wwPDB X-ray Structure Validation Summary Report (i)

Nov 7, 2023 – 03:31 am GMT

PDB ID	:	8A5B
Title	:	Crystal structure of human cathepsin L in complex with covalently bound MG-101
Authors	:	Falke, S.; Lieske, J.; Guenther, S.; Reinke, P.Y.A.; Ewert, W.; Loboda, J.; Karnicar, K.; Usenik, A.; Lindic, N.; Sekirnik, A.; Chapman, H.N.; Hinrichs, W.; Turk, D.; Meents, A.
Deposited on	:	2022-06-14
Resolution	:	1.80 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:

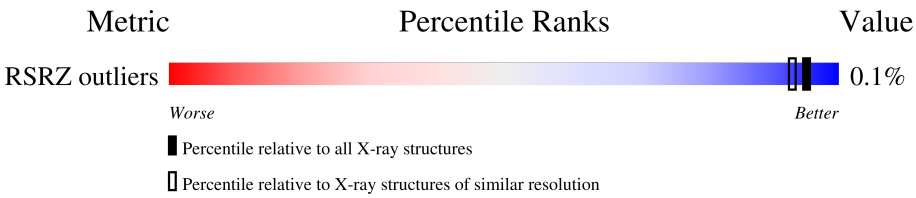
MolProbity	:	FAILED
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 1.80 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))
RSRZ outliers	127900	5850 (1.80-1.80)

MolProbity failed to run properly - the sequence quality summary graphics cannot be shown.

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 13746 atoms, of which 6530 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 L.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	220	3327	1078	1604	287	344	14	0	4	0
1	B	220	3280	1065	1577	282	342	14	0	1	0
1	C	216	3245	1054	1564	279	333	15	0	2	0
1	D	220	3311	1074	1593	284	345	15	0	3	0

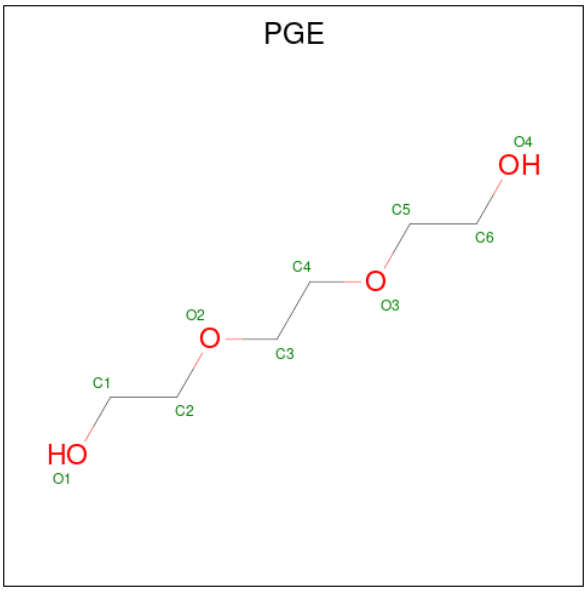
There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	110	ALA	THR	engineered mutation	UNP P07711
B	110	ALA	THR	engineered mutation	UNP P07711
C	110	ALA	THR	engineered mutation	UNP P07711
D	110	ALA	THR	engineered mutation	UNP P07711

- Molecule 2 is a protein called Calpain Inhibitor I.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	H	N	O			
2	E	4	64	20	37	3	4	0	0	0
2	F	4	64	20	37	3	4	0	0	0
2	G	4	64	20	37	3	4	0	0	0
2	H	4	64	20	37	3	4	0	0	0

- Molecule 3 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).

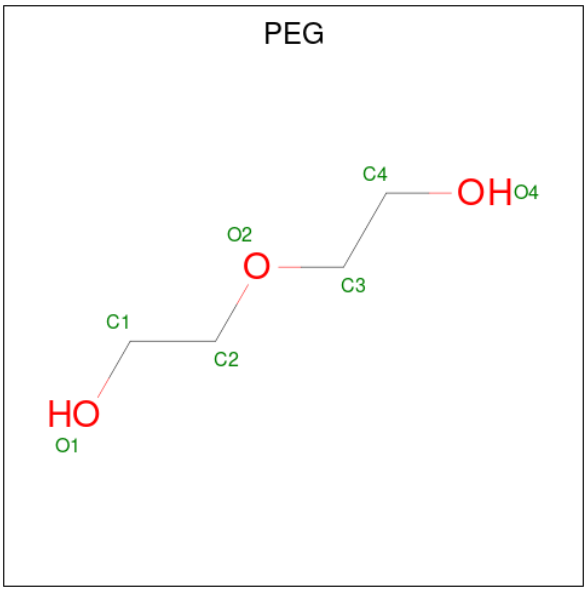


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
3	A	1	24	6	14	4	0	0

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Na		
4	A	1	1	1	0	0
			Total	Na		
4	B	2	2	2	0	0
			Total	Na		
4	D	1	1	1	0	0

- Molecule 5 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
5	C	1	17	4	10	3	0	0
			Total	C	H	O		
5	D	1	17	4	10	3	0	0
			Total	C	H	O		
5	D	1	17	4	10	3	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	A	62	62	62	0	0
			Total	O		
6	B	63	63	63	0	0
			Total	O		
6	C	45	45	45	0	0
			Total	O		
6	D	77	77	77	0	0
			Total	O		
6	F	1	1	1	0	0

MolProbity failed to run properly - this section is therefore empty.

