



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 08:21 am BST

PDB ID : 6EK2  
Title : CRYSTAL STRUCTURE OF HUMAN CD81 LARGE EXTRACELLULAR LOOP IN COMPLEX WITH SINGLE CHAIN FV FRAGMENT 10  
Authors : Harris, S.F.; Villasenor, A.; Kuglstatter, A.  
Deposited on : 2017-09-25  
Resolution : 2.65 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](mailto: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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Xtrriage (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

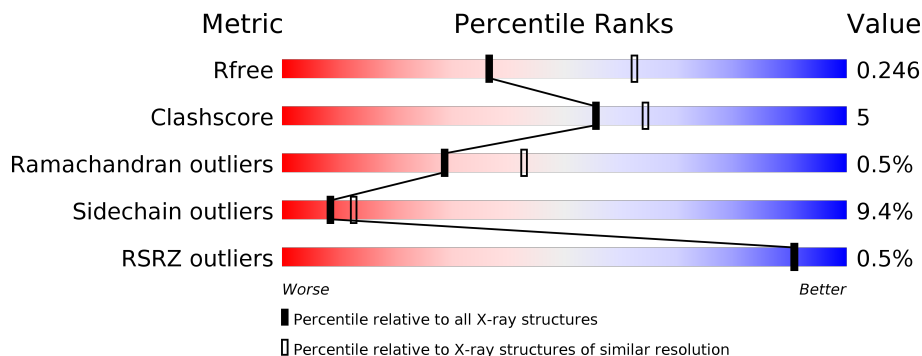
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.65 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	1332 (2.68-2.64)
Clashscore	141614	1374 (2.68-2.64)
Ramachandran outliers	138981	1349 (2.68-2.64)
Sidechain outliers	138945	1349 (2.68-2.64)
RSRZ outliers	127900	1318 (2.68-2.64)

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  $\geq 3$ , 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	98	 % 82% 8% • 8%
1	B	98	 % 83% 8% 9%
2	H	247	 72% 18% • 8%
2	I	247	 71% 18% • 9%

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 5162 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
			Total	C	N	O	S			
1	A	90	693	430	118	141	4	0	0	0
1	B	89	683	424	115	140	4	0	0	0

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	202	HIS	-	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
B	110	GLY	-	expression tag	UNP P60033
B	111	SER	-	expression tag	UNP P60033
B	202	HIS	-	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

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	H	227	1767	1120	293	346	8	0	0	0
2	I	226	1762	1117	292	345	8	0	0	0

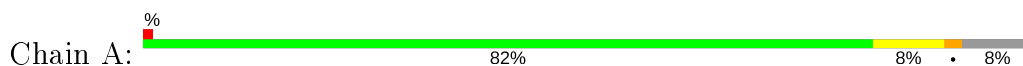
- Molecule 3 is water.

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
3	A	39	Total 39	O 39	0	0
3	B	29	Total 29	O 29	0	0
3	H	118	Total 118	O 118	0	0
3	I	71	Total 71	O 71	0	0

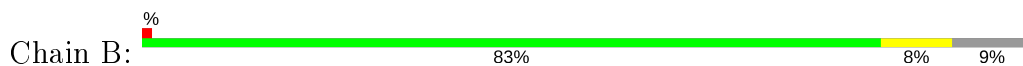
### 3 Residue-property plots [i](#)

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.

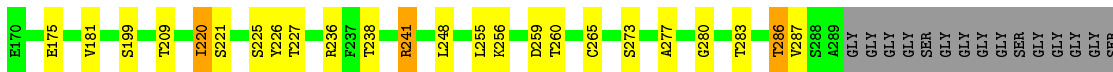
- Molecule 1: CD81 antigen



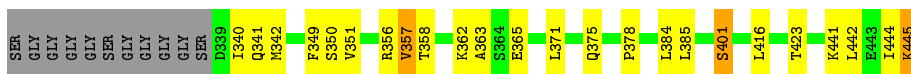
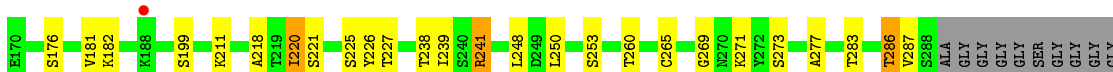
- Molecule 1: CD81 antigen



