Full wwPDB X-ray Structure Validation Report (i)

Jun 24, 2024 – 02:16 PM EDT

PDB ID	:	5WKZ
Title	:	VH1-69 germline antibody predicted from CR6261
Authors	:	Lang, S.; Lee, P.S.
Deposited on	:	2017-07-25
Resolution	:	1.85 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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:

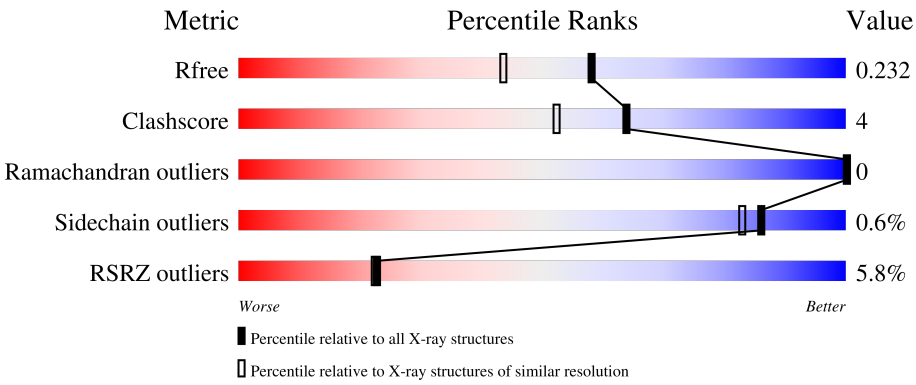
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 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}	130704	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (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 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	H	250	6% 83% • • 12%
2	L	217	4% 88% 10% •

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3559 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 Immunoglobulin heavy variable 1-69D,IgG H chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	H	221	1638	1029	276	325	8	0	0	0

There are 41 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
H	99	HIS	-	linker	UNP A0A0B4J2H0
H	100	MET	-	linker	UNP A0A0B4J2H0
H	101	GLY	-	linker	UNP A0A0B4J2H0
H	102	TYR	-	linker	UNP A0A0B4J2H0
H	103	GLN	-	linker	UNP A0A0B4J2H0
H	104	LEU	-	linker	UNP A0A0B4J2H0
H	105	ARG	-	linker	UNP A0A0B4J2H0
H	106	GLU	-	linker	UNP A0A0B4J2H0
H	107	THR	-	linker	UNP A0A0B4J2H0
H	108	MET	-	linker	UNP A0A0B4J2H0
H	109	ASP	-	linker	UNP A0A0B4J2H0
H	110	VAL	-	linker	UNP A0A0B4J2H0
H	111	TRP	-	linker	UNP A0A0B4J2H0
H	112	GLY	-	linker	UNP A0A0B4J2H0
H	113	LYS	-	linker	UNP A0A0B4J2H0
H	225	GLY	-	expression tag	UNP S6BAM6
H	226	SER	-	expression tag	UNP S6BAM6
H	227	GLY	-	expression tag	UNP S6BAM6
H	228	GLY	-	expression tag	UNP S6BAM6
H	229	GLY	-	expression tag	UNP S6BAM6
H	230	LEU	-	expression tag	UNP S6BAM6
H	231	ASN	-	expression tag	UNP S6BAM6
H	232	ASP	-	expression tag	UNP S6BAM6
H	233	ILE	-	expression tag	UNP S6BAM6
H	234	PHE	-	expression tag	UNP S6BAM6
H	235	GLU	-	expression tag	UNP S6BAM6
H	236	ALA	-	expression tag	UNP S6BAM6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
H	237	GLN	-	expression tag	UNP S6BAM6
H	238	LYS	-	expression tag	UNP S6BAM6
H	239	ILE	-	expression tag	UNP S6BAM6
H	240	GLU	-	expression tag	UNP S6BAM6
H	241	TRP	-	expression tag	UNP S6BAM6
H	242	HIS	-	expression tag	UNP S6BAM6
H	243	GLU	-	expression tag	UNP S6BAM6
H	244	CYS	-	expression tag	UNP S6BAM6
H	245	HIS	-	expression tag	UNP S6BAM6
H	246	HIS	-	expression tag	UNP S6BAM6
H	247	HIS	-	expression tag	UNP S6BAM6
H	248	HIS	-	expression tag	UNP S6BAM6
H	249	HIS	-	expression tag	UNP S6BAM6
H	250	HIS	-	expression tag	UNP S6BAM6

