



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 05:36 am GMT

PDB ID : 4Q4Z
Title : Thermus thermophilus RNA polymerase de novo transcription initiation complex
Authors : Murakami, K.S.
Deposited on : 2014-04-15
Resolution : 2.90 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

<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 : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : trunk28620
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
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) : recalc28949

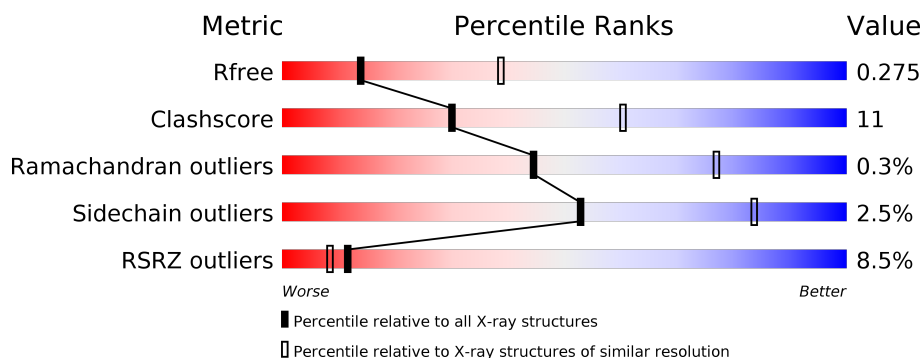
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.90 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}	100719	1586 (2.90-2.90)
Clashscore	112137	1807 (2.90-2.90)
Ramachandran outliers	110173	1768 (2.90-2.90)
Sidechain outliers	110143	1770 (2.90-2.90)
RSRZ outliers	101464	1596 (2.90-2.90)

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. 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>4%</div> <div>59%14%27%</div> </div>
1	B	315	<div> <div>4%</div> <div>60%12%28%</div> </div>
2	C	1119	<div> <div>6%</div> <div>79%19%..</div> </div>
3	D	1524	<div> <div>9%</div> <div>76%20%..</div> </div>
4	E	99	<div> <div>7%</div> <div>82%13%5%</div> </div>
5	F	423	<div> <div>14%</div> <div>66%15%18%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	G	22	
7	H	27	

2 Entry composition

There are 12 unique types of molecules in this entry. The entry contains 28755 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
			8774	5550	1565	1635	24			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1494	Total	C	N	O	S	0	1	0
			11808	7484	2083	2205	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	346	Total	C	N	O	S	0	0	0
			2807	1770	509	524	4			

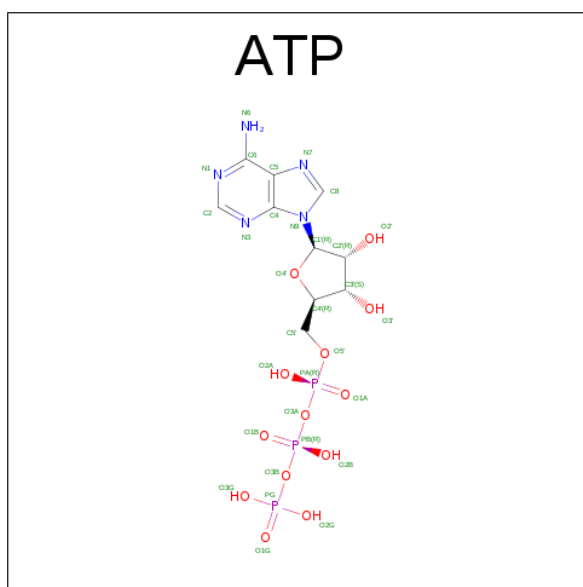
- Molecule 6 is a DNA chain called DNA (5'-D(*CP*CP*TP*GP*CP*AP*TP*CP*CP*GP*TP*GP*AP*GP*TP*GP*CP*AP*GP*CP*CP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	18	Total	C	N	O	P	0	0	0
			370	175	68	109	18			

- Molecule 7 is a DNA chain called DNA (25-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	H	25	Total	C	N	O	P	0	0	0
			516	246	99	147	24			

- Molecule 8 is ADENOSINE-5'-TRIPHOSPHATE (three-letter code: ATP) (formula: $C_{10}H_{16}N_5O_{13}P_3$).



