



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 08:28 am BST

PDB ID : 5U0C  
Title : Structure of Zika virus NS5 RNA polymerase domain  
Authors : Zhao, B.; Du, F.  
Deposited on : 2016-11-23  
Resolution : 3.00 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

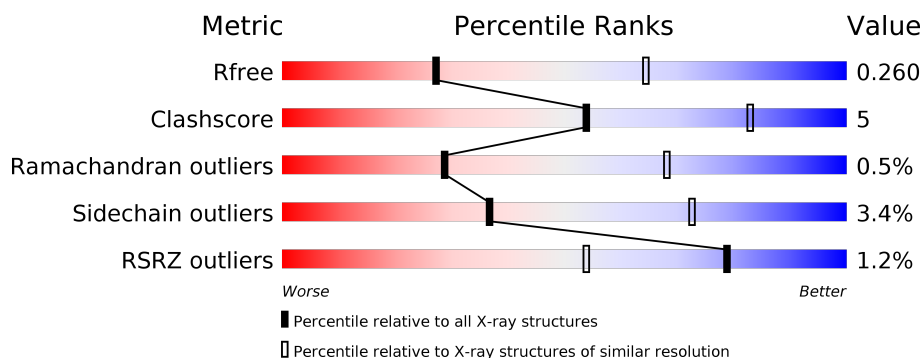
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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	639	<div> <div>%</div> <div> <div></div> <div>82%</div> <div>15%</div> <div>..</div> </div> </div>
1	B	639	<div> <div>%</div> <div> <div></div> <div>81%</div> <div>15%</div> <div>..</div> </div> </div>
1	C	639	<div> <div>%</div> <div> <div></div> <div>82%</div> <div>13%</div> <div>..</div> </div> </div>
1	D	639	<div> <div>%</div> <div> <div></div> <div>83%</div> <div>14%</div> <div>.</div> </div> </div>
1	E	639	<div> <div>2%</div> <div> <div></div> <div>80%</div> <div>16%</div> <div>..</div> </div> </div>
1	F	639	<div> <div>%</div> <div> <div></div> <div>78%</div> <div>18%</div> <div>..</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	639	<div><div><div></div><div></div><div></div></div><div><div>2%</div><div>87%</div><div>10%</div><div>••</div></div></div>
1	H	639	<div><div><div></div><div></div><div></div></div><div><div>2%</div><div>84%</div><div>12%</div><div>••</div></div></div>

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 40565 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 NS5 RNA polymerase domain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	624	Total	C	N	O	S	0	0	0
			5085	3203	924	925	33			
1	B	621	Total	C	N	O	S	0	0	0
			5058	3187	917	921	33			
1	C	621	Total	C	N	O	S	0	0	0
			5058	3187	917	921	33			
1	D	623	Total	C	N	O	S	0	0	0
			5070	3195	919	923	33			
1	E	622	Total	C	N	O	S	0	0	0
			5067	3192	919	923	33			
1	F	622	Total	C	N	O	S	0	0	0
			5068	3194	919	922	33			
1	G	624	Total	C	N	O	S	0	0	0
			5075	3198	920	924	33			
1	H	623	Total	C	N	O	S	0	0	0
			5068	3194	919	922	33			

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	G	2	Total	Zn	0	0
			2	2		
2	D	2	Total	Zn	0	0
			2	2		
2	E	2	Total	Zn	0	0
			2	2		
2	H	2	Total	Zn	0	0
			2	2		
2	B	2	Total	Zn	0	0
			2	2		
2	C	2	Total	Zn	0	0
			2	2		

