wwPDB EM Validation Summary Report (i)

Oct 16, 2024 – 12:48 AM JST

PDB ID	:	8KHR
EMDB ID	:	EMD-37249
Title	:	Cryo-EM structure of EBV gH/gL-gp42 in complex with fab 2C1
Authors	:	Fang, X.Y.; Zhao, G.X.; Zeng, M.S.; Liu, Z.
Deposited on	:	2023-08-22
Resolution	:	3.25 Å(reported)

This is a wwPDB EM 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/EMValidationReportHelp
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:

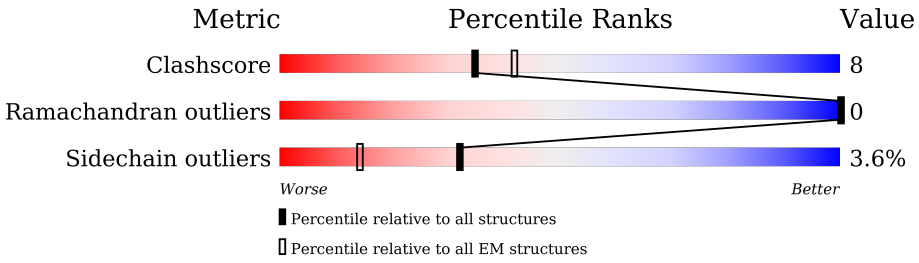
EMDB validation analysis	:	FAILED
MolProbity	:	4.02b-467
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
MapQ	:	FAILED
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 3.25 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)	EM structures (#Entries)
Clashscore	210492	15764
Ramachandran outliers	207382	16835
Sidechain outliers	206894	16415

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the 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%

Mol	Chain	Length	Quality of chain
1	C	196	70% 17% • 12%
2	H	111	76% 22% •
3	L	107	82% 17% •
4	A	655	83% 16% ••
5	B	107	68% 22% 9%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 8320 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 Soluble gp42.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	C	173	Total	C	N	O	S	0	0
			1267	821	216	227	3		

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	224	HIS	-	expression tag	UNP P0C6Z5
C	225	HIS	-	expression tag	UNP P0C6Z5
C	226	HIS	-	expression tag	UNP P0C6Z5
C	227	HIS	-	expression tag	UNP P0C6Z5
C	228	HIS	-	expression tag	UNP P0C6Z5
C	229	HIS	-	expression tag	UNP P0C6Z5

- Molecule 2 is a protein called 2C1 heavy chain.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	H	111	Total	C	N	O	S	0	0
			704	437	128	136	3		

- Molecule 3 is a protein called 2C1 light chain.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	L	107	Total	C	N	O	S	0	0
			703	434	127	140	2		

- Molecule 4 is a protein called Envelope glycoprotein H.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	A	650	Total	C	N	O	S	0	0
			4938	3178	825	909	26		

- Molecule 5 is a protein called Envelope glycoprotein L.

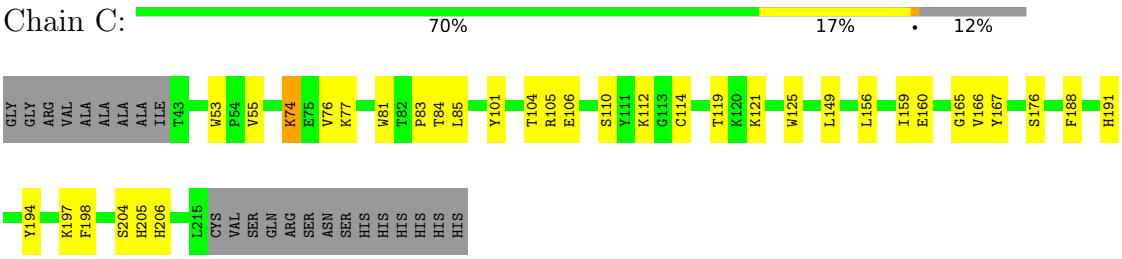
Mol	Chain	Residues	Atoms					AltConf	Trace
5	B	97	Total	C	N	O	S	0	0
			708	447	122	137	2		

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. 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. 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.

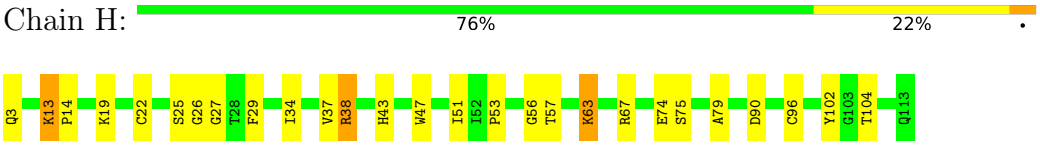
- Molecule 1: Soluble gp42

Chain C:



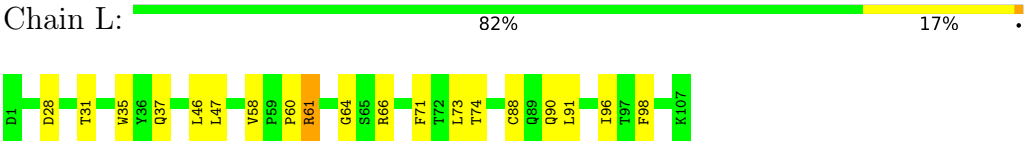
- Molecule 2: 2C1 heavy chain

Chain H:



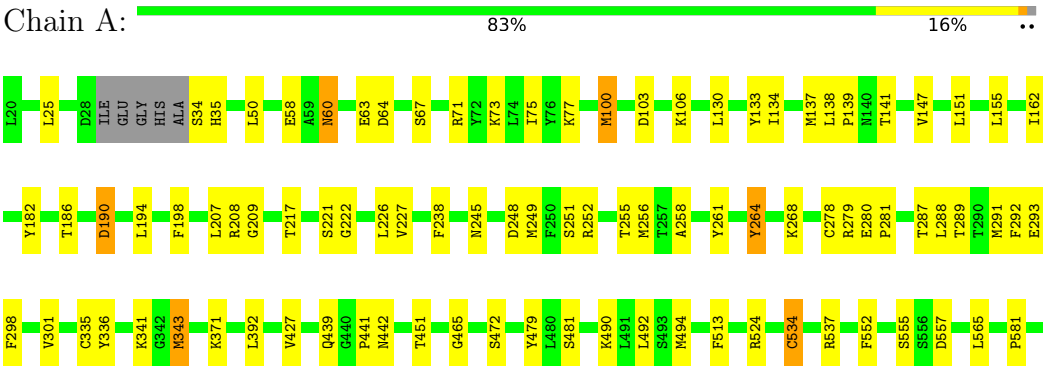
- Molecule 3: 2C1 light chain

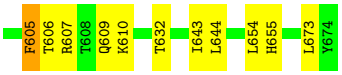
Chain L:



- Molecule 4: Envelope glycoprotein H

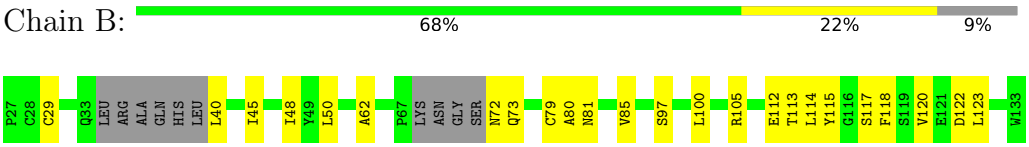
Chain A:





• Molecule 5: Envelope glycoprotein L

Chain B:



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	227428	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	50	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	2000	Depositor
Magnification	Not provided	
Image detector	GATAN K3 (6k x 4k)	Depositor

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	0.25	0/1311	0.49	0/1805
2	H	0.26	0/716	0.54	0/980
3	L	0.27	0/714	0.57	0/986
4	A	0.26	0/5044	0.50	0/6857
5	B	0.25	0/718	0.53	0/975
All	All	0.26	0/8503	0.51	0/11603

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

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	1267	0	1100	24	0
2	H	704	0	562	14	0
3	L	703	0	601	12	0
4	A	4938	0	4857	71	0
5	B	708	0	657	14	0
All	All	8320	0	7777	125	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 8.

The worst 5 of 125 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)
4:A:60:ASN:O	4:A:60:ASN:ND2	2.12	0.80
3:L:58:VAL:HG12	3:L:60:PRO:HD2	1.67	0.77
3:L:31:THR:HG21	3:L:90:GLN:HE21	1.50	0.76
3:L:61:ARG:HH21	4:A:73:LYS:HZ1	1.41	0.69
1:C:188:PHE:HD2	1:C:206:HIS:HB3	1.57	0.69

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 PDB entries followed by that with respect to all EM entries.

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	171/196 (87%)	152 (89%)	19 (11%)	0	100	100
2	H	109/111 (98%)	84 (77%)	25 (23%)	0	100	100
3	L	105/107 (98%)	94 (90%)	11 (10%)	0	100	100
4	A	646/655 (99%)	616 (95%)	30 (5%)	0	100	100
5	B	91/107 (85%)	87 (96%)	4 (4%)	0	100	100
All	All	1122/1176 (95%)	1033 (92%)	89 (8%)	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 PDB entries followed by that with respect to all EM entries.

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	116/175 (66%)	108 (93%)	8 (7%)	13	37

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	H	48/88 (54%)	40 (83%)	8 (17%)	2 8
3	L	59/93 (63%)	57 (97%)	2 (3%)	32 57
4	A	513/561 (91%)	504 (98%)	9 (2%)	54 72
5	B	72/94 (77%)	70 (97%)	2 (3%)	38 62
All	All	808/1011 (80%)	779 (96%)	29 (4%)	32 55

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	H	96	CYS
5	B	115	TYR
3	L	98	PHE
4	A	513	PHE
3	L	61	ARG

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

Mol	Chain	Res	Type
3	L	90	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 oligosaccharides in this entry.

5.6 Ligand geometry

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.