



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 11, 2023 – 03:47 PM EDT

PDB ID : 4MQ9
Title : Crystal structure of Thermus thermophilus RNA polymerase holoenzyme in complex with GE23077
Authors : Ho, M.X.; Arnold, E.; Ebright, R.H.; Zhang, Y.; Tuske, S.
Deposited on : 2013-09-16
Resolution : 3.35 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

<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
Mogul : 1.8.5 (274361), CSD as541be (2020)
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

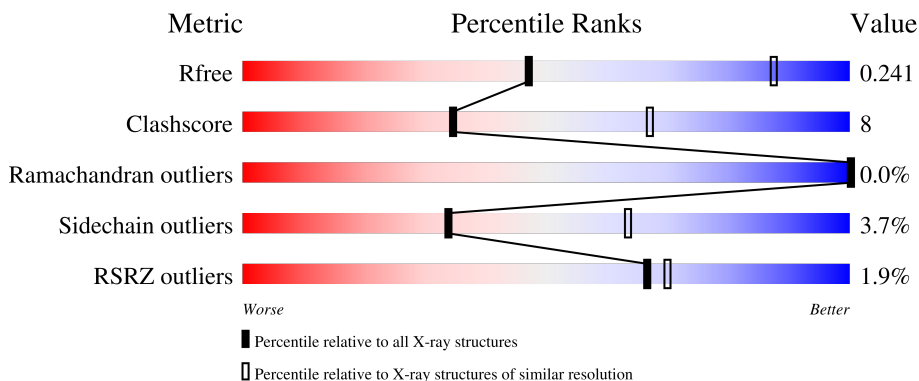
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.35 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.



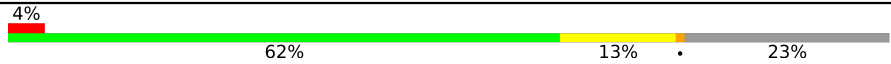
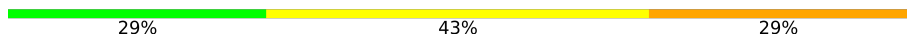
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1558 (3.42-3.30)
Clashscore	141614	1627 (3.42-3.30)
Ramachandran outliers	138981	1599 (3.42-3.30)
Sidechain outliers	138945	1598 (3.42-3.30)
RSRZ outliers	127900	1507 (3.42-3.30)

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 $\leq 5\%$. 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	314	
1	B	314	
2	C	1119	
3	D	1524	
4	E	99	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	F	443	
6	I	7	

2 Entry composition [i](#)

There are 10 unique types of molecules in this entry. The entry contains 26552 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
			1777	1135	309	331	2			
1	B	224	Total	C	N	O	S	0	0	0
			1750	1118	303	327	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	1100	Total	C	N	O	S	0	0	0
			8677	5487	1552	1614	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	1365	Total	C	N	O	S	0	0	0
			10781	6821	1912	2014	34			

- 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	93	Total	C	N	O	S	0	0	0
			754	481	131	138	4			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	F	339	Total	C	N	O	S	0	0	0
			2754	1736	501	513	4			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-19	MET	-	expression tag	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 protein called GE23077.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
6	I	7	Total	C	N	O	0	0	0
			50	26	9	15			

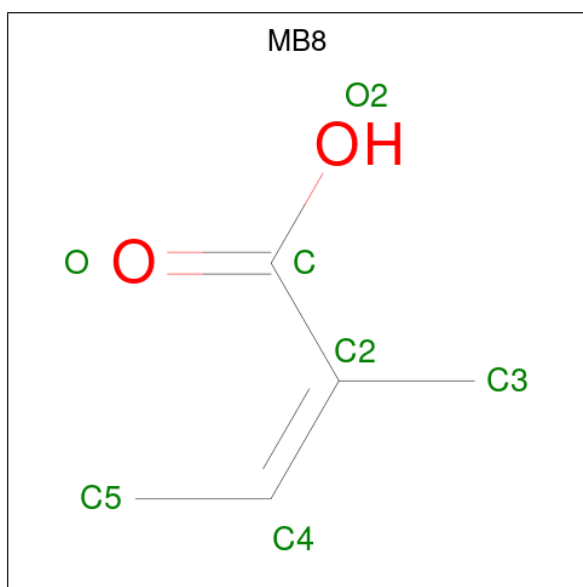
- Molecule 7 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

