wwPDB X-ray Structure Validation Summary Report (i)

Nov 13, 2023 – 01:28 PM JST

PDB ID	:	5XHF
Title	:	Crystal structure of Trastuzumab Fab fragment bearing p-azido-L-phenylalanine
Authors	:	Kuratani, M.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S.
Deposited on	:	2017-04-20
Resolution	:	3.21 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 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:

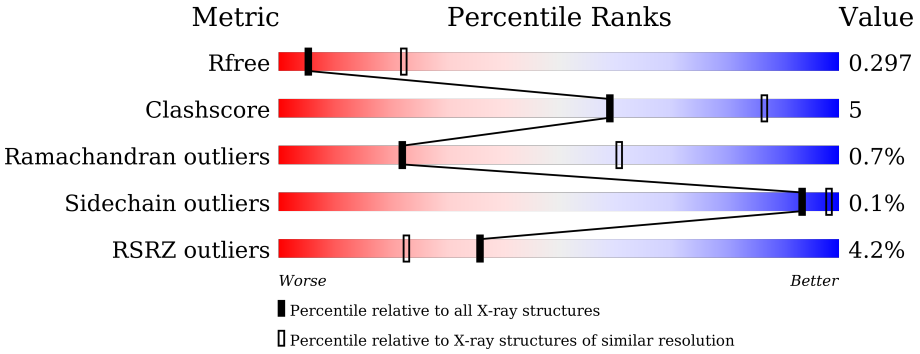
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 3.21 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	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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 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	214	4% 86% 14%
1	C	214	8% 89% 10%
2	B	220	% 85% 12% •
2	D	220	4% 85% 15%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 6565 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 polypeptide (L chain).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	213	1648	1033	278	332	5	0	0	0
1	C	214	1654	1036	279	333	6	0	0	0

- Molecule 2 is a protein called polypeptide (H chain).

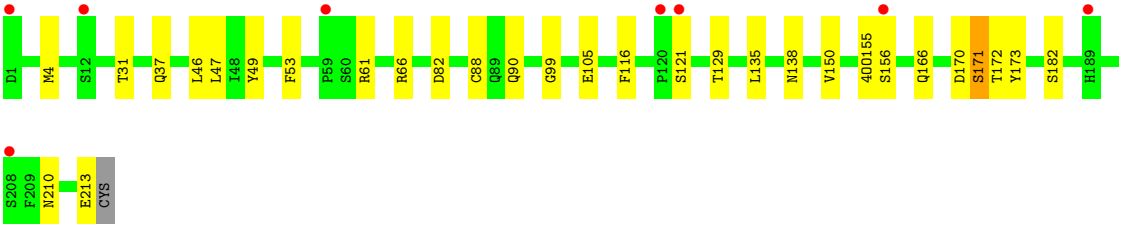
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	214	1611	1021	271	313	6	0	0	0
2	D	220	1652	1044	278	323	7	0	1	0

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.

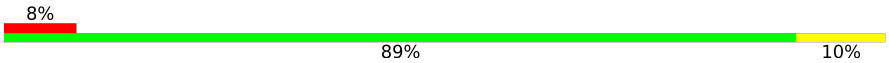
• Molecule 1: polypeptide (L chain)

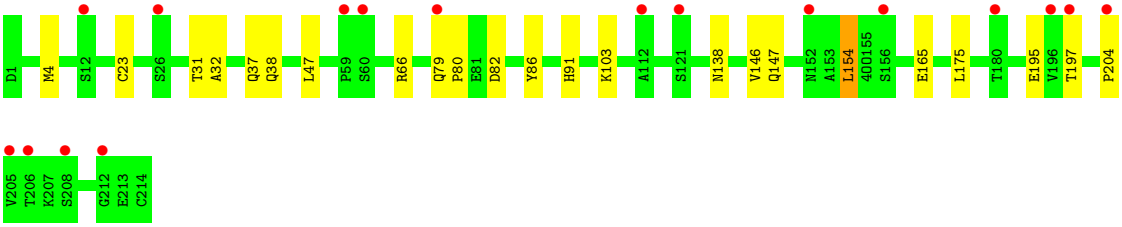
Chain A:



• Molecule 1: polypeptide (L chain)

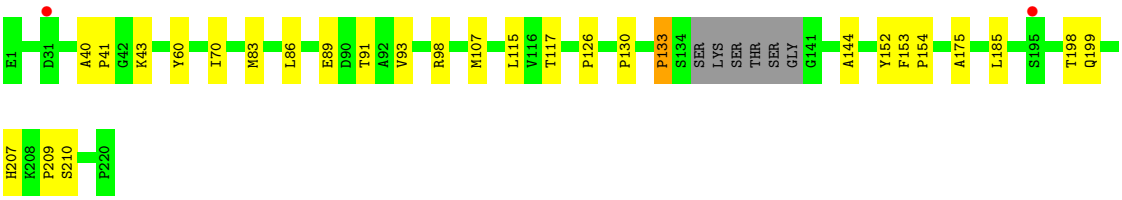
Chain C:





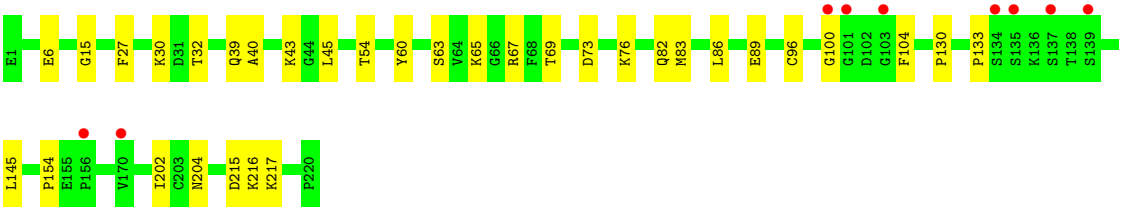
• Molecule 2: polypeptide (H chain)

Chain B:



• Molecule 2: polypeptide (H chain)

Chain D:


E1
E6
G15
F27
K30
D31
T32
Q39
A40
K43
G44
L45
T54
Y60
S63
V64
K65
G66
R67
F68
T69
D73
K76
Q82
M83
L86
E89
C96
G100
G101
D102
G103
F104
P130
P133
S134
S135
K136
S137
T138
S139
L145
P154
E155
P156
V170
I202
C203
N204
D215
K216
K217
P220

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	38.84Å 80.07Å 85.79Å 113.67° 92.35° 102.78°	Depositor
Resolution (Å)	42.44 – 3.21 42.44 – 3.21	Depositor EDS
% Data completeness (in resolution range)	98.4 (42.44-3.21) 98.4 (42.44-3.21)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.60 (at 3.19Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155)	Depositor
R, R_{free}	0.219 , 0.297 0.220 , 0.297	Depositor DCC
R_{free} test set	1473 reflections (9.98%)	wwPDB-VP
Wilson B-factor (Å^2)	44.7	Xtriage
Anisotropy	0.421	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 44.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.26$	Xtriage
Estimated twinning fraction	0.064 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	6565	wwPDB-VP
Average B, all atoms (Å^2)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.17% 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: 4OO

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	0.24	0/1670	0.43	0/2269
1	C	0.24	0/1676	0.43	0/2277
2	B	0.24	0/1652	0.44	0/2253
2	D	0.24	0/1697	0.45	0/2314
All	All	0.24	0/6695	0.44	0/9113

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	1648	0	1595	16	0
1	C	1654	0	1600	12	0
2	B	1611	0	1567	16	0
2	D	1652	0	1610	19	0
All	All	6565	0	6372	61	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 61 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:D:83:MET:HB3	2:D:86:LEU:HD21	1.71	0.70
2:D:100:GLY:HA3	2:D:104:PHE:HB3	1.73	0.68
1:C:37:GLN:HB2	1:C:47:LEU:HD11	1.78	0.66
1:A:170:ASP:O	1:A:172:THR:N	2.32	0.63
2:D:202:ILE:HG12	2:D:217:LYS:HA	1.83	0.61

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	210/214 (98%)	194 (92%)	14 (7%)	2 (1%)	15	54
1	C	211/214 (99%)	196 (93%)	14 (7%)	1 (0%)	29	67
2	B	210/220 (96%)	196 (93%)	12 (6%)	2 (1%)	15	54
2	D	219/220 (100%)	207 (94%)	11 (5%)	1 (0%)	29	67
All	All	850/868 (98%)	793 (93%)	51 (6%)	6 (1%)	22	61

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	171	SER
1	C	138	ASN
1	A	138	ASN
2	B	133	PRO
2	B	41	PRO

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	187/188 (100%)	187 (100%)	0	100	100
1	C	188/188 (100%)	187 (100%)	1 (0%)	88	95
2	B	176/181 (97%)	176 (100%)	0	100	100
2	D	182/181 (101%)	182 (100%)	0	100	100
All	All	733/738 (99%)	732 (100%)	1 (0%)	93	98

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	154	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

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, 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	212/214 (99%)	0.38	8 (3%)	40	26	16, 37, 57, 77	0
1	C	213/214 (99%)	0.53	17 (7%)	12	6	17, 38, 64, 87	0
2	B	214/220 (97%)	0.26	2 (0%)	84	75	13, 28, 59, 76	0
2	D	220/220 (100%)	0.35	9 (4%)	37	24	13, 30, 68, 93	0
All	All	859/868 (98%)	0.38	36 (4%)	36	23	13, 33, 62, 93	0

The worst 5 of 36 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	112	ALA	4.7
2	D	135	SER	4.2
2	D	137	SER	4.2
2	D	139	SER	4.1
2	D	101	GLY	4.0

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.