wwPDB X-ray Structure Validation Summary Report (i)

Dec 13, 2022 – 12:06 pm GMT

PDB ID	:	7QHK
Title	:	Peptide QLRQQE 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	:	2021-12-13
Resolution	:	1.83 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:

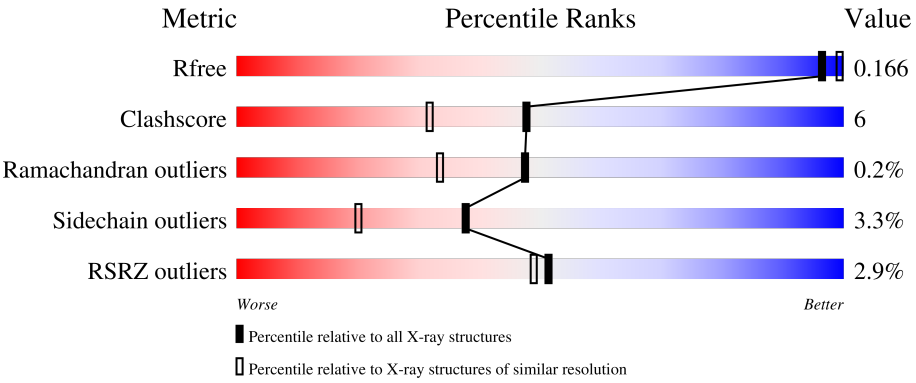
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.31.3
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.3

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 1.83 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	4003 (1.86-1.82)
Clashscore	141614	4233 (1.86-1.82)
Ramachandran outliers	138981	4185 (1.86-1.82)
Sidechain outliers	138945	4186 (1.86-1.82)
RSRZ outliers	127900	3957 (1.86-1.82)

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	90% 10%
1	BA	221	5% 87% 12%
2	PA	6	50% 17% 50% 33%
2	PB	6	33% 17% 50%

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
4	MPD	BA	501	-	-	X	-
4	MPD	BA	505	-	-	X	-
5	GOL	AA	304	-	-	-	X

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 8161 atoms, of which 4176 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	C	H	N	O	S	1660	6	0
			3390	1089	1660	297	334	10			
1	BA	221	Total	C	H	N	O	S	1667	7	0
			3401	1092	1666	299	334	10			

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 GLN-LEU-ARG-GLN.

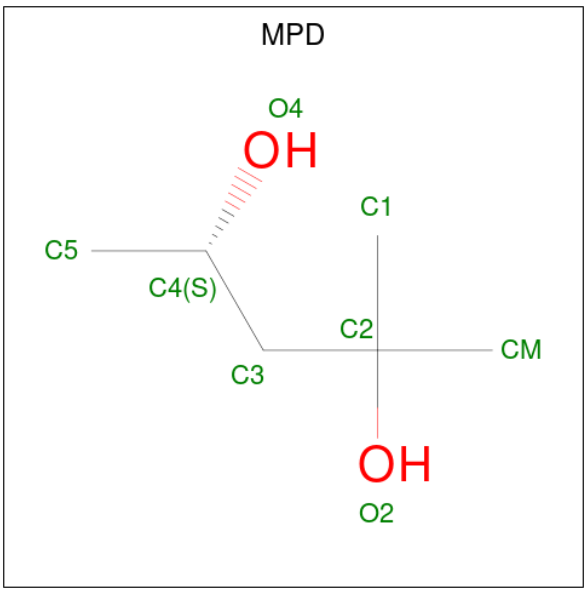
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	PA	4	Total	C	H	N	O	42	0	0
			79	22	42	9	6			
2	PB	3	Total	C	H	N	O	22	0	0
			50	15	22	5	8			

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	AA	2	Total	Cl	0	0
			2	2		
3	BA	1	Total	Cl	0	0
			1	1		

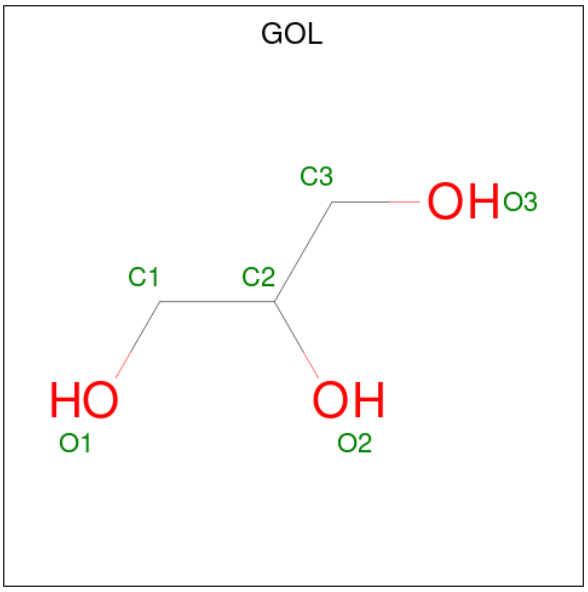
- Molecule 4 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula:

$C_6H_{14}O_2$).



