



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 25, 2024 – 04:15 pm BST

PDB ID : 9F98
Title : Crystal structure of MUS81-EME1, apo form.
Authors : Collie, G.W.
Deposited on : 2024-05-07
Resolution : 2.15 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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.13
EDS : 2.37.1
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) : 2.37.1

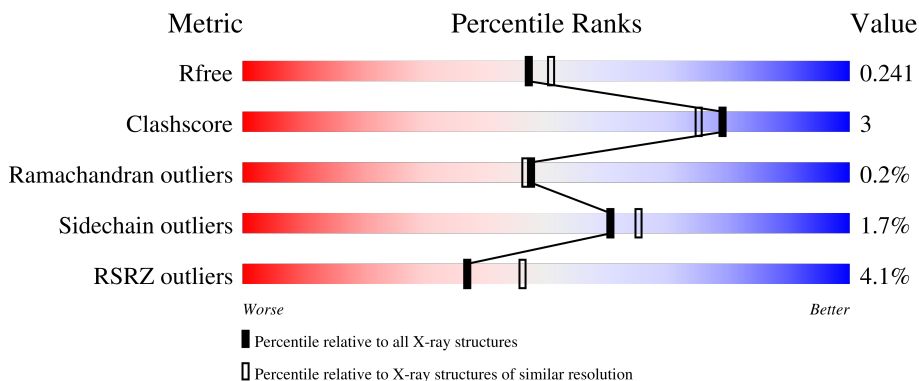
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION





The reported resolution of this entry is 2.15 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	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)
RSRZ outliers	127900	1456 (2.16-2.16)

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 $\leq 5\%$. 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	308	
1	C	308	
2	B	326	
2	D	326	

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 5178 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 Crossover junction endonuclease MUS81.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	185	1464	929	268	263	4	0	0	0
1	C	185	1426	908	252	262	4	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	244	GLY	-	expression tag	UNP Q96NY9
A	245	SER	-	expression tag	UNP Q96NY9
C	244	GLY	-	expression tag	UNP Q96NY9
C	245	SER	-	expression tag	UNP Q96NY9

- Molecule 2 is a protein called Crossover junction endonuclease EME1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	144	1069	683	172	206	8	0	0	0
2	D	138	1038	661	172	198	7	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	245	GLY	-	expression tag	UNP Q96AY2
D	245	GLY	-	expression tag	UNP Q96AY2

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	76	Total	O	0	0
			76	76		

Continued on next page...

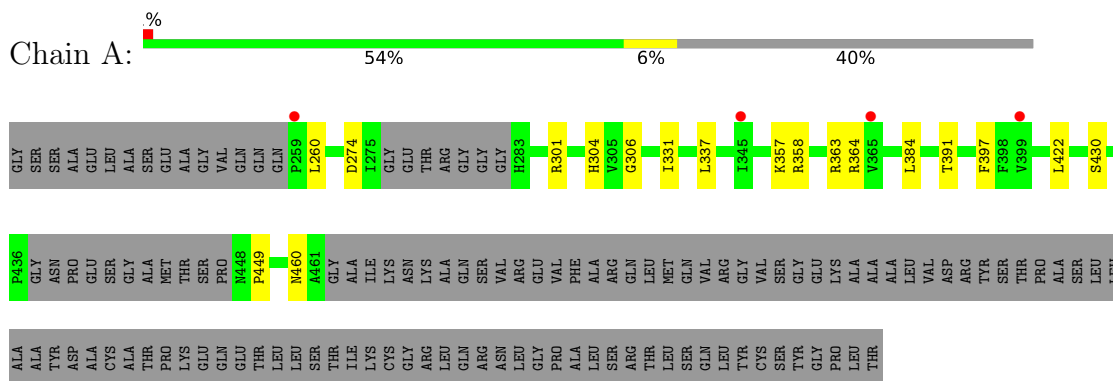
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	21	Total 21	O 21	0	0
3	C	50	Total 50	O 50	0	0
3	D	34	Total 34	O 34	0	0

