



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 04:58 am BST

PDB ID : 4GZY  
Title : Crystal structures of bacterial RNA Polymerase paused elongation complexes  
Authors : Weixlbaumer, A.; Leon, K.; Landick, R.; Darst, S.A.  
Deposited on : 2012-09-06  
Resolution : 3.51 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

<https://www.wwpdb.org/validation/2017/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
Xtriage (Phenix)	:	1.13
EDS	:	2.11
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)	:	2.11

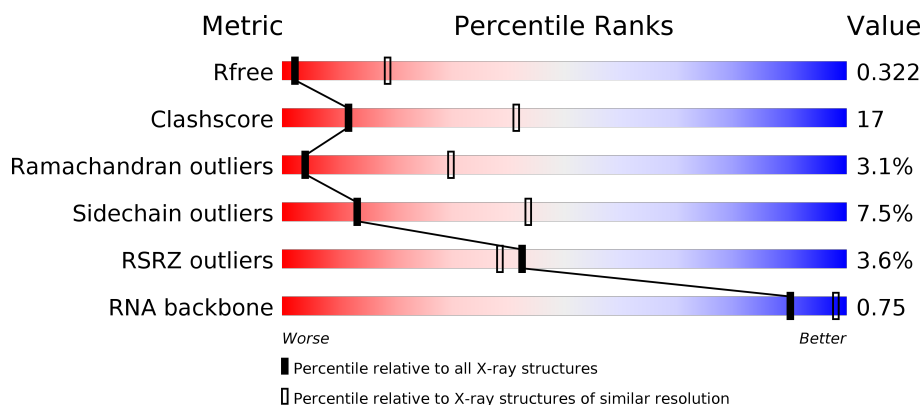
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.51 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	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)
RSRZ outliers	127900	1559 (3.60-3.40)
RNA backbone	3102	1002 (4.00-3.00)

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 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	
1	B	315	
2	C	1119	
3	D	1534	

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	99	<div><div></div><div>5%</div><div>55%</div><div>31%</div><div>8%</div><div>6%</div></div>
5	N	13	<div><div></div><div>69%</div><div>69%</div><div>15%</div><div>15%</div></div>
6	R	29	<div><div></div><div>3%</div><div>10%</div><div>17%</div><div>69%</div></div>
7	T	22	<div><div></div><div>27%</div><div>64%</div><div>36%</div></div>

## 2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 24400 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	223	Total	C	N	O	S	0	0	0
			1759	1123	306	328	2			
1	B	223	Total	C	N	O	S	0	0	0
			1759	1123	306	328	2			

- Molecule 2 is a protein called DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1083	Total	C	N	O	S	0	0	0
			8548	5412	1524	1588	24			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1358	Total	C	N	O	S	0	0	0
			10714	6780	1900	2001	33			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	1525	HIS	-	EXPRESSION TAG	UNP Q8RQE8
D	1526	HIS	-	EXPRESSION TAG	UNP Q8RQE8
D	1527	HIS	-	EXPRESSION TAG	UNP Q8RQE8
D	1528	HIS	-	EXPRESSION TAG	UNP Q8RQE8
D	1529	HIS	-	EXPRESSION TAG	UNP Q8RQE8
D	1530	HIS	-	EXPRESSION TAG	UNP Q8RQE8
D	1531	HIS	-	EXPRESSION TAG	UNP Q8RQE8
D	1532	HIS	-	EXPRESSION TAG	UNP Q8RQE8
D	1533	HIS	-	EXPRESSION TAG	UNP Q8RQE8
D	1534	HIS	-	EXPRESSION TAG	UNP Q8RQE8

- Molecule 4 is a protein called DNA-directed RNA polymerase subunit omega.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	93	Total	C	N	O	S	0	0	0
			754	481	131	138	4			

- Molecule 5 is a DNA chain called non-template DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	N	11	Total	C	N	O	P	0	0	0
			225	107	43	64	11			

- Molecule 6 is a RNA chain called RNA transcript.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	R	9	Total	C	N	O	P	0	0	0
			191	85	31	66	9			

- Molecule 7 is a DNA chain called template DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	T	22	Total	C	N	O	P	0	0	0
			447	213	81	131	22			

