



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 7, 2026 – 02:15 AM UTC

PDB ID : 5E17 / pdb_00005e17
Title : T. thermophilus transcription initiation complex having a RRR discriminator sequence and a nontemplate-strand length corresponding to TSS selection at position 7 (RPO-GGG-7)
Authors : Zhang, Y.; Ebright, R.H.
Deposited on : 2015-09-29
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 ⓘ 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	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

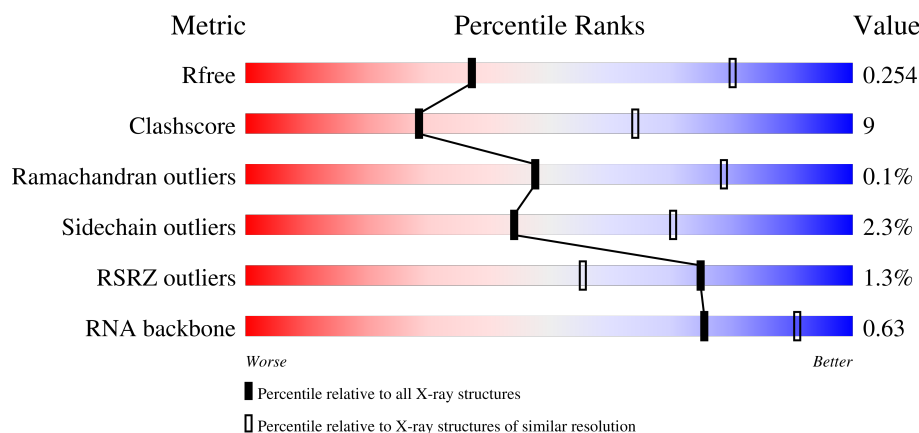
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.20 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}	180053	1466 (3.20-3.20)
Clashscore	190562	1573 (3.20-3.20)
Ramachandran outliers	187476	1548 (3.20-3.20)
Sidechain outliers	187428	1547 (3.20-3.20)
RSRZ outliers	180081	1466 (3.20-3.20)
RNA backbone	3983	1222 (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 $\leq 5\%$. 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	56% 17% 27%
1	B	315	50% 20% 30%
2	C	1119	75% 23% ..
3	D	1524	76% 21% ..

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	99	% 79% 16% 5%
5	F	443	2% 62% 14% 24%
6	G	21	48% 38% 14%
7	H	27	7% 41% 48% 11%
8	I	7	14% 71% 14% 14%

2 Entry composition

There are 11 unique types of molecules in this entry. The entry contains 28784 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	222	Total	C	N	O	S	0	0	0
			1750	1120	303	325	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
			11759	7458	2070	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
			758	483	132	139	4			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	337	Total	C	N	O	S	0	0	0
			2737	1726	499	508	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*T
P*GP*AP*GP*TP*CP*GP*AP*GP*GP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	18	Total	C	N	O	P	0	0	0
			372	176	73	106	17			

- Molecule 7 is a DNA chain called DNA (27-MER).

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(*CP*CP*CP*UP*CP*GP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	I	7	Total	C	N	O	P	0	0	0
			142	65	24	47	6			

- Molecule 9 is MAGNESIUM ION (CCD ID: 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 (CCD ID: ZN) (formula: Zn).