- Molecule 8 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

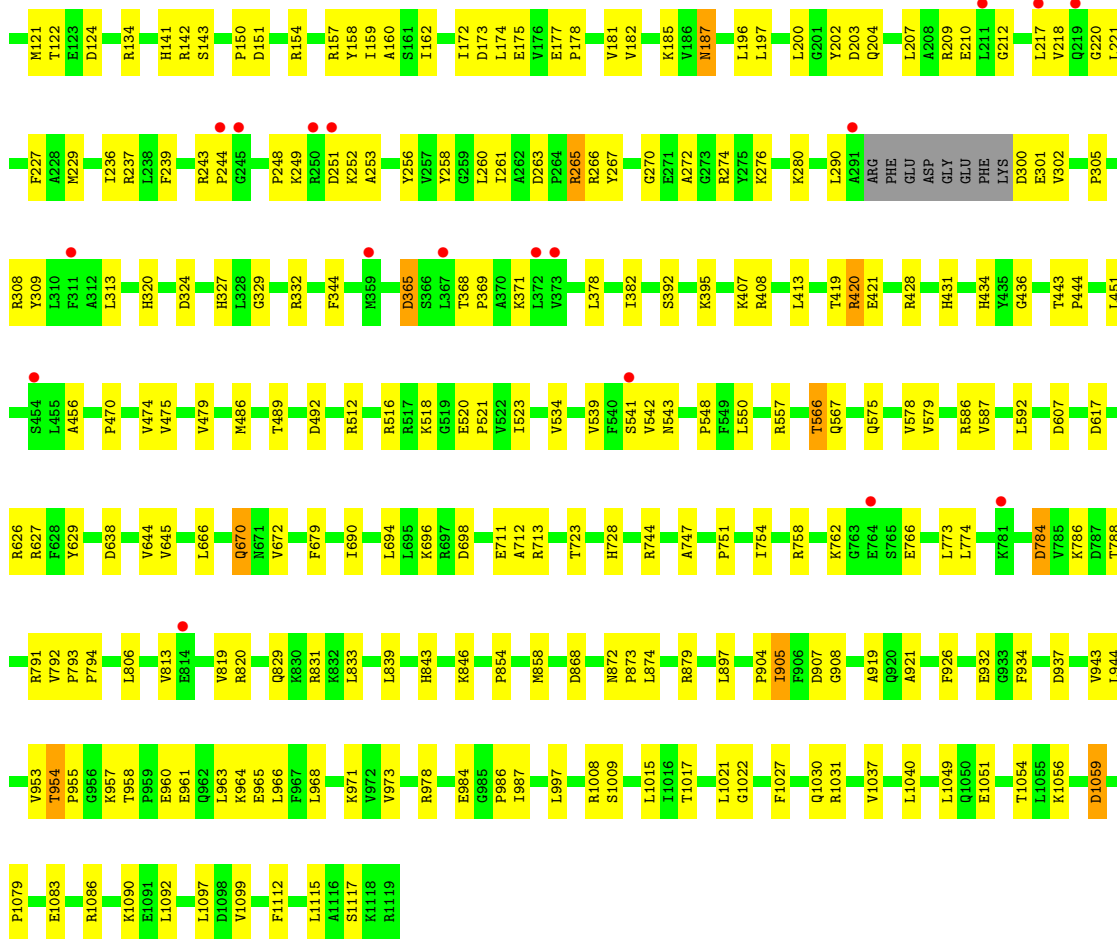
- Molecule 9 is (2Z)-2-methylbut-2-enoic acid (three-letter code: MB8) (formula: C₅H₈O₂).

