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)

Feb 15, 2017 – 01:23 am GMT

PDB ID	:	4J3O
Title	:	Crystal structure of the FimD usher traversed by the pilus tip complex assembly composed of FimC:FimF:FimG:FimH
Authors	:	Geibel, S.; Waksman, G.
Deposited on	:	2013-02-06
Resolution	:	3.80 Å(reported)

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

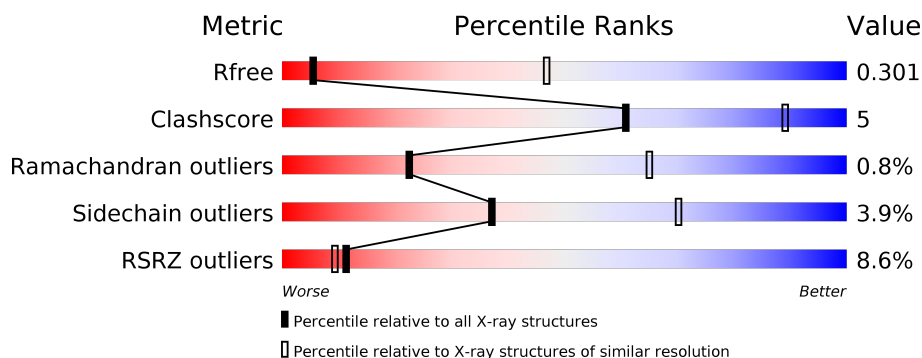|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

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 3.80 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		0.301
Clashscore		5
Ramachandran outliers		0.8%
Sidechain outliers		3.9%
RSRZ outliers		8.6%

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}	100719	1019 (4.08-3.52)
Clashscore	112137	1030 (4.04-3.56)
Ramachandran outliers	110173	1011 (4.06-3.54)
Sidechain outliers	110143	1005 (4.06-3.54)
RSRZ outliers	101464	1032 (4.08-3.52)

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	G	144	3% / 87% / 12% ·
2	H	279	6% / 83% / 16% ·
3	C	211	8% / 82% / 12% · 6%
4	F	154	16% / 83% / 16% ·
5	D	843	9% / 76% / 16% · 8%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 11732 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 Protein FimG.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	G	144	Total	C	N	O	S	0	0	0
			1043	640	176	224	3			

- Molecule 2 is a protein called Protein FimH.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	279	Total	C	N	O	S	0	0	0
			2051	1297	342	408	4			

- Molecule 3 is a protein called Chaperone protein FimC.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	199	Total	C	N	O	S	0	0	0
			1550	981	268	295	6			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	206	HIS	-	EXPRESSION TAG	UNP P31697
C	207	HIS	-	EXPRESSION TAG	UNP P31697
C	208	HIS	-	EXPRESSION TAG	UNP P31697
C	209	HIS	-	EXPRESSION TAG	UNP P31697
C	210	HIS	-	EXPRESSION TAG	UNP P31697
C	211	HIS	-	EXPRESSION TAG	UNP P31697

- Molecule 4 is a protein called Protein FimF.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	F	154	Total	C	N	O	S	0	0	0
			1137	711	196	226	4			

- Molecule 5 is a protein called Outer membrane usher protein FimD.

| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	D	778	5951	3714	1055	1163	19	0	0	0

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	348	PRO	THR	CONFLICT	UNP P30130
D	834	SER	-	EXPRESSION TAG	UNP P30130
D	835	ALA	-	EXPRESSION TAG	UNP P30130
D	836	TRP	-	EXPRESSION TAG	UNP P30130
D	837	SER	-	EXPRESSION TAG	UNP P30130
D	838	HIS	-	EXPRESSION TAG	UNP P30130
D	839	PRO	-	EXPRESSION TAG	UNP P30130
D	840	GLN	-	EXPRESSION TAG	UNP P30130
D	841	PHE	-	EXPRESSION TAG	UNP P30130
D	842	GLU	-	EXPRESSION TAG	UNP P30130
D	843	LYS	-	EXPRESSION TAG	UNP P30130