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

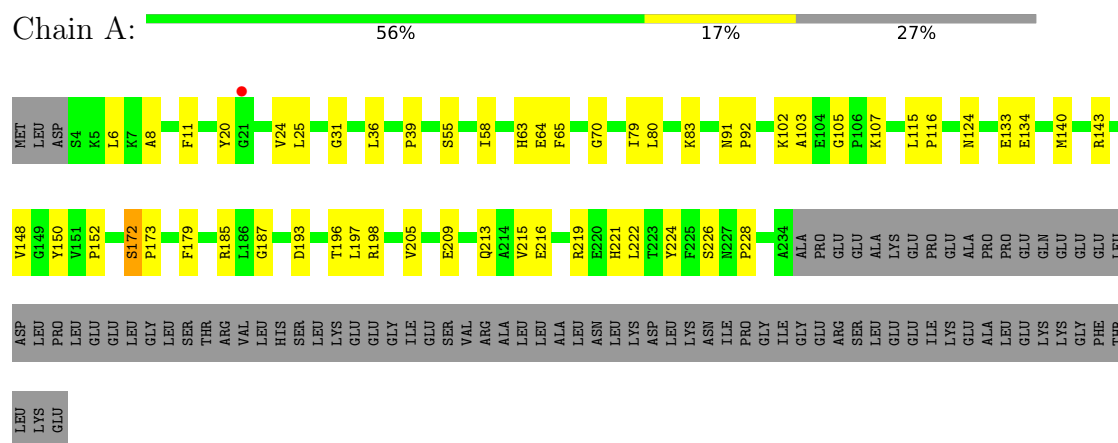
- Molecule 11 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
11	A	12	Total O 12 12	0	0
11	B	12	Total O 12 12	0	0
11	C	49	Total O 49 49	0	0
11	D	62	Total O 62 62	0	0
11	E	5	Total O 5 5	0	0
11	F	8	Total O 8 8	0	0
11	G	5	Total O 5 5	0	0
11	H	4	Total O 4 4	0	0
11	I	2	Total O 2 2	0	0

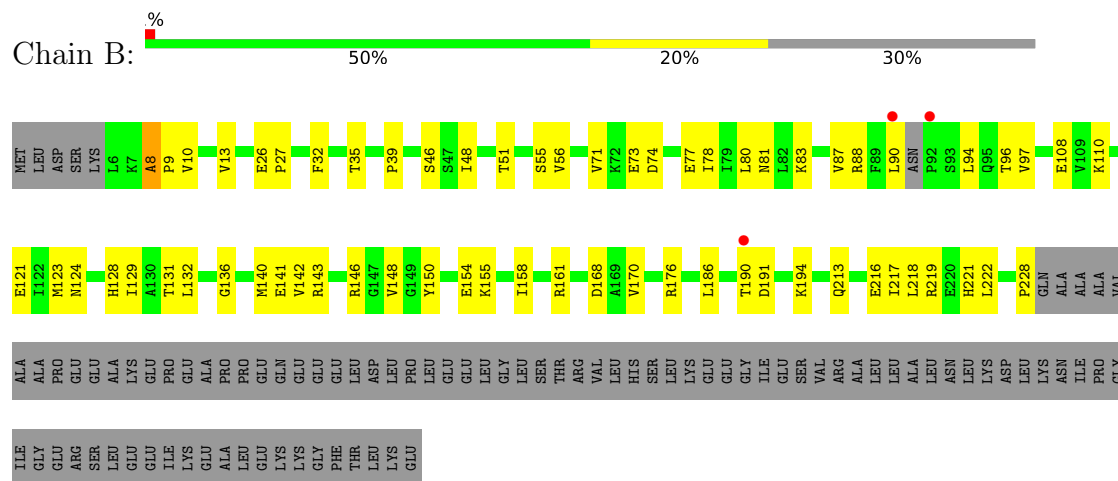
3 Residue-property plots [i](#)

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.

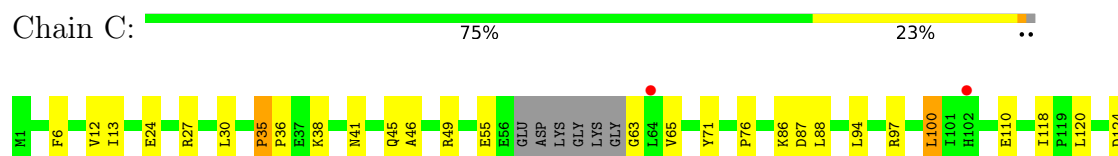
• Molecule 1: DNA-directed RNA polymerase subunit alpha

