wwPDB X-ray Structure Validation Summary Report (i)

May 26, 2020 – 09:59 am BST

PDB ID	:	6EJM
Title	:	CRYSTAL STRUCTURE OF HUMAN CD81 LARGE EXTRACELLULAR LOOP IN COMPLEX WITH SINGLE CHAIN FV FRAGMENT 5
Authors	:	Kuglstatter, A.; Harris, S.F.; Villasenor, A.
Deposited on	:	2017-09-22
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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

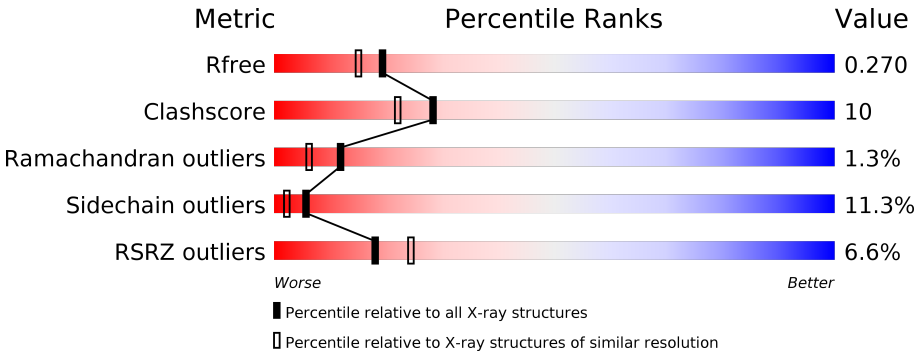
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.15 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 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% 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	99	15% 68% 22% • 8%
1	B	99	11% 65% 21% •• 11%
2	H	249	4% 69% 19% •• 8%
2	I	249	3% 70% 16% •• 9%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 5072 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 CD81 antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	91	Total	C	N	O	S	0	0	0
			705	440	123	138	4			
1	B	88	Total	C	N	O	S	0	0	0
			680	425	116	135	4			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	110	GLY	-	expression tag	UNP P60033
A	111	SER	-	expression tag	UNP P60033
A	203	HIS	-	expression tag	UNP P60033
A	204	HIS	-	expression tag	UNP P60033
A	205	HIS	-	expression tag	UNP P60033
A	206	HIS	-	expression tag	UNP P60033
A	207	HIS	-	expression tag	UNP P60033
A	208	HIS	-	expression tag	UNP P60033
B	110	GLY	-	expression tag	UNP P60033
B	111	SER	-	expression tag	UNP P60033
B	203	HIS	-	expression tag	UNP P60033
B	204	HIS	-	expression tag	UNP P60033
B	205	HIS	-	expression tag	UNP P60033
B	206	HIS	-	expression tag	UNP P60033
B	207	HIS	-	expression tag	UNP P60033
B	208	HIS	-	expression tag	UNP P60033

- Molecule 2 is a protein called SINGLE CHAIN FV FRAGMENT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	229	Total	C	N	O	S	0	0	0
			1733	1100	289	337	7			
2	I	227	Total	C	N	O	S	0	0	0
			1725	1096	287	335	7			

- Molecule 3 is water.

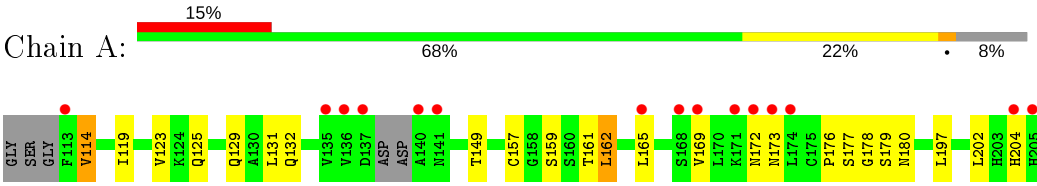
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	16	Total O 16 16	0	0
3	B	17	Total O 17 17	0	0
3	H	100	Total O 100 100	0	0
3	I	96	Total O 96 96	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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.

