W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report ⓘ

Apr 29, 2025 – 11:08 AM EDT

PDB ID	:	4K5U / pdb_00004k5u
Title	:	Recognition of the BG-H Antigen by a Lamprey Variable Lymphocyte Receptor
Authors	:	Luo, M.; Velikovsky, C.A.; Yang, X.B.; Mariuzza, R.A.
Deposited on	:	2013-04-15
Resolution	:	1.70 Å(reported)

The following versions of software and data (see references ⓘ) were used in the production of this report:

|		
MolProbity	:	4-5-2 with Phenix2.0rc1
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0rc1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.006 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.43.1

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 1.70 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.
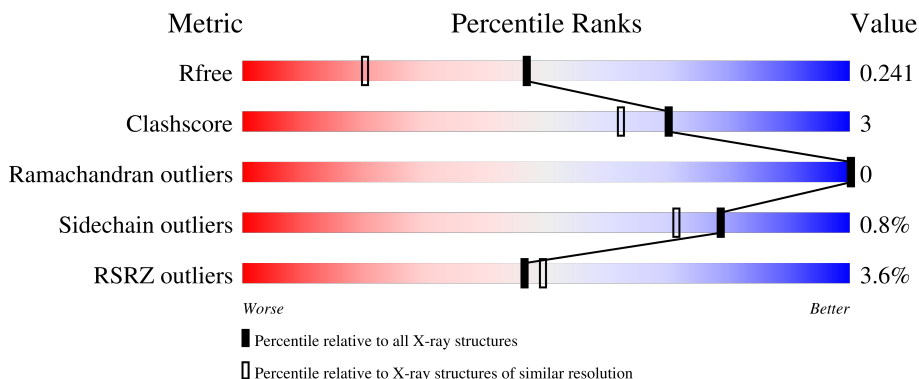
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	5161 (1.70-1.70)
Clashscore	180529	5671 (1.70-1.70)
Ramachandran outliers	177936	5594 (1.70-1.70)
Sidechain outliers	177891	5594 (1.70-1.70)
RSRZ outliers	164620	5159 (1.70-1.70)

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	220	3% / 89% / 10% ••
1	B	220	3% / 85% / 12% ••
1	C	220	5% / 90% / 9% •
1	D	220	3% / 85% / 13% •

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 7074 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 Variable lymphocyte receptor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	218	Total	C	N	O	S	0	0	0
			1664	1049	292	312	11			
1	A	218	Total	C	N	O	S	0	0	0
			1664	1049	292	312	11			
1	C	217	Total	C	N	O	S	0	0	0
			1657	1044	291	311	11			
1	D	217	Total	C	N	O	S	0	0	0
			1657	1044	291	311	11			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MET	-	initiating methionine	UNP K0IE77
A	1	MET	-	initiating methionine	UNP K0IE77
C	1	MET	-	initiating methionine	UNP K0IE77
D	1	MET	-	initiating methionine	UNP K0IE77

- Molecule 2 is beta-D-galactopyranose (CCD ID: GAL) (formula: $C_6H_{12}O_6$).
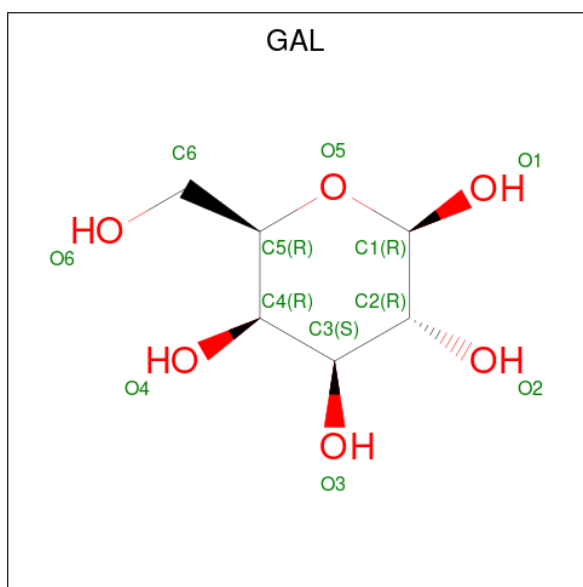
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	B	1	Total 12	C 6	O 6	0	0
2	A	1	Total 12	C 6	O 6	0	0
2	C	1	Total 12	C 6	O 6	0	0
2	D	1	Total 12	C 6	O 6	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	95	Total 95	O 95	0	0
3	A	108	Total 108	O 108	0	0
3	C	83	Total 83	O 83	0	0
3	D	98	Total 98	O 98	0	0

