



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 21, 2020 – 12:59 PM BST

PDB ID : 6N62  
Title : Escherichia coli RNA polymerase sigma70-holoenzyme bound to upstream fork promoter DNA  
Authors : Braffman, N.; Hauver, J.; Campbell, E.A.; Darst, S.A.  
Deposited on : 2018-11-24  
Resolution : 3.80 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	:	2.13.1
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)	:	2.13.1

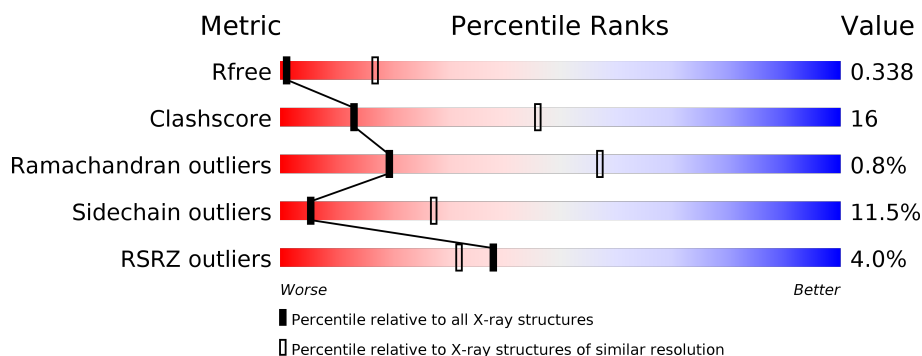
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.80 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	1212 (4.00-3.60)
Clashscore	141614	1288 (4.00-3.60)
Ramachandran outliers	138981	1243 (4.00-3.60)
Sidechain outliers	138945	1237 (4.00-3.60)
RSRZ outliers	127900	1121 (4.00-3.60)

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	239	<div> <div>10%</div> <div>58%</div> <div>34%</div> <div>• 5%</div> </div>
1	B	239	<div> <div>3%</div> <div>56%</div> <div>30%</div> <div>• 9%</div> </div>
2	C	1342	<div> <div>%</div> <div>60%</div> <div>34%</div> <div>5%</div> </div>
3	D	1409	<div> <div>4%</div> <div>51%</div> <div>31%</div> <div>5%</div> <div>13%</div> </div>
4	E	91	<div> <div>2%</div> <div>51%</div> <div>25%</div> <div>11%</div> <div>13%</div> </div>
5	F	613	<div> <div>7%</div> <div>50%</div> <div>24%</div> <div>•</div> <div>23%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	N	29	<div><div></div><div>28%</div><div>69%</div><div></div></div>
7	T	24	<div><div></div><div>38%</div><div>63%</div><div></div></div>

## 2 Entry composition [i](#)

There are 9 unique types of molecules in this entry. The entry contains 29075 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	228	Total	C	N	O	S	0	0	0
			1750	1088	312	344	6			
1	B	217	Total	C	N	O	S	0	0	0
			1667	1041	293	327	6			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	235	GLU	-	expression tag	UNP P0A7Z6
A	236	VAL	-	expression tag	UNP P0A7Z6
A	237	LEU	-	expression tag	UNP P0A7Z6
A	238	PHE	-	expression tag	UNP P0A7Z6
A	239	GLN	-	expression tag	UNP P0A7Z6
B	235	GLU	-	expression tag	UNP P0A7Z6
B	236	VAL	-	expression tag	UNP P0A7Z6
B	237	LEU	-	expression tag	UNP P0A7Z6
B	238	PHE	-	expression tag	UNP P0A7Z6
B	239	GLN	-	expression tag	UNP P0A7Z6

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1337	Total	C	N	O	S	0	0	0
			10545	6616	1838	2048	43			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1226	Total	C	N	O	S	0	0	0
			9557	6006	1714	1791	46			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	1	VAL	-	expression tag	UNP P0A8T8
D	1408	LEU	-	expression tag	UNP P0A8T8
D	1409	GLU	-	expression tag	UNP P0A8T8

