Full wwPDB EM Validation Report (i)

Oct 14, 2023 – 08:08 AM EDT

PDB ID	:	7SW1
EMDB ID	:	EMD-25459
Title	:	MicroED structure of proteinase K from a 115 nm thick lamella measured at 200 kV
Authors	:	Martynowycz, M.W.; Clabbers, M.T.B.; Unge, J.; Hattne, J.; Gonen, T.
Deposited on	:	2021-11-19
Resolution	:	1.85 Å(reported)
Based on initial model	:	6CL7

This is a Full wwPDB EM Validation 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/EMValidationReportHelp
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:

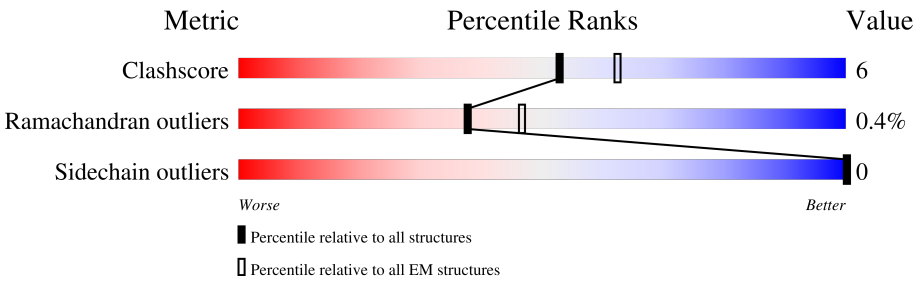
EMDB validation analysis	:	0.0.1.dev70
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	FAILED
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON CRYSTALLOGRAPHY

The reported resolution of this entry is 1.85 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)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the 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%

Mol	Chain	Length	Quality of chain
1	A	279	90% 8% ••

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2174 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 Proteinase K.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	279	2029	1247	357	415	10	0	0

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		AltConf
			Total	O	
2	A	145	145	145	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. 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. 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: Proteinase K

Chain A:

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	CRYSTALLOGRAPHY	Depositor
Imposed symmetry	3D CRYSTAL, a=67.810 Å, b=67.810 Å, c=102.66 Å, α=90°, β=90°, γ=90°, space group=P43212	Depositor
Number of images used	Not provided	
Resolution determination method	DIFFRACTION PATTERN/LAYERLINES	Depositor
CTF correction method	NONE	Depositor
Microscope	FEI TALOS ARCTICA	Depositor
Voltage (kV)	200	Depositor
Electron dose (e^-/Å^2)	0.01	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	FEI CETA (4k x 4k)	Depositor

5 Model quality (i)

5.1 Standard geometry (i)

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	0.52	2/2068 (0.1%)	0.94	12/2810 (0.4%)

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	A	0	3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	204	ASN	CG-ND2	5.40	1.46	1.32
1	A	209	TYR	CE1-CZ	5.29	1.45	1.38

All (12) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	209	TYR	CB-CG-CD2	-17.79	110.33	121.00
1	A	169	ARG	CD-NE-CZ	13.99	143.18	123.60
1	A	272	ARG	CD-NE-CZ	12.14	140.60	123.60
1	A	169	ARG	NE-CZ-NH2	-11.02	114.79	120.30
1	A	330	MET	CG-SD-CE	-7.97	87.45	100.20
1	A	209	TYR	CB-CG-CD1	6.63	124.98	121.00
1	A	272	ARG	NE-CZ-NH2	-6.51	117.05	120.30
1	A	246	VAL	CA-CB-CG2	-5.83	102.16	110.90
1	A	169	ARG	CB-CG-CD	5.76	126.57	111.60
1	A	169	ARG	N-CA-C	-5.55	96.02	111.00
1	A	169	ARG	NE-CZ-NH1	5.36	122.98	120.30
1	A	169	ARG	CA-CB-CG	5.07	124.55	113.40

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	169	ARG	Sidechain
1	A	209	TYR	Sidechain
1	A	272	ARG	Sidechain

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	2029	0	1932	22	0
2	A	145	0	0	6	1
All	All	2174	0	1932	22	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 6.

All (22) 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:A:209:TYR:CD2	1:A:246:VAL:HG21	2.17	0.80
1:A:132:GLU:OE2	2:A:401:HOH:O	2.07	0.71
1:A:159:GLN:HE21	1:A:196:PHE:HE1	1.37	0.71
1:A:169:ARG:NH1	2:A:403:HOH:O	2.29	0.65
1:A:362:ASN:HB3	1:A:365:ASP:OD1	1.98	0.63
1:A:237:SER:HB3	1:A:329:SER:OG	2.06	0.56
1:A:153:GLU:HB3	1:A:184:SER:HB2	1.90	0.54
1:A:177:HIS:ND1	1:A:330:MET:SD	2.80	0.54
1:A:287:ALA:HB1	1:A:310:GLY:HA3	1.89	0.53
1:A:206:SER:HA	2:A:446:HOH:O	2.10	0.52
1:A:169:ARG:HD2	2:A:502:HOH:O	2.11	0.51
1:A:381:ASN:ND2	2:A:411:HOH:O	2.44	0.50
1:A:159:GLN:HG2	2:A:473:HOH:O	2.14	0.47
1:A:209:TYR:HD2	1:A:246:VAL:HG21	1.75	0.47
1:A:112:PRO:HD2	1:A:115:LEU:HD12	1.97	0.47
1:A:144:ASP:OD2	1:A:174:HIS:HD2	1.98	0.47
1:A:159:GLN:NE2	1:A:196:PHE:HE1	2.11	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:177:HIS:CD2	1:A:315:SER:HB3	2.53	0.44
1:A:209:TYR:HD2	1:A:246:VAL:CG2	2.32	0.42
1:A:174:HIS:CE1	1:A:329:SER:HG	2.37	0.42
1:A:236:LEU:HB2	1:A:262:VAL:HG12	2.03	0.41
1:A:209:TYR:CD2	1:A:246:VAL:CG2	2.96	0.40

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)
2:A:527:HOH:O	2:A:539:HOH:O[6_445]	2.01	0.19

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 PDB entries followed by that with respect to all EM entries.

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	277/279 (99%)	267 (96%)	9 (3%)	1 (0%)	34 19

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	144	ASP

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 PDB entries followed by that with respect to all EM entries.

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	213/213 (100%)	213 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	159	GLN
1	A	194	GLN
1	A	381	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)

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 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-25459. These allow visual inspection of the internal detail of the map and identification of artifacts.

No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections (i)

This section was not generated.

6.2 Central slices (i)

This section was not generated.

6.3 Largest variance slices (i)

This section was not generated.

6.4 Orthogonal standard-deviation projections (False-color) (i)

This section was not generated.

6.5 Orthogonal surface views (i)

This section was not generated.

6.6 Mask visualisation (i)

This section was not generated. No masks/segmentation were deposited.

7 Map analysis (i)

This section contains the results of statistical analysis of the map.

7.1 Map-value distribution (i)

This section was not generated.

7.2 Volume estimate versus contour level (i)

This section was not generated.

7.3 Rotationally averaged power spectrum (i)

This section was not generated. The rotationally averaged power spectrum had issues being displayed.

8 Fourier-Shell correlation (i)

This section was not generated. No FSC curve or half-maps provided.

9 Map-model fit (i)

This section was not generated.