



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 7, 2023 – 06:47 PM EDT

PDB ID : 1NIK  
Title : Wild Type RNA Polymerase II  
Authors : Bushnell, D.A.; Kornberg, R.D.  
Deposited on : 2002-12-24  
Resolution : 4.10 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 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  
Xtriage (Phenix) : 1.13  
EDS : 2.35  
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

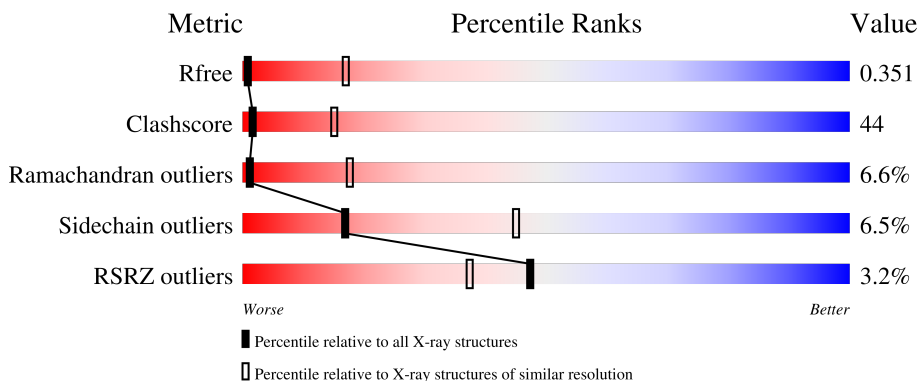
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 4.10 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	1193 (4.50-3.70)
Clashscore	141614	1003 (4.44-3.76)
Ramachandran outliers	138981	1005 (4.48-3.72)
Sidechain outliers	138945	1199 (4.50-3.70)
RSRZ outliers	127900	1034 (4.50-3.70)

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  $\geq 3$ , 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	1733	 2% 32% 40% 7% 20%
2	B	1224	 2% 32% 51% 7% 10%
3	C	318	 2% 37% 40% 7% 16%
4	D	161	 91% 5%
5	E	215	 4% 42% 55%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
6	F	155	<p>%</p> <p>18% 32% 46%</p>
7	G	170	<p>95% 5%</p>
8	H	146	<p>5% 30% 52% 9% 9%</p>
9	I	122	<p>13% 35% 51% 11%</p>
10	J	70	<p>4% 31% 51% 9% 7%</p>
11	K	120	<p>2% 41% 50% 5%</p>
12	L	70	<p>4% 17% 33% 13% 34%</p>

## 2 Entry composition [i](#)

There are 13 unique types of molecules in this entry. The entry contains 28295 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 RPB1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1388	10864	6858	1899	2046	61	0	0	7

- Molecule 2 is a protein called ORF YOR151c.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	1097	8721	5526	1523	1618	54	0	0	0

- Molecule 3 is a protein called DNA-directed RNA polymerase II, chain RPB3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	266	2095	1317	348	417	13	0	0	0

- Molecule 4 is a protein called DNA-directed RNA polymerase II, chain RPB4.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
4	D	153	Total 153 C 153	0	0	153

- Molecule 5 is a protein called DNA-directed RNA polymerase II, chain RPB5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	E	214	1752	1111	309	321	11	0	0	0

- Molecule 6 is a protein called DNA-directed RNA polymerase I, II and III 23 kDa polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	F	84	679	434	115	127	3	0	0	0

- Molecule 7 is a protein called DNA-directed RNA polymerase II, chain RPB7.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf	Trace
7	G	170	Total	C	0	0	170
			170	170			

- Molecule 8 is a protein called DNA-directed RNA polymerase subunit RPB8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	H	133	Total	C	N	O	S	0	0	0
			1068	673	180	211	4			

- Molecule 9 is a protein called DNA-directed RNA polymerase II, chain RPB9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
9	I	119	Total	C	N	O	S	0	0	0
			971	596	179	186	10			

