



Full wwPDB X-ray Structure Validation Report ⓘ

Nov 14, 2022 – 12:06 pm GMT

PDB ID : 7Q8N
Title : Peptide KKYDAFLA 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 : 2021-11-11
Resolution : 2.00 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at 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 : 2.31.2
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) : 2.31.2

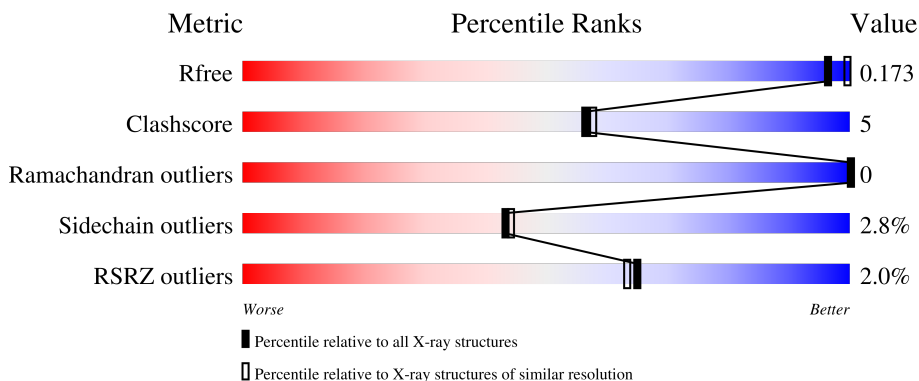
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 ≥ 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	222	 2% 89% 10%
1	BA	222	 91% 9%
2	PA	10	 10% 30% 10% 60%
2	PB	10	 30% 40% 60%

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 7777 atoms, of which 3963 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	222	3356	1079	1643	294	330	10	1646	2	0
1	BA	221	3311	1065	1622	290	324	10	1622	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AA	26	ALA	CYS	engineered mutation	UNP O60911
AA	109	GLN	ASN	engineered mutation	UNP O60911
AA	180	GLN	ASN	engineered mutation	UNP O60911
BA	247	ALA	CYS	engineered mutation	UNP O60911
BA	330	GLN	ASN	engineered mutation	UNP O60911
BA	401	GLN	ASN	engineered mutation	UNP O60911

- Molecule 2 is a protein called KKYDAFLA Peptide.

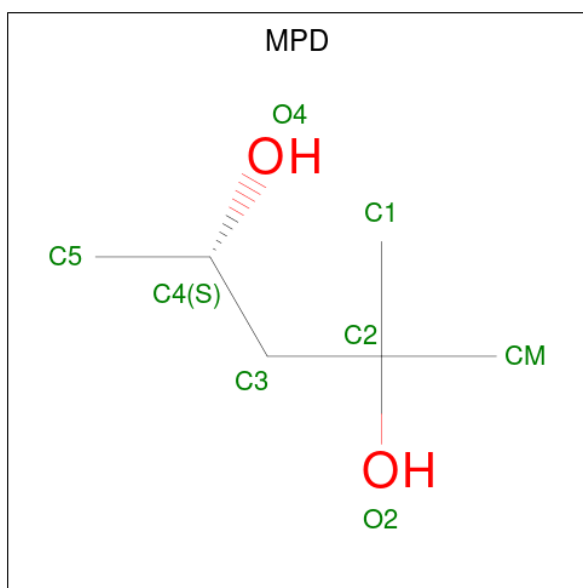
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	H	N	O			
2	PA	4	52	18	27	4	3	27	0	1
2	PB	4	52	18	27	4	3	27	0	1

