wwPDB X-ray Structure Validation Summary Report (i)

Nov 14, 2022 – 12:07 pm GMT

PDB ID	:	7Q8M
Title	:	Peptide KPKKKTK 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	:	1.57 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (i)) were used in the production of this report:

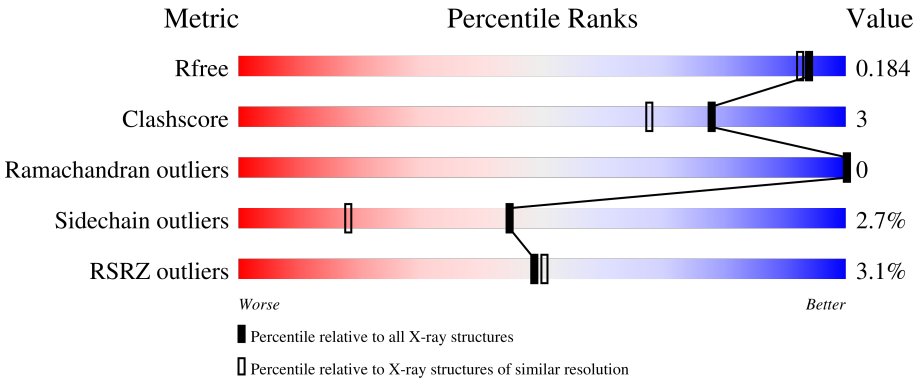
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.31.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

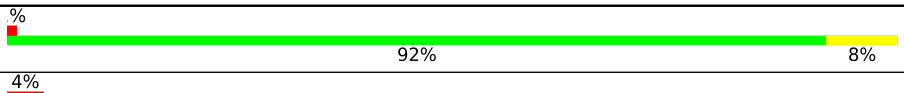
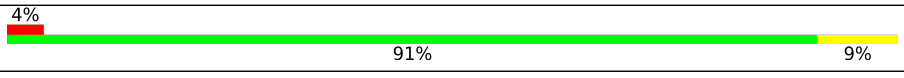
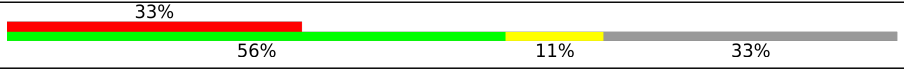
The reported resolution of this entry is 1.57 Å.

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	5534 (1.60-1.56)
Clashscore	141614	5861 (1.60-1.56)
Ramachandran outliers	138981	5708 (1.60-1.56)
Sidechain outliers	138945	5703 (1.60-1.56)
RSRZ outliers	127900	5431 (1.60-1.56)

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	222	
1	BA	222	
2	PA	9	
2	PB	9	

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 8153 atoms, of which 4183 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	222	Total	C	H	N	O	S	1662	3	0
			3378	1085	1656	296	331	10			
1	BA	222	Total	C	H	N	O	S	1662	2	0
			3354	1079	1641	293	331	10			

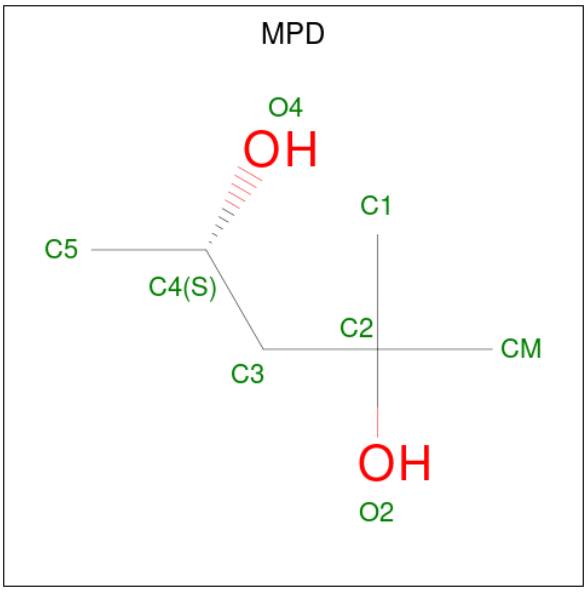
There are 6 discrepancies between the modelled and reference sequences:

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	248	ALA	CYS	engineered mutation	UNP O60911
BA	331	GLN	ASN	engineered mutation	UNP O60911
BA	402	GLN	ASN	engineered mutation	UNP O60911

- Molecule 2 is a protein called KPKKKTK Peptide.