- Molecule 10 is a protein called DNA-directed RNA polymerase II, chain RPB10.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
10	J	65	Total	C	N	O	S	0	0	0
			532	339	93	94	6			

- Molecule 11 is a protein called DNA-directed RNA polymerase II, chain RPB11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
11	K	114	Total	C	N	O	S	0	0	0
			919	590	156	171	2			

- Molecule 12 is a protein called DNA-directed RNA polymerase II, chain RPB12.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
12	L	46	Total	C	N	O	S	0	0	0
			364	224	72	64	4			

- Molecule 13 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
13	A	2	Total	Zn	0	0
			2	2		
13	B	1	Total	Zn	0	0
			1	1		

*Continued on next page...*

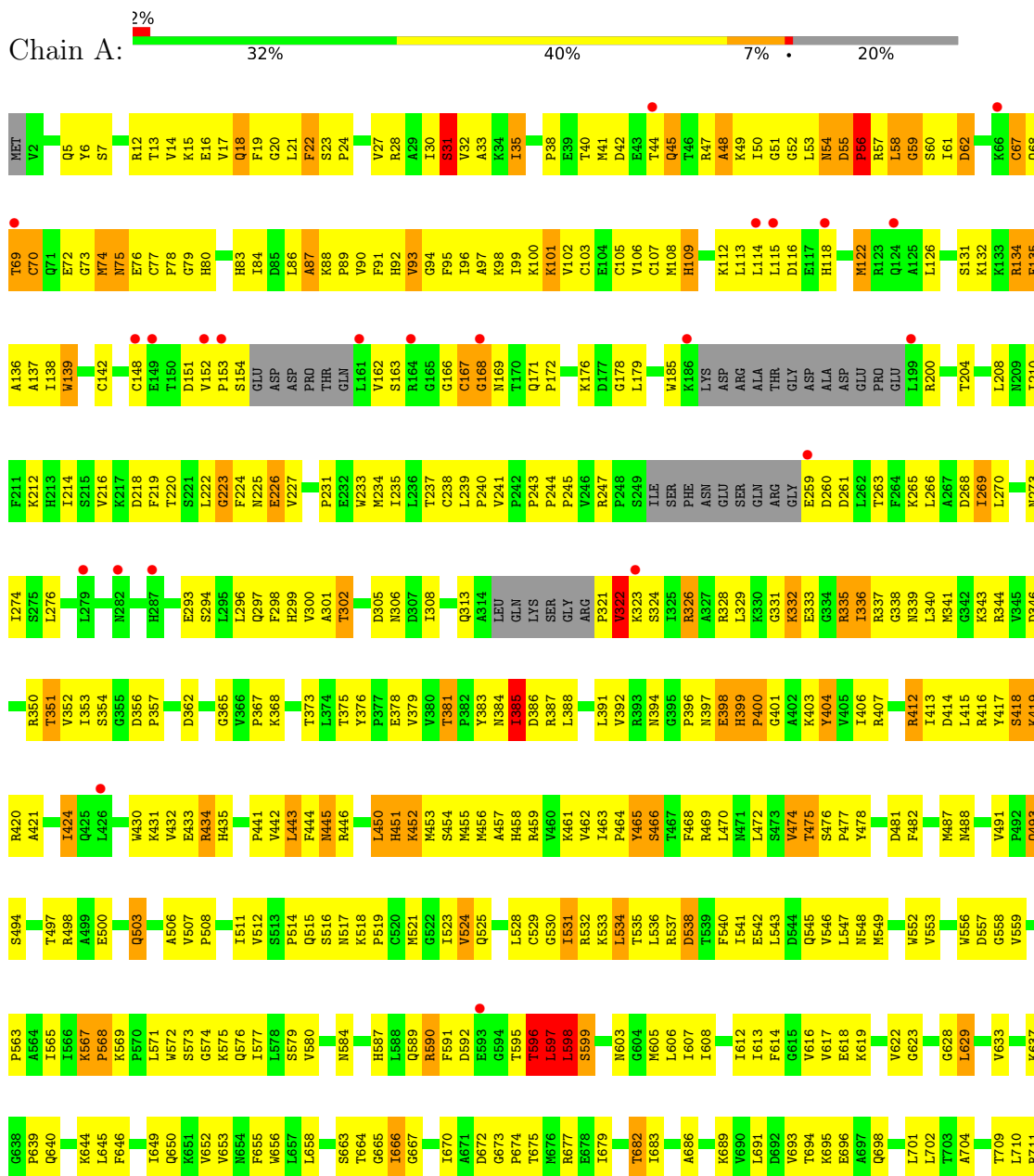
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
13	C	1	Total 1	Zn 1	0	0
13	I	2	Total 2	Zn 2	0	0
13	J	1	Total 1	Zn 1	0	0

