wwPDB X-ray Structure Validation Summary Report (i)

Nov 14, 2022 – 12:09 pm GMT

|		
PDB ID	:	7Q8F
Title	:	Peptide GNYKEAKK 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	:	1.49 Å(reported)

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 1.49 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.
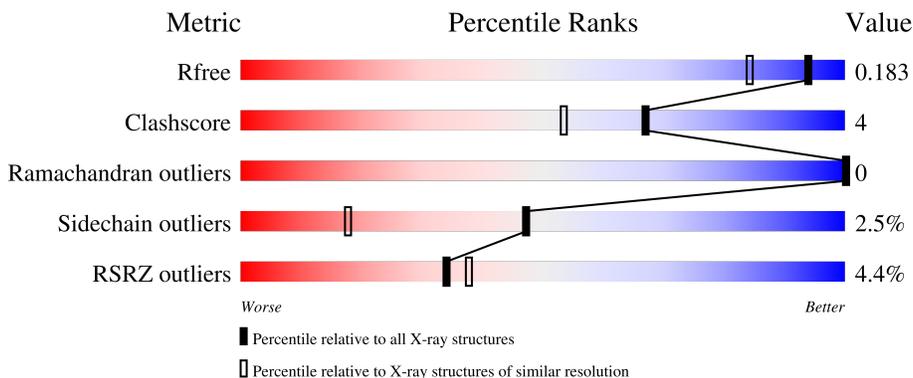
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2936 (1.50-1.50)
Clashscore	141614	3144 (1.50-1.50)
Ramachandran outliers	138981	3066 (1.50-1.50)
Sidechain outliers	138945	3064 (1.50-1.50)
RSRZ outliers	127900	2884 (1.50-1.50)

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 >=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 <=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	2% / 95% / 5%
1	BA	221	4% / 94% / 6%
2	PA	10	30% / 70% / 10% / 10% / 10%
2	PB	10	30% / 60% / 10% / 30%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard

residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
6	GOL	AA	808	-	-	X	X

2 Entry composition ⓘ

There are 7 unique types of molecules in this entry. The entry contains 8304 atoms, of which 4248 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
1	AA	221	Total 3336	C 1072	H 1633	N 293	O 328	S 10	1640	2	0
1	BA	221	Total 3323	C 1069	H 1626	N 291	O 327	S 10	1627	1	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 GNYKEAKK Peptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	PB	7	Total 115	C 35	H 61	N 10	O 9	61	0	1
2	PA	9	Total 136	C 41	H 70	N 13	O 12	74	0	1

- Molecule 3 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula: $C_6H_{14}O_2$).

MPD

O4

OH

C1

C5

C4(S)

C2

CM

C3

OH

O2
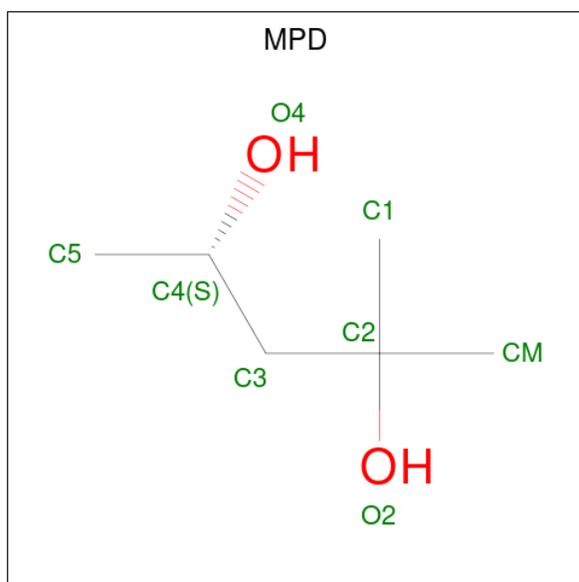
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	AA	1	Total 8	C 6	O 2	0	0
3	AA	1	Total 8	C 6	O 2	0	0
3	AA	1	Total 8	C 6	O 2	0	0
3	AA	1	Total 8	C 6	O 2	0	0
3	BA	1	Total 8	C 6	O 2	0	0
3	BA	1	Total 8	C 6	O 2	0	0
3	BA	1	Total 8	C 6	O 2	0	0

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

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

- Molecule 5 is trifluoroacetic acid (three-letter code: TFA) (formula: $C_2HF_3O_2$).

