
RNA backbone	3102	1007 (3.16-2.64)

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	315	<div> <div>61%</div> <div>12%</div> <div>27%</div> </div>
1	B	315	<div> <div>3%</div> <div>58%</div> <div>12%</div> <div>28%</div> </div>
2	C	1119	<div> <div>5%</div> <div>84%</div> <div>15%</div> <div>..</div> </div>
3	D	1524	<div> <div>6%</div> <div>81%</div> <div>16%</div> <div>..</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	99	<div><div></div><div>89%</div><div>5% • 5%</div></div>
5	F	443	<div><div>6%</div><div>67%</div><div>11%</div><div>22%</div></div>
6	G	21	<div><div>62%</div><div>19%</div><div>19%</div></div>
7	H	26	<div><div>46%</div><div>46%</div><div>8%</div></div>
8	I	4	<div><div>50%</div><div>50%</div></div>

## 2 Entry composition [i](#)

There are 11 unique types of molecules in this entry. The entry contains 28900 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	231	Total	C	N	O	S	0	0	0
			1809	1155	315	337	2			
1	B	227	Total	C	N	O	S	0	0	0
			1789	1143	310	334	2			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1112	Total	C	N	O	S	0	0	0
			8762	5544	1559	1635	24			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1486	Total	C	N	O	S	0	1	0
			11746	7446	2070	2195	35			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	94	Total	C	N	O	S	0	0	0
			759	484	132	139	4			

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

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

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(\*CP\*CP\*T\*GP\*CP\*AP\*TP\*CP\*CP\*GP\*T  
P\*GP\*AP\*GP\*TP\*CP\*GP\*AP\*GP\*GP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	17	Total	C	N	O	P	0	0	0
			350	166	68	100	16			

- Molecule 7 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\*GP\*CP\*AP\*G\*)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	H	24	Total	C	N	O	P	0	0	0
			495	236	94	142	23			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	I	4	Total	C	N	O	P	0	0	0
			86	38	15	29	4			

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

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

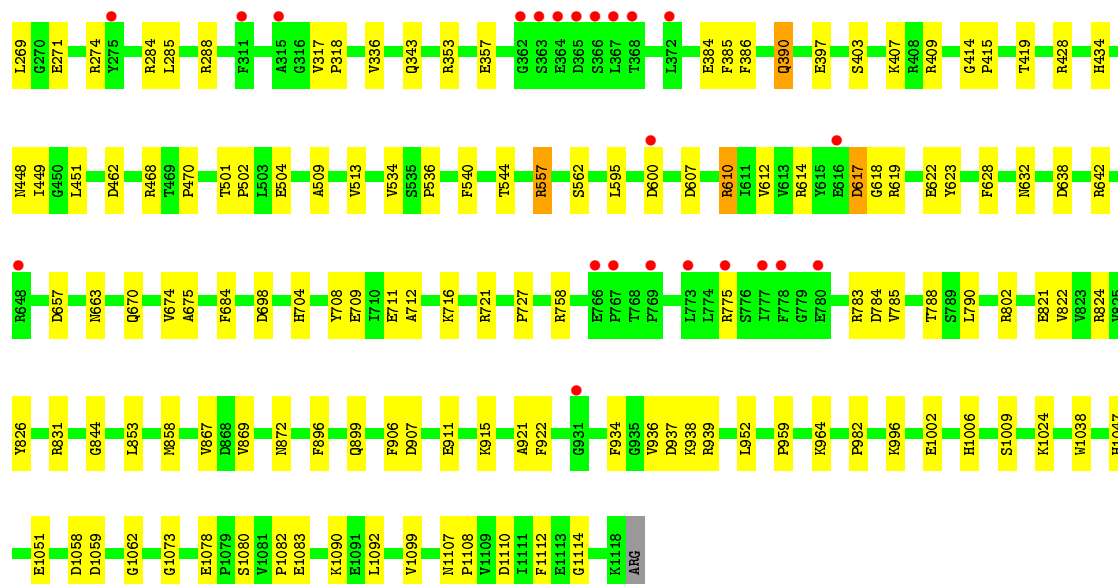
- Molecule 10 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
10	D	3	Total Mg 3 3	0	0
10	F	1	Total Mg 1 1	0	0

