



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 13, 2018 – 11:51 pm GMT

PDB ID : 3S3M  
Title : Crystal structure of the Prototype Foamy Virus (PFV) intasome in complex with magnesium and Dolutegravir (S/GSK1349572)  
Authors : Hare, S.; Cherepanov, P.  
Deposited on : 2011-05-18  
Resolution : 2.49 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  
Mogul : 1.7.3 (157068), CSD as539be (2018)  
Xtriage (Phenix) : 1.13  
EDS : trunk31020  
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) : trunk31020

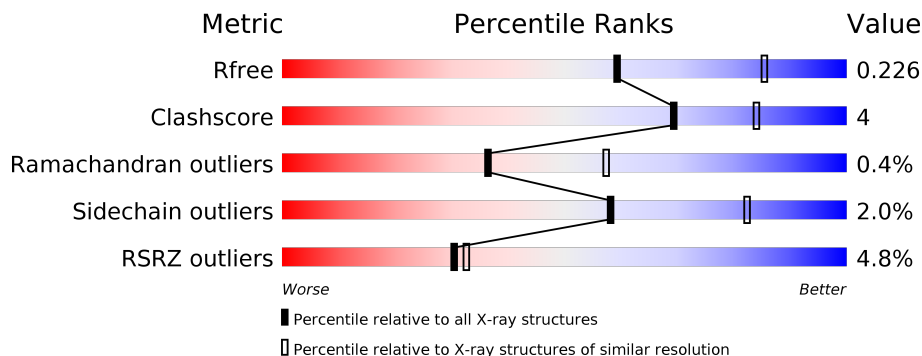
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.49 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	4155 (2.50-2.50)
Clashscore	122126	4827 (2.50-2.50)
Ramachandran outliers	120053	4735 (2.50-2.50)
Sidechain outliers	120020	4737 (2.50-2.50)
RSRZ outliers	108989	4058 (2.50-2.50)

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 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	395	<div> <div>3%</div> <div>82%</div> <div>11%</div> <div>7%</div> </div>
1	B	395	<div> <div>4%</div> <div>42%</div> <div>53%</div> </div>
2	C	19	<div> <div>37%</div> <div>42%</div> <div>21%</div> </div>
3	D	17	<div> <div>59%</div> <div>35%</div> <div>6%</div> </div>

## 2 Entry composition

There are 11 unique types of molecules in this entry. The entry contains 5424 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 PFV integrase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	368	Total	C	N	O	S	0	1	0
			2940	1885	518	533	4			
1	B	184	Total	C	N	O	S	0	0	0
			1441	935	234	271	1			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	EXPRESSION TAG	UNP P14350
A	-1	PRO	-	EXPRESSION TAG	UNP P14350
A	0	GLY	-	EXPRESSION TAG	UNP P14350
A	217	SER	GLY	VARIANT	UNP P14350
A	218	GLY	SER	VARIANT	UNP P14350
B	-2	GLY	-	EXPRESSION TAG	UNP P14350
B	-1	PRO	-	EXPRESSION TAG	UNP P14350
B	0	GLY	-	EXPRESSION TAG	UNP P14350
B	217	SER	GLY	VARIANT	UNP P14350
B	218	GLY	SER	VARIANT	UNP P14350

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	19	Total	C	N	O	P	0	0	0
			387	187	68	114	18			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	17	Total	C	N	O	P	0	0	0
			345	166	65	98	16			

