wwPDB X-ray Structure Validation Summary Report (i)

Sep 13, 2020 – 07:17 PM BST

PDB ID	:	5EJQ
Title	:	Structure of Dictyostelium Discoideum Myosin VII MyTH4-FERM MF1 domain, mutant 2
Authors	:	Planelles-Herrero, V.J.; Sirkia, H.; Sourigues, Y.; Titus, M.A.; Houdusse, A.
Deposited on	:	2015-11-02
Resolution	:	2.19 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 following versions of software and data (see references (i)) were used in the production of this report:

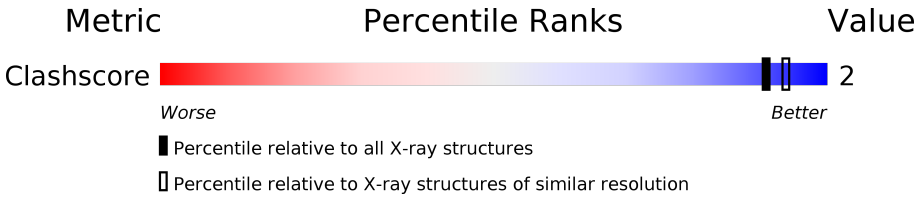
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	FAILED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.14.4.dev1

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 2.19 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	141614	5594 (2.20-2.20)

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 on 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%

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	A	511	93%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 4607 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.

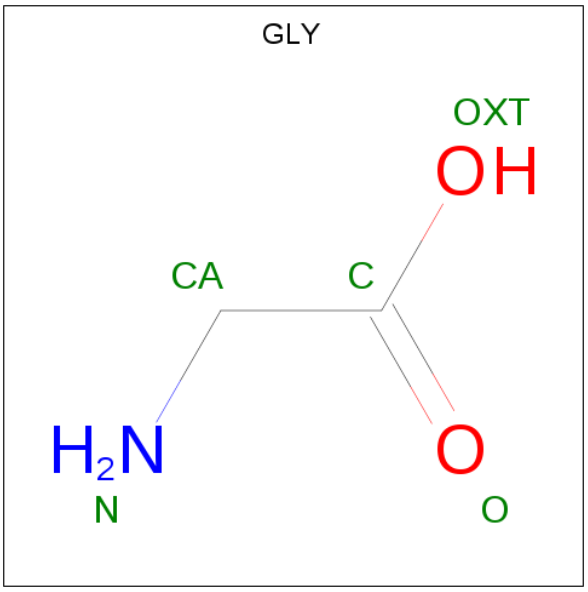
- Molecule 1 is a protein called Myosin-I heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	498	Total	C	N	O	S	0	5	0
			4073	2589	681	791	12			

There are 13 discrepancies between the modelled and reference sequences:

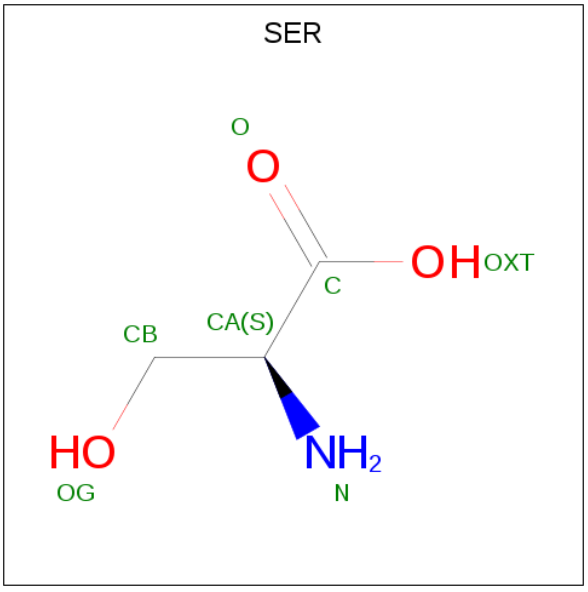
Chain	Residue	Modelled	Actual	Comment	Reference
A	1157	GLU	LYS	engineered mutation	UNP Q9U1M8
A	1159	GLU	HIS	conflict	UNP Q9U1M8
A	1161	GLU	LYS	engineered mutation	UNP Q9U1M8
A	1174	GLU	LYS	engineered mutation	UNP Q9U1M8
A	1574J	SER	-	insertion	UNP Q9U1M8
A	1621	LEU	-	expression tag	UNP Q9U1M8
A	1622	GLU	-	expression tag	UNP Q9U1M8
A	1623	HIS	-	expression tag	UNP Q9U1M8
A	1624	HIS	-	expression tag	UNP Q9U1M8
A	1625	HIS	-	expression tag	UNP Q9U1M8
A	1626	HIS	-	expression tag	UNP Q9U1M8
A	1627	HIS	-	expression tag	UNP Q9U1M8
A	1628	HIS	-	expression tag	UNP Q9U1M8

- Molecule 2 is GLYCINE (three-letter code: GLY) (formula: $C_2H_5NO_2$).