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

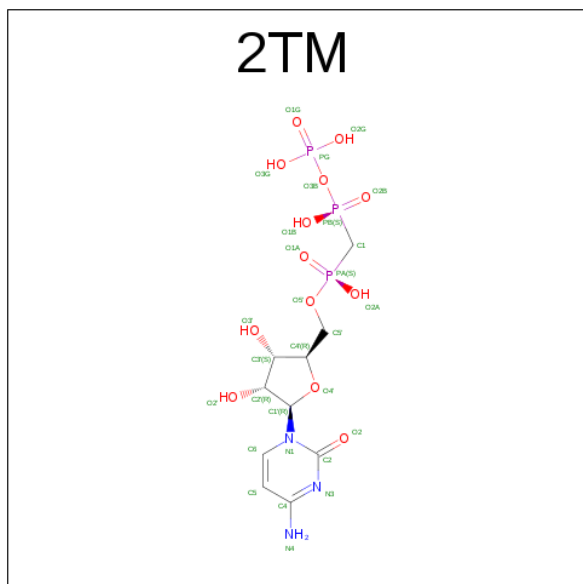
- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

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

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

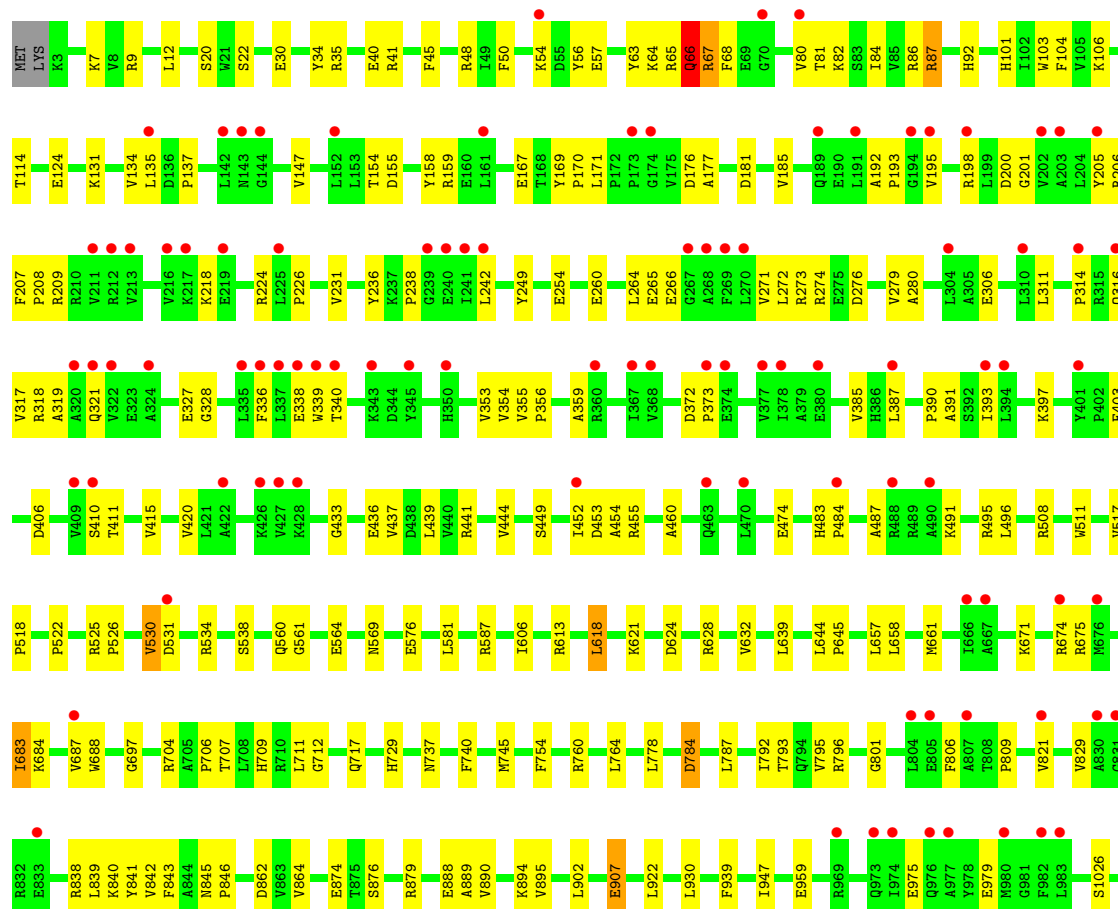
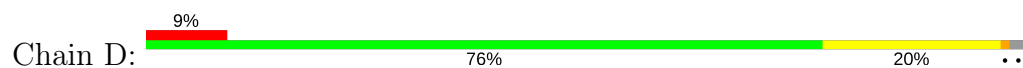
- Molecule 11 is 5'-O-[(S)-HYDROXY{[(S)-HYDROXY(PHOSPHONOOXY)PHOSPHORYL]METHYL}PHOSPHORYL]CYTIDINE (three-letter code: 2TM) (formula: C₁₀H₁₈N₃O₁₃P₃).