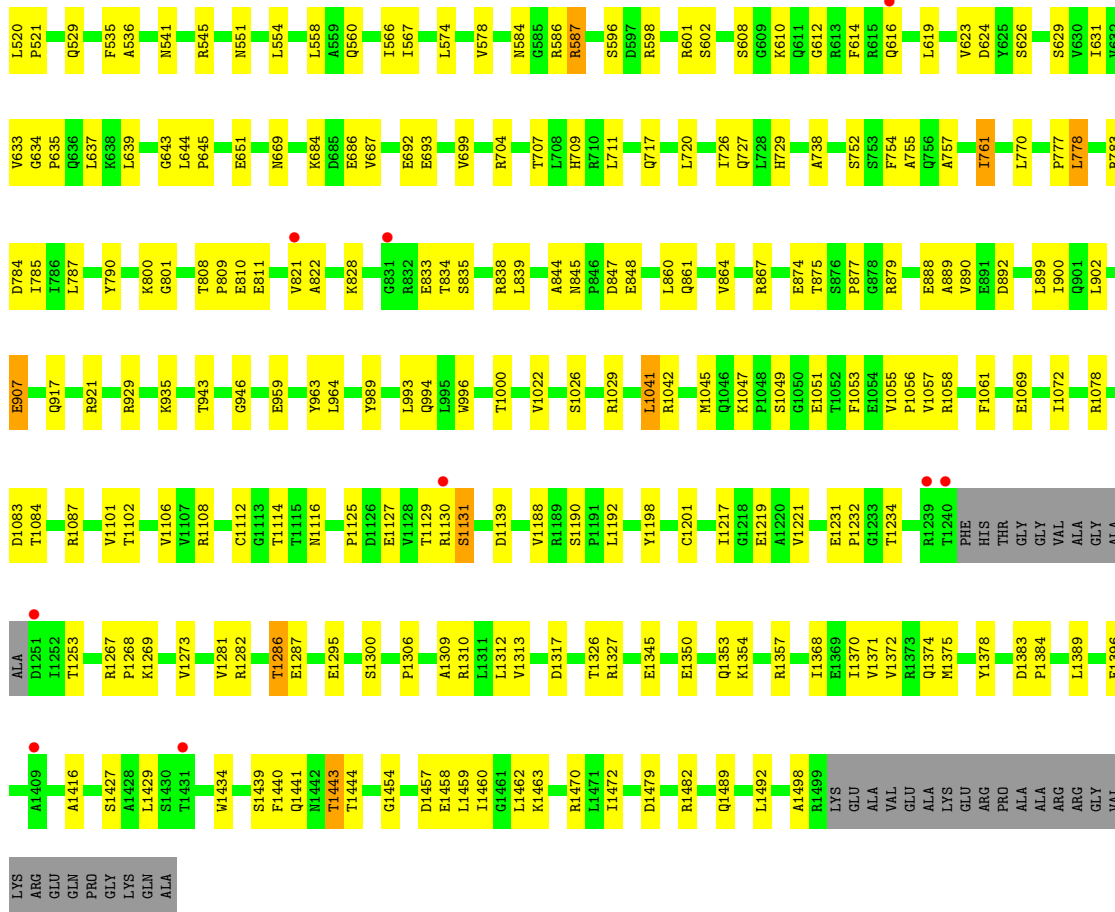


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
9	I	1	Total	C	O	0	0
			2	1	1		

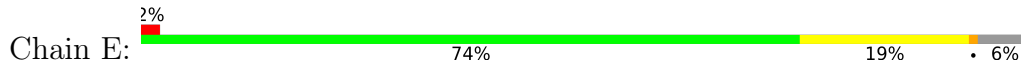
- Molecule 10 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	D	4	Total	O	0	0
			4	4		

