



# wwPDB X-ray Structure Validation Summary Report

Dec 12, 2022 – 03:30 pm GMT

PDB ID : 7QFF  
Title : Peptide VACKSSQP in complex with human cathepsin V C25A mutant  
Authors : Loboda, J.; Sosnowski, P.; Tusar, L.; Vidmar, R.; Vizovisek, M.; Horvat, J.;  
Kosec, G.; Impens, F.; Demol, H.; Turk, B.; Gevaert, K.; Turk, D.  
Deposited on : 2021-12-06  
Resolution : 1.50 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 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 : 1.8.4, CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.31.3  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0267  
CCP4 : 7.1.010 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.31.3

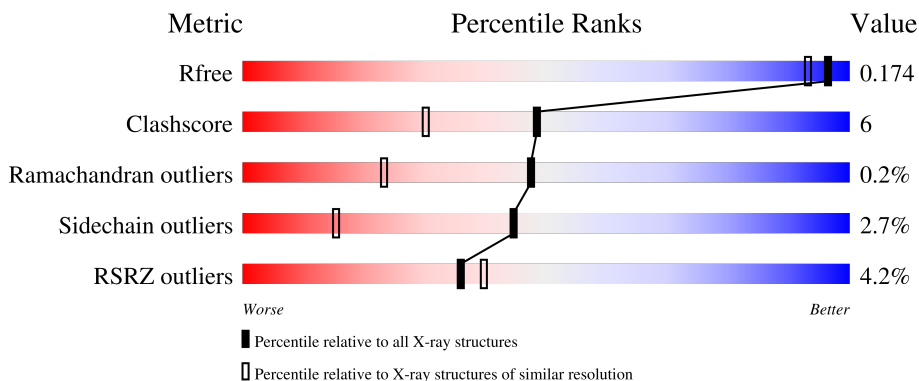
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.50 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))
$R_{free}$	130704	2936 (1.50-1.50)
Clashscore	141614	3144 (1.50-1.50)
Ramachandran outliers	138981	3066 (1.50-1.50)
Sidechain outliers	138945	3064 (1.50-1.50)
RSRZ outliers	127900	2884 (1.50-1.50)

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	AA	221	 92% 8%
1	BA	221	 5% 93% 7%
2	PA	10	 20% 40% 10% 50%
2	PB	10	 20% 20% 30% 50%
3	PAB	3	 33% 67%

## 2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 8634 atoms, of which 4482 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 Cathepsin L2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
1	AA	221	3412	1095	1674	299	334	10	1674	6	0
1	BA	221	3373	1083	1658	294	327	11	1658	4	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AA	25	ALA	CYS	engineered mutation	UNP O60911
AA	108	GLN	ASN	engineered mutation	UNP O60911
AA	179	GLN	ASN	engineered mutation	UNP O60911
BA	246	ALA	CYS	engineered mutation	UNP O60911
BA	329	GLN	ASN	engineered mutation	UNP O60911
BA	400	GLN	ASN	engineered mutation	UNP O60911

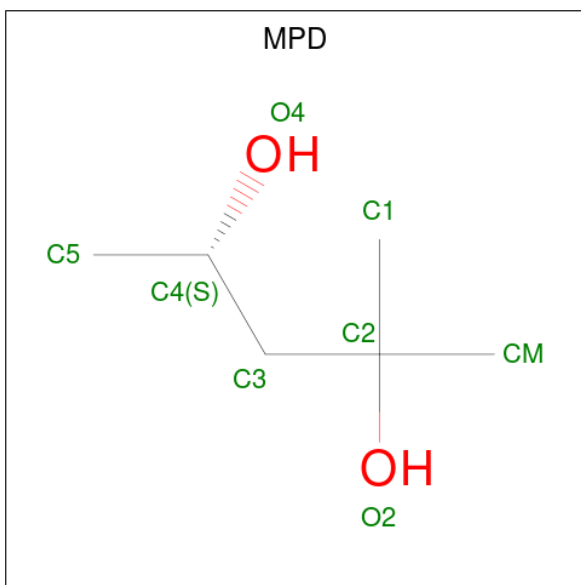
- Molecule 2 is a protein called ACE-VAL-ALA-CYS-LYS.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
2	PA	5	65	19	35	5	5	1	35	0	0
2	PB	5	65	19	35	5	5	1	35	0	0