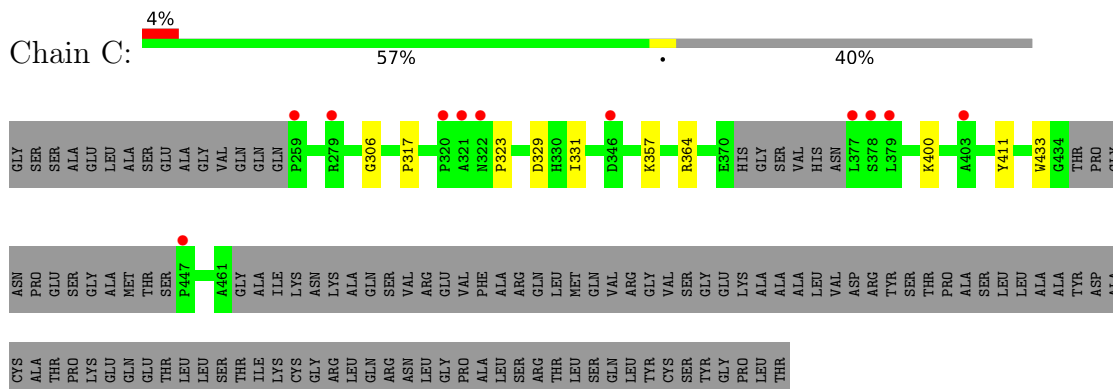
3 Residue-property plots

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.

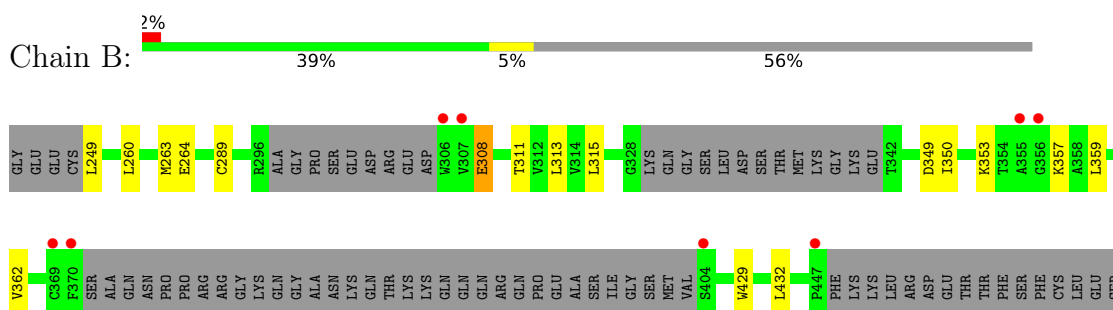
- Molecule 1: Crossover junction endonuclease MUS81



- Molecule 1: Crossover junction endonuclease MUS81



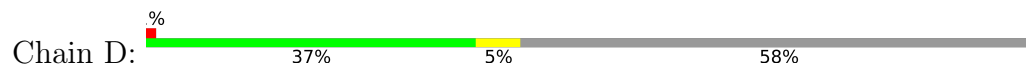
- Molecule 2: Crossover junction endonuclease EME1



ASP TRP ALA GLY VAL LYS VAL ASP ALA GLY ARG GLY VAL LEU ASN LEU LEU LEU ASP ILE GLN VAL ARG GLY ARG GLY VAL LEU THR ALA THR SER TRP ARG ARG ARG ILE ILE PRO GLN LEU LEU ASN ARG ARG ARG ILE TYR LEU MET ALA THR THR VAL VAL ASN ALA HIS LEU TYR PRO SER LEU SER PRO GLN LEU LEU VAL GLN ALA TYR GLN CYS PHE SER ASP LYS GLU ARG

GLN ASN LEU LEU ALA ASP ILE GLN VAL ARG GLY ARG GLY VAL THR ALA THR SER ARG ARG ARG ILE ILE PRO GLN LEU LEU SER ARG ARG ARG TYR LEU MET ALA THR THR VAL VAL ASN HIS LEU SER LEU SER ASP

● Molecule 2: Crossover junction endonuclease EME1



GLY GLU CYS L249 L252 L260 V287 P288 C289 R295 ARG ALA GLY THR SER PRO GLN ASP W306 V307 E309 P310 L313 L316 E319 V322 G328 LYS GLY SER LEU ASP THR MET LYS GLY THR T342 K357 V362 I363

Q366 GLU CYS PHE ALA GLN ASN PRO ARG GLY LYS GLN ALA ASN LYS GLN ALA THR SER LYS THR THR LYS GLN GLN GLN ARG GLN PRO GLU ALA ILE ILE SER MET VAL S404 Q424 A446 PRO PHE LYS LEU ARG ASP THR PHE SER CYS LEU SER

