



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 9, 2018 – 10:34 am GMT

PDB ID : 3KLH  
Title : Crystal structure of AZT-Resistant HIV-1 Reverse Transcriptase crosslinked to post-translocation AZTMP-Terminated DNA (COMPLEX P)  
Authors : Tu, X.; Sarafianos, S.G.; Arnold, E.  
Deposited on : 2009-11-07  
Resolution : 2.90 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	:	trunk30967
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk30967

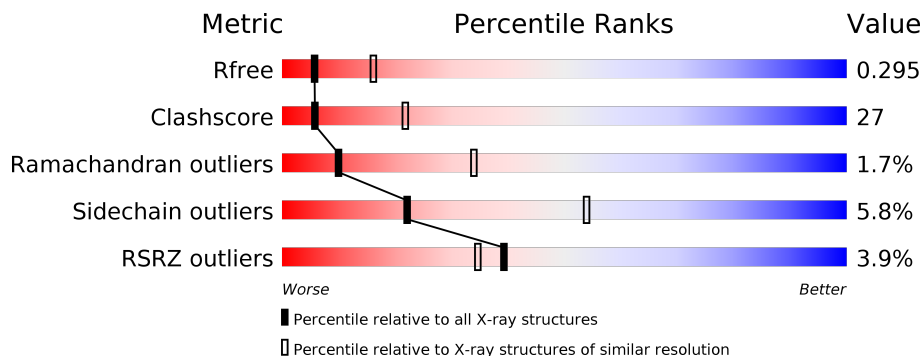
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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))
$R_{free}$	111664	1716 (2.90-2.90)
Clashscore	122126	1924 (2.90-2.90)
Ramachandran outliers	120053	1884 (2.90-2.90)
Sidechain outliers	120020	1886 (2.90-2.90)
RSRZ outliers	108989	1669 (2.90-2.90)

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	564	<div> <div>5%</div> <div> <div></div> <div>55%</div> <div>37%</div> <div>7%</div> </div> </div>
2	B	437	<div> <div>3%</div> <div> <div></div> <div>53%</div> <div>40%</div> <div>5%</div> </div> </div>
3	C	211	<div> <div></div> <div> <div></div> <div>61%</div> <div>36%</div> </div> </div>
4	D	225	<div> <div>4%</div> <div> <div></div> <div>60%</div> <div>38%</div> </div> </div>
5	E	27	<div> <div>7%</div> <div> <div></div> <div>78%</div> <div>11%</div> </div> </div>
6	F	21	<div> <div>10%</div> <div> <div></div> <div>29%</div> <div>71%</div> </div> </div>

## 2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 12472 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 Reverse transcriptase/ribonuclease H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	558	Total	C	N	O	S	0	0	0
			4545	2942	760	836	7			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	EXPRESSION TAG	UNP P03366
A	0	VAL	-	EXPRESSION TAG	UNP P03366
A	41	LEU	MET	ENGINEERED	UNP P03366
A	67	ASN	ASP	ENGINEERED	UNP P03366
A	70	ARG	LYS	ENGINEERED	UNP P03366
A	215	TYR	THR	ENGINEERED	UNP P03366
A	219	GLN	LYS	ENGINEERED	UNP P03366
A	258	CYS	GLN	ENGINEERED	UNP P03366
A	280	SER	CYS	ENGINEERED	UNP P03366
A	561	VAL	PHE	SEE REMARK 999	UNP P03366

- Molecule 2 is a protein called p51 RT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	428	Total	C	N	O	S	0	0	0
			3538	2305	586	640	7			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	ENGINEERED	UNP P03366
B	429	GLY	-	EXPRESSION TAG	UNP P03366
B	430	GLY	-	EXPRESSION TAG	UNP P03366
B	431	HIS	-	EXPRESSION TAG	UNP P03366
B	432	HIS	-	EXPRESSION TAG	UNP P03366
B	433	HIS	-	EXPRESSION TAG	UNP P03366

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	434	HIS	-	EXPRESSION TAG	UNP P03366
B	435	HIS	-	EXPRESSION TAG	UNP P03366
B	436	HIS	-	EXPRESSION TAG	UNP P03366
B	437	HIS	-	EXPRESSION TAG	UNP P03366

- Molecule 3 is a protein called monoclonal antibody, heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	211	Total	C	N	O	S	0	0	0
			1643	1025	270	342	6			

