



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:52 PM GMT

PDB ID : 4Q5S  
Title : Thermus thermophilus RNA polymerase initially transcribing complex containing 6-mer RNA  
Authors : Murakami, K.S.  
Deposited on : 2014-04-17  
Resolution : 3.00 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  
<http://wwpdb.org/validation/2016/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 : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

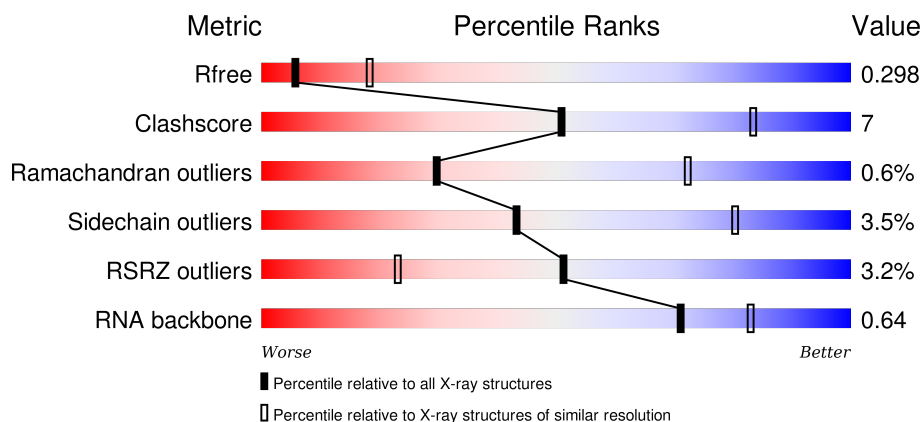
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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))
$R_{free}$	91344	1578 (3.00-3.00)
Clashscore	102246	1912 (3.00-3.00)
Ramachandran outliers	100387	1853 (3.00-3.00)
Sidechain outliers	100360	1856 (3.00-3.00)
RSRZ outliers	91569	1592 (3.00-3.00)
RNA backbone	2183	1036 (3.40-2.60)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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>3%</div> <div>59%12%28%</div> </div>
1	B	315	<div> <div>60%9%30%</div> </div>
2	C	1119	<div> <div>3%</div> <div>79%20%..</div> </div>
3	D	1524	<div> <div>2%</div> <div>78%19%..</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	99	<div><div></div><div>2%</div><div>81%</div><div>14%</div><div>5%</div></div>
5	F	423	<div><div></div><div>8%</div><div>62%</div><div>16%</div><div>20%</div></div>
6	G	22	<div><div></div><div>32%</div><div>27%</div><div>5%</div><div>36%</div></div>
7	H	27	<div><div></div><div>37%</div><div>22%</div><div>41%</div></div>
8	I	5	<div><div></div><div>60%</div><div>40%</div></div>

## 2 Entry composition

There are 11 unique types of molecules in this entry. The entry contains 28290 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	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
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
3	D	1484	Total	C	N	O	S	0	0	0
			11722	7430	2064	2192	36			

- 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
			761	486	132	139	4			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	338	Total	C	N	O	S	0	0	0
			2747	1736	500	507	4			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	14	Total	C	N	O	P	0	0	0
			289	137	52	86	14			

- 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\*GP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	H	16	Total	C	N	O	P	0	0	0
			333	159	66	93	15			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	I	5	Total	C	N	O	P	0	0	0
			102	46	16	35	5			

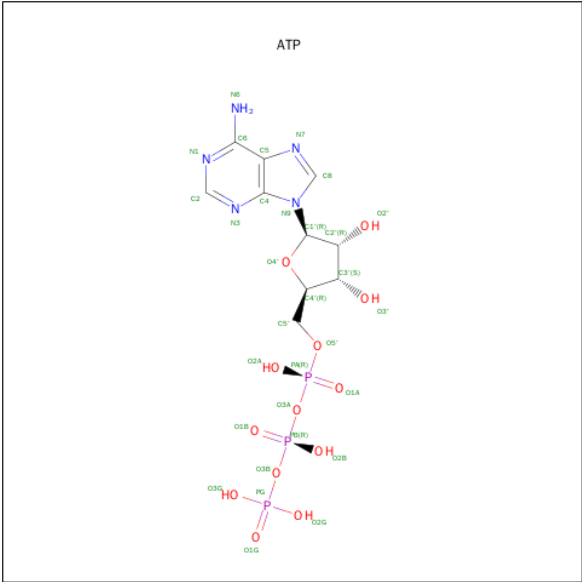
- Molecule 9 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

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

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

- Molecule 11 is ADENOSINE-5'-TRIPHOSPHATE (three-letter code: ATP) (formula: C<sub>10</sub>H<sub>16</sub>N<sub>5</sub>O<sub>13</sub>P<sub>3</sub>).