*Continued on next page...*

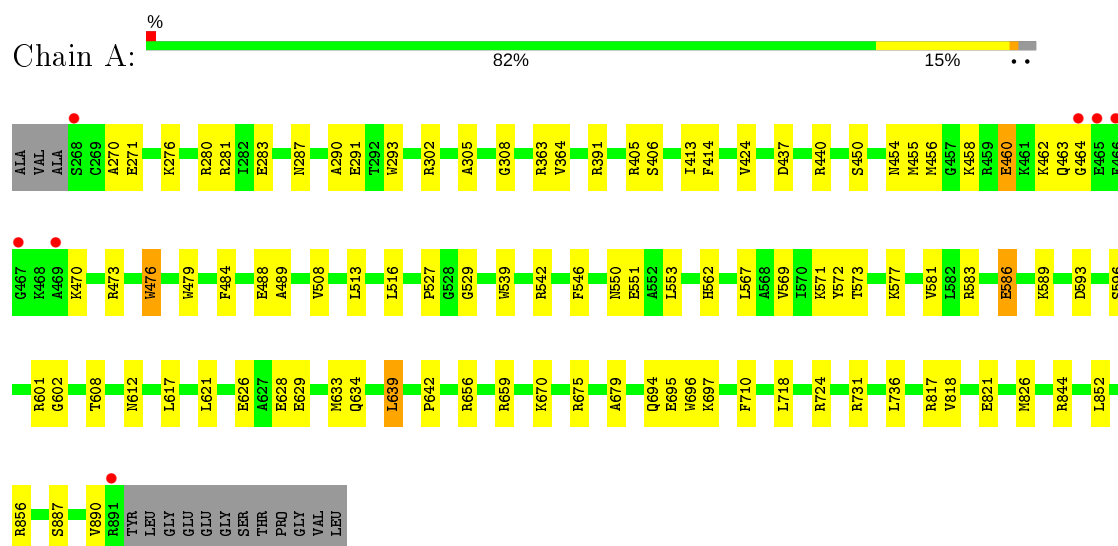
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	2	Total 2	Zn 2	0	0
2	F	2	Total 2	Zn 2	0	0

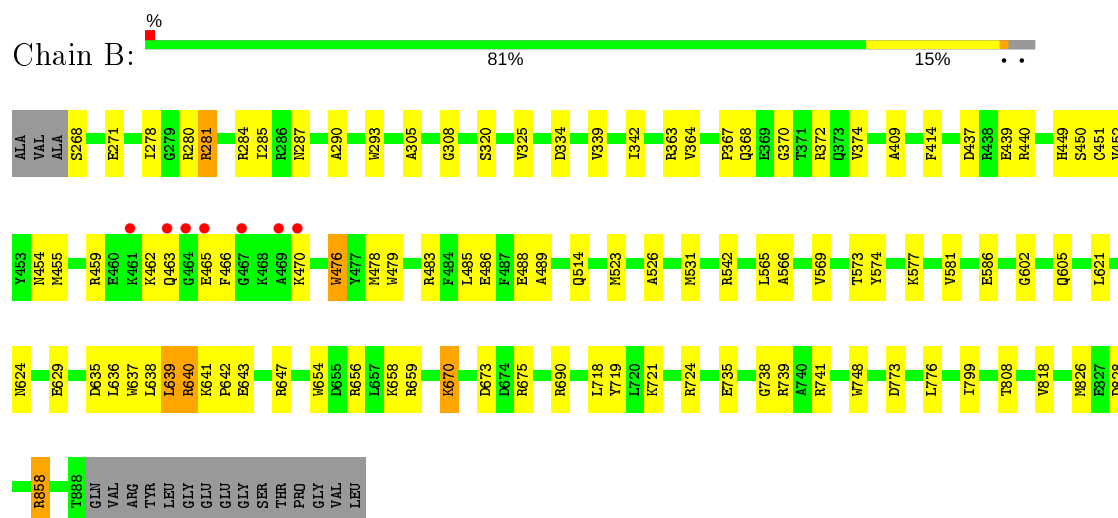
### 3 Residue-property plots

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.

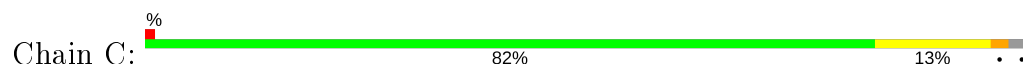
#### • Molecule 1: NS5 RNA polymerase domain