3 Residue-property plots ⓘ

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: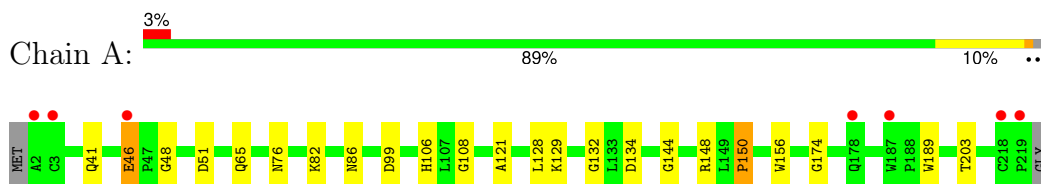 green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.
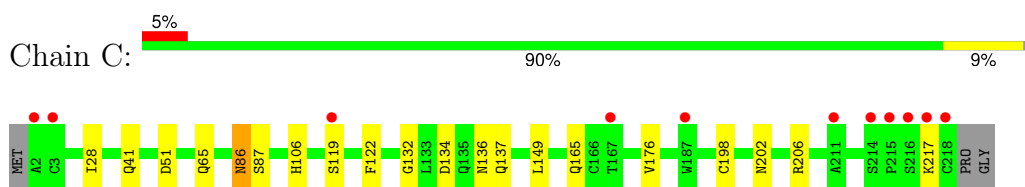
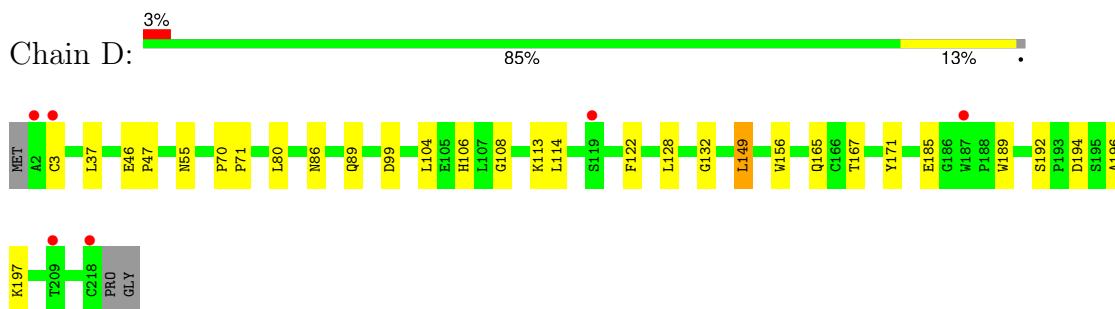• Molecule 1: Variable lymphocyte receptor
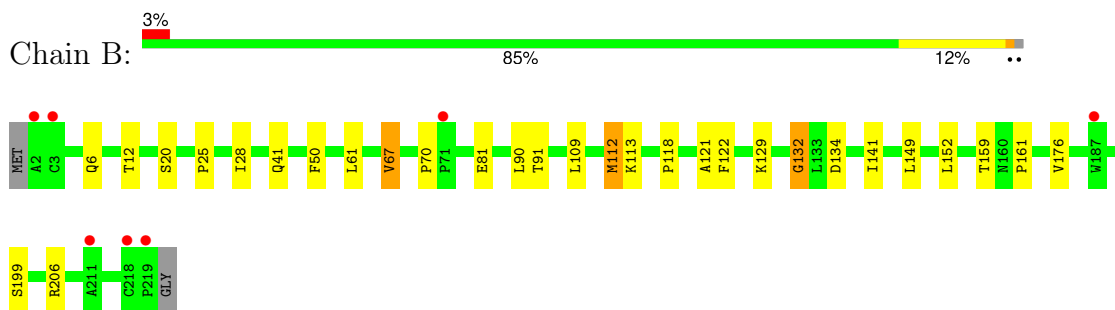
Chain B:

• Molecule 1: Variable lymphocyte receptor

Chain A:

• Molecule 1: Variable lymphocyte receptor

Chain C:

• Molecule 1: Variable lymphocyte receptor

Chain D:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	57.71Å 62.78Å 63.37Å 72.03° 71.39° 84.35°	Depositor
Resolution (Å)	59.72 – 1.70 59.72 – 1.70	Depositor EDS
% Data completeness (in resolution range)	93.4 (59.72-1.70) 93.4 (59.72-1.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.81 (at 1.70Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.206 , 0.247 0.201 , 0.241	Depositor DCC
R_{free} test set	4154 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	22.6	Xtriage
Anisotropy	0.270	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 26.8	EDS
L-test for twinning[2]	$< |L| > = 0.52, < L^2 > = 0.35$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	7074	wwPDB-VP
Average B, all atoms (Å^2)	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.86% 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 ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: GAL

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	1.58	12/1706 (0.7%)	1.35	3/2332 (0.1%)
1	B	1.57	6/1706 (0.4%)	1.30	4/2332 (0.2%)
1	C	1.44	4/1698 (0.2%)	1.32	5/2320 (0.2%)
1	D	1.59	9/1698 (0.5%)	1.36	9/2320 (0.4%)
All	All	1.54	31/6808 (0.5%)	1.33	21/9304 (0.2%)

The worst 5 of 31 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	141	ILE	CA-CB	8.59	1.61	1.54
1	D	149	LEU	CA-CB	7.86	1.61	1.53
1	B	132	GLY	N-CA	6.56	1.53	1.44
1	A	132	GLY	N-CA	6.35	1.54	1.45
1	B	112	MET	N-CA	6.27	1.54	1.46

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	198	CYS	CA-C-N	6.88	129.39	120.44
1	C	198	CYS	C-N-CA	6.88	129.39	120.44
1	A	46	GLU	CA-C-N	-6.73	112.84	119.78
1	A	46	GLU	C-N-CA	-6.73	112.84	119.78
1	B	28	ILE	N-CA-C	-6.68	102.41	108.95

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	1664	0	1643	6	0
1	B	1664	0	1641	16	0
1	C	1657	0	1634	8	0
1	D	1657	0	1634	8	0
2	A	12	0	12	0	0
2	B	12	0	12	2	0
2	C	12	0	12	2	0
2	D	12	0	12	2	0
3	A	108	0	0	1	0
3	B	95	0	0	4	0
3	C	83	0	0	1	0
3	D	98	0	0	1	0
All	All	7074	0	6600	40	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 3.