- Molecule 2: SINGLE CHAIN FV FRAGMENT



- Molecule 2: SINGLE CHAIN FV FRAGMENT



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	52.32Å 118.06Å 130.13Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.72 – 2.65 43.72 – 2.65	Depositor EDS
% Data completeness (in resolution range)	72.6 (43.72-2.65) 73.2 (43.72-2.65)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.18	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.72 (at 2.65Å)	Xtrriage
Refinement program	BUSTER 2.9.2	Depositor
R, $R_{free}$	0.179 , 0.247 0.185 , 0.246	Depositor DCC
$R_{free}$ test set	933 reflections (5.29%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	42.0	Xtrriage
Anisotropy	0.166	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 51.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	5162	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	46.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.09% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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	0.50	0/702	0.74	0/948
1	B	0.50	0/691	0.74	0/933
2	H	0.48	0/1810	0.77	1/2449 (0.0%)
2	I	0.46	0/1805	0.77	1/2442 (0.0%)
All	All	0.48	0/5008	0.76	2/6772 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	I	273	SER	N-CA-C	-5.67	95.68	111.00
2	H	273	SER	N-CA-C	-5.26	96.81	111.00

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	693	0	672	4	0
1	B	683	0	665	4	0
2	H	1767	0	1698	19	0
2	I	1762	0	1693	24	0
3	A	39	0	0	0	0
3	B	29	0	0	0	0
3	H	118	0	0	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	I	71	0	0	2	0
All	All	5162	0	4728	49	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 5.

The worst 5 of 49 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:H:236:ARG:NH1	2:H:259:ASP:OD2	2.19	0.73
2:I:220:ILE:HB	2:I:239:ILE:HG21	1.74	0.70
2:I:220:ILE:HB	2:I:239:ILE:CG2	2.28	0.63
2:H:181:VAL:HG11	2:H:255:LEU:HD13	1.80	0.63
2:I:378:PRO:HD3	3:I:508:HOH:O	2.01	0.60

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	88/98 (90%)	83 (94%)	5 (6%)	0	100	100
1	B	87/98 (89%)	79 (91%)	6 (7%)	2 (2%)	6	8
2	H	223/247 (90%)	212 (95%)	10 (4%)	1 (0%)	34	48
2	I	222/247 (90%)	209 (94%)	13 (6%)	0	100	100
All	All	620/690 (90%)	583 (94%)	34 (6%)	3 (0%)	29	43

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	138	ASP

Continued on next page...



*Continued from previous page...*

Mol	Chain	Res	Type
1	B	183	SER
2	H	345	SER

### 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	80/86 (93%)	76 (95%)	4 (5%)	24	38
1	B	79/86 (92%)	78 (99%)	1 (1%)	69	82
2	H	191/195 (98%)	169 (88%)	22 (12%)	5	7
2	I	191/195 (98%)	167 (87%)	24 (13%)	4	6
All	All	541/562 (96%)	490 (91%)	51 (9%)	8	13

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	H	371	LEU
2	I	182	LYS
2	I	365	GLU
2	H	414	THR
2	I	199	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
2	H	270	ASN
2	I	341	GLN
2	I	243	ASN
2	H	243	ASN
2	H	341	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 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 [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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	90/98 (91%)	-0.35	1 (1%) 80 79	21, 36, 84, 96	0
1	B	89/98 (90%)	-0.41	1 (1%) 80 79	24, 38, 80, 100	0
2	H	227/247 (91%)	-0.52	0 100 100	22, 38, 57, 85	0
2	I	226/247 (91%)	-0.29	1 (0%) 92 93	29, 55, 75, 96	0
All	All	632/690 (91%)	-0.40	3 (0%) 91 91	21, 43, 75, 100	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	139	ASP	3.2
2	I	188	LYS	2.2
1	A	140	ALA	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 carbohydrates 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.