



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 04:52 PM GMT

PDB ID : 4GDF  
Title : A Crystal Structure of SV40 Large T Antigen  
Authors : Chang, Y.P.; Xu, M.; Chen, X.S.  
Deposited on : 2012-07-31  
Resolution : 2.80 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  
<http://wwpdb.org/validation/2016/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 : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

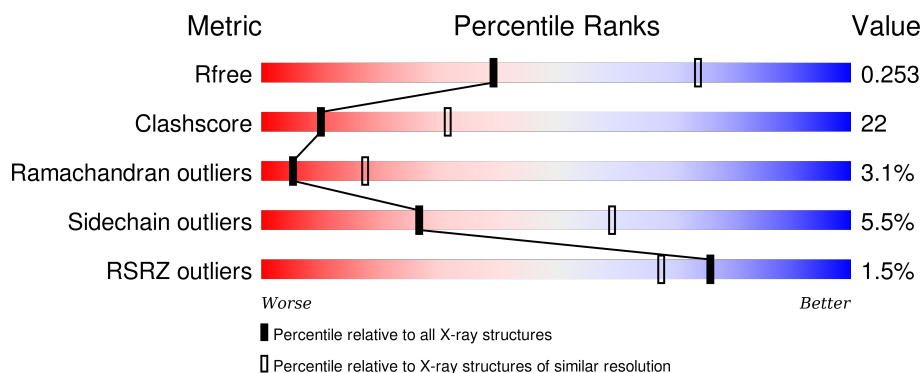
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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.




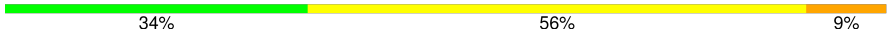

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	91344	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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. 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	497	<div> <div> <div>0%</div> <div>58%</div> <div>37%</div> <div>5%</div> </div> </div>
1	B	497	<div> <div> <div>0%</div> <div>60%</div> <div>36%</div> <div>•</div> </div> </div>
1	E	497	<div> <div> <div>0%</div> <div>58%</div> <div>37%</div> <div>•</div> </div> </div>
1	F	497	<div> <div> <div>2%</div> <div>59%</div> <div>38%</div> <div>•</div> </div> </div>
2	C	32	<div> <div> <div>3%</div> <div>34%</div> <div>53%</div> <div>13%</div> </div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	G	32	 3% 34% 53% 13%
3	D	32	 34% 56% 9%
3	H	32	 34% 56% 9%

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 18835 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 Large T antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	497	Total	C	N	O	S	0	0	0
			4035	2601	675	733	26			
1	B	497	Total	C	N	O	S	0	0	0
			4035	2601	675	733	26			
1	E	497	Total	C	N	O	S	0	0	0
			4035	2601	675	733	26			
1	F	497	Total	C	N	O	S	0	0	0
			4035	2601	675	733	26			

- Molecule 2 is a DNA chain called DNA (32-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	32	Total	C	N	O	P	0	0	0
			657	312	126	188	31			
2	G	32	Total	C	N	O	P	0	0	0
			657	312	126	188	31			

- Molecule 3 is a DNA chain called DNA (32-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	32	Total	C	N	O	P	0	0	0
			649	310	116	192	31			
3	H	32	Total	C	N	O	P	0	0	0
			649	310	116	192	31			

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	F	1	Total 1	Zn 1	0	0
4	E	1	Total 1	Zn 1	0	0