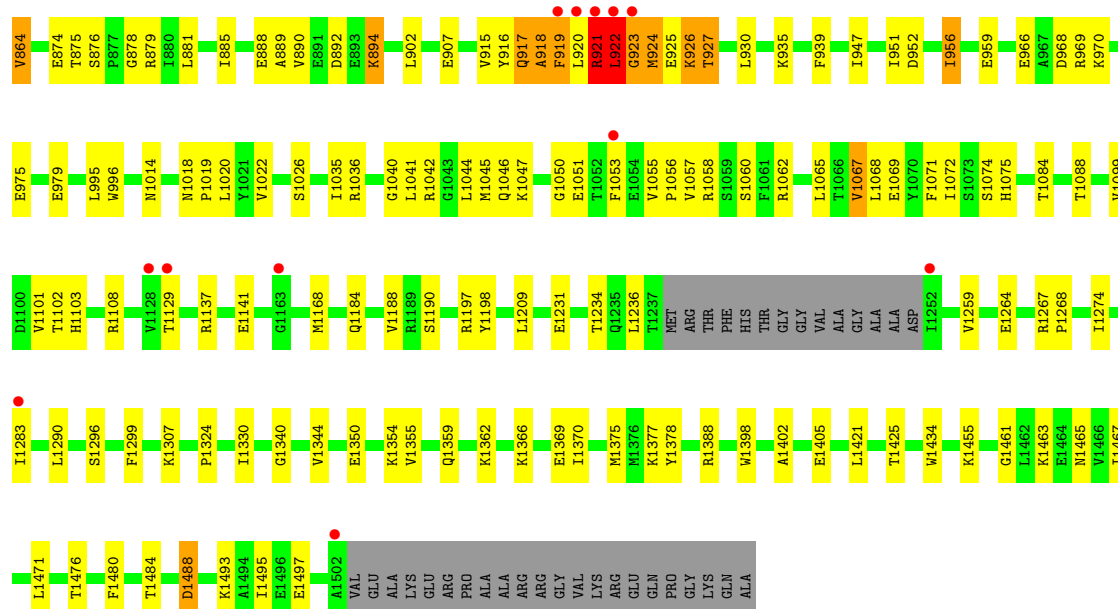


• Molecule 1: DNA-directed RNA polymerase subunit alpha

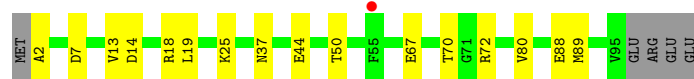
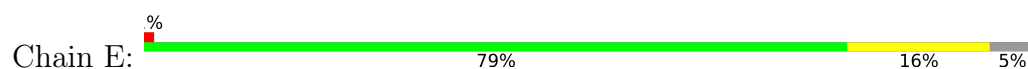