- Molecule 8 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

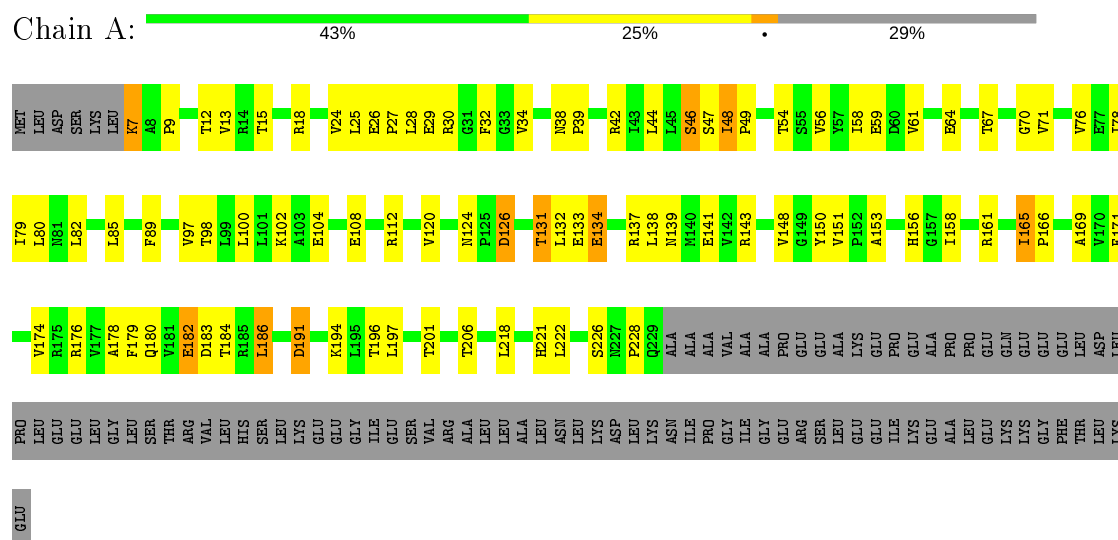
- Molecule 9 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

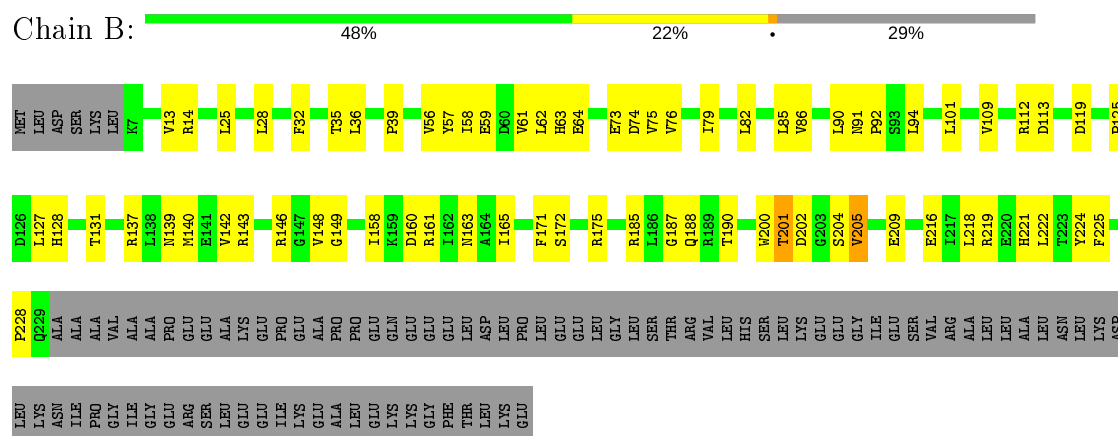
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA 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