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

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

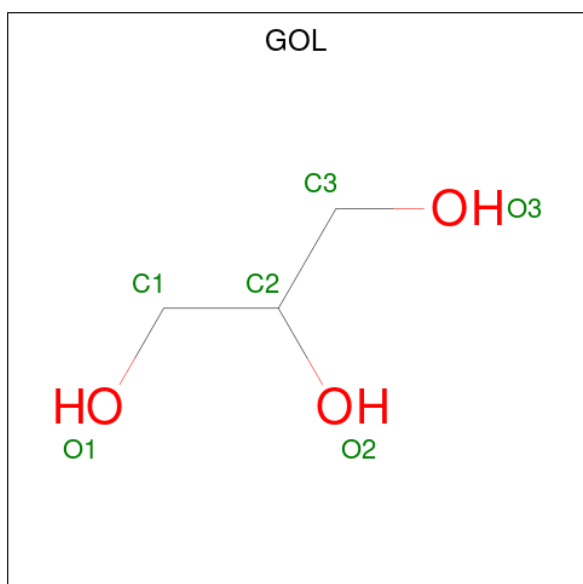
- Molecule 4 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula:

C₆H₁₄O₂).



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	BA	1	Total C O 8 6 2	0	0
4	BA	1	Total C O 8 6 2	0	0

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



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

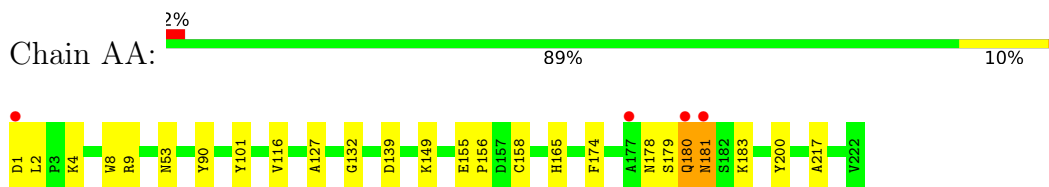
- Molecule 6 is water.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	AA	163	Total	H	O	326	0
			489	326	163		
6	BA	158	Total	H	O	316	0
			474	316	158		
6	PB	1	Total	H	O	2	0
			3	2	1		

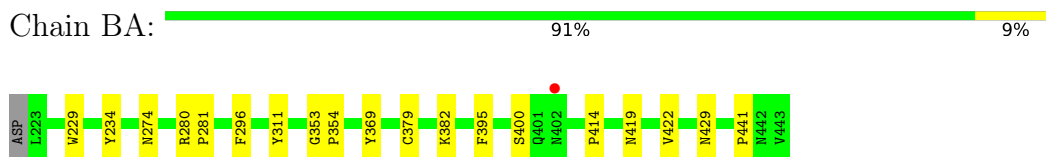
3 Residue-property plots [i](#)

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.