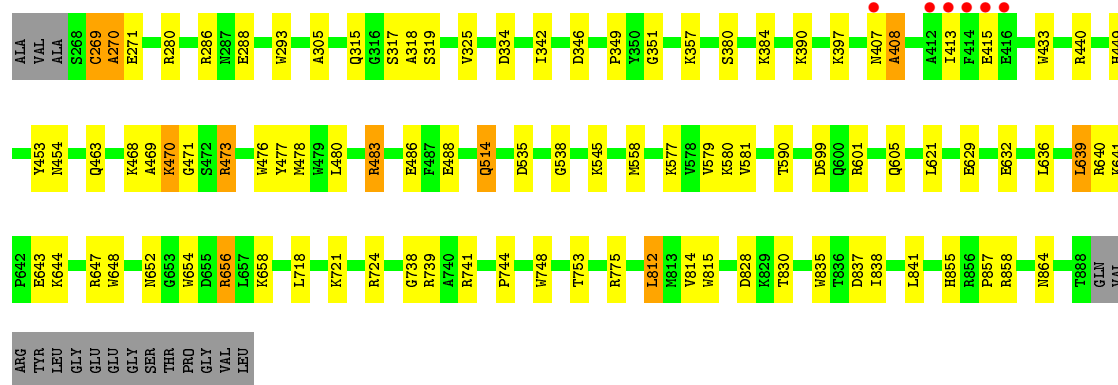


#### • Molecule 1: NS5 RNA polymerase domain

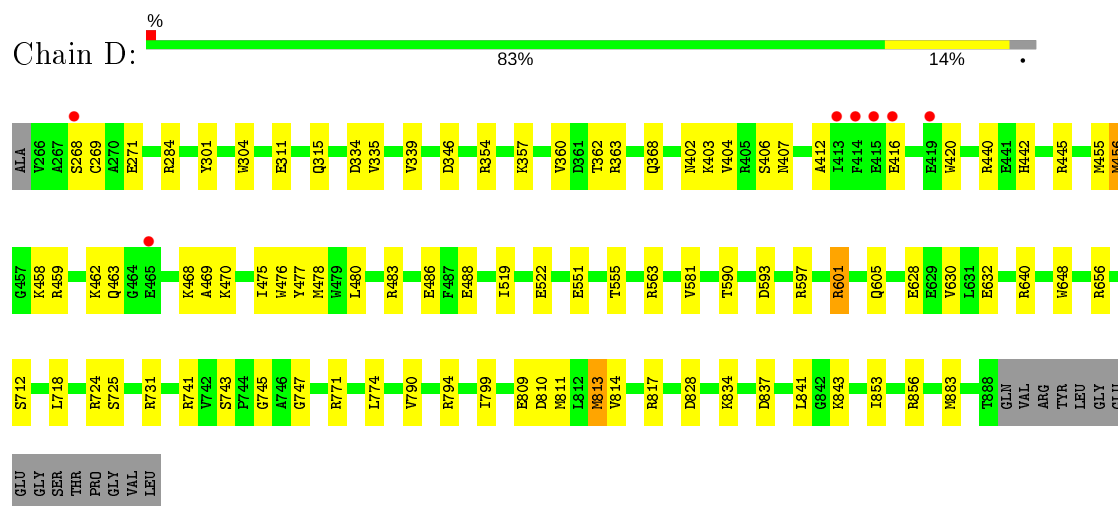


#### • Molecule 1: NS5 RNA polymerase domain

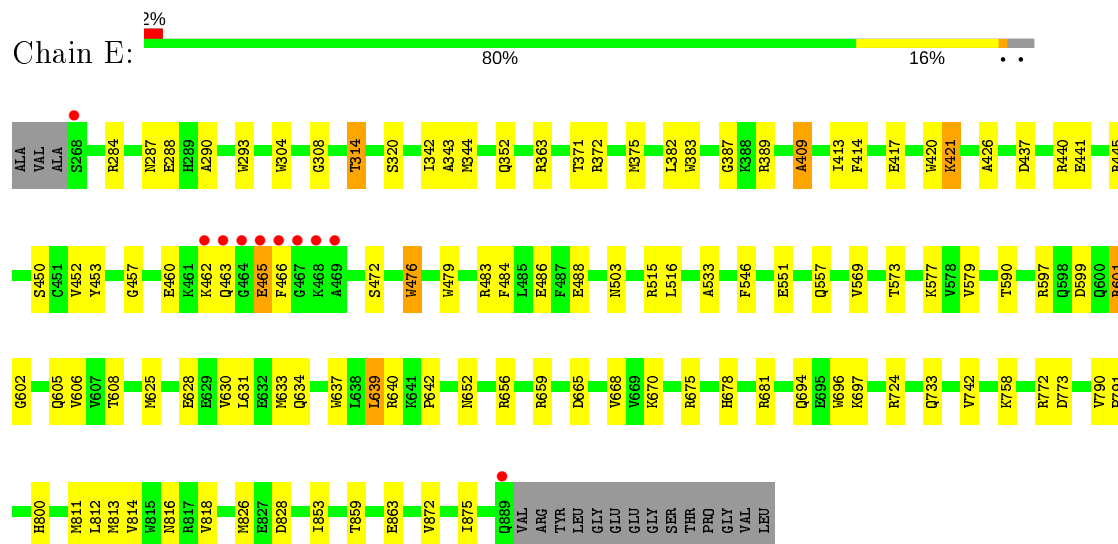