# 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 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.
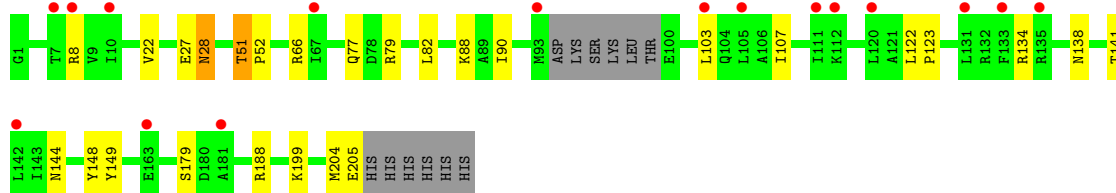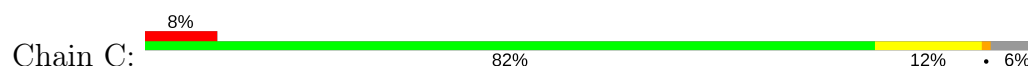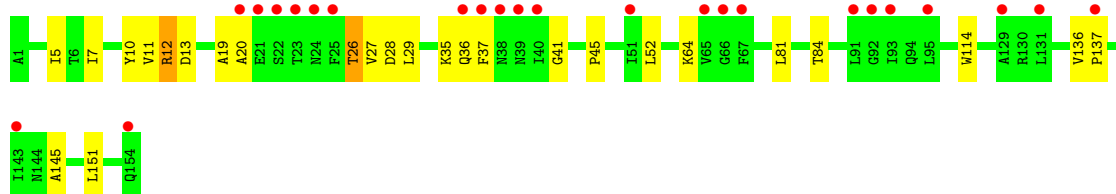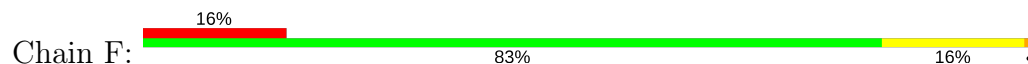
- Molecule 1: Protein FimG
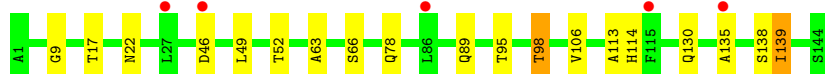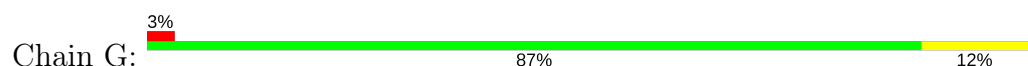
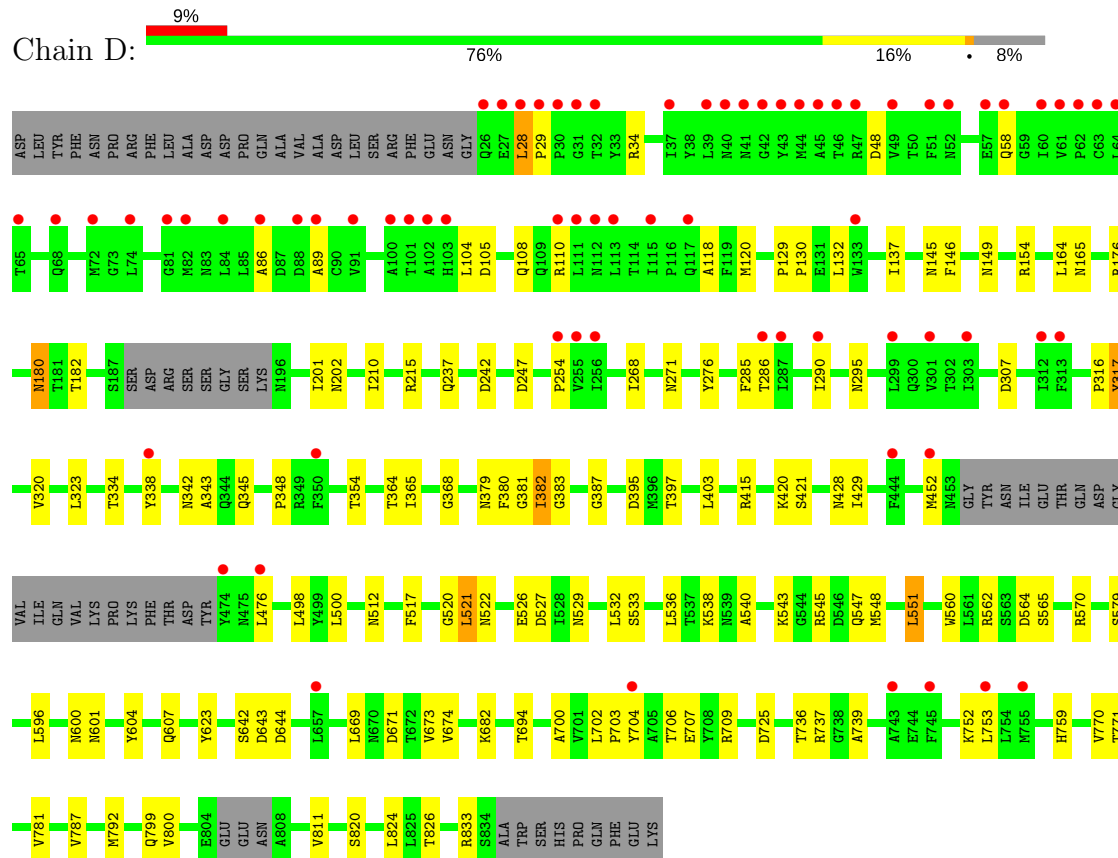

