



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 09:13 am BST

PDB ID : 5YOX  
Title : HD domain-containing protein YGK1(YGL101W)  
Authors : Yang, J.; Wang, F.; Gao, Z.; Zhou, K.; Liu, Q.  
Deposited on : 2017-10-31  
Resolution : 2.61 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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
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

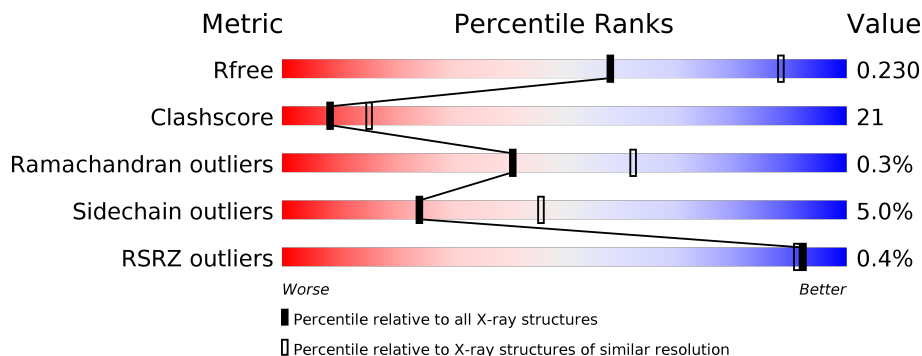
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.61 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.




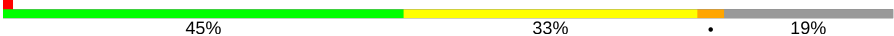
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	3797 (2.64-2.60)
Clashscore	141614	4168 (2.64-2.60)
Ramachandran outliers	138981	4093 (2.64-2.60)
Sidechain outliers	138945	4093 (2.64-2.60)
RSRZ outliers	127900	3731 (2.64-2.60)

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	215	
1	B	215	
1	C	215	
1	D	215	
1	E	215	
1	F	215	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	G	215	 42% 36% 20%
1	H	215	 45% 33% 19%

## 2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 11919 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 HD domain-containing protein YGL101W.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	183	1520	968	257	286	5	4	0	0	0
1	B	174	1458	930	246	273	5	4	0	0	0
1	C	175	1463	933	246	275	5	4	0	0	0
1	D	176	1475	939	250	277	5	4	0	0	0
1	E	176	1475	939	250	277	5	4	0	0	0
1	F	174	1451	923	246	273	5	4	0	0	0
1	G	173	1449	925	243	272	5	4	0	0	0
1	H	175	1463	933	246	275	5	4	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	initiating methionine	UNP P53144
B	1	MSE	-	initiating methionine	UNP P53144
C	1	MSE	-	initiating methionine	UNP P53144
D	1	MSE	-	initiating methionine	UNP P53144
E	1	MSE	-	initiating methionine	UNP P53144
F	1	MSE	-	initiating methionine	UNP P53144
G	1	MSE	-	initiating methionine	UNP P53144
H	1	MSE	-	initiating methionine	UNP P53144

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	G	1	Total Zn 1 1	0	0
2	D	1	Total Zn 1 1	0	0
2	E	1	Total Zn 1 1	0	0
2	H	1	Total Zn 1 1	0	0
2	B	1	Total Zn 1 1	0	0
2	C	1	Total Zn 1 1	0	0
2	A	1	Total Zn 1 1	0	0
2	F	1	Total Zn 1 1	0	0

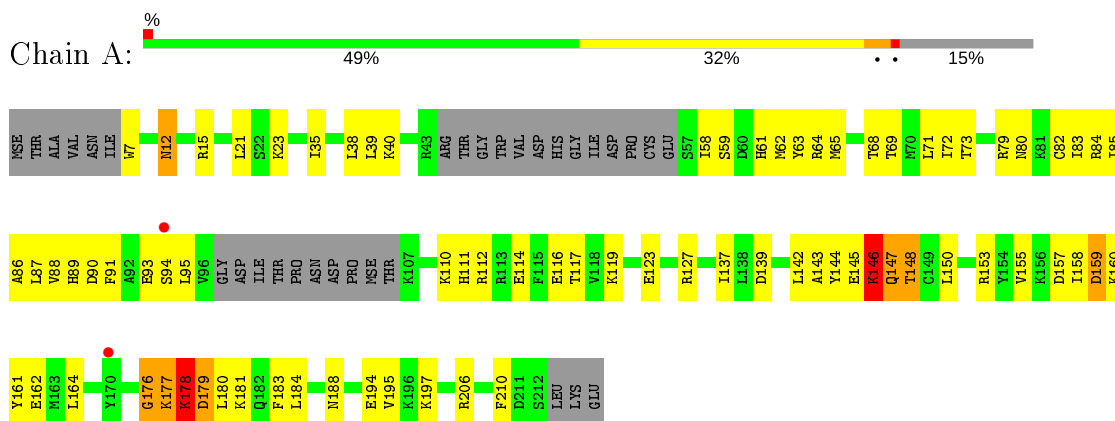
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	15	Total O 15 15	0	0
3	B	18	Total O 18 18	0	0
3	C	22	Total O 22 22	0	0
3	D	27	Total O 27 27	0	0
3	E	24	Total O 24 24	0	0
3	F	21	Total O 21 21	0	0
3	G	20	Total O 20 20	0	0
3	H	10	Total O 10 10	0	0

