wwPDB X-ray Structure Validation Summary Report (i)

Nov 1, 2023 – 10:30 AM EDT

PDB ID	:	3KSE
Title	:	Unreduced cathepsin L in complex with stefin A
Authors	:	Renko, M.; Turk, D.
Deposited on	:	2009-11-22
Resolution	:	1.71 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:

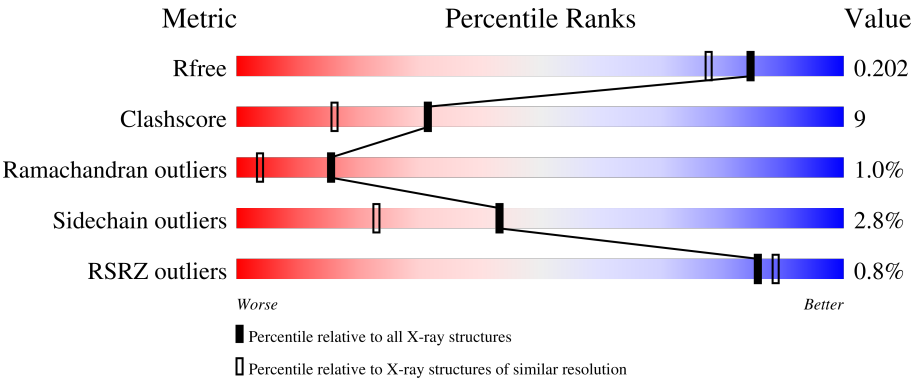
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 1.71 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	5722 (1.74-1.70)
Clashscore	141614	6152 (1.74-1.70)
Ramachandran outliers	138981	6051 (1.74-1.70)
Sidechain outliers	138945	6051 (1.74-1.70)
RSRZ outliers	127900	5629 (1.74-1.70)

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	A	220	82% 17%
1	B	220	2% 85% 15%
1	C	220	83% 15%
2	D	98	% 89% 11%
2	E	98	% 85% 14%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	98	

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
1	SCH	B	25[A]	-	-	X	-
1	SCH	B	25[B]	-	-	X	-
1	SCH	C	25[A]	-	-	X	-
1	SCH	C	25[B]	-	-	X	-

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 8650 atoms, of which 0 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 L1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	220	Total	C	N	O	S	12	9	0
			1736	1088	285	347	16			
1	B	220	Total	C	N	O	S	7	6	0
			1720	1076	281	346	17			
1	C	220	Total	C	N	O	S	8	7	0
			1721	1077	281	345	18			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	110	ALA	THR	engineered mutation	UNP P07711
A	179	ASP	ASN	engineered mutation	UNP P07711
B	110	ALA	THR	engineered mutation	UNP P07711
B	179	ASP	ASN	engineered mutation	UNP P07711
C	110	ALA	THR	engineered mutation	UNP P07711
C	179	ASP	ASN	engineered mutation	UNP P07711

- Molecule 2 is a protein called Cystatin-A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	98	Total	C	N	O	S	6	0	0
			775	492	126	155	2			
2	E	98	Total	C	N	O	S	0	0	0
			775	492	126	155	2			
2	F	98	Total	C	N	O	S	0	1	0
			780	496	127	155	2			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	261	Total	O	0	0
			261	261		

Continued on next page...

Continued from previous page...

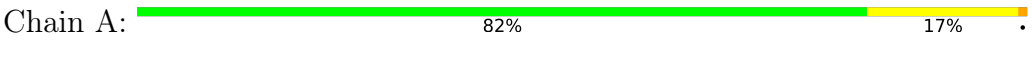
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	276	Total O 276 276	0	0
3	C	269	Total O 269 269	0	0
3	D	112	Total O 112 112	0	0
3	E	116	Total O 116 116	0	0
3	F	109	Total O 109 109	0	0

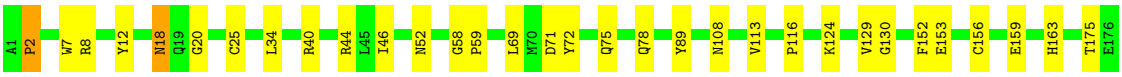
3 Residue-property plots (i)

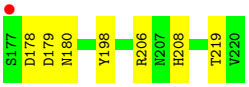
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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 L1

Chain A:





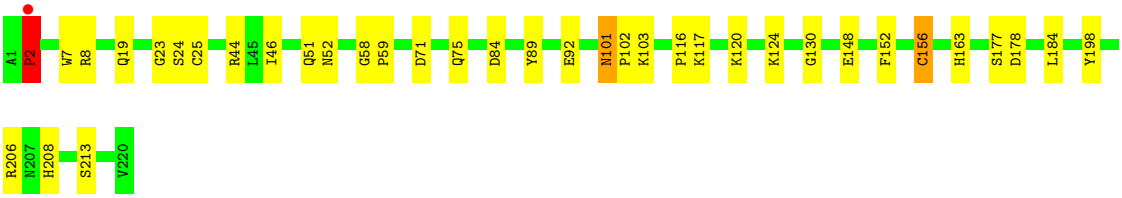


• Molecule 1: Cathepsin L1

Chain B:

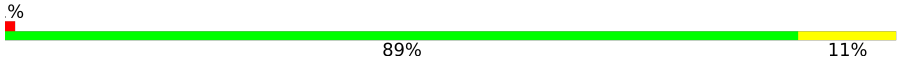
• Molecule 1: Cathepsin L1

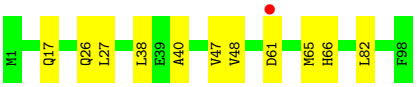
Chain C:



• Molecule 2: Cystatin-A

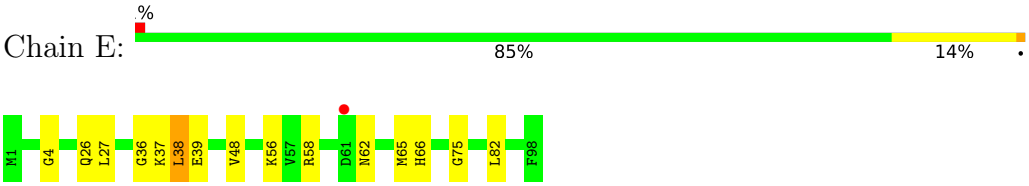
Chain D:





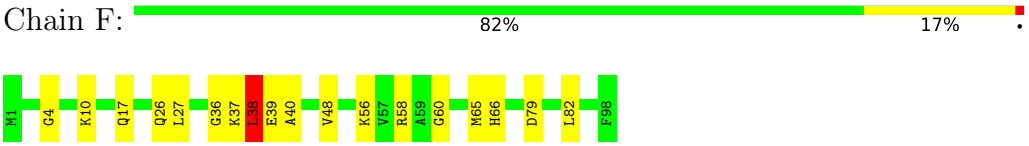
• Molecule 2: Cystatin-A

Chain E:



• Molecule 2: Cystatin-A

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	35.23Å 83.95Å 83.91Å 118.07° 98.04° 98.04°	Depositor
Resolution (Å)	15.33 – 1.71 15.34 – 1.71	Depositor EDS
% Data completeness (in resolution range)	100.0 (15.33-1.71) 95.2 (15.34-1.71)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.36 (at 1.71Å)	Xtriage
Refinement program	MAIN	Depositor
R, R_{free}	0.153 , 0.204 0.151 , 0.202	Depositor DCC
R_{free} test set	4292 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	12.9	Xtriage
Anisotropy	0.136	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 41.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.478 for h,l,-h-k-l 0.478 for h,-h-k-l,k 0.016 for -h,-k,h+k+l 0.017 for -h,h+k+l,-l 0.016 for -h,-l,-k	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	8650	wwPDB-VP
Average B, all atoms (Å^2)	18.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.87% 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: SCH

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	A	1.20	0/1801	1.02	1/2432 (0.0%)
1	B	1.21	3/1769 (0.2%)	1.08	0/2388
1	C	1.25	0/1774	1.10	3/2394 (0.1%)
2	D	1.19	0/788	1.09	1/1062 (0.1%)
2	E	1.13	0/788	1.12	1/1062 (0.1%)
2	F	1.14	0/797	1.14	2/1074 (0.2%)
All	All	1.20	3/7717 (0.0%)	1.08	8/10412 (0.1%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	159	GLU	CG-CD	5.58	1.60	1.51
1	B	113	VAL	CB-CG1	5.32	1.64	1.52
1	B	67	GLY	C-O	-5.08	1.15	1.23

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	2	PRO	N-CA-C	7.45	131.47	112.10
2	F	79	ASP	CB-CG-OD1	5.78	123.50	118.30
2	F	38	LEU	N-CA-C	-5.57	95.95	111.00
1	C	184	LEU	CB-CG-CD2	-5.14	102.26	111.00
1	A	124	LYS	CD-CE-NZ	-5.11	99.94	111.70

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	A	1736	0	1614	25	0
1	B	1720	0	1596	40	0
1	C	1721	0	1600	39	0
2	D	775	0	780	8	0
2	E	775	0	780	10	0
2	F	780	0	789	15	0
3	A	261	0	0	2	1
3	B	276	0	0	2	1
3	C	269	0	0	4	0
3	D	112	0	0	1	0
3	E	116	0	0	1	0
3	F	109	0	0	2	0
All	All	8650	0	7159	131	1

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 9.