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	79	Total	C	N	O	S	0	0	0
			627	382	118	126	1			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	471	Total	C	N	O	S	0	0	0
			3839	2405	684	727	23			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	149	ASN	ASP	conflict	UNP Q0P6L9

- Molecule 6 is a DNA chain called non-template strand DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	N	29	Total	C	N	O	P	0	0	0
			595	284	106	176	29			

- Molecule 7 is a DNA chain called template strand DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	T	24	Total	C	N	O	P	0	0	0
			492	233	94	141	24			

- 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		

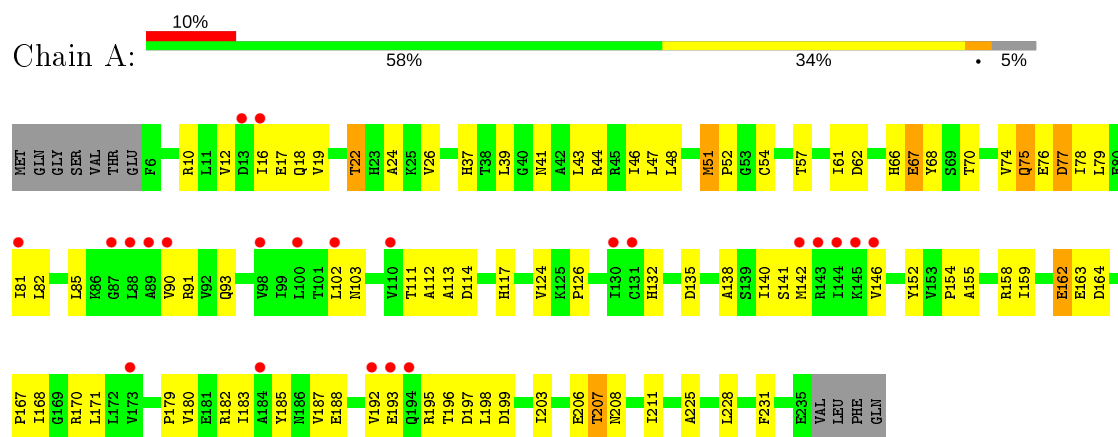
- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

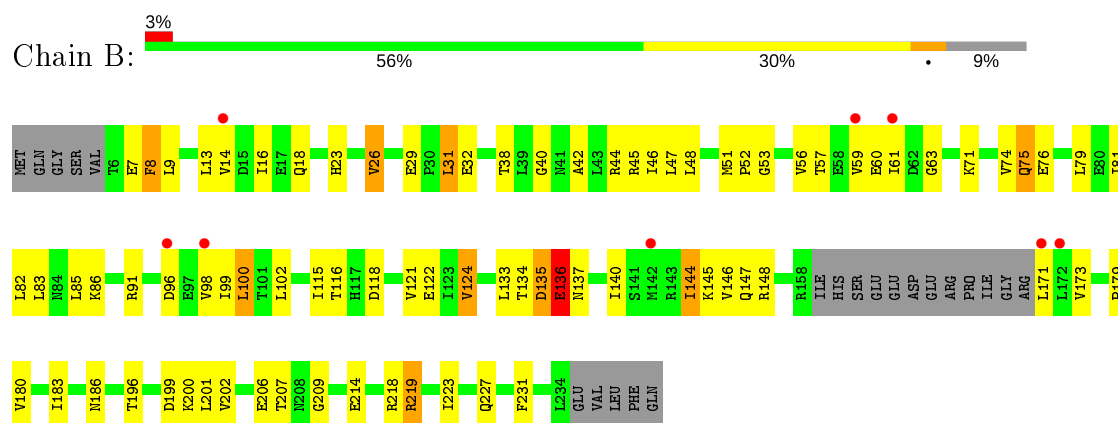
### 3 Residue-property plots

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.