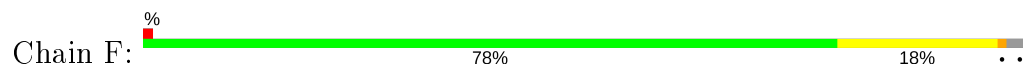
• Molecule 1: NS5 RNA polymerase domain

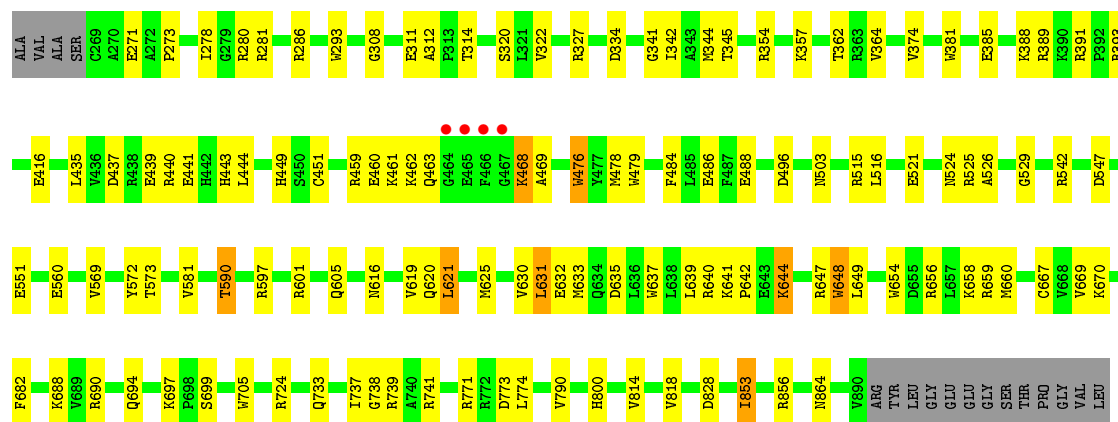


• Molecule 1: NS5 RNA polymerase domain

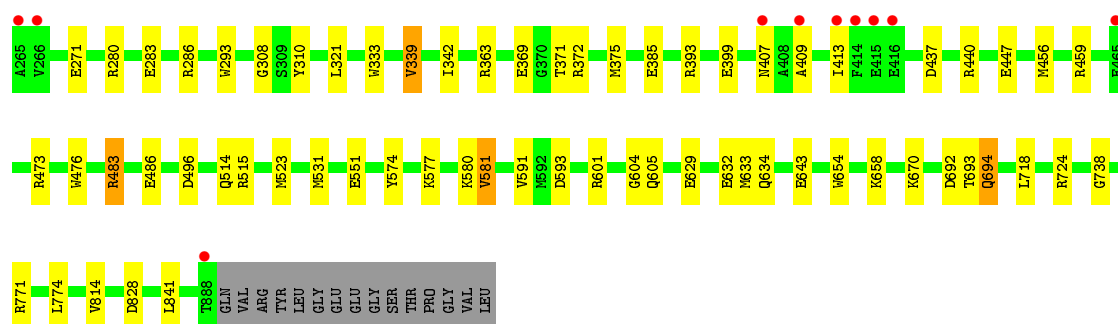
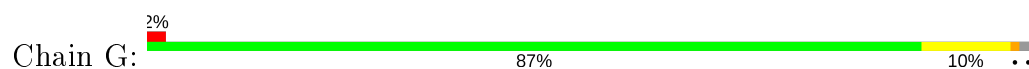


