



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 8, 2018 – 08:42 pm GMT

PDB ID : 1J5O  
Title : CRYSTAL STRUCTURE OF MET184ILE MUTANT OF HIV-1 REVERSE TRANSCRIPTASE IN COMPLEX WITH DOUBLE STRANDED DNA TEMPLATE-PRIMER  
Authors : Sarafianos, S.G.; Das, K.; Arnold, E.  
Deposited on : 2002-05-24  
Resolution : 3.50 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

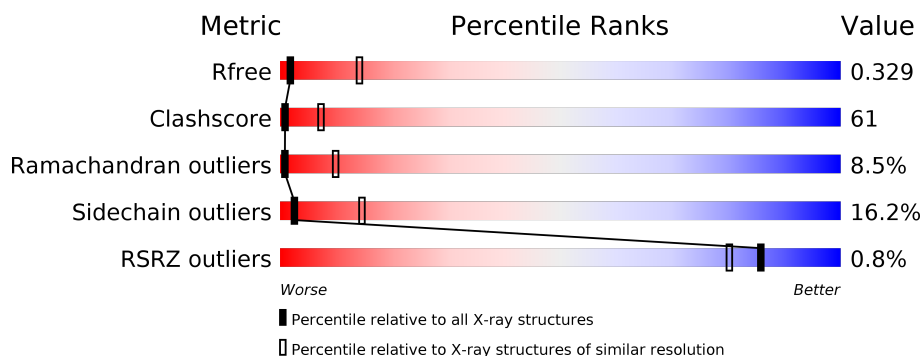
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.50 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	1391 (3.60-3.40)
Clashscore	122126	1485 (3.60-3.40)
Ramachandran outliers	120053	1446 (3.60-3.40)
Sidechain outliers	120020	1447 (3.60-3.40)
RSRZ outliers	108989	1303 (3.60-3.40)

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	T	19	<div> <div>11%</div> <div>5% 95%</div> </div>
2	P	18	<div> <div>6%</div> <div>89% 11%</div> </div>
3	A	558	<div> <div>28% 56% 14%</div> </div>
4	B	430	<div> <div>% 25% 56% 17%</div> </div>
5	L	214	<div> <div>% 25% 59% 15%</div> </div>
6	H	220	<div> <div>27% 61% 11%</div> </div>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 11720 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 DNA chain called 5'-D(\*AP\*TP\*GP\*GP\*CP\*GP\*CP\*CP\*GP\*AP\*AP\*CP\*AP\*GP\*GP\*GP\*AP\*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	T	19	Total	C	N	O	P	0	0	0
			390	184	80	108	18			

- Molecule 2 is a DNA chain called 5'-D(\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*GP\*CP\*GP\*CP\*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	P	18	Total	C	N	O	P	0	0	0
			363	173	64	109	17			

- Molecule 3 is a protein called Reverse transcriptase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	558	Total	C	N	O	S	0	0	0
			4292	2778	713	795	6			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	184	ILE	MET	ENGINEERED	UNP P03366
A	280	SER	CYS	ENGINEERED	UNP P03366

- Molecule 4 is a protein called Reverse transcriptase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	B	430	Total	C	N	O	S	0	0	0
			3411	2217	568	620	6			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	184	ILE	MET	ENGINEERED	UNP P03366
B	280	SER	CYS	ENGINEERED	UNP P03366

- Molecule 5 is a protein called ANTIBODY (LIGHT CHAIN).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	L	214	Total	C	N	O	S	0	0	0
			1616	1010	256	343	7			

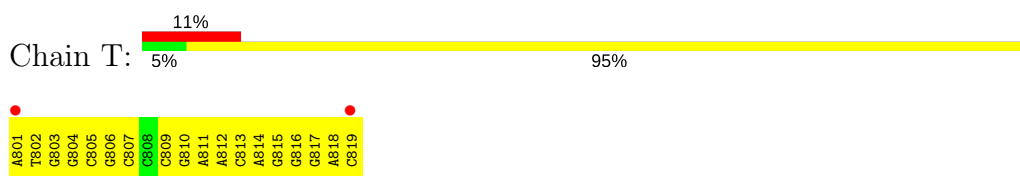
- Molecule 6 is a protein called ANTIBODY (HEAVY CHAIN).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	H	220	Total	C	N	O	S	0	0	0
			1648	1037	270	333	8			

