



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 2, 2017 – 07:14 PM EDT

PDB ID : 3V6D  
Title : Crystal structure of HIV-1 reverse transcriptase (RT) cross-linked with AZT-terminated DNA  
Authors : Das, K.; Martinez, S.E.; Arnold, E.  
Deposited on : unknown  
Resolution : 2.70 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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20030345  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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) : rb-20030345

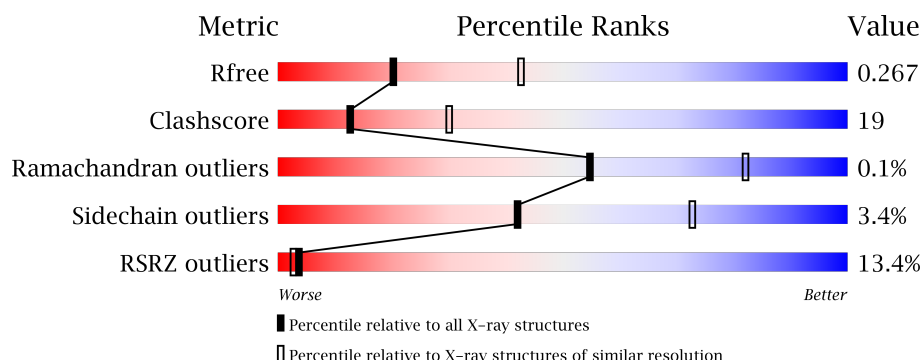
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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))
$R_{free}$	100719	2259 (2.70-2.70)
Clashscore	112137	2590 (2.70-2.70)
Ramachandran outliers	110173	2550 (2.70-2.70)
Sidechain outliers	110143	2550 (2.70-2.70)
RSRZ outliers	101464	2275 (2.70-2.70)

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 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	556	<div> <div>17%</div> <div> <div>62%</div> <div>35%</div> <div>•</div> </div> </div>
1	C	556	<div> <div>16%</div> <div> <div>58%</div> <div>40%</div> <div>•</div> </div> </div>
2	B	428	<div> <div>7%</div> <div> <div>65%</div> <div>30%</div> <div>• •</div> </div> </div>
2	D	428	<div> <div>10%</div> <div> <div>60%</div> <div>36%</div> <div>•</div> </div> </div>
3	E	27	<div> <div>19%</div> <div> <div>30%</div> <div>37%</div> <div>22%</div> <div>11%</div> </div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	T	27	
4	F	21	
4	P	21	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 17627 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 HIV-1 REVERSE TRANSCRIPTASE P66 subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	555	Total	C	N	O	S	0	0	0
			4511	2920	751	832	8			
1	C	554	Total	C	N	O	S	0	0	0
			4506	2917	750	831	8			

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	258	CYS	GLN	ENGINEERED MUTATION	UNP P03366
A	280	SER	CYS	ENGINEERED MUTATION	UNP P03366
A	498	ASN	ASP	ENGINEERED MUTATION	UNP P03366
C	-1	MET	-	EXPRESSION TAG	UNP P03366
C	0	VAL	-	EXPRESSION TAG	UNP P03366
C	258	CYS	GLN	ENGINEERED MUTATION	UNP P03366
C	280	SER	CYS	ENGINEERED MUTATION	UNP P03366
C	498	ASN	ASP	ENGINEERED MUTATION	UNP P03366

- Molecule 2 is a protein called HIV-1 REVERSE TRANSCRIPTASE P51 subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	412	Total	C	N	O	S	0	0	0
			3400	2212	563	619	6			
2	D	412	Total	C	N	O	S	0	0	0
			3400	2212	563	619	6			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	ENGINEERED MUTATION	UNP P03366
D	280	SER	CYS	ENGINEERED MUTATION	UNP P03366

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	T	24	Total	C	N	O	P	0	0	0
			497	234	102	138	23			
3	E	24	Total	C	N	O	P	0	0	0
			497	234	102	138	23			