The worst 5 of 131 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:C:19:GLN:NE2	1:C:25[B]:SCH:HE2	1.45	1.30
1:C:19:GLN:HE22	1:C:25[B]:SCH:CE	1.49	1.26
1:C:19:GLN:NE2	1:C:25[B]:SCH:CE	2.19	0.95
1:A:175:THR:HG22	1:A:180:ASN:OD1	1.68	0.93
1:B:175:THR:CG2	1:B:176:GLU:H	1.83	0.91

All (1) 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)
3:A:861:HOH:O	3:B:471:HOH:O[1_554]	2.18	0.02

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	A	226/220 (103%)	221 (98%)	4 (2%)	1 (0%)	34 18
1	B	222/220 (101%)	215 (97%)	6 (3%)	1 (0%)	29 13
1	C	223/220 (101%)	218 (98%)	4 (2%)	1 (0%)	34 18
2	D	96/98 (98%)	93 (97%)	2 (2%)	1 (1%)	15 4
2	E	96/98 (98%)	92 (96%)	2 (2%)	2 (2%)	7 1
2	F	97/98 (99%)	92 (95%)	2 (2%)	3 (3%)	4 0
All	All	960/954 (101%)	931 (97%)	20 (2%)	9 (1%)	15 4

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	2	PRO
1	B	2	PRO
1	C	2	PRO
2	F	60	GLY
2	E	36	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	A	186/177 (105%)	175 (94%)	11 (6%)	19 5
1	B	182/177 (103%)	179 (98%)	3 (2%)	62 47
1	C	183/177 (103%)	177 (97%)	6 (3%)	38 17

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	D	85/85 (100%)	85 (100%)	0	100 100
2	E	85/85 (100%)	83 (98%)	2 (2%)	49 29
2	F	86/85 (101%)	84 (98%)	2 (2%)	50 31
All	All	807/786 (103%)	783 (97%)	24 (3%)	43 21

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	2	PRO
1	C	116	PRO
1	C	101	ASN
1	C	148	GLU
1	A	116	PRO

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 32 such sidechains are listed below:

Mol	Chain	Res	Type
2	F	52	ASN
2	F	66	HIS
1	C	19	GLN
1	B	208	HIS
2	F	76	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

5 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	$\#\|Z\| > 2$
1	SCH	B	25[A]	-	6,7,8	0.49	0	3,7,9	0.28	0
1	SCH	A	25	1	6,7,8	0.36	0	3,7,9	0.66	0
1	SCH	C	25[B]	-	6,7,8	0.61	0	3,7,9	0.51	0
1	SCH	B	25[B]	-	6,7,8	0.38	0	3,7,9	0.32	0
1	SCH	C	25[A]	-	6,7,8	0.55	0	3,7,9	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
1	SCH	B	25[A]	-	-	0/2/6/8	-
1	SCH	A	25	1	-	0/2/6/8	-
1	SCH	C	25[B]	-	-	0/2/6/8	-
1	SCH	B	25[B]	-	-	0/2/6/8	-
1	SCH	C	25[A]	-	-	0/2/6/8	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5 monomers are involved in 34 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	B	25[A]	SCH	8	0
1	A	25	SCH	3	0
1	C	25[B]	SCH	13	0
1	B	25[B]	SCH	6	0
1	C	25[A]	SCH	4	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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	A	217/220 (98%)	-0.60	1 (0%)	91	92	8, 14, 26, 44	0
1	B	218/220 (99%)	-0.52	4 (1%)	68	72	8, 14, 27, 47	1 (0%)
1	C	218/220 (99%)	-0.57	1 (0%)	91	92	8, 13, 27, 41	1 (0%)
2	D	98/98 (100%)	-0.40	1 (1%)	82	85	10, 17, 32, 37	3 (3%)
2	E	98/98 (100%)	-0.39	1 (1%)	82	85	10, 16, 33, 43	0
2	F	98/98 (100%)	-0.46	0	100	100	10, 16, 35, 41	0
All	All	947/954 (99%)	-0.52	8 (0%)	86	89	8, 14, 29, 47	5 (0%)

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	175	THR	3.2
1	B	177	SER	3.0
2	E	61	ASP	2.7
1	C	2	PRO	2.4
2	D	61	ASP	2.3

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

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
1	SCH	A	25	8/9	0.95	0.10	6,9,12,16	0
1	SCH	C	25[A]	8/9	0.95	0.10	8,9,13,18	4
1	SCH	C	25[B]	8/9	0.95	0.10	8,9,13,18	4
1	SCH	B	25[A]	8/9	0.96	0.12	8,11,19,22	4

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
1	SCH	B	25[B]	8/9	0.96	0.12	8,11,19,22	4

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.