



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 12:53 PM EST

PDB ID : 5H5V
Title : Crystal structure of the flagellar cap protein FliD D1-D2-D3 domains from Escherichia coli
Authors : Song, W.S.; Cho, S.Y.; Hong, H.J.; Yoon, S.I.
Deposited on : 2016-11-09
Resolution : 3.00 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 ⓘ symbol.

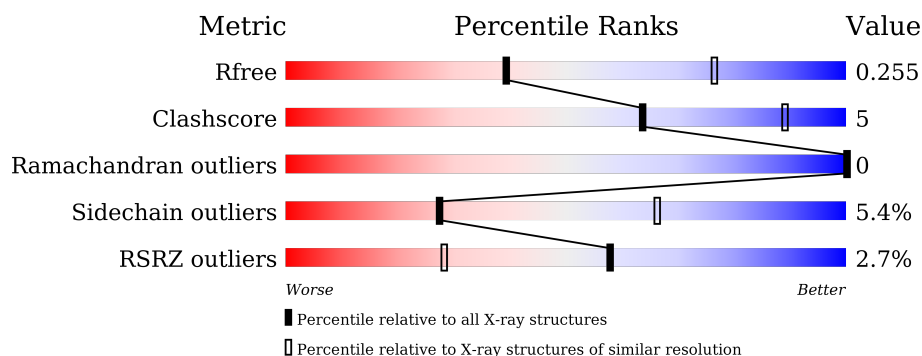
The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028442
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)	:	rb-20028442

i







X-RAY DIFFRACTION

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	1578 (3.00-3.00)
Clashscore	102246	1912 (3.00-3.00)
Ramachandran outliers	100387	1853 (3.00-3.00)
Sidechain outliers	100360	1856 (3.00-3.00)
RSRZ outliers	91569	1592 (3.00-3.00)

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 $\leq 5\%$. 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	380	 79%13% • 7%
1	B	380	 77%11% • 11%
1	C	380	 80%12% • 6%
1	D	380	 82%8% • 9%
1	E	380	 74%9% • • 21%
1	F	380	 76%9% • • 13%

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 13919 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 Flagellar hook-associated protein 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	352	Total	C	N	O	Se	0	0	0
			2497	1524	423	546	4			
1	B	339	Total	C	N	O	Se	0	0	0
			2345	1424	399	518	4			
1	C	356	Total	C	N	O	Se	0	0	0
			2513	1536	425	548	4			
1	D	345	Total	C	N	O	Se	0	0	0
			2363	1438	400	521	4			
1	E	299	Total	C	N	O	Se	0	0	0
			1921	1151	344	423	3			
1	F	330	Total	C	N	O	Se	0	0	0
			2280	1386	392	498	4			

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	37	GLY	-	expression tag	UNP A0A094VPA5
A	38	SER	-	expression tag	UNP A0A094VPA5
A	39	ALA	-	expression tag	UNP A0A094VPA5
A	40	LYS	-	expression tag	UNP A0A094VPA5
A	41	ASP	-	expression tag	UNP A0A094VPA5
A	42	LEU	-	expression tag	UNP A0A094VPA5
B	37	GLY	-	expression tag	UNP A0A094VPA5
B	38	SER	-	expression tag	UNP A0A094VPA5
B	39	ALA	-	expression tag	UNP A0A094VPA5
B	40	LYS	-	expression tag	UNP A0A094VPA5
B	41	ASP	-	expression tag	UNP A0A094VPA5
B	42	LEU	-	expression tag	UNP A0A094VPA5
C	37	GLY	-	expression tag	UNP A0A094VPA5
C	38	SER	-	expression tag	UNP A0A094VPA5
C	39	ALA	-	expression tag	UNP A0A094VPA5
C	40	LYS	-	expression tag	UNP A0A094VPA5
C	41	ASP	-	expression tag	UNP A0A094VPA5

Continued on next page...