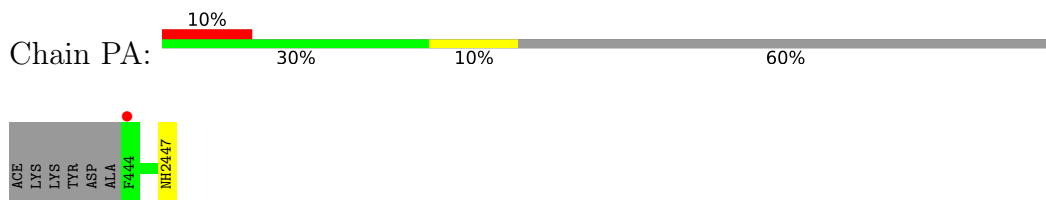
- Molecule 1: Cathepsin L2



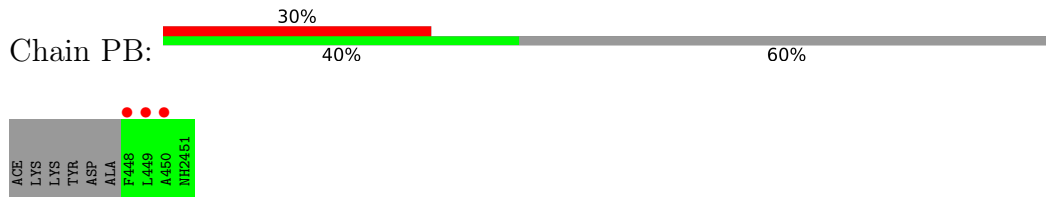
- Molecule 1: Cathepsin L2



- Molecule 2: KKYDAFLA Peptide



- Molecule 2: KKYDAFLA Peptide



4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α , β , γ	94.15Å 94.15Å 126.10Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.78 – 2.00 45.78 – 2.00	Depositor EDS
% Data completeness (in resolution range)	100.0 (45.78-2.00) 100.0 (45.78-2.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.51 (at 2.00Å)	Xtriage
Refinement program	MAIN	Depositor
R, R_{free}	0.173 , 0.200 0.171 , 0.173	Depositor DCC
R_{free} test set	1890 reflections (4.86%)	wwPDB-VP
Wilson B-factor (Å ²)	30.2	Xtriage
Anisotropy	0.019	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	7777	wwPDB-VP
Average B, all atoms (Å ²)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.59% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: NH2, GOL, MPD, 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	0.75	0/1755	0.75	0/2372
1	BA	0.71	0/1731	0.74	0/2339
2	PA	0.58	0/24	0.75	0/31
2	PB	0.23	0/24	0.44	0/31
All	All	0.73	0/3534	0.74	0/4773

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	1713	1643	1647	11	2
1	BA	1689	1622	1621	9	1
2	PA	25	27	24	1	0
2	PB	25	27	24	0	0
3	AA	1	0	0	0	0
3	BA	1	0	0	0	0
4	AA	16	0	28	3	0
4	BA	16	0	28	6	0
5	BA	6	0	8	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	AA	163	326	0	0	0
6	BA	158	316	0	1	0
6	PB	1	2	0	0	0
All	All	3814	3963	3380	28	2

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 5.

All (28) 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)
4:BA:501:MPD:HM2	4:BA:501:MPD:H52	1.53	0.90
1:AA:127:ALA:HA	4:AA:303:MPD:HM3	1.73	0.70
4:BA:501:MPD:HM2	4:BA:501:MPD:C5	2.22	0.67
1:AA:180:GLN:OE1	1:AA:183:LYS:NZ	2.32	0.62
4:BA:502:MPD:H13	4:BA:502:MPD:C5	2.32	0.59
1:AA:8:TRP:CE2	1:AA:132:GLY:HA2	2.37	0.59
1:AA:165:HIS:HD1	2:PA:447:NH2:N	2.05	0.55
1:AA:178:ASN:HB3	1:AA:181[B]:ASN:HB2	1.93	0.51
4:BA:502:MPD:H13	4:BA:502:MPD:H53	1.91	0.51
4:AA:303:MPD:O4	4:AA:303:MPD:HM1	2.11	0.50
1:BA:395:PHE:CG	1:BA:400:SER:HA	2.47	0.49
1:AA:116:VAL:HB	1:AA:217:ALA:HB3	1.94	0.48
1:AA:9:ARG:HD3	1:AA:200:TYR:CZ	2.49	0.47
1:BA:229:TRP:CE2	1:BA:353:GLY:HA2	2.50	0.46
4:AA:302:MPD:H52	4:AA:302:MPD:O2	2.16	0.46
4:BA:501:MPD:CM	6:BA:710:HOH:O	2.64	0.45
4:BA:501:MPD:C5	4:BA:501:MPD:CM	2.91	0.45
1:BA:395:PHE:CD1	1:BA:400:SER:HA	2.53	0.44
1:BA:280:ARG:N	1:BA:281:PRO:CD	2.81	0.43
1:AA:174:PHE:CD2	1:AA:179:SER:HA	2.54	0.43
1:AA:180:GLN:OE1	1:AA:183:LYS:CE	2.66	0.43
1:BA:229:TRP:CG	1:BA:354:PRO:HD3	2.54	0.42
1:AA:1:ASP:CG	1:AA:2:LEU:N	2.73	0.42
1:BA:414:PRO:O	1:BA:419:ASN:HA	2.21	0.41
1:AA:155:GLU:HA	1:AA:156:PRO:HD2	1.92	0.41
1:BA:229:TRP:HB3	1:BA:234:TYR:HB2	2.03	0.41
1:BA:296:PHE:CZ	1:BA:441:PRO:HD3	2.55	0.41
1:BA:369:TYR:CZ	1:BA:422:VAL:HG23	2.56	0.41

