wwPDB X-ray Structure Validation Summary Report (i)

Nov 16, 2023 – 03:35 AM JST

PDB ID	:	6KQM
Title	:	Thermus thermophilus initial transcription complex comprising sigma A and 5'-triphosphate RNA of 5 nt
Authors	:	Zhang, Y.; Li, L.; Ebright, R.H.
Deposited on	:	2019-08-18
Resolution	:	3.20 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 (i) 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 (i)) were used in the production of this report:

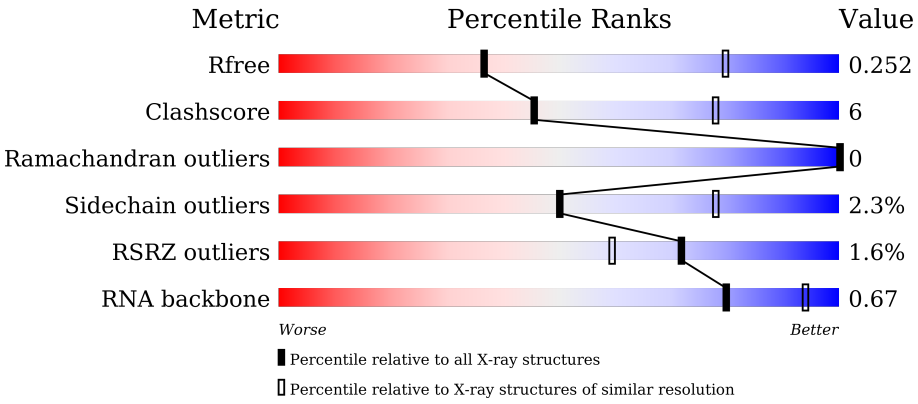
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.20 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	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)
RNA backbone	3102	1010 (3.50-2.90)

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 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 <=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% 16% 27%
1	B	315	59% 11% 29%
2	C	1119	81% 17%
3	D	1524	81% 16%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	99	
5	F	443	
6	G	21	
7	H	27	
8	I	5	

2 Entry composition (i)

There are 11 unique types of molecules in this entry. The entry contains 28789 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	1	0
			1814	1158	316	338	2			
1	B	223	Total	C	N	O	S	0	0	0
			1758	1124	305	327	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	3	0
			8792	5562	1570	1636	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	3	0
			11753	7455	2067	2195	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			

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*TP*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*GP*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(*(CTP))-R(P*UP*CP*GP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	I	5	Total	C	N	O	P	0	0	0
			106	47	18	36	5			

- 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	2	Total Mg 2 2	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

- Molecule 11 is water.

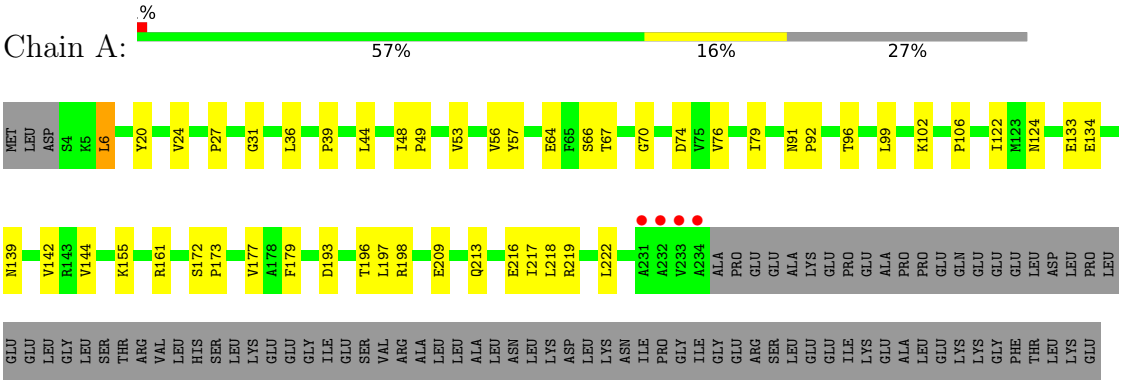
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
11	A	12	Total O 12 12	0	0
11	B	6	Total O 6 6	0	0
11	C	44	Total O 44 44	0	0
11	D	61	Total O 61 61	0	0
11	E	2	Total O 2 2	0	0
11	F	6	Total O 6 6	0	0
11	G	9	Total O 9 9	0	0
11	H	2	Total O 2 2	0	0
11	I	5	Total O 5 5	0	0

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.