- Molecule 4 is a protein called monoclonal antibody, light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	225	Total	C	N	O	S	0	0	0
			1685	1060	276	340	9			

- Molecule 5 is a DNA chain called DNA (5'-D(\*AP\*T\*GP\*CP\*TP\*AP\*GP\*GP\*CP\*GP\*C  
P\*CP\*CP\*GP\*AP\*AP\*CP\*AP\*GP\*GP\*GP\*AP\*CP\*TP\*GP\*TP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	24	Total	C	N	O	P	0	0	0
			493	233	97	140	23			

- Molecule 6 is a DNA chain called DNA (5'-D(\*AP\*CP\*AP\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*  
TP\*TP\*CP\*GP\*GP\*(MRG)P\*CP\*GP\*CP\*CP\*(ATM))-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	F	21	Total	C	N	O	P	0	0	0
			429	205	77	126	20			

- Molecule 7 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	1	Total	Mg	0	0
			1	1		

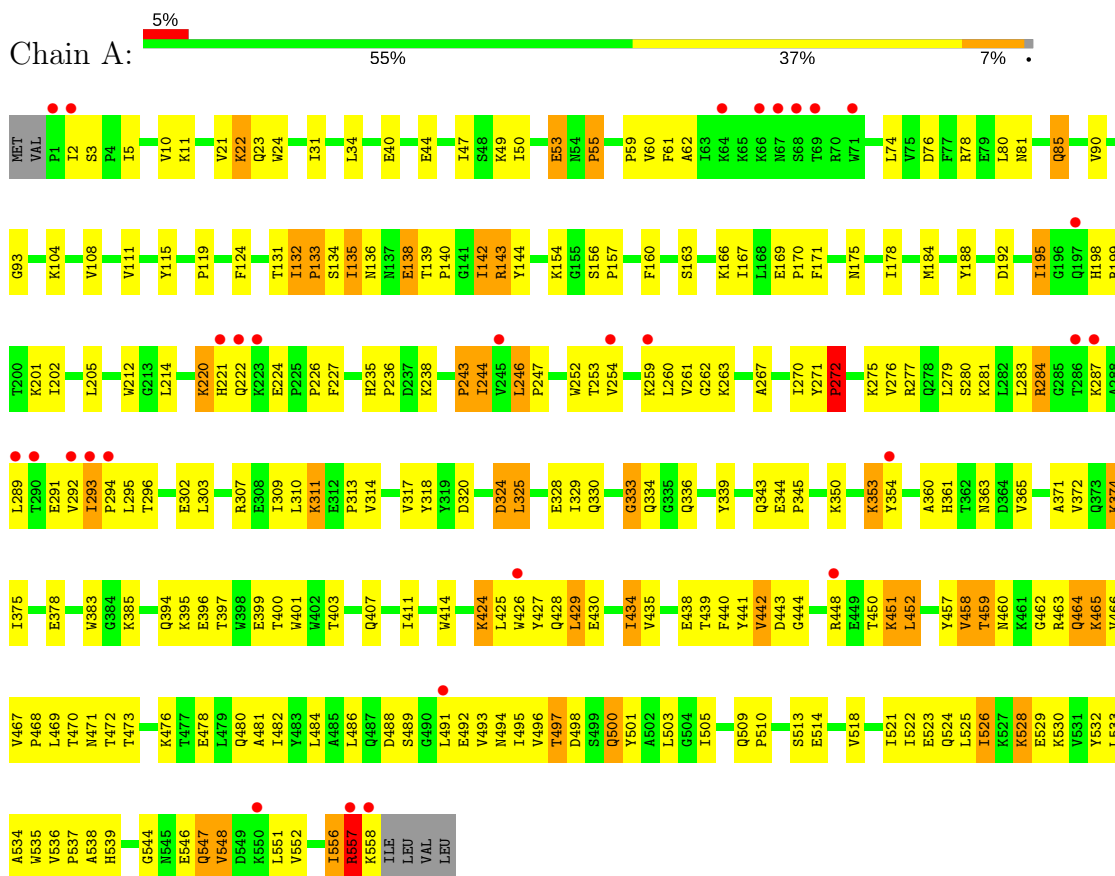
- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	41	Total 41	O 41	0	0
8	B	37	Total 37	O 37	0	0
8	C	22	Total 22	O 22	0	0
8	D	25	Total 25	O 25	0	0
8	E	5	Total 5	O 5	0	0
8	F	8	Total 8	O 8	0	0