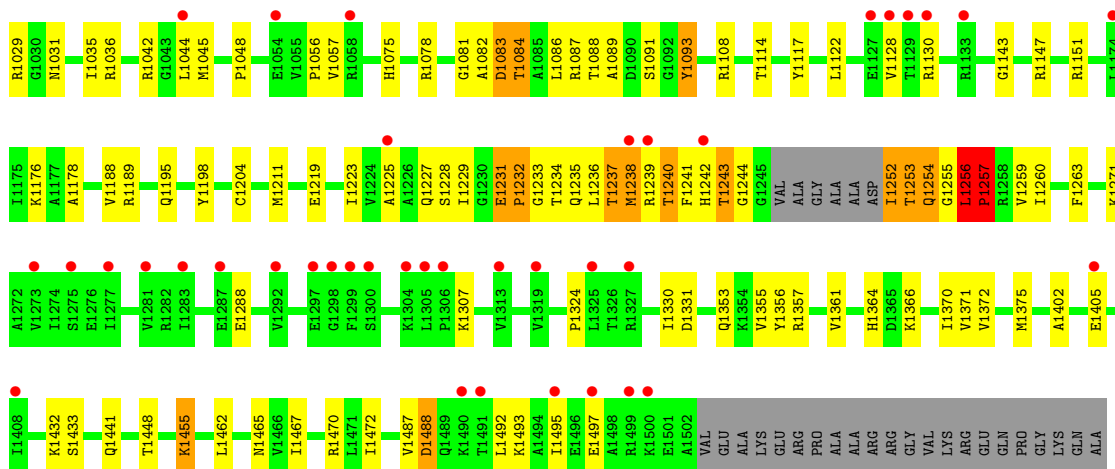


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
11	D	1	Total	C	N	O	P	0	0
			29	10	3	13	3		

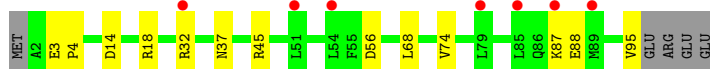
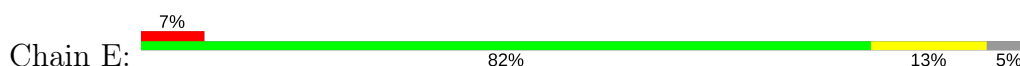
- Molecule 12 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
12	A	4	Total	O	0	0
			4	4		
12	B	1	Total	O	0	0
			1	1		
12	C	22	Total	O	0	0
			22	22		
12	D	26	Total	O	0	0
			26	26		
12	E	2	Total	O	0	0
			2	2		
12	G	1	Total	O	0	0
			1	1		

