



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 19, 2016 – 09:56 PM GMT

PDB ID : 5C0S
Title : Crystal structure of a generation 4 influenza hemagglutinin stabilized stem in complex with the broadly neutralizing antibody CR6261
Authors : Boyington, J.C.; Kwong, P.D.; Nabel, G.J.; Mascola, J.R.
Deposited on : 2015-06-12
Resolution : 4.30 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-20026982
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-20026982

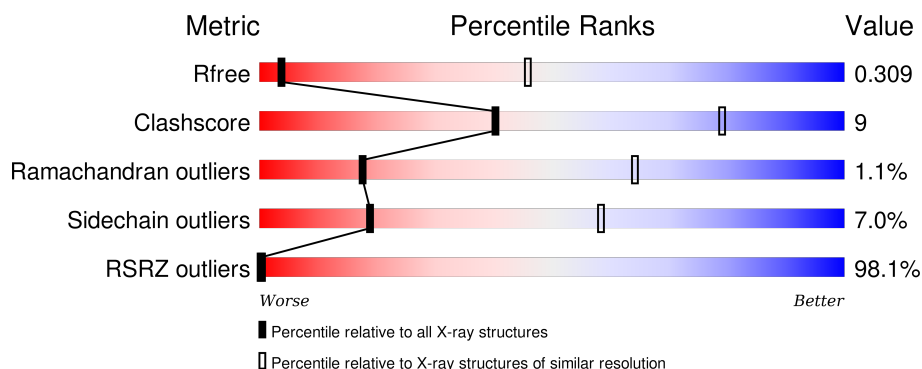
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 4.30 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	1059 (5.00-3.60)
Clashscore	102246	1166 (5.00-3.60)
Ramachandran outliers	100387	1106 (5.00-3.60)
Sidechain outliers	100360	1089 (5.00-3.60)
RSRZ outliers	91569	1062 (5.00-3.60)

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	A	305	<div> <div>78%</div> <div>56% 21% • 21%</div> </div>
2	H	226	<div> <div>95%</div> <div>78% 17% • •</div> </div>
3	L	221	<div> <div>95%</div> <div>77% 16% • •</div> </div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 5158 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 Hemagglutinin, Envelope glycoprotein, Fibrin fusion protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	242	Total	C	N	O	S	0	0	0
			1930	1198	336	388	8			

There are 37 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	33	GLY	-	linker	UNP Q6WG00
A	34	SER	-	linker	UNP Q6WG00
A	35	GLY	-	linker	UNP Q6WG00
A	47	GLN	SER	conflict	UNP Q6WG00
A	47A	ARG	ILE	conflict	UNP Q6WG00
A	47B	GLU	GLN	conflict	UNP Q6WG00
A	47C	THR	SER	conflict	UNP Q6WG00
A	108	GLY	-	linker	UNP Q6WG00
A	109	GLY	-	linker	UNP Q6WG00
A	110	ASP	TRP	conflict	UNP P04578
A	111	PRO	MET	conflict	UNP P04578
A	123	ILE	LEU	conflict	UNP P04578
A	125	TYR	HIS	conflict	UNP P04578
A	137	ASN	-	linker	UNP P04578
A	138	GLY	-	linker	UNP P04578
A	139	THR	-	linker	UNP P04578
A	140	GLY	-	linker	UNP P04578
A	141	GLY	-	linker	UNP P04578
A	142	GLY	-	linker	UNP P04578
A	257	PRO	-	linker	UNP Q6WG00
A	258	GLY	-	linker	UNP Q6WG00
A	259	SER	-	linker	UNP Q6WG00
A	288	GLY	-	expression tag	UNP D9IEJ2
A	289	ARG	-	expression tag	UNP D9IEJ2
A	290	LEU	-	expression tag	UNP D9IEJ2
A	291	VAL	-	expression tag	UNP D9IEJ2
A	292	PRO	-	expression tag	UNP D9IEJ2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	293	ARG	-	expression tag	UNP D9IEJ2
A	294	GLY	-	expression tag	UNP D9IEJ2
A	295	SER	-	expression tag	UNP D9IEJ2
A	296	GLY	-	expression tag	UNP D9IEJ2
A	297	HIS	-	expression tag	UNP D9IEJ2
A	298	HIS	-	expression tag	UNP D9IEJ2
A	299	HIS	-	expression tag	UNP D9IEJ2
A	300	HIS	-	expression tag	UNP D9IEJ2
A	301	HIS	-	expression tag	UNP D9IEJ2
A	302	HIS	-	expression tag	UNP D9IEJ2