TFA

OXT

HO

F2 F

C1　　O　O

C2

F

F1

F

F3
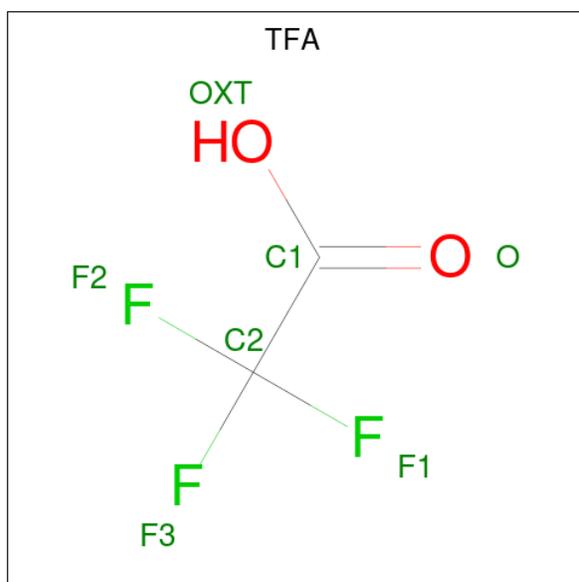
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	AA	1	Total 7	C 2	F 3	O 2	0	0

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).
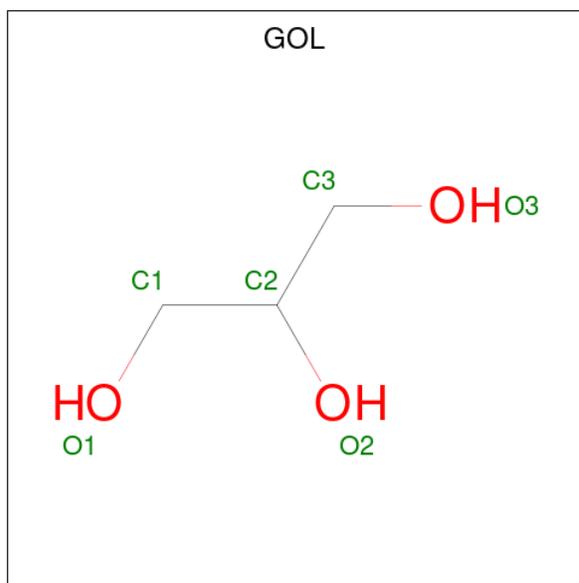
GOL

C3

OH O3

C1　　C2

HO

O1

OH

O2

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

Continued on next page...

W O R L D W I D E

PDB

PROTEIN DATA BANK

Continued from previous page...

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

- Molecule 7 is water.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	AA	206	Total 618	H 412	O 206	413	0
7	BA	213	Total 639	H 426	O 213	426	0
7	PB	4	Total 12	H 8	O 4	8	0
7	PA	6	Total 18	H 12	O 6	12	0

3 Residue-property plots ⓘ

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

Chain AA:
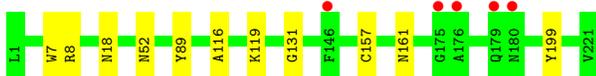
- Molecule 1: Cathepsin L2

Chain BA:
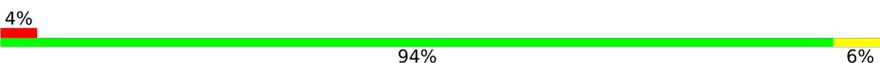
- Molecule 2: GNYKEAKK Peptide

Chain PB:

- Molecule 2: GNYKEAKK Peptide

Chain PA:

4 Data and refinement statistics ⓘ

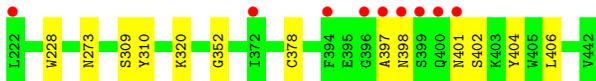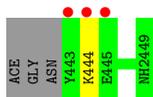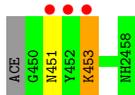| Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	94.32Å 94.32Å 125.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.16 – 1.49 47.16 – 1.49	Depositor EDS
% Data completeness (in resolution range)	100.0 (47.16-1.49) 99.9 (47.16-1.49)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.37 (at 1.48Å)	Xtriage
Refinement program	MAIN	Depositor
R, R_{free}	0.175 , 0.197 0.177 , 0.183	Depositor DCC
R_{free} test set	4707 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	22.5	Xtriage
Anisotropy	0.005	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< |L| > = 0.51, < L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	8304	wwPDB-VP
Average B, all atoms (Å^2)	14.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.92% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, NH2, MPD, TFA, 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	0.81	0/1745	0.85	0/2358
1	BA	0.76	0/1739	0.84	0/2350
2	PA	0.52	0/65	0.95	0/83
2	PB	0.52	0/53	0.77	0/67
All	All	0.78	0/3602	0.85	0/4858

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 ⓘ

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	1703	1633	1631	9	0
1	BA	1697	1626	1622	7	0
2	PA	66	70	67	2	0
2	PB	54	61	58	0	0
3	AA	32	0	56	8	0
3	BA	24	0	42	3	0
4	AA	1	0	0	0	0
4	BA	1	0	0	0	0
5	AA	7	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	AA	12	0	16	5	0
6	BA	30	0	40	3	0
7	AA	206	412	0	4	0
7	BA	213	426	0	0	0
7	PA	6	12	0	0	0
7	PB	4	8	0	0	0
All	All	4056	4248	3532	28	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 4.

The worst 5 of 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)
1:AA:18:ASN:H	3:AA:802:MPD:H12	1.31	0.95
3:AA:801:MPD:H12	6:AA:808:GOL:H31	1.49	0.95
6:AA:808:GOL:H12	7:AA:1029:HOH:O	1.70	0.90
3:BA:501:MPD:H53	3:BA:501:MPD:HM1	1.56	0.88
3:AA:801:MPD:C1	6:AA:808:GOL:H31	2.17	0.73

There are no symmetry-related clashes.

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	AA	221/221 (100%)	217 (98%)	4 (2%)	0	100	100
1	BA	220/221 (100%)	214 (97%)	6 (3%)	0	100	100
2	PA	7/10 (70%)	6 (86%)	1 (14%)	0	100	100
2	PB	5/10 (50%)	5 (100%)	0	0	100	100
All	All	453/462 (98%)	442 (98%)	11 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	AA	180/178 (101%)	177 (98%)	3 (2%)	60	33
1	BA	179/178 (101%)	176 (98%)	3 (2%)	60	33
2	PA	6/6 (100%)	4 (67%)	2 (33%)	0	0
2	PB	5/6 (83%)	4 (80%)	1 (20%)	1	0
All	All	370/368 (100%)	361 (98%)	9 (2%)	47	19

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	PA	451	ASN
2	PA	453	LYS
1	BA	273	ASN
1	BA	310	TYR
1	BA	378	CYS

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

Mol	Chain	Res	Type
1	AA	161	ASN
1	BA	325	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5　Carbohydrates　ⓘ

There are no monosaccharides in this entry.

5.6　Ligand geometry　ⓘ

Of 17 ligands modelled in this entry, 2 are monoatomic - leaving 15 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
3	MPD	AA	804	-	7,7,7	0.40	0	9,10,10	0.44	0
6	GOL	BA	503	-	5,5,5	0.56	0	5,5,5	0.54	0
6	GOL	BA	509	-	5,5,5	0.39	0	5,5,5	0.63	0
6	GOL	BA	507	-	5,5,5	0.15	0	5,5,5	0.84	0
6	GOL	AA	807	-	5,5,5	0.14	0	5,5,5	0.13	0
6	GOL	BA	506	-	5,5,5	0.16	0	5,5,5	0.23	0
3	MPD	BA	502	-	7,7,7	0.31	0	9,10,10	0.28	0
3	MPD	AA	805	-	7,7,7	0.21	0	9,10,10	0.22	0
3	MPD	AA	802	-	7,7,7	0.44	0	9,10,10	0.54	0
6	GOL	AA	808	-	5,5,5	0.75	0	5,5,5	0.55	0
3	MPD	AA	801	-	7,7,7	0.23	0	9,10,10	0.35	0
5	TFA	AA	806	-	6,6,6	1.19	1 (16%)	9,9,9	0.94	0
6	GOL	BA	505	-	5,5,5	0.31	0	5,5,5	1.15	1 (20%)
3	MPD	BA	501	-	7,7,7	0.47	0	9,10,10	0.55	0
3	MPD	BA	508	-	7,7,7	0.30	0	9,10,10	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
3	MPD	AA	804	-	-	0/5/5/5	-
6	GOL	BA	503	-	-	0/4/4/4	-
6	GOL	BA	509	-	-	2/4/4/4	-
6	GOL	BA	507	-	-	2/4/4/4	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
6	GOL	AA	807	-	-	0/4/4/4	-
6	GOL	BA	506	-	-	0/4/4/4	-
3	MPD	BA	502	-	-	0/5/5/5	-
3	MPD	AA	805	-	-	1/5/5/5	-
3	MPD	AA	802	-	-	1/5/5/5	-
6	GOL	AA	808	-	-	0/4/4/4	-
3	MPD	AA	801	-	-	0/5/5/5	-
5	TFA	AA	806	-	-	0/6/6/6	-
6	GOL	BA	505	-	-	1/4/4/4	-
3	MPD	BA	501	-	-	1/5/5/5	-
3	MPD	BA	508	-	-	1/5/5/5	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	AA	806	TFA	OXT-C1	-2.85	1.19	1.30

