



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 8, 2026 – 03:57 PM UTC

PDB ID : 9MMU / pdb_00009mmu
Title : M5 protein with Factor H 6-7 domain
Authors : Kumar, A.; Ghosh, P.
Deposited on : 2024-12-20
Resolution : 2.25 Å(reported)

This is a Full wwPDB X-ray Structure 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/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 : 4-5-2 with Phenix2.0
Mogul : **NOT EXECUTED**
Xtrriage (Phenix) : 2.0
EDS : **NOT EXECUTED**
Buster-report : **NOT EXECUTED**
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

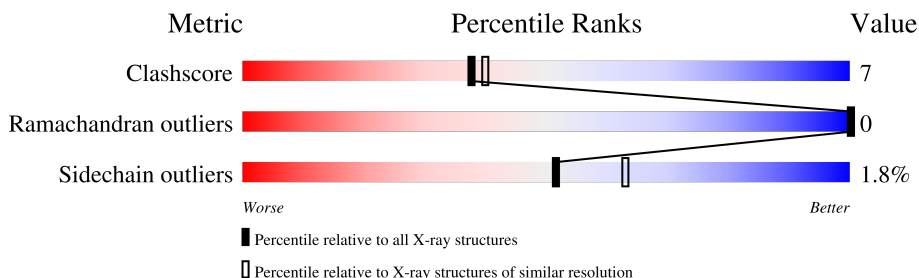
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.25 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))
Clashscore	190562	2005 (2.26-2.26)
Ramachandran outliers	187476	1965 (2.26-2.26)
Sidechain outliers	187428	1966 (2.26-2.26)

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 $\leq 5\%$

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	86	57% (green), 5% (yellow), 38% (grey)
1	B	86	52% (green), 7% (yellow), 41% (grey)
1	E	86	58% (green), 7% (yellow), 35% (grey)
1	F	86	49% (green), 8% (yellow), 43% (grey)
2	C	125	75% (green), 14% (yellow), 10% (grey)
2	D	125	67% (green), 14% (yellow), 18% (grey)
2	G	125	76% (green), 14% (yellow), 10% (grey)
2	H	125	67% (green), 13% (yellow), 20% (grey)

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 4614 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 M protein, serotype 5.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
1	A	53	355	210	69	76	0	0	0
1	B	51	338	199	64	75	0	0	0
1	E	56	378	226	70	82	0	0	0
1	F	49	336	198	62	76	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	90	GLY	-	expression tag	UNP P02977
A	91	PRO	-	expression tag	UNP P02977
A	92	GLY	-	expression tag	UNP P02977
A	93	SER	-	expression tag	UNP P02977
B	90	GLY	-	expression tag	UNP P02977
B	91	PRO	-	expression tag	UNP P02977
B	92	GLY	-	expression tag	UNP P02977
B	93	SER	-	expression tag	UNP P02977
E	90	GLY	-	expression tag	UNP P02977
E	91	PRO	-	expression tag	UNP P02977
E	92	GLY	-	expression tag	UNP P02977
E	93	SER	-	expression tag	UNP P02977
F	90	GLY	-	expression tag	UNP P02977
F	91	PRO	-	expression tag	UNP P02977
F	92	GLY	-	expression tag	UNP P02977
F	93	SER	-	expression tag	UNP P02977

- Molecule 2 is a protein called Complement factor H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	112	Total	C	N	O	S	0	0	0
			845	538	142	156	9			
2	D	103	Total	C	N	O	S	0	0	0
			739	465	124	141	9			
2	G	112	Total	C	N	O	S	0	0	0
			836	533	140	154	9			
2	H	100	Total	C	N	O	S	0	0	0
			717	453	118	137	9			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	319	MET	-	initiating methionine	UNP P08603
C	320	GLY	-	expression tag	UNP P08603
D	319	MET	-	initiating methionine	UNP P08603
D	320	GLY	-	expression tag	UNP P08603
G	319	MET	-	initiating methionine	UNP P08603
G	320	GLY	-	expression tag	UNP P08603
H	319	MET	-	initiating methionine	UNP P08603
H	320	GLY	-	expression tag	UNP P08603

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	5	Total	O	0	0
			5	5		
3	B	5	Total	O	0	0
			5	5		
3	C	23	Total	O	0	0
			23	23		
3	D	6	Total	O	0	0
			6	6		
3	E	5	Total	O	0	0
			5	5		
3	F	5	Total	O	0	0
			5	5		
3	G	14	Total	O	0	0
			14	14		
3	H	7	Total	O	0	0
			7	7		

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.

Note EDS was not executed.

