



wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 9, 2016 – 11:41 AM EST

PDB ID : 5F72
Title : De novo design and crystallographic validation of antibodies targeting a pre-selected epitope
Authors : Liu, X.; Taylor, R.D.; Griffin, L.; Coker, S.; Adams, R.; Ceska, T.; Shi, J.; Lawson, A.D.G.; Baker, T.
Deposited on : 2015-12-07
Resolution : 1.85 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

<http://wwpdb.org/validation/2016/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
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028442
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028442

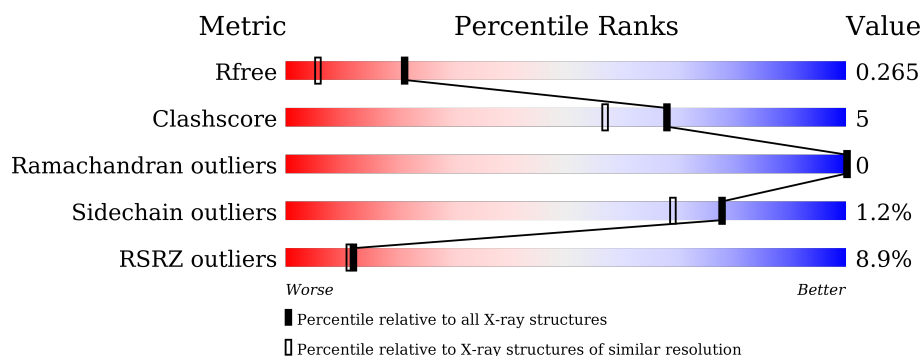
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.85 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}	91344	1745 (1.86-1.86)
Clashscore	102246	1898 (1.86-1.86)
Ramachandran outliers	100387	1875 (1.86-1.86)
Sidechain outliers	100360	1875 (1.86-1.86)
RSRZ outliers	91569	1747 (1.86-1.86)

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. 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	C	296	<div> <div>16%</div> <div> <div></div> <div>79%</div> <div>15%</div> <div>• •</div> </div> </div>
1	K	296	<div> <div>6%</div> <div> <div></div> <div>81%</div> <div>14%</div> <div>• •</div> </div> </div>
2	S	257	<div> <div>6%</div> <div> <div></div> <div>77%</div> <div>11%</div> <div>• 12%</div> </div> </div>
2	T	257	<div> <div>4%</div> <div> <div></div> <div>84%</div> <div>• • 11%</div> </div> </div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 8333 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 Kelch-like ECH-associated protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	K	285	Total	C	N	O	S	0	0	0
			2190	1361	394	420	15			
1	C	285	Total	C	N	O	S	0	0	0
			2190	1361	394	420	15			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
K	316	GLY	-	expression tag	UNP Q14145
K	317	SER	-	expression tag	UNP Q14145
K	318	MET	-	expression tag	UNP Q14145
K	319	GLY	-	expression tag	UNP Q14145
K	320	HIS	-	expression tag	UNP Q14145
K	354	ASP	ARG	conflict	UNP Q14145
C	316	GLY	-	expression tag	UNP Q14145
C	317	SER	-	expression tag	UNP Q14145
C	318	MET	-	expression tag	UNP Q14145
C	319	GLY	-	expression tag	UNP Q14145
C	320	HIS	-	expression tag	UNP Q14145
C	354	ASP	ARG	conflict	UNP Q14145

- Molecule 2 is a protein called Single chain Fv from a Fab.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	T	229	Total	C	N	O	S	0	0	0
			1707	1069	288	342	8			
2	S	226	Total	C	N	O	S	0	0	0
			1682	1054	281	339	8			

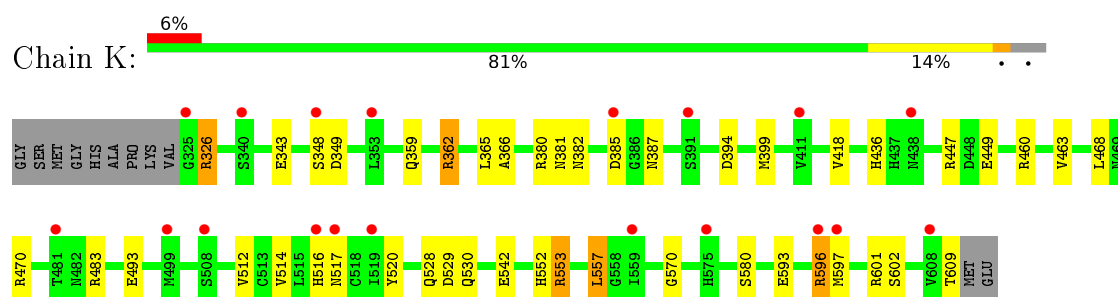
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	K	181	Total 181	O 181	0	0
3	C	155	Total 155	O 155	0	0
3	T	136	Total 136	O 136	0	0
3	S	92	Total 92	O 92	0	0