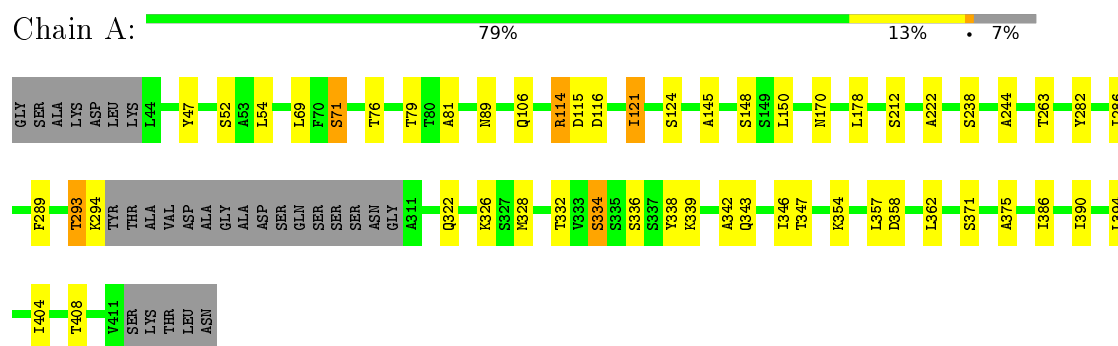
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	42	LEU	-	expression tag	UNP A0A094VPA5
D	37	GLY	-	expression tag	UNP A0A094VPA5
D	38	SER	-	expression tag	UNP A0A094VPA5
D	39	ALA	-	expression tag	UNP A0A094VPA5
D	40	LYS	-	expression tag	UNP A0A094VPA5
D	41	ASP	-	expression tag	UNP A0A094VPA5
D	42	LEU	-	expression tag	UNP A0A094VPA5
E	37	GLY	-	expression tag	UNP A0A094VPA5
E	38	SER	-	expression tag	UNP A0A094VPA5
E	39	ALA	-	expression tag	UNP A0A094VPA5
E	40	LYS	-	expression tag	UNP A0A094VPA5
E	41	ASP	-	expression tag	UNP A0A094VPA5
E	42	LEU	-	expression tag	UNP A0A094VPA5
F	37	GLY	-	expression tag	UNP A0A094VPA5
F	38	SER	-	expression tag	UNP A0A094VPA5
F	39	ALA	-	expression tag	UNP A0A094VPA5
F	40	LYS	-	expression tag	UNP A0A094VPA5
F	41	ASP	-	expression tag	UNP A0A094VPA5
F	42	LEU	-	expression tag	UNP A0A094VPA5

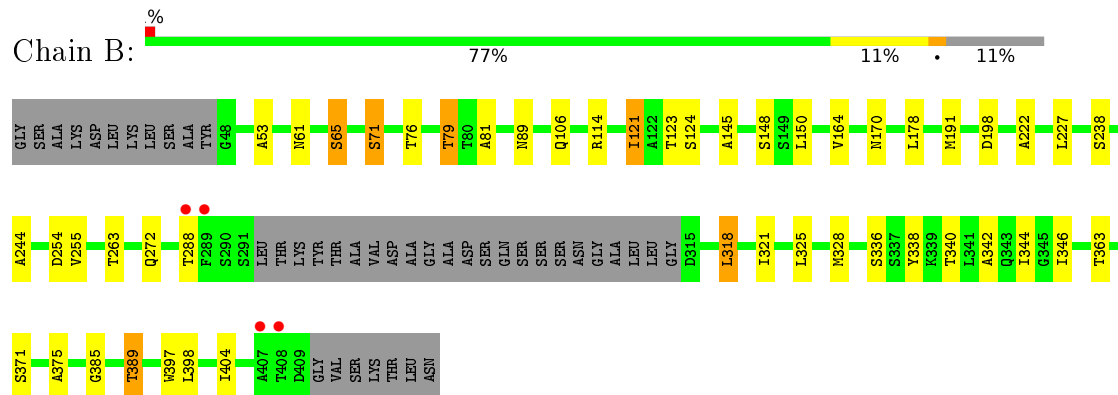
3 Residue-property plots [i](#)

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.

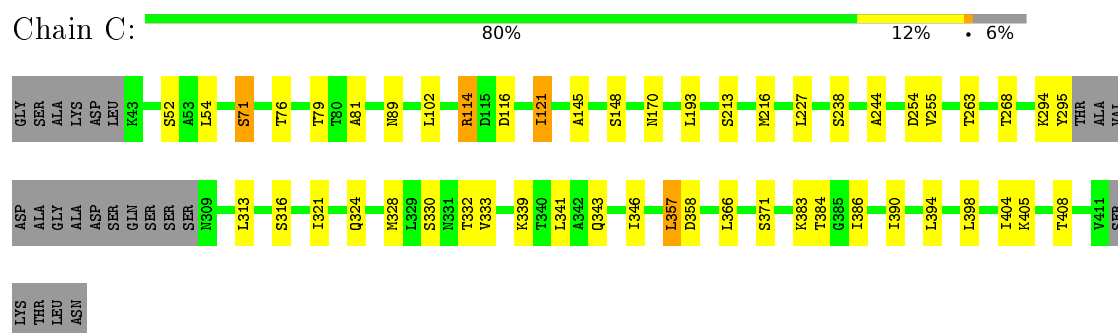
• Molecule 1: Flagellar hook-associated protein 2