- Molecule 2: Protein FimH
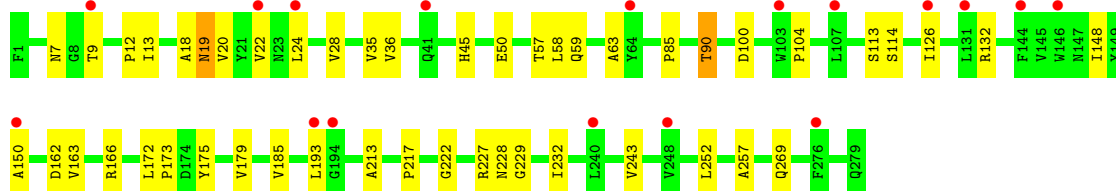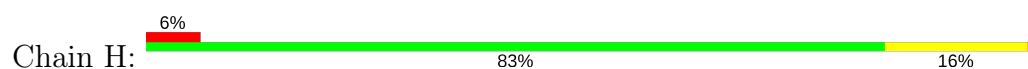
- Molecule 3: Chaperone protein FimC

- Molecule 4: Protein FimF

• Molecule 5: Outer membrane usher protein FimD

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 42 21 2	Depositor
Cell constants a, b, c, α, β, γ	122.36Å 122.36Å 328.46Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	83.67 – 3.80 83.67 – 3.80	Depositor EDS
% Data completeness (in resolution range)	99.2 (83.67-3.80) 86.0 (83.67-3.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.39 (at 3.78Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, R_{free}	0.248 , 0.299 0.249 , 0.301	Depositor DCC
R_{free} test set	1252 reflections (4.95%)	DCC
Wilson B-factor (Å^2)	159.4	Xtriage
Anisotropy	0.279	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 127.3	EDS
L-test for twinning[2]	$< |L| > = 0.42, < L^2 > = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	11732	wwPDB-VP
Average B, all atoms (Å^2)	194.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.03% 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	G	0.21	0/1057	0.39	0/1445
2	H	0.21	0/2096	0.38	0/2881
3	C	0.20	0/1578	0.38	0/2146
4	F	0.22	0/1157	0.40	0/1584
5	D	0.20	0/6079	0.36	0/8276
All	All	0.21	0/11967	0.37	0/16332

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 (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	G	1043	0	1009	12	0
2	H	2051	0	2007	26	0
3	C	1550	0	1582	13	0
4	F	1137	0	1126	13	0
5	D	5951	0	5656	66	0
All	All	11732	0	11380	120	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 5.

The worst 5 of 120 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)
5:D:562:ARG:HG2	5:D:564:ASP:H	1.53	0.72
5:D:165:ASN:HA	5:D:182:THR:HG22	1.74	0.69
2:H:12:PRO:HA	2:H:18:ALA:HB2	1.75	0.68
5:D:538:LYS:HD2	5:D:545:ARG:HG2	1.76	0.68
2:H:163:VAL:HG12	2:H:185:VAL:HG22	1.76	0.67

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	G	142/144 (99%)	130 (92%)	11 (8%)	1 (1%)	25	68
2	H	277/279 (99%)	250 (90%)	23 (8%)	4 (1%)	13	54
3	C	195/211 (92%)	181 (93%)	14 (7%)	0	100	100
4	F	152/154 (99%)	135 (89%)	15 (10%)	2 (1%)	14	56
5	D	770/843 (91%)	701 (91%)	64 (8%)	5 (1%)	28	70
All	All	1536/1631 (94%)	1397 (91%)	127 (8%)	12 (1%)	22	65

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	F	27	VAL
2	H	228	ASN
1	G	98	THR
2	H	7	ASN
5	D	452	MET

5.3.2 Protein sidechains ⓘ

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	G	115/115 (100%)	112 (97%)	3 (3%)	51	78
2	H	226/226 (100%)	221 (98%)	5 (2%)	57	81
3	C	170/182 (93%)	162 (95%)	8 (5%)	30	66
4	F	124/124 (100%)	117 (94%)	7 (6%)	25	62
5	D	621/694 (90%)	595 (96%)	26 (4%)	34	68
All	All	1256/1341 (94%)	1207 (96%)	49 (4%)	37	70

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	F	151	LEU
5	D	149	ASN
5	D	706	THR
5	D	105	ASP
5	D	180	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	G	78	GLN
2	H	59	GLN
5	D	295	ASN
5	D	428	ASN

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 ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

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, 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(Å²)	Q<0.9
1	G	144/144 (100%)	0.31	5 (3%)	44	35	119, 153, 181, 201	1 (0%)
2	H	279/279 (100%)	0.25	17 (6%)	22	16	118, 187, 225, 290	2 (0%)
3	C	199/211 (94%)	0.40	16 (8%)	13	10	138, 202, 249, 262	1 (0%)
4	F	154/154 (100%)	0.60	24 (15%)	2	3	136, 211, 304, 331	0
5	D	778/843 (92%)	0.36	72 (9%)	9	8	110, 180, 318, 369	2 (0%)
All	All	1554/1631 (95%)	0.36	134 (8%)	11	9	110, 184, 289, 369	6 (0%)

The worst 5 of 134 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	D	102	ALA	14.9
5	D	474	TYR	10.9
5	D	27	GLU	10.1
5	D	62	PRO	10.0
4	F	25	PHE	9.0

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 carbohydrates in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.