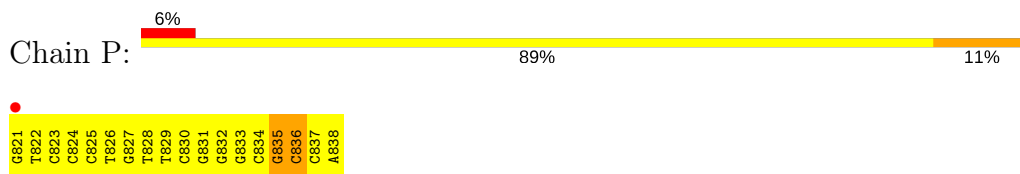
### 3 Residue-property plots

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.

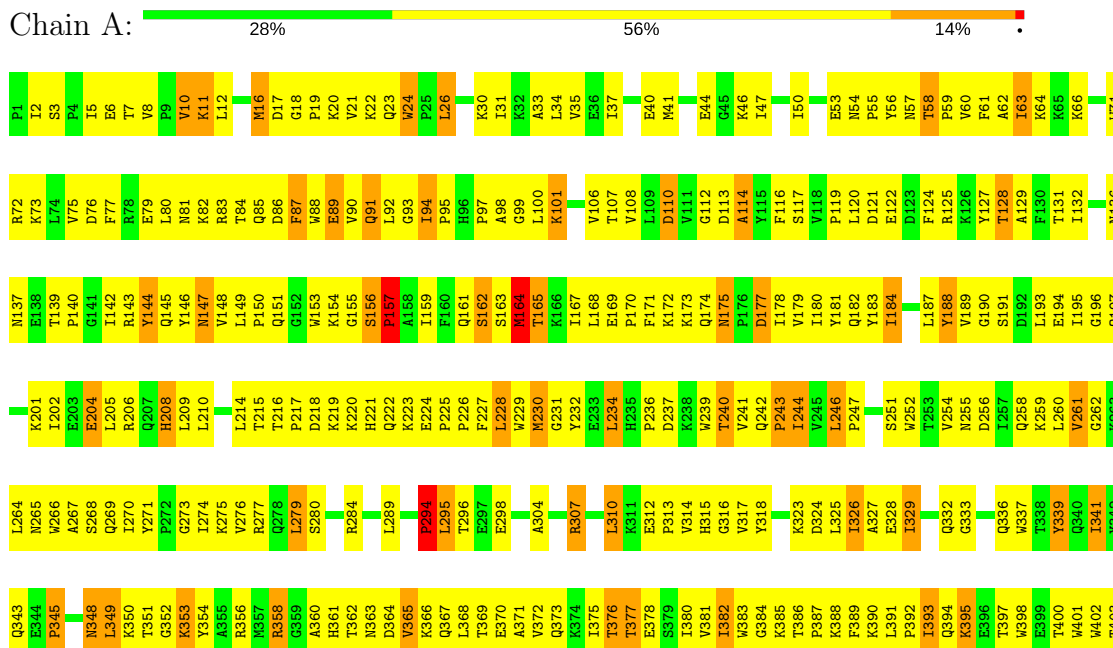
- Molecule 1: 5'-D(\*AP\*TP\*GP\*GP\*CP\*GP\*CP\*CP\*GP\*AP\*AP\*CP\*AP\*GP\*GP\*GP\*AP\*C)-3'

