



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 10, 2023 – 03:31 AM EDT

PDB ID : 7MLB  
Title : Crystal structure of Thermus thermophilus transcription initiation complex with 5nt RNA  
Authors : Liu, Y.; Ebright, R.H.  
Deposited on : 2021-04-28  
Resolution : 3.60 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 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) : 1.13  
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

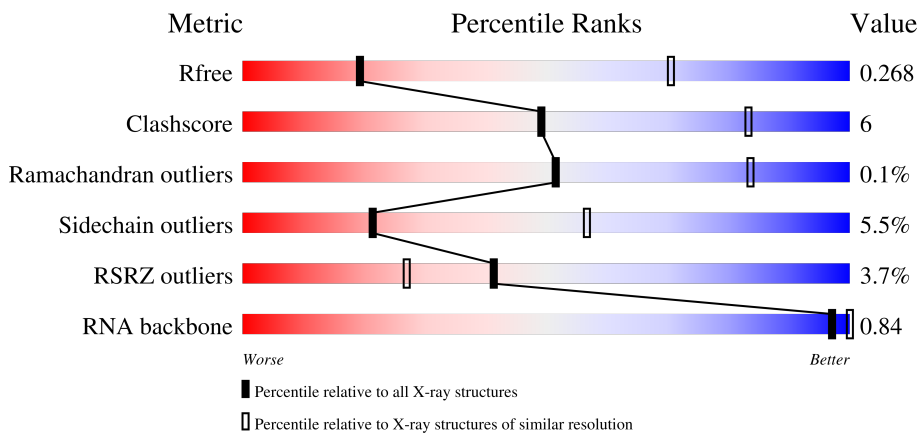
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.60 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	1257 (3.70-3.50)
Clashscore	141614	1353 (3.70-3.50)
Ramachandran outliers	138981	1307 (3.70-3.50)
Sidechain outliers	138945	1307 (3.70-3.50)
RSRZ outliers	127900	1161 (3.70-3.50)
RNA backbone	3102	1017 (4.20-3.00)

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  $\geq 3$ , 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	 57% 15% 28%
1	B	315	 59% 10% 30%
2	C	1119	 3% 81% 16%
3	D	1524	 5% 79% 17%

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
4	E	99	<p>%</p> <p>77% 18% 5%</p>
5	F	443	<p>4%</p> <p>64% 14% 22%</p>
6	G	21	<p>19% 52% 29%</p>
7	H	27	<p>26% 41% 33%</p>
8	I	5	<p>60% 40%</p>

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	2001	-	-	-	X

## 2 Entry composition [i](#)

There are 10 unique types of molecules in this entry. The entry contains 28400 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
			Total	C	N	O	S			
1	A	226	Total	C	N	O	S	0	0	0
			1782	1138	310	332	2			
1	B	222	Total	C	N	O	S	0	0	0
			1750	1118	304	326	2			

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

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

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

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

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

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

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
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(P\*CP\*TP\*GP\*CP\*AP\*TP\*CP\*CP\*CP\*TP\*GP\*AP\*GP\*TP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
6	G	15	305	145	53	92	15	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
7	H	18	376	179	76	104	17	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
8	I	5	104	47	18	34	5	0	0	0

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

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

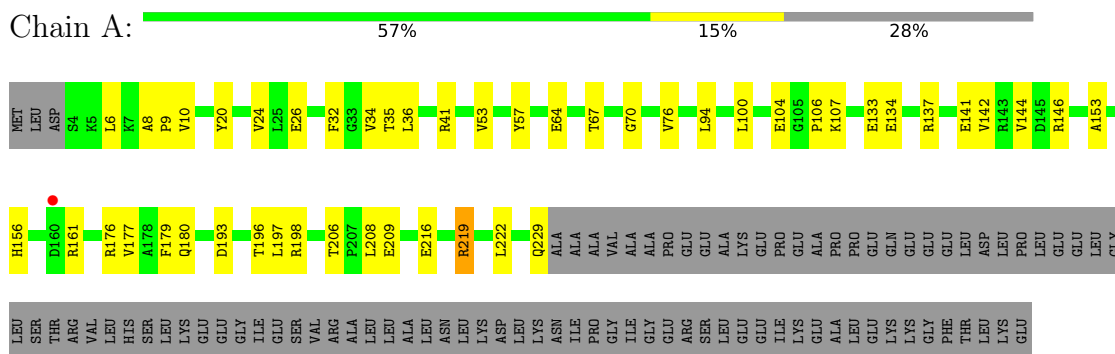
- Molecule 10 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