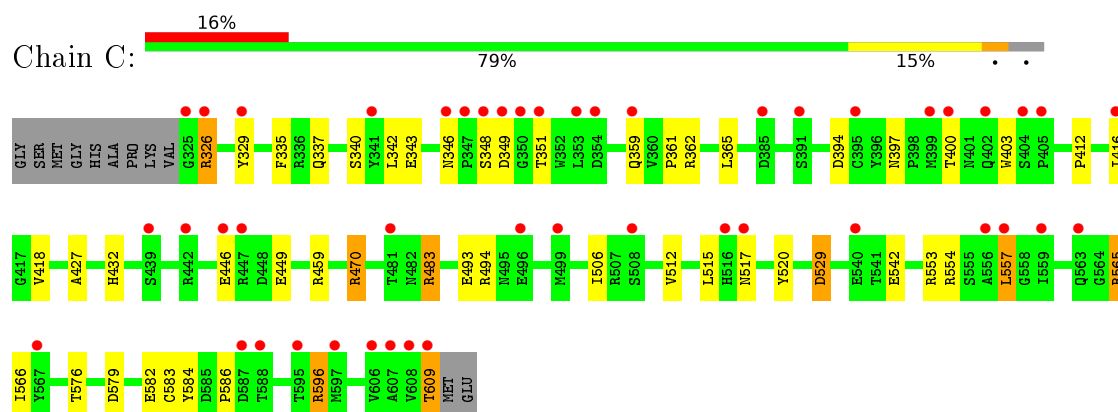
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

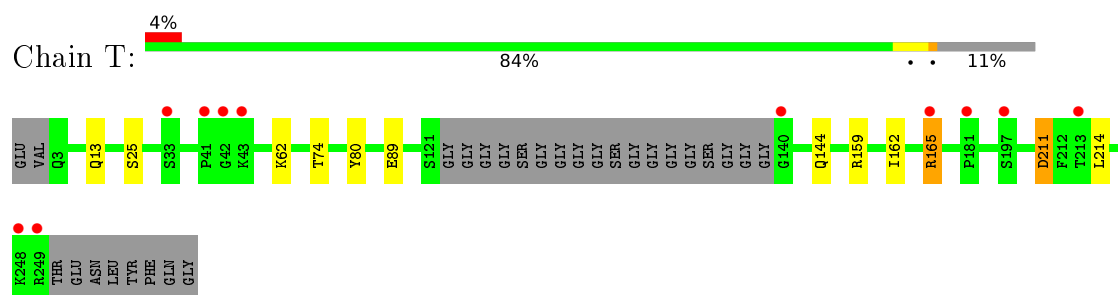
- Molecule 1: Kelch-like ECH-associated protein 1



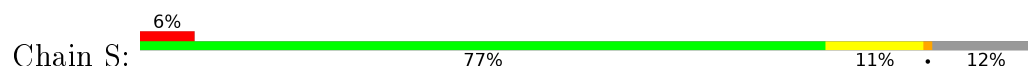
- Molecule 1: Kelch-like ECH-associated protein 1

