wwPDB X-ray Structure Validation Summary Report (i)

May 13, 2020 – 01:31 am BST

PDB ID	:	5X0T
Title	:	Crystal structure of CD147 C2 domain in complex with Fab of its monoclonal antibody 6H8
Authors	:	Lin, P.; Zhang, M.-Y.; Ye, S.; Chen, X.; Yu, X.-L.; Zhang, R.-G.; Zhu, P.; Chen, Z.-N.
Deposited on	:	2017-01-23
Resolution	:	2.50 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:

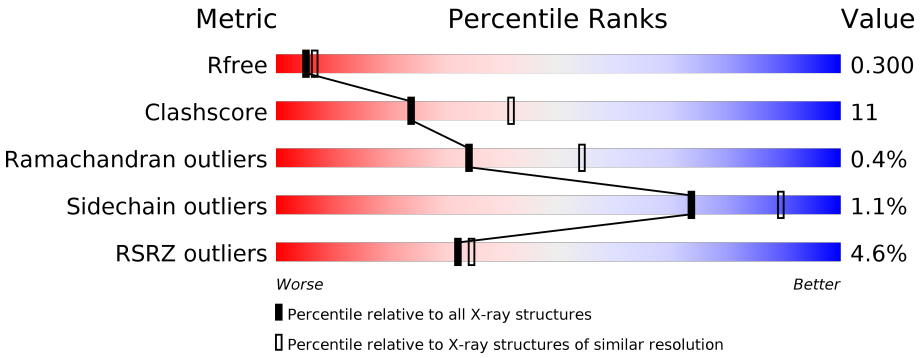
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	241	3% 72% 16% • 11%
1	C	241	6% 74% 15% • 11%
2	B	234	% 70% 21% 9%
2	D	234	3% 73% 18% 9%
3	E	89	8% 61% 27% 12%
3	F	89	9% 57% 30% 12%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 7981 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 6H8 Fab fragment heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	215	Total	C	N	O	S	0	0	0
			1645	1042	271	324	8			
1	C	215	Total	C	N	O	S	0	0	0
			1645	1042	271	324	8			

- Molecule 2 is a protein called 6H8 Fab fragment light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	213	Total	C	N	O	S	0	0	0
			1652	1027	276	341	8			
2	D	213	Total	C	N	O	S	0	0	0
			1652	1027	276	341	8			

- Molecule 3 is a protein called Basigin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	78	Total	C	N	O	S	0	0	0
			599	380	95	121	3			
3	F	78	Total	C	N	O	S	0	0	0
			599	380	95	121	3			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	21	MET	-	expression tag	UNP P35613
E	102	LEU	-	expression tag	UNP P35613
E	103	GLU	-	expression tag	UNP P35613
E	104	HIS	-	expression tag	UNP P35613
E	105	HIS	-	expression tag	UNP P35613
E	106	HIS	-	expression tag	UNP P35613
E	107	HIS	-	expression tag	UNP P35613
E	108	HIS	-	expression tag	UNP P35613

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	109	HIS	-	expression tag	UNP P35613
F	21	MET	-	expression tag	UNP P35613
F	102	LEU	-	expression tag	UNP P35613
F	103	GLU	-	expression tag	UNP P35613
F	104	HIS	-	expression tag	UNP P35613
F	105	HIS	-	expression tag	UNP P35613
F	106	HIS	-	expression tag	UNP P35613
F	107	HIS	-	expression tag	UNP P35613
F	108	HIS	-	expression tag	UNP P35613
F	109	HIS	-	expression tag	UNP P35613

- Molecule 4 is water.

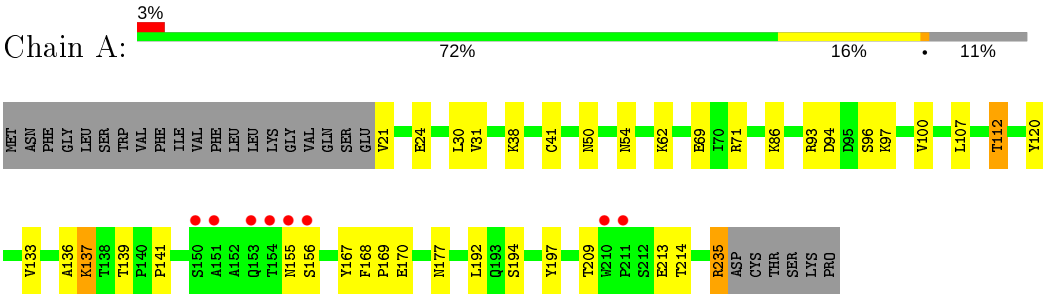
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	39	Total O 39 39	0	0
4	C	34	Total O 34 34	0	0
4	B	47	Total O 47 47	0	0
4	D	42	Total O 42 42	0	0
4	E	13	Total O 13 13	0	0
4	F	14	Total O 14 14	0	0

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.