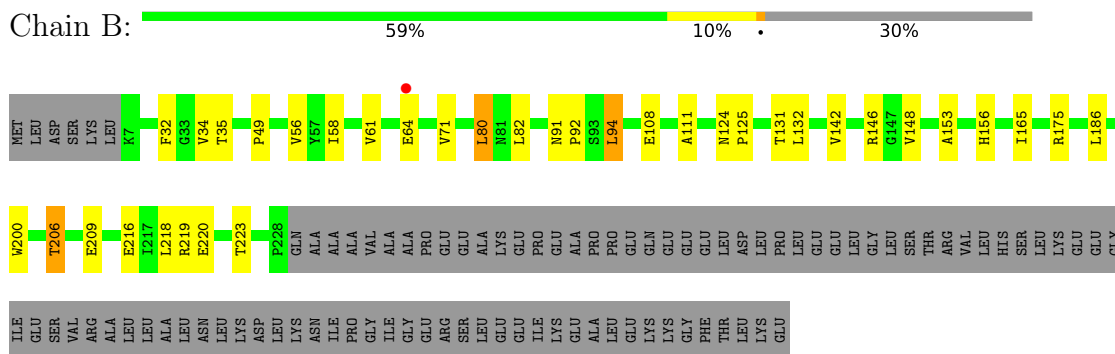
### 3 Residue-property plots [i](#)

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.

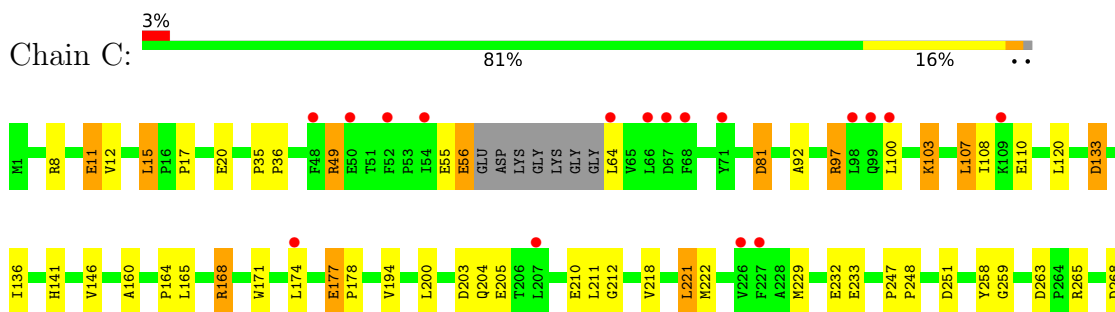
- Molecule 1: DNA-directed RNA polymerase subunit alpha

