



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 3, 2023 – 07:46 AM EDT

PDB ID : 1HYS  
Title : CRYSTAL STRUCTURE OF HIV-1 REVERSE TRANSCRIPTASE IN COMPLEX WITH A POLYPURINE TRACT RNA:DNA  
Authors : Sarafianos, S.G.; Das, K.; Tantillo, C.; Clark Jr., A.D.; Ding, J.; Whitcomb, J.; Boyer, P.L.; Hughes, S.H.; Arnold, E.  
Deposited on : 2001-01-22  
Resolution : 3.00 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  
Xtriage (Phenix) : 1.13  
EDS : 2.34  
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) : 2.34

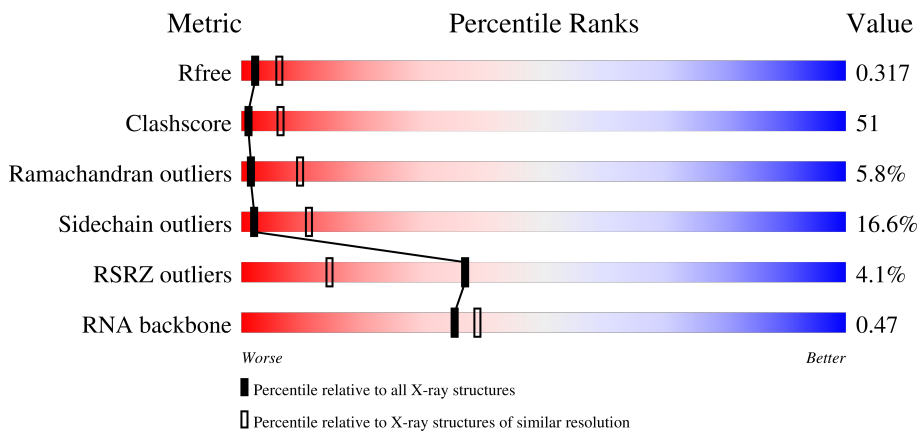
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

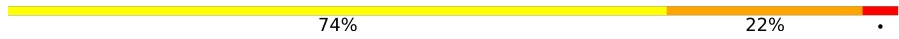



The reported resolution of this entry is 3.00 Å.

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.



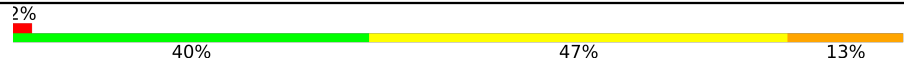

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)
RNA backbone	3102	1173 (3.30-2.70)

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 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	E	23	 74% 22% .
2	F	22	 86% 14%
3	A	553	 7% 33% 52% 15% .
4	B	425	 4% 30% 55% 14% .

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
5	C	214	 2% 40% 47% 13%
6	D	220	 33% 54% 13%

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 12139 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 RNA chain called 5'-R(\*UP\*CP\*AP\*GP\*CP\*CP\*AP\*CP\*UP\*UP\*UP\*UP\*UP\*AP\*AP\*AP\*AP\*AP\*GP\*AP\*AP\*AP\*AP\*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	E	23	486	220	89	155	22	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	F	22	447	217	74	135	21	0	0	0

- Molecule 3 is a protein called HIV-1 REVERSE TRANSCRIPTASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	553	4465	2894	739	825	7	249	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	280	SER	CYS	engineered mutation	UNP P03366

- Molecule 4 is a protein called HIV-1 REVERSE TRANSCRIPTASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	B	425	3481	2265	578	632	6	77	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	engineered mutation	UNP P03366

- Molecule 5 is a protein called FAB-28 MONOCLONAL ANTIBODY FRAGMENT LIGHT CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	C	214	1612	1008	256	341	7	0	0	0

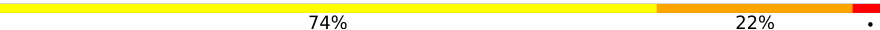
- Molecule 6 is a protein called FAB-28 MONOCLONAL ANTIBODY FRAGMENT HEAVY CHAIN.