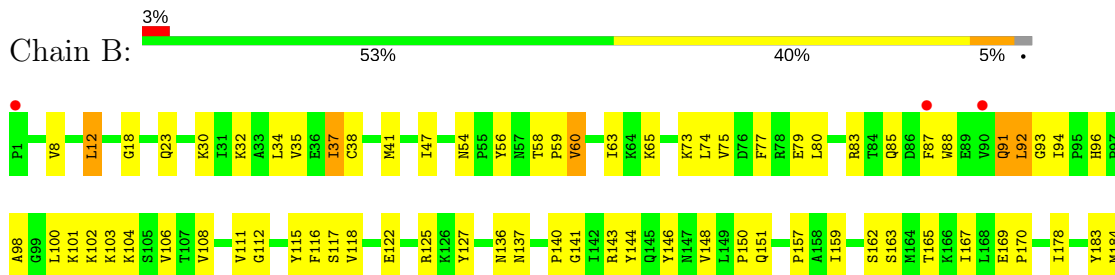
### 3 Residue-property plots

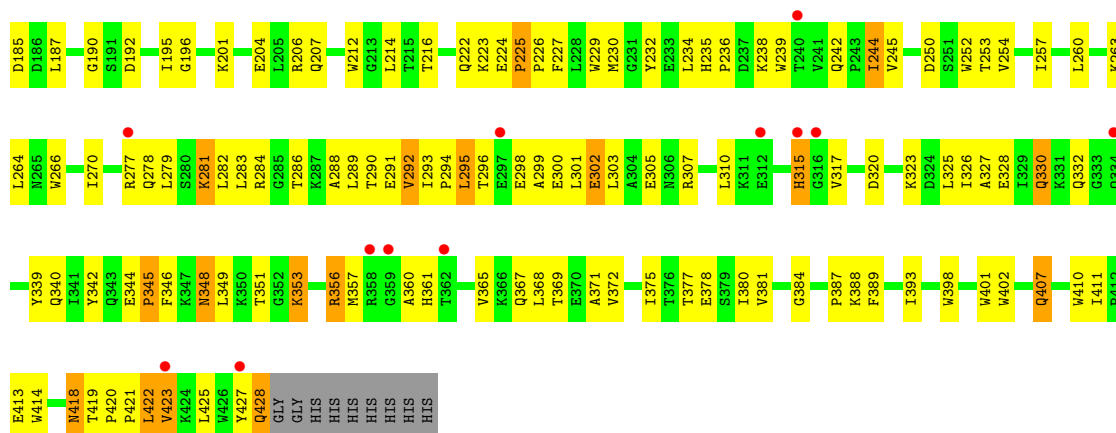
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: Reverse transcriptase/ribonuclease H