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
6	BA	505	GOL	C3-C2-C1	-2.46	102.14	111.70

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	BA	505	GOL	O1-C1-C2-C3
6	BA	507	GOL	C1-C2-C3-O3
3	BA	508	MPD	O2-C2-C3-C4
6	BA	509	GOL	O2-C2-C3-O3
3	AA	802	MPD	C2-C3-C4-C5

There are no ring outliers.

8 monomers are involved in 17 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	BA	503	GOL	2	0
6	BA	507	GOL	1	0
6	AA	807	GOL	1	0
3	AA	805	MPD	1	0
3	AA	802	MPD	5	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	AA	808	GOL	4	0
3	AA	801	MPD	2	0
3	BA	501	MPD	3	0

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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	221/221 (100%)	0.29	5 (2%)	60	65	14, 20, 37, 61	4 (1%)
1	BA	221/221 (100%)	0.49	9 (4%)	37	41	15, 23, 42, 98	1 (0%)
2	PA	7/10 (70%)	2.89	3 (42%)	0	0	33, 37, 70, 85	0
2	PB	6/10 (60%)	6.69	3 (50%)	0	0	39, 43, 86, 95	2 (33%)
All	All	455/462 (98%)	0.51	20 (4%)	34	38	14, 22, 43, 98	7 (1%)

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	PB	443	TYR	25.9
2	PB	444	LYS	9.7
2	PA	453	LYS	6.5
2	PA	452	TYR	5.4
1	BA	396	GLY	5.3

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

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(Å²)	Q<0.9
3	MPD	AA	802	8/8	0.48	0.22	31,54,64,65	0
6	GOL	AA	808	6/6	<mark>0.54</mark>	<mark>0.54</mark>	58,62,67,68	0
6	GOL	BA	506	6/6	0.55	0.21	90,93,96,97	0
3	MPD	AA	805	8/8	0.57	0.39	62,68,71,71	0
6	GOL	BA	509	6/6	0.65	0.21	55,61,67,70	0
6	GOL	BA	505	6/6	0.66	0.23	123,128,130,131	0
6	GOL	BA	503	6/6	0.68	0.19	34,47,55,57	0
3	MPD	BA	502	8/8	0.68	0.23	36,47,63,68	0
3	MPD	AA	801	8/8	0.70	0.17	63,70,97,101	0
6	GOL	AA	807	6/6	0.72	0.17	58,75,80,82	0
5	TFA	AA	806	7/7	0.80	0.20	46,53,68,72	0
3	MPD	BA	501	8/8	0.82	0.17	39,48,53,54	0
3	MPD	BA	508	8/8	0.90	0.14	43,56,87,95	0
3	MPD	AA	804	8/8	0.90	0.12	29,38,62,73	0
6	GOL	BA	507	6/6	0.92	0.14	76,79,82,83	0
4	CL	BA	504	1/1	0.96	0.07	29,29,29,29	0
4	CL	AA	803	1/1	1.00	0.06	24,24,24,24	0

6.5 Other polymers ⓘ

There are no such residues in this entry.