- Molecule 2 is a protein called CR6261 antibody heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	221	Total	C	N	O	S	0	0	0
			1643	1039	274	321	9			

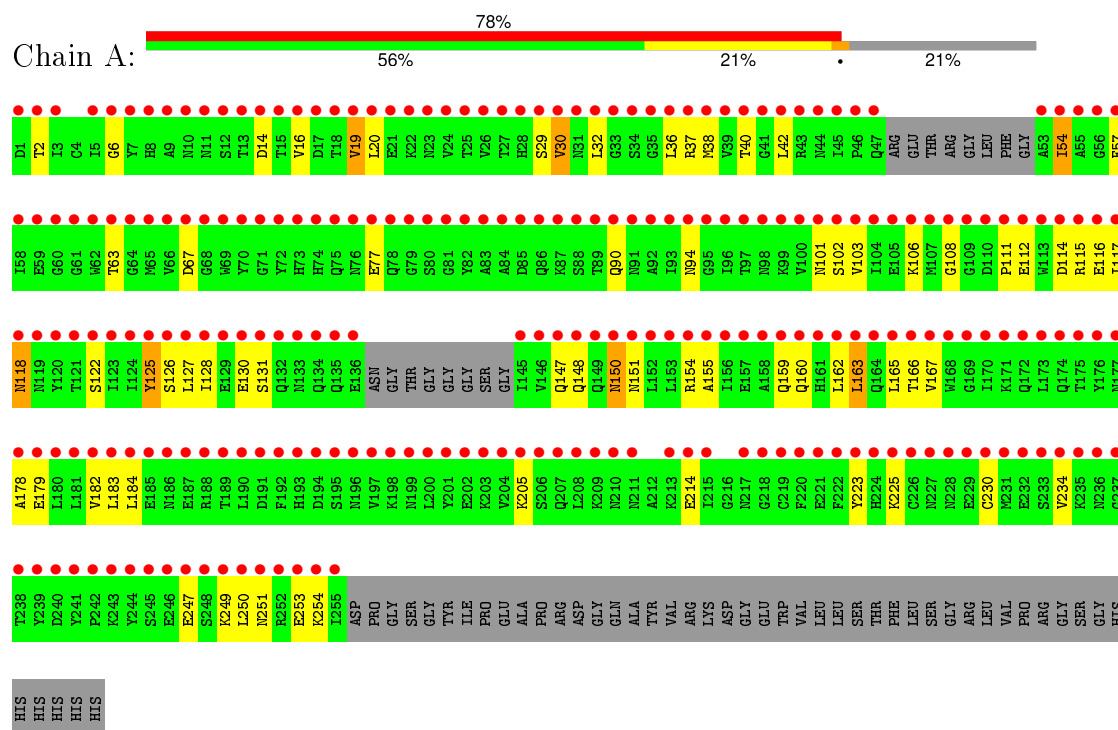
- Molecule 3 is a protein called CR6261 antibody light chain.

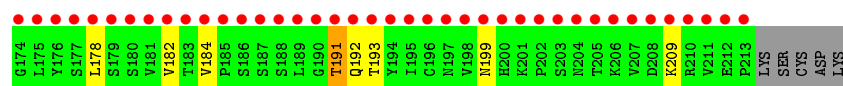
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	L	212	Total	C	N	O	S	0	0	0
			1585	995	266	320	4			