The worst 5 of 40 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:165:GLN:HB3	1:C:217:LYS:HD3	1.62	0.80
2:C:301:GAL:H61	3:C:483:HOH:O	1.86	0.75
1:C:132:GLY:HA3	2:C:301:GAL:H62	1.80	0.62
1:A:41:GLN:HA	1:A:65:GLN:OE1	2.00	0.62
1:C:165:GLN:HB3	1:C:217:LYS:CD	2.33	0.57

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	216/220 (98%)	208 (96%)	8 (4%)	0	100	100
1	B	216/220 (98%)	207 (96%)	9 (4%)	0	100	100
1	C	215/220 (98%)	206 (96%)	9 (4%)	0	100	100
1	D	215/220 (98%)	209 (97%)	6 (3%)	0	100	100
All	All	862/880 (98%)	830 (96%)	32 (4%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2   Protein sidechains ⓘ

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	189/190 (100%)	187 (99%)	2 (1%)	70	60
1	B	189/190 (100%)	188 (100%)	1 (0%)	86	82
1	C	188/190 (99%)	187 (100%)	1 (0%)	86	82
1	D	188/190 (99%)	186 (99%)	2 (1%)	70	60
All	All	754/760 (99%)	748 (99%)	6 (1%)	79	71

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	86	ASN
1	D	86	ASN
1	D	114	LEU
1	A	86	ASN
1	B	67	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	41	GLN
1	C	60	HIS
1	C	137	GLN
1	C	179	HIS
1	D	60	HIS

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 oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

4 ligands are 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
2	GAL	A	301	-	12,12,12	1.05	0	17,17,17	1.20	1 (5%)
2	GAL	D	301	-	12,12,12	1.05	1 (8%)	17,17,17	1.76	3 (17%)
2	GAL	C	301	-	12,12,12	0.87	0	17,17,17	2.01	7 (41%)
2	GAL	B	301	-	12,12,12	0.96	0	17,17,17	1.68	3 (17%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GAL	A	301	-	-	0/2/22/22	0/1/1/1
2	GAL	D	301	-	-	2/2/22/22	0/1/1/1
2	GAL	C	301	-	-	2/2/22/22	0/1/1/1
2	GAL	B	301	-	-	2/2/22/22	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	301	GAL	O4-C4	-2.20	1.37	1.43

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	301	GAL	C1-O5-C5	4.06	121.51	113.65
2	C	301	GAL	O3-C3-C4	3.66	119.02	110.38
2	B	301	GAL	O3-C3-C2	-3.66	101.74	110.38
2	C	301	GAL	O2-C2-C3	-3.24	102.73	110.38
2	D	301	GAL	C1-O5-C5	3.16	119.77	113.65

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	301	GAL	C4-C5-C6-O6
2	B	301	GAL	O5-C5-C6-O6
2	D	301	GAL	O5-C5-C6-O6
2	C	301	GAL	C4-C5-C6-O6
2	C	301	GAL	O5-C5-C6-O6

There are no ring outliers.

3 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	301	GAL	2	0
2	C	301	GAL	2	0
2	B	301	GAL	2	0

5.7 Other polymers ⓘ

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 ⓘ

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, 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(Å²)	Q<0.9
1	A	218/220 (99%)	-0.00	7 (3%)	50	53	11, 19, 34, 44	0
1	B	218/220 (99%)	0.12	7 (3%)	50	53	13, 21, 34, 47	0
1	C	217/220 (98%)	0.30	11 (5%)	34	37	12, 23, 40, 64	0
1	D	217/220 (98%)	0.11	6 (2%)	55	58	12, 20, 35, 45	0
All	All	870/880 (98%)	0.13	31 (3%)	46	49	11, 21, 36, 64	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	218	CYS	7.2
1	C	215	PRO	5.3
1	D	2	ALA	4.6
1	D	218	CYS	4.2
1	C	214	SER	3.4

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 monosaccharides in this entry.

6.4 Ligands ⓘ

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(Å²)	Q<0.9
2	GAL	D	301	12/12	0.82	0.12	24,28,34,35	0
2	GAL	B	301	12/12	0.83	0.12	21,29,34,36	0
2	GAL	C	301	12/12	0.84	0.11	24,28,30,42	0
2	GAL	A	301	12/12	0.85	0.11	20,25,32,34	0

6.5 Other polymers ⓘ

There are no such residues in this entry.