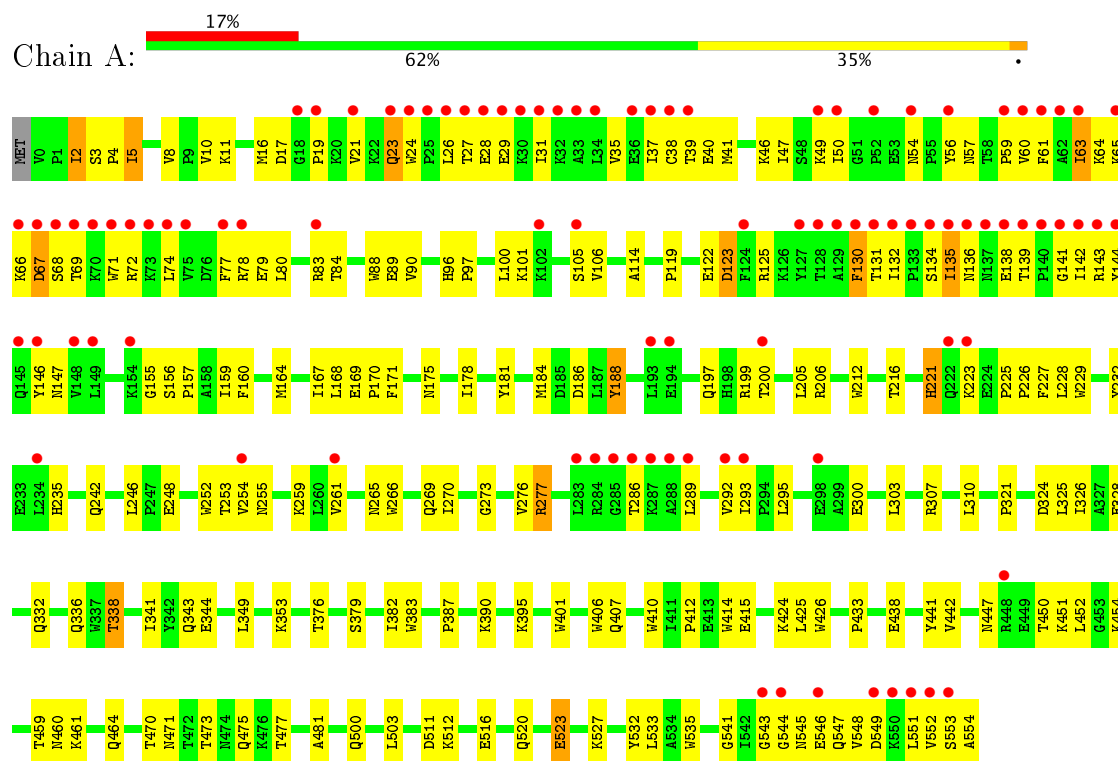
- Molecule 4 is a DNA chain called DNA (5'-D(\*AP\*CP\*AP\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*(MRG)P\*CP\*GP\*CP\*CP\*(ATM))-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
4	P	20	Total	C	N	O	P	S	0	0	0
			408	195	72	121	19	1			
4	F	20	Total	C	N	O	P	S	0	0	0
			408	195	72	121	19	1			

