wwPDB X-ray Structure Validation Summary Report (i)

May 16, 2020 – 05:42 pm BST

PDB ID	:	5CJQ
Title	:	Crystal structure of a trimeric influenza hemagglutinin stem in complex with an broadly neutralizing antibody CR9114
Authors	:	Zhu, X.; Wilson, I.A.
Deposited on	:	2015-07-14
Resolution	:	3.60 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 following versions of software and data (see references (1)) were used in the production of this report:

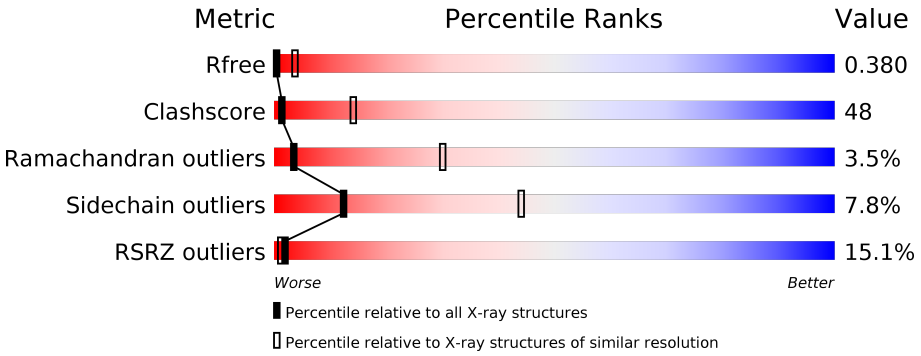
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 3.60 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	1257 (3.70-3.50)
Clashscore	141614	1353 (3.70-3.50)
Ramachandran outliers	138981	1307 (3.70-3.50)
Sidechain outliers	138945	1307 (3.70-3.50)
RSRZ outliers	127900	1161 (3.70-3.50)

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 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	L	215	25% 34% 48% 13%
2	H	230	13% 49% 40% 6%
3	A	66	3% 30% 45% 21%
4	B	193	6% 34% 39% 6% 20%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 4812 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 CR9114 light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	L	211	Total	C	N	O	S	0	0	0
			1568	978	266	320	4			

- Molecule 2 is a protein called CR9114 heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	217	Total	C	N	O	S	0	0	0
			1613	1017	269	320	7			

- Molecule 3 is a protein called Designed influenza hemagglutinin stem #4900, HA1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	52	Total	C	N	O	S	0	0	0
			397	242	72	80	3			

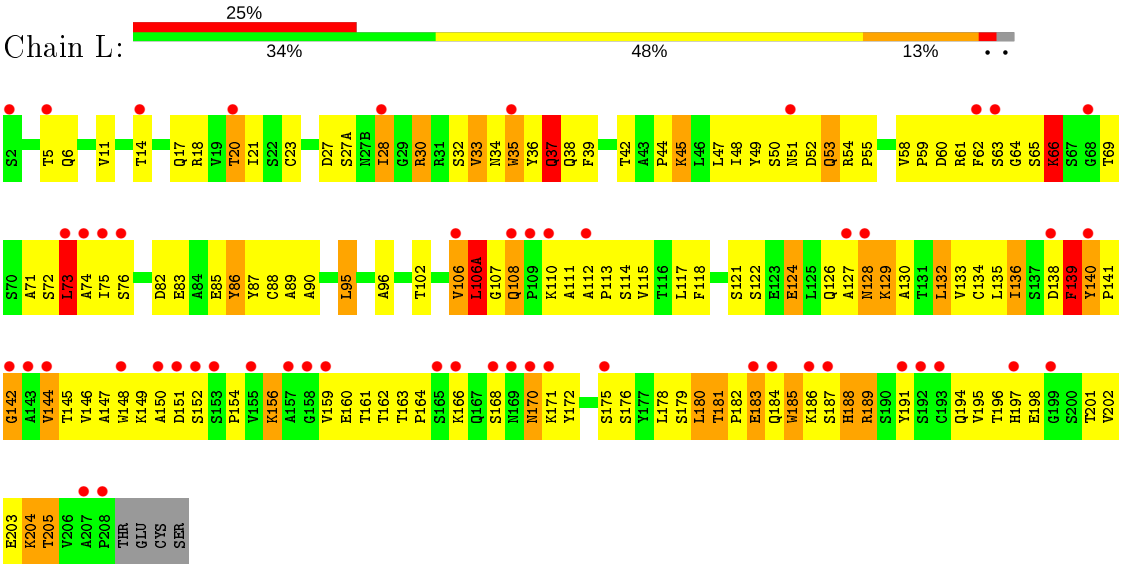
- Molecule 4 is a protein called Designed influenza hemagglutinin stem #4900, HA2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	B	154	Total	C	N	O	S	0	0	0
			1234	767	208	251	8			