- Molecule 1: M protein, serotype 5

Chain A: 



- Molecule 1: M protein, serotype 5

Chain B: 



- Molecule 1: M protein, serotype 5

Chain E: 




- Molecule 1: M protein, serotype 5

Chain F: 



- Molecule 2: Complement factor H

Chain C: 



- Molecule 2: Complement factor H

4 Data and refinement statistics i

EDS was not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, α , β , γ	62.77Å 62.77Å 199.34Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	47.72 – 2.25	Depositor
% Data completeness (in resolution range)	99.3 (47.72-2.25)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.18 (at 2.24Å)	Xtrriage
Refinement program	PHENIX 1.19.2_4158, PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.253 , 0.287	Depositor
Wilson B-factor (Å ²)	51.2	Xtrriage
Anisotropy	0.303	Xtrriage
L-test for twinning ²	$\langle L \rangle = 0.39$, $\langle L^2 \rangle = 0.22$	Xtrriage
Estimated twinning fraction	0.437 for -h,-k,l 0.459 for h,-h-k,-l 0.439 for -k,-h,-l	Xtrriage
Total number of atoms	4614	wwPDB-VP
Average B, all atoms (Å ²)	49.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.29% 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	0.12	0/355	0.23	0/485
1	B	0.17	0/338	0.28	0/462
1	E	0.12	0/378	0.23	0/517
1	F	0.12	0/336	0.20	0/459
2	C	0.11	0/880	0.29	0/1208
2	D	0.09	0/766	0.27	0/1050
2	G	0.11	0/871	0.28	0/1197
2	H	0.11	0/743	0.31	0/1018
All	All	0.12	0/4667	0.27	0/6396

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	355	0	265	6	0
1	B	338	0	240	6	0
1	E	378	0	288	6	0
1	F	336	0	252	9	0
2	C	845	0	693	12	0
2	D	739	0	539	11	0
2	G	836	0	674	11	0
2	H	717	0	512	9	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	5	0	0	1	0
3	B	5	0	0	0	0
3	C	23	0	0	4	0
3	D	6	0	0	0	0
3	E	5	0	0	2	0
3	F	5	0	0	0	0
3	G	14	0	0	3	0
3	H	7	0	0	0	0
All	All	4614	0	3463	56	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 7.

All (56) 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:125:LEU:HD21	1:B:126:GLU:HG3	1.66	0.77
1:E:126:GLU:OE2	3:E:201:HOH:O	2.04	0.73
1:A:134:TYR:O	2:C:441:ARG:NH2	2.22	0.72
2:G:354:SER:O	3:G:502:HOH:O	2.13	0.65
2:G:423:PRO:O	3:G:501:HOH:O	2.13	0.65
2:D:430:THR:HB	2:D:432:MET:HE2	1.79	0.64
2:G:401:ASN:OD1	2:G:404:ARG:NH1	2.30	0.64
2:G:349:VAL:HG13	2:G:375:THR:HA	1.82	0.62
2:H:326:ASP:O	2:H:328:PRO:HD3	1.99	0.62
1:A:126:GLU:OE2	3:A:201:HOH:O	2.16	0.60
2:C:336:TYR:N	3:C:501:HOH:O	2.35	0.60
1:E:128:GLU:OE2	3:E:202:HOH:O	2.17	0.59
2:C:411:SER:HG	2:C:430:THR:HG1	1.52	0.58
1:F:127:ARG:HH12	2:H:403:GLY:C	2.12	0.57
1:E:150:LEU:HD13	1:F:150:LEU:HA	1.87	0.56
2:C:330:ILE:HG23	2:C:384:PRO:HG2	1.88	0.55
2:G:330:ILE:HG23	2:G:384:PRO:HG2	1.88	0.54
2:G:426:GLN:N	3:G:501:HOH:O	2.27	0.54
1:B:134:TYR:OH	2:D:440:PRO:O	2.25	0.53
2:D:332:HIS:HB3	2:D:358:ASP:HB2	1.90	0.53
1:F:134:TYR:OH	2:H:440:PRO:O	2.23	0.52
2:G:386:LEU:HB3	2:G:408:GLN:HB2	1.91	0.52
2:D:387:ARG:NH1	2:D:389:CYS:SG	2.83	0.51
2:H:430:THR:HB	2:H:432:MET:HE2	1.93	0.50
1:B:117:THR:O	1:B:120:THR:OG1	2.27	0.49