• Molecule 1: NS5 RNA polymerase domain

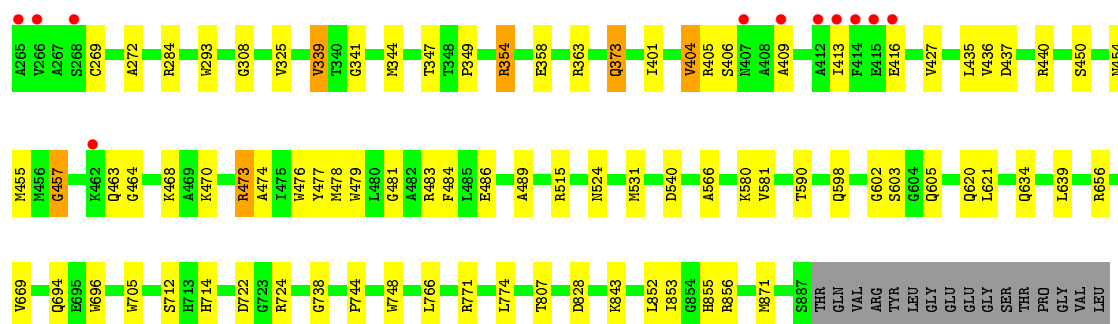
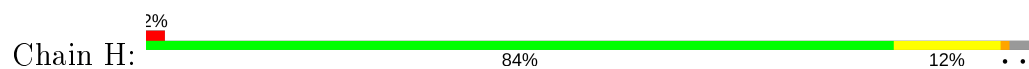




• Molecule 1: NS5 RNA polymerase domain



• Molecule 1: NS5 RNA polymerase domain





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	121.52Å 188.71Å 192.54Å 90.00° 91.99° 90.00°	Depositor
Resolution (Å)	89.14 – 3.00 96.21 – 3.00	Depositor EDS
% Data completeness (in resolution range)	97.8 (89.14-3.00) 89.5 (96.21-3.00)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.45 (at 3.01Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, $R_{free}$	0.226 , 0.260 0.226 , 0.260	Depositor DCC
$R_{free}$ test set	8552 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	40.3	Xtriage
Anisotropy	0.470	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 3.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.000 for -h,l,k 0.000 for -h,-l,-k 0.069 for h,-k,-l	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	40565	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 42.40 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.0502e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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	0.23	0/5209	0.41	0/7044
1	B	0.23	0/5182	0.42	0/7008
1	C	0.23	0/5182	0.41	0/7008
1	D	0.24	0/5194	0.42	0/7025
1	E	0.23	0/5191	0.42	0/7020
1	F	0.23	0/5192	0.43	0/7022
1	G	0.23	0/5199	0.41	0/7032
1	H	0.23	0/5192	0.41	0/7022
All	All	0.23	0/41541	0.42	0/56181

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 [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	5085	0	4983	63	0
1	B	5058	0	4953	57	0
1	C	5058	0	4953	53	1
1	D	5070	0	4967	44	0
1	E	5067	0	4961	64	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	F	5068	0	4965	76	0
1	G	5075	0	4972	35	0
1	H	5068	0	4965	43	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
2	C	2	0	0	0	0
2	D	2	0	0	0	0
2	E	2	0	0	0	0
2	F	2	0	0	0	0
2	G	2	0	0	0	0
2	H	2	0	0	0	0
All	All	40565	0	39719	423	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 5.

The worst 5 of 423 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)
1:D:745:GLY:HA3	1:D:747:GLY:H	1.29	0.95
1:B:773:ASP:OD1	1:B:858:ARG:NH1	2.06	0.88
1:A:529:GLY:O	1:A:670:LYS:NZ	2.06	0.88
1:A:628:GLU:OE2	1:A:656:ARG:NH1	2.06	0.88
1:D:632:GLU:OE1	1:D:640:ARG:NH1	2.13	0.81

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:C:280:ARG:NH1	1:E:288:GLU:O[2_656]	2.14	0.06