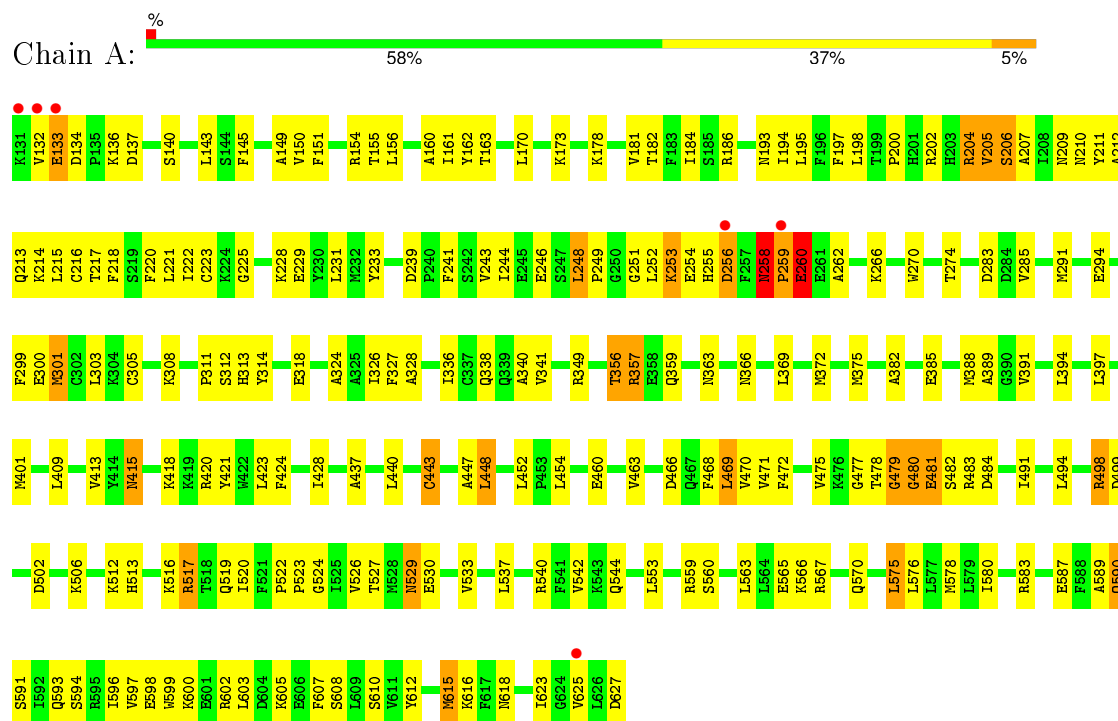
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	18	Total 18	O 18	0	0
5	B	12	Total 12	O 12	0	0
5	C	9	Total 9	O 9	0	0
5	D	6	Total 6	O 6	0	0
5	E	14	Total 14	O 14	0	0
5	F	7	Total 7	O 7	0	0
5	G	6	Total 6	O 6	0	0
5	H	7	Total 7	O 7	0	0

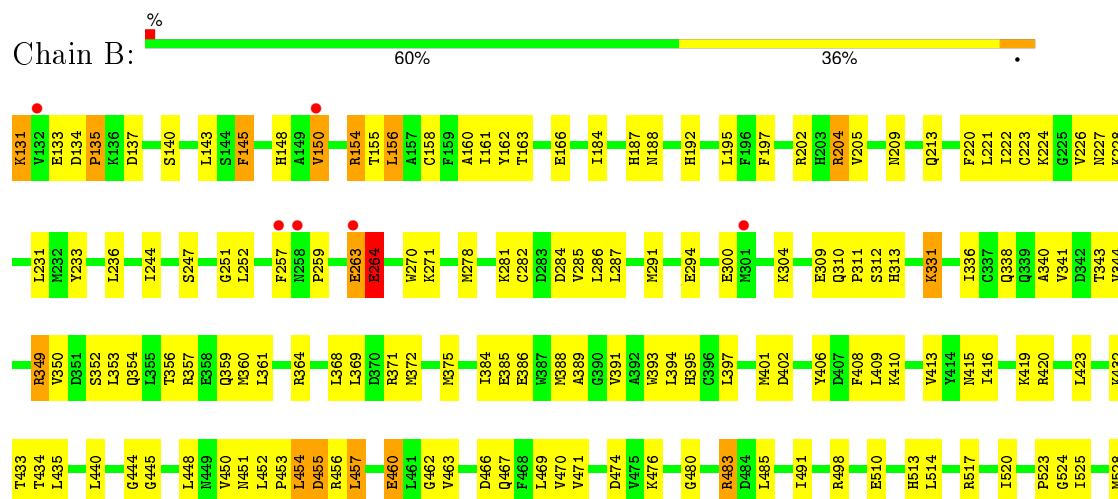
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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 ( $\text{RSRZ} > 2$ ). 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: Large T antigen