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:F:127:ARG:NH1	2:H:403:GLY:O	2.46	0.49
2:C:354:SER:O	3:C:501:HOH:O	2.20	0.49
1:F:128:GLU:O	1:F:132:THR:HG23	2.12	0.48
1:B:129:VAL:O	1:B:133:GLN:HG3	2.14	0.47
2:H:332:HIS:HB3	2:H:358:ASP:HB2	1.95	0.47
2:C:399:ASN:ND2	3:C:503:HOH:O	2.47	0.47
1:F:127:ARG:HE	1:F:127:ARG:HB3	1.35	0.47
2:C:335:LEU:HD21	2:C:372:ILE:HD11	1.97	0.47
1:E:125:LEU:HB3	1:F:125:LEU:HB3	1.96	0.46
2:D:374:CYS:HA	2:D:379:TRP:HA	1.98	0.46
2:H:437:SER:HA	2:H:438:PRO:HA	1.82	0.45
2:C:325:CYS:O	2:C:347:VAL:N	2.37	0.45
2:C:399:ASN:OD1	3:C:502:HOH:O	2.21	0.45
1:E:128:GLU:O	1:E:132:THR:HG23	2.17	0.44
2:G:429:VAL:HG12	2:G:438:PRO:HG2	1.99	0.44
2:G:336:TYR:CZ	2:G:356:TYR:HB3	2.52	0.44
2:H:334:GLY:HA2	2:H:356:TYR:H	1.83	0.44
1:A:125:LEU:HD11	2:D:359:GLU:HB3	2.00	0.43
2:D:437:SER:HA	2:D:438:PRO:HA	1.83	0.43
1:A:126:GLU:HG3	2:C:390:TYR:HE2	1.83	0.43
2:G:364:PRO:HG2	2:G:381:PRO:HG3	2.01	0.43
2:C:420:TYR:HB3	2:C:442:CYS:HB3	2.00	0.42
2:D:334:GLY:HA2	2:D:356:TYR:H	1.84	0.42
2:G:381:PRO:HG2	2:G:384:PRO:HA	2.01	0.42
1:E:129:VAL:HG23	1:F:129:VAL:HG23	2.01	0.42
1:B:126:GLU:HB3	2:D:390:TYR:HE2	1.86	0.41
1:F:127:ARG:HD2	2:H:390:TYR:CD1	2.56	0.41
2:D:364:PRO:HG3	2:D:381:PRO:HG3	2.03	0.41
1:A:129:VAL:HG23	1:B:129:VAL:HG23	2.03	0.40
2:D:359:GLU:H	2:D:359:GLU:HG3	1.56	0.40
2:C:387:ARG:HB2	2:C:433:GLU:HA	2.04	0.40

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	51/86 (59%)	50 (98%)	1 (2%)	0	100	100
1	B	49/86 (57%)	49 (100%)	0	0	100	100
1	E	54/86 (63%)	52 (96%)	2 (4%)	0	100	100
1	F	47/86 (55%)	47 (100%)	0	0	100	100
2	C	108/125 (86%)	102 (94%)	6 (6%)	0	100	100
2	D	93/125 (74%)	85 (91%)	8 (9%)	0	100	100
2	G	108/125 (86%)	102 (94%)	6 (6%)	0	100	100
2	H	88/125 (70%)	82 (93%)	6 (7%)	0	100	100
All	All	598/844 (71%)	569 (95%)	29 (5%)	0	100	100

There are no Ramachandran outliers to report.

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	26/79 (33%)	26 (100%)	0	100	100
1	B	24/79 (30%)	24 (100%)	0	100	100
1	E	29/79 (37%)	29 (100%)	0	100	100
1	F	27/79 (34%)	27 (100%)	0	100	100
2	C	81/109 (74%)	80 (99%)	1 (1%)	63	74
2	D	62/109 (57%)	59 (95%)	3 (5%)	23	25
2	G	78/109 (72%)	78 (100%)	0	100	100
2	H	60/109 (55%)	57 (95%)	3 (5%)	22	24
All	All	387/752 (52%)	380 (98%)	7 (2%)	51	63

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	380	SER
2	D	427	THR
2	D	430	THR
2	D	439	THR
2	H	325	CYS
2	H	374	CYS
2	H	442	CYS

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

Mol	Chain	Res	Type
2	D	396	ASN
1	F	113	GLN
2	G	434	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](#)

Mogul was not executed - this section is therefore empty.

5.5 Carbohydrates [i](#)

Mogul was not executed - this section is therefore empty.

5.6 Ligand geometry [i](#)

Mogul was not executed - this section is therefore empty.

5.7 Other polymers [i](#)

Mogul was not executed - this section is therefore empty.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates [i](#)

EDS was not executed - this section is therefore empty.

6.4 Ligands [i](#)

EDS was not executed - this section is therefore empty.

6.5 Other polymers [i](#)

EDS was not executed - this section is therefore empty.