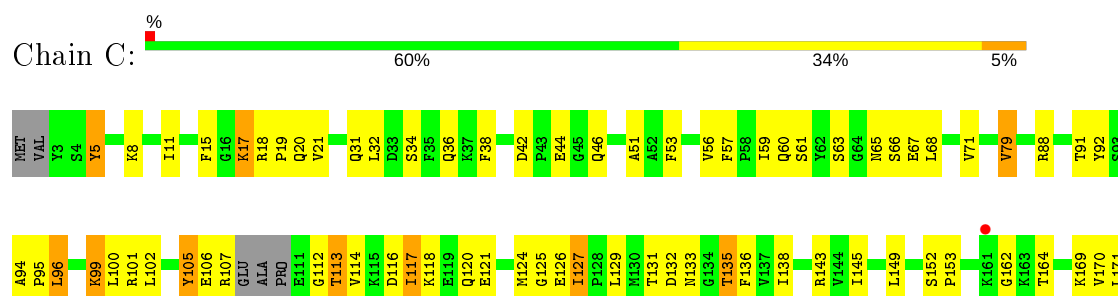
#### • Molecule 1: DNA-directed RNA polymerase subunit alpha



#### • Molecule 1: DNA-directed RNA polymerase subunit alpha



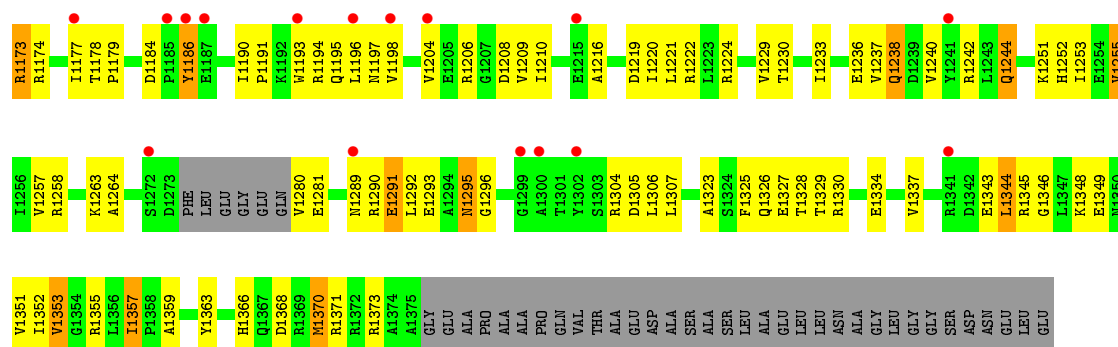
#### • Molecule 2: DNA-directed RNA polymerase subunit beta



