Full wwPDB X-ray Structure Validation Report (i)

Jan 8, 2024 – 09:53 am GMT

PDB ID	:	5MV9
Title	:	Structure of human Myosin 7a C-terminal MyTH4-FERM domain in complex with harmonin-a PDZ3 domain
Authors	:	Yu, I.; Sourigues, Y.; Titus, M.A.; Houdusse, A.
Deposited on	:	2017-01-16
Resolution	:	2.60 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 (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:

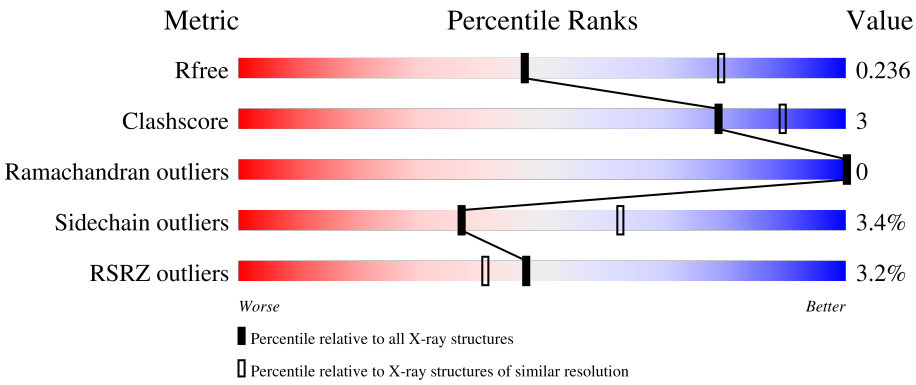
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 2.60 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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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	527	2%, 82%, 9%, 9%
2	B	130	7%, 82%, 9%, • 8%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5073 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 Unconventional myosin-VIIa.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	480	3858	2517	622	706	13	0	1	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1689	MET	-	initiating methionine	UNP Q13402
A	1690	GLY	-	expression tag	UNP Q13402
A	1691	SER	-	expression tag	UNP Q13402
A	1692	SER	-	expression tag	UNP Q13402
A	1693	HIS	-	expression tag	UNP Q13402
A	1694	HIS	-	expression tag	UNP Q13402
A	1695	HIS	-	expression tag	UNP Q13402
A	1696	HIS	-	expression tag	UNP Q13402
A	1697	HIS	-	expression tag	UNP Q13402
A	1698	HIS	-	expression tag	UNP Q13402
A	1699	SER	-	expression tag	UNP Q13402
A	1700	SER	-	expression tag	UNP Q13402
A	1701	GLY	-	expression tag	UNP Q13402

- Molecule 2 is a protein called cDNA FLJ51329, highly similar to Harmonin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	120	879	566	135	174	4	0	2	0

There are 5 discrepancies between the modelled and reference sequences:

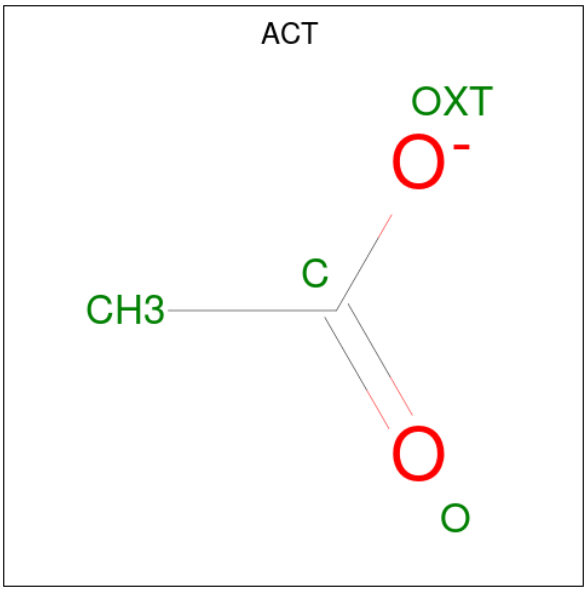
Chain	Residue	Modelled	Actual	Comment	Reference
B	423	GLY	-	expression tag	UNP B4DV53
B	424	ALA	-	expression tag	UNP B4DV53
B	425	MET	-	expression tag	UNP B4DV53

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	426	GLY	-	expression tag	UNP B4DV53
B	427	SER	-	expression tag	UNP B4DV53

- Molecule 3 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
3	A	1	4	2	2	0	0
			Total	C	O		
3	A	1	4	2	2	0	0

- Molecule 4 is water.