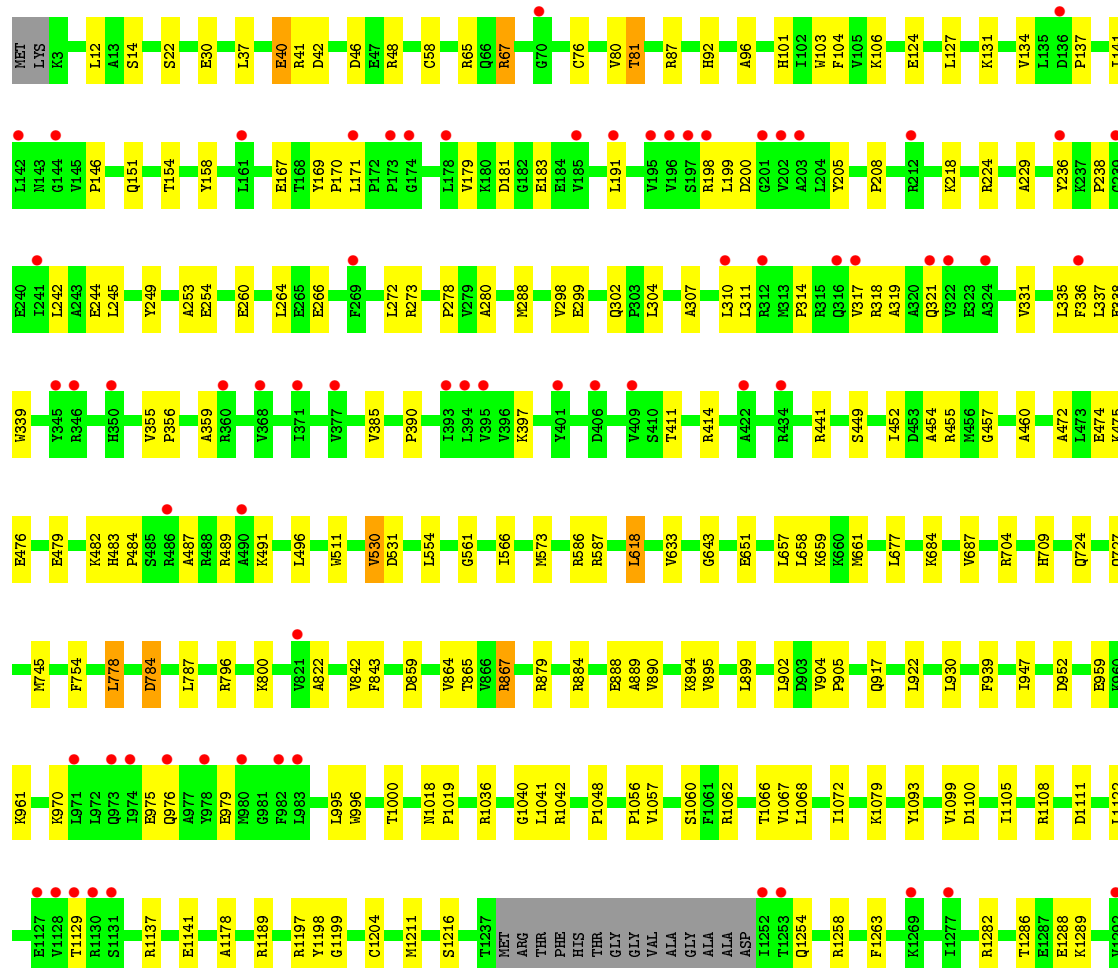
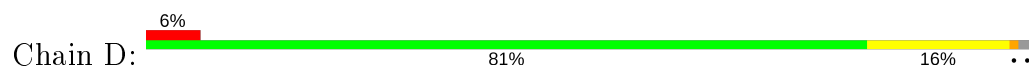
- Molecule 11 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
11	A	11	Total O 11 11	0	0
11	B	8	Total O 8 8	0	0
11	C	96	Total O 96 96	0	0
11	D	128	Total O 128 128	0	0
11	E	8	Total O 8 8	0	0
11	F	26	Total O 26 26	0	0
11	G	10	Total O 10 10	0	0
11	I	4	Total O 4 4	0	0