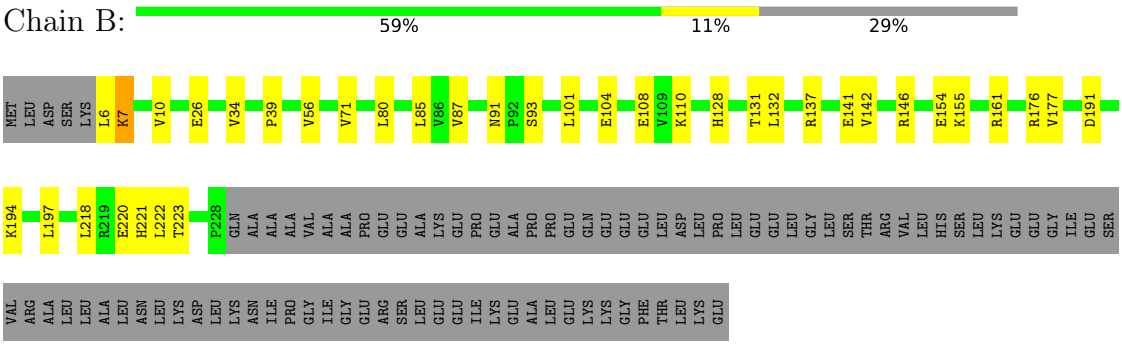
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain A:



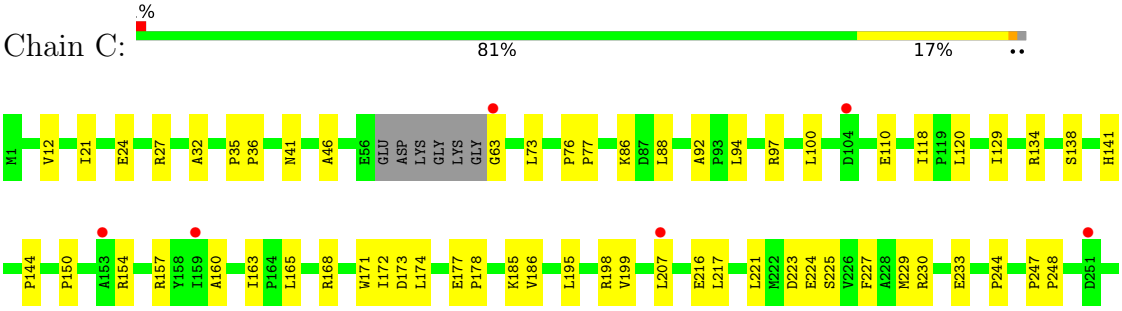
• Molecule 1: DNA-directed RNA polymerase subunit alpha

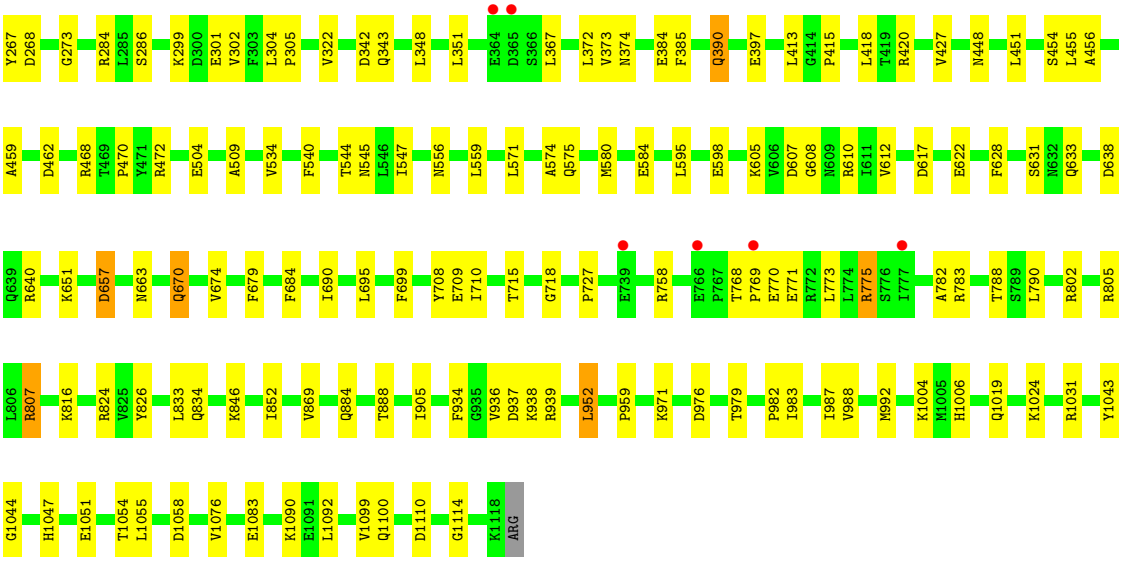
Chain B:



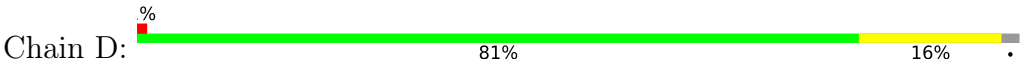
• Molecule 2: DNA-directed RNA polymerase subunit beta

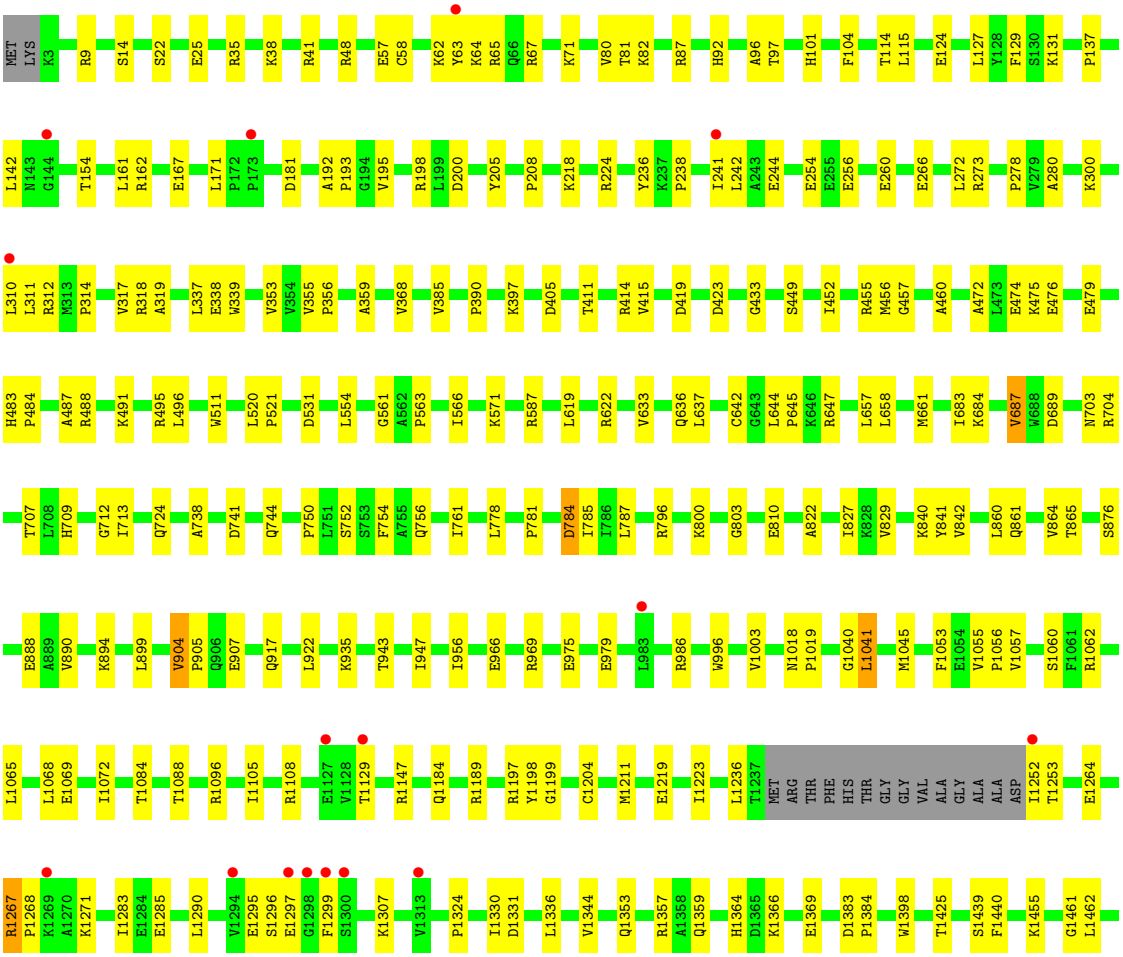
Chain C:

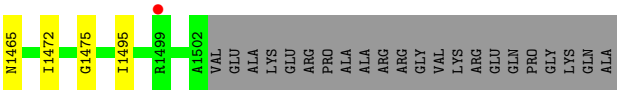




• Molecule 3: DNA-directed RNA polymerase subunit beta'

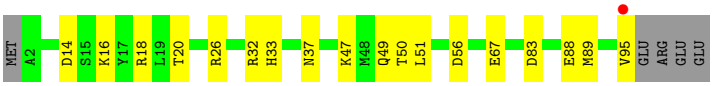






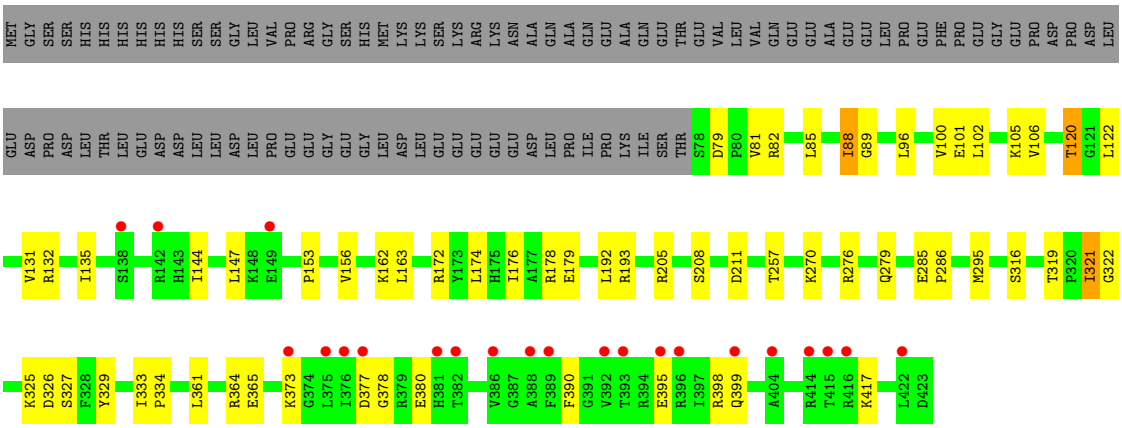
• Molecule 4: DNA-directed RNA polymerase subunit omega

Chain E:



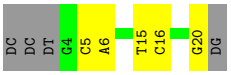
• Molecule 5: RNA polymerase sigma factor SigA

Chain F:



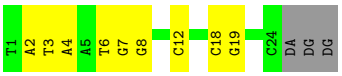
• Molecule 6: DNA (5'-D(*CP*CP*T*GP*CP*AP*TP*CP*CP*GP*TP*GP*AP*GP*TP*CP*GP*AP*GP*GP*G)-3')

Chain G:



• 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')

Chain H:



• Molecule 8: RNA (5'-R(*(CTP))-R(P*UP*CP*GP*A)-3')

Chain I:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	184.73Å 102.75Å 295.71Å 90.00° 98.86° 90.00°	Depositor
Resolution (Å)	48.89 – 3.20 48.89 – 3.20	Depositor EDS
% Data completeness (in resolution range)	95.1 (48.89-3.20) 95.1 (48.89-3.20)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.09 (at 3.19Å)	Xtriage
Refinement program	PHENIX 1.12_2829	Depositor
R, R_{free}	0.205 , 0.252 0.205 , 0.252	Depositor DCC
R_{free} test set	4310 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	50.0	Xtriage
Anisotropy	0.064	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 36.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	28789	wwPDB-VP
Average B, all atoms (Å^2)	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.19% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: CTP, ZN, MG

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	0/1849	0.44	0/2515
1	B	0.25	0/1790	0.47	0/2435
2	C	0.24	0/8969	0.45	0/12129
3	D	0.26	0/11969	0.45	0/16182
4	E	0.23	0/775	0.43	0/1045
5	F	0.24	0/2852	0.41	0/3837
6	G	0.48	0/393	0.93	0/606
7	H	0.40	0/556	0.91	0/858
8	I	0.28	0/94	0.75	0/144
All	All	0.26	0/29247	0.48	0/39751