### 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.

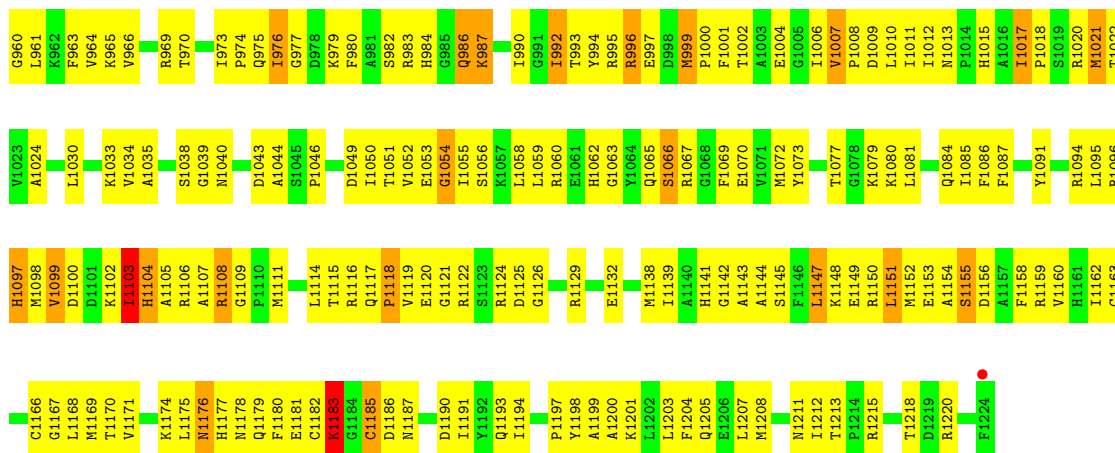
#### ● Molecule 1: RPB1



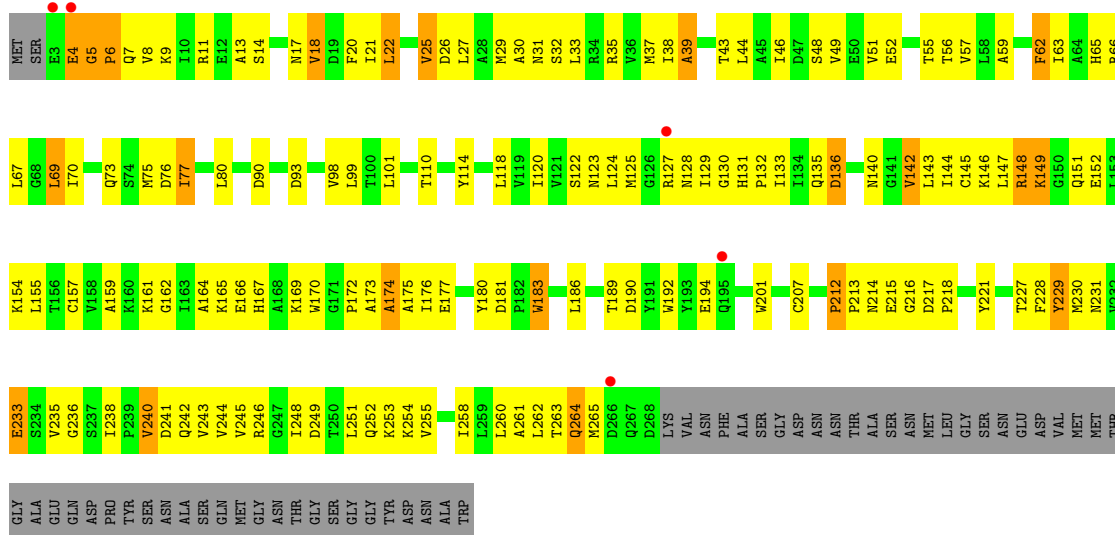




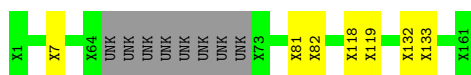
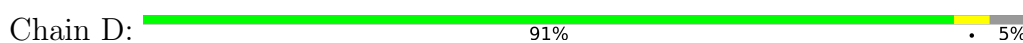




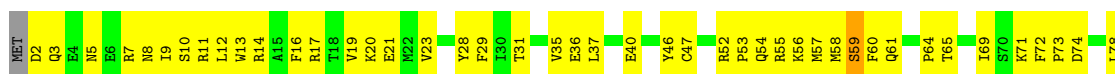
• Molecule 3: DNA-directed RNA polymerase II, chain RPB3



