



# wwPDB X-ray Structure Validation Summary Report

Aug 28, 2023 – 05:24 AM EDT

PDB ID : 3KLG  
Title : Crystal structure of AZT-resistant HIV-1 Reverse Transcriptase crosslinked to pre-translocation AZTMP-Terminated DNA (COMPLEX N)  
Authors : Tu, X.; Sarafianos, S.G.; Arnold, E.  
Deposited on : 2009-11-07  
Resolution : 3.65 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.35  
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) : 2.35

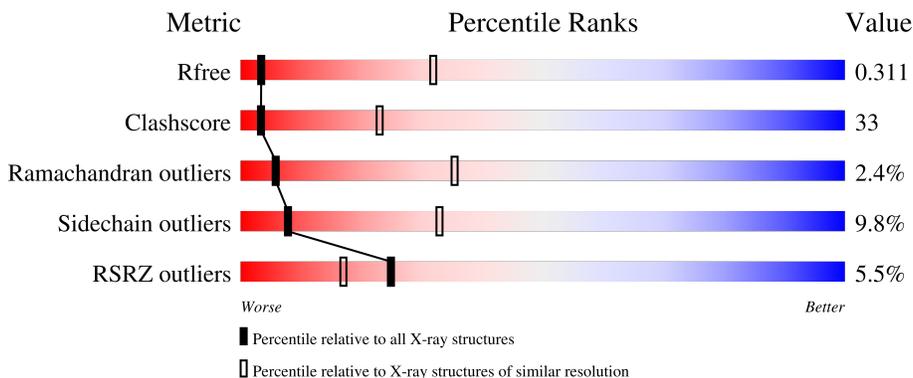
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.65 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.



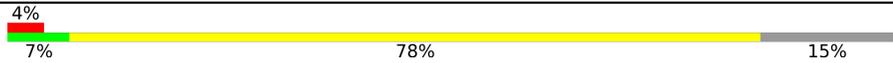
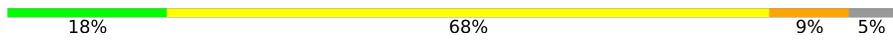
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1557 (3.82-3.50)
Clashscore	141614	1037 (3.80-3.52)
Ramachandran outliers	138981	1004 (3.80-3.52)
Sidechain outliers	138945	1002 (3.80-3.52)
RSRZ outliers	127900	1441 (3.82-3.50)

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 of 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	562	
1	E	562	
2	B	437	
2	F	437	
3	C	27	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	G	27	
4	D	22	
4	H	22	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	ATM	H	823	-	-	X	-

## 2 Entry composition i

There are 4 unique types of molecules in this entry. The entry contains 17640 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
			Total	C	N	O	S			
1	A	554	4513	2922	752	832	7	0	0	0
1	E	554	4513	2922	752	832	7	0	0	0

There are 20 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 mutation	UNP P03366
A	67	ASN	ASP	engineered mutation	UNP P03366
A	70	ARG	LYS	engineered mutation	UNP P03366
A	215	TYR	THR	engineered mutation	UNP P03366
A	219	GLN	LYS	engineered mutation	UNP P03366
A	258	CYS	GLN	engineered mutation	UNP P03366
A	280	SER	CYS	engineered mutation	UNP P03366
A	559	VAL	ILE	SEE REMARK 999	UNP P03366
E	-1	MET	-	expression tag	UNP P03366
E	0	VAL	-	expression tag	UNP P03366
E	41	LEU	MET	engineered mutation	UNP P03366
E	67	ASN	ASP	engineered mutation	UNP P03366
E	70	ARG	LYS	engineered mutation	UNP P03366
E	215	TYR	THR	engineered mutation	UNP P03366
E	219	GLN	LYS	engineered mutation	UNP P03366
E	258	CYS	GLN	engineered mutation	UNP P03366
E	280	SER	CYS	engineered mutation	UNP P03366
E	559	VAL	ILE	SEE REMARK 999	UNP P03366

- Molecule 2 is a protein called p51 RT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	413	3405	2215	564	620	6	0	0	0
2	F	413	3405	2215	564	620	6	0	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	engineered mutation	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
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
F	280	SER	CYS	engineered mutation	UNP P03366
F	429	GLY	-	expression tag	UNP P03366
F	430	GLY	-	expression tag	UNP P03366
F	431	HIS	-	expression tag	UNP P03366
F	432	HIS	-	expression tag	UNP P03366
F	433	HIS	-	expression tag	UNP P03366
F	434	HIS	-	expression tag	UNP P03366
F	435	HIS	-	expression tag	UNP P03366
F	436	HIS	-	expression tag	UNP P03366
F	437	HIS	-	expression tag	UNP P03366