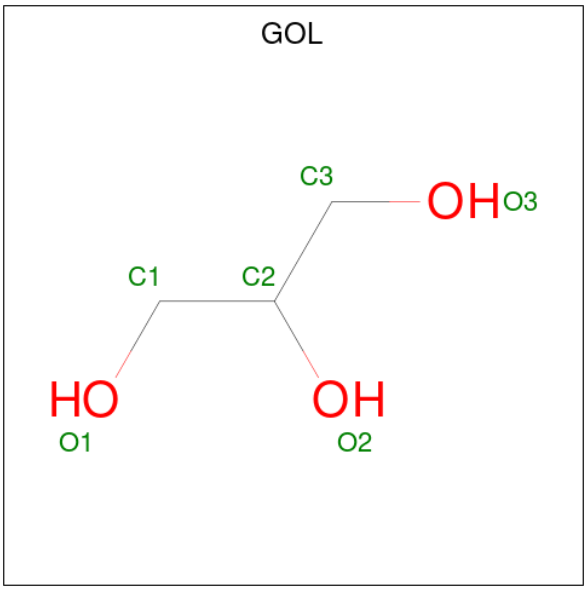
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	PA	4	Total	C	H	N	O	35	0	1
			61	16	35	6	4			
2	PB	6	Total	C	H	N	O	71	0	1
			105	28	61	10	6			

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



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

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



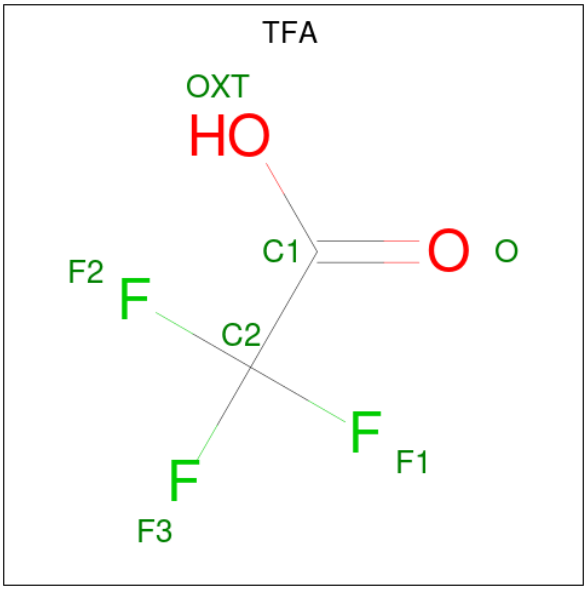
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	AA	1	Total C O 6 3 3	0	0

Continued on next page...

Continued from previous page...

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

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



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

- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	AA	203	Total H O 609 406 203	406	0
6	BA	189	Total H O 567 378 189	378	0
6	PA	1	Total H O 3 2 1	2	0

Continued on next page...

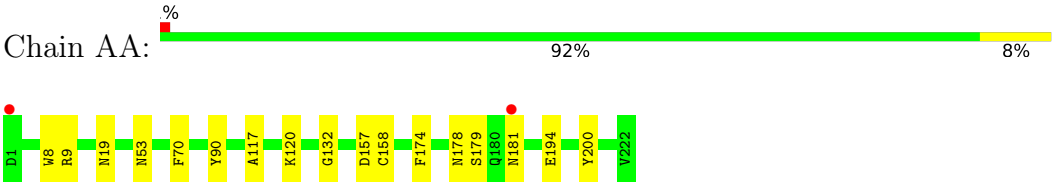
Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	PB	2	Total H O 6 4 2	4	0

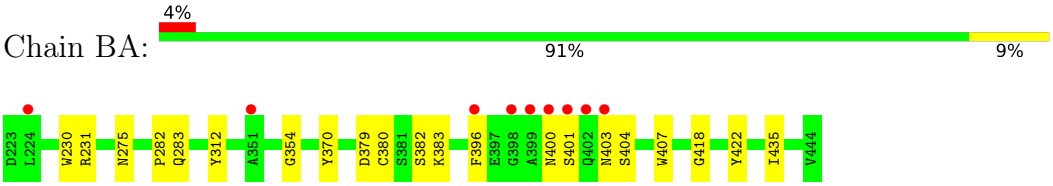
3 Residue-property plots (i)

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.

