wwPDB X-ray Structure Validation Summary Report (i)

May 13, 2020 – 01:38 am BST

PDB ID	:	4WEU
Title	:	Co-complex structure of the F4 fimbrial adhesin FaeG variant ad with llama single domain antibody V3
Authors	:	Moonens, K.; Van den Broeck, I.; Pardon, E.; De Kerpel, M.; Remaut, H.; De Greve, H.
Deposited on	:	2014-09-11
Resolution	:	2.61 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:

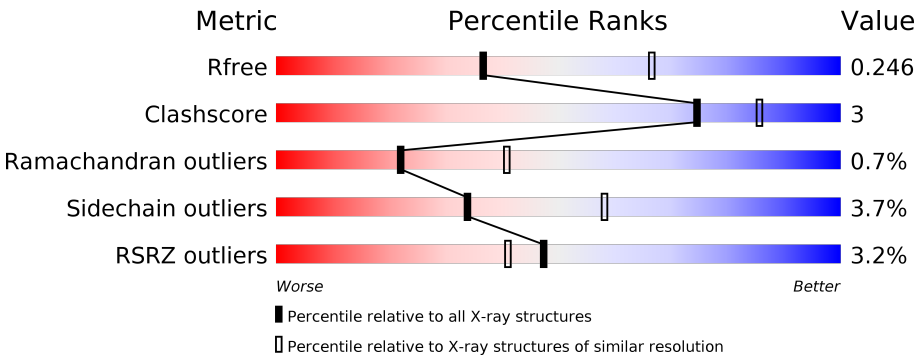
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.61 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	3797 (2.64-2.60)
Clashscore	141614	4168 (2.64-2.60)
Ramachandran outliers	138981	4093 (2.64-2.60)
Sidechain outliers	138945	4093 (2.64-2.60)
RSRZ outliers	127900	3731 (2.64-2.60)

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 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	D	127	87% 7% • 5%
1	E	127	% 85% 9% • 5%
2	A	277	3% 77% 11% 12%
2	B	277	5% 73% 7% 19%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 5316 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 Anti-F4+ETEC bacteria VHH variable region.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	D	121	926	587	157	178	4	0	0	0
1	E	121	930	589	158	179	4	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	922	HIS	-	expression tag	UNP R9W3F6
D	923	HIS	-	expression tag	UNP R9W3F6
D	924	HIS	-	expression tag	UNP R9W3F6
D	925	HIS	-	expression tag	UNP R9W3F6
D	926	HIS	-	expression tag	UNP R9W3F6
D	927	HIS	-	expression tag	UNP R9W3F6
E	922	HIS	-	expression tag	UNP R9W3F6
E	923	HIS	-	expression tag	UNP R9W3F6
E	924	HIS	-	expression tag	UNP R9W3F6
E	925	HIS	-	expression tag	UNP R9W3F6
E	926	HIS	-	expression tag	UNP R9W3F6
E	927	HIS	-	expression tag	UNP R9W3F6

- Molecule 2 is a protein called K88 fimbrial protein AD.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	245	1794	1128	304	359	3	0	0	0
2	B	223	1637	1032	279	324	2	0	0	0