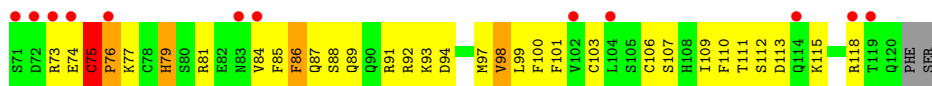
• Molecule 4: DNA-directed RNA polymerase II, chain RPB4



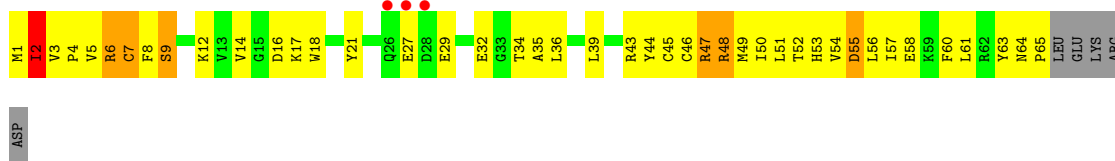
• Molecule 5: DNA-directed RNA polymerase II, chain RPB5



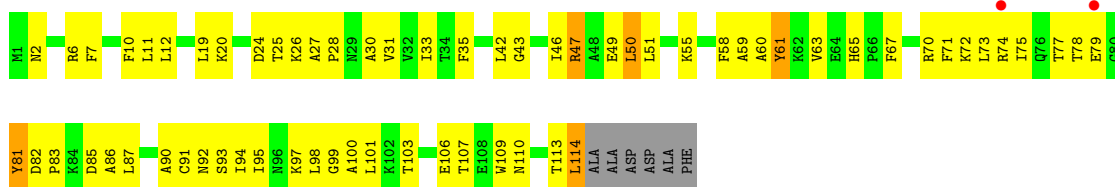
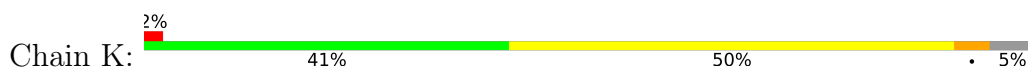




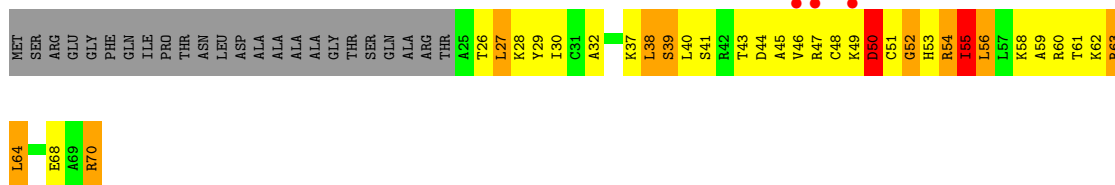
● Molecule 10: DNA-directed RNA polymerase II, chain RPB10