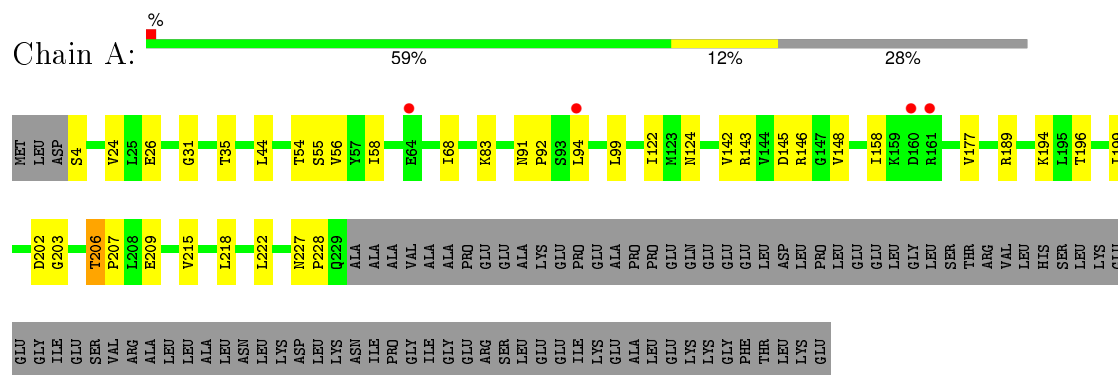


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
11	I	1	Total	C	N	O	P	0	0
			31	10	5	13	3		

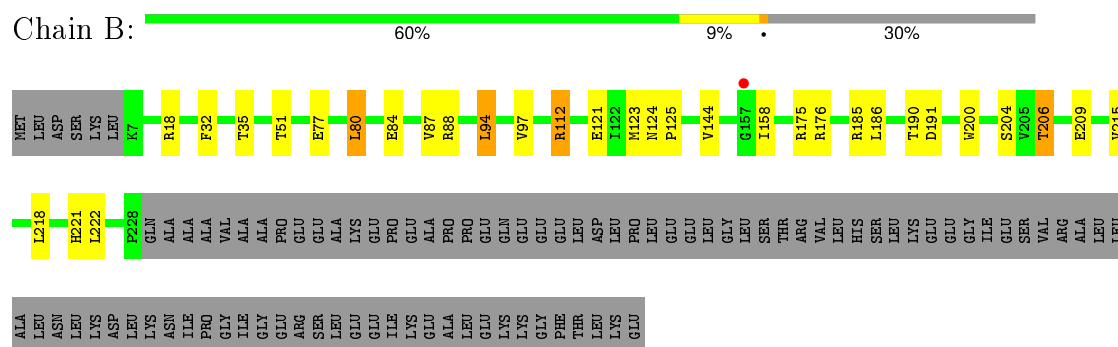
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

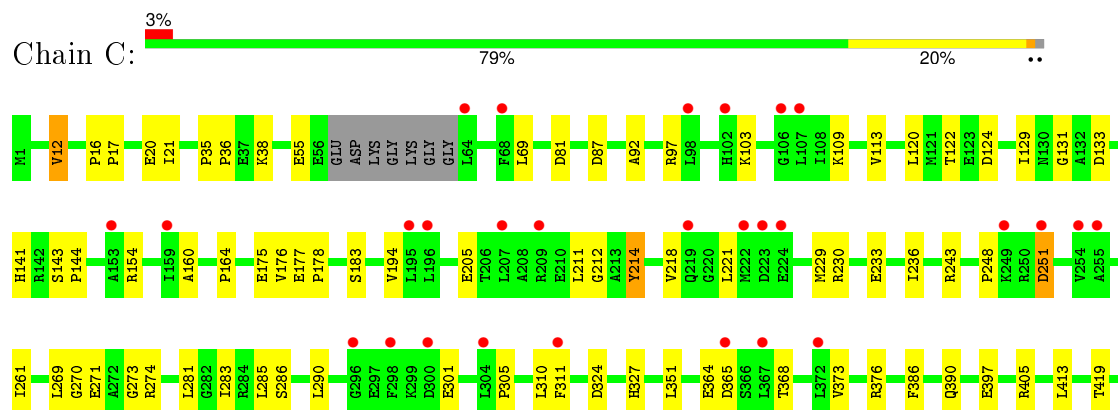
- Molecule 1: DNA-directed RNA polymerase subunit alpha