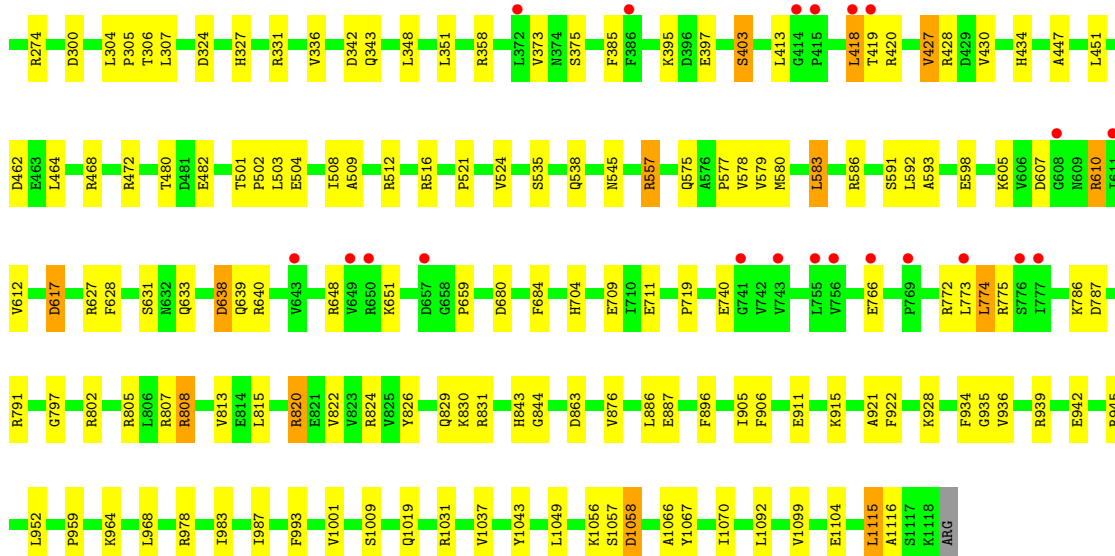


- Molecule 1: DNA-directed RNA polymerase subunit alpha

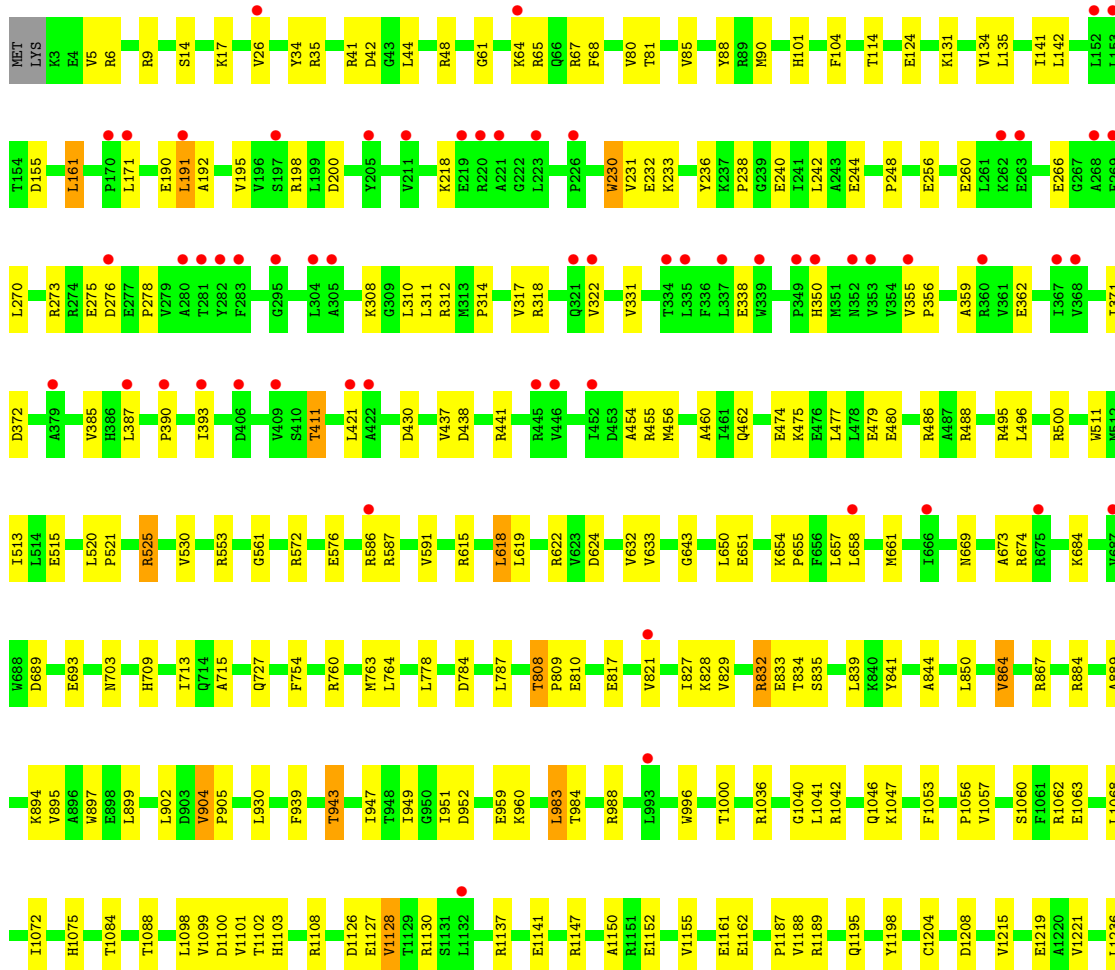
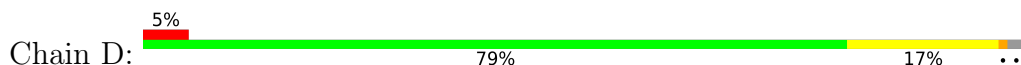