● Molecule 11: DNA-directed RNA polymerase II, chain RPB11



● Molecule 12: DNA-directed RNA polymerase II, chain RPB12



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	224.10Å 394.46Å 284.34Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.00 – 4.10 39.97 – 4.11	Depositor EDS
% Data completeness (in resolution range)	(Not available) (40.00-4.10) 86.4 (39.97-4.11)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.11	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.58 (at 4.13Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.334 , 0.360 0.330 , 0.351	Depositor DCC
$R_{free}$ test set	3272 reflections (3.38%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	114.0	Xtriage
Anisotropy	0.378	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 168.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	0.044 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.047 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
$F_o, F_c$ correlation	0.78	EDS
Total number of atoms	28295	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	194.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.28% 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

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	0.41	0/11048	0.71	6/14936 (0.0%)
2	B	0.46	0/8891	0.71	1/11990 (0.0%)
3	C	0.48	0/2133	0.76	2/2891 (0.1%)
5	E	0.37	0/1788	0.65	0/2406
6	F	0.40	0/691	0.64	0/933
8	H	0.40	0/1086	0.73	0/1470
9	I	0.48	0/989	0.76	1/1331 (0.1%)
10	J	0.53	0/541	0.78	0/727
11	K	0.46	0/937	0.68	0/1265
12	L	0.49	0/366	0.78	0/485
All	All	0.44	0/28470	0.71	10/38434 (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
2	B	0	1

There are no bond length outliers.

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1392	SER	N-CA-C	6.34	128.13	111.00
3	C	39	ALA	N-CA-C	6.06	127.35	111.00
1	A	398	GLU	N-CA-C	-5.77	95.43	111.00
3	C	183	TRP	N-CA-C	-5.61	95.86	111.00
2	B	647	GLY	N-CA-C	5.22	126.15	113.10

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	811	TYR	Sidechain

## 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	10864	0	10959	1032	1
2	B	8721	0	8746	900	0
3	C	2095	0	2052	162	1
4	D	153	0	0	4	0
5	E	1752	0	1776	129	0
6	F	679	0	701	66	0
7	G	170	0	0	5	0
8	H	1068	0	1040	134	0
9	I	971	0	933	101	0
10	J	532	0	544	77	0
11	K	919	0	929	84	0
12	L	364	0	390	50	0
13	A	2	0	0	0	0
13	B	1	0	0	0	0
13	C	1	0	0	0	0
13	I	2	0	0	1	0
13	J	1	0	0	1	0
All	All	28295	0	28070	2499	1

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 44.

The worst 5 of 2499 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:D:132:UNK:CA	4:D:133:UNK:CA	2.01	1.36
12:L:60:ARG:HG3	12:L:61:THR:H	1.05	1.17
1:A:567:LYS:HB2	1:A:568:PRO:HD2	1.18	1.15

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
9:I:111:THR:HG22	9:I:113:ASP:H	1.05	1.14
1:A:1364:ASN:ND2	1:A:1366:ARG:HG2	1.65	1.11

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:419:LYS:NZ	3:C:90:ASP:N[4_555]	1.74	0.46

## 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	1365/1733 (79%)	1022 (75%)	251 (18%)	92 (7%)	1	17
2	B	1077/1224 (88%)	839 (78%)	169 (16%)	69 (6%)	1	18
3	C	264/318 (83%)	208 (79%)	40 (15%)	16 (6%)	1	19
5	E	212/215 (99%)	170 (80%)	33 (16%)	9 (4%)	3	24
6	F	82/155 (53%)	64 (78%)	15 (18%)	3 (4%)	3	27
8	H	129/146 (88%)	93 (72%)	21 (16%)	15 (12%)	0	6
9	I	117/122 (96%)	93 (80%)	15 (13%)	9 (8%)	1	15
10	J	63/70 (90%)	47 (75%)	12 (19%)	4 (6%)	1	19
11	K	112/120 (93%)	96 (86%)	15 (13%)	1 (1%)	17	54
12	L	44/70 (63%)	22 (50%)	13 (30%)	9 (20%)	0	2
All	All	3465/4173 (83%)	2654 (77%)	584 (17%)	227 (7%)	1	18