- Molecule 1: DNA-directed RNA polymerase subunit alpha

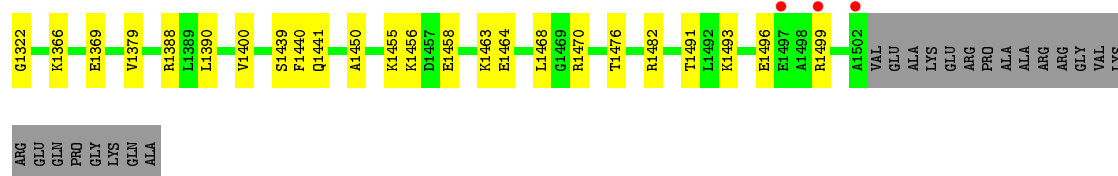


- Molecule 2: DNA-directed RNA polymerase subunit beta

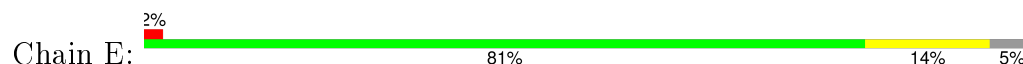




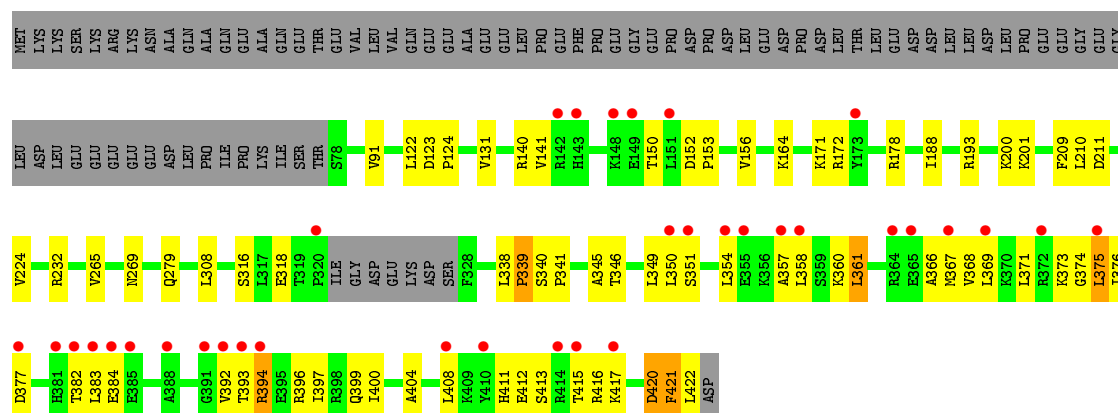




- Molecule 4: DNA-directed RNA polymerase subunit omega



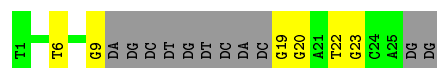
- Molecule 5: RNA polymerase sigma factor SigA



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



- 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\*GP\*G)-3')



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





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	187.09Å 102.08Å 297.26Å 90.00° 98.14° 90.00°	Depositor
Resolution (Å)	41.98 – 3.00 41.98 – 2.97	Depositor EDS
% Data completeness (in resolution range)	95.3 (41.98-3.00) 89.6 (41.98-2.97)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.83 (at 2.95Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.269 , 0.294 0.280 , 0.298	Depositor DCC
$R_{free}$ test set	1755 reflections (1.74%)	DCC
Wilson B-factor (Å <sup>2</sup> )	56.3	Xtriage
Anisotropy	0.493	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 29.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.42$ , $\langle L^2 \rangle = 0.25$	Xtriage
Outliers	0 of 107824 reflections	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	28290	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	108.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.36% 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, 0.375 respectively for untwinned datasets, and 0.333, 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, ATP

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	0.21	0/1814	0.40	0/2466
1	B	0.21	0/1782	0.40	0/2424
2	C	0.21	0/8937	0.39	0/12087
3	D	0.21	0/11928	0.39	0/16127
4	E	0.21	0/775	0.38	0/1045
5	F	0.23	1/2791 (0.0%)	0.38	0/3754
6	G	0.55	0/323	1.02	1/497 (0.2%)
7	H	0.46	0/374	1.03	0/575
8	I	1.07	0/112	1.03	0/171
All	All	0.23	1/28836 (0.0%)	0.43	1/39146 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	F	339	PRO	C-N	-5.72	1.20	1.34

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	G	14	DG	O3'-P-O5'	-5.61	93.35	104.00

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	24	0
1	B	1750	0	1797	20	0
2	C	8770	0	8874	149	0
3	D	11722	0	11949	160	0
4	E	761	0	778	9	0
5	F	2747	0	2831	79	0
6	G	289	0	159	9	0
7	H	333	0	183	6	0
8	I	102	0	55	3	0
9	D	1	0	0	0	0
10	D	2	0	0	0	0
11	I	31	0	11	2	0
All	All	28290	0	28471	390	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 7.