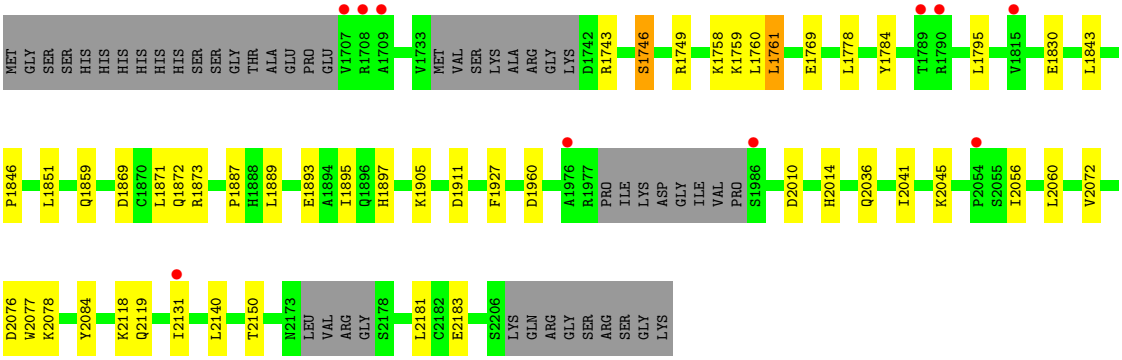
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	291	291	291	0	0
			Total	O		
4	B	37	37	37	0	0

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.

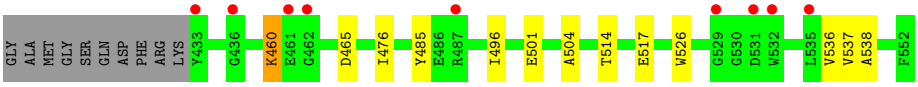
• Molecule 1: Unconventional myosin-VIIa

Chain A:



• Molecule 2: cDNA FLJ51329, highly similar to Harmonin

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	121.52Å 151.71Å 100.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.84 – 2.60 19.84 – 2.60	Depositor EDS
% Data completeness (in resolution range)	96.8 (19.84-2.60) 97.3 (19.84-2.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.05	Depositor
$< I/\sigma(I) >$ [1]	1.24 (at 2.59Å)	Xtriage
Refinement program	BUSTER 2.10.2	Depositor
R, R_{free}	0.180 , 0.237 0.175 , 0.236	Depositor DCC
R_{free} test set	1400 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	65.2	Xtriage
Anisotropy	0.396	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 64.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	5073	wwPDB-VP
Average B, all atoms (Å^2)	75.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.23% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: ACT

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	0.49	0/3965	0.68	0/5384
2	B	0.45	0/903	0.72	0/1231
All	All	0.49	0/4868	0.69	0/6615

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	3858	0	3759	20	0
2	B	879	0	824	7	0
3	A	8	0	6	0	0
4	A	291	0	0	1	0
4	B	37	0	0	0	0
All	All	5073	0	4589	26	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.