• Molecule 2: DNA-directed RNA polymerase subunit beta

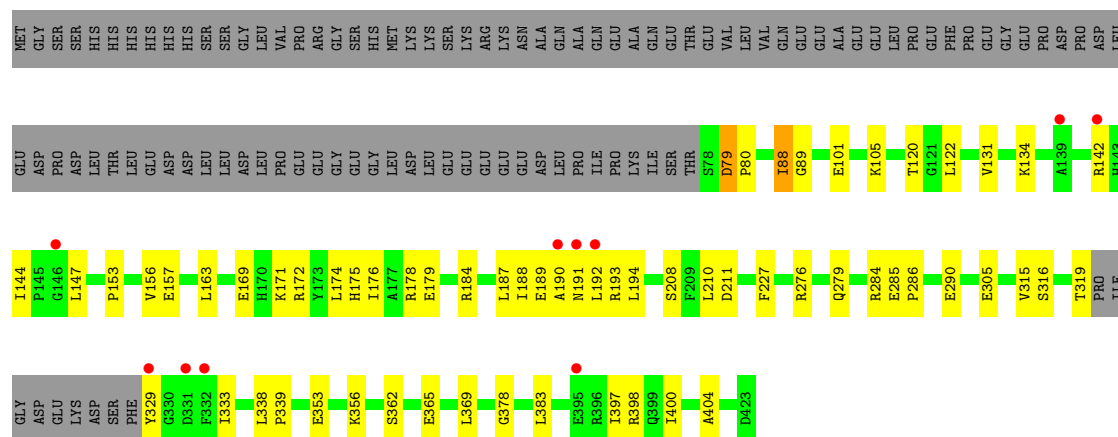




- Molecule 4: DNA-directed RNA polymerase subunit omega



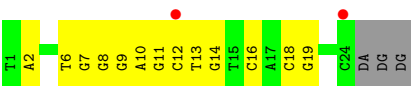
- Molecule 5: RNA polymerase sigma factor SigA



- 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')



● Molecule 7: DNA (27-MER)



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



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	183.01Å 103.57Å 294.94Å 90.00° 99.20° 90.00°	Depositor
Resolution (Å)	39.82 – 3.20 39.82 – 3.20	Depositor EDS
% Data completeness (in resolution range)	93.2 (39.82-3.20) 93.8 (39.82-3.20)	Depositor EDS
R_{merge}	0.18	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.69 (at 3.18Å)	Xtriage
Refinement program	PHENIX 1.9	Depositor
R, R_{free}	0.215 , 0.260 0.214 , 0.254	Depositor DCC
R_{free} test set	4268 reflections (4.64%)	wwPDB-VP
Wilson B-factor (Å ²)	61.3	Xtriage
Anisotropy	0.469	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 47.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.026 for $1/2^*h-3/2^*k,-1/2^*h-1/2^*k,-1/2^*h+1/2^*k-l$ 0.023 for $1/2^*h+3/2^*k,1/2^*h-1/2^*k,-1/2^*h-1/2^*k-l$	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	28784	wwPDB-VP
Average B, all atoms (Å ²)	62.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.13% 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: 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	0.32	0/1849	0.88	7/2515 (0.3%)
1	B	0.29	0/1781	0.79	2/2420 (0.1%)
2	C	0.32	0/8969	0.78	8/12129 (0.1%)
3	D	0.41	6/11975 (0.1%)	0.83	21/16189 (0.1%)
4	E	0.29	0/772	0.76	0/1040
5	F	0.41	1/2779 (0.0%)	0.78	5/3737 (0.1%)
6	G	0.20	0/418	0.32	0/645
7	H	0.21	0/556	0.40	0/858
8	I	0.16	0/157	0.45	0/242
All	All	0.36	7/29256 (0.0%)	0.79	43/39775 (0.1%)

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
3	D	0	1

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	921	ARG	C-O	13.25	1.40	1.24
3	D	917	GLN	C-O	9.23	1.34	1.24
3	D	923	GLY	CA-C	-7.96	1.39	1.52
3	D	921	ARG	CA-C	6.72	1.61	1.52
3	D	925	GLU	N-CA	5.40	1.53	1.46

The worst 5 of 43 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	922	LEU	N-CA-C	-12.70	95.97	111.33
3	D	923	GLY	CA-C-N	-10.27	106.67	120.54
3	D	923	GLY	C-N-CA	-10.27	106.67	120.54
3	D	924	MET	CA-C-N	-10.19	103.17	121.81
3	D	924	MET	C-N-CA	-10.19	103.17	121.81

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	D	922	LEU	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	1814	0	1869	38	0
1	B	1750	0	1802	45	0
2	C	8792	0	8902	165	0
3	D	11759	0	12002	238	0
4	E	758	0	770	12	0
5	F	2737	0	2819	42	0
6	G	372	0	203	10	0
7	H	495	0	272	13	0
8	I	142	0	78	2	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	1	0
11	B	12	0	0	6	0
11	C	49	0	0	0	0
11	D	62	0	0	5	0
11	E	5	0	0	1	0
11	F	8	0	0	0	0
11	G	5	0	0	0	0
11	H	4	0	0	2	0
11	I	2	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	28784	0	28717	503	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 9.