The worst 5 of 390 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)
2:C:776:SER:CB	5:F:373:LYS:NZ	2.02	1.23
2:C:769:PRO:HG2	5:F:374:GLY:O	1.40	1.19
2:C:776:SER:HB2	5:F:373:LYS:NZ	1.58	1.19
2:C:769:PRO:CG	5:F:374:GLY:O	1.93	1.17
3:D:233:LYS:NZ	3:D:240:GLU:OE2	1.88	1.04

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%)	209 (93%)	15 (7%)	0	100	100
1	B	220/315 (70%)	207 (94%)	12 (6%)	1 (0%)	34	76
2	C	1107/1119 (99%)	1029 (93%)	70 (6%)	8 (1%)	26	70
3	D	1480/1524 (97%)	1393 (94%)	77 (5%)	10 (1%)	26	70
4	E	92/99 (93%)	89 (97%)	3 (3%)	0	100	100
5	F	334/423 (79%)	316 (95%)	16 (5%)	2 (1%)	30	72
All	All	3457/3795 (91%)	3243 (94%)	193 (6%)	21 (1%)	30	72

5 of 21 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
5	F	421	PHE
2	C	12	VAL
2	C	211	LEU
3	D	448	GLU
3	D	484	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	199/273 (73%)	197 (99%)	2 (1%)	82	95
1	B	195/273 (71%)	188 (96%)	7 (4%)	42	79
2	C	936/941 (100%)	908 (97%)	28 (3%)	48	83
3	D	1251/1279 (98%)	1196 (96%)	55 (4%)	35	74
4	E	83/88 (94%)	82 (99%)	1 (1%)	78	94
5	F	294/371 (79%)	282 (96%)	12 (4%)	37	76
All	All	2958/3225 (92%)	2853 (96%)	105 (4%)	43	80

5 of 105 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	272	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	D	709	HIS
5	F	172	ARG
3	D	312	ARG
3	D	488	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	294	HIS
3	D	350	HIS
3	D	1124	GLN
2	C	1050	GLN
3	D	1359	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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 3 are monoatomic - leaving 1 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
11	ATP	I	101	-	24,33,33	1.67	1 (4%)	31,52,52	1.39	3 (9%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
11	ATP	I	101	-	-	0/18/38/38	0/3/3/3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
11	I	101	ATP	C5'-C4'	-7.66	1.26	1.51

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
11	I	101	ATP	PA-O3A-PB	-3.71	122.30	132.73
11	I	101	ATP	C5'-C4'-C3'	-3.58	100.98	115.21
11	I	101	ATP	PB-O3B-PG	-3.49	120.97	132.67

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
11	I	101	ATP	2	0

## 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, 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%)	-0.05	4 (1%) 71 43	76, 116, 149, 155	0
1	B	222/315 (70%)	-0.25	1 (0%) 91 76	63, 90, 119, 133	0
2	C	1111/1119 (99%)	0.07	34 (3%) 52 24	28, 107, 187, 229	0
3	D	1484/1524 (97%)	0.02	36 (2%) 62 32	30, 90, 161, 225	1 (0%)
4	E	94/99 (94%)	-0.04	2 (2%) 67 36	64, 100, 177, 190	0
5	F	338/423 (79%)	0.40	35 (10%) 8 3	94, 135, 226, 265	0
6	G	14/22 (63%)	0.02	0 100 100	78, 106, 188, 200	0
7	H	16/27 (59%)	-0.26	0 100 100	126, 135, 179, 183	0
8	I	5/5 (100%)	-0.63	0 100 100	60, 65, 75, 76	0
All	All	3510/3849 (91%)	0.05	112 (3%) 51 23	28, 104, 177, 265	1 (0%)

The worst 5 of 112 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	F	391	GLY	8.6
2	C	300	ASP	5.5
5	F	149	GLU	5.1
5	F	392	VAL	4.8
2	C	311	PHE	4.7

### 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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
10	ZN	D	2002	1/1	0.88	0.23	-0.22	238,238,238,238	0
11	ATP	I	101	31/31	0.82	0.24	-0.25	72,98,200,233	0
10	ZN	D	2003	1/1	0.98	0.15	-0.59	47,47,47,47	0
9	MG	D	2001	1/1	0.93	0.10	-1.87	24,24,24,24	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.