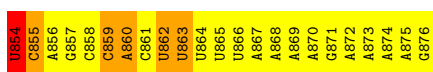
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	D	220	1648	1037	270	333	8	5	0	0

### 3 Residue-property plots [i](#)

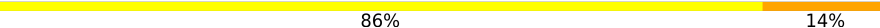
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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'-R(\*UP\*CP\*AP\*GP\*CP\*CP\*AP\*CP\*UP\*UP\*UP\*UP\*UP\*AP\*AP\*AP\*AP\*GP\*P\*AP\*AP\*AP\*AP\*G)-3'

Chain E:  74% 22%



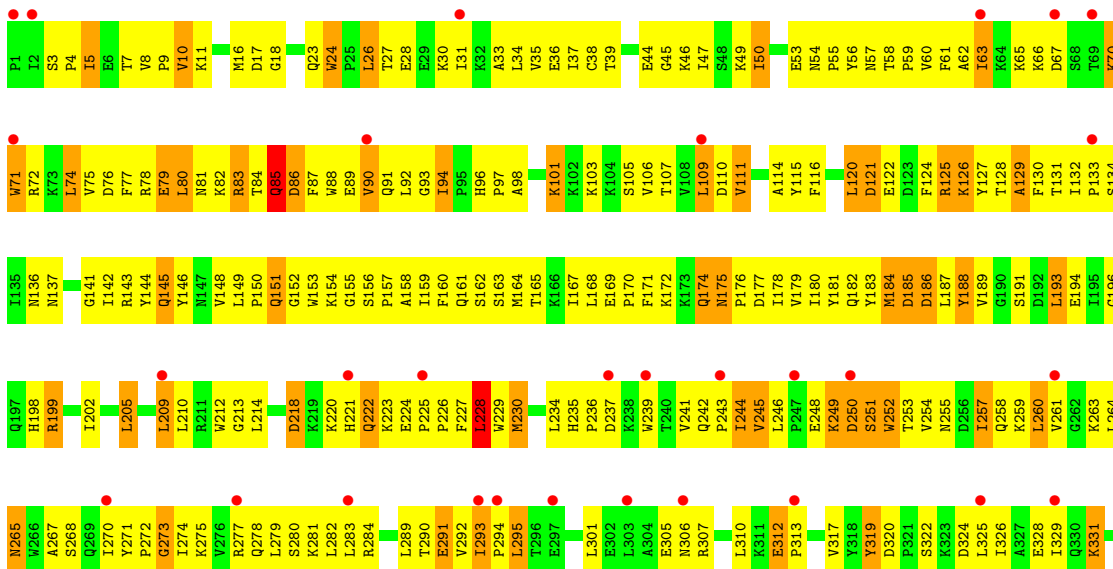
- Molecule 2: 5'-D(\*CP\*TP\*TP\*TP\*TP\*CP\*TP\*TP\*TP\*TP\*AP\*AP\*AP\*AP\*AP\*GP\*TP\*GP\*GP\*CP\*TP\*G)-3'