3 Residue-property plots

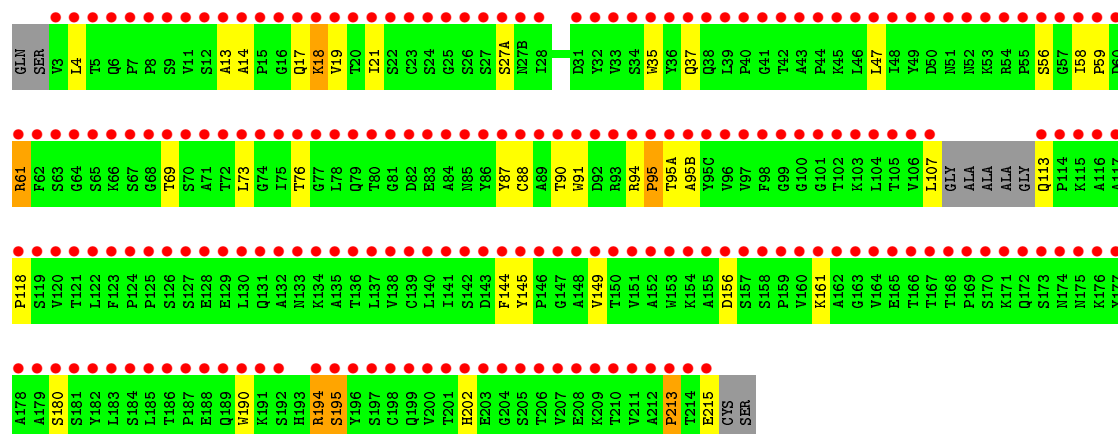
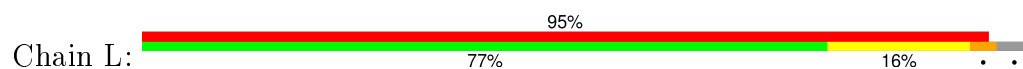
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: Hemagglutinin, Envelope glycoprotein, Fibritin fusion protein





● Molecule 3: CR6261 antibody light chain



4 Data and refinement statistics

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α , β , γ	101.65Å 101.65Å 186.02Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	27.06 – 4.30 27.06 – 4.30	Depositor EDS
% Data completeness (in resolution range)	99.0 (27.06-4.30) 99.1 (27.06-4.30)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.41 (at 4.25Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: 1.8.2_1309)	Depositor
R, R_{free}	0.238 , 0.310 0.237 , 0.309	Depositor DCC
R_{free} test set	240 reflections (5.00%)	DCC
Wilson B-factor (Å ²)	166.1	Xtriage
Anisotropy	0.325	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	508.12 , -10.0	EDS
Estimated twinning fraction	0.073 for h,-h-k,-l	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.27$	Xtriage
Outliers	0 of 4799 reflections	Xtriage
F_o, F_c correlation	0.65	EDS
Total number of atoms	5158	wwPDB-VP
Average B, all atoms (Å ²)	86.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.61% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	A	0.30	0/1961	0.52	0/2650
2	H	0.27	0/1684	0.50	0/2292
3	L	0.26	0/1624	0.53	0/2220
All	All	0.28	0/5269	0.51	0/7162

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	A	1930	0	1850	50	0
2	H	1643	0	1625	36	0
3	L	1585	0	1547	20	0
All	All	5158	0	5022	89	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 9.

The worst 5 of 89 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:A:90:GLN:NE2	2:H:31:SER:O	2.07	0.87

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:114:ASP:O	1:A:118:ASN:ND2	2.16	0.79
2:H:119:PRO:HB3	2:H:145:TYR:HB3	1.67	0.75
1:A:36:LEU:HD11	1:A:178:ALA:HB1	1.72	0.71
1:A:101:ASN:HD21	2:H:29:PHE:HB2	1.55	0.71

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	236/305 (77%)	216 (92%)	18 (8%)	2 (1%)	24	70
2	H	219/226 (97%)	210 (96%)	7 (3%)	2 (1%)	21	67
3	L	208/221 (94%)	192 (92%)	13 (6%)	3 (1%)	14	59
All	All	663/752 (88%)	618 (93%)	38 (6%)	7 (1%)	17	64

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	115	ARG
3	L	95	PRO
1	A	54	ILE
2	H	133	GLY
3	L	213	PRO

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	A	213/260 (82%)	192 (90%)	21 (10%)	10	42
2	H	184/189 (97%)	175 (95%)	9 (5%)	31	68
3	L	178/182 (98%)	168 (94%)	10 (6%)	26	65
All	All	575/631 (91%)	535 (93%)	40 (7%)	19	58

5 of 40 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	166	THR
2	H	64	GLN
3	L	180	SER
1	A	249	LYS
2	H	98	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	101	ASN
1	A	118	ASN
1	A	147	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 [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	A	242/305 (79%)	10.53	239 (98%) 0 0	25, 75, 165, 251	0
2	H	221/226 (97%)	9.13	214 (96%) 0 0	23, 61, 140, 194	0
3	L	212/221 (95%)	9.73	209 (98%) 0 0	37, 93, 177, 253	0
All	All	675/752 (89%)	9.82	662 (98%) 0 0	23, 77, 168, 253	0

The worst 5 of 662 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	233	SER	43.2
1	A	226	CYS	40.4
1	A	109	GLY	33.2
3	L	50	ASP	32.6
1	A	79	GLY	32.5

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.