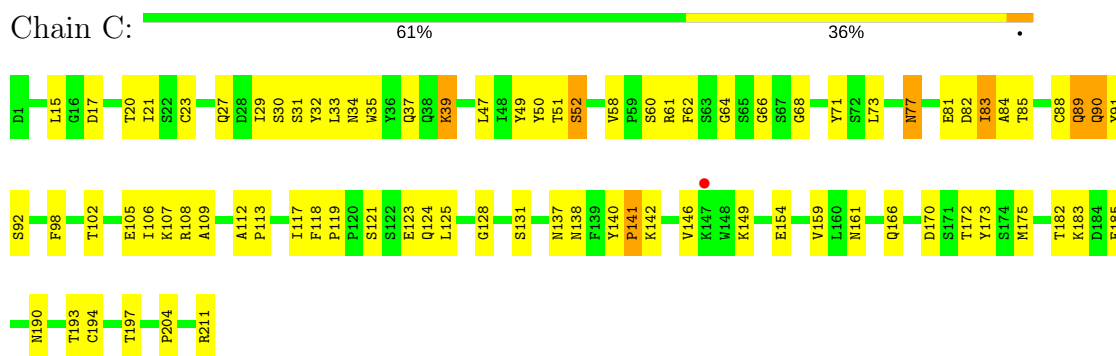


#### • Molecule 2: p51 RT

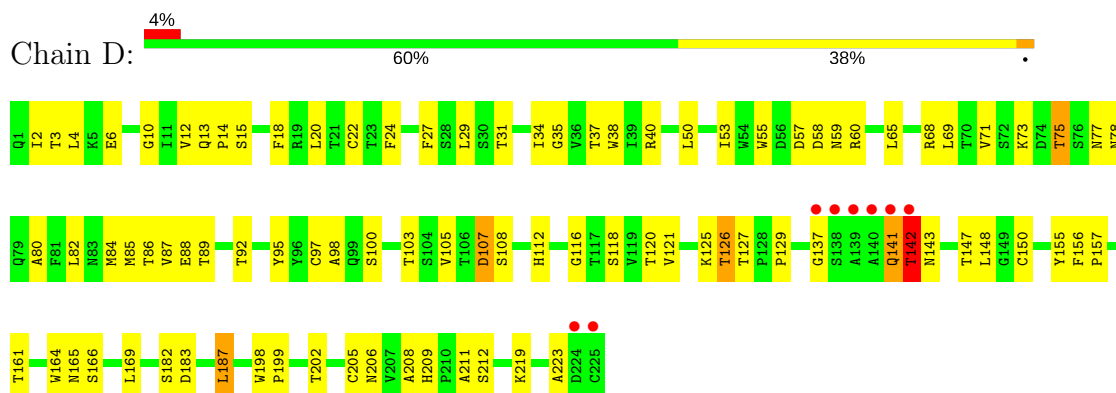




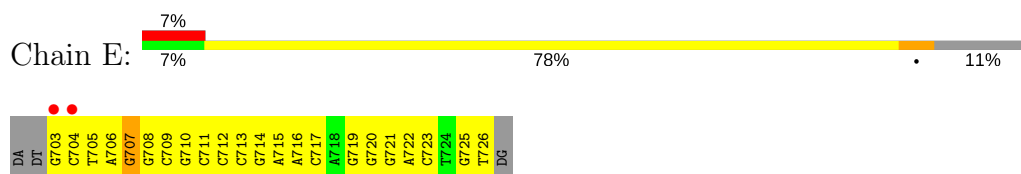
- Molecule 3: monoclonal antibody, heavy chain



- Molecule 4: monoclonal antibody, light chain



- Molecule 5: DNA (5'-D(\*AP\*T\*GP\*CP\*TP\*AP\*GP\*GP\*CP\*GP\*CP\*CP\*CP\*GP\*AP\*AP\*CP\*AP\*GP\*GP\*GP\*AP\*CP\*TP\*GP\*TP\*G)-3')



- Molecule 6: DNA (5'-D(\*AP\*CP\*AP\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*(M RG)P\*CP\*GP\*CP\*CP\*(ATM))-3')





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32 1 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	166.47Å 166.47Å 220.82Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	24.79 – 2.90 31.68 – 2.91	Depositor EDS
% Data completeness (in resolution range)	86.5 (24.79-2.90) 86.6 (31.68-2.91)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.54 (at 2.90Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, $R_{free}$	0.260 , 0.294 0.260 , 0.295	Depositor DCC
$R_{free}$ test set	2742 reflections (4.11%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	62.1	Xtriage
Anisotropy	0.050	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 32.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	0.043 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	12472	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	74.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.07% 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: ATM, MG, MRG

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	0.51	0/4664	0.69	3/6336 (0.0%)
2	B	0.59	0/3643	0.72	0/4952
3	C	0.49	0/1681	0.73	1/2283 (0.0%)
4	D	0.51	0/1729	0.74	0/2372
5	E	0.52	0/554	0.80	0/854
6	F	0.51	0/424	0.79	0/649
All	All	0.53	0/12695	0.72	4/17446 (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
5	E	0	1

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	52	SER	N-CA-CB	-5.46	102.31	110.50
1	A	458	VAL	N-CA-C	-5.43	96.35	111.00
1	A	557	ARG	NE-CZ-NH2	5.37	122.98	120.30
1	A	442	VAL	CB-CA-C	-5.06	101.79	111.40

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
5	E	707	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	4545	0	4607	281	0
2	B	3538	0	3576	209	0
3	C	1643	0	1565	68	0
4	D	1685	0	1640	78	0
5	E	493	0	269	30	0
6	F	429	0	242	23	0
7	A	1	0	0	0	0
8	A	41	0	0	11	0
8	B	37	0	0	17	0
8	C	22	0	0	6	0
8	D	25	0	0	6	0
8	E	5	0	0	3	0
8	F	8	0	0	2	0
All	All	12472	0	11899	653	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 27.