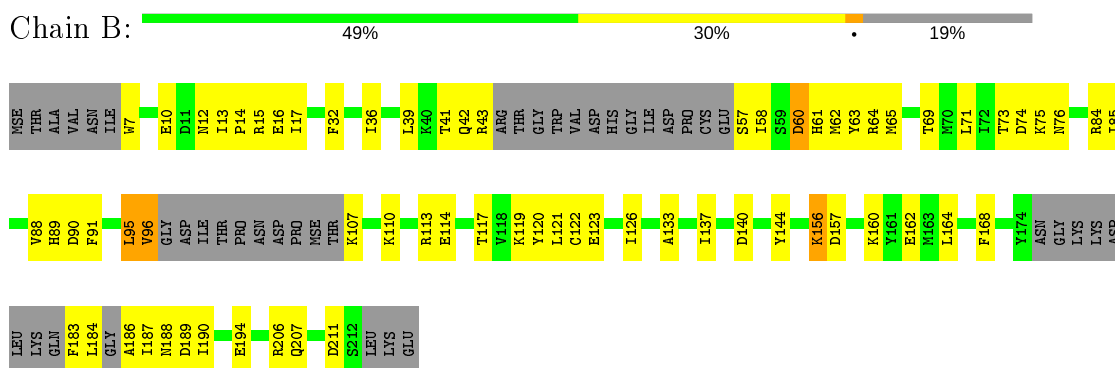
### 3 Residue-property plots [i](#)

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.

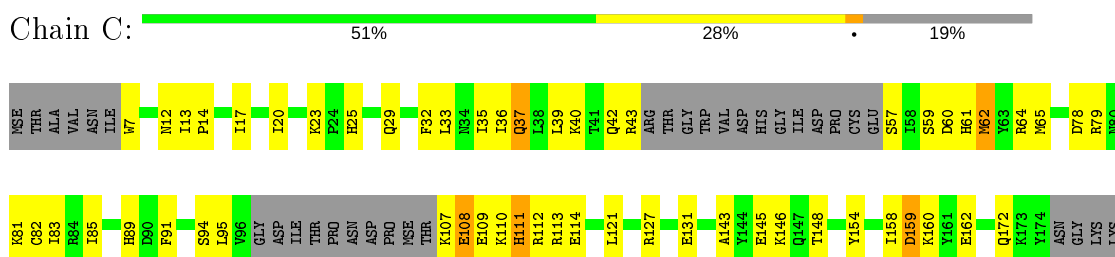
- Molecule 1: HD domain-containing protein YGL101W



- Molecule 1: HD domain-containing protein YGL101W



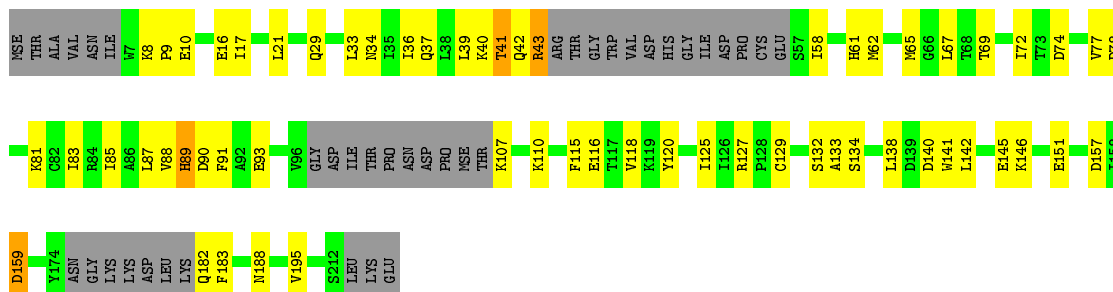
- Molecule 1: HD domain-containing protein YGL101W