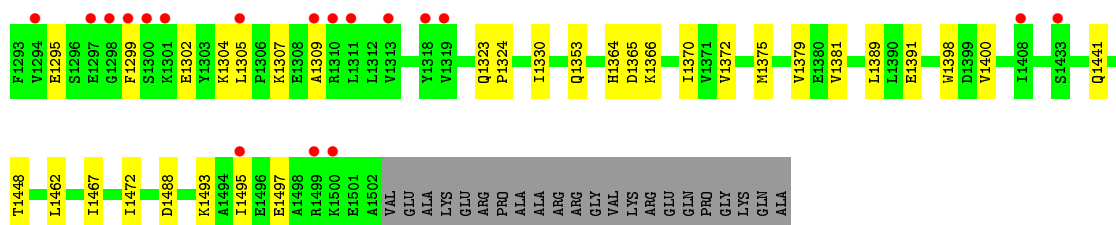




• Molecule 3: DNA-directed RNA polymerase subunit beta'







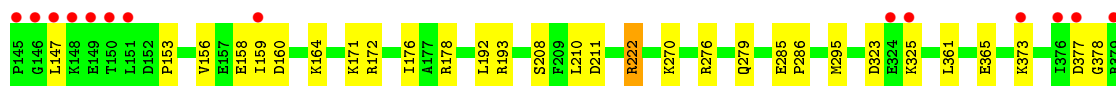
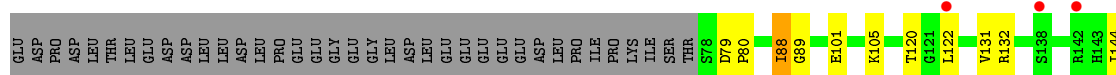
- Molecule 4: DNA-directed RNA polymerase subunit omega

Chain E: 89% 5% 5%



- Molecule 5: RNA polymerase sigma factor SigA

Chain F: 6% 67% 11% 22%



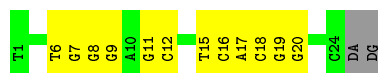
- Molecule 6: DNA (5'-D(\*CP\*CP\*T\*GP\*CP\*AP\*TP\*CP\*CP\*GP\*TP\*GP\*AP\*GP\*TP\*CP\*GP\*AP\*GP\*GP\*G)-3')

Chain G: 62% 19% 19%



- Molecule 7: 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\*GP\*CP\*AP\*G\*)-3')

Chain H: 46% 46% 8%



- Molecule 8: RNA (5'-R\*(UTP))-R(P\*CP\*GP\*A)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	183.80Å 103.17Å 295.56Å 90.00° 99.12° 90.00°	Depositor
Resolution (Å)	44.84 – 2.89 44.84 – 2.89	Depositor EDS
% Data completeness (in resolution range)	98.9 (44.84-2.89) 99.0 (44.84-2.89)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.13 (at 2.90Å)	Xtriage
Refinement program	PHENIX 1.12_2829	Depositor
R, $R_{free}$	0.207 , 0.246 0.207 , 0.246	Depositor DCC
$R_{free}$ test set	2680 reflections (2.21%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	67.8	Xtriage
Anisotropy	0.504	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 46.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.013 for $1/2^*h-3/2^*k,-1/2^*h-1/2^*k,-1/2^*h+1/2^*k-l$ 0.011 for $1/2^*h+3/2^*k,1/2^*h-1/2^*k,-1/2^*h-1/2^*k-l$	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	28900	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	81.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.03% 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: UTP, MG, ZN

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	0/1841	0.45	0/2504
1	B	0.24	0/1821	0.47	0/2476
2	C	0.24	0/8929	0.44	0/12078
3	D	0.24	0/11955	0.45	0/16163
4	E	0.23	0/773	0.41	0/1042
5	F	0.23	0/2852	0.39	0/3837
6	G	0.44	0/393	0.90	0/606
7	H	0.42	0/556	0.95	0/858
8	I	0.25	0/72	0.71	0/110
All	All	0.25	0/29192	0.47	0/39674