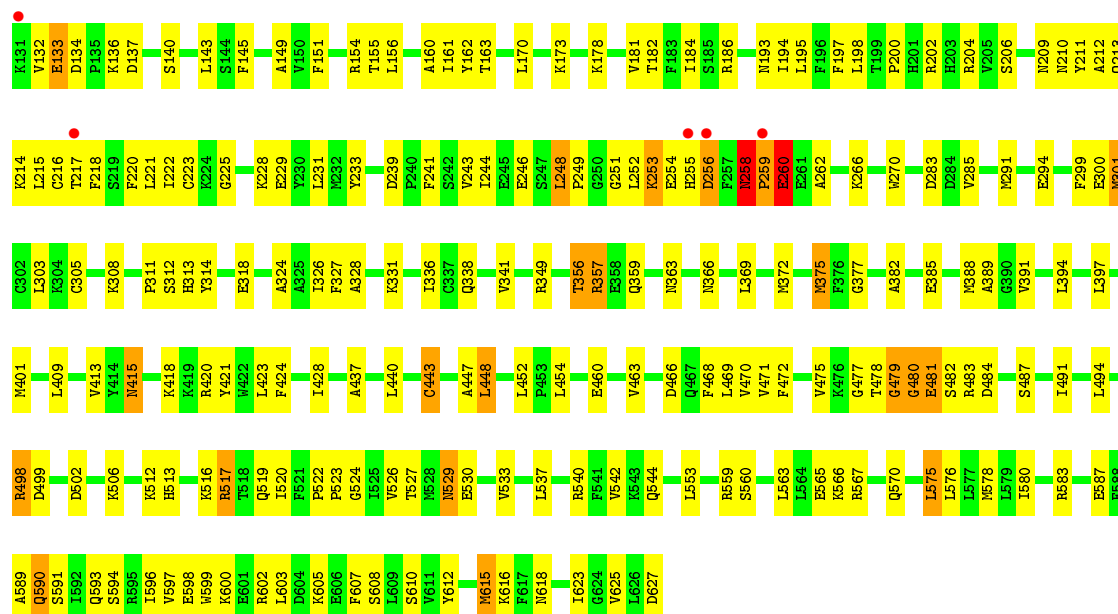


#### • Molecule 1: Large T antigen

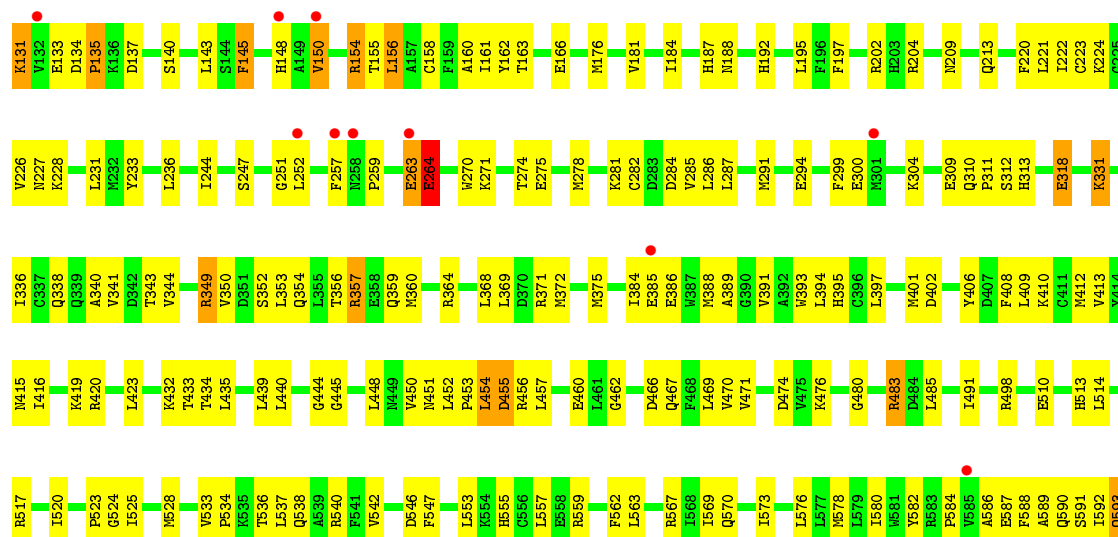




• Molecule 1: Large T antigen



• Molecule 1: Large T antigen





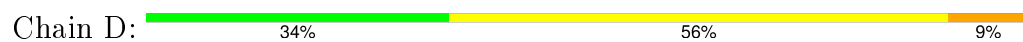
- Molecule 2: DNA (32-MER)



- Molecule 2: DNA (32-MER)



- Molecule 3: DNA (32-MER)



- Molecule 3: DNA (32-MER)





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	73.63Å 128.31Å 166.12Å 90.00° 89.91° 90.00°	Depositor
Resolution (Å)	50.00 – 2.80 48.37 – 2.81	Depositor EDS
% Data completeness (in resolution range)	(Not available) (50.00-2.80) 86.5 (48.37-2.81)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.73 (at 2.81Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.227 , 0.255 0.224 , 0.253	Depositor DCC
$R_{free}$ test set	3317 reflections (5.08%)	DCC
Wilson B-factor (Å <sup>2</sup> )	51.4	Xtriage
Anisotropy	0.376	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 26.3	EDS
Estimated twinning fraction	0.460 for h,-k,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 70152 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	18835	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.16% 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, 0.375 respectively for untwinned datasets, and 0.333, 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

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	0.47	0/4124	0.63	3/5562 (0.1%)
1	B	0.45	0/4124	0.62	0/5562
1	E	0.47	0/4124	0.63	3/5562 (0.1%)
1	F	0.45	0/4124	0.62	0/5562
2	C	0.54	0/738	1.00	5/1138 (0.4%)
2	G	0.55	0/738	1.02	5/1138 (0.4%)
3	D	0.54	0/726	0.95	1/1118 (0.1%)
3	H	0.55	0/726	0.96	1/1118 (0.1%)
All	All	0.48	0/19424	0.70	18/26760 (0.1%)

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
1	F	0	1
2	C	0	2
2	G	0	2
3	D	0	3
3	H	0	3
All	All	0	12

There are no bond length outliers.

