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 (i)

May 13, 2020 – 06:41 am BST

PDB ID	:	4FQY
Title	:	Crystal structure of broadly neutralizing antibody CR9114 bound to H3 influenza hemagglutinin
Authors	:	Ekiert, D.C.; Dreyfus, C.; Wilson, I.A.
Deposited on	:	2012-06-25
Resolution	:	5.25 Å(reported)

The following versions of software and data (see references (1)) were used in the production of this report:

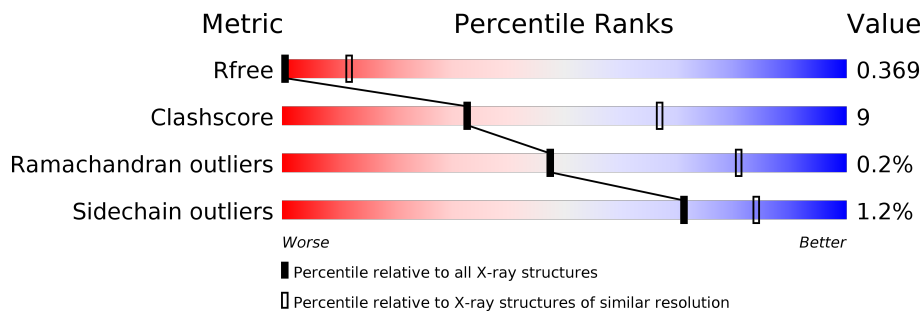|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 5.25 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	Percentile Ranks	Value
Rfree		0.369
Clashscore		9
Ramachandran outliers		0.2%
Sidechain outliers		1.2%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	1175 (6.72-3.80)
Clashscore	141614	1012 (6.64-3.84)
Ramachandran outliers	138981	1181 (6.72-3.80)
Sidechain outliers	138945	1155 (6.72-3.80)

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%

Mol	Chain	Length	Quality of chain
1	A	323	93% 5% ·
2	B	174	86% 13% ··
3	H	224	84% 10% 5%
4	L	216	86% 9% ··

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 7150 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 HA1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	318	Total 2553	C 1598	N 450	O 491	S 14	0	12	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	7	ALA	-	EXPRESSION TAG	UNP Q91MA7
A	8	ASP	-	EXPRESSION TAG	UNP Q91MA7
A	9	PRO	-	EXPRESSION TAG	UNP Q91MA7
A	10	GLY	-	EXPRESSION TAG	UNP Q91MA7

- Molecule 2 is a protein called Hemagglutinin HA2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	172	Total 1446	C 901	N 251	O 287	S 7	0	7	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	123	GLY	ARG	SEE REMARK 999	UNP Q91MA7

- Molecule 3 is a protein called Antibody CR9114 heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	H	213	Total 1596	C 1007	N 266	O 316	S 7	0	2	0

- Molecule 4 is a protein called Antibody CR9114 light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	L	209	1555	971	263	317	4	0	0	0

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. 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: Hemagglutinin HA1

Chain A:

- Molecule 2: Hemagglutinin HA2

Chain B:

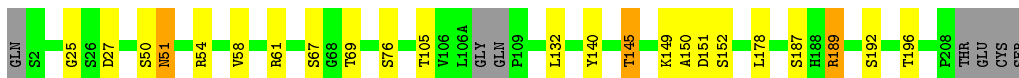- Molecule 3: Antibody CR9114 heavy chain

Chain H:

- Molecule 4: Antibody CR9114 light chain

Chain L:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 21 3	Depositor
Cell constants a, b, c, α, β, γ	203.71Å　　203.71Å　　203.71Å 90.00°　　90.00°　　90.00°	Depositor
Resolution (Å)	48.02　–　5.25 48.01　–　5.25	Depositor EDS
% Data completeness (in resolution range)	99.7 (48.02-5.25) 99.4 (48.01-5.25)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$<I/\sigma(I)>$ [1]	1.55 (at 5.39Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, R_{free}	0.366　,　0.374 0.374　,　0.369	Depositor DCC
R_{free} test set	528 reflections (9.82%)	wwPDB-VP
Wilson B-factor (Å^2)	302.5	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 194.7	EDS
L-test for twinning[2]	$<|L|> = 0.44, <L^2> = 0.27$	Xtriage
Estimated twinning fraction	0.061 for -l,-k,-h	Xtriage
F_o,F_c correlation	0.78	EDS
Total number of atoms	7150	wwPDB-VP
Average B, all atoms (Å^2)	166.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.14% 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)

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	0.52	1/2612 (0.0%)	0.63	0/3558
2	B	0.62	0/1470	0.65	0/1975
3	H	0.40	0/1634	0.78	3/2226 (0.1%)
4	L	0.35	0/1592	0.50	0/2172
All	All	0.49	1/7308 (0.0%)	0.65	3/9931 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	323	VAL	CB-CG2	-5.12	1.42	1.52

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	H	114	ALA	O-C-N	17.76	151.12	122.70
3	H	114	ALA	CA-C-N	-14.06	86.27	117.20
3	H	114	ALA	C-N-CA	-11.47	93.04	121.70

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	2553	0	2494	14	1
2	B	1446	0	1371	77	26

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	H	1596	0	1543	103	5
4	L	1555	0	1506	9	13
All	All	7150	0	6914	122	40

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 122 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:B:21:TRP:CD1	3:H:54:PHE:HE1	1.08	1.61
2:B:21:TRP:CD1	3:H:54:PHE:CE1	1.86	1.60
2:B:21:TRP:NE1	3:H:54:PHE:CD1	1.81	1.47
3:H:11:VAL:HG11	3:H:146:PHE:CE2	1.67	1.30
2:B:41:THR:HG21	3:H:98:TYR:CZ	1.67	1.29

The worst 5 of 40 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)
2:B:76:ARG:NH2	2:B:77[B]:ILE:CB[6_555]	0.47	1.73
2:B:76:ARG:NH2	2:B:77[A]:ILE:CB[6_555]	0.49	1.71
3:H:210:ARG:NH2	4:L:189:ARG:NH2[9_555]	0.70	1.50
4:L:25:GLY:O	4:L:27:ASP:OD2[14_555]	1.11	1.09
2:B:76:ARG:NH2	2:B:77[A]:ILE:CG2[6_555]	1.16	1.04

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	328/323 (102%)	321 (98%)	6 (2%)	1 (0%)	41	76

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	177/174 (102%)	171 (97%)	6 (3%)	0	100	100
3	H	209/224 (93%)	206 (99%)	3 (1%)	0	100	100
4	L	205/216 (95%)	201 (98%)	3 (2%)	1 (0%)	29	68
All	All	919/937 (98%)	899 (98%)	18 (2%)	2 (0%)	47	81

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	L	51	ASN
1	A	62	ILE

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	291/283 (103%)	289 (99%)	2 (1%)	84	90
2	B	153/148 (103%)	152 (99%)	1 (1%)	84	90
3	H	178/187 (95%)	176 (99%)	2 (1%)	73	85
4	L	174/180 (97%)	170 (98%)	4 (2%)	50	70
All	All	796/798 (100%)	787 (99%)	9 (1%)	71	85

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	H	189	LEU
4	L	189	ARG
4	L	145	THR
2	B	60	ASN
4	L	67	SER

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

Mol	Chain	Res	Type
1	A	53	ASN
1	A	171	ASN
2	B	53	ASN
2	B	60	ASN
2	B	125	GLN

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 carbohydrates in this entry.

5.6 Ligand geometry (i)

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)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.