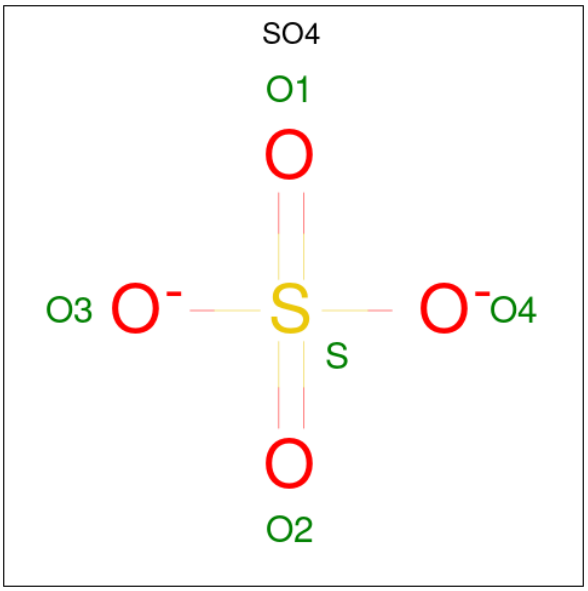
- Molecule 2 is a protein called Lambda-chain (AA -20 to 215).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	L	212	1573	989	260	320	4	0	1	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
L	97	TYR	-	insertion	UNP A2NUT2
L	98	VAL	GLY	conflict	UNP A2NUT2

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
3	H	1	5	4	1	0	0

- Molecule 4 is water.

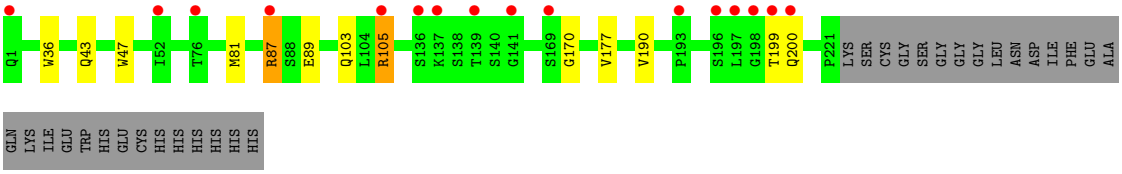
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	H	158	158	158	0	0
			Total	O		
4	L	185	185	185	0	0

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.

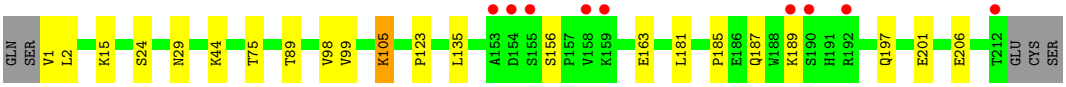
• Molecule 1: Immunoglobulin heavy variable 1-69D,IgG H chain

Chain H:



• Molecule 2: Lambda-chain (AA -20 to 215)

Chain L:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 42 21 2	Depositor
Cell constants a, b, c, α, β, γ	84.23Å 84.23Å 177.54Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.42 – 1.85 48.42 – 1.85	Depositor EDS
% Data completeness (in resolution range)	99.6 (48.42-1.85) 99.6 (48.42-1.85)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.26 (at 1.84Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, R_{free}	0.205 , 0.232 0.204 , 0.232	Depositor DCC
R_{free} test set	2811 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	30.7	Xtriage
Anisotropy	0.236	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 45.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3559	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 3.72% 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: SO4

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	H	0.40	0/1675	0.61	1/2280 (0.0%)
2	L	0.39	0/1615	0.59	0/2209
All	All	0.40	0/3290	0.60	1/4489 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	H	87	ARG	NE-CZ-NH1	-7.82	116.39	120.30

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	H	1638	0	1617	10	0
2	L	1573	0	1541	20	0
3	H	5	0	0	0	0
4	H	158	0	0	1	2
4	L	185	0	0	7	2
All	All	3559	0	3158	27	2

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 4.

