wwPDB X-ray Structure Validation Summary Report (i)

Apr 23, 2024 – 12:08 PM EDT

PDB ID	:	8CXC
Title	:	Novel Anti-Mesothelin Antibodies Enable Crystallography of the Intact Mesothelin Ectodo- main and Engineering of Potent, T cell-engaging Bispecific Therapeutics
Authors	:	Bandaranayake, A.D.; Rupert, P.B.; Lin, I.; Pilat, K.; Ruff, R.O.; Friend, D.J.; Chan, M.K.; Clarke, M.; Carter, J.; Meshinchi, S.; Mehlin, C.; Olson, J.M.; Strong, R.K.; Correnti, C.E.
Deposited on	:	2022-05-20
Resolution	:	4.31 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 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:

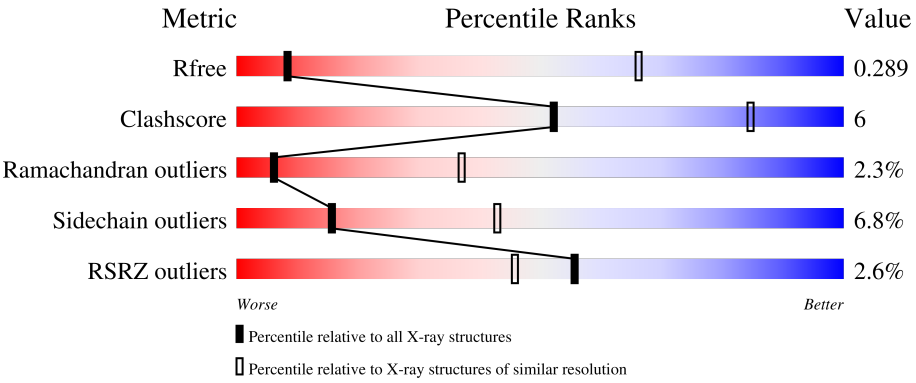
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 4.31 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	1014 (4.80-3.80)
Clashscore	141614	1077 (4.80-3.80)
Ramachandran outliers	138981	1029 (4.80-3.80)
Sidechain outliers	138945	1012 (4.80-3.80)
RSRZ outliers	127900	1075 (4.90-3.70)

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	L	213	% 73% 21% 6%
2	H	218	75% 22% .
3	M	327	77% 11% . 12%
4	A	234	9% 90% 5% 6%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 6984 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 3F2 Antibody light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	L	212	Total	C	N	O	S	0	0	0
			1605	1001	266	332	6			

- Molecule 2 is a protein called 3F2 Antibody heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	217	Total	C	N	O	S	0	0	0
			1571	993	263	309	6			

- Molecule 3 is a protein called Mesothelin, cleaved form.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	M	288	Total	C	N	O	S	0	0	0
			2136	1369	349	408	10			

There are 17 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
M	606	GLY	-	expression tag	UNP Q13421
M	607	SER	-	expression tag	UNP Q13421
M	608	GLY	-	expression tag	UNP Q13421
M	609	LEU	-	expression tag	UNP Q13421
M	610	ASN	-	expression tag	UNP Q13421
M	611	ASP	-	expression tag	UNP Q13421
M	612	ILE	-	expression tag	UNP Q13421
M	613	PHE	-	expression tag	UNP Q13421
M	614	GLU	-	expression tag	UNP Q13421
M	615	ALA	-	expression tag	UNP Q13421
M	616	GLN	-	expression tag	UNP Q13421
M	617	LYS	-	expression tag	UNP Q13421
M	618	ILE	-	expression tag	UNP Q13421
M	619	GLU	-	expression tag	UNP Q13421

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
M	620	TRP	-	expression tag	UNP Q13421
M	621	HIS	-	expression tag	UNP Q13421
M	622	GLU	-	expression tag	UNP Q13421