5 of 227 Ramachandran outliers are listed below:



Mol	Chain	Res	Type
1	A	31	SER
1	A	48	ALA
1	A	55	ASP
1	A	56	PRO
1	A	74	MET

### 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	1206/1520 (79%)	1128 (94%)	78 (6%)	17	45
2	B	952/1061 (90%)	886 (93%)	66 (7%)	15	43
3	C	234/274 (85%)	222 (95%)	12 (5%)	24	52
5	E	196/197 (100%)	189 (96%)	7 (4%)	35	60
6	F	74/137 (54%)	68 (92%)	6 (8%)	11	38
8	H	117/128 (91%)	112 (96%)	5 (4%)	29	56
9	I	113/116 (97%)	104 (92%)	9 (8%)	12	39
10	J	60/65 (92%)	56 (93%)	4 (7%)	16	44
11	K	99/102 (97%)	90 (91%)	9 (9%)	9	33
12	L	40/57 (70%)	35 (88%)	5 (12%)	4	22
All	All	3091/3657 (84%)	2890 (94%)	201 (6%)	17	45

5 of 201 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	732	SER
2	B	1185	CYS
12	L	68	GLU
2	B	791	THR
2	B	996	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 74 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	123	ASN
11	K	65	HIS
3	C	242	GLN
5	E	114	ASN
1	A	1364	ASN

### 5.3.3 RNA [i](#)

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 7 ligands modelled in this entry, 7 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 [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, 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	1388/1733 (80%)	-0.07	38 (2%) 54 44	20, 181, 283, 321	0
2	B	1097/1224 (89%)	-0.02	28 (2%) 56 45	74, 189, 297, 321	0
3	C	266/318 (83%)	0.00	5 (1%) 66 58	87, 180, 262, 321	0
4	D	0/161	-	-	-	-
5	E	214/215 (99%)	0.20	9 (4%) 36 29	88, 223, 305, 321	0
6	F	84/155 (54%)	0.05	2 (2%) 59 49	90, 163, 254, 321	0
7	G	0/170	-	-	-	-
8	H	133/146 (91%)	0.40	7 (5%) 26 23	139, 242, 318, 321	0
9	I	119/122 (97%)	0.71	16 (13%) 3 4	139, 241, 318, 321	0
10	J	65/70 (92%)	-0.11	3 (4%) 32 27	68, 173, 284, 313	0
11	K	114/120 (95%)	-0.03	2 (1%) 68 59	98, 170, 245, 304	0
12	L	46/70 (65%)	0.19	3 (6%) 18 15	134, 233, 304, 321	0
All	All	3526/4504 (78%)	0.02	113 (3%) 47 37	20, 189, 294, 321	0

The worst 5 of 113 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	199	LEU	5.2
1	A	114	LEU	5.1
2	B	869	SER	4.7
2	B	715	ALA	4.6
5	E	88	VAL	4.6

### 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 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(Å <sup>2</sup> )	Q<0.9
13	ZN	A	1735	1/1	0.90	0.11	172,172,172,172	0
13	ZN	I	203	1/1	0.90	0.19	172,172,172,172	0
13	ZN	I	204	1/1	0.92	0.25	172,172,172,172	0
13	ZN	A	1734	1/1	0.93	0.17	172,172,172,172	0
13	ZN	C	319	1/1	0.94	0.06	172,172,172,172	0
13	ZN	J	101	1/1	0.94	0.13	172,172,172,172	0
13	ZN	B	1307	1/1	0.99	0.06	172,172,172,172	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.