- Molecule 2: DNA-directed RNA polymerase subunit beta





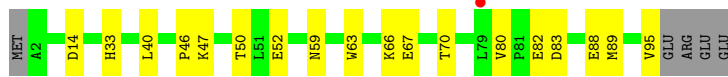
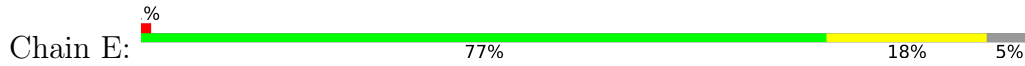
- Molecule 3: DNA-directed RNA polymerase subunit beta'



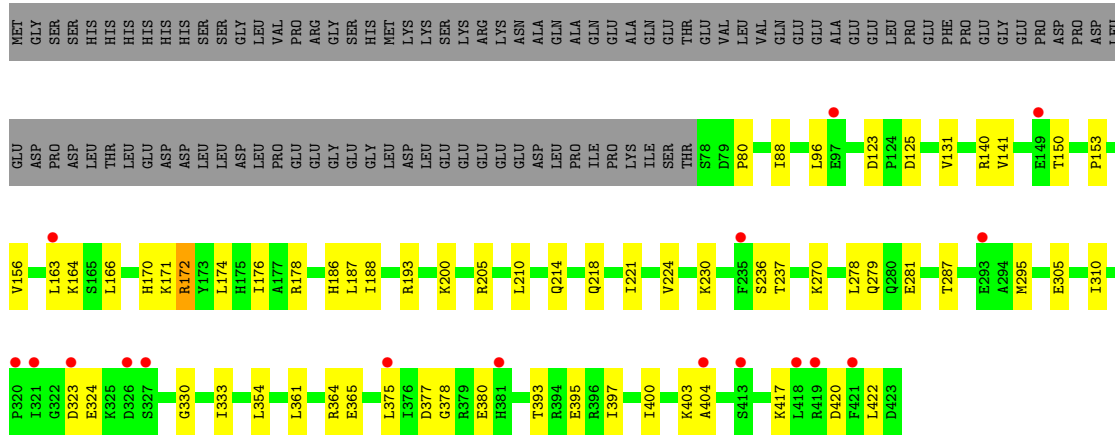




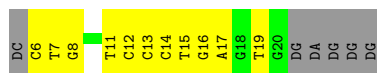
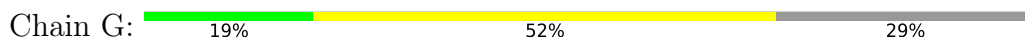
• Molecule 4: DNA-directed RNA polymerase subunit omega



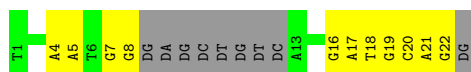
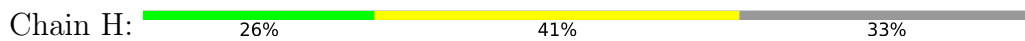
• Molecule 5: RNA polymerase sigma factor SigA



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



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



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

Chain I:  60% 40%



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	186.17Å 104.27Å 299.02Å 90.00° 97.95° 90.00°	Depositor
Resolution (Å)	44.62 – 3.60 49.36 – 3.60	Depositor EDS
% Data completeness (in resolution range)	98.3 (44.62-3.60) 98.5 (49.36-3.60)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	0.12	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.23 (at 3.57Å)	Xtrriage
Refinement program	PHENIX 1.19.2_4158, PHENIX 1.19.2_4158	Depositor
R, $R_{free}$	0.233 , 0.270 0.233 , 0.268	Depositor DCC
$R_{free}$ test set	2001 reflections (3.07%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	115.3	Xtrriage
Anisotropy	0.514	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 91.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.27$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	28400	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	144.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.44% 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: 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	0.25	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	0.54	0/340	0.92	0/522
7	H	0.50	0/423	0.87	0/651
8	I	0.31	0/115	0.92	0/176
All	All	0.26	0/28982	0.53	0/39357