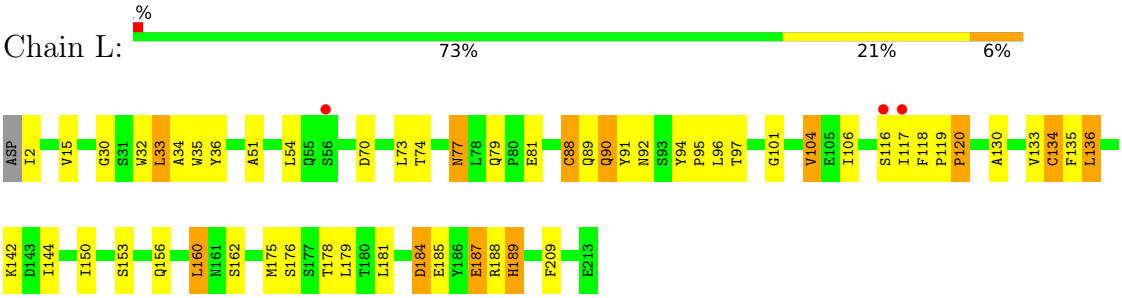
- Molecule 4 is a protein called scFv Amatuximab.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	A	221	1672	1050	273	340	9	0	0	0

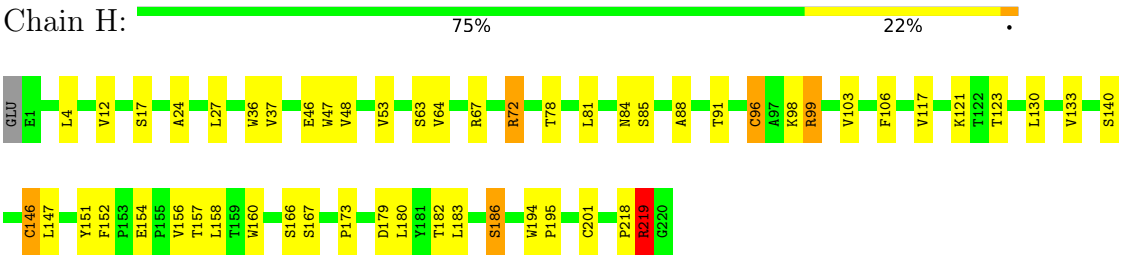
3 Residue-property plots (i)

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 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.