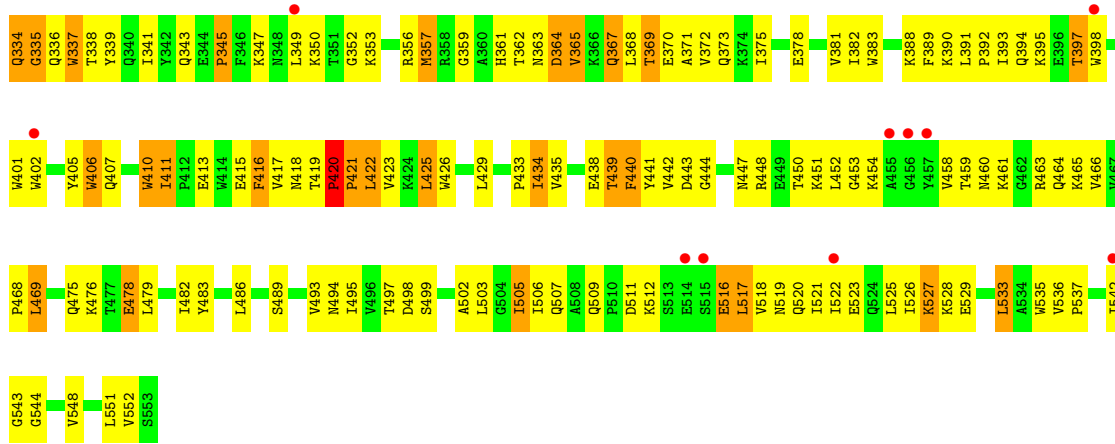
Chain F:  86% 14%



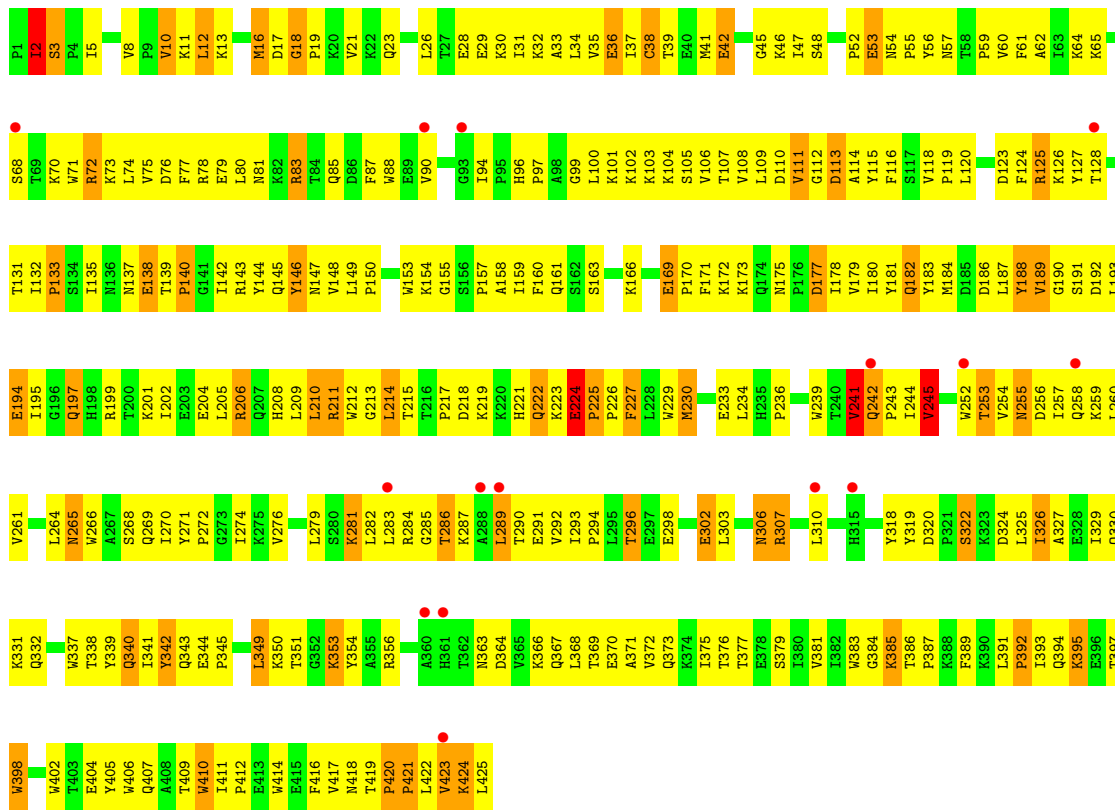
- Molecule 3: HIV-1 REVERSE TRANSCRIPTASE