3 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	57.24Å 62.20Å 67.24Å 105.35° 93.25° 115.81°	Depositor
Resolution (Å)	49.25 – 1.80 49.25 – 1.56	Depositor EDS
% Data completeness (in resolution range)	99.4 (49.25-1.80) 88.7 (49.25-1.56)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	-0.09 (at 1.56Å)	Xtriage
Refinement program	PHENIX 1.18_3861	Depositor
R, R_{free}	0.178 , 0.225 (Not available) , (Not available)	Depositor DCC
R_{free} test set	1994 reflections (1.84%)	wwPDB-VP
Wilson B-factor (Å^2)	21.3	Xtriage
Anisotropy	0.128	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 40.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	13746	wwPDB-VP
Average B, all atoms (Å^2)	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.60% 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.

4 Model quality (i)

4.1 Standard geometry (i)

MolProbity failed to run properly - this section is therefore empty.

4.2 Too-close contacts (i)

MolProbity failed to run properly - this section is therefore empty.

4.3 Torsion angles (i)

4.3.1 Protein backbone (i)

MolProbity failed to run properly - this section is therefore empty.

4.3.2 Protein sidechains (i)

MolProbity failed to run properly - this section is therefore empty.

4.3.3 RNA (i)

MolProbity failed to run properly - this section is therefore empty.

4.4 Non-standard residues in protein, DNA, RNA chains (i)

4 non-standard protein/DNA/RNA residues 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	$\#\|Z\| > 2$
2	2DO	G	4	2,1	7,7,8	0.43	0	6,7,9	1.25	1 (16%)
2	2DO	E	4	2,1	7,7,8	0.38	0	6,7,9	0.80	0
2	2DO	H	4	2,1	7,7,8	0.43	0	6,7,9	1.83	1 (16%)
2	2DO	F	4	2,1	7,7,8	0.20	0	6,7,9	0.80	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
2	2DO	G	4	2,1	-	2/6/6/8	-
2	2DO	E	4	2,1	-	2/6/6/8	-
2	2DO	H	4	2,1	-	3/6/6/8	-
2	2DO	F	4	2,1	-	2/6/6/8	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	H	4	2DO	CB-CA-C	-4.35	106.30	112.25
2	G	4	2DO	CB-CA-C	2.70	115.95	112.25

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	H	4	2DO	CE-CD-CG-CB
2	F	4	2DO	CE-CD-CG-CB
2	G	4	2DO	CE-CD-CG-CB
2	E	4	2DO	CE-CD-CG-CB
2	E	4	2DO	O-C-CA-N

There are no ring outliers.

No monomer is involved in short contacts.

4.5 Carbohydrates (i)

There are no monosaccharides in this entry.

4.6 Ligand geometry (i)

Of 8 ligands modelled in this entry, 4 are monoatomic - leaving 4 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	$\#\|Z\| > 2$
5	PEG	D	301	-	6,6,6	0.24	0	5,5,5	0.39	0
5	PEG	C	301	-	6,6,6	0.28	0	5,5,5	0.29	0
3	PGE	A	301	-	9,9,9	0.33	0	8,8,8	0.40	0
5	PEG	D	302	-	6,6,6	0.29	0	5,5,5	0.25	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
5	PEG	D	301	-	-	3/4/4/4	-
5	PEG	C	301	-	-	2/4/4/4	-
3	PGE	A	301	-	-	2/7/7/7	-
5	PEG	D	302	-	-	3/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 10 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	301	PGE	O1-C1-C2-O2
3	A	301	PGE	O3-C5-C6-O4
5	D	301	PEG	O2-C3-C4-O4
5	D	301	PEG	O1-C1-C2-O2
5	D	302	PEG	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

4.7 Other polymers (i)

There are no such residues in this entry.

4.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

5 Fit of model and data (i)

5.1 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	A	220/220 (100%)	-0.48	0	100	100	20, 28, 43, 61	0
1	B	220/220 (100%)	-0.48	0	100	100	18, 24, 46, 61	0
1	C	216/220 (98%)	-0.36	1 (0%)	91	89	21, 29, 49, 62	0
1	D	220/220 (100%)	-0.57	0	100	100	18, 23, 39, 51	0
2	E	2/4 (50%)	0.86	0	100	100	37, 37, 37, 44	0
2	F	2/4 (50%)	0.26	0	100	100	35, 35, 35, 37	0
2	G	2/4 (50%)	0.61	0	100	100	33, 33, 33, 41	0
2	H	2/4 (50%)	0.85	0	100	100	31, 31, 31, 35	0
All	All	884/896 (98%)	-0.46	1 (0%)	95	93	18, 26, 44, 62	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	1	ALA	4.3

5.2 Non-standard residues in protein, DNA, RNA chains (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
2	2DO	F	4	8/9	0.88	0.18	25,41,55,55	0
2	2DO	G	4	8/9	0.90	0.17	33,44,52,52	0
2	2DO	H	4	8/9	0.92	0.17	27,37,47,47	0
2	2DO	E	4	8/9	0.94	0.20	29,44,58,58	0

5.3 Carbohydrates (i)

There are no monosaccharides in this entry.

5.4 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
5	PEG	D	301	7/7	0.80	0.12	39,52,60,64	0
5	PEG	C	301	7/7	0.81	0.12	41,50,60,60	0
5	PEG	D	302	7/7	0.82	0.13	30,44,58,58	0
4	NA	A	302	1/1	0.90	0.10	38,38,38,38	0
3	PGE	A	301	10/10	0.90	0.19	36,48,65,70	0
4	NA	B	302	1/1	0.95	0.07	41,41,41,41	0
4	NA	D	303	1/1	0.97	0.08	45,45,45,45	0
4	NA	B	301	1/1	0.98	0.06	33,33,33,33	0

5.5 Other polymers (i)

There are no such residues in this entry.