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	G	1	DA	O4'-C1'-N9	-12.03	99.58	108.00
2	C	3	DT	N1-C1'-C2'	10.90	133.31	112.60

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	G	3	DT	N1-C1'-C2'	10.71	132.94	112.60
2	C	1	DA	O4'-C1'-N9	-10.46	100.68	108.00
3	D	31	DG	N9-C1'-C2'	8.13	128.06	112.60

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	309	GLU	Mainchain
2	C	16	DG	Sidechain
2	C	17	DC	Sidechain
3	D	5	DT	Sidechain
3	D	8	DG	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	4035	0	4068	179	0
1	B	4035	0	4068	180	0
1	E	4035	0	4068	177	0
1	F	4035	0	4068	182	0
2	C	657	0	360	31	0
2	G	657	0	360	29	0
3	D	649	0	362	26	0
3	H	649	0	362	26	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	E	1	0	0	0	0
4	F	1	0	0	0	0
5	A	18	0	0	0	0
5	B	12	0	0	1	0
5	C	9	0	0	0	0
5	D	6	0	0	0	0
5	E	14	0	0	1	0
5	F	7	0	0	0	0
5	G	6	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	H	7	0	0	0	0
All	All	18835	0	17716	798	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 22.

The worst 5 of 798 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:G:1:DA:H2'	2:G:2:DC:C5	1.63	1.31
2:C:1:DA:H2'	2:C:2:DC:C5	1.64	1.30
2:G:1:DA:C2'	2:G:2:DC:C5	2.21	1.23
2:C:1:DA:C2'	2:C:2:DC:C5	2.21	1.22
1:F:498:ARG:HD2	1:F:540:ARG:HH21	1.11	1.09

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	495/497 (100%)	423 (86%)	52 (10%)	20 (4%)	4	12
1	B	495/497 (100%)	416 (84%)	68 (14%)	11 (2%)	8	28
1	E	495/497 (100%)	421 (85%)	55 (11%)	19 (4%)	4	13
1	F	495/497 (100%)	417 (84%)	67 (14%)	11 (2%)	8	28
All	All	1980/1988 (100%)	1677 (85%)	242 (12%)	61 (3%)	5	17

5 of 61 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	258	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	357	ARG
1	A	479	GLY
1	A	480	GLY
1	A	481	GLU

### 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	446/446 (100%)	422 (95%)	24 (5%)	27	60
1	B	446/446 (100%)	420 (94%)	26 (6%)	25	57
1	E	446/446 (100%)	423 (95%)	23 (5%)	29	62
1	F	446/446 (100%)	421 (94%)	25 (6%)	26	59
All	All	1784/1784 (100%)	1686 (94%)	98 (6%)	27	59

5 of 98 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	546	ASP
1	E	260	GLU
1	F	517	ARG
1	B	593	GLN
1	B	618	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 46 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	513	HIS
1	E	415	ASN
1	F	492	ASN
1	B	538	GLN
1	E	192	HIS

### 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 4 ligands modelled in this entry, 4 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	497/497 (100%)	-0.18	6 (1%) 81 73	34, 56, 97, 138	0
1	B	497/497 (100%)	-0.05	7 (1%) 78 69	29, 59, 107, 126	0
1	E	497/497 (100%)	-0.14	5 (1%) 84 77	33, 56, 96, 138	0
1	F	497/497 (100%)	-0.06	11 (2%) 65 54	29, 59, 107, 126	0
2	C	32/32 (100%)	-0.26	1 (3%) 52 40	29, 55, 160, 182	0
2	G	32/32 (100%)	-0.42	1 (3%) 52 40	30, 55, 160, 182	0
3	D	32/32 (100%)	-0.37	0 100 100	30, 48, 107, 139	0
3	H	32/32 (100%)	-0.48	0 100 100	32, 48, 107, 140	0
All	All	2116/2116 (100%)	-0.12	31 (1%) 76 68	29, 58, 106, 182	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	259	PRO	5.0
1	B	257	PHE	4.8
2	C	1	DA	4.6
1	B	263	GLU	4.5
1	B	301	MET	4.0

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
4	ZN	A	700	1/1	0.99	0.16	-0.64	58,58,58,58	0
4	ZN	E	700	1/1	0.97	0.14	-0.81	58,58,58,58	0
4	ZN	B	700	1/1	0.95	0.13	-0.85	68,68,68,68	0
4	ZN	F	700	1/1	0.94	0.14	-	69,69,69,69	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.