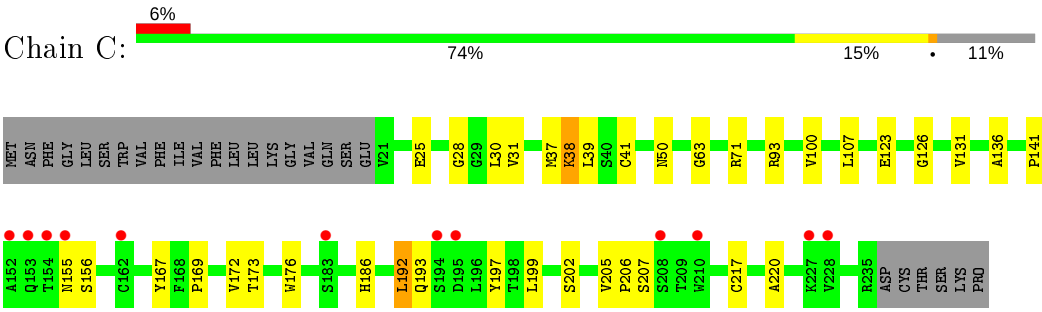
- Molecule 1: 6H8 Fab fragment heavy chain

Chain A:



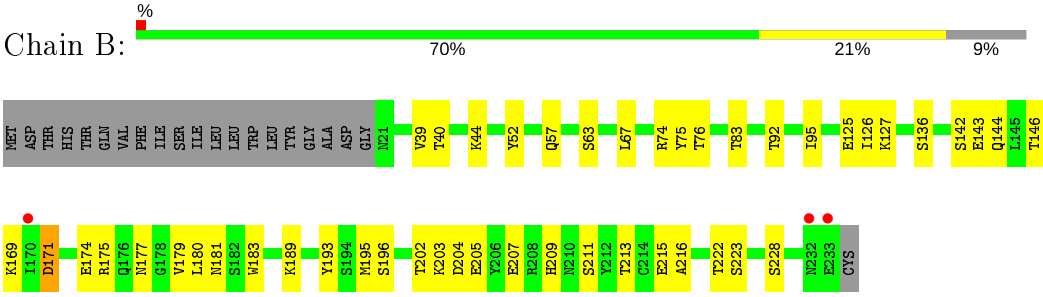
- Molecule 1: 6H8 Fab fragment heavy chain

Chain C:



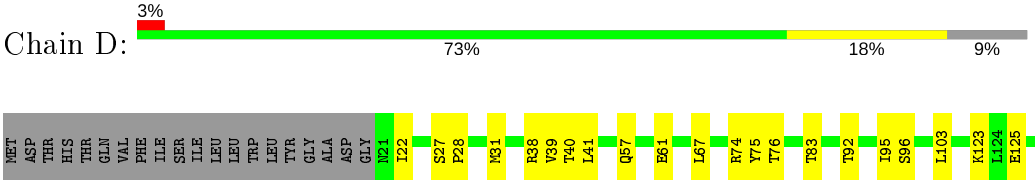
- Molecule 2: 6H8 Fab fragment light chain

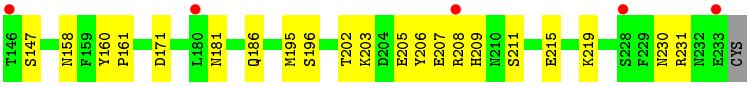
Chain B:



- Molecule 2: 6H8 Fab fragment light chain

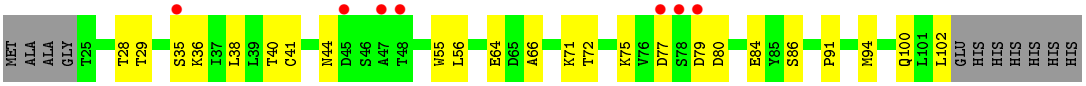
Chain D:





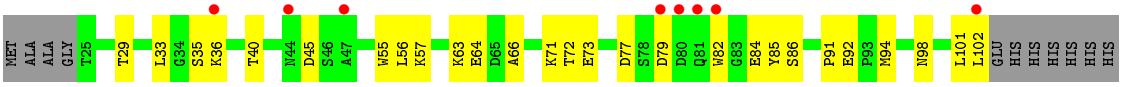
- Molecule 3: Basigin

Chain E:



- Molecule 3: Basigin

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	46.99Å 83.49Å 130.19Å 90.00° 92.92° 90.00°	Depositor
Resolution (Å)	37.16 – 2.50 37.16 – 2.50	Depositor EDS
% Data completeness (in resolution range)	96.8 (37.16-2.50) 83.6 (37.16-2.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	6.26 (at 2.51Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.246 , 0.300 0.247 , 0.300	Depositor DCC
R_{free} test set	1688 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	24.4	Xtriage
Anisotropy	0.486	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 14.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	0.148 for h,-k,-l	Xtriage
F_o,F_c correlation	0.86	EDS
Total number of atoms	7981	wwPDB-VP
Average B, all atoms (Å^2)	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.51% 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	0.30	0/1689	0.54	0/2307
1	C	0.35	0/1689	0.55	0/2307
2	B	0.33	0/1689	0.57	0/2292
2	D	0.39	0/1689	0.57	0/2292
3	E	0.39	0/610	0.64	0/829
3	F	0.31	0/610	0.66	0/829
All	All	0.34	0/7976	0.57	0/10856

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
3	E	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	E	79	ASP	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	1645	0	1595	32	0
1	C	1645	0	1595	26	0
2	B	1652	0	1573	40	0
2	D	1652	0	1573	28	0
3	E	599	0	586	23	0
3	F	599	0	586	25	0
4	A	39	0	0	11	1
4	B	47	0	0	6	1
4	C	34	0	0	2	0
4	D	42	0	0	8	0
4	E	13	0	0	5	0
4	F	14	0	0	6	0
All	All	7981	0	7508	169	1

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 11.

The worst 5 of 169 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:C:193:GLN:NE2	4:C:301:HOH:O	1.88	1.04
2:B:216:ALA:HA	4:B:313:HOH:O	1.58	1.02
3:F:45:ASP:OD2	4:F:301:HOH:O	1.85	0.94
1:A:21:VAL:N	4:A:305:HOH:O	2.04	0.91
2:D:61:GLU:N	4:D:301:HOH:O	1.91	0.90

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:A:319:HOH:O	4:B:306:HOH:O[2_759]	1.87	0.33

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	213/241 (88%)	207 (97%)	6 (3%)	0	100	100
1	C	213/241 (88%)	207 (97%)	6 (3%)	0	100	100
2	B	211/234 (90%)	202 (96%)	8 (4%)	1 (0%)	29	48
2	D	211/234 (90%)	201 (95%)	9 (4%)	1 (0%)	29	48
3	E	76/89 (85%)	63 (83%)	12 (16%)	1 (1%)	12	21
3	F	76/89 (85%)	65 (86%)	10 (13%)	1 (1%)	12	21
All	All	1000/1128 (89%)	945 (94%)	51 (5%)	4 (0%)	34	54

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	171	ASP
2	D	171	ASP
3	E	77	ASP
3	F	77	ASP

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	187/211 (89%)	183 (98%)	4 (2%)	53	78
1	C	187/211 (89%)	185 (99%)	2 (1%)	73	89
2	B	190/208 (91%)	189 (100%)	1 (0%)	88	96
2	D	190/208 (91%)	187 (98%)	3 (2%)	62	84
3	E	68/76 (90%)	68 (100%)	0	100	100
3	F	68/76 (90%)	68 (100%)	0	100	100
All	All	890/990 (90%)	880 (99%)	10 (1%)	73	89

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	38	LYS
1	C	192	LEU
2	D	41	LEU
1	A	235	ARG
2	B	44	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	D	99	GLN

5.3.3 RNA

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

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	215/241 (89%)	0.35	8 (3%) 41 45	8, 23, 56, 97	0
1	C	215/241 (89%)	0.42	15 (6%) 16 16	9, 25, 59, 91	0
2	B	213/234 (91%)	0.32	3 (1%) 75 77	10, 21, 50, 82	1 (0%)
2	D	213/234 (91%)	0.36	6 (2%) 53 56	12, 27, 47, 79	1 (0%)
3	E	78/89 (87%)	0.65	7 (8%) 9 9	13, 32, 57, 66	0
3	F	78/89 (87%)	0.66	8 (10%) 6 6	13, 30, 53, 69	0
All	All	1012/1128 (89%)	0.41	47 (4%) 32 34	8, 25, 55, 97	2 (0%)

The worst 5 of 47 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	150	SER	9.2
1	A	154	THR	7.7
1	A	151	ALA	6.4
1	C	150	SER	6.2
2	D	233	GLU	5.3

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.