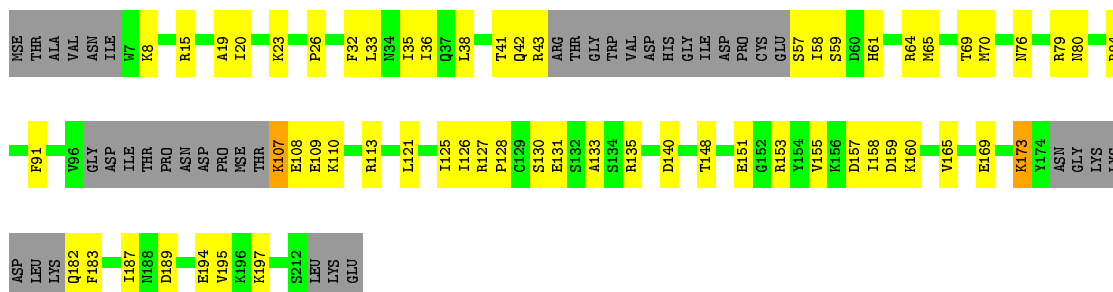
- Molecule 1: HD domain-containing protein YGL101W

Chain D: 54% 26% 18%



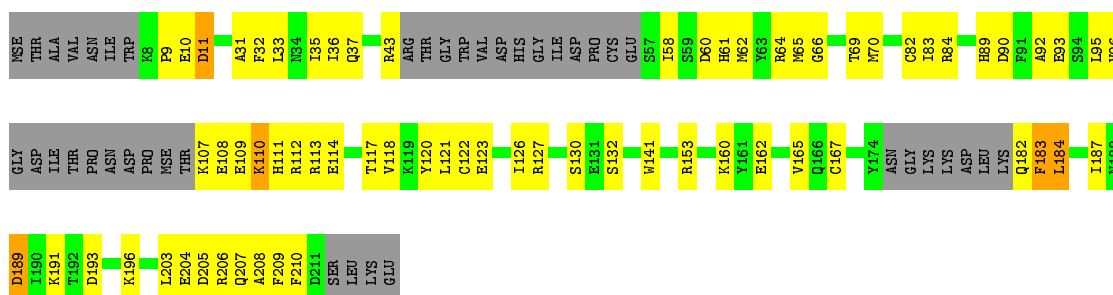
- Molecule 1: HD domain-containing protein YGL101W

Chain E: 54% 27% 18%



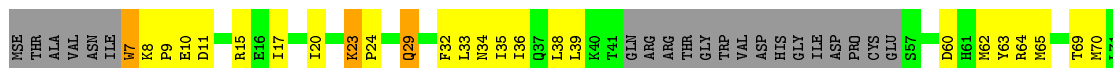
- Molecule 1: HD domain-containing protein YGL101W

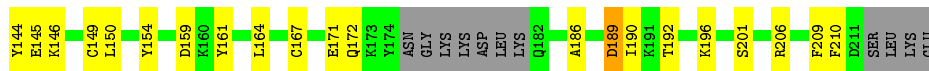
Chain F: 49% 29% 19%



- Molecule 1: HD domain-containing protein YGL101W

Chain G: 42% 36% 20%





• Molecule 1: HD domain-containing protein YGL101W





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	52.63Å 142.18Å 112.27Å 90.00° 89.99° 90.00°	Depositor
Resolution (Å)	24.77 – 2.61 24.77 – 2.61	Depositor EDS
% Data completeness (in resolution range)	98.2 (24.77-2.61) 98.1 (24.77-2.61)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.54 (at 2.60Å)	Xtriage
Refinement program	PHENIX 1.12_2829	Depositor
R, $R_{free}$	0.207 , 0.245 0.200 , 0.230	Depositor DCC
$R_{free}$ test set	2510 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	49.8	Xtriage
Anisotropy	0.329	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 34.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.478 for h,-k,-l	Xtriage
Reported twinning fraction	0.500 for h,-k,-l	Depositor
Outliers	0 of 49325 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	11919	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.14% 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	0.57	0/1543	0.63	3/2067 (0.1%)
1	B	0.46	0/1479	0.51	0/1980
1	C	0.53	0/1485	0.49	0/1990
1	D	0.32	0/1497	0.48	0/2005
1	E	0.28	0/1497	0.42	0/2005
1	F	0.54	0/1471	0.54	0/1969
1	G	0.56	0/1471	0.55	0/1971
1	H	0.39	0/1485	0.51	0/1990
All	All	0.47	0/11928	0.52	3/15977 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	146	LYS	N-CA-C	-7.69	90.25	111.00
1	A	179	ASP	CB-CG-OD2	5.23	123.01	118.30
1	A	178	LYS	C-N-CA	5.16	134.59	121.70

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	1520	0	1513	82	0
1	B	1458	0	1448	60	0
1	C	1463	0	1450	59	0
1	D	1475	0	1466	65	0
1	E	1475	0	1466	42	0
1	F	1451	0	1445	61	0
1	G	1449	0	1440	84	0
1	H	1463	0	1450	88	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	F	1	0	0	0	0
2	G	1	0	0	0	0
2	H	1	0	0	0	0
3	A	15	0	0	1	0
3	B	18	0	0	1	0
3	C	22	0	0	1	0
3	D	27	0	0	0	0
3	E	24	0	0	2	0
3	F	21	0	0	1	0
3	G	20	0	0	3	0
3	H	10	0	0	1	0
All	All	11919	0	11678	487	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 21.