- Molecule 3 is a protein called GLU-TYR-SER.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	H	N	O			
3	PAB	3	48	17	20	3	8	20	0	0

- Molecule 4 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula: C<sub>6</sub>H<sub>14</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	AA	1	Total C O 8 6 2	0	0
4	AA	1	Total C O 8 6 2	0	0
4	AA	1	Total C O 8 6 2	0	0
4	AA	1	Total C O 8 6 2	0	0
4	BA	1	Total C O 8 6 2	0	0
4	BA	1	Total C O 8 6 2	0	0

- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

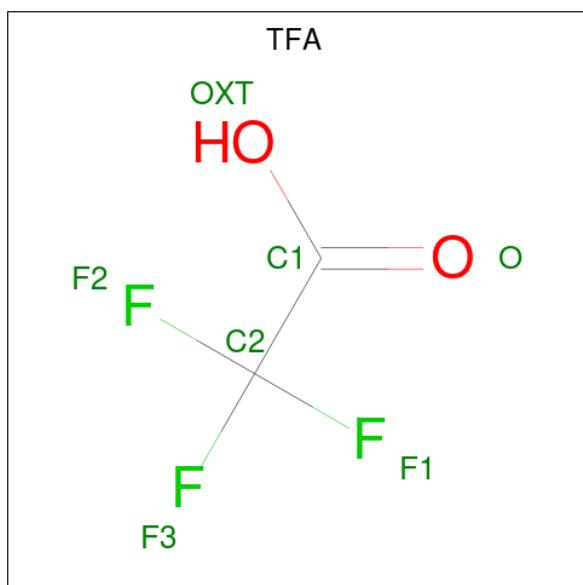
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	AA	1	Total Cl 1 1	0	0
5	BA	1	Total Cl 1 1	0	0

- 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	AA	1	Total	C	O	0	0
			6	3	3		
6	AA	1	Total	C	O	0	0
			6	3	3		
6	AA	1	Total	C	O	0	0
			6	3	3		
6	BA	1	Total	C	O	0	0
			6	3	3		

- Molecule 7 is trifluoroacetic acid (three-letter code: TFA) (formula:  $C_2HF_3O_2$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
7	BA	1	Total	C	F	O	0	0
			7	2	3	2		

- Molecule 8 is water.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
8	AA	237	Total	H	O	474	0
			711	474	237		
8	BA	289	Total	H	O	578	0
			867	578	289		
8	PA	1	Total	H	O	2	0
			3	2	1		
8	PAB	1	Total	H	O	2	0
			3	2	1		
8	PB	2	Total	H	O	4	0
			6	4	2		

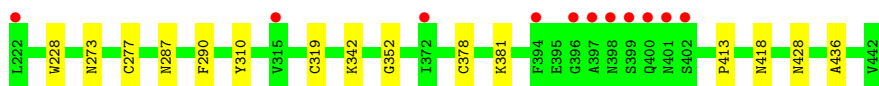
### 3 Residue-property plots [i](#)

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.

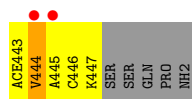
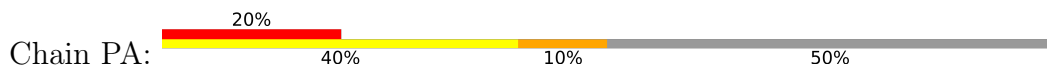
- Molecule 1: Cathepsin L2



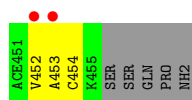
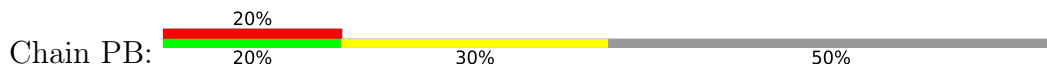
- Molecule 1: Cathepsin L2