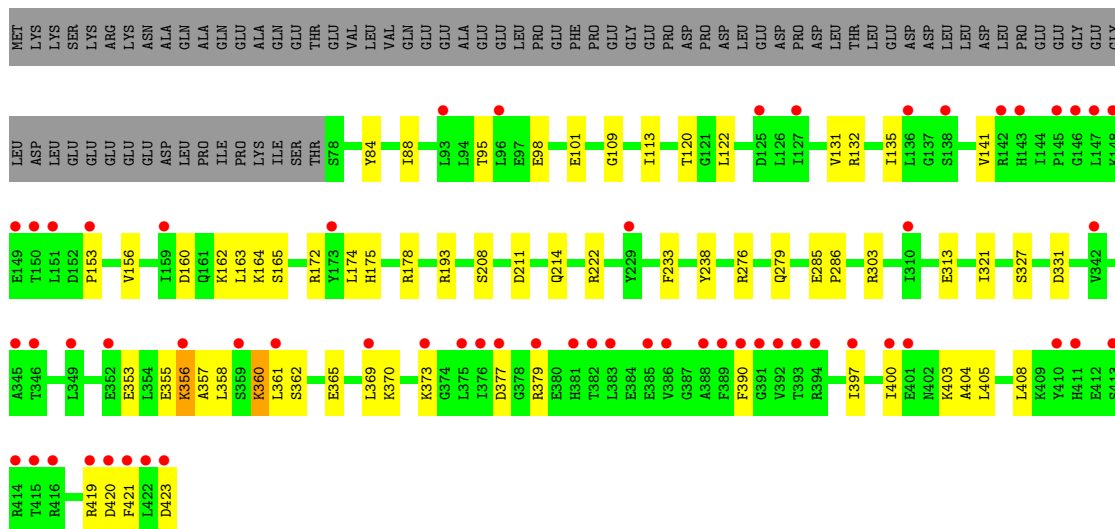




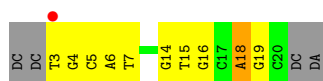
- Molecule 4: DNA-directed RNA polymerase subunit omega



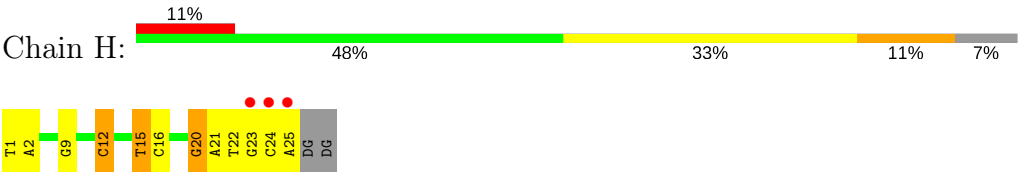
- Molecule 5: RNA polymerase sigma factor SigA



- 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 (25-MER)