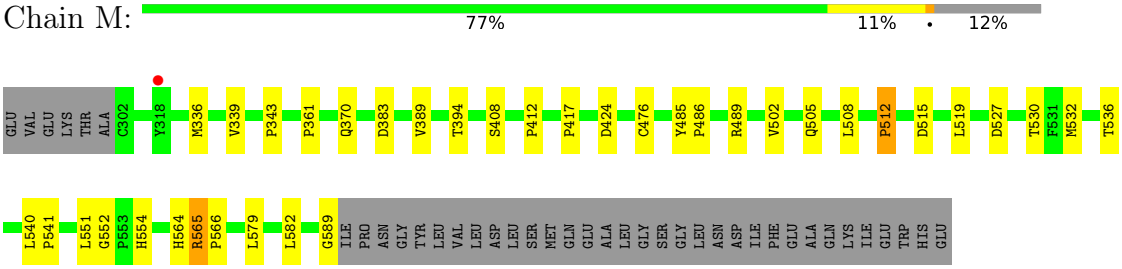
• Molecule 1: 3F2 Antibody light chain



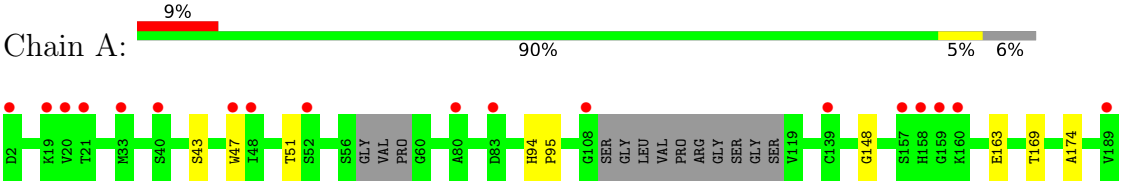
• Molecule 2: 3F2 Antibody heavy chain

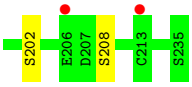


• Molecule 3: Mesothelin, cleaved form



• Molecule 4: scFv Amatuximab




S202
E206
D207
S208
C213
S235

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	185.30Å 204.18Å 123.11Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.78 – 4.31 45.74 – 4.31	Depositor EDS
% Data completeness (in resolution range)	84.0 (45.78-4.31) 84.1 (45.74-4.31)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.07 (at 4.28Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.243 , 0.300 0.240 , 0.289	Depositor DCC
R_{free} test set	667 reflections (4.93%)	wwPDB-VP
Wilson B-factor (Å^2)	181.2	Xtriage
Anisotropy	0.286	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 328.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	6984	wwPDB-VP
Average B, all atoms (Å^2)	289.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.33% 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	1.02	4/1643 (0.2%)	1.19	4/2238 (0.2%)
2	H	1.06	2/1607 (0.1%)	1.16	2/2196 (0.1%)
3	M	0.83	0/2184	0.95	2/2992 (0.1%)
4	A	0.81	0/1712	0.89	0/2315
All	All	0.93	6/7146 (0.1%)	1.05	8/9741 (0.1%)

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	H	133	VAL	C-O	9.78	1.42	1.23
1	L	185	GLU	CD-OE2	-8.94	1.15	1.25
1	L	2	ILE	N-CA	7.24	1.60	1.46
1	L	185	GLU	CD-OE1	-7.07	1.17	1.25
1	L	91	TYR	C-O	5.90	1.34	1.23

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	L	185	GLU	OE1-CD-OE2	-9.20	112.26	123.30
3	M	589	GLY	CA-C-O	8.65	136.18	120.60
2	H	67	ARG	NE-CZ-NH1	6.91	123.76	120.30
2	H	219	ARG	C-N-CA	6.07	135.05	122.30
1	L	189	HIS	CA-CB-CG	5.87	123.57	113.60

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	L	1605	0	1498	38	0
2	H	1571	0	1514	31	0
3	M	2136	0	2015	16	0
4	A	1672	0	1583	3	0
All	All	6984	0	6610	75	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 6.

The worst 5 of 75 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)
3:M:565:ARG:HB2	3:M:566:PRO:HD3	1.61	0.83
3:M:485:TYR:HB3	3:M:486:PRO:HD3	1.63	0.79
1:L:35:TRP:CZ3	1:L:88:CYS:HB3	2.19	0.77
1:L:136:LEU:HD12	1:L:136:LEU:N	1.99	0.77
1:L:96:LEU:HD13	2:H:106:PHE:HZ	1.54	0.72

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	210/213 (99%)	179 (85%)	24 (11%)	7 (3%)	4 30
2	H	213/218 (98%)	187 (88%)	22 (10%)	4 (2%)	8 41
3	M	286/327 (88%)	239 (84%)	41 (14%)	6 (2%)	7 39
4	A	215/234 (92%)	179 (83%)	32 (15%)	4 (2%)	8 41
All	All	924/992 (93%)	784 (85%)	119 (13%)	21 (2%)	6 37

5 of 21 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	L	101	GLY
1	L	184	ASP
2	H	219	ARG
3	M	417	PRO
1	L	30	GLY

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	179/188 (95%)	157 (88%)	22 (12%)	4 22
2	H	168/180 (93%)	150 (89%)	18 (11%)	6 27
3	M	217/286 (76%)	209 (96%)	8 (4%)	34 59
4	A	183/194 (94%)	180 (98%)	3 (2%)	62 79
All	All	747/848 (88%)	696 (93%)	51 (7%)	16 43

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	H	98	LYS
2	H	157	THR
4	A	47	TRP
2	H	99	ARG
2	H	146	CYS

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

Mol	Chain	Res	Type
3	M	522	GLN
3	M	554	HIS
4	A	152	ASN
2	H	77	ASN
2	H	82	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 monosaccharides 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

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	H	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	H	133:VAL	C	137:THR	N	7.86

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	212/213 (99%)	-0.20	3 (1%) 75 66	146, 222, 301, 360	0
2	H	217/218 (99%)	-0.29	0 100 100	134, 215, 303, 353	0
3	M	288/327 (88%)	-0.33	1 (0%) 94 90	188, 312, 426, 516	0
4	A	221/234 (94%)	0.41	20 (9%) 9 8	207, 373, 548, 639	0
All	All	938/992 (94%)	-0.12	24 (2%) 56 46	134, 273, 470, 639	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	A	108	GLY	6.0
4	A	159	GLY	5.4
4	A	157	SER	4.0
4	A	20	VAL	3.7
4	A	158	HIS	3.7

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 monosaccharides 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.