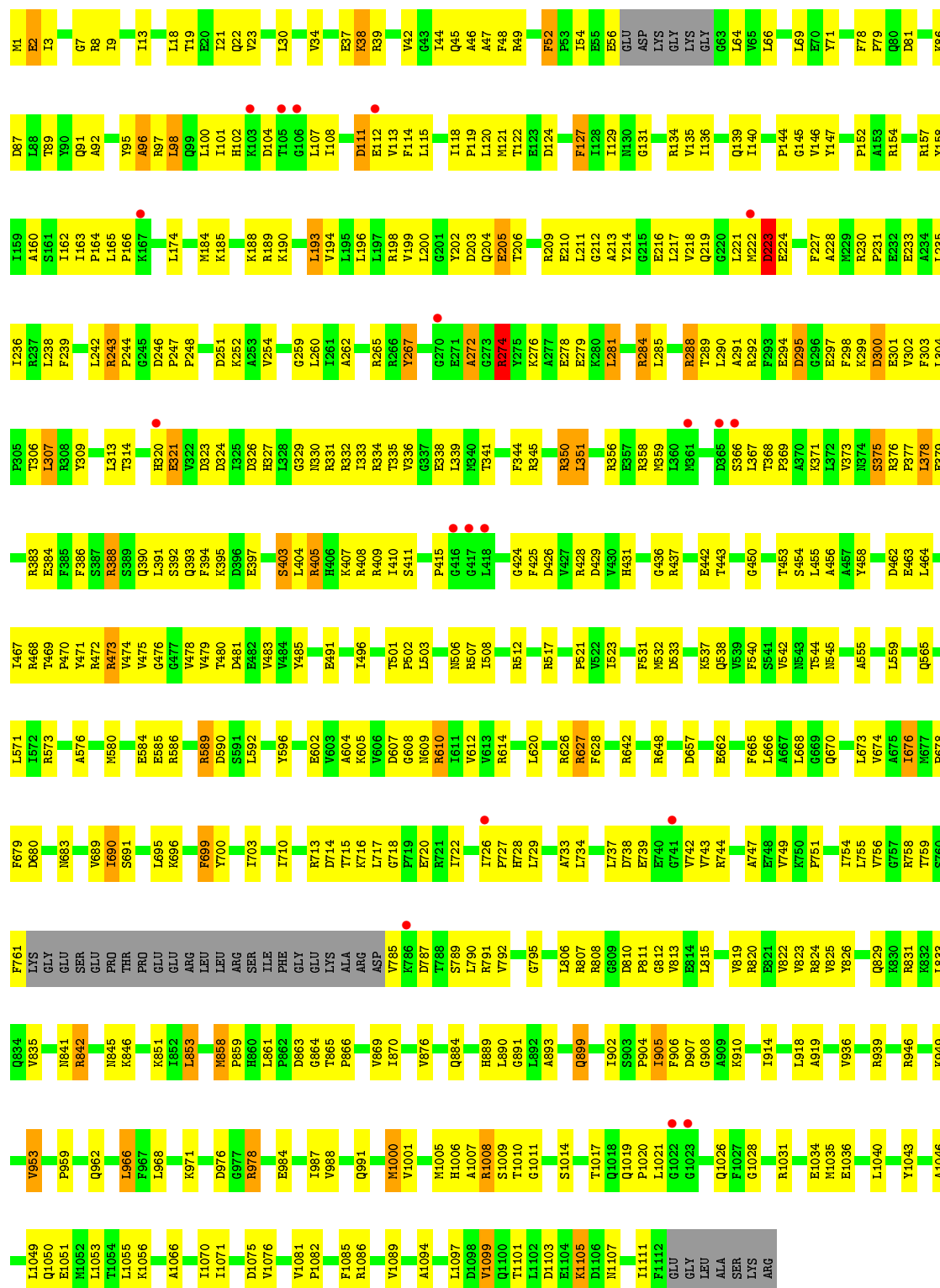


- Molecule 1: DNA-directed RNA polymerase subunit alpha



- Molecule 2: DNA-directed RNA polymerase subunit beta



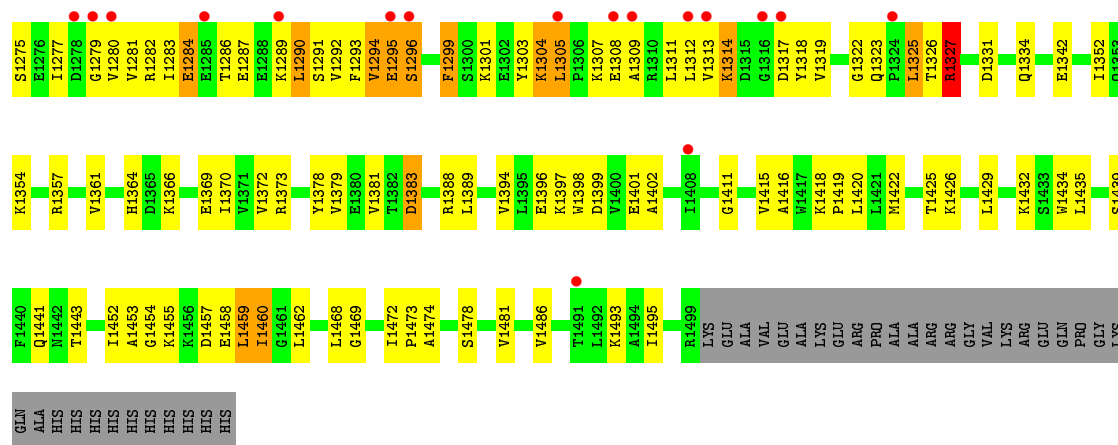


• Molecule 3: DNA-directed RNA polymerase subunit beta'