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	184.49Å 102.16Å 294.72Å 90.00° 98.96° 90.00°	Depositor
Resolution (Å)	29.78 – 2.90 43.54 – 2.85	Depositor EDS
% Data completeness (in resolution range)	89.5 (29.78-2.90) 84.3 (43.54-2.85)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.90 (at 2.86Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.256 , 0.275 0.258 , 0.275	Depositor DCC
R_{free} test set	1689 reflections (1.57%)	DCC
Wilson B-factor (Å ²)	42.8	Xtriage
Anisotropy	0.728	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 18.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.41$, $\langle L^2 \rangle = 0.24$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.88	EDS
Total number of atoms	28755	wwPDB-VP
Average B, all atoms (Å ²)	70.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.77% 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 ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: 2TM, MG, ZN, 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	0.24	0/1841	0.46	0/2504
1	B	0.22	0/1821	0.44	0/2476
2	C	0.31	1/8941 (0.0%)	0.50	6/12092 (0.0%)
3	D	0.35	1/12019 (0.0%)	0.50	4/16248 (0.0%)
4	E	0.24	0/775	0.41	0/1045
5	F	0.23	0/2852	0.40	0/3837
6	G	0.72	2/414 (0.5%)	1.16	3/637 (0.5%)
7	H	0.57	0/580	1.13	3/895 (0.3%)
All	All	0.33	4/29243 (0.0%)	0.52	16/39734 (0.0%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	1093	TYR	CE2-CZ	-6.38	1.30	1.38
6	G	18	DA	O3'-P	-5.67	1.54	1.61
6	G	3	DT	O3'-P	-5.63	1.54	1.61
2	C	769	PRO	N-CD	5.43	1.55	1.47

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	761	PHE	O-C-N	-10.43	106.01	122.70
6	G	18	DA	O4'-C1'-N9	9.34	114.54	108.00
3	D	1257	PRO	CA-N-CD	-8.41	99.72	111.50
3	D	1232	PRO	CA-N-CD	-7.58	100.88	111.50
2	C	761	PHE	C-N-CA	7.11	139.47	121.70

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	1809	0	1863	38	0
1	B	1789	0	1841	24	0
2	C	8774	0	8877	226	3
3	D	11808	0	12041	258	4
4	E	761	0	778	9	1
5	F	2807	0	2882	101	2
6	G	370	0	202	19	0
7	H	516	0	283	31	1
8	C	31	0	11	2	0
9	D	2	0	0	0	0
10	D	3	0	0	0	0
11	D	29	0	14	2	0
12	A	4	0	0	9	0
12	B	1	0	0	0	0
12	C	22	0	0	29	0
12	D	26	0	0	26	0
12	E	2	0	0	1	0
12	G	1	0	0	0	0
All	All	28755	0	28792	605	6

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 11.

The worst 5 of 605 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:773:LEU:CD2	5:F:373:LYS:HG3	1.14	1.60
5:F:358:LEU:HD13	5:F:370:LYS:NZ	1.33	1.41
2:C:773:LEU:CD2	5:F:373:LYS:CG	2.03	1.36
2:C:778:PHE:HE2	5:F:419:ARG:NH2	1.25	1.35
2:C:764:GLU:O	2:C:766:GLU:N	1.62	1.32

The worst 5 of 6 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)
3:D:306:GLU:OE1	7:H:1:DT:O5'[4_745]	1.70	0.50
2:C:318:PRO:CB	4:E:87:LYS:CG[1_545]	1.77	0.43
3:D:34:TYR:OH	3:D:327:GLU:OE1[4_755]	1.89	0.31
3:D:328:GLY:O	5:F:303:ARG:NH2[4_745]	2.11	0.09
2:C:37:GLU:OE1	3:D:1151:ARG:NH1[3_545]	2.12	0.08

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%)	222 (99%)	3 (1%)	0	100	100
2	C	1108/1119 (99%)	1076 (97%)	26 (2%)	6 (0%)	32	68
3	D	1491/1524 (98%)	1454 (98%)	33 (2%)	4 (0%)	44	77
4	E	92/99 (93%)	89 (97%)	3 (3%)	0	100	100
5	F	344/423 (81%)	336 (98%)	7 (2%)	1 (0%)	44	77
All	All	3489/3795 (92%)	3403 (98%)	75 (2%)	11 (0%)	44	77

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	425	PHE
2	C	765	SER
2	C	769	PRO
3	D	1240	THR
2	C	779	GLY

