wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 09:50 PM GMT

PDB ID	:	4WEN
Title	:	Co-complex structure of the F4 fimbrial adhesin FaeG variant ac with llama single domain antibody V2
Authors	:	Moonens, K.; Van den Broeck, I.; Pardon, E.; De Kerpel, M.; Remaut, H.; De Greve, H.
Deposited on	:	2014-09-10
Resolution	:	1.89 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 (i) symbol.

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

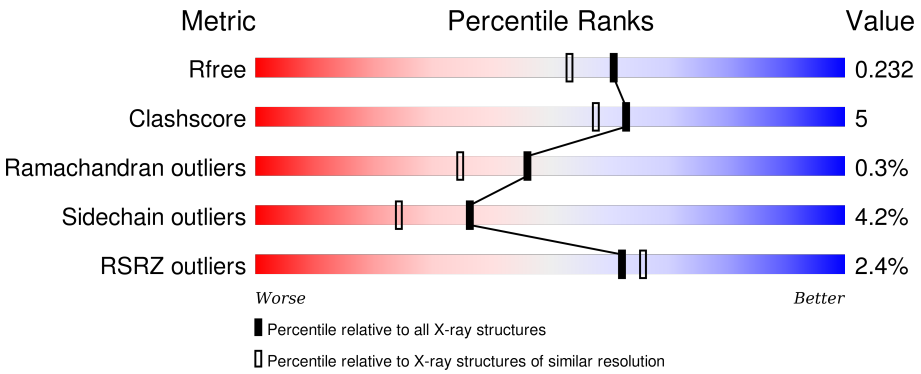
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 1.89 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	4755 (1.90-1.90)
Clashscore	102246	5398 (1.90-1.90)
Ramachandran outliers	100387	5338 (1.90-1.90)
Sidechain outliers	100360	5339 (1.90-1.90)
RSRZ outliers	91569	4766 (1.90-1.90)

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 <=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	275	2% 82% 9% • 8%
2	B	127	2% 83% 11% • 5%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2858 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 K88 fimbrial protein AC.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	253	1851	1164	317	367	3	0	0	0

There are 31 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	9	TRP	-	expression tag	UNP L7XD53
A	10	MET	-	expression tag	UNP L7XD53
A	11	THR	-	expression tag	UNP L7XD53
A	12	GLY	-	expression tag	UNP L7XD53
A	13	HIS	-	expression tag	UNP L7XD53
A	14	HIS	-	expression tag	UNP L7XD53
A	15	HIS	-	expression tag	UNP L7XD53
A	16	HIS	-	expression tag	UNP L7XD53
A	17	HIS	-	expression tag	UNP L7XD53
A	18	HIS	-	expression tag	UNP L7XD53
A	263	ASP	-	expression tag	UNP L7XD53
A	264	ASN	-	expression tag	UNP L7XD53
A	265	LYS	-	expression tag	UNP L7XD53
A	266	GLN	-	expression tag	UNP L7XD53
A	267	MET	-	expression tag	UNP L7XD53
A	268	THR	-	expression tag	UNP L7XD53
A	269	GLY	-	expression tag	UNP L7XD53
A	270	ASP	-	expression tag	UNP L7XD53
A	271	PHE	-	expression tag	UNP L7XD53
A	272	ASN	-	expression tag	UNP L7XD53
A	273	GLY	-	expression tag	UNP L7XD53
A	274	SER	-	expression tag	UNP L7XD53
A	275	VAL	-	expression tag	UNP L7XD53
A	276	ASP	-	expression tag	UNP L7XD53
A	277	ILE	-	expression tag	UNP L7XD53
A	278	GLY	-	expression tag	UNP L7XD53
A	279	GLY	-	expression tag	UNP L7XD53

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	280	SER	-	expression tag	UNP L7XD53
A	281	ILE	-	expression tag	UNP L7XD53
A	282	THR	-	expression tag	UNP L7XD53
A	283	ALA	-	expression tag	UNP L7XD53

- Molecule 2 is a protein called Anti-F4+ETEC bacteria VHH variable region.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	121	902	561	158	179	4	0	1	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	922	HIS	-	expression tag	UNP R9VYW2
B	923	HIS	-	expression tag	UNP R9VYW2
B	924	HIS	-	expression tag	UNP R9VYW2
B	925	HIS	-	expression tag	UNP R9VYW2
B	926	HIS	-	expression tag	UNP R9VYW2
B	927	HIS	-	expression tag	UNP R9VYW2

- Molecule 3 is water.