• Molecule 1: Flagellar hook-associated protein 2



• Molecule 1: Flagellar hook-associated protein 2



- W397 T400 T401 GLY ILE ILE LYS ALA ALA THR ASP GLY VAL SER LYS THR LEU ASN

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	81.06Å 181.58Å 110.46Å 90.00° 105.23° 90.00°	Depositor
Resolution (Å)	30.00 – 3.00 29.79 – 2.99	Depositor EDS
% Data completeness (in resolution range)	99.7 (30.00-3.00) 99.2 (29.79-2.99)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.66 (at 3.00Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.228 , 0.258 0.227 , 0.255	Depositor DCC
R_{free} test set	3107 reflections (5.33%)	DCC
Wilson B-factor (Å ²)	62.1	Xtriage
Anisotropy	0.129	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 36.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	13919	wwPDB-VP
Average B, all atoms (Å ²)	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.90% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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	A	0.67	0/2509	0.74	0/3415
1	B	0.60	0/2356	0.70	0/3218
1	C	0.68	0/2526	0.74	0/3439
1	D	0.64	0/2375	0.71	0/3249
1	E	0.54	0/1922	0.66	0/2631
1	F	0.67	2/2290 (0.1%)	0.80	7/3127 (0.2%)
All	All	0.64	2/13978 (0.0%)	0.73	7/19079 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	216	MSE	CG-SE	-8.62	1.66	1.95
1	F	216	MSE	CB-CG	-7.43	1.30	1.52

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	F	216	MSE	CB-CG-SE	16.46	162.07	112.70
1	F	115	ASP	CB-CG-OD1	7.71	125.24	118.30
1	F	115	ASP	CB-CG-OD2	-6.60	112.36	118.30
1	F	188	ASP	CB-CG-OD1	5.92	123.63	118.30
1	F	198	ASP	CB-CG-OD1	-5.17	113.64	118.30

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	A	2497	0	2459	35	0
1	B	2345	0	2208	28	0
1	C	2513	0	2459	35	0
1	D	2363	0	2186	21	0
1	E	1921	0	1629	8	0
1	F	2280	0	2132	19	0
All	All	13919	0	13073	141	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 141 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:C:339:LYS:H	1:C:343:GLN:HE22	1.22	0.88
1:C:54:LEU:HD11	1:C:321:ILE:HD13	1.68	0.75
1:F:321:ILE:HG23	1:F:397:TRP:CZ3	2.23	0.73
1:D:145:ALA:O	1:D:148:SER:HB3	1.91	0.71
1:F:325:LEU:HD23	1:F:328:MSE:HE3	1.72	0.71

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	A	348/380 (92%)	335 (96%)	13 (4%)	0	100	100
1	B	335/380 (88%)	326 (97%)	9 (3%)	0	100	100
1	C	352/380 (93%)	342 (97%)	10 (3%)	0	100	100
1	D	341/380 (90%)	333 (98%)	8 (2%)	0	100	100
1	E	293/380 (77%)	287 (98%)	6 (2%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	F	326/380 (86%)	318 (98%)	8 (2%)	0	100	100
All	All	1995/2280 (88%)	1941 (97%)	54 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	A	272/302 (90%)	258 (95%)	14 (5%)	29	69
1	B	242/302 (80%)	228 (94%)	14 (6%)	25	63
1	C	270/302 (89%)	255 (94%)	15 (6%)	26	65
1	D	235/302 (78%)	224 (95%)	11 (5%)	32	72
1	E	161/302 (53%)	152 (94%)	9 (6%)	26	65
1	F	229/302 (76%)	216 (94%)	13 (6%)	25	64
All	All	1409/1812 (78%)	1333 (95%)	76 (5%)	27	66

5 of 76 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	238	SER
1	C	371	SER
1	F	188	ASP
1	C	255	VAL
1	C	324	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	343	GLN
1	D	272	GLN
1	F	104	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	104	GLN
1	E	104	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](#)

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	348/380 (91%)	-0.52	0 100 100	34, 51, 83, 104	0
1	B	335/380 (88%)	-0.36	4 (1%) 81 55	32, 60, 134, 197	0
1	C	352/380 (92%)	-0.51	0 100 100	33, 48, 87, 111	0
1	D	341/380 (89%)	-0.23	8 (2%) 64 33	27, 62, 149, 188	0
1	E	295/380 (77%)	0.31	36 (12%) 5 2	37, 77, 289, 348	0
1	F	326/380 (85%)	-0.26	6 (1%) 71 43	34, 66, 145, 189	0
All	All	1997/2280 (87%)	-0.28	54 (2%) 58 28	27, 57, 223, 348	0

The worst 5 of 54 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	278	TRP	5.6
1	D	46	ALA	5.4
1	E	406	ALA	4.7
1	E	402	GLY	4.6
1	E	399	SER	4.6

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.