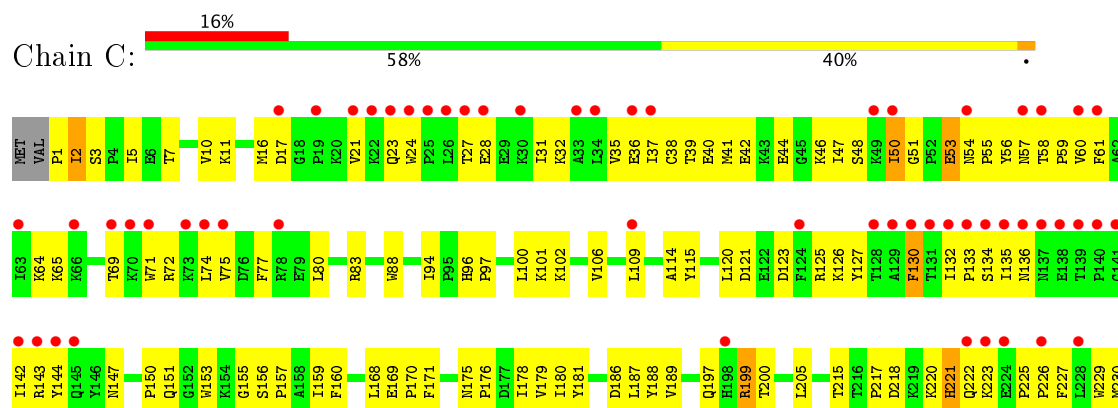
### 3 Residue-property plots

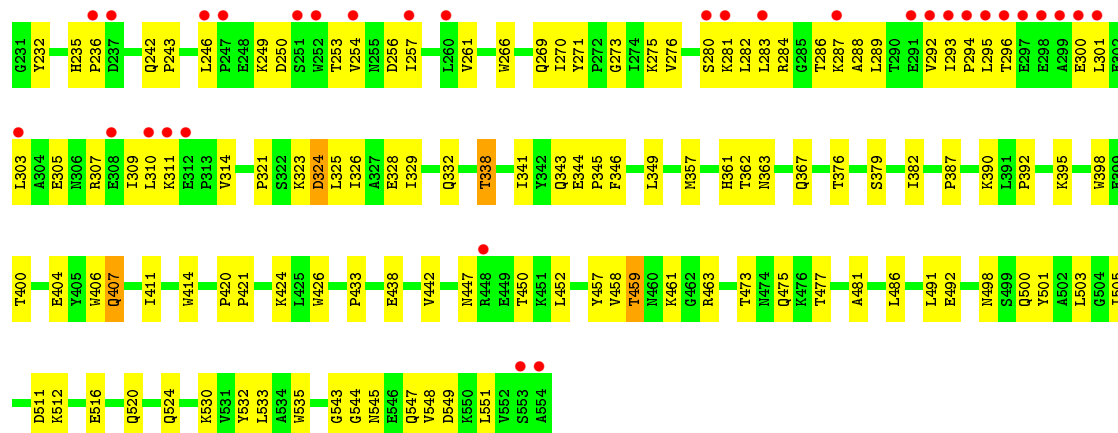
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: HIV-1 REVERSE TRANSCRIPTASE P66 subunit