All (26) 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:1778:LEU:HB3	1:A:1784:TYR:HB2	1.80	0.64
1:A:1869:ASP:O	1:A:1873:ARG:HG3	1.99	0.62
2:B:460:LYS:HA	2:B:526:TRP:CH2	2.42	0.55
1:A:2181:LEU:HD21	2:B:476:ILE:HG22	1.92	0.51
1:A:1846:PRO:HG2	1:A:1851:LEU:HA	1.93	0.50
2:B:465:ASP:HB3	2:B:485:TYR:HB2	1.92	0.49
2:B:504:ALA:HB3	2:B:536:VAL:HB	1.94	0.49
1:A:1893:GLU:O	1:A:1897:HIS:HD2	1.96	0.49
1:A:2131:ILE:HD12	1:A:2140:LEU:HD22	1.94	0.49
2:B:501:GLU:HB3	2:B:538:ALA:HB3	1.96	0.47
1:A:2060:LEU:HD21	1:A:2077:TRP:CD1	2.50	0.47
1:A:1743:ARG:HB2	1:A:1746:SER:HB2	1.97	0.47
1:A:1758:LYS:O	1:A:1761:LEU:HB2	2.15	0.47
1:A:2056:ILE:HB	1:A:2078:LYS:HG3	1.98	0.46
1:A:2118:LYS:HB3	1:A:2183:GLU:HB2	1.98	0.45
1:A:2010:ASP:HA	1:A:2014:HIS:HB2	2.00	0.44
1:A:2140:LEU:HG	1:A:2150:THR:HG23	1.98	0.43
2:B:496:ILE:HD12	2:B:537:VAL:HG11	2.01	0.43
1:A:1893:GLU:O	1:A:1897:HIS:CD2	2.71	0.43
1:A:2036:GLN:HG2	1:A:2084:TYR:CD1	2.53	0.43
1:A:1895:ILE:HD12	1:A:1895:ILE:HA	1.93	0.42
1:A:1749:ARG:HD2	4:A:2481:HOH:O	2.18	0.42
2:B:514:THR:HG23	2:B:517:GLU:H	1.84	0.42
1:A:1843:LEU:HD11	1:A:1887:PRO:HD3	2.02	0.41
1:A:1859:GLN:HG3	1:A:1871:LEU:HD11	2.02	0.41
1:A:2041:ILE:HG22	1:A:2045:LYS:HD2	2.02	0.41

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	473/527 (90%)	454 (96%)	19 (4%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	120/130 (92%)	110 (92%)	10 (8%)	0	100	100
All	All	593/657 (90%)	564 (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	415/478 (87%)	398 (96%)	17 (4%)	30	56
2	B	86/103 (84%)	85 (99%)	1 (1%)	71	87
All	All	501/581 (86%)	483 (96%)	18 (4%)	37	61

All (18) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1746	SER
1	A	1759	LYS
1	A	1760	LEU
1	A	1761	LEU
1	A	1769[A]	GLU
1	A	1769[B]	GLU
1	A	1795	LEU
1	A	1830	GLU
1	A	1872	GLN
1	A	1889	LEU
1	A	1905	LYS
1	A	1911	ASP
1	A	1927	PHE
1	A	1960	ASP
1	A	2072	VAL
1	A	2076	ASP
1	A	2119	GLN
2	B	460	LYS

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	1880	ASN
1	A	1897	HIS
1	A	2119	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)

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)

2 ligands 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	$\#\|Z\| > 2$
3	ACT	A	2301	-	3,3,3	1.15	0	3,3,3	1.19	0
3	ACT	A	2302	-	3,3,3	1.28	0	3,3,3	0.79	0

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.

No monomer is involved in short contacts.

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 (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	480/527 (91%)	-0.26	10 (2%)	63	58	46, 68, 102, 121	0
2	B	120/130 (92%)	0.11	9 (7%)	14	10	60, 88, 135, 159	0
All	All	600/657 (91%)	-0.18	19 (3%)	47	40	46, 71, 113, 159	0

All (19) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	462	GLY	4.6
1	A	1709	ALA	4.2
2	B	529	GLY	4.1
1	A	1707	VAL	3.2
2	B	531	ASP	3.2
1	A	1708	ARG	3.1
1	A	1976	ALA	2.9
1	A	1790	ARG	2.8
2	B	461	GLU	2.7
1	A	1789	THR	2.6
1	A	1986	SER	2.6
1	A	2131	ILE	2.4
2	B	532	TRP	2.4
2	B	535	LEU	2.3
1	A	2054	PRO	2.3
1	A	1815	VAL	2.3
2	B	433	TYR	2.1
2	B	487	ARG	2.1
2	B	436	GLY	2.1

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)

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
3	ACT	A	2302	4/4	0.45	0.31	97,98,99,100	0
3	ACT	A	2301	4/4	0.86	0.24	59,65,65,70	0

6.5 Other polymers (i)

There are no such residues in this entry.