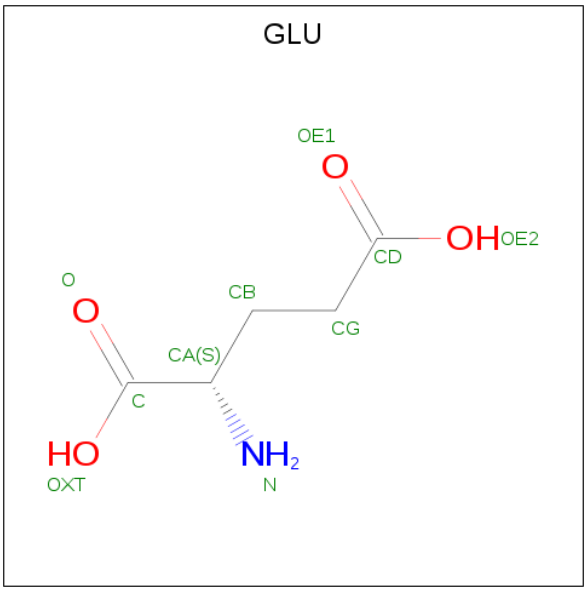
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	5	2	1	2	0	0

- Molecule 3 is SERINE (three-letter code: SER) (formula: $C_3H_7NO_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
3	A	1	7	3	1	3	0	0

- Molecule 4 is GLUTAMIC ACID (three-letter code: GLU) (formula: $C_5H_9NO_4$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	N	O	0	0
			10	5	1	4		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	512	Total	O	0	0
			512	512		

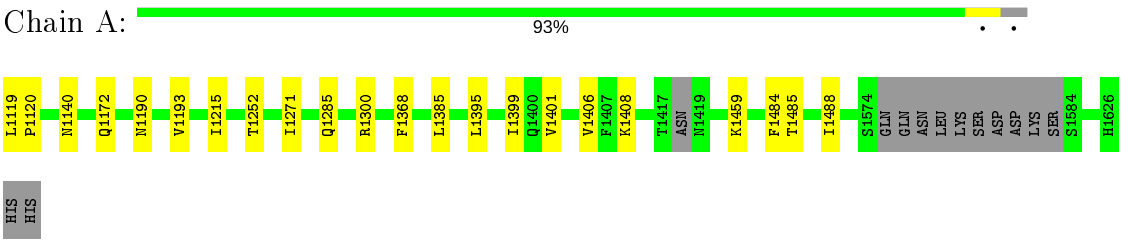
3 Residue-property plots (i)

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 failed to run properly.

- Molecule 1: Myosin-I heavy chain

Chain A:



4 Data and refinement statistics (i)

EDS failed to run properly - this section is therefore incomplete.

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	53.32Å 61.48Å 172.62Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.01 – 2.19	Depositor
% Data completeness (in resolution range)	100.0 (19.01-2.19)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	213.41 (at 2.19Å)	Xtriage
Refinement program	BUSTER-TNT 2.10.1	Depositor
R, R_{free}	0.185 , 0.217	Depositor
Wilson B-factor (Å^2)	26.1	Xtriage
Anisotropy	0.030	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.53$, $< L^2 > = 0.38$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	4607	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.03% 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

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	0.50	0/4151	0.63	0/5604

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

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	4073	0	4032	14	0
2	A	5	0	2	1	0
3	A	7	0	4	0	0
4	A	10	0	5	1	0
5	A	512	0	0	1	0
All	All	4607	0	4043	14	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 2.

The worst 5 of 14 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:1140:ASN:HD21	4:A:1703:GLU:HB2	1.62	0.64

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1395:LEU:HD11	1:A:1399:ILE:H	1.69	0.58
1:A:1385:LEU:HB3	2:A:1701:GLY:HA3	1.88	0.54
1:A:1395:LEU:CD1	1:A:1399:ILE:HG13	2.41	0.51
1:A:1190:ASN:HB3	1:A:1193:VAL:HG23	1.93	0.51

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

There are no protein backbone outliers to report in this entry.

5.3.2 Protein sidechains (i)

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

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)

3 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	SER	A	1702	-	3,6,6	0.17	0	1,7,7	0.90	0
4	GLU	A	1703	-	2,9,9	0.18	0	2,11,11	1.45	0
2	GLY	A	1701	-	1,4,4	0.03	0	0,4,4	0.00	-

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
3	SER	A	1702	-	-	0/2/6/6	-
4	GLU	A	1703	-	-	0/3/9/9	-
2	GLY	A	1701	-	-	0/0/2/2	-

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.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1703	GLU	1	0
2	A	1701	GLY	1	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	1625:HIS	C	1626:HIS	N	3.94

6 Fit of model and data

6.1 Protein, DNA and RNA chains

EDS failed to run properly - this section is therefore empty.

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

EDS failed to run properly - this section is therefore empty.

6.3 Carbohydrates

EDS failed to run properly - this section is therefore empty.

6.4 Ligands

EDS failed to run properly - this section is therefore empty.

6.5 Other polymers

EDS failed to run properly - this section is therefore empty.