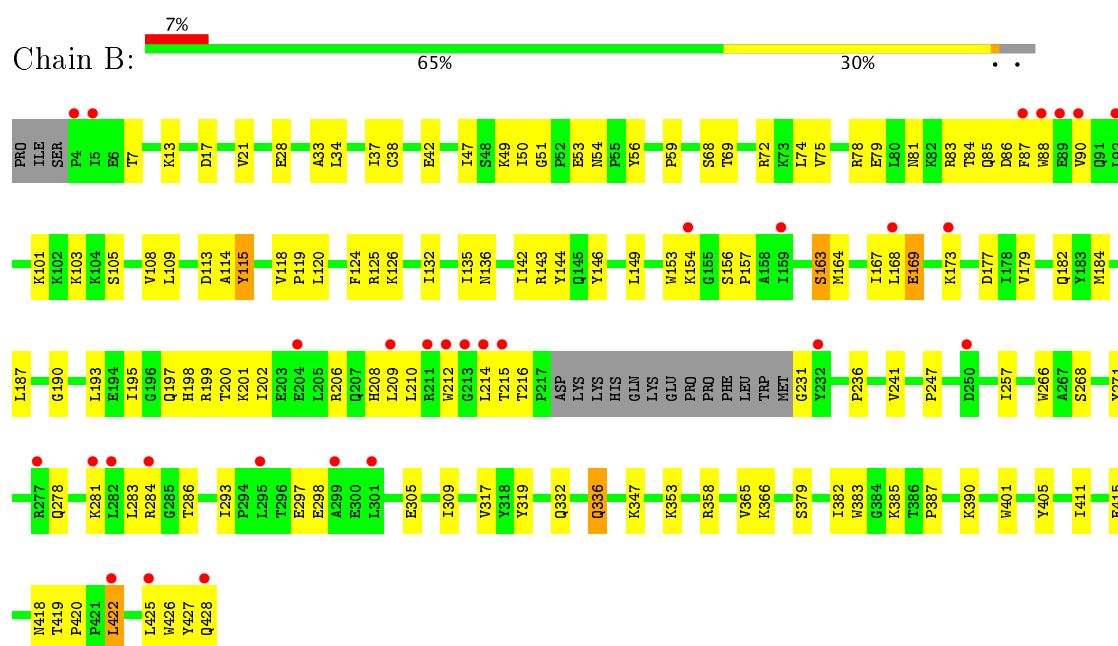


- Molecule 1: HIV-1 REVERSE TRANSCRIPTASE P66 subunit

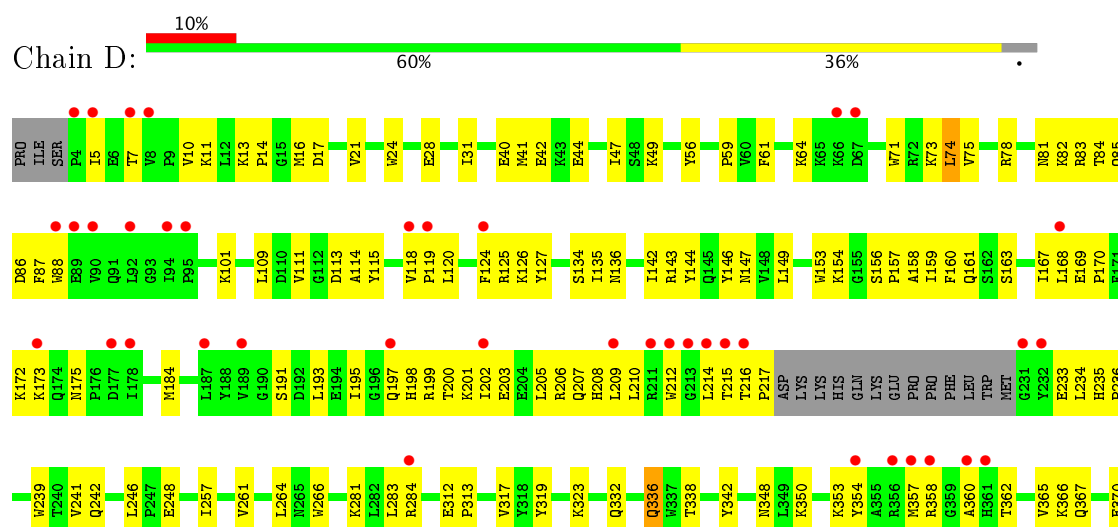


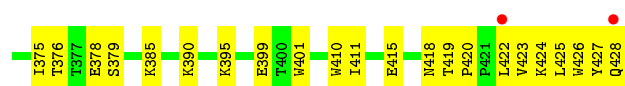


• Molecule 2: HIV-1 REVERSE TRANSCRIPTASE P51 subunit



• Molecule 2: HIV-1 REVERSE TRANSCRIPTASE P51 subunit

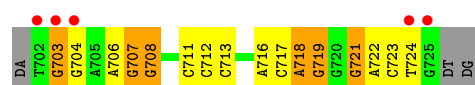




- Molecule 3: DNA (5'-D(\*AP\*TP\*GP\*GP\*AP\*AP\*GP\*GP\*CP\*GP\*CP\*CP\*CP\*GP\*AP\*AP\*CP\*AP\*GP\*GP\*GP\*AP\*CP\*TP\*GP\*TP\*G)-3')