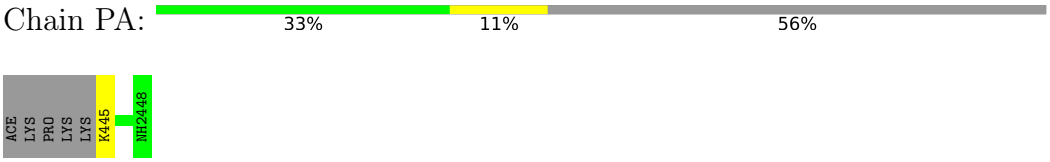
• Molecule 1: Cathepsin L2



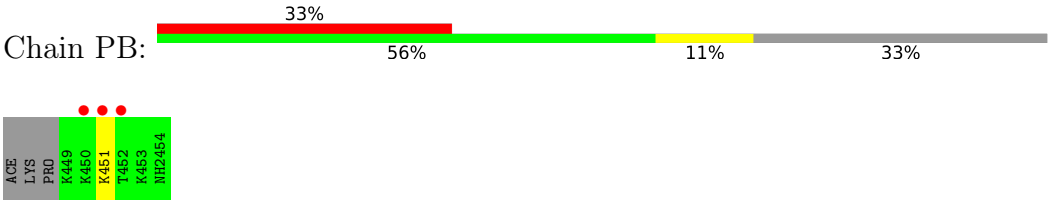
• Molecule 1: Cathepsin L2



• Molecule 2: KPKKKTK Peptide



• Molecule 2: KPKKKTK Peptide



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	94.23Å 94.23Å 125.61Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.11 – 1.57 47.11 – 1.57	Depositor EDS
% Data completeness (in resolution range)	100.0 (47.11-1.57) 99.9 (47.11-1.57)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.46 (at 1.57Å)	Xtriage
Refinement program	MAIN	Depositor
R, R_{free}	0.185 , 0.208 0.183 , 0.184	Depositor DCC
R_{free} test set	3968 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	24.8	Xtriage
Anisotropy	0.116	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	8153	wwPDB-VP
Average B, all atoms (Å^2)	16.0	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: NH2, TFA, MPD, GOL

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	0.69	0/1764	0.79	0/2383
1	BA	0.64	0/1755	0.78	0/2372
2	PA	0.62	0/24	0.86	0/29
2	PB	0.48	0/42	0.69	0/51
All	All	0.66	0/3585	0.78	0/4835

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	1722	1656	1653	8	1
1	BA	1713	1641	1636	11	1
2	PA	26	35	32	1	0
2	PB	44	61	58	0	0
3	AA	16	0	28	3	0
3	BA	8	0	14	0	0
4	AA	6	0	8	0	0
4	BA	12	0	16	5	0
5	AA	28	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	203	406	0	0	1
6	BA	189	378	0	1	1
6	PA	1	2	0	0	0
6	PB	2	4	0	0	0
All	All	3970	4183	3445	20	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 3.

The worst 5 of 20 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:19:ASN:H	3:AA:301:MPD:HM3	1.35	0.90
1:BA:283:GLN:HE22	4:BA:503:GOL:H2	1.58	0.68
1:BA:231:ARG:HH22	4:BA:502:GOL:H31	1.62	0.65
1:BA:400:ASN:HB3	1:BA:403:ASN:OD1	2.02	0.59
1:AA:70:PHE:CZ	2:PA:445:LYS:HA	2.38	0.58

All (2) 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)
6:AA:564:HOH:O	6:BA:691:HOH:O[6_555]	1.78	0.42
1:AA:157:ASP:OD1	1:BA:382:SER:H[4_445]	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	223/222 (100%)	217 (97%)	6 (3%)	0	100	100
1	BA	222/222 (100%)	215 (97%)	7 (3%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	PA	2/9 (22%)	2 (100%)	0	0	100	100
2	PB	4/9 (44%)	4 (100%)	0	0	100	100
All	All	451/462 (98%)	438 (97%)	13 (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	182/179 (102%)	178 (98%)	4 (2%)	52	25
1	BA	181/179 (101%)	176 (97%)	5 (3%)	43	17
2	PA	3/7 (43%)	3 (100%)	0	100	100
2	PB	5/7 (71%)	4 (80%)	1 (20%)	1	0
All	All	371/372 (100%)	361 (97%)	10 (3%)	44	18

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

Mol	Chain	Res	Type
1	BA	380	CYS
1	BA	404	SER
2	PB	451	LYS
1	AA	194	GLU
1	BA	275	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)

10 ligands are modelled in this entry.