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 [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	1809	0	1863	23	0
1	B	1789	0	1841	27	0
2	C	8762	0	8855	113	0
3	D	11746	0	11984	150	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	E	759	0	771	5	0
5	F	2807	0	2882	35	0
6	G	350	0	192	3	0
7	H	495	0	272	20	0
8	I	86	0	42	3	0
9	D	2	0	0	0	0
10	D	3	0	0	0	0
10	F	1	0	0	0	0
11	A	11	0	0	0	0
11	B	8	0	0	0	0
11	C	96	0	0	1	0
11	D	128	0	0	2	0
11	E	8	0	0	1	0
11	F	26	0	0	1	0
11	G	10	0	0	0	0
11	I	4	0	0	0	0
All	All	28900	0	28702	336	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 336 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:16:DC:H2''	7:H:17:DA:C8	1.73	1.23
2:C:7:GLY:CA	2:C:907:ASP:OD1	2.23	0.87
2:C:628:PHE:H	2:C:638:ASP:HB3	1.45	0.82
7:H:16:DC:H2''	7:H:17:DA:H8	1.38	0.78
7:H:19:DG:H1'	7:H:20:DG:H5'	1.70	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	229/315 (73%)	226 (99%)	3 (1%)	0	100	100
1	B	225/315 (71%)	221 (98%)	3 (1%)	1 (0%)	34	66
2	C	1108/1119 (99%)	1085 (98%)	23 (2%)	0	100	100
3	D	1483/1524 (97%)	1448 (98%)	35 (2%)	0	100	100
4	E	92/99 (93%)	91 (99%)	1 (1%)	0	100	100
5	F	344/443 (78%)	340 (99%)	4 (1%)	0	100	100
All	All	3481/3815 (91%)	3411 (98%)	69 (2%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	8	ALA

### 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	200/273 (73%)	197 (98%)	3 (2%)	65	87
1	B	200/273 (73%)	192 (96%)	8 (4%)	31	65
2	C	934/941 (99%)	914 (98%)	20 (2%)	53	81
3	D	1254/1279 (98%)	1231 (98%)	23 (2%)	59	85
4	E	82/88 (93%)	81 (99%)	1 (1%)	71	91
5	F	301/388 (78%)	295 (98%)	6 (2%)	55	82
All	All	2971/3242 (92%)	2910 (98%)	61 (2%)	53	81

5 of 61 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	670	GLN
3	D	81	THR

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
5	F	158	GLU
2	C	775	ARG
2	C	1078	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	724	GLN
3	D	1441	GLN
3	D	1124	GLN
2	C	390	GLN
3	D	1195	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	I	2/4 (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 ⓘ

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

## 5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

## 5.6 Ligand geometry ⓘ

Of 6 ligands modelled in this entry, 6 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, 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	231/315 (73%)	0.04	4 (1%) 70 69	56, 75, 103, 157	0
1	B	227/315 (72%)	0.25	9 (3%) 38 33	58, 85, 120, 151	0
2	C	1112/1119 (99%)	0.18	51 (4%) 32 29	38, 70, 137, 161	0
3	D	1486/1524 (97%)	0.33	85 (5%) 23 19	36, 70, 140, 167	0
4	E	94/99 (94%)	-0.01	0 100 100	47, 73, 114, 125	0
5	F	346/443 (78%)	0.42	27 (7%) 13 10	46, 88, 145, 172	0
6	G	17/21 (80%)	-0.03	0 100 100	49, 86, 175, 182	0
7	H	24/26 (92%)	-0.29	0 100 100	75, 114, 160, 180	0
8	I	3/4 (75%)	-0.62	0 100 100	51, 51, 55, 66	0
All	All	3540/3866 (91%)	0.25	176 (4%) 28 25	36, 75, 138, 182	0

The worst 5 of 176 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	F	149	GLU	8.7
5	F	422	LEU	7.2
5	F	146	GLY	6.7
3	D	173	PRO	6.4
3	D	1499	ARG	6.3

### 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, 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
10	MG	D	2005	1/1	0.68	0.13	74,74,74,74	0
10	MG	F	2001	1/1	0.75	0.08	71,71,71,71	0
10	MG	D	2004	1/1	0.91	0.52	72,72,72,72	0
10	MG	D	2003	1/1	0.98	0.15	41,41,41,41	0
9	ZN	D	2002	1/1	0.98	0.10	103,103,103,103	0
9	ZN	D	2001	1/1	1.00	0.19	53,53,53,53	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.