The worst 5 of 653 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:B:225:PRO:HB2	2:B:226:PRO:HD3	1.16	1.12
1:A:139:THR:HG23	1:A:140:PRO:HD2	1.20	1.11
1:A:556:ILE:HG22	1:A:557:ARG:H	0.98	1.08
1:A:498:ASP:HB2	1:A:538:ALA:HB2	1.33	1.08
1:A:486:LEU:HB3	1:A:524:GLN:HG2	1.35	1.07

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	556/564 (99%)	483 (87%)	61 (11%)	12 (2%)	7	27
2	B	426/437 (98%)	377 (88%)	41 (10%)	8 (2%)	9	31
3	C	209/211 (99%)	185 (88%)	22 (10%)	2 (1%)	17	49
4	D	223/225 (99%)	197 (88%)	24 (11%)	2 (1%)	19	52
All	All	1414/1437 (98%)	1242 (88%)	148 (10%)	24 (2%)	10	34

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	85	GLN
1	A	272	PRO
2	B	225	PRO
2	B	292	VAL
2	B	345	PRO

### 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	498/504 (99%)	461 (93%)	37 (7%)	15	41
2	B	390/397 (98%)	366 (94%)	24 (6%)	20	51
3	C	190/190 (100%)	184 (97%)	6 (3%)	42	76
4	D	196/196 (100%)	189 (96%)	7 (4%)	38	73
All	All	1274/1287 (99%)	1200 (94%)	74 (6%)	22	54

5 of 74 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	526	ILE
2	B	92	LEU
4	D	75	THR
1	A	547	GLN
2	B	12	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 44 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	137	ASN
2	B	255	ASN
4	D	141	GLN
2	B	151	GLN
2	B	235	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

2 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
6	MRG	F	817	1,5,6	22,28,29	1.68	3 (13%)	24,39,42	3.88	8 (33%)
6	ATM	F	822	5,6	14,23,24	1.33	3 (21%)	17,32,35	3.57	2 (11%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
6	MRG	F	817	1,5,6	-	0/8/26/27	0/3/3/3
6	ATM	F	822	5,6	-	0/6/24/25	0/2/2/2

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	F	817	MRG	C21-N2	-5.62	1.33	1.45
6	F	822	ATM	N4'-N3'	-2.93	1.16	1.23
6	F	817	MRG	C8-N7	-2.27	1.30	1.34
6	F	822	ATM	C6-C5	-2.21	1.34	1.40
6	F	822	ATM	C4-N3	2.77	1.38	1.33

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	F	817	MRG	C21-N2-C2	-14.01	99.20	123.73
6	F	817	MRG	C5-C6-N1	-8.29	111.67	123.47
6	F	817	MRG	C23-C22-C21	-4.67	98.69	112.66
6	F	817	MRG	C2-N3-C4	-2.96	111.74	115.11
6	F	817	MRG	C2'-C1'-N9	-2.42	108.63	114.27

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	F	817	MRG	5	0
6	F	822	ATM	1	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 1 ligands modelled in this entry, 1 is 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

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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	558/564 (98%)	0.29	29 (5%) 27 22	43, 79, 112, 125	0
2	B	428/437 (97%)	0.15	15 (3%) 44 38	36, 62, 108, 122	0
3	C	211/211 (100%)	0.16	1 (0%) 90 90	52, 69, 102, 109	0
4	D	225/225 (100%)	0.03	8 (3%) 42 37	42, 64, 96, 115	0
5	E	24/27 (88%)	0.23	2 (8%) 11 8	57, 109, 140, 157	0
6	F	19/21 (90%)	0.19	2 (10%) 6 4	71, 93, 130, 137	0
All	All	1465/1485 (98%)	0.19	57 (3%) 39 35	36, 70, 110, 157	0

The worst 5 of 57 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	315	HIS	6.1
4	D	141	GLN	5.6
1	A	67	ASN	5.2
4	D	225	CYS	4.6
1	A	557	ARG	4.4

### 6.2 Non-standard residues in protein, DNA, RNA chains [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
6	MRG	F	817	26/27	0.92	0.22	84,95,114,122	0
6	ATM	F	822	22/23	0.94	0.20	84,92,96,100	0



### 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( $\text{\AA}^2$ )	Q<0.9
7	MG	A	1001	1/1	0.95	0.25	28,28,28,28	0

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

There are no such residues in this entry.