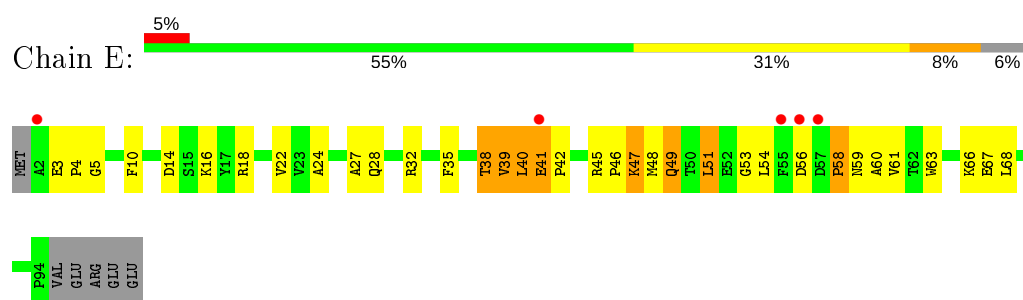




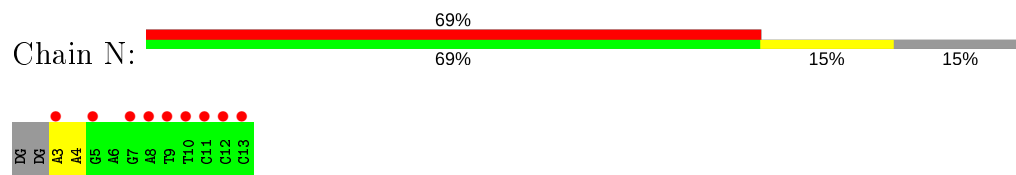




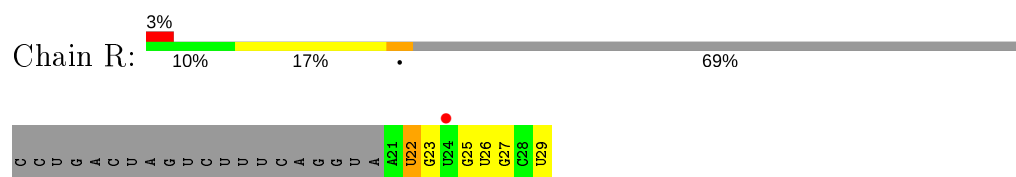
- Molecule 4: DNA-directed RNA polymerase subunit omega



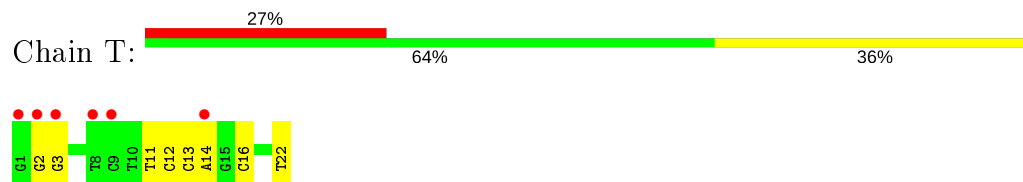
- Molecule 5: non-template DNA



- Molecule 6: RNA transcript



- Molecule 7: template DNA



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	207.21Å 207.21Å 203.22Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.77 – 3.51 49.77 – 3.51	Depositor EDS
% Data completeness (in resolution range)	99.7 (49.77-3.51) 87.9 (49.77-3.51)	Depositor EDS
$R_{merge}$	0.15	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.76 (at 3.48Å)	Xtriage
Refinement program	PHENIX 1.8_1069, CNS	Depositor
R, $R_{free}$	0.263 , 0.322 0.265 , 0.322	Depositor DCC
$R_{free}$ test set	3199 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	95.9	Xtriage
Anisotropy	0.360	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 55.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.038 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	24400	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	115.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.70% 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, 0.333 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: 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	0.28	0/1791	0.49	0/2436
1	B	0.29	0/1791	0.48	0/2436
2	C	0.30	0/8711	0.52	1/11784 (0.0%)
3	D	0.30	0/10897	0.50	1/14726 (0.0%)
4	E	0.29	0/768	0.54	0/1035
5	N	0.43	0/252	1.11	0/386
6	R	0.22	0/212	0.77	0/328
7	T	0.43	0/500	1.09	0/768
All	All	0.30	0/24922	0.54	2/33899 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	705	ALA	C-N-CD	5.90	140.78	128.40
2	C	853	LEU	CA-CB-CG	5.11	127.05	115.30