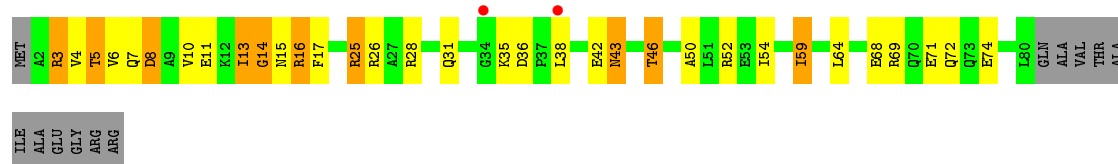




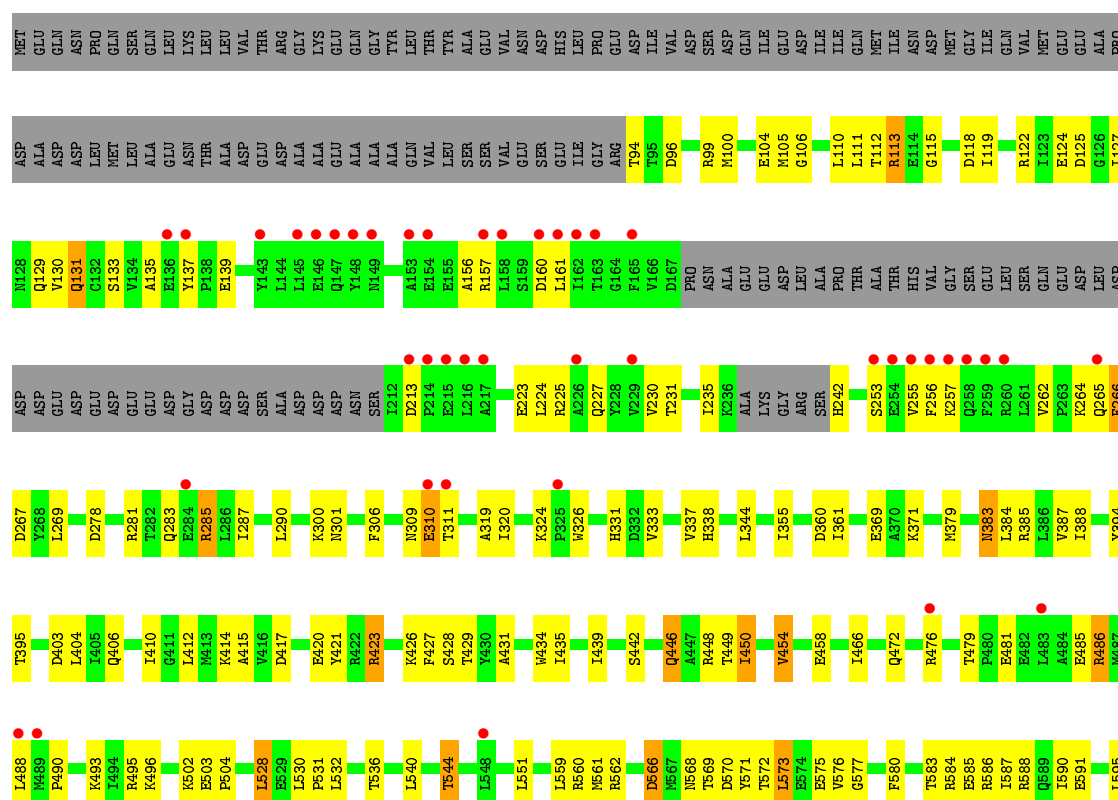




• Molecule 4: DNA-directed RNA polymerase subunit omega

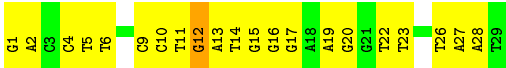


• Molecule 5: RNA polymerase sigma factor RpoD

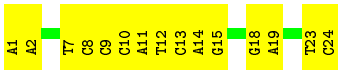




● Molecule 6: non-template strand DNA



● Molecule 7: template strand DNA



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	173.77Å 173.77Å 388.77Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	24.88 – 3.80 49.77 – 3.80	Depositor EDS
% Data completeness (in resolution range)	99.7 (24.88-3.80) 99.7 (49.77-3.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.08 (at 3.77Å)	Xtriage
Refinement program	PHENIX (1.13 _2998: ???)	Depositor
R, $R_{free}$	0.286 , 0.334 0.297 , 0.338	Depositor DCC
$R_{free}$ test set	1999 reflections (3.38%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	171.7	Xtriage
Anisotropy	0.039	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 101.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	29075	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	164.0	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

### 5.1 Standard geometry

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	0.26	0/1771	0.51	0/2401
1	B	0.30	0/1686	0.53	0/2285
2	C	0.28	0/10712	0.50	3/14450 (0.0%)
3	D	0.27	0/9702	0.48	1/13092 (0.0%)
4	E	0.26	0/629	0.48	0/847
5	F	0.27	0/3891	0.44	0/5231
6	N	0.58	1/666 (0.2%)	0.91	0/1026
7	T	0.56	0/552	0.83	0/849
All	All	0.29	1/29609 (0.0%)	0.51	4/40181 (0.0%)

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
1	B	0	1
2	C	0	7
3	D	0	2
All	All	0	10

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	N	12	DG	C1'-N9	-7.88	1.36	1.47

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	1161	LEU	CA-CB-CG	6.94	131.26	115.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	1151	LEU	CA-CB-CG	6.74	130.80	115.30
2	C	1151	LEU	CB-CG-CD2	6.11	121.39	111.00
3	D	1343	GLU	C-N-CA	5.53	135.52	121.70

There are no chirality outliers.