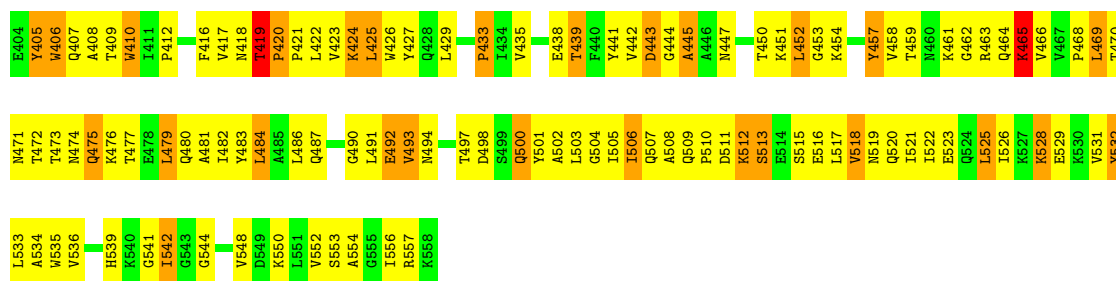


- Molecule 2: 5'-D(\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*GP\*CP\*GP\*CP\*CP\*A)-3'

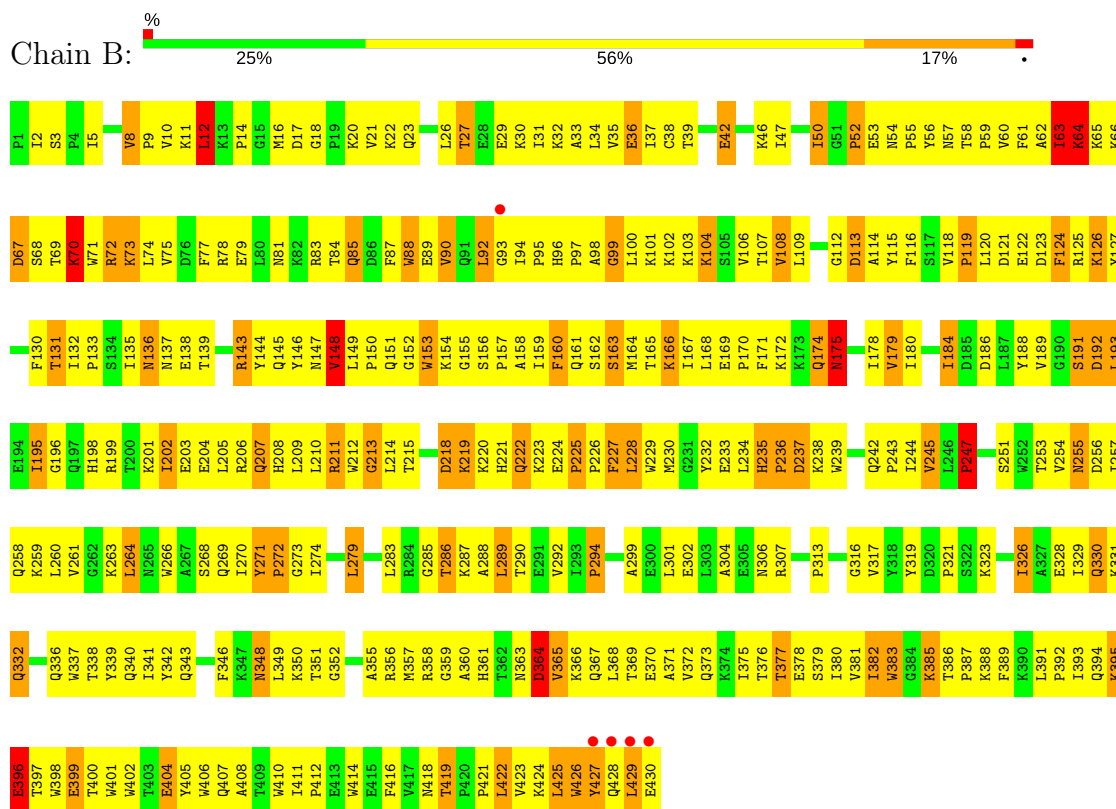


- Molecule 3: Reverse transcriptase

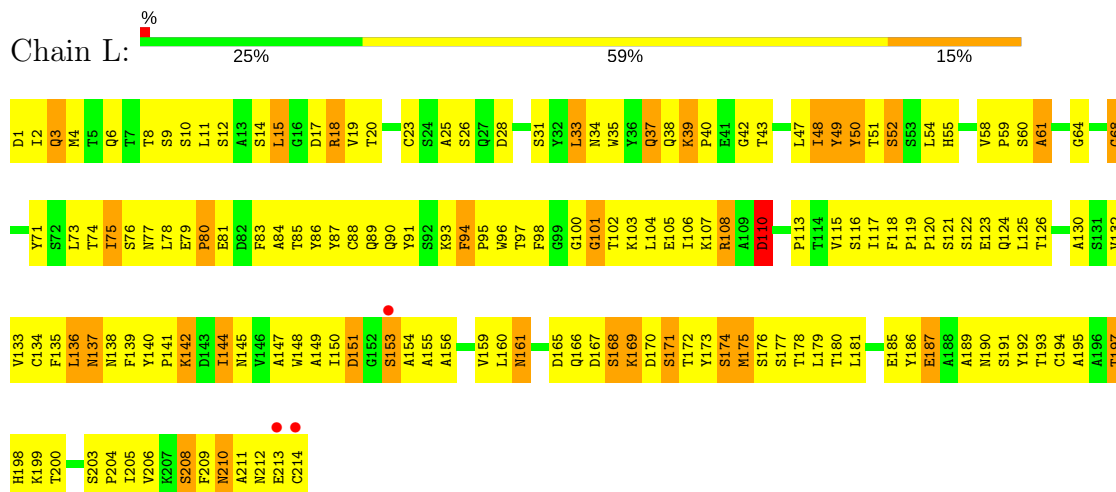




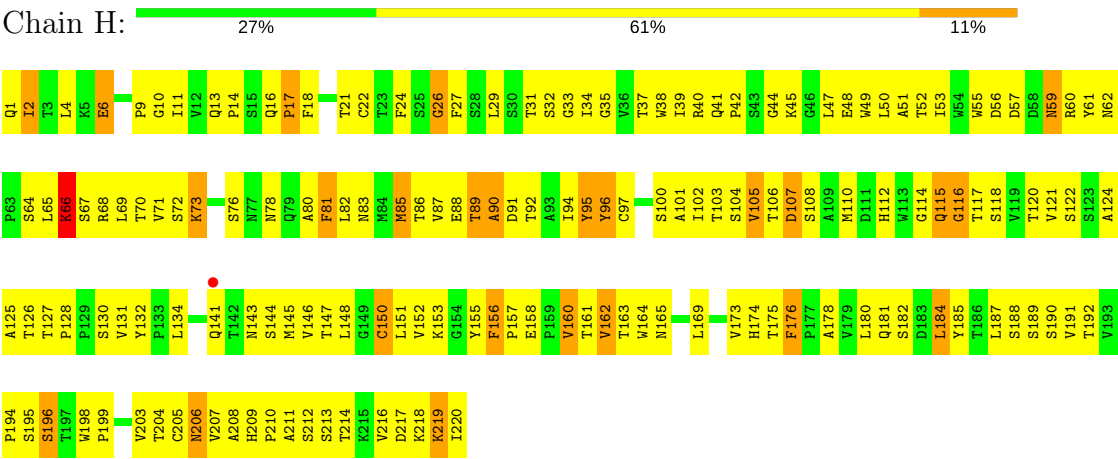
### • Molecule 4: Reverse transcriptase



### • Molecule 5: ANTIBODY (LIGHT CHAIN)



● Molecule 6: ANTIBODY (HEAVY CHAIN)



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32 1 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	169.18Å 169.18Å 221.91Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	10.00 – 3.50 39.97 – 3.50	Depositor EDS
% Data completeness (in resolution range)	91.5 (10.00-3.50) 91.4 (39.97-3.50)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.71 (at 3.48Å)	Xtriage
Refinement program	X-PLOR 3.843	Depositor
R, $R_{free}$	0.262 , 0.338 0.259 , 0.329	Depositor DCC
$R_{free}$ test set	1272 reflections (3.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	98.8	Xtriage
Anisotropy	0.181	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 114.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.049 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.85	EDS
Total number of atoms	11720	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.19% 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