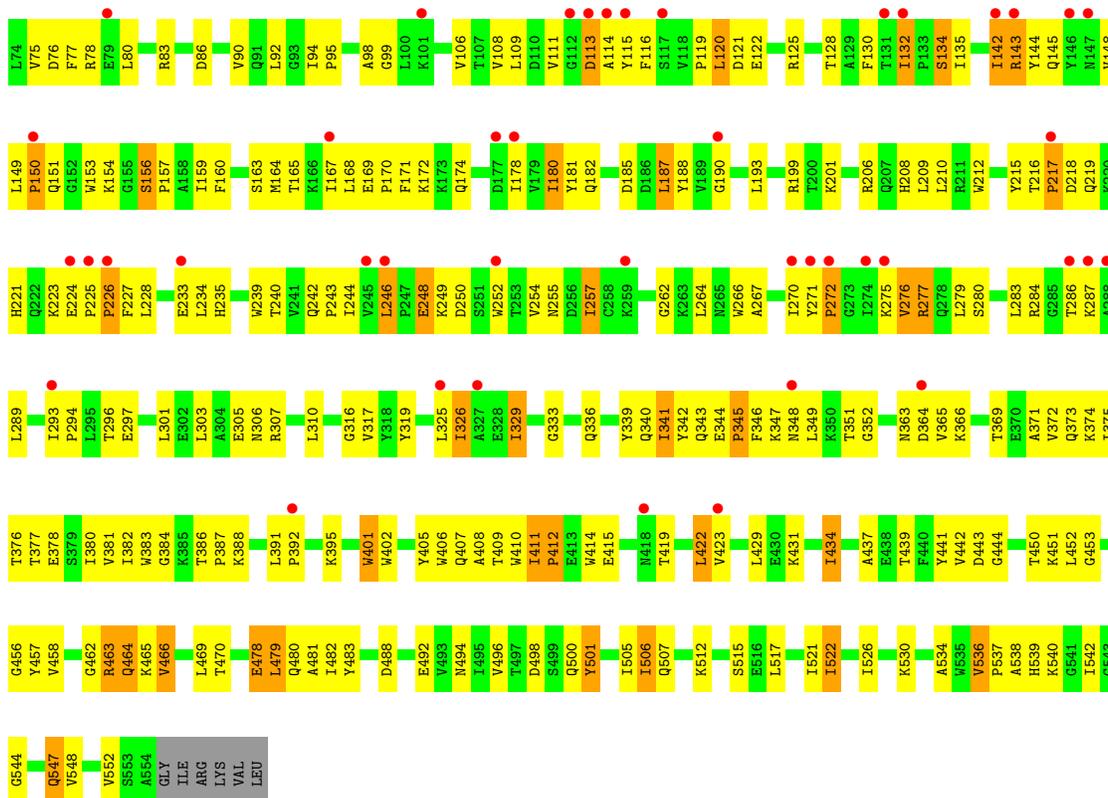
- Molecule 3 is a DNA chain called DNA (5'-D(\*AP\*T\*GP\*CP\*AP\*TP\*GP\*GP\*CP\*GP\*C  
P\*CP\*CP\*GP\*AP\*AP\*CP\*AP\*GP\*GP\*GP\*AP\*CP\*TP\*GP\*TP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	23	473	223	95	133	22	0	0	0
3	G	23	473	223	95	133	22	0	0	0