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	1782	0	1834	27	0
1	B	1750	0	1797	19	0
2	C	8770	0	8874	113	0
3	D	11738	0	11971	146	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	E	761	0	778	14	0
5	F	2807	0	2882	36	0
6	G	305	0	170	11	0
7	H	376	0	205	14	0
8	I	104	0	54	7	0
9	B	1	0	0	0	0
9	D	3	0	0	0	0
9	F	1	0	0	0	0
10	D	2	0	0	0	0
All	All	28400	0	28565	338	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 338 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:G:15:DT:H2'	6:G:16:DG:C8	2.18	0.78
1:B:206:THR:HG22	1:B:209:GLU:H	1.50	0.76
2:C:1116:ALA:HB2	3:D:88:TYR:HB3	1.69	0.74
3:D:1491:THR:HG21	4:E:89:MET:HG2	1.73	0.71
2:C:628:PHE:H	2:C:638:ASP:HB2	1.55	0.70

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	224/315 (71%)	222 (99%)	2 (1%)	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%)	1088 (98%)	18 (2%)	1 (0%)	51	83
3	D	1482/1524 (97%)	1459 (98%)	21 (1%)	2 (0%)	51	83
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	3469/3815 (91%)	3415 (98%)	51 (2%)	3 (0%)	51	83

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	592	LEU
3	D	1440	PHE
3	D	530	VAL

### 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	A	199/273 (73%)	192 (96%)	7 (4%)	36	68
1	B	195/273 (71%)	190 (97%)	5 (3%)	46	74
2	C	936/941 (100%)	877 (94%)	59 (6%)	18	53
3	D	1253/1279 (98%)	1179 (94%)	74 (6%)	19	55
4	E	83/88 (94%)	82 (99%)	1 (1%)	71	87
5	F	301/388 (78%)	284 (94%)	17 (6%)	21	56
All	All	2967/3242 (92%)	2804 (94%)	163 (6%)	21	57

5 of 163 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	867	ARG
3	D	1496	GLU
3	D	983	LEU
3	D	1219	GLU

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type
5	F	150	THR

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

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

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	I	4/5 (80%)	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 [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, 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	226/315 (71%)	-0.09	1 (0%) 92 86	99, 144, 168, 177	0
1	B	222/315 (70%)	-0.18	1 (0%) 91 83	98, 143, 178, 192	0
2	C	1111/1119 (99%)	0.09	38 (3%) 45 30	87, 140, 186, 211	0
3	D	1486/1524 (97%)	0.17	71 (4%) 30 19	87, 136, 187, 216	1 (0%)
4	E	94/99 (94%)	-0.17	1 (1%) 80 68	97, 159, 197, 210	0
5	F	346/443 (78%)	0.17	17 (4%) 29 18	101, 155, 209, 223	0
6	G	15/21 (71%)	-0.22	0 100 100	116, 162, 249, 258	0
7	H	18/27 (66%)	-0.30	0 100 100	131, 162, 242, 256	0
8	I	5/5 (100%)	0.28	0 100 100	116, 117, 128, 147	0
All	All	3523/3868 (91%)	0.09	129 (3%) 41 27	87, 142, 189, 258	1 (0%)

The worst 5 of 129 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	322	VAL	6.7
2	C	415	PRO	6.0
3	D	335	LEU	5.8
3	D	262	LYS	5.8
3	D	337	LEU	5.6

### 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, 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(Å <sup>2</sup> )	Q<0.9
9	MG	B	2001	1/1	0.08	0.88	169,169,169,169	0
9	MG	D	2005	1/1	0.84	0.37	112,112,112,112	0
10	ZN	D	2002	1/1	0.90	0.08	215,215,215,215	0
9	MG	F	2001	1/1	0.91	0.06	132,132,132,132	0
9	MG	D	2003	1/1	0.93	0.33	86,86,86,86	0
9	MG	D	2004	1/1	0.94	0.55	86,86,86,86	0
10	ZN	D	2001	1/1	0.95	0.23	118,118,118,118	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.