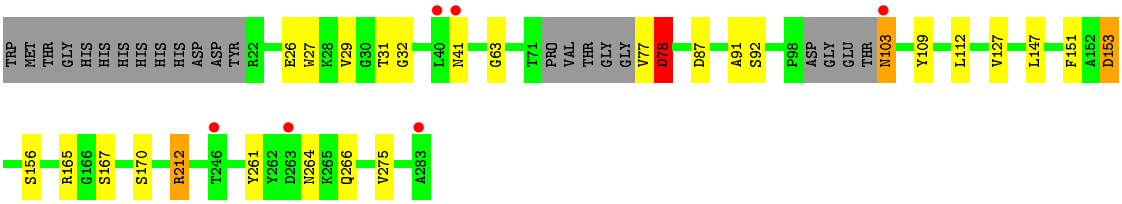
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	86	86	86	0	0
			Total	O		
3	B	19	19	19	0	0

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 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.

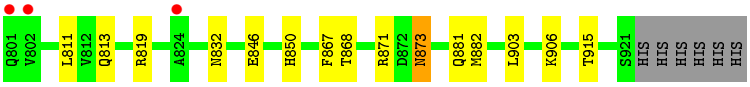
- Molecule 1: K88 fimbrial protein AC

Chain A:



- Molecule 2: Anti-F4+ETEC bacteria VHH variable region

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 3 2 1	Depositor
Cell constants a, b, c, α, β, γ	145.83Å 145.83Å 37.88Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	47.74 – 1.89 47.74 – 1.89	Depositor EDS
% Data completeness (in resolution range)	99.4 (47.74-1.89) 99.5 (47.74-1.89)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.41 (at 1.90Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.188 , 0.226 0.194 , 0.232	Depositor DCC
R_{free} test set	1837 reflections (5.27%)	DCC
Wilson B-factor (Å^2)	36.1	Xtriage
Anisotropy	0.045	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 46.8	EDS
Estimated twinning fraction	0.037 for -h,-k,l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Outliers	0 of 36708 reflections	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	2858	wwPDB-VP
Average B, all atoms (Å^2)	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.93% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	1.02	1/1879 (0.1%)	1.05	7/2549 (0.3%)
2	B	0.92	0/921	1.00	0/1249
All	All	0.99	1/2800 (0.0%)	1.04	7/3798 (0.2%)

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	A	0	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	212	ARG	CD-NE	-6.31	1.35	1.46

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	212	ARG	NE-CZ-NH2	-8.57	116.02	120.30
1	A	153	ASP	CB-CG-OD2	6.28	123.95	118.30
1	A	112	LEU	CA-CB-CG	5.91	128.89	115.30
1	A	147	LEU	CB-CG-CD1	-5.44	101.75	111.00
1	A	112	LEU	CB-CG-CD2	-5.42	101.79	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	103	ASN	Peptide
1	A	32	GLY	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	A	1851	0	1827	16	0
2	B	902	0	881	15	0
3	A	86	0	0	3	1
3	B	19	0	0	1	0
All	All	2858	0	2708	29	1

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 5.

The worst 5 of 29 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:871:ARG:HD2	2:B:873:ASN:ND2	1.93	0.83
1:A:153:ASP:OD1	1:A:212:ARG:HD3	1.85	0.75
1:A:31:THR:CG2	2:B:906:LYS:HG2	2.16	0.75
2:B:871:ARG:CD	2:B:873:ASN:ND2	2.60	0.65
1:A:103:ASN:N	3:A:319:HOH:O	2.32	0.63

All (1) 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)
3:A:308:HOH:O	3:A:308:HOH:O[5_556]	2.19	0.01

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	247/275 (90%)	236 (96%)	10 (4%)	1 (0%)	39	27
2	B	120/127 (94%)	115 (96%)	5 (4%)	0	100	100
All	All	367/402 (91%)	351 (96%)	15 (4%)	1 (0%)	46	35

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	78	ASP

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	191/212 (90%)	184 (96%)	7 (4%)	41	29
2	B	97/104 (93%)	92 (95%)	5 (5%)	29	17
All	All	288/316 (91%)	276 (96%)	12 (4%)	36	24

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	156	SER
1	A	167	SER
2	B	846	GLU
1	A	92	SER
2	B	813	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	233	ASN
2	B	881	GLN
1	A	272	ASN
1	A	198	ASN
1	A	256	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	A	253/275 (92%)	0.13	6 (2%)	62	66	26, 47, 81, 127	0
2	B	121/127 (95%)	0.12	3 (2%)	61	64	33, 52, 80, 97	0
All	All	374/402 (93%)	0.12	9 (2%)	62	66	26, 49, 81, 127	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	801	GLN	3.7
2	B	802	VAL	3.5
1	A	246	THR	3.2
2	B	824	ALA	2.9
1	A	283	ALA	2.6

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.