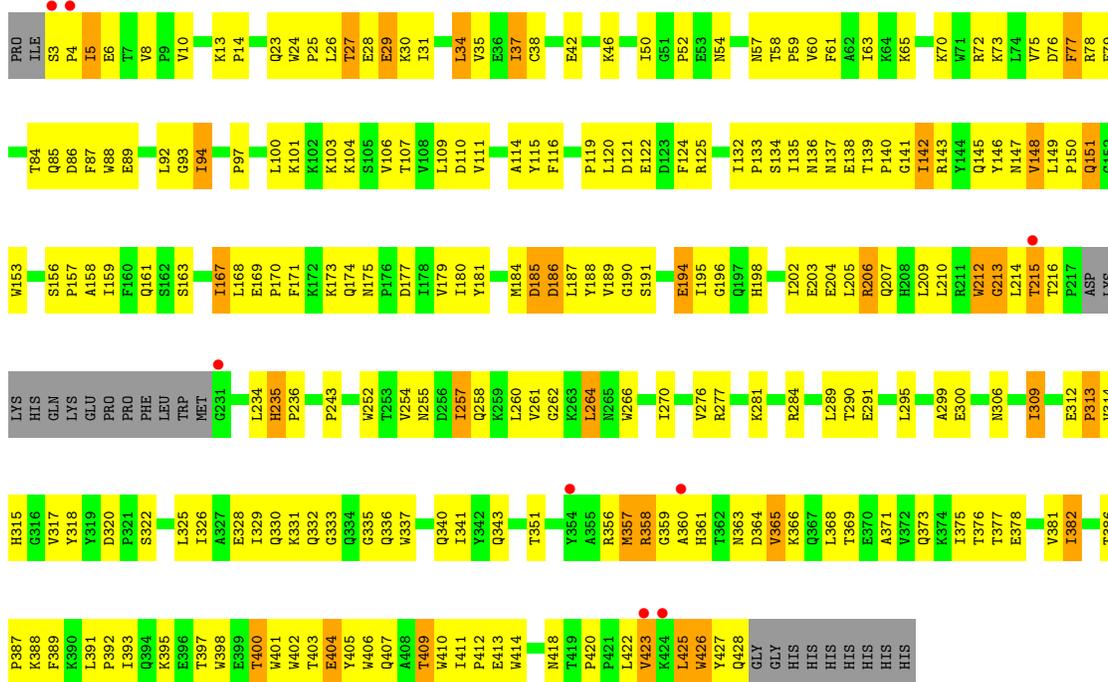
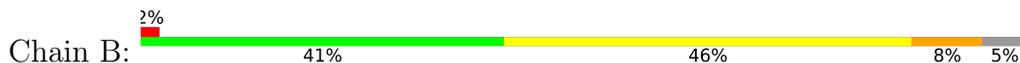
- Molecule 4 is a DNA chain called DNA (5'-D(\*A\*CP\*AP\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*T  
P\*TP\*CP\*GP\*GP\*(MRG)P\*CP\*GP\*CP\*CP\*AP\*(ATM))-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	P	S			
4	D	21	Total 429	205	77	126	20	1	0	0	0
4	H	21	Total 429	205	77	126	20	1	0	0	0



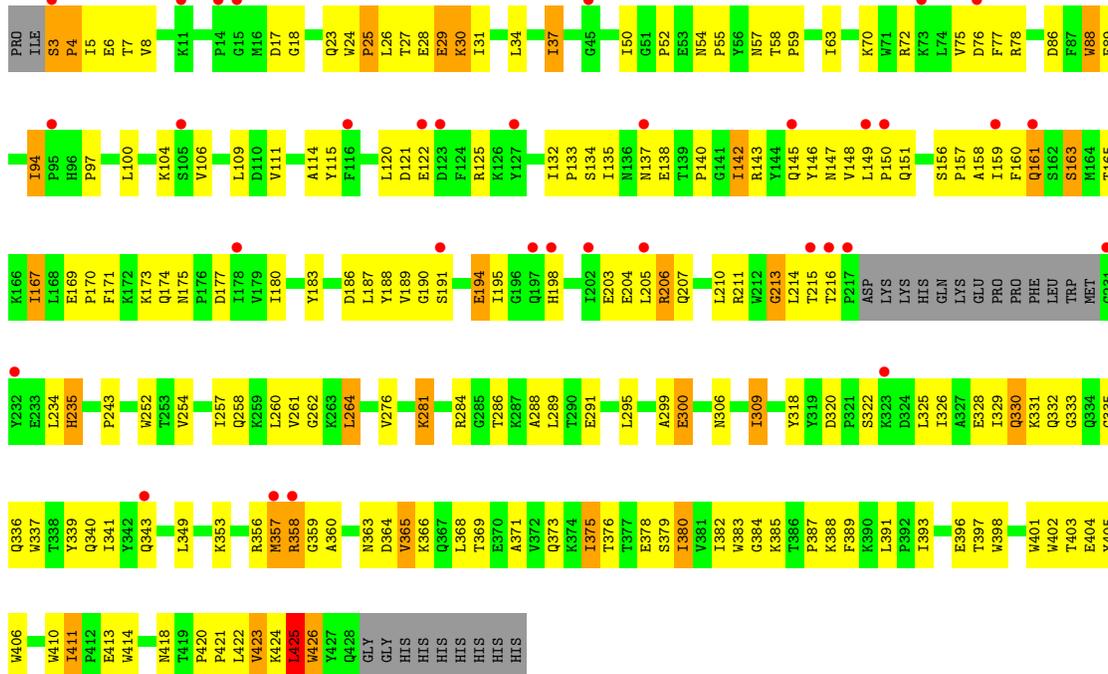


• Molecule 2: p51 RT



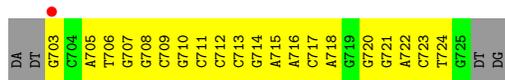
• Molecule 2: p51 RT

Chain F:



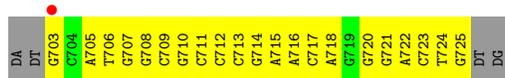
• Molecule 3: DNA (5'-D(\*AP\*T\*GP\*CP\*AP\*TP\*GP\*GP\*CP\*GP\*CP\*CP\*CP\*GP\*AP\*AP\*CP\*AP\*GP\*GP\*GP\*AP\*CP\*TP\*GP\*TP\*G)-3')

Chain C:



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

Chain G:



• Molecule 4: DNA (5'-D(\*A\*CP\*AP\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*(MR G)P\*CP\*GP\*CP\*CP\*AP\*(ATM))-3')

Chain D:



• Molecule 4: DNA (5'-D(\*A\*CP\*AP\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*(MR G)P\*CP\*GP\*CP\*CP\*AP\*(ATM))-3')

Chain H: 32% 55% 9% 5%

DA	C803	A804	G805	T806	C807	C808	C809	T810	G816	G817	C818	G819	C820	C821	A822	T823
----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	79.12Å 154.40Å 277.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.97 – 3.65 39.97 – 3.65	Depositor EDS
% Data completeness (in resolution range)	85.0 (39.97-3.65) 83.4 (39.97-3.65)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.47 (at 3.66Å)	Xtrriage
Refinement program	CNS 1.2	Depositor
R, $R_{free}$	0.312 , 0.350 0.292 , 0.311	Depositor DCC
$R_{free}$ test set	652 reflections (2.02%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	120.0	Xtrriage
Anisotropy	0.483	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 78.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.26$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	17640	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	120.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.43% 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: MRG, ATM

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	0.54	0/4632	0.68	0/6295
1	E	0.43	1/4632 (0.0%)	0.64	2/6295 (0.0%)
2	B	0.62	1/3502 (0.0%)	0.76	4/4760 (0.1%)
2	F	0.41	0/3502	0.63	1/4760 (0.0%)
3	C	0.83	0/532	0.92	0/820
3	G	0.61	0/532	0.86	0/820
4	D	0.75	0/424	0.88	0/649
4	H	0.64	0/424	0.89	0/649
All	All	0.53	2/18180 (0.0%)	0.70	7/25048 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	478	GLU	CD-OE2	5.98	1.32	1.25
2	B	79	GLU	CG-CD	5.07	1.59	1.51

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	34	LEU	CB-CA-C	-5.98	98.84	110.20
2	F	353	LYS	N-CA-CB	-5.98	99.84	110.60
2	B	213	GLY	N-CA-C	5.57	127.03	113.10
1	E	539	HIS	N-CA-C	-5.54	96.05	111.00
2	B	315	HIS	CB-CA-C	5.48	121.37	110.40

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	4513	0	4568	350	0
1	E	4513	0	4568	293	0
2	B	3405	0	3437	244	0
2	F	3405	0	3437	207	0
3	C	473	0	257	37	0
3	G	473	0	257	32	0
4	D	429	0	243	36	0
4	H	429	0	243	35	0
All	All	17640	0	17010	1127	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 33.

The worst 5 of 1127 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:E:142:ILE:H	1:E:142:ILE:HD13	1.01	1.16
1:A:24:TRP:HA	3:C:703:DG:H21	1.09	1.10
2:B:5:ILE:HG22	2:B:6:GLU:N	1.53	1.10
2:B:5:ILE:HG22	2:B:6:GLU:H	0.94	1.08
1:E:540:LYS:O	1:E:542:ILE:HD12	1.53	1.08

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	552/562 (98%)	469 (85%)	71 (13%)	12 (2%)	6	37
1	E	552/562 (98%)	460 (83%)	78 (14%)	14 (2%)	5	35
2	B	409/437 (94%)	352 (86%)	45 (11%)	12 (3%)	4	32
2	F	409/437 (94%)	348 (85%)	53 (13%)	8 (2%)	7	39
All	All	1922/1998 (96%)	1629 (85%)	247 (13%)	46 (2%)	6	35