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

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	1814	0	1869	32	0
1	B	1758	0	1808	25	0
2	C	8792	0	8902	125	0
3	D	11753	0	11992	149	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	16	0
5	F	2807	0	2882	43	0
6	G	350	0	192	4	0
7	H	495	0	272	8	0
8	I	106	0	54	1	0
9	B	1	0	0	0	0
9	D	2	0	0	0	0
9	F	1	0	0	0	0
10	D	2	0	0	0	0
11	A	12	0	0	0	0
11	B	6	0	0	0	0
11	C	44	0	0	1	0
11	D	61	0	0	0	0
11	E	2	0	0	0	0
11	F	6	0	0	1	0
11	G	9	0	0	0	0
11	H	2	0	0	0	0
11	I	5	0	0	0	0
All	All	28789	0	28749	363	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 363 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)
4:E:33:HIS:HA	4:E:95:VAL:CG2	1.78	1.13
4:E:33:HIS:HA	4:E:95:VAL:HG21	1.48	0.92
2:C:768:THR:CG2	2:C:769:PRO:HD2	2.06	0.85
2:C:628:PHE:H	2:C:638:ASP:HB3	1.41	0.85
4:E:33:HIS:HA	4:E:95:VAL:HG22	1.54	0.85

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	230/315 (73%)	225 (98%)	5 (2%)	0	100	100
1	B	221/315 (70%)	217 (98%)	4 (2%)	0	100	100
2	C	1111/1119 (99%)	1078 (97%)	33 (3%)	0	100	100
3	D	1485/1524 (97%)	1447 (97%)	38 (3%)	0	100	100
4	E	92/99 (93%)	89 (97%)	3 (3%)	0	100	100
5	F	344/443 (78%)	338 (98%)	6 (2%)	0	100	100
All	All	3483/3815 (91%)	3394 (97%)	89 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	201/273 (74%)	197 (98%)	4 (2%)	55	80
1	B	196/273 (72%)	190 (97%)	6 (3%)	40	72
2	C	939/941 (100%)	918 (98%)	21 (2%)	52	79
3	D	1255/1279 (98%)	1229 (98%)	26 (2%)	53	79
4	E	83/88 (94%)	81 (98%)	2 (2%)	49	77
5	F	301/388 (78%)	291 (97%)	10 (3%)	38	71
All	All	2975/3242 (92%)	2906 (98%)	69 (2%)	50	78

5 of 69 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	E	49	GLN
5	F	88	ILE
5	F	326	ASP
2	C	670	GLN
2	C	657	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	1124	GLN
3	D	1184	GLN
5	F	218	GLN
3	D	1195	GLN
3	D	611	GLN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	I	3/5 (60%)	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 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 (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, 95^{th} 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%)	-0.19	4 (1%) 70 57	26, 50, 76, 114	0
1	B	223/315 (70%)	-0.25	0 100 100	24, 53, 89, 105	0
2	C	1112/1119 (99%)	-0.28	12 (1%) 80 69	7, 42, 100, 146	0
3	D	1486/1524 (97%)	-0.22	17 (1%) 80 69	5, 39, 102, 136	1 (0%)
4	E	94/99 (94%)	-0.36	1 (1%) 80 69	16, 44, 84, 93	0
5	F	346/443 (78%)	0.05	22 (6%) 19 11	20, 59, 129, 145	0
6	G	17/21 (80%)	-0.47	0 100 100	9, 29, 134, 134	0
7	H	24/27 (88%)	-0.47	0 100 100	43, 64, 106, 139	0
8	I	4/5 (80%)	-0.62	0 100 100	10, 13, 18, 23	0
All	All	3537/3868 (91%)	-0.22	56 (1%) 72 59	5, 46, 105, 146	1 (0%)

The worst 5 of 56 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	63	GLY	4.4
5	F	375	LEU	4.2
1	A	232	ALA	4.0
5	F	381	HIS	3.9
5	F	414	ARG	3.9

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 monosaccharides in this entry.

6.4 Ligands

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^{th} 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
9	MG	F	2001	1/1	0.88	0.11	48,48,48,48	0
9	MG	B	2001	1/1	0.93	0.47	54,54,54,54	0
10	ZN	D	2002	1/1	0.96	0.06	103,103,103,103	0
9	MG	D	2004	1/1	0.97	0.24	28,28,28,28	0
10	ZN	D	2001	1/1	0.99	0.13	17,17,17,17	0
9	MG	D	2003	1/1	0.99	0.21	8,8,8,8	0

6.5 Other polymers

There are no such residues in this entry.