## 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	622/639 (97%)	595 (96%)	26 (4%)	1 (0%)	47	82
1	B	619/639 (97%)	588 (95%)	28 (4%)	3 (0%)	29	68
1	C	619/639 (97%)	584 (94%)	30 (5%)	5 (1%)	19	57
1	D	621/639 (97%)	575 (93%)	44 (7%)	2 (0%)	41	76
1	E	620/639 (97%)	587 (95%)	29 (5%)	4 (1%)	25	64
1	F	620/639 (97%)	581 (94%)	34 (6%)	5 (1%)	19	57
1	G	622/639 (97%)	587 (94%)	34 (6%)	1 (0%)	47	82
1	H	621/639 (97%)	587 (94%)	32 (5%)	2 (0%)	41	76
All	All	4964/5112 (97%)	4684 (94%)	257 (5%)	23 (0%)	29	68

5 of 23 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	640	ARG
1	F	271	GLU
1	G	633	MET
1	A	508	VAL
1	C	577	LYS

### 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	539/549 (98%)	525 (97%)	14 (3%)	46	78
1	B	536/549 (98%)	512 (96%)	24 (4%)	27	64
1	C	536/549 (98%)	515 (96%)	21 (4%)	32	69
1	D	537/549 (98%)	519 (97%)	18 (3%)	37	72
1	E	537/549 (98%)	521 (97%)	16 (3%)	41	75
1	F	537/549 (98%)	519 (97%)	18 (3%)	37	72
1	G	537/549 (98%)	522 (97%)	15 (3%)	43	77

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	H	536/549 (98%)	518 (97%)	18 (3%)	37	72
All	All	4295/4392 (98%)	4151 (97%)	144 (3%)	37	72

5 of 144 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	420	TRP
1	E	382	LEU
1	H	621	LEU
1	D	480	LEU
1	D	799	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	442	HIS
1	E	557	GLN
1	H	373	GLN
1	D	368	GLN
1	H	524	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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 16 ligands modelled in this entry, 16 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	624/639 (97%)	-0.22	7 (1%) 80 56	29, 59, 86, 109	0
1	B	621/639 (97%)	-0.31	7 (1%) 80 56	22, 45, 69, 110	0
1	C	621/639 (97%)	-0.30	6 (0%) 82 59	26, 46, 85, 107	0
1	D	623/639 (97%)	-0.34	7 (1%) 80 56	26, 39, 68, 103	0
1	E	622/639 (97%)	-0.20	10 (1%) 72 44	28, 59, 90, 130	0
1	F	622/639 (97%)	-0.28	4 (0%) 89 72	27, 53, 80, 113	0
1	G	624/639 (97%)	-0.31	10 (1%) 72 44	31, 44, 80, 99	0
1	H	623/639 (97%)	-0.32	11 (1%) 68 40	24, 37, 70, 94	0
All	All	4980/5112 (97%)	-0.29	62 (1%) 79 54	22, 45, 81, 130	0

The worst 5 of 62 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	416	GLU	4.6
1	G	414	PHE	4.5
1	D	415	GLU	4.4
1	E	464	GLY	4.4
1	E	466	PHE	4.1

### 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(Å <sup>2</sup> )	Q<0.9
2	ZN	B	1001	1/1	0.99	0.21	54,54,54,54	0
2	ZN	A	1001	1/1	0.99	0.16	58,58,58,58	0
2	ZN	C	1001	1/1	0.99	0.20	69,69,69,69	0
2	ZN	E	1002	1/1	0.99	0.20	52,52,52,52	0
2	ZN	A	1002	1/1	0.99	0.21	42,42,42,42	0
2	ZN	G	1002	1/1	0.99	0.20	44,44,44,44	0
2	ZN	H	1001	1/1	0.99	0.19	49,49,49,49	0
2	ZN	C	1002	1/1	0.99	0.21	42,42,42,42	0
2	ZN	D	1002	1/1	0.99	0.21	44,44,44,44	0
2	ZN	B	1002	1/1	0.99	0.21	47,47,47,47	0
2	ZN	E	1001	1/1	0.99	0.17	65,65,65,65	0
2	ZN	F	1002	1/1	1.00	0.19	46,46,46,46	0
2	ZN	G	1001	1/1	1.00	0.15	54,54,54,54	0
2	ZN	D	1001	1/1	1.00	0.20	48,48,48,48	0
2	ZN	F	1001	1/1	1.00	0.18	46,46,46,46	0
2	ZN	H	1002	1/1	1.00	0.21	45,45,45,45	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.