ASP TRP ALA GLY VAL LYS VAL ASP ALA GLY ARG GLY LEU LEU THR VAL TRP ARG ARG ARG ILE ILE PRO GLN LEU LEU ASN ARG ARG ARG TYR LEU MET ALA THR THR VAL VAL ASN ALA HIS LEU TYR PRO SER LEU SER PRO GLN LEU LEU VAL GLN ALA TYR GLN CYS PHE SER ASP LYS GLU ARG

GLN ASN LEU LEU ASP ILE GLN VAL ARG GLY ARG GLY VAL THR ALA THR SER TRP ARG ARG ARG ILE ILE PRO GLN LEU LEU SER ARG ARG ARG TYR LEU MET ALA THR THR VAL VAL ASN HIS LEU TYR PRO SER LEU SER ASP

4 Data and refinement statistics i

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	57.19Å 59.38Å 79.57Å 98.73° 105.06° 115.00°	Depositor
Resolution (Å)	36.66 – 2.15 36.66 – 2.15	Depositor EDS
% Data completeness (in resolution range)	97.1 (36.66-2.15) 97.1 (36.66-2.15)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.77 (at 2.16Å)	Xtrriage
Refinement program	BUSTER 2.11.8	Depositor
R, R_{free}	0.216 , 0.255 0.209 , 0.241	Depositor DCC
R_{free} test set	2151 reflections (4.66%)	wwPDB-VP
Wilson B-factor (Å ²)	39.9	Xtrriage
Anisotropy	0.153	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 49.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.020 for k,h,-h-k-l	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	5178	wwPDB-VP
Average B, all atoms (Å ²)	45.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.26% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

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	0.41	0/1494	0.57	0/2028
1	C	0.37	0/1454	0.59	0/1976
2	B	0.37	0/1082	0.53	0/1474
2	D	0.39	0/1049	0.56	0/1426
All	All	0.39	0/5079	0.56	0/6904

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 [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	1464	0	1449	10	0
1	C	1426	0	1392	6	0
2	B	1069	0	1055	7	0
2	D	1038	0	1046	7	0
3	A	76	0	0	0	0
3	B	21	0	0	0	0
3	C	50	0	0	0	0
3	D	34	0	0	0	0
All	All	5178	0	4942	28	0

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 28 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)
2:D:260:LEU:HB2	2:D:289:CYS:HA	1.72	0.71
1:A:337:LEU:HD22	1:A:384:LEU:HD13	1.81	0.63
2:B:349:ASP:HB3	2:B:353:LYS:NZ	2.15	0.62
2:B:350:ILE:HG21	2:B:359:LEU:HD13	1.95	0.49
2:D:313:LEU:HD11	2:D:362:VAL:HG23	1.95	0.48

There are no symmetry-related clashes.

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	179/308 (58%)	176 (98%)	3 (2%)	0	100	100
1	C	179/308 (58%)	177 (99%)	1 (1%)	1 (1%)	25	18
2	B	136/326 (42%)	132 (97%)	4 (3%)	0	100	100
2	D	130/326 (40%)	129 (99%)	1 (1%)	0	100	100
All	All	624/1268 (49%)	614 (98%)	9 (1%)	1 (0%)	47	46

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	329	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 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	158/260 (61%)	157 (99%)	1 (1%)	86	90
1	C	150/260 (58%)	150 (100%)	0	100	100
2	B	113/275 (41%)	109 (96%)	4 (4%)	36	34
2	D	111/275 (40%)	107 (96%)	4 (4%)	35	33
All	All	532/1070 (50%)	523 (98%)	9 (2%)	60	65

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	308	GLU
2	D	424	GLN
2	B	311	THR
2	B	432	LEU
2	D	249	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	295	HIS
1	C	448	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

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 [i](#)

6.1 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	185/308 (60%)	0.07	4 (2%) 62 69	26, 40, 56, 61	0
1	C	185/308 (60%)	0.17	11 (5%) 22 30	31, 44, 62, 71	0
2	B	144/326 (44%)	0.35	8 (5%) 24 33	33, 48, 70, 84	0
2	D	138/326 (42%)	0.07	4 (2%) 51 61	32, 45, 67, 71	0
All	All	652/1268 (51%)	0.16	27 (4%) 37 46	26, 44, 63, 84	0

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	306	TRP	4.2
2	B	307	VAL	3.9
1	A	259	PRO	3.4
1	A	399	VAL	3.3
1	C	321	ALA	3.0

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

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.