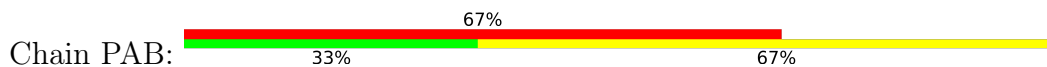
- Molecule 2: ACE-VAL-ALA-CYS-LYS



- Molecule 2: ACE-VAL-ALA-CYS-LYS



- Molecule 3: GLU-TYR-SER



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	94.40Å 94.40Å 126.83Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.20 – 1.50 47.20 – 1.50	Depositor EDS
% Data completeness (in resolution range)	100.0 (47.20-1.50) 99.9 (47.20-1.50)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.97 (at 1.50Å)	Xtrriage
Refinement program	MAIN	Depositor
R, $R_{free}$	0.169 , 0.194 0.170 , 0.174	Depositor DCC
$R_{free}$ test set	4586 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	22.1	Xtrriage
Anisotropy	0.028	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	(Not available) , (Not available)	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	8634	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	14.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.31% 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: MPD, ACE, TFA, GOL, CL

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	AA	0.71	0/1780	0.80	0/2405
1	BA	0.67	0/1761	0.77	0/2379
2	PA	0.86	0/27	1.14	0/35
2	PB	0.90	0/27	0.96	0/35
3	PAB	0.63	0/28	0.83	0/35
All	All	0.70	0/3623	0.79	0/4889

There are no bond length outliers.

There are no bond angle outliers.

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	AA	1738	1674	1667	19	0
1	BA	1715	1658	1650	9	0
2	PA	30	35	33	11	0
2	PB	30	35	35	5	0
3	PAB	28	20	17	3	0
4	AA	32	0	56	9	0
4	BA	16	0	28	2	0
5	AA	1	0	0	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	BA	1	0	0	0	0
6	AA	18	0	24	3	0
6	BA	6	0	8	1	0
7	BA	7	0	0	0	0
8	AA	237	474	0	4	1
8	BA	289	578	0	5	1
8	PA	1	2	0	0	0
8	PAB	1	2	0	0	0
8	PB	2	4	0	0	0
All	All	4152	4482	3518	43	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 6.

The worst 5 of 43 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:PA:443:ACE:O	3:PAB:448:GLU:HB2	1.49	1.11
1:AA:18:ASN:H	4:AA:301:MPD:H12	1.31	0.95
2:PA:443:ACE:O	3:PAB:448:GLU:CB	2.05	0.94
1:AA:179:GLN:O	1:AA:180[B]:ASN:HB2	1.79	0.82
4:AA:302:MPD:H12	8:AA:606:HOH:O	1.79	0.81

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)
8:AA:401:HOH:O	8:BA:814:HOH:O[4_445]	2.15	0.05

## 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	AA	225/221 (102%)	214 (95%)	11 (5%)	0	100	100
1	BA	223/221 (101%)	217 (97%)	6 (3%)	0	100	100
2	PA	3/10 (30%)	1 (33%)	1 (33%)	1 (33%)	0	0
2	PB	3/10 (30%)	2 (67%)	1 (33%)	0	100	100
3	PAB	1/3 (33%)	1 (100%)	0	0	100	100
All	All	455/465 (98%)	435 (96%)	19 (4%)	1 (0%)	47	23

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	PA	444	VAL

### 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	AA	184/178 (103%)	177 (96%)	7 (4%)	33	7
1	BA	182/178 (102%)	177 (97%)	5 (3%)	44	15
2	PA	3/7 (43%)	3 (100%)	0	100	100
2	PB	3/7 (43%)	3 (100%)	0	100	100
3	PAB	3/3 (100%)	3 (100%)	0	100	100
All	All	375/373 (100%)	363 (97%)	12 (3%)	44	10

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	BA	273	ASN
1	BA	287	ASN
1	BA	428	ASN
1	BA	310	TYR
1	AA	148[B]	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are

no such sidechains identified.