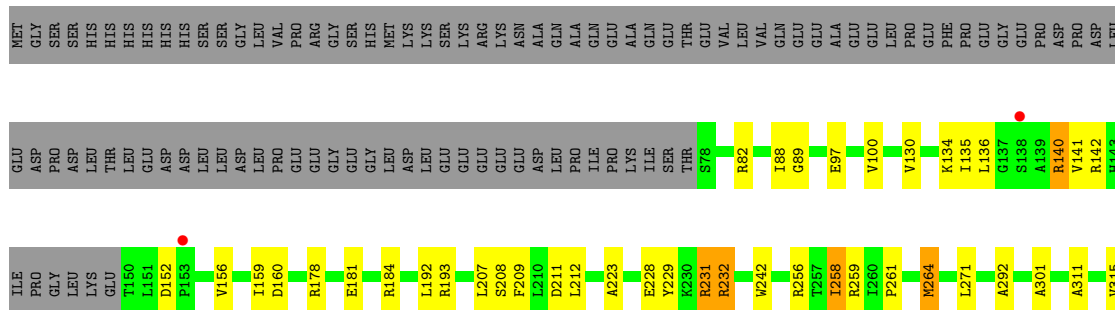


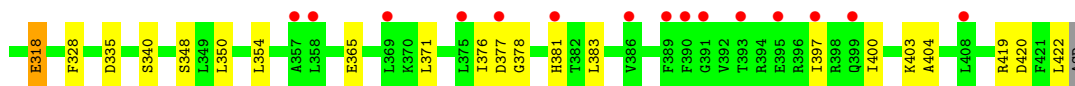


• Molecule 4: DNA-directed RNA polymerase subunit omega

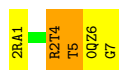


• Molecule 5: RNA polymerase sigma factor





- Molecule 6: GE23077



4 Data and refinement statistics i

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, α , β , γ	236.57Å 236.57Å 252.31Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	38.97 – 3.35 38.97 – 3.35	Depositor EDS
% Data completeness (in resolution range)	98.9 (38.97-3.35) 98.9 (38.97-3.35)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.58 (at 3.32Å)	Xtrriage
Refinement program	PHENIX (PHENIX.REFINE: 1.7.3_928)	Depositor
R, R_{free}	0.214 , 0.242 0.215 , 0.241	Depositor DCC
R_{free} test set	2266 reflections (2.00%)	wwPDB-VP
Wilson B-factor (Å ²)	126.0	Xtrriage
Anisotropy	0.350	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 83.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	0.032 for h,-h-k,-l	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	26552	wwPDB-VP
Average B, all atoms (Å ²)	140.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.79% 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 2TL, MB8, DSN, ZN, FGL, DVA, MG, R2T, 2RA, 0QZ

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	0/1809	0.47	0/2461
1	B	0.23	0/1781	0.44	0/2426
2	C	0.25	0/8841	0.46	0/11956
3	D	0.25	0/10966	0.47	0/14820
4	E	0.23	0/768	0.41	0/1035
5	F	0.23	0/2797	0.41	0/3761
All	All	0.25	0/26962	0.46	0/36459

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
6	I	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
6	I	5	2TL	Peptide

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	1777	0	1826	32	0
1	B	1750	0	1794	35	0
2	C	8677	0	8791	165	0
3	D	10781	0	10998	206	0
4	E	754	0	769	15	0
5	F	2754	0	2826	42	0
6	I	50	0	37	2	0
7	D	2	0	0	0	0
8	D	1	0	0	0	0
9	I	2	0	0	0	0
10	D	4	0	0	0	0
All	All	26552	0	27041	443	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 8.

The worst 5 of 443 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)
1:B:104:GLU:HB2	1:B:137:ARG:HA	1.60	0.82
2:C:102:HIS:HB3	2:C:105:THR:HB	1.65	0.77
3:D:508:ARG:HD3	3:D:509:PRO:HD2	1.66	0.77
2:C:904:PRO:HB2	2:C:907:ASP:HB3	1.69	0.73
3:D:95:LEU:HD11	3:D:517:VAL:HG23	1.70	0.73

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	224/314 (71%)	223 (100%)	1 (0%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	220/314 (70%)	217 (99%)	3 (1%)	0	100	100
2	C	1094/1119 (98%)	1059 (97%)	35 (3%)	0	100	100
3	D	1359/1524 (89%)	1317 (97%)	41 (3%)	1 (0%)	51	82
4	E	91/99 (92%)	90 (99%)	1 (1%)	0	100	100
5	F	335/443 (76%)	328 (98%)	7 (2%)	0	100	100
All	All	3323/3813 (87%)	3234 (97%)	88 (3%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	406	ASP

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	197/272 (72%)	188 (95%)	9 (5%)	27	59
1	B	193/272 (71%)	187 (97%)	6 (3%)	40	69
2	C	926/941 (98%)	895 (97%)	31 (3%)	38	67
3	D	1155/1279 (90%)	1115 (96%)	40 (4%)	36	66
4	E	82/88 (93%)	80 (98%)	2 (2%)	49	74
5	F	295/388 (76%)	279 (95%)	16 (5%)	22	54
All	All	2848/3240 (88%)	2744 (96%)	104 (4%)	34	63

5 of 104 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	388	HIS
3	D	783	ARG
5	F	271	LEU
3	D	407	VAL
3	D	502	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	709	HIS
3	D	1441	GLN
5	F	217	ASN
3	D	1442	ASN
2	C	829	GLN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

7 non-standard protein/DNA/RNA residues are modelled in this entry.