- Molecule 3: DNA (5'-D(\*AP\*TP\*GP\*GP\*AP\*AP\*GP\*GP\*CP\*GP\*CP\*CP\*CP\*GP\*AP\*AP\*CP\*AP\*GP\*GP\*GP\*AP\*CP\*TP\*GP\*TP\*G)-3')



- Molecule 4: 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')



- Molecule 4: 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 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	89.48Å 133.17Å 139.90Å 90.00° 98.67° 90.00°	Depositor
Resolution (Å)	44.30 – 2.70 48.21 – 2.71	Depositor EDS
% Data completeness (in resolution range)	97.1 (44.30-2.70) 97.8 (48.21-2.71)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.05 (at 2.73Å)	Xtriage
Refinement program	PHENIX 1.7.1 _743	Depositor
R, $R_{free}$	0.233 , 0.268 0.234 , 0.267	Depositor DCC
$R_{free}$ test set	2606 reflections (3.12%)	DCC
Wilson B-factor (Å <sup>2</sup> )	67.4	Xtriage
Anisotropy	0.099	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 56.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	17627	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	80.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.69% 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: ATM, 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	0.46	0/4629	0.61	1/6290 (0.0%)
1	C	0.45	0/4624	0.58	0/6282
2	B	0.48	0/3497	0.60	0/4751
2	D	0.46	0/3497	0.61	0/4751
3	E	0.79	0/560	1.44	10/864 (1.2%)
3	T	0.80	0/560	1.48	11/864 (1.3%)
4	F	0.82	0/400	1.58	9/612 (1.5%)
4	P	0.78	0/400	1.52	6/612 (1.0%)
All	All	0.51	0/18167	0.76	37/25026 (0.1%)

There are no bond length outliers.

The worst 5 of 37 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	P	810	DT	O4'-C1'-N1	9.89	114.92	108.00
4	F	807	DC	O4'-C1'-N1	8.96	114.27	108.00
4	P	816	DG	O4'-C1'-N9	8.78	114.14	108.00
3	E	721	DG	O4'-C1'-N9	-8.59	101.99	108.00
3	T	703	DG	O4'-C4'-C3'	-7.75	101.35	106.00

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	4511	0	4570	174	0
1	C	4506	0	4568	182	0
2	B	3400	0	3433	117	0
2	D	3400	0	3433	135	0
3	E	497	0	268	14	0
3	T	497	0	268	15	0
4	F	408	0	231	27	0
4	P	408	0	231	20	0
All	All	17627	0	17002	643	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 19.

The worst 5 of 643 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:C:223:LYS:O	1:C:225:PRO:HD3	1.53	1.09
1:A:541:GLY:HA2	1:A:546:GLU:HB2	1.41	1.01
1:C:500:GLN:HG2	2:D:422:LEU:HD22	1.44	1.00
1:A:459:THR:HG22	1:A:461:LYS:H	1.23	0.99
1:A:2:ILE:HD12	1:A:2:ILE:H	1.26	0.99

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	553/556 (100%)	529 (96%)	23 (4%)	1 (0%)	51	79
1	C	552/556 (99%)	528 (96%)	24 (4%)	0	100	100
2	B	408/428 (95%)	399 (98%)	9 (2%)	0	100	100
2	D	408/428 (95%)	391 (96%)	17 (4%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1921/1968 (98%)	1847 (96%)	73 (4%)	1 (0%)	55 82

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	135	ILE

### 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	495/497 (100%)	476 (96%)	19 (4%)	38 68
1	C	495/497 (100%)	475 (96%)	20 (4%)	36 67
2	B	374/390 (96%)	362 (97%)	12 (3%)	44 75
2	D	374/390 (96%)	366 (98%)	8 (2%)	59 85
All	All	1738/1774 (98%)	1679 (97%)	59 (3%)	42 73