The worst 5 of 503 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)
3:D:918:ALA:O	3:D:923:GLY:N	1.80	1.12
3:D:919:PHE:O	3:D:923:GLY:HA2	1.51	1.09
3:D:919:PHE:C	3:D:923:GLY:HA2	1.98	0.89
3:D:921:ARG:O	3:D:923:GLY:HA3	1.80	0.82
3:D:921:ARG:C	3:D:923:GLY:HA3	2.07	0.79

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%)	228 (99%)	2 (1%)	0	100	100
1	B	218/315 (69%)	210 (96%)	7 (3%)	1 (0%)	24	59
2	C	1111/1119 (99%)	1081 (97%)	29 (3%)	1 (0%)	48	79
3	D	1485/1524 (97%)	1445 (97%)	40 (3%)	0	100	100
4	E	92/99 (93%)	90 (98%)	2 (2%)	0	100	100
5	F	333/443 (75%)	329 (99%)	4 (1%)	0	100	100
All	All	3469/3815 (91%)	3383 (98%)	84 (2%)	2 (0%)	48	79

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	925	TYR
1	B	8	ALA

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%)	198 (98%)	3 (2%)	57	76
1	B	195/273 (71%)	190 (97%)	5 (3%)	40	69
2	C	939/941 (100%)	909 (97%)	30 (3%)	34	65
3	D	1256/1279 (98%)	1229 (98%)	27 (2%)	47	71
4	E	82/88 (93%)	81 (99%)	1 (1%)	63	79
5	F	293/388 (76%)	291 (99%)	2 (1%)	76	83
All	All	2966/3242 (92%)	2898 (98%)	68 (2%)	44	70

5 of 68 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	1067	VAL
3	D	1188	VAL
4	E	50	THR
2	C	513	VAL
2	C	429	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	724	GLN
3	D	1124	GLN
5	F	83	GLN
3	D	1033	GLN
3	D	66	GLN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	I	6/7 (85%)	1 (16%)	0

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
8	I	2	C

There are no RNA pucker outliers to report.

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 ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

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 ⓘ

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, 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%)	-0.10	1 (0%) 88 79	35, 54, 82, 135	1 (0%)
1	B	222/315 (70%)	0.05	3 (1%) 73 54	38, 61, 96, 118	0
2	C	1112/1119 (99%)	-0.11	4 (0%) 88 79	20, 49, 113, 153	3 (0%)
3	D	1486/1524 (97%)	0.05	25 (1%) 69 49	18, 53, 118, 151	4 (0%)
4	E	94/99 (94%)	-0.15	1 (1%) 78 61	24, 49, 86, 107	0
5	F	337/443 (76%)	0.31	10 (2%) 52 33	42, 73, 124, 149	0
6	G	18/21 (85%)	0.33	0 100 100	36, 64, 143, 151	0
7	H	24/27 (88%)	0.53	2 (8%) 17 11	59, 83, 139, 168	0
8	I	7/7 (100%)	0.10	1 (14%) 6 5	33, 35, 85, 108	0
All	All	3531/3870 (91%)	0.01	47 (1%) 75 55	18, 55, 116, 168	8 (0%)

The worst 5 of 47 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	923	GLY	6.9
3	D	922	LEU	6.4
5	F	191	ASN	5.1
3	D	184	GLU	4.5
5	F	146	GLY	4.4

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 oligosaccharides 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($\text{\AA}^2$)	Q<0.9
9	MG	B	2001	1/1	0.69	0.15	64,64,64,64	0
9	MG	F	2001	1/1	0.94	0.06	62,62,62,62	0
9	MG	D	1601	1/1	0.95	0.10	37,37,37,37	0
10	ZN	D	1603	1/1	0.97	0.15	45,45,45,45	0
9	MG	D	1604	1/1	0.99	0.03	19,19,19,19	0
10	ZN	D	1602	1/1	1.00	0.14	6,6,6,6	0

6.5 Other polymers [i](#)

There are no such residues in this entry.