The worst 5 of 487 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:G:8:LYS:HD2	1:G:11:ASP:OD1	1.27	1.25
1:A:63:TYR:CE2	1:G:39:LEU:HD13	1.79	1.17
1:B:168:PHE:HB2	1:B:183:PHE:HZ	1.15	1.11
1:D:67:LEU:HD23	1:H:36:ILE:HD11	1.29	1.08
1:H:96:VAL:HG11	1:H:110:LYS:HE3	1.35	1.04

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	177/215 (82%)	155 (88%)	21 (12%)	1 (1%)	25	45
1	B	164/215 (76%)	154 (94%)	9 (6%)	1 (1%)	25	45
1	C	167/215 (78%)	151 (90%)	16 (10%)	0	100	100
1	D	168/215 (78%)	151 (90%)	16 (10%)	1 (1%)	25	45
1	E	168/215 (78%)	158 (94%)	10 (6%)	0	100	100
1	F	166/215 (77%)	146 (88%)	20 (12%)	0	100	100
1	G	165/215 (77%)	150 (91%)	15 (9%)	0	100	100
1	H	167/215 (78%)	146 (87%)	20 (12%)	1 (1%)	25	45
All	All	1342/1720 (78%)	1211 (90%)	127 (10%)	4 (0%)	41	62

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	187	ILE
1	D	41	THR
1	H	183	PHE
1	A	176	GLY

### 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	169/193 (88%)	158 (94%)	11 (6%)	17	33

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	163/193 (84%)	155 (95%)	8 (5%)	25	46
1	C	163/193 (84%)	152 (93%)	11 (7%)	16	31
1	D	165/193 (86%)	161 (98%)	4 (2%)	49	72
1	E	165/193 (86%)	159 (96%)	6 (4%)	35	59
1	F	162/193 (84%)	152 (94%)	10 (6%)	18	35
1	G	162/193 (84%)	153 (94%)	9 (6%)	21	40
1	H	163/193 (84%)	156 (96%)	7 (4%)	29	53
All	All	1312/1544 (85%)	1246 (95%)	66 (5%)	24	46

5 of 66 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	205	ASP
1	E	107	LYS
1	H	82	CYS
1	D	43	ARG
1	D	188	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	89	HIS
1	H	147	GLN
1	G	29	GLN
1	D	42	GLN
1	G	182	GLN

### 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 8 ligands modelled in this entry, 8 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	179/215 (83%)	-0.50	2 (1%) 80 78	27, 50, 67, 80	0
1	B	170/215 (79%)	-0.54	0 100 100	31, 45, 63, 75	0
1	C	171/215 (79%)	-0.53	0 100 100	27, 46, 66, 82	0
1	D	172/215 (80%)	-0.55	0 100 100	37, 50, 66, 73	0
1	E	172/215 (80%)	-0.56	0 100 100	29, 46, 62, 72	0
1	F	170/215 (79%)	-0.59	0 100 100	31, 44, 61, 70	0
1	G	169/215 (78%)	-0.40	1 (0%) 89 88	33, 58, 70, 79	0
1	H	171/215 (79%)	-0.40	2 (1%) 79 76	41, 60, 78, 91	0
All	All	1374/1720 (79%)	-0.51	5 (0%) 92 91	27, 50, 69, 91	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	94	SER	2.9
1	A	170	TYR	2.3
1	H	184	LEU	2.2
1	A	94	SER	2.1
1	G	77	VAL	2.0

### 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 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	300	1/1	0.48	0.20	59,59,59,59	1
2	ZN	E	300	1/1	0.74	0.19	40,40,40,40	1
2	ZN	A	300	1/1	0.80	0.13	42,42,42,42	1
2	ZN	H	300	1/1	0.81	0.14	64,64,64,64	1
2	ZN	D	300	1/1	0.87	0.12	57,57,57,57	1
2	ZN	G	300	1/1	0.89	0.12	59,59,59,59	1
2	ZN	F	300	1/1	0.94	0.12	45,45,45,45	1
2	ZN	C	300	1/1	0.94	0.14	34,34,34,34	1

## 6.5 Other polymers [i](#)

There are no such residues in this entry.