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	$\# Z > 2$
6	2TL	I	5	6	5,6,7	1.16	1 (20%)	6,7,9	1.00	0
6	2RA	I	1	9,6	3,5,6	0.57	0	1,5,7	0.26	0
6	R2T	I	4	6	8,10,11	1.86	2 (25%)	6,13,15	0.82	0
6	0QZ	I	6	6	4,5,6	1.59	1 (25%)	2,5,7	0.92	0
6	FGL	I	7	6	5,6,7	1.42	1 (20%)	1,7,9	1.84	0

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
6	2TL	I	5	6	-	1/5/6/8	-
6	2RA	I	1	9,6	-	0/2/4/6	-
6	R2T	I	4	6	-	7/13/14/16	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
6	0QZ	I	6	6	-	2/3/4/6	-
6	FGL	I	7	6	-	3/4/6/8	-

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	I	4	R2T	CD-NE2	4.24	1.43	1.32
6	I	6	0QZ	OB-CA	-3.02	1.38	1.43
6	I	4	R2T	OB1-CB	-2.22	1.37	1.43
6	I	5	2TL	OG1-CB	-2.07	1.37	1.43
6	I	7	FGL	OG2-CB	2.07	1.28	1.22

There are no bond angle outliers.

There are no chirality outliers.

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	I	4	R2T	CA-CB-CG-CD
6	I	4	R2T	CA-CB-CG-OG1
6	I	4	R2T	OB1-CB-CG-CD
6	I	4	R2T	OB1-CB-CG-OG1
6	I	4	R2T	OE1-CD-CG-OG1

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	I	1	2RA	1	0
6	I	4	R2T	1	0

5.5 Carbohydrates [i](#)

There are no monosaccharides 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	$\# Z > 2$
9	MB8	I	101	6	0,1,6	-	-	-		

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, 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	226/314 (71%)	-0.24	3 (1%) 77 80	84, 110, 169, 198	0
1	B	224/314 (71%)	0.13	4 (1%) 68 71	97, 165, 233, 259	0
2	C	1100/1119 (98%)	0.01	24 (2%) 62 65	67, 135, 240, 281	0
3	D	1365/1524 (89%)	-0.07	13 (0%) 82 86	68, 114, 213, 249	0
4	E	93/99 (93%)	-0.08	2 (2%) 62 65	96, 142, 212, 216	0
5	F	339/443 (76%)	0.06	17 (5%) 28 31	89, 153, 249, 267	0
6	I	0/7	-	-	-	-
All	All	3347/3820 (87%)	-0.03	63 (1%) 66 70	67, 129, 230, 281	0

The worst 5 of 63 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	372	LEU	4.6
2	C	217	LEU	4.4
2	C	98	LEU	4.4
2	C	250	ARG	4.2
2	C	99	GLN	4.1

6.2 Non-standard residues in protein, DNA, RNA chains [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(Å ²)	Q<0.9
6	DVA	I	3	7/8	0.95	0.26	81,85,91,93	0
6	FGL	I	7	7/8	0.96	0.15	82,93,106,108	0
6	2RA	I	1	6/7	0.96	0.17	85,96,99,111	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
6	DSN	I	2	6/7	0.97	0.11	80,85,93,97	0
6	R2T	I	4	11/12	0.98	0.17	77,87,94,97	0
6	2TL	I	5	7/8	0.98	0.20	78,84,91,93	0
6	0QZ	I	6	6/7	0.99	0.19	75,79,86,87	0

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, 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(\AA^2)	Q<0.9
9	MB8	I	101	2/7	0.82	0.36	102,102,102,107	0
7	ZN	D	1601	1/1	0.97	0.07	159,159,159,159	0
8	MG	D	1603	1/1	0.99	0.11	75,75,75,75	0
7	ZN	D	1602	1/1	0.99	0.19	105,105,105,105	0

6.5 Other polymers [i](#)

There are no such residues in this entry.