There are 62 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	9	TRP	-	expression tag	UNP P14191
A	10	MET	-	expression tag	UNP P14191
A	11	THR	-	expression tag	UNP P14191
A	12	GLY	-	expression tag	UNP P14191
A	13	HIS	-	expression tag	UNP P14191
A	14	HIS	-	expression tag	UNP P14191
A	15	HIS	-	expression tag	UNP P14191
A	16	HIS	-	expression tag	UNP P14191
A	17	HIS	-	expression tag	UNP P14191
A	18	HIS	-	expression tag	UNP P14191
A	265	ASP	-	expression tag	UNP P14191
A	266	ASN	-	expression tag	UNP P14191
A	267	LYS	-	expression tag	UNP P14191
A	268	GLN	-	expression tag	UNP P14191
A	269	MET	-	expression tag	UNP P14191
A	270	THR	-	expression tag	UNP P14191
A	271	GLY	-	expression tag	UNP P14191
A	272	ASP	-	expression tag	UNP P14191
A	273	PHE	-	expression tag	UNP P14191
A	274	ASN	-	expression tag	UNP P14191
A	275	GLY	-	expression tag	UNP P14191
A	276	SER	-	expression tag	UNP P14191
A	277	VAL	-	expression tag	UNP P14191
A	278	ASP	-	expression tag	UNP P14191
A	279	ILE	-	expression tag	UNP P14191
A	280	GLY	-	expression tag	UNP P14191
A	281	GLY	-	expression tag	UNP P14191
A	282	SER	-	expression tag	UNP P14191
A	283	ILE	-	expression tag	UNP P14191
A	284	THR	-	expression tag	UNP P14191
A	285	ALA	-	expression tag	UNP P14191
B	9	TRP	-	expression tag	UNP P14191
B	10	MET	-	expression tag	UNP P14191
B	11	THR	-	expression tag	UNP P14191
B	12	GLY	-	expression tag	UNP P14191
B	13	HIS	-	expression tag	UNP P14191
B	14	HIS	-	expression tag	UNP P14191
B	15	HIS	-	expression tag	UNP P14191
B	16	HIS	-	expression tag	UNP P14191
B	17	HIS	-	expression tag	UNP P14191
B	18	HIS	-	expression tag	UNP P14191
B	265	ASP	-	expression tag	UNP P14191
B	266	ASN	-	expression tag	UNP P14191

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	267	LYS	-	expression tag	UNP P14191
B	268	GLN	-	expression tag	UNP P14191
B	269	MET	-	expression tag	UNP P14191
B	270	THR	-	expression tag	UNP P14191
B	271	GLY	-	expression tag	UNP P14191
B	272	ASP	-	expression tag	UNP P14191
B	273	PHE	-	expression tag	UNP P14191
B	274	ASN	-	expression tag	UNP P14191
B	275	GLY	-	expression tag	UNP P14191
B	276	SER	-	expression tag	UNP P14191
B	277	VAL	-	expression tag	UNP P14191
B	278	ASP	-	expression tag	UNP P14191
B	279	ILE	-	expression tag	UNP P14191
B	280	GLY	-	expression tag	UNP P14191
B	281	GLY	-	expression tag	UNP P14191
B	282	SER	-	expression tag	UNP P14191
B	283	ILE	-	expression tag	UNP P14191
B	284	THR	-	expression tag	UNP P14191
B	285	ALA	-	expression tag	UNP P14191

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	D	7	7	7	0	0
			Total	O		
3	A	13	13	13	0	0
			Total	O		
3	B	7	7	7	0	0
			Total	O		
3	E	2	2	2	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 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: Anti-F4+ETEC bacteria VHH variable region

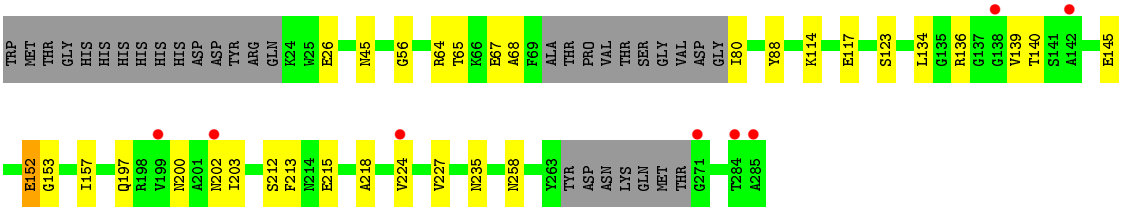
Chain D:

- Molecule 1: Anti-F4+ETEC bacteria VHH variable region

Chain E:

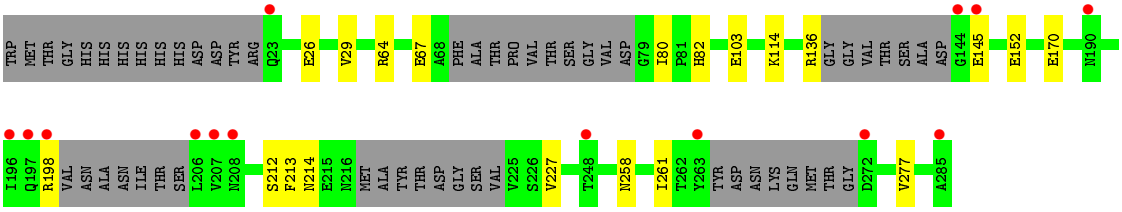
- Molecule 2: K88 fimbrial protein AD

Chain A:



- Molecule 2: K88 fimbrial protein AD

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, α, β, γ	79.80Å 95.20Å 113.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.54 – 2.61 29.54 – 2.61	Depositor EDS
% Data completeness (in resolution range)	98.3 (29.54-2.61) 98.4 (29.54-2.61)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.37 (at 2.61Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.204 , 0.243 0.208 , 0.246	Depositor DCC
R_{free} test set	1326 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	56.6	Xtriage
Anisotropy	0.077	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 35.0	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	0.94	EDS
Total number of atoms	5316	wwPDB-VP
Average B, all atoms (Å^2)	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.78% 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	D	0.67	0/948	0.82	0/1288
1	E	0.65	0/952	0.86	1/1293 (0.1%)
2	A	0.64	0/1823	0.78	0/2473
2	B	0.58	0/1661	0.76	1/2248 (0.0%)
All	All	0.63	0/5384	0.80	2/7302 (0.0%)

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	D	0	1
1	E	0	1
2	A	0	1
2	B	0	1
All	All	0	4

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	865	LYS	CA-CB-CG	6.75	128.25	113.40
2	B	214	ASN	N-CA-C	-5.07	97.31	111.00

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	A	213	PHE	Peptide
2	B	213	PHE	Peptide
1	D	854	THR	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	E	854	THR	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	D	926	0	891	6	0
1	E	930	0	897	6	0
2	A	1794	0	1763	16	0
2	B	1637	0	1606	9	0
3	A	13	0	0	1	0
3	B	7	0	0	0	0
3	D	7	0	0	1	0
3	E	2	0	0	0	0
All	All	5316	0	5157	33	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 3.

The worst 5 of 33 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:D:829:PHE:O	1:D:830:ASP:HB2	1.63	0.97
2:A:68:ALA:HB1	2:A:224:VAL:HG22	1.66	0.77
2:A:123:SER:OG	3:A:305:HOH:O	2.03	0.74
2:A:258:ASN:HD21	1:E:903:ASN:HD21	1.42	0.68
1:D:903:ASN:HD21	2:B:258:ASN:HD21	1.51	0.59

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	D	119/127 (94%)	118 (99%)	0	1 (1%)	19	36
1	E	119/127 (94%)	118 (99%)	0	1 (1%)	19	36
2	A	239/277 (86%)	228 (95%)	9 (4%)	2 (1%)	19	36
2	B	211/277 (76%)	203 (96%)	7 (3%)	1 (0%)	29	50
All	All	688/808 (85%)	667 (97%)	16 (2%)	5 (1%)	22	41

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	830	ASP
2	A	218	ALA
2	B	80	ILE
1	E	830	ASP
2	A	140	THR

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	D	94/101 (93%)	90 (96%)	4 (4%)	29	53
1	E	95/101 (94%)	91 (96%)	4 (4%)	30	53
2	A	187/215 (87%)	180 (96%)	7 (4%)	34	58
2	B	169/215 (79%)	164 (97%)	5 (3%)	41	66
All	All	545/632 (86%)	525 (96%)	20 (4%)	34	58

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	A	212	SER
2	A	227	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	E	857	SER
2	A	145	GLU
2	A	152	GLU

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

Mol	Chain	Res	Type
2	A	235	ASN
2	A	258	ASN
2	B	168	ASN
2	A	200	ASN
2	B	38	ASN

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)

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	D	121/127 (95%)	-0.42	0	100	100	37, 50, 72, 86	0
1	E	121/127 (95%)	-0.34	1 (0%)	86	84	40, 54, 79, 102	0
2	A	245/277 (88%)	-0.24	8 (3%)	46	40	37, 52, 91, 116	0
2	B	223/277 (80%)	-0.03	14 (6%)	20	15	39, 68, 119, 141	0
All	All	710/808 (87%)	-0.22	23 (3%)	47	41	37, 56, 96, 141	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	A	285	ALA	5.8
2	B	285	ALA	3.5
2	A	138	GLY	3.4
2	B	207	VAL	3.2
2	A	199	VAL	2.9

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.