### 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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 13 ligands modelled in this entry, 2 are monoatomic - leaving 11 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	$\# Z  > 2$
4	MPD	AA	307	-	7,7,7	0.20	0	9,10,10	0.37	0
4	MPD	BA	501	-	7,7,7	0.22	0	9,10,10	0.34	0
4	MPD	BA	503	-	7,7,7	0.20	0	9,10,10	0.24	0
6	GOL	AA	305	-	5,5,5	0.14	0	5,5,5	0.19	0
6	GOL	AA	306	-	5,5,5	0.26	0	5,5,5	0.28	0
7	TFA	BA	505	-	6,6,6	0.89	1 (16%)	9,9,9	0.80	0
6	GOL	BA	504	-	5,5,5	0.14	0	5,5,5	0.10	0
6	GOL	AA	308	-	5,5,5	0.15	0	5,5,5	0.19	0
4	MPD	AA	302	-	7,7,7	0.41	0	9,10,10	0.36	0
4	MPD	AA	301	-	7,7,7	0.35	0	9,10,10	0.40	0
4	MPD	AA	303	-	7,7,7	0.21	0	9,10,10	0.39	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
4	MPD	AA	307	-	-	1/5/5/5	-
4	MPD	BA	501	-	-	0/5/5/5	-
4	MPD	BA	503	-	-	2/5/5/5	-
6	GOL	AA	305	-	-	0/4/4/4	-
6	GOL	AA	306	-	-	0/4/4/4	-
7	TFA	BA	505	-	-	2/6/6/6	-
6	GOL	BA	504	-	-	0/4/4/4	-
6	GOL	AA	308	-	-	0/4/4/4	-
4	MPD	AA	302	-	-	0/5/5/5	-
4	MPD	AA	301	-	-	2/5/5/5	-
4	MPD	AA	303	-	-	0/5/5/5	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	BA	505	TFA	OXT-C1	-2.14	1.22	1.30

There are no bond angle outliers.

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	AA	307	MPD	C2-C3-C4-C5
4	BA	503	MPD	C1-C2-C3-C4
7	BA	505	TFA	O-C1-C2-F2
7	BA	505	TFA	OXT-C1-C2-F2
4	AA	301	MPD	O2-C2-C3-C4

There are no ring outliers.

8 monomers are involved in 15 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	AA	307	MPD	1	0
4	BA	501	MPD	2	0
6	AA	306	GOL	2	0
6	BA	504	GOL	1	0
6	AA	308	GOL	1	0
4	AA	302	MPD	3	0
4	AA	301	MPD	4	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	AA	303	MPD	1	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	AA	221/221 (100%)	-0.03	2 (0%) 84 87	15, 21, 40, 69	1 (0%)
1	BA	221/221 (100%)	0.20	11 (4%) 28 31	16, 23, 43, 91	0
2	PA	4/10 (40%)	3.96	2 (50%) 0 0	44, 49, 63, 67	4 (100%)
2	PB	4/10 (40%)	4.72	2 (50%) 0 0	35, 45, 70, 111	0
3	PAB	3/3 (100%)	2.04	2 (66%) 0 0	26, 26, 31, 64	3 (100%)
All	All	453/465 (97%)	0.18	19 (4%) 36 40	15, 22, 44, 111	8 (1%)

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	PB	452	VAL	14.0
2	PA	444	VAL	8.0
1	AA	63	GLN	7.7
1	BA	401	ASN	6.5
1	BA	397	ALA	5.5

### 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](#)

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	MPD	AA	301	8/8	0.58	0.22	28,48,58,60	0
6	GOL	AA	308	6/6	0.61	0.30	74,87,91,91	0
4	MPD	AA	303	8/8	0.70	0.18	61,72,81,85	0
6	GOL	BA	504	6/6	0.76	0.28	89,95,96,97	0
6	GOL	AA	306	6/6	0.77	0.32	69,77,89,91	0
6	GOL	AA	305	6/6	0.85	0.16	79,84,89,92	0
4	MPD	AA	302	8/8	0.85	0.21	26,37,54,66	0
4	MPD	BA	501	8/8	0.88	0.17	53,57,62,62	0
4	MPD	AA	307	8/8	0.89	0.18	70,75,93,94	0
7	TFA	BA	505	7/7	0.90	0.21	105,108,117,118	0
4	MPD	BA	503	8/8	0.94	0.08	40,47,73,78	0
5	CL	BA	502	1/1	0.98	0.06	27,27,27,27	0
5	CL	AA	304	1/1	0.99	0.05	25,25,25,25	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.