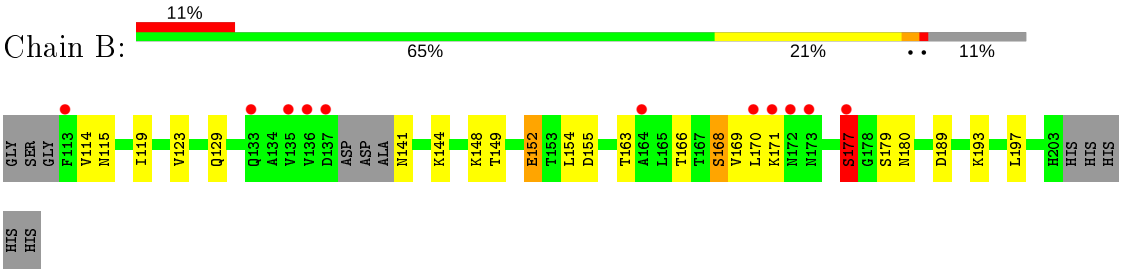
- Molecule 1: CD81 antigen

Chain A:



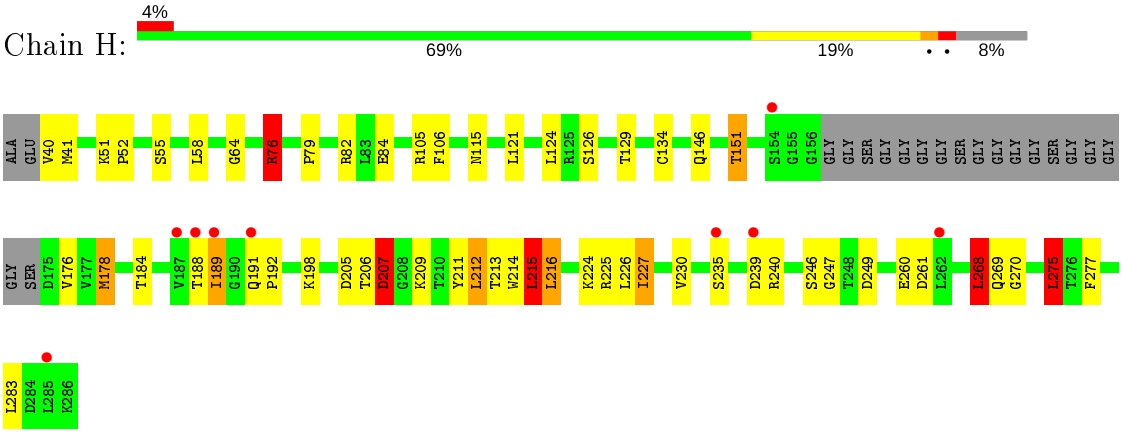
- Molecule 1: CD81 antigen

Chain B:



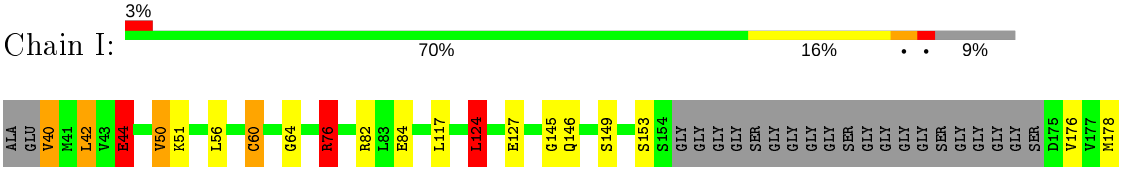
- Molecule 2: SINGLE CHAIN FV FRAGMENT

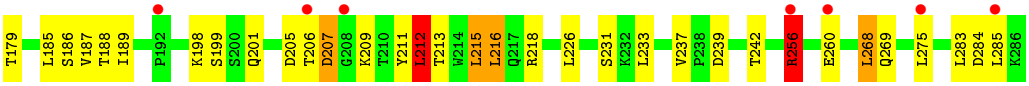
Chain H:



- Molecule 2: SINGLE CHAIN FV FRAGMENT

Chain I:




T179
L185
S186
V187
T188
I189
P192
K198
S199
S200
Q201
D205
T206
D207
G208
K209
T210
Y211
L212
T213
W214
L215
L216
Q217
R218
L226
S231
K232
L233
V237
P238
D239
T242
R250
E260
L268
Q269
L275
L283
D284
L285
K286