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	1759	0	1805	59	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	1759	0	1805	53	0
2	C	8548	0	8650	377	0
3	D	10714	0	10936	395	0
4	E	754	0	769	24	0
5	N	225	0	124	1	0
6	R	191	0	95	7	0
7	T	447	0	248	6	0
8	D	2	0	0	0	0
9	D	1	0	0	0	0
All	All	24400	0	24432	851	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 17.

The worst 5 of 851 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:610:ARG:HH11	2:C:610:ARG:HG3	1.16	1.08
3:D:346:ARG:NH1	3:D:347:VAL:O	1.89	1.05
2:C:405:ARG:NH1	2:C:442:GLU:OE2	1.90	1.04
2:C:274:ARG:NH1	2:C:284:ARG:HH12	1.54	1.03
3:D:550:ARG:HH11	3:D:573:MET:HB3	1.25	0.98

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	221/315 (70%)	193 (87%)	19 (9%)	9 (4%)	3	23
1	B	221/315 (70%)	199 (90%)	21 (10%)	1 (0%)	29	68

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	C	1077/1119 (96%)	901 (84%)	149 (14%)	27 (2%)	5	34
3	D	1352/1534 (88%)	1112 (82%)	189 (14%)	51 (4%)	3	25
4	E	91/99 (92%)	74 (81%)	14 (15%)	3 (3%)	4	28
All	All	2962/3382 (88%)	2479 (84%)	392 (13%)	91 (3%)	4	30

5 of 91 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	184	THR
2	C	2	GLU
2	C	111	ASP
2	C	164	PRO
2	C	213	ALA

### 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	196/273 (72%)	187 (95%)	9 (5%)	27	61
1	B	196/273 (72%)	191 (97%)	5 (3%)	46	74
2	C	912/941 (97%)	837 (92%)	75 (8%)	11	40
3	D	1147/1289 (89%)	1058 (92%)	89 (8%)	12	42
4	E	82/88 (93%)	70 (85%)	12 (15%)	3	18
All	All	2533/2864 (88%)	2343 (92%)	190 (8%)	13	43

5 of 190 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	1021	LEU
3	D	178	LEU
3	D	1460	ILE
2	C	1105	LYS
3	D	79	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	670	GLN
2	C	829	GLN
3	D	101	HIS
2	C	609	ASN
2	C	884	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
6	R	8/29 (27%)	1 (12%)	0

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
6	R	22	U

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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 3 ligands modelled in this entry, 3 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<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	223/315 (70%)	-0.31	0	100	100	45, 90, 147, 194	0
1	B	223/315 (70%)	-0.38	0	100	100	40, 86, 139, 173	0
2	C	1083/1119 (96%)	-0.11	19 (1%)	68	62	32, 101, 186, 255	0
3	D	1358/1534 (88%)	0.02	68 (5%)	28	25	34, 102, 194, 264	0
4	E	93/99 (93%)	0.10	5 (5%)	25	23	58, 107, 184, 204	0
5	N	11/13 (84%)	2.78	9 (81%)	0	0	369, 412, 462, 465	0
6	R	9/29 (31%)	1.02	1 (11%)	5	6	178, 188, 215, 227	0
7	T	22/22 (100%)	1.70	6 (27%)	0	0	202, 334, 453, 466	0
All	All	3022/3446 (87%)	-0.05	108 (3%)	42	38	32, 100, 193, 466	0

The worst 5 of 108 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	E	56	ASP	10.4
3	D	802	ALA	10.0
3	D	810	GLU	9.9
2	C	417	GLY	6.8
7	T	9	DC	6.2

### 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. 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	B-factors( $\text{\AA}^2$ )	Q<0.9
8	ZN	D	1601	1/1	0.82	0.27	188,188,188,188	0
9	MG	D	1603	1/1	0.96	0.28	46,46,46,46	0
8	ZN	D	1602	1/1	0.98	0.12	80,80,80,80	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.