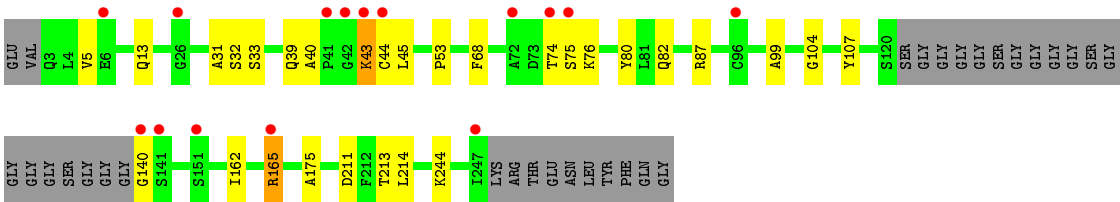


- Molecule 2: Single chain Fv from a Fab



- Molecule 2: Single chain Fv from a Fab





4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	70.50Å 69.81Å 99.56Å 90.00° 92.23° 90.00°	Depositor
Resolution (Å)	29.69 – 1.85 29.69 – 1.85	Depositor EDS
% Data completeness (in resolution range)	99.2 (29.69-1.85) 99.2 (29.69-1.85)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.05 (at 1.85Å)	Xtriage
Refinement program	REFMAC 5.8.0103	Depositor
R, R_{free}	0.223 , 0.265 0.223 , 0.265	Depositor DCC
R_{free} test set	4108 reflections (5.28%)	DCC
Wilson B-factor (Å ²)	21.8	Xtriage
Anisotropy	0.037	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 45.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.011 for k,h,-l 0.013 for -k,-h,-l 0.025 for h,-k,-l	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	8333	wwPDB-VP
Average B, all atoms (Å ²)	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.72% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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

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	C	1.30	15/2243 (0.7%)	1.16	21/3055 (0.7%)
1	K	1.28	14/2243 (0.6%)	1.14	17/3055 (0.6%)
2	S	1.10	6/1718 (0.3%)	1.01	5/2335 (0.2%)
2	T	1.14	6/1743 (0.3%)	0.95	2/2366 (0.1%)
All	All	1.22	41/7947 (0.5%)	1.08	45/10811 (0.4%)

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
1	K	0	1

The worst 5 of 41 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	326	ARG	NE-CZ	20.09	1.59	1.33
1	K	326	ARG	CZ-NH2	16.48	1.54	1.33
1	K	326	ARG	CZ-NH1	-13.90	1.15	1.33
1	C	326	ARG	CD-NE	12.94	1.68	1.46
1	C	517	ASN	CG-OD1	12.23	1.50	1.24

The worst 5 of 45 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	553	ARG	NE-CZ-NH2	9.75	125.17	120.30
1	K	385	ASP	CB-CG-OD1	-9.54	109.72	118.30
1	C	596	ARG	NE-CZ-NH2	-9.00	115.80	120.30
1	K	362	ARG	NE-CZ-NH1	-8.96	115.82	120.30
1	K	385	ASP	CB-CG-OD2	8.92	126.33	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	K	529	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	C	2190	0	2075	28	0
1	K	2190	0	2075	22	0
2	S	1682	0	1617	17	0
2	T	1707	0	1648	6	0
3	C	155	0	0	3	0
3	K	181	0	0	2	0
3	S	92	0	0	2	0
3	T	136	0	0	3	0
All	All	8333	0	7415	71	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 71 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:S:43:LYS:CD	2:S:43:LYS:CE	1.76	1.55
1:C:326:ARG:NE	1:C:326:ARG:CD	1.68	1.51
1:K:359:GLN:HG3	3:K:863:HOH:O	1.62	0.99
1:C:609:THR:HG23	1:C:609:THR:O	1.67	0.93
1:C:609:THR:O	1:C:609:THR:CG2	2.29	0.78

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	C	283/296 (96%)	273 (96%)	10 (4%)	0	100	100
1	K	283/296 (96%)	276 (98%)	7 (2%)	0	100	100
2	S	222/257 (86%)	214 (96%)	8 (4%)	0	100	100
2	T	225/257 (88%)	220 (98%)	5 (2%)	0	100	100
All	All	1013/1106 (92%)	983 (97%)	30 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	C	231/239 (97%)	228 (99%)	3 (1%)	76	65
1	K	231/239 (97%)	227 (98%)	4 (2%)	68	54
2	S	182/197 (92%)	180 (99%)	2 (1%)	80	72
2	T	185/197 (94%)	184 (100%)	1 (0%)	92	90
All	All	829/872 (95%)	819 (99%)	10 (1%)	78	69

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

Mol	Chain	Res	Type
1	C	483	ARG
1	C	529	ASP
2	T	25	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	K	580	SER
1	C	557	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	K	387	ASN
2	S	35	HIS

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, 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	C	285/296 (96%)	0.93	46 (16%) 3 3	9, 25, 47, 67	23 (8%)
1	K	285/296 (96%)	0.40	19 (6%) 21 19	6, 21, 39, 67	19 (6%)
2	S	226/257 (87%)	0.30	15 (6%) 22 20	13, 26, 43, 64	2 (0%)
2	T	229/257 (89%)	0.03	11 (4%) 34 32	10, 19, 36, 61	3 (1%)
All	All	1025/1106 (92%)	0.44	91 (8%) 12 11	6, 23, 43, 67	47 (4%)

The worst 5 of 91 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	416	ILE	13.1
2	S	96	CYS	7.7
2	S	41	PRO	6.6
1	C	481	THR	6.3
2	T	213	THR	5.9

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.