Chain A:  7% 33% 52% 15%

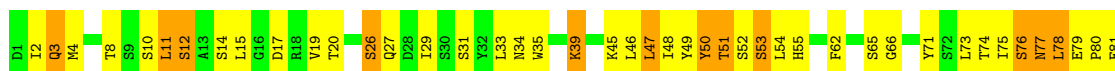


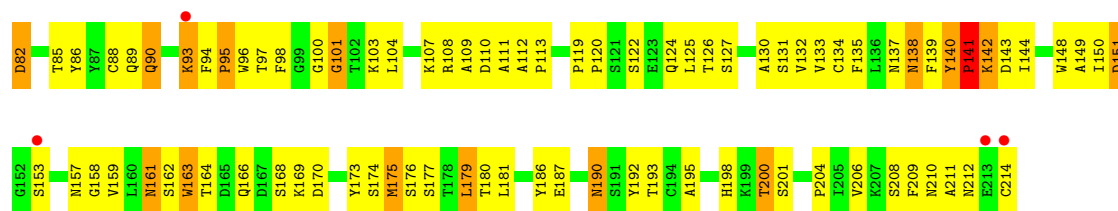


• Molecule 4: HIV-1 REVERSE TRANSCRIPTASE



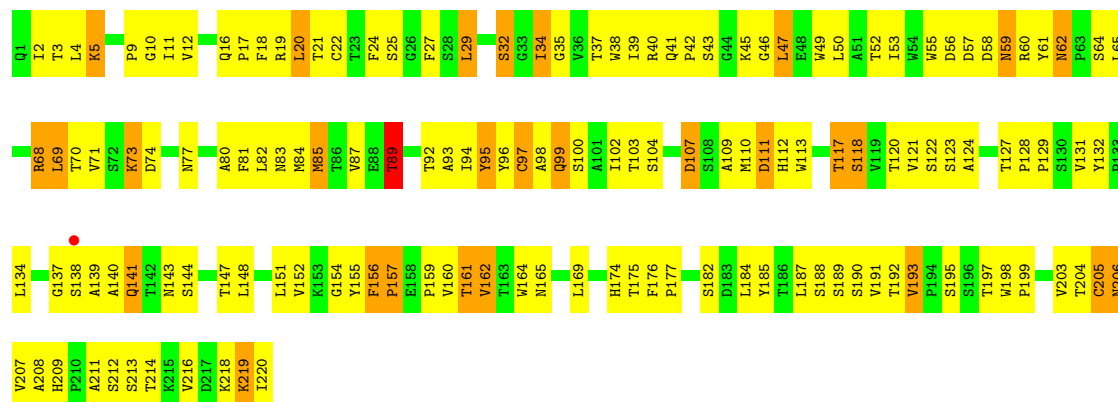
• Molecule 5: FAB-28 MONOCLONAL ANTIBODY FRAGMENT LIGHT CHAIN





● Molecule 6: FAB-28 MONOCLONAL ANTIBODY FRAGMENT HEAVY CHAIN

Chain D: 33% 54% 13%





## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 32 1 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	166.16Å 166.16Å 218.81Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	8.00 – 3.00 24.98 – 3.00	Depositor EDS
% Data completeness (in resolution range)	74.1 (8.00-3.00) 75.2 (24.98-3.00)	Depositor EDS
$R_{merge}$	0.15	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.35 (at 2.99Å)	Xtrriage
Refinement program	X-PLOR 3.843	Depositor
R, $R_{free}$	0.274 , 0.316 0.274 , 0.317	Depositor DCC
$R_{free}$ test set	1286 reflections (2.21%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	55.8	Xtrriage
Anisotropy	0.030	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 14.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.25$	Xtrriage
Estimated twinning fraction	0.058 for -h,-k,l	Xtrriage
$F_o, F_c$ correlation	0.81	EDS
Total number of atoms	12139	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	41.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.68% 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	E	0.76	1/544 (0.2%)	1.11	7/845 (0.8%)
2	F	0.66	0/499	1.05	5/769 (0.7%)
3	A	0.55	0/4583	0.80	6/6232 (0.1%)
4	B	0.79	2/3583 (0.1%)	1.10	8/4871 (0.2%)
5	C	0.55	0/1650	0.80	1/2251 (0.0%)
6	D	0.64	0/1691	0.88	0/2320
All	All	0.65	3/12550 (0.0%)	0.93	27/17288 (0.2%)

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
4	B	0	5
6	D	0	1
All	All	0	6

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	211	ARG	CZ-NH1	22.86	1.62	1.33
1	E	854	U	C5-C6	7.03	1.40	1.34
4	B	211	ARG	CZ-NH2	-5.46	1.25	1.33