5 of 46 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	3	SER
1	A	134	SER
2	B	358	ARG
1	E	134	SER
1	E	345	PRO

### 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	495/502 (99%)	445 (90%)	50 (10%)	7	31
1	E	495/502 (99%)	446 (90%)	49 (10%)	8	32
2	B	375/397 (94%)	339 (90%)	36 (10%)	8	34
2	F	375/397 (94%)	340 (91%)	35 (9%)	9	36
All	All	1740/1798 (97%)	1570 (90%)	170 (10%)	8	33

5 of 170 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	329	ILE
2	F	135	ILE
1	E	402	TRP
1	E	501	TYR
2	F	206	ARG

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

Mol	Chain	Res	Type
1	E	336	GLN
2	F	54	ASN
1	E	343	GLN
1	E	509	GLN
2	F	147	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](#)

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	ATM	H	823	3,4	19,23,24	0.70	1 (5%)	27,32,35	0.69	1 (3%)
4	MRG	H	817	3,4	21,28,29	1.52	3 (14%)	19,39,42	3.06	3 (15%)
4	MRG	D	817	3,4	21,28,29	1.62	3 (14%)	19,39,42	3.18	4 (21%)
4	ATM	D	823	3,4	19,23,24	0.59	0	27,32,35	0.80	1 (3%)

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	ATM	H	823	3,4	-	1/10/24/25	0/2/2/2
4	MRG	H	817	3,4	-	4/8/26/27	0/3/3/3
4	MRG	D	817	3,4	-	4/8/26/27	0/3/3/3
4	ATM	D	823	3,4	-	0/10/24/25	0/2/2/2

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	817	MRG	C21-N2	-5.15	1.34	1.46
4	H	817	MRG	C21-N2	-4.99	1.34	1.46
4	D	817	MRG	C5-C6	-2.89	1.41	1.47
4	H	817	MRG	C5-C6	-2.71	1.41	1.47
4	H	823	ATM	N4'-N3'	-2.59	1.16	1.23

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	D	817	MRG	C21-N2-C2	-12.42	100.14	123.41
4	H	817	MRG	C21-N2-C2	-11.97	100.99	123.41
4	H	817	MRG	C23-C22-C21	-4.14	99.34	112.65
4	D	817	MRG	C23-C22-C21	-4.13	99.39	112.65
4	D	823	ATM	C2'-C3'-N3'	3.05	116.96	108.71

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	D	817	MRG	N3-C2-N2-C21
4	D	817	MRG	N1-C2-N2-C21
4	D	817	MRG	N2-C21-C22-C23
4	H	817	MRG	N3-C2-N2-C21
4	H	817	MRG	N1-C2-N2-C21

There are no ring outliers.

4 monomers are involved in 25 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	H	823	ATM	14	0
4	H	817	MRG	3	0
4	D	817	MRG	2	0
4	D	823	ATM	6	0

## 5.5 Carbohydrates [i](#)

There are no monosaccharides 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 [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	554/562 (98%)	-0.24	4 (0%) 87 80	26, 89, 115, 135	0
1	E	554/562 (98%)	0.63	62 (11%) 5 3	32, 154, 164, 172	0
2	B	413/437 (94%)	-0.16	8 (1%) 66 53	54, 84, 119, 143	0
2	F	413/437 (94%)	0.48	34 (8%) 11 8	127, 153, 165, 170	0
3	C	23/27 (85%)	-0.16	1 (4%) 35 24	65, 90, 125, 141	0
3	G	23/27 (85%)	0.29	1 (4%) 35 24	151, 159, 167, 178	0
4	D	19/22 (86%)	-0.52	0 100 100	75, 100, 127, 136	0
4	H	19/22 (86%)	0.03	0 100 100	149, 159, 170, 171	0
All	All	2018/2096 (96%)	0.17	110 (5%) 25 17	26, 122, 162, 178	0

The worst 5 of 110 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	231	GLY	8.6
1	E	131	THR	6.4
1	E	327	ALA	5.8
1	E	25	PRO	5.2
1	E	27	THR	5.1

### 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
4	ATM	H	823	22/23	0.81	0.32	155,157,158,158	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
4	MRG	H	817	26/27	0.86	0.26	154,157,167,167	0
4	MRG	D	817	26/27	0.93	0.18	89,93,101,108	0
4	ATM	D	823	22/23	0.95	0.28	87,90,92,94	0

### 6.3 Carbohydrates [i](#)

There are no monosaccharides 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.