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Zn	0	0
			1	1		

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



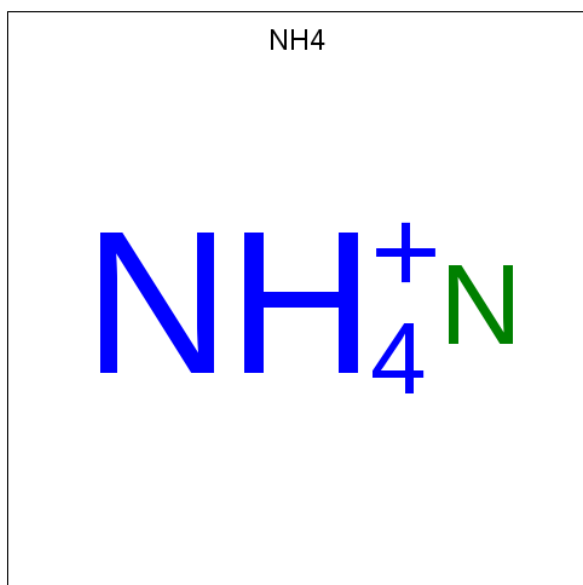
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	O	S	0	0
			5	4	1		
5	B	1	Total	O	S	0	0
			5	4	1		

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	C	O	0	0
			6	3	3		
6	A	1	Total	C	O	0	0
			6	3	3		
6	A	1	Total	C	O	0	0
			6	3	3		
6	A	1	Total	C	O	0	0
			6	3	3		

- Molecule 7 is AMMONIUM ION (three-letter code: NH4) (formula: H<sub>4</sub>N).

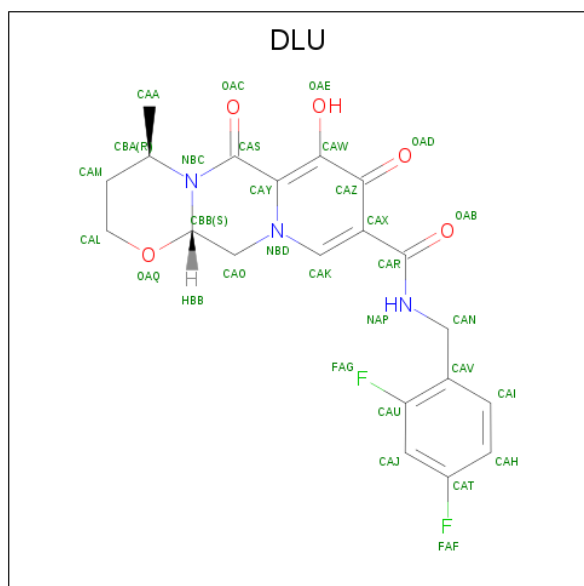


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	1	Total N 1 1	0	0

- Molecule 8 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

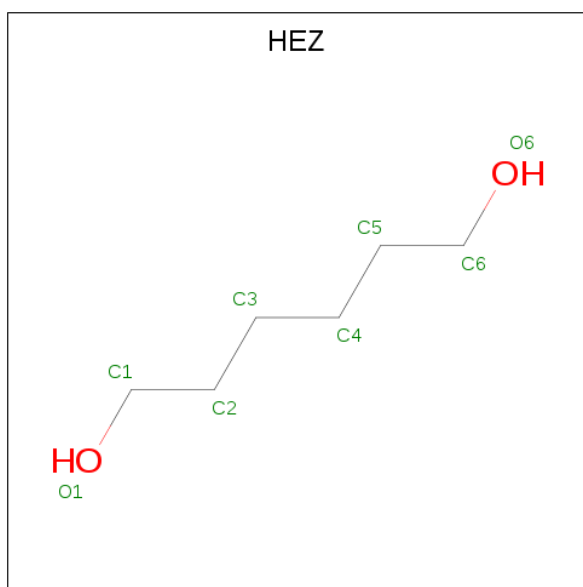
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	A	2	Total Mg 2 2	0	0

- Molecule 9 is (4R,12aS)-N-(2,4-difluorobenzyl)-7-hydroxy-4-methyl-6,8-dioxo-3,4,6,8,12,12a-hexahydro-2H-pyrido[1',2':4,5]pyrazino[2,1-b][1,3]oxazine-9-carboxamide (three-letter code: DLU) (formula: C<sub>20</sub>H<sub>19</sub>F<sub>2</sub>N<sub>3</sub>O<sub>5</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
9	A	1	Total 30	C 20	F 2	N 3	O 5	0	0

- Molecule 10 is HEXANE-1,6-DIOL (three-letter code: HEZ) (formula:  $C_6H_{14}O_2$ ).