All (27) 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:H:103:GLN:HB3	1:H:105:ARG:HD2	1.68	0.75
2:L:156:SER:OG	4:L:301:HOH:O	2.08	0.70
2:L:1:VAL:HG13	2:L:99:VAL:HG21	1.79	0.65
2:L:15:LYS:HG2	2:L:75:THR:HG22	1.78	0.65
1:H:103:GLN:HB3	1:H:105:ARG:CD	2.34	0.57
2:L:185:PRO:O	2:L:189:LYS:HG3	2.05	0.56
2:L:189:LYS:O	4:L:303:HOH:O	2.18	0.55
2:L:135:LEU:HD12	2:L:181:LEU:HD23	1.87	0.55
2:L:187:GLN:OE1	4:L:302:HOH:O	2.17	0.54
1:H:199:THR:HG23	1:H:200:GLN:OE1	2.09	0.53
2:L:197:GLN:HG2	2:L:206:GLU:HG3	1.92	0.52
2:L:105:LYS:HD3	2:L:105:LYS:C	2.31	0.51
1:H:43:GLN:NE2	4:H:405:HOH:O	2.36	0.49
2:L:44:LYS:NZ	4:L:309:HOH:O	2.45	0.49
1:H:177:VAL:HG21	2:L:163:GLU:HB3	1.96	0.48
2:L:105:LYS:NZ	4:L:311:HOH:O	2.47	0.47
1:H:177:VAL:CG2	2:L:163:GLU:HB3	2.46	0.46
1:H:87:ARG:HB3	1:H:89:GLU:OE1	2.15	0.46
2:L:156:SER:N	4:L:301:HOH:O	2.49	0.45
2:L:2:LEU:HD11	2:L:89:THR:HG22	1.99	0.45
1:H:170:GLY:O	1:H:190:VAL:HA	2.18	0.44
2:L:1:VAL:CG1	2:L:99:VAL:HG21	2.47	0.43
2:L:201:GLU:OE2	4:L:304:HOH:O	2.22	0.43
2:L:123:PRO:HD3	2:L:135:LEU:CD2	2.49	0.42
1:H:47:TRP:CG	2:L:98:VAL:HB	2.55	0.42
1:H:36:TRP:CE2	1:H:81:MET:HB2	2.54	0.42
2:L:24:SER:O	2:L:29:ASN:HB2	2.20	0.42

All (2) 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:H:431:HOH:O	4:L:385:HOH:O[2_675]	2.10	0.10
4:H:554:HOH:O	4:L:432:HOH:O[5_557]	2.13	0.07

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	H	219/250 (88%)	219 (100%)	0	0	100	100
2	L	211/217 (97%)	207 (98%)	4 (2%)	0	100	100
All	All	430/467 (92%)	426 (99%)	4 (1%)	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	H	184/208 (88%)	183 (100%)	1 (0%)	88	86
2	L	178/182 (98%)	177 (99%)	1 (1%)	86	83
All	All	362/390 (93%)	360 (99%)	2 (1%)	86	83

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	H	105	ARG
2	L	105	LYS

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)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	SO4	H	301	-	4,4,4	0.12	0	6,6,6	0.17	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	H	221/250 (88%)	0.27	16 (7%)	15	15	20, 34, 57, 69	0
2	L	212/217 (97%)	0.09	9 (4%)	36	34	21, 32, 51, 64	0
All	All	433/467 (92%)	0.19	25 (5%)	23	22	20, 33, 55, 69	0

All (25) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	199	THR	3.6
2	L	212	THR	3.5
1	H	1	GLN	3.5
1	H	141	GLY	3.4
2	L	158	VAL	3.0
2	L	155	SER	3.0
2	L	159	LYS	2.9
1	H	197	LEU	2.8
2	L	190	SER	2.8
2	L	153	ALA	2.8
1	H	198	GLY	2.8
2	L	192	ARG	2.6
1	H	137	LYS	2.6
1	H	76	THR	2.5
2	L	154	ASP	2.5
1	H	136	SER	2.4
2	L	189	LYS	2.3
1	H	196	SER	2.3
1	H	193	PRO	2.3
1	H	169	SER	2.3
1	H	139	THR	2.2
1	H	52	ILE	2.2
1	H	87	ARG	2.2
1	H	200	GLN	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	H	105	ARG	2.0

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)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	SO4	H	301	5/5	0.95	0.15	48,54,61,63	0

6.5 Other polymers (i)

There are no such residues in this entry.