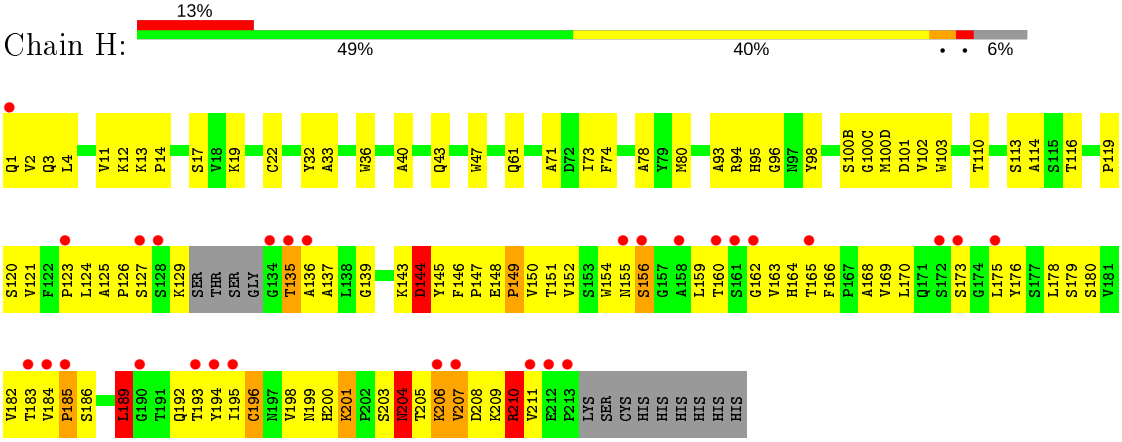
3 Residue-property plots (i)

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

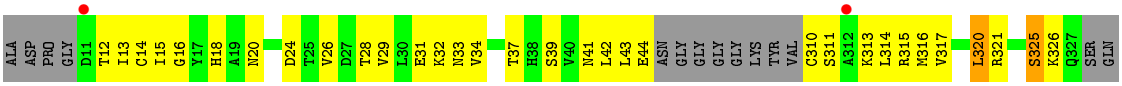
- Molecule 1: CR9114 light chain



- Molecule 2: CR9114 heavy chain

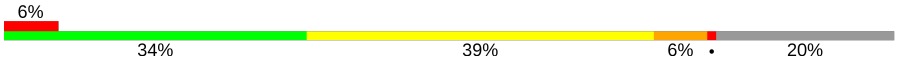


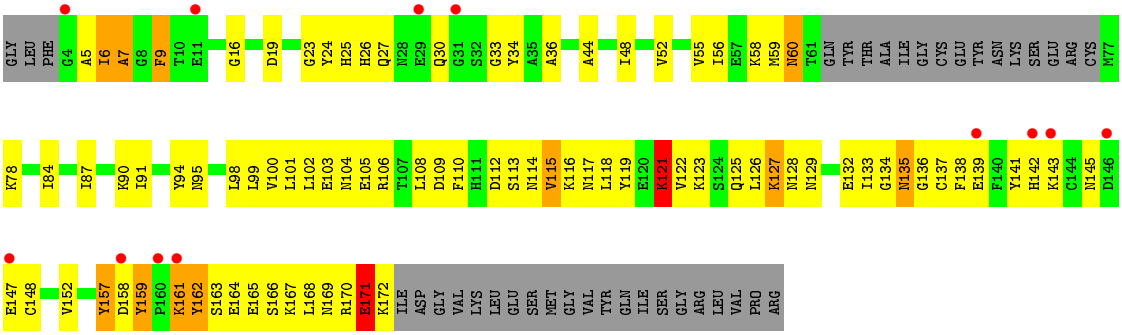
- Molecule 3: Designed influenza hemagglutinin stem #4900, HA1



• Molecule 4: Designed influenza hemagglutinin stem #4900, HA2

Chain B:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, α, β, γ	157.70Å 157.70Å 202.65Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	38.85 – 3.60 28.32 – 3.58	Depositor EDS
% Data completeness (in resolution range)	99.1 (38.85-3.60) 99.6 (28.32-3.58)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.10	Depositor
$< I/\sigma(I) >$ [1]	1.68 (at 3.55Å)	Xtriage
Refinement program	PHENIX 1.8.4_1496	Depositor
R, R_{free}	0.346 , 0.369 0.349 , 0.380	Depositor DCC
R_{free} test set	550 reflections (4.74%)	wwPDB-VP
Wilson B-factor (Å^2)	138.2	Xtriage
Anisotropy	0.241	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 125.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.030 for -2/3*h-1/3*k+2/3*l,-1/3*h-2/3*k-2/3*l,2/3*h-2/3*k+1/3*l 0.003 for -h,1/3*h-1/3*k+2/3*l,2/3*h+4/3*k+1/3*l 0.000 for -1/3*h+1/3*k-2/3*l,-k,-4/3*h-2/3*k+1/3*l	Xtriage
F_o,F_c correlation	0.85	EDS
Total number of atoms	4812	wwPDB-VP
Average B, all atoms (Å^2)	158.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.15% 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