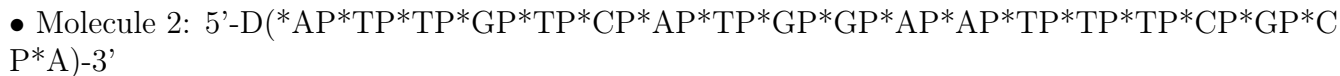
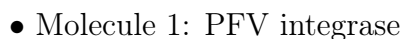
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
10	B	1	Total	C	O	0	0
			8	6	2		

- Molecule 11 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
11	A	150	Total	O	0	0
			150	150		
11	B	36	Total	O	0	0
			36	36		
11	C	30	Total	O	0	0
			30	30		
11	D	19	Total	O	0	0
			19	19		



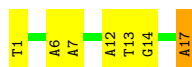
- Molecule 1: PFV integrase





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

Chain D: 59% 35% 6%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	160.14Å 160.14Å 123.43Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.15 – 2.49 39.15 – 2.49	Depositor EDS
% Data completeness (in resolution range)	98.3 (39.15-2.49) 98.3 (39.15-2.49)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.83 (at 2.48Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.206 , 0.232 0.200 , 0.226	Depositor DCC
$R_{free}$ test set	2832 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	57.4	Xtriage
Anisotropy	0.326	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 40.1	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	0.95	EDS
Total number of atoms	5424	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.79% 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: GOL, ZN, MG, HEZ, NH4, SO4, DLU

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	0.56	0/3022	0.64	0/4121
1	B	0.59	0/1481	0.59	0/2025
2	C	0.99	0/433	1.75	13/667 (1.9%)
3	D	1.15	0/387	1.78	8/595 (1.3%)
All	All	0.67	0/5323	0.93	21/7408 (0.3%)

There are no bond length outliers.

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	8	DT	O4'-C4'-C3'	-8.66	100.81	106.00
2	C	5	DT	O4'-C1'-N1	-8.46	102.08	108.00
3	D	1	DT	P-O3'-C3'	8.27	129.63	119.70
2	C	1	DA	P-O3'-C3'	7.69	128.93	119.70
3	D	17	DA	O4'-C1'-N9	7.13	112.99	108.00

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	2940	0	2986	26	0
1	B	1441	0	1421	8	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	387	0	218	4	0
3	D	345	0	193	3	0
4	A	1	0	0	0	0
5	A	5	0	0	0	0
5	B	5	0	0	0	0
6	A	24	0	32	3	0
7	A	1	0	0	0	0
8	A	2	0	0	0	0
9	A	30	0	18	0	0
10	B	8	0	14	0	0
11	A	150	0	0	0	1
11	B	36	0	0	1	0
11	C	30	0	0	1	0
11	D	19	0	0	0	0
All	All	5424	0	4882	40	1

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 4.

The worst 5 of 40 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:A:97:GLN:HE22	1:A:339:LYS:HG2	1.28	0.96
1:A:97:GLN:NE2	1:A:339:LYS:HG2	1.90	0.85
1:A:137:GLN:H	6:A:802:GOL:H2	1.43	0.83
1:B:258:SER:O	1:B:261:LEU:O	2.03	0.76
1:B:215:GLN:NE2	11:B:427:HOH:O	2.23	0.72

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
11:A:504:HOH:O	11:A:525:HOH:O[8_554]	2.07	0.13

## 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	367/395 (93%)	354 (96%)	12 (3%)	1 (0%)	43	64
1	B	182/395 (46%)	171 (94%)	10 (6%)	1 (0%)	31	51
All	All	549/790 (70%)	525 (96%)	22 (4%)	2 (0%)	36	57

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	10	ALA
1	B	298	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	331/354 (94%)	326 (98%)	5 (2%)	67	87
1	B	159/354 (45%)	154 (97%)	5 (3%)	43	70
All	All	490/708 (69%)	480 (98%)	10 (2%)	58	82

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	164	SER
1	B	149	MET
1	B	216	SER
1	A	149	MET

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	164	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	97	GLN
1	A	304	HIS

### 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](#)

Of 12 ligands modelled in this entry, 3 are monoatomic and 1 is modelled with single atom - leaving 8 for Mogul analysis.