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	73.35Å 100.50Å 116.67Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	41.57 – 2.15 41.57 – 2.15	Depositor EDS
% Data completeness (in resolution range)	85.1 (41.57-2.15) 85.1 (41.57-2.15)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	1.43 (at 2.16Å)	Xtriage
Refinement program	REFMAC 5.5.0072	Depositor
R, R_{free}	0.215 , 0.278 0.209 , 0.270	Depositor DCC
R_{free} test set	2064 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	36.2	Xtriage
Anisotropy	0.370	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 50.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	5072	wwPDB-VP
Average B, all atoms (Å^2)	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.99% 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	0.97	1/715 (0.1%)	0.83	0/964
1	B	0.98	0/688	0.84	2/927 (0.2%)
2	H	0.97	2/1769 (0.1%)	1.03	9/2397 (0.4%)
2	I	0.98	3/1761 (0.2%)	1.04	11/2387 (0.5%)
All	All	0.97	6/4933 (0.1%)	0.98	22/6675 (0.3%)

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
2	H	0	1

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	I	60	CYS	CB-SG	-6.99	1.70	1.82
2	I	44	GLU	CG-CD	-6.58	1.42	1.51
2	H	260	GLU	CG-CD	6.23	1.61	1.51
1	A	157	CYS	CB-SG	5.38	1.91	1.82
2	I	40	VAL	CA-CB	5.34	1.66	1.54

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	I	76	ARG	NE-CZ-NH1	8.21	124.41	120.30
2	H	215	LEU	CA-CB-CG	7.54	132.65	115.30
2	H	212	LEU	CA-CB-CG	-6.98	99.24	115.30
2	I	56	LEU	CA-CB-CG	6.82	130.97	115.30
2	H	178	MET	CG-SD-CE	-6.78	89.35	100.20

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	H	206	THR	Peptide

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	705	0	688	18	0
1	B	680	0	669	21	0
2	H	1733	0	1720	28	0
2	I	1725	0	1714	36	0
3	A	16	0	0	1	0
3	B	17	0	0	0	0
3	H	100	0	0	1	0
3	I	96	0	0	3	0
All	All	5072	0	4791	94	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 10.

The worst 5 of 94 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:123:VAL:HG22	1:B:119:ILE:HD11	1.42	1.00
2:H:227:ILE:HD11	2:H:230:VAL:O	1.74	0.86
1:A:132:GLN:HG3	1:A:165:LEU:HD11	1.55	0.85
2:H:178:MET:CE	2:H:269:GLN:HG2	2.09	0.83
1:A:162:LEU:HD12	1:A:165:LEU:HD12	1.63	0.81

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	87/99 (88%)	82 (94%)	4 (5%)	1 (1%)	14 8
1	B	84/99 (85%)	77 (92%)	5 (6%)	2 (2%)	6 1
2	H	225/249 (90%)	214 (95%)	8 (4%)	3 (1%)	12 6
2	I	223/249 (90%)	218 (98%)	3 (1%)	2 (1%)	17 11
All	All	619/696 (89%)	591 (96%)	20 (3%)	8 (1%)	12 6

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	177	SER
2	H	207	ASP
2	I	207	ASP
1	B	114	VAL
2	H	247	GLY

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	81/87 (93%)	74 (91%)	7 (9%)	10 6
1	B	79/87 (91%)	73 (92%)	6 (8%)	13 8
2	H	191/196 (97%)	166 (87%)	25 (13%)	4 1
2	I	191/196 (97%)	168 (88%)	23 (12%)	5 2
All	All	542/566 (96%)	481 (89%)	61 (11%)	6 2

5 of 61 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	H	215	LEU
2	H	246	SER
2	I	242	THR
2	H	216	LEU
2	H	226	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	129	GLN
2	I	221	GLN
1	B	173	ASN
1	A	173	ASN
2	H	115	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 carbohydrates 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

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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, 95^{th} 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	91/99 (91%)	0.74	15 (16%) 1 2	24, 39, 79, 84	0
1	B	88/99 (88%)	0.51	11 (12%) 3 5	21, 37, 70, 74	0
2	H	229/249 (91%)	0.22	9 (3%) 39 48	18, 33, 56, 60	0
2	I	227/249 (91%)	0.18	7 (3%) 49 58	20, 36, 60, 69	0
All	All	635/696 (91%)	0.32	42 (6%) 18 24	18, 35, 64, 84	0

The worst 5 of 42 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	169	VAL	6.8
1	B	136	VAL	6.8
1	A	172	ASN	5.1
1	A	141	ASN	4.9
2	I	285	LEU	4.9

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

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

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.