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	T	0.55	0/439	0.76	0/676
2	P	0.56	1/405 (0.2%)	0.81	0/623
3	A	0.56	0/4404	0.78	2/6017 (0.0%)
4	B	0.63	0/3510	0.84	4/4784 (0.1%)
5	L	0.54	0/1654	0.85	0/2256
6	H	0.58	0/1691	0.85	0/2320
All	All	0.58	1/12103 (0.0%)	0.82	6/16676 (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
2	P	0	1
3	A	0	1
All	All	0	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	P	835	DG	O3'-P	-5.99	1.53	1.61

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	247	PRO	N-CA-CB	6.75	111.39	103.30
4	B	243	PRO	N-CA-CB	6.31	110.87	103.30
3	A	294	PRO	N-CA-CB	5.91	110.39	103.30
3	A	247	PRO	N-CA-CB	5.63	110.06	103.30
4	B	313	PRO	N-CA-CB	5.37	109.74	103.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	144	TYR	Sidechain
2	P	836	DC	Sidechain

## 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	T	390	0	212	55	0
2	P	363	0	204	103	0
3	A	4292	0	4127	482	0
4	B	3411	0	3300	446	0
5	L	1616	0	1517	174	0
6	H	1648	0	1602	184	0
All	All	11720	0	10962	1380	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 61.

The worst 5 of 1380 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:P:822:DT:H2''	2:P:823:DC:C6	1.75	1.22
1:T:803:DG:H2''	1:T:804:DG:C5'	1.77	1.14
2:P:831:DG:H2''	2:P:832:DG:H5'	1.17	1.11
3:A:20:LYS:HA	3:A:57:ASN:H	1.15	1.10
3:A:333:GLY:H	3:A:336:GLN:HB2	0.99	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	
3	A	556/558 (100%)	394 (71%)	121 (22%)	41 (7%)	1	12
4	B	428/430 (100%)	284 (66%)	96 (22%)	48 (11%)	0	6
5	L	212/214 (99%)	162 (76%)	35 (16%)	15 (7%)	1	13
6	H	218/220 (99%)	160 (73%)	42 (19%)	16 (7%)	1	13
All	All	1414/1422 (99%)	1000 (71%)	294 (21%)	120 (8%)	1	10

5 of 120 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	63	ILE
3	A	66	LYS
3	A	136	ASN
3	A	195	ILE
3	A	223	LYS

### 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	
3	A	434/498 (87%)	364 (84%)	70 (16%)	2	15
4	B	350/392 (89%)	290 (83%)	60 (17%)	2	12
5	L	182/182 (100%)	148 (81%)	34 (19%)	1	8
6	H	191/191 (100%)	168 (88%)	23 (12%)	5	26
All	All	1157/1263 (92%)	970 (84%)	187 (16%)	2	14

5 of 187 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	B	73	LYS
4	B	211	ARG

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
6	H	81	PHE
4	B	104	LYS
4	B	163	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 36 such sidechains are listed below:

Mol	Chain	Res	Type
4	B	207	GLN
4	B	306	ASN
6	H	62	ASN
4	B	222	GLN
4	B	330	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](#)

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	T	19/19 (100%)	0.55	2 (10%) 6 7	59, 76, 94, 96	0
2	P	18/18 (100%)	0.12	1 (5%) 24 21	61, 74, 86, 87	0
3	A	558/558 (100%)	-0.34	0 100 100	11, 52, 75, 92	0
4	B	430/430 (100%)	-0.45	5 (1%) 79 72	6, 32, 75, 93	0
5	L	214/214 (100%)	-0.35	3 (1%) 75 69	17, 44, 66, 76	0
6	H	220/220 (100%)	-0.46	1 (0%) 90 87	14, 33, 57, 68	0
All	All	1459/1459 (100%)	-0.38	12 (0%) 86 80	6, 43, 75, 96	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	B	429	LEU	5.5
5	L	214	CYS	5.4
1	T	801	DA	3.8
4	B	427	TYR	3.5
4	B	93	GLY	3.4

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.