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	$\# Z  > 2$
5	SO4	A	394	-	4,4,4	0.18	0	6,6,6	0.27	0
9	DLU	A	398	8	26,33,33	1.12	2 (7%)	34,49,49	2.05	13 (38%)
6	GOL	A	801	-	5,5,5	0.46	0	5,5,5	0.57	0
6	GOL	A	802	-	5,5,5	0.53	0	5,5,5	0.54	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
6	GOL	A	803	-	5,5,5	0.36	0	5,5,5	0.35	0
6	GOL	A	805	-	5,5,5	0.41	0	5,5,5	0.23	0
5	SO4	B	393	-	4,4,4	0.18	0	6,6,6	0.31	0
10	HEZ	B	394	-	7,7,7	0.40	0	6,6,6	0.30	0

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
5	SO4	A	394	-	-	0/0/0/0	0/0/0/0
9	DLU	A	398	8	-	0/9/35/35	0/3/4/4
6	GOL	A	801	-	-	0/4/4/4	0/0/0/0
6	GOL	A	802	-	-	0/4/4/4	0/0/0/0
6	GOL	A	803	-	-	0/4/4/4	0/0/0/0
6	GOL	A	805	-	-	0/4/4/4	0/0/0/0
5	SO4	B	393	-	-	0/0/0/0	0/0/0/0
10	HEZ	B	394	-	-	0/5/5/5	0/0/0/0

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
9	A	398	DLU	CAO-CBB	-2.84	1.47	1.51
9	A	398	DLU	CAV-CAU	2.09	1.41	1.38

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
9	A	398	DLU	CAL-OAQ-CBB	-3.57	105.46	110.84
9	A	398	DLU	CAM-CBA-NBC	-3.19	106.22	109.41
9	A	398	DLU	CAO-NBD-CAK	-3.00	116.03	118.21
9	A	398	DLU	OAB-CAR-CAX	-2.97	117.52	121.72
9	A	398	DLU	CAK-NBD-CAY	-2.55	120.57	122.36

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	A	801	GOL	1	0
6	A	802	GOL	2	0

## 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	368/395 (93%)	0.02	12 (3%) 46 50	36, 51, 99, 131	0
1	B	184/395 (46%)	0.08	16 (8%) 10 10	43, 60, 138, 152	0
2	C	19/19 (100%)	-0.43	0 100 100	39, 57, 79, 101	0
3	D	17/17 (100%)	-0.50	0 100 100	46, 49, 82, 99	0
All	All	588/826 (71%)	0.01	28 (4%) 30 32	36, 54, 113, 152	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	8	LEU	5.8
1	B	237	GLY	4.9
1	B	280	ASN	4.7
1	B	279	ALA	4.5
1	A	9	ASP	4.2

### 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](#)

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
6	GOL	A	802	6/6	0.73	0.23	67,71,74,74	0
6	GOL	A	801	6/6	0.82	0.15	73,76,76,76	0
7	NH4	A	395	1/1	0.86	0.46	37,37,37,37	0
6	GOL	A	803	6/6	0.86	0.26	74,77,77,81	0
6	GOL	A	805	6/6	0.90	0.23	90,91,91,92	0
10	HEZ	B	394	8/8	0.94	0.18	65,69,70,71	0
5	SO4	B	393	5/5	0.95	0.14	81,81,83,83	0
5	SO4	A	394	5/5	0.96	0.12	80,81,83,83	0
9	DLU	A	398	30/30	0.97	0.12	36,41,44,44	0
8	MG	A	397	1/1	0.98	0.12	39,39,39,39	0
8	MG	A	396	1/1	0.99	0.11	43,43,43,43	0
4	ZN	A	393	1/1	1.00	0.15	47,47,47,47	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.