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the sym-

metry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:AA:139:ASP:OD2	1:BA:382:LYS:HZ2[3_444]	1.54	0.06
1:AA:4:LYS:HZ2	1:AA:101:TYR:O[3_544]	1.55	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	222/222 (100%)	216 (97%)	6 (3%)	0	100	100
1	BA	219/222 (99%)	213 (97%)	6 (3%)	0	100	100
2	PA	2/10 (20%)	1 (50%)	1 (50%)	0	100	100
2	PB	2/10 (20%)	1 (50%)	1 (50%)	0	100	100
All	All	445/464 (96%)	431 (97%)	14 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	181/179 (101%)	174 (96%)	7 (4%)	32	30
1	BA	178/179 (99%)	174 (98%)	4 (2%)	52	55
2	PA	2/6 (33%)	2 (100%)	0	100	100
2	PB	2/6 (33%)	2 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	363/370 (98%)	352 (97%)	11 (3%)	43 41

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AA	53	ASN
1	AA	90	TYR
1	AA	149	LYS
1	AA	158	CYS
1	AA	180	GLN
1	AA	181[A]	ASN
1	AA	181[B]	ASN
1	BA	274	ASN
1	BA	311	TYR
1	BA	379	CYS
1	BA	429	ASN

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 7 ligands modelled in this entry, 2 are monoatomic - leaving 5 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	BA	501	-	7,7,7	0.21	0	9,10,10	0.36	0
5	GOL	BA	503	-	5,5,5	0.12	0	5,5,5	0.08	0
4	MPD	AA	303	-	7,7,7	0.25	0	9,10,10	0.24	0
4	MPD	BA	502	-	7,7,7	0.21	0	9,10,10	0.39	0
4	MPD	AA	302	-	7,7,7	0.27	0	9,10,10	0.33	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	BA	501	-	-	1/5/5/5	-
5	GOL	BA	503	-	-	0/4/4/4	-
4	MPD	AA	303	-	-	0/5/5/5	-
4	MPD	BA	502	-	-	1/5/5/5	-
4	MPD	AA	302	-	-	1/5/5/5	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	BA	502	MPD	C2-C3-C4-C5
4	AA	302	MPD	C2-C3-C4-C5
4	BA	501	MPD	C2-C3-C4-C5

There are no ring outliers.

4 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	BA	501	MPD	4	0
4	AA	303	MPD	2	0
4	BA	502	MPD	2	0
4	AA	302	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, 95th 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 ²)	Q<0.9
1	AA	222/222 (100%)	-0.55	4 (1%) 68 66	20, 28, 49, 86	1 (0%)
1	BA	221/222 (99%)	-0.57	1 (0%) 91 90	22, 32, 55, 110	0
2	PA	3/10 (30%)	1.21	1 (33%) 0 0	53, 53, 60, 98	0
2	PB	3/10 (30%)	3.63	3 (100%) 0 0	57, 57, 79, 138	0
All	All	449/464 (96%)	-0.52	9 (2%) 65 63	20, 30, 56, 138	1 (0%)

All (9) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	PB	448	PHE	5.1
1	AA	181[A]	ASN	5.0
1	AA	1	ASP	3.8
2	PB	450	ALA	3.3
1	AA	180	GLN	2.6
2	PB	449	LEU	2.4
1	AA	177	ALA	2.3
2	PA	444	PHE	2.3
1	BA	402	ASN	2.0

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, 95th 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 ²)	Q<0.9
4	MPD	BA	502	8/8	0.87	0.28	73,86,93,101	0
5	GOL	BA	503	6/6	0.87	0.19	55,68,82,88	0
4	MPD	AA	302	8/8	0.89	0.19	63,70,94,95	0
4	MPD	AA	303	8/8	0.91	0.33	104,112,116,118	0
4	MPD	BA	501	8/8	0.93	0.31	70,80,95,95	0
3	CL	BA	504	1/1	0.98	0.06	41,41,41,41	0
3	CL	AA	301	1/1	0.99	0.06	35,35,35,35	0

6.5 Other polymers [i](#)

There are no such residues in this entry.