The worst 5 of 27 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	211	ARG	NE-CZ-NH2	34.23	137.41	120.30
4	B	211	ARG	NE-CZ-NH1	-28.40	106.10	120.30
4	B	211	ARG	NH1-CZ-NH2	-15.12	102.77	119.40
1	E	854	U	N1-C1'-C2'	8.67	125.28	114.00
2	F	892	DA	N9-C1'-C2'	8.29	128.36	112.60

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	B	146	TYR	Sidechain
4	B	183	TYR	Sidechain
4	B	188	TYR	Sidechain
4	B	211	ARG	Sidechain
4	B	342	TYR	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	E	486	0	249	75	0
2	F	447	0	254	98	0
3	A	4465	0	4497	436	0
4	B	3481	0	3505	336	0
5	C	1612	0	1516	144	0
6	D	1648	0	1602	176	0
All	All	12139	0	11623	1181	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 51.

The worst 5 of 1181 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)
3:A:439:THR:HG21	4:B:289:LEU:H	1.15	1.10
2:F:878:DC:H1'	2:F:879:DT:H5''	1.35	1.08
2:F:897:DC:H2''	2:F:898:DT:H5'	1.28	1.08
1:E:868:A:H2'	1:E:869:A:C8	1.89	1.08
2:F:884:DT:C2'	2:F:885:DT:H71	1.82	1.07

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	551/553 (100%)	431 (78%)	94 (17%)	26 (5%)	2	14
4	B	423/425 (100%)	319 (75%)	73 (17%)	31 (7%)	1	5
5	C	212/214 (99%)	171 (81%)	25 (12%)	16 (8%)	1	5
6	D	218/220 (99%)	181 (83%)	29 (13%)	8 (4%)	3	19
All	All	1404/1412 (99%)	1102 (78%)	221 (16%)	81 (6%)	1	10

5 of 81 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	63	ILE
3	A	245	VAL
3	A	248	GLU
3	A	273	GLY
3	A	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	
3	A	486/495 (98%)	399 (82%)	87 (18%)	2	9
4	B	381/387 (98%)	325 (85%)	56 (15%)	3	15
5	C	181/181 (100%)	149 (82%)	32 (18%)	2	9
6	D	191/191 (100%)	160 (84%)	31 (16%)	2	12
All	All	1239/1254 (99%)	1033 (83%)	206 (17%)	2	11

5 of 206 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	B	221	HIS
5	C	3	GLN
6	D	162	VAL
4	B	244	ILE
4	B	318	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 41 such sidechains are listed below:

Mol	Chain	Res	Type
4	B	418	ASN
6	D	41	GLN
5	C	38	GLN
5	C	190	ASN
6	D	99	GLN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	E	23/23 (100%)	1 (4%)	1 (4%)

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	E	855	C

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	E	854	U

## 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 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	E	23/23 (100%)	-0.36	0 <a href="#">100</a> <a href="#">100</a>	62, 66, 76, 78	0
2	F	22/22 (100%)	-0.18	0 <a href="#">100</a> <a href="#">100</a>	52, 70, 80, 86	0
3	A	553/553 (100%)	0.48	40 (7%) <a href="#">15</a> <a href="#">4</a>	10, 51, 75, 87	61 (11%)
4	B	425/425 (100%)	0.34	15 (3%) <a href="#">44</a> <a href="#">18</a>	8, 25, 69, 83	20 (4%)
5	C	214/214 (100%)	0.27	4 (1%) <a href="#">66</a> <a href="#">37</a>	9, 37, 66, 75	0
6	D	220/220 (100%)	0.23	1 (0%) <a href="#">91</a> <a href="#">75</a>	8, 27, 49, 70	2 (0%)
All	All	1457/1457 (100%)	0.35	60 (4%) <a href="#">37</a> <a href="#">14</a>	8, 39, 71, 87	83 (5%)

The worst 5 of 60 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	2	ILE	5.8
3	A	456	GLY	5.7
4	B	315	HIS	5.5
5	C	213	GLU	4.7
3	A	542	ILE	4.1

### 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 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.