5 of 10 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	135	ASP	Peptide
2	C	169	LYS	Peptide
2	C	572	ILE	Peptide
2	C	811	ASN	Peptide
2	C	985	GLU	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	1750	0	1764	47	0
1	B	1667	0	1689	45	0
2	C	10545	0	10559	371	0
3	D	9557	0	9764	356	0
4	E	627	0	634	28	0
5	F	3839	0	3904	97	0
6	N	595	0	329	51	0
7	T	492	0	269	32	0
8	D	1	0	0	0	0
9	D	2	0	0	0	0
All	All	29075	0	28912	946	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 16.

The worst 5 of 946 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:116:ASP:O	2:C:117:ILE:HG23	1.37	1.20
2:C:116:ASP:O	2:C:117:ILE:CG2	1.94	1.14
6:N:13:DA:H1'	6:N:14:DT:H5'	1.27	1.07
2:C:106:GLU:O	2:C:114:VAL:HG22	1.55	1.06
6:N:12:DG:H4'	6:N:13:DA:OP1	1.52	1.05

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	226/239 (95%)	193 (85%)	31 (14%)	2 (1%)	17	54
1	B	213/239 (89%)	183 (86%)	26 (12%)	4 (2%)	8	42
2	C	1333/1342 (99%)	1179 (88%)	140 (10%)	14 (1%)	14	51
3	D	1218/1409 (86%)	1093 (90%)	118 (10%)	7 (1%)	25	62
4	E	77/91 (85%)	69 (90%)	7 (9%)	1 (1%)	12	48
5	F	465/613 (76%)	426 (92%)	37 (8%)	2 (0%)	34	70
All	All	3532/3933 (90%)	3143 (89%)	359 (10%)	30 (1%)	19	57

5 of 30 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	163	GLU
2	C	399	ALA
2	C	573	ASN
2	C	1150	ASP
2	C	1151	LEU

### 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	191/206 (93%)	168 (88%)	23 (12%)	5	25
1	B	183/206 (89%)	163 (89%)	20 (11%)	6	29
2	C	1152/1157 (100%)	1027 (89%)	125 (11%)	6	29
3	D	1026/1170 (88%)	901 (88%)	125 (12%)	5	25
4	E	67/75 (89%)	55 (82%)	12 (18%)	2	12
5	F	420/540 (78%)	376 (90%)	44 (10%)	7	30
All	All	3039/3354 (91%)	2690 (88%)	349 (12%)	5	27

5 of 349 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	1198	LEU
3	D	227	PHE
5	F	383	ASN
2	C	1265	PHE
3	D	86	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 73 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	1220	GLN
3	D	294	ASN
5	F	331	HIS
2	C	1314	GLN
3	D	341	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.



## 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 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	228/239 (95%)	0.54	23 (10%) 7 6	144, 165, 185, 215	0
1	B	217/239 (90%)	0.07	8 (3%) 41 34	145, 167, 185, 189	0
2	C	1337/1342 (99%)	0.03	17 (1%) 77 70	30, 155, 181, 201	0
3	D	1226/1409 (87%)	0.25	50 (4%) 37 31	144, 160, 192, 206	0
4	E	79/91 (86%)	-0.07	2 (2%) 57 49	146, 154, 181, 188	0
5	F	471/613 (76%)	0.19	43 (9%) 9 7	144, 174, 195, 207	0
6	N	29/29 (100%)	-0.61	0 100 100	112, 180, 203, 205	0
7	T	24/24 (100%)	-0.56	0 100 100	108, 189, 204, 206	0
All	All	3611/3986 (90%)	0.15	143 (3%) 38 32	30, 161, 191, 215	0

The worst 5 of 143 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	193	GLU	9.2
1	A	89	ALA	8.7
1	A	194	GLN	7.8
5	F	259	PHE	7.2
1	A	90	VAL	6.5

### 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 [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
9	ZN	D	1502	1/1	0.87	0.07	161,161,161,161	0
8	MG	D	1501	1/1	0.94	0.61	144,144,144,144	0
9	ZN	D	1503	1/1	0.95	0.14	160,160,160,160	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.