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	200/273 (73%)	196 (98%)	4 (2%)	60	87
1	B	200/273 (73%)	196 (98%)	4 (2%)	60	87
2	C	936/941 (100%)	907 (97%)	29 (3%)	45	79
3	D	1260/1279 (98%)	1229 (98%)	31 (2%)	53	83
4	E	83/88 (94%)	83 (100%)	0	100	100
5	F	301/371 (81%)	296 (98%)	5 (2%)	66	89
All	All	2980/3225 (92%)	2907 (98%)	73 (2%)	53	84

5 of 73 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	778	PHE
3	D	134	VAL
3	D	1488	ASP
3	D	66	GLN
3	D	354	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	560	GLN
3	D	737	ASN
3	D	1441	GLN
3	D	669	ASN
3	D	717	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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 [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 7 ligands modelled in this entry, 5 are monoatomic - leaving 2 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	$\# Z > 2$
8	ATP	C	1201	10	27,33,33	0.98	1 (3%)	25,52,52	1.48	2 (8%)
11	2TM	D	2006	-	25,30,30	2.38	9 (36%)	27,47,47	1.94	7 (25%)

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
8	ATP	C	1201	10	-	0/18/38/38	0/3/3/3
11	2TM	D	2006	-	-	0/15/38/38	0/2/2/2

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
11	D	2006	2TM	C6-N1	-4.40	1.29	1.35
11	D	2006	2TM	PB-O1B	-3.95	1.46	1.56
11	D	2006	2TM	C2'-C1'	-3.63	1.47	1.53
11	D	2006	2TM	PA-O2A	-3.40	1.48	1.56
11	D	2006	2TM	C5-C4	-3.34	1.32	1.41

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	C	1201	ATP	N3-C2-N1	-5.59	123.99	128.86
11	D	2006	2TM	C6-N1-C2	-3.19	116.11	121.28

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
11	D	2006	2TM	O2B-PB-C1	2.34	114.76	108.97
11	D	2006	2TM	O2A-PA-C1	2.50	117.16	106.54
8	C	1201	ATP	N6-C6-N1	2.50	123.73	118.77

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
8	C	1201	ATP	2	0
11	D	2006	2TM	2	0

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	A	231/315 (73%)	0.44	12 (5%) 28 23	36, 68, 92, 101	0
1	B	227/315 (72%)	0.43	12 (5%) 27 23	41, 69, 100, 122	0
2	C	1112/1119 (99%)	0.48	71 (6%) 20 15	25, 58, 115, 135	0
3	D	1494/1524 (98%)	0.59	135 (9%) 10 7	19, 58, 122, 214	0
4	E	94/99 (94%)	0.53	7 (7%) 15 11	34, 73, 111, 114	0
5	F	346/423 (81%)	0.93	60 (17%) 2 1	43, 79, 152, 176	0
6	G	18/22 (81%)	0.18	1 (5%) 25 20	55, 81, 139, 140	0
7	H	25/27 (92%)	0.39	3 (12%) 5 3	70, 100, 149, 169	0
All	All	3547/3844 (92%)	0.57	301 (8%) 11 8	19, 64, 124, 214	0

The worst 5 of 301 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	173	PRO	8.7
2	C	311	PHE	7.6
2	C	300	ASP	7.4
3	D	211	VAL	7.0
2	C	207	LEU	6.9

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 [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, 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	LLDF	B-factors(Å ²)	Q<0.9
10	MG	D	2005	1/1	0.96	0.32	1.65	45,45,45,45	0
8	ATP	C	1201	31/31	0.88	0.19	-0.78	38,47,75,83	0
11	2TM	D	2006	29/29	0.84	0.19	-0.85	38,51,69,94	0
9	ZN	D	2001	1/1	0.97	0.17	-1.80	38,38,38,38	0
9	ZN	D	2002	1/1	0.97	0.02	-2.17	59,59,59,59	0
10	MG	D	2003	1/1	0.95	0.26	-	38,38,38,38	0
10	MG	D	2004	1/1	0.40	0.66	-	54,54,54,54	0

6.5 Other polymers [i](#)

There are no such residues in this entry.