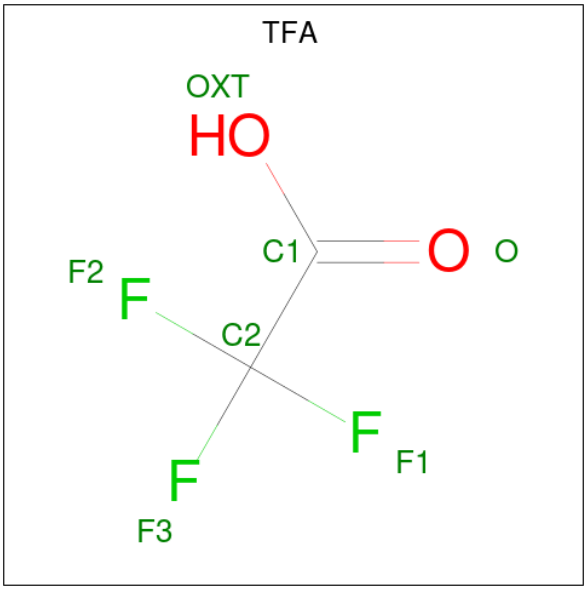
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	AA	1	Total C O 8 6 2	0	0
4	BA	1	Total C O 8 6 2	0	0
4	BA	1	Total C O 8 6 2	0	0
4	BA	1	Total C O 8 6 2	0	0
4	BA	1	Total C O 8 6 2	0	0

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

- Molecule 6 is trifluoroacetic acid (three-letter code: TFA) (formula: $C_2HF_3O_2$).



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

- Molecule 7 is water.

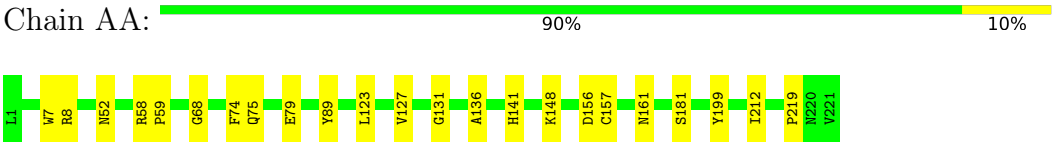
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	AA	191	Total H O 573 382 191	382	0
7	BA	201	Total H O 603 402 201	402	0
7	PB	1	Total H O 3 2 1	2	0

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.

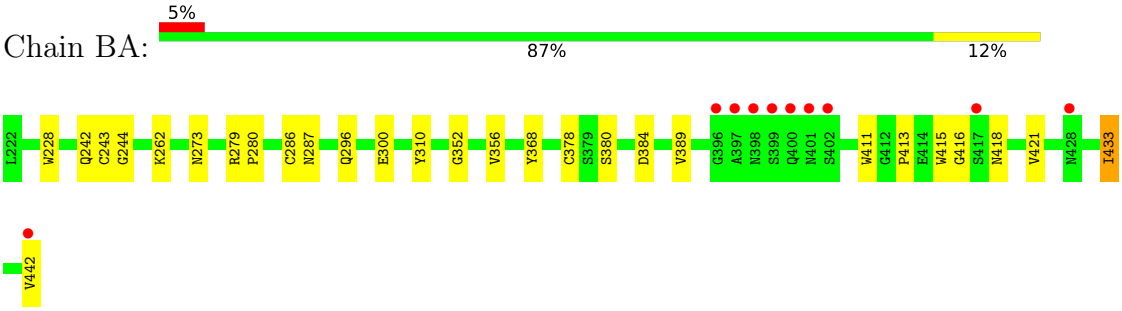
• Molecule 1: Cathepsin L2

Chain AA:



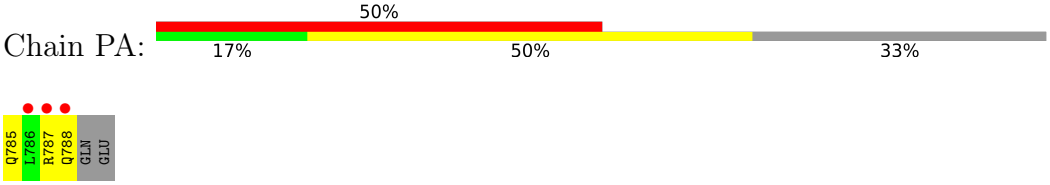
• Molecule 1: Cathepsin L2

Chain BA:



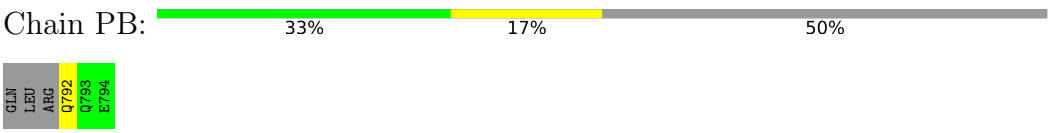
• Molecule 2: GLN-LEU-ARG-GLN

Chain PA:



• Molecule 2: GLN-LEU-ARG-GLN

Chain PB:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	93.69Å 93.69Å 124.75Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.42 – 1.83 45.41 – 1.83	Depositor EDS
% Data completeness (in resolution range)	100.0 (45.42-1.83) 96.7 (45.41-1.83)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.68 (at 1.83Å)	Xtriage
Refinement program	MAIN	Depositor
R, R_{free}	0.173 , 0.200 0.173 , 0.166	Depositor DCC
R_{free} test set	2101 reflections (4.35%)	wwPDB-VP
Wilson B-factor (Å^2)	22.0	Xtriage
Anisotropy	0.164	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	8161	wwPDB-VP
Average B, all atoms (Å^2)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.46% 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: TFA, CL, 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	0.81	0/1776	0.82	0/2399
1	BA	0.76	0/1777	0.80	0/2401
2	PA	0.63	0/36	0.74	0/46
2	PB	0.48	0/27	0.57	0/33
All	All	0.78	0/3616	0.81	0/4879

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	BA	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	BA	415	TRP	Mainchain

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	1730	1660	1655	15	1
1	BA	1735	1666	1660	19	1
2	PA	37	42	39	1	0
2	PB	28	22	21	0	0
3	AA	2	0	0	1	0
3	BA	1	0	0	0	0
4	AA	8	0	14	4	0
4	BA	32	0	56	15	1
5	AA	6	0	8	1	0
5	BA	6	0	8	0	0
6	AA	7	0	0	0	0
7	AA	191	382	0	4	4
7	BA	201	402	0	2	2
7	PB	1	2	0	0	0
All	All	3985	4176	3461	43	5

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)
4:BA:501:MPD:H53	4:BA:501:MPD:H12	1.25	1.10
4:BA:504:MPD:H53	4:BA:505:MPD:H52	1.38	1.06
1:AA:141:HIS:HD2	7:AA:434:HOH:O	1.60	0.84
4:BA:504:MPD:H53	4:BA:505:MPD:C5	2.09	0.83
1:BA:384:ASP:OD2	4:BA:501:MPD:H51	1.84	0.77

All (5) 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)
4:BA:502:MPD:CM	7:AA:407:HOH:O[6_455]	1.96	0.24
7:AA:558:HOH:O	7:BA:636:HOH:O[6_555]	2.02	0.18
7:AA:438:HOH:O	7:AA:509:HOH:O[4_455]	2.07	0.13
7:AA:407:HOH:O	7:BA:648:HOH:O[6_555]	2.14	0.06
1:AA:156[A]:ASP:OD1	1:BA:380:SER:N[4_445]	2.17	0.03

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%)	217 (96%)	8 (4%)	0	100 100
1	BA	226/221 (102%)	217 (96%)	7 (3%)	2 (1%)	17 6
2	PA	2/6 (33%)	2 (100%)	0	0	100 100
2	PB	1/6 (17%)	1 (100%)	0	0	100 100
All	All	454/454 (100%)	437 (96%)	15 (3%)	2 (0%)	47 20

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	BA	416[A]	GLY
1	BA	416[B]	GLY