5 of 59 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	241	VAL
1	C	5	ILE
2	D	248	GLU
2	B	257	ILE
2	B	336	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	222	GLN
1	C	367	GLN
1	C	547	GLN
2	B	336	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	464	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

4 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
4	MRG	F	817	1,3,4	21,28,29	3.19	10 (47%)	24,39,42	1.93	7 (29%)
4	ATM	F	822	3,4	14,23,24	1.44	1 (7%)	18,32,35	2.02	5 (27%)
4	MRG	P	817	1,3,4	21,28,29	3.27	10 (47%)	24,39,42	2.02	8 (33%)
4	ATM	P	822	3,4	14,23,24	1.40	1 (7%)	18,32,35	2.34	5 (27%)

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
4	MRG	F	817	1,3,4	-	0/8/26/27	0/3/3/3
4	ATM	F	822	3,4	-	0/6/24/25	0/2/2/2
4	MRG	P	817	1,3,4	-	0/8/26/27	0/3/3/3
4	ATM	P	822	3,4	-	0/6/24/25	0/2/2/2

The worst 5 of 22 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	P	817	MRG	O5'-C5'	-4.67	1.38	1.44
4	F	817	MRG	O5'-C5'	-4.63	1.38	1.44

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	P	817	MRG	C2'-C3'	-3.43	1.43	1.52
4	F	817	MRG	C2'-C3'	-3.40	1.43	1.52
4	P	817	MRG	O3'-C3'	-3.24	1.36	1.43

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	P	822	ATM	C2'-C1'-N1	-5.96	100.15	114.23
4	F	822	ATM	C5-C4-N3	-4.93	119.80	125.24
4	P	817	MRG	N3-C2-N1	-4.10	120.03	126.23
4	F	817	MRG	N3-C2-N1	-4.03	120.14	126.23
4	P	822	ATM	C5-C4-N3	-3.69	121.17	125.24

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	F	817	MRG	4	0
4	F	822	ATM	2	0
4	P	817	MRG	1	0
4	P	822	ATM	4	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 5.7 Other polymers [i](#)

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 ⓘ

### 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	555/556 (99%)	1.07	96 (17%) 2 1	36, 83, 138, 149	0
1	C	554/556 (99%)	1.02	89 (16%) 2 1	34, 81, 136, 148	0
2	B	412/428 (96%)	0.68	30 (7%) 16 13	34, 66, 118, 131	0
2	D	412/428 (96%)	0.76	41 (9%) 8 6	40, 69, 125, 132	0
3	E	24/27 (88%)	0.98	5 (20%) 1 1	67, 104, 156, 164	0
3	T	24/27 (88%)	1.23	5 (20%) 1 1	69, 106, 158, 161	0
4	F	18/21 (85%)	0.49	2 (11%) 6 5	60, 92, 141, 142	0
4	P	18/21 (85%)	0.47	2 (11%) 6 5	68, 90, 142, 144	0
All	All	2017/2064 (97%)	0.90	270 (13%) 4 3	34, 75, 133, 164	0

The worst 5 of 270 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	214	LEU	14.1
1	A	133	PRO	14.0
1	C	142	ILE	13.4
1	A	74	LEU	12.5
1	C	133	PRO	12.3

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

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( $\text{\AA}^2$ )	Q<0.9
4	MRG	F	817	26/27	0.91	0.14	-	84,99,117,127	0
4	ATM	F	822	22/23	0.93	0.18	-	71,78,87,95	0
4	MRG	P	817	26/27	0.90	0.18	-	88,102,117,121	0
4	ATM	P	822	22/23	0.94	0.16	-	71,82,95,95	0

### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [i](#)

There are no ligands in this entry.

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

There are no such residues in this entry.