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	GOL	AA	303	-	5,5,5	0.20	0	5,5,5	0.15	0
4	GOL	BA	502	-	5,5,5	0.42	0	5,5,5	0.87	0
5	TFA	AA	304	-	6,6,6	0.96	1 (16%)	9,9,9	0.94	0
3	MPD	AA	301	-	7,7,7	0.35	0	9,10,10	0.29	0
5	TFA	AA	305	-	6,6,6	0.85	1 (16%)	9,9,9	0.92	1 (11%)
3	MPD	AA	302	-	7,7,7	0.25	0	9,10,10	0.30	0
3	MPD	BA	501	-	7,7,7	0.30	0	9,10,10	0.27	0
5	TFA	AA	306	-	6,6,6	0.96	0	9,9,9	0.84	0
4	GOL	BA	503	-	5,5,5	0.27	0	5,5,5	0.43	0
5	TFA	AA	307	-	6,6,6	0.90	1 (16%)	9,9,9	0.84	1 (11%)

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	GOL	AA	303	-	-	0/4/4/4	-
4	GOL	BA	502	-	-	2/4/4/4	-
5	TFA	AA	304	-	-	0/6/6/6	-
3	MPD	AA	301	-	-	0/5/5/5	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	TFA	AA	305	-	-	0/6/6/6	-
3	MPD	AA	302	-	-	1/5/5/5	-
3	MPD	BA	501	-	-	1/5/5/5	-
5	TFA	AA	306	-	-	1/6/6/6	-
4	GOL	BA	503	-	-	0/4/4/4	-
5	TFA	AA	307	-	-	2/6/6/6	-

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	AA	304	TFA	OXT-C1	-2.16	1.22	1.30
5	AA	307	TFA	OXT-C1	-2.15	1.22	1.30
5	AA	305	TFA	OXT-C1	-2.06	1.22	1.30

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	AA	305	TFA	OXT-C1-C2	2.33	122.17	112.78
5	AA	307	TFA	OXT-C1-C2	2.04	121.00	112.78

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	BA	502	GOL	C1-C2-C3-O3
3	AA	302	MPD	C2-C3-C4-C5
3	BA	501	MPD	CM-C2-C3-C4
4	BA	502	GOL	O2-C2-C3-O3
5	AA	307	TFA	O-C1-C2-F1

There are no ring outliers.

4 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	BA	502	GOL	2	0
3	AA	301	MPD	2	0
3	AA	302	MPD	1	0
4	BA	503	GOL	3	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

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, 95^{th} 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%)	-0.19	2 (0%) 84 85	17, 24, 45, 72	2 (0%)
1	BA	221/222 (99%)	0.10	9 (4%) 37 37	19, 29, 53, 86	6 (2%)
2	PA	3/9 (33%)	0.96	0 100 100	43, 43, 46, 68	0
2	PB	4/9 (44%)	2.21	3 (75%) 0 0	49, 51, 70, 77	1 (25%)
All	All	450/462 (97%)	-0.02	14 (3%) 49 50	17, 26, 51, 86	9 (2%)

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	BA	399	ALA	7.6
1	BA	398	GLY	5.7
1	BA	403	ASN	4.7
1	AA	1	ASP	4.7
1	BA	400	ASN	4.7

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, 95^{th} 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
3	MPD	AA	301	8/8	0.64	0.25	37,58,87,94	0
4	GOL	BA	503	6/6	0.72	0.20	62,69,77,82	0
4	GOL	BA	502	6/6	0.73	0.22	58,63,69,70	0
3	MPD	BA	501	8/8	0.78	0.22	49,59,82,86	0
3	MPD	AA	302	8/8	0.79	0.26	61,68,88,90	0
4	GOL	AA	303	6/6	0.80	0.14	44,49,53,53	0
5	TFA	AA	305	7/7	0.88	0.12	37,47,50,55	0
5	TFA	AA	307	7/7	0.90	0.23	57,64,84,86	0
5	TFA	AA	306	7/7	0.91	0.18	31,41,62,63	0
5	TFA	AA	304	7/7	0.92	0.11	27,33,46,46	0

6.5 Other polymers (i)

There are no such residues in this entry.