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	183/178 (103%)	177 (97%)	6 (3%)	38 20
1	BA	183/178 (103%)	178 (97%)	5 (3%)	44 28
2	PA	4/6 (67%)	2 (50%)	2 (50%)	0 0
2	PB	3/6 (50%)	2 (67%)	1 (33%)	0 0
All	All	373/368 (101%)	359 (96%)	14 (4%)	38 15

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	BA	310	TYR
1	BA	378	CYS
2	PB	792	GLN
2	PA	787	ARG
2	PA	788	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	AA	104	ASN
1	AA	141	HIS
1	AA	145	GLN
1	BA	242	GLN
1	BA	441	ASN

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 11 ligands modelled in this entry, 3 are monoatomic - 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	$\#\|Z\|>2$
4	MPD	BA	501	-	7,7,7	0.35	0	9,10,10	0.28	0
5	GOL	AA	304	-	5,5,5	0.11	0	5,5,5	0.19	0
6	TFA	AA	305	-	6,6,6	1.09	0	9,9,9	1.08	1 (11%)
4	MPD	AA	302	-	7,7,7	0.30	0	9,10,10	0.41	0
4	MPD	BA	505	-	7,7,7	1.14	1 (14%)	9,10,10	0.55	0
4	MPD	BA	502	-	7,7,7	0.45	0	9,10,10	0.33	0
5	GOL	BA	506	-	5,5,5	0.30	0	5,5,5	0.31	0
4	MPD	BA	504	-	7,7,7	0.23	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
4	MPD	BA	501	-	-	1/5/5/5	-
5	GOL	AA	304	-	-	0/4/4/4	-
6	TFA	AA	305	-	-	0/6/6/6	-
4	MPD	AA	302	-	-	0/5/5/5	-
4	MPD	BA	505	-	-	1/5/5/5	-
4	MPD	BA	502	-	-	0/5/5/5	-
5	GOL	BA	506	-	-	1/4/4/4	-
4	MPD	BA	504	-	-	2/5/5/5	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	BA	505	MPD	C1-C2	-2.69	1.44	1.52

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
6	AA	305	TFA	OXT-C1-C2	2.05	121.04	112.78

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	BA	501	MPD	C2-C3-C4-C5
5	BA	506	GOL	O1-C1-C2-C3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
4	BA	504	MPD	O2-C2-C3-C4
4	BA	505	MPD	CM-C2-C3-C4
4	BA	504	MPD	C2-C3-C4-O4

There are no ring outliers.

6 monomers are involved in 21 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	BA	501	MPD	6	0
5	AA	304	GOL	1	0
4	AA	302	MPD	4	0
4	BA	505	MPD	9	0
4	BA	502	MPD	0	1
4	BA	504	MPD	4	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, 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	221/221 (100%)	-0.38	0 100 100	13, 19, 39, 59	1 (0%)
1	BA	221/221 (100%)	-0.11	10 (4%) 33 30	14, 25, 51, 104	1 (0%)
2	PA	4/6 (66%)	3.36	3 (75%) 0 0	55, 63, 127, 129	0
2	PB	3/6 (50%)	0.54	0 100 100	52, 52, 62, 95	0
All	All	449/454 (98%)	-0.21	13 (2%) 51 49	13, 22, 49, 129	2 (0%)

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	PA	788	GLN	5.3
1	BA	401	ASN	4.9
1	BA	396	GLY	4.6
2	PA	786	LEU	3.7
1	BA	398	ASN	3.3

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, 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
5	GOL	AA	304	6/6	0.53	0.74	97,99,102,107	0
5	GOL	BA	506	6/6	0.57	0.29	97,107,112,112	0
4	MPD	AA	302	8/8	0.82	0.27	59,74,81,88	0
4	MPD	BA	505	8/8	0.85	0.23	64,76,93,97	0
4	MPD	BA	504	8/8	0.86	0.24	79,84,90,91	0
4	MPD	BA	502	8/8	0.88	0.14	37,41,45,46	0
4	MPD	BA	501	8/8	0.91	0.16	39,44,51,61	0
6	TFA	AA	305	7/7	0.91	0.17	83,91,96,98	0
3	CL	AA	301	1/1	0.96	0.13	33,33,33,33	0
3	CL	BA	503	1/1	0.98	0.06	31,31,31,31	0
3	CL	AA	303	1/1	0.99	0.07	29,29,29,29	0

6.5 Other polymers (i)

There are no such residues in this entry.