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	L	0.75	1/1606 (0.1%)	1.43	20/2193 (0.9%)
2	H	0.69	5/1652 (0.3%)	1.08	9/2251 (0.4%)
3	A	0.61	0/399	1.02	1/539 (0.2%)
4	B	0.44	0/1256	0.75	0/1687
All	All	0.65	6/4913 (0.1%)	1.14	30/6670 (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	L	0	2
2	H	0	2
All	All	0	4

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	H	206	LYS	CA-CB	7.92	1.71	1.53
2	H	206	LYS	CD-CE	6.83	1.68	1.51
2	H	206	LYS	CB-CG	5.88	1.68	1.52
2	H	207	VAL	CB-CG2	-5.75	1.40	1.52
1	L	35	TRP	CB-CG	-5.61	1.40	1.50

The worst 5 of 30 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	L	142	GLY	N-CA-C	13.49	146.82	113.10
2	H	206	LYS	CB-CG-CD	12.76	144.77	111.60
1	L	139	PHE	CB-CA-C	-8.75	92.90	110.40
1	L	132	LEU	CB-CG-CD1	-8.52	96.51	111.00
2	H	206	LYS	N-CA-C	-8.29	88.61	111.00

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	H	203	SER	Peptide
2	H	206	LYS	Mainchain
1	L	129	LYS	Peptide
1	L	188	HIS	Peptide

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	L	1568	0	1517	213	0
2	H	1613	0	1568	123	0
3	A	397	0	401	40	0
4	B	1234	0	1153	112	0
All	All	4812	0	4639	455	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 48.

The worst 5 of 455 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:H:114:ALA:HB1	2:H:146:PHE:CE2	1.56	1.40
2:H:114:ALA:CB	2:H:146:PHE:HE2	1.43	1.32
2:H:114:ALA:CB	2:H:146:PHE:CE2	2.21	1.19
1:L:140:TYR:CE1	1:L:171:LYS:HD2	1.82	1.14
2:H:146:PHE:HB2	2:H:175:LEU:HD11	1.28	1.10

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	L	209/215 (97%)	200 (96%)	4 (2%)	5 (2%)	6 37
2	H	213/230 (93%)	204 (96%)	6 (3%)	3 (1%)	11 48
3	A	48/66 (73%)	44 (92%)	3 (6%)	1 (2%)	7 40
4	B	150/193 (78%)	112 (75%)	25 (17%)	13 (9%)	1 9
All	All	620/704 (88%)	560 (90%)	38 (6%)	22 (4%)	3 30

5 of 22 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	L	107	GLY
1	L	128	ASN
2	H	189	LEU
2	H	204	ASN
3	A	325	SER

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	L	175/179 (98%)	152 (87%)	23 (13%)	4 23
2	H	181/193 (94%)	170 (94%)	11 (6%)	18 53
3	A	48/56 (86%)	45 (94%)	3 (6%)	18 53
4	B	133/167 (80%)	128 (96%)	5 (4%)	33 66
All	All	537/595 (90%)	495 (92%)	42 (8%)	12 44

5 of 42 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	L	181	THR
2	H	19	LYS
4	B	121	LYS
1	L	183	GLU
1	L	204	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	L	188	HIS
2	H	155	ASN
2	H	200	HIS
1	L	184	GLN
2	H	192	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 carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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, 95^{th} 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	L	211/215 (98%)	1.35	53 (25%) 0 0	104, 181, 245, 333	0
2	H	217/230 (94%)	0.78	29 (13%) 3 2	74, 135, 238, 337	0
3	A	52/66 (78%)	0.29	2 (3%) 40 26	91, 123, 170, 200	0
4	B	154/193 (79%)	0.31	12 (7%) 13 8	90, 141, 191, 220	0
All	All	634/704 (90%)	0.81	96 (15%) 2 1	74, 160, 234, 337	0

The worst 5 of 96 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	L	208	PRO	13.6
1	L	169	ASN	12.1
2	H	184	VAL	9.6
2	H	161	SER	7.8
1	L	171	LYS	7.7

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 carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.