



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 8, 2017 – 06:46 AM EDT

PDB ID : 3F85  
Title : Structure of fusion complex of homo trimeric major pilin subunits CfaB of CFA/I fimbriae from ETEC E. coli  
Authors : Xia, D.; Li, Y.F.  
Deposited on : unknown  
Resolution : 2.10 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto: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  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20029824  
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) : rb-20029824

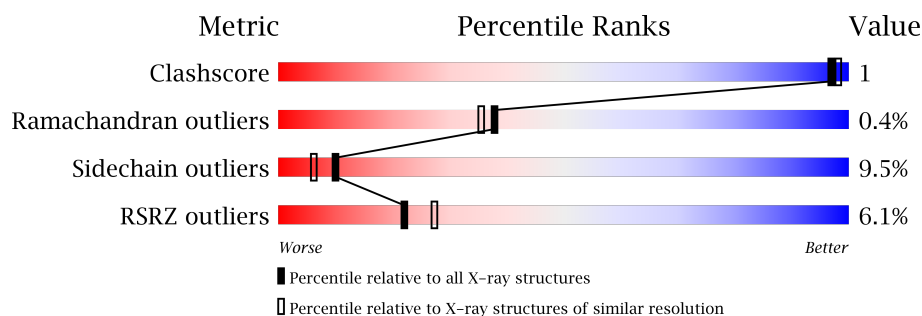
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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))
Clashscore	112137	4788 (2.10-2.10)
Ramachandran outliers	110173	4740 (2.10-2.10)
Sidechain outliers	110143	4741 (2.10-2.10)
RSRZ outliers	101464	4275 (2.10-2.10)

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  $\geq 3$ , 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	463	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	MG	A	464	-	-	-	X

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3400 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 homo trimeric fusion of CFA/I fimbrial subunits B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	456	Total	C	N	O	S	0	0	0
			3288	2055	543	681	9			

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	INITIATING METHIONINE	UNP P02971
A	135	ASP	-	LINKER	UNP P02971
A	136	ASN	-	LINKER	UNP P02971
A	137	LYS	-	LINKER	UNP P02971
A	138	GLN	-	LINKER	UNP P02971
A	286	ASP	-	LINKER	UNP P02971
A	287	ASN	-	LINKER	UNP P02971
A	288	LYS	-	LINKER	UNP P02971
A	289	GLN	-	LINKER	UNP P02971
A	437	ASP	-	LINKER	UNP P02971
A	438	ASN	-	LINKER	UNP P02971
A	439	LYS	-	LINKER	UNP P02971
A	440	GLN	-	LINKER	UNP P02971
A	456	LEU	-	EXPRESSION TAG	UNP P02971
A	457	GLU	-	EXPRESSION TAG	UNP P02971
A	458	HIS	-	EXPRESSION TAG	UNP P02971
A	459	HIS	-	EXPRESSION TAG	UNP P02971
A	460	HIS	-	EXPRESSION TAG	UNP P02971
A	461	HIS	-	EXPRESSION TAG	UNP P02971
A	462	HIS	-	EXPRESSION TAG	UNP P02971
A	463	HIS	-	EXPRESSION TAG	UNP P02971

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	4	Total 4	Mg 4	0	0

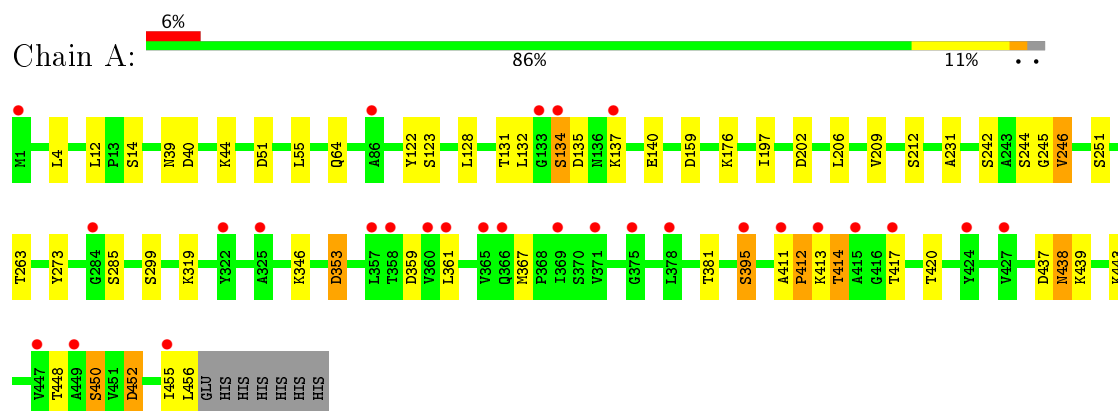
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	108	Total 108	O 108	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 ( $\text{RSRZ} > 2$ ). 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: homo trimeric fusion of CFA/I fimbrial subunits B



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	127.41 Å   44.98 Å   97.88 Å 90.00°   125.35°   90.00°	Depositor
Resolution (Å)	20.00 – 2.10 35.22 – 2.10	Depositor EDS
% Data completeness (in resolution range)	93.4 (20.00-2.10) 83.7 (35.22-2.10)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.83 (at 2.10 Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.233   ,   0.283 0.236   ,   (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	40.9	Xtriage
Anisotropy	0.660	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 53.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	3400	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.43% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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

Bond lengths and bond angles in the following residue types are not validated in this section: MG

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	0.89	2/3335 (0.1%)	0.81	7/4559 (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	8

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	209	VAL	CB-CG1	-5.91	1.40	1.52
1	A	246	VAL	CA-CB	5.28	1.65	1.54

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	159	ASP	CB-CG-OD2	8.60	126.04	118.30
1	A	437	ASP	CB-CG-OD2	7.58	125.13	118.30
1	A	40	ASP	CB-CG-OD2	6.82	124.44	118.30
1	A	51	ASP	CB-CG-OD2	6.41	124.07	118.30
1	A	353	ASP	CB-CG-OD2	5.57	123.31	118.30
1	A	452	ASP	CB-CG-OD2	5.16	122.95	118.30
1	A	359	ASP	CB-CG-OD2	5.11	122.90	118.30

There are no chirality outliers.

All (8) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	132	LEU	Peptide
1	A	134	SER	Peptide
1	A	245	GLY	Peptide
1	A	246	VAL	Peptide
1	A	395	SER	Peptide
1	A	411	ALA	Peptide
1	A	414	THR	Peptide
1	A	450	SER	Peptide

## 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	3288	0	3323	7	0
2	A	4	0	0	0	0
3	A	108	0	0	0	0
All	All	3400	0	3323	7	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 1.

All (7) 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:A:206:LEU:HG	1:A:273:TYR:HB3	1.88	0.54
1:A:4:LEU:HD23	1:A:12:LEU:HD11	1.91	0.52
1:A:55:LEU:HB3	1:A:122:TYR:HB3	1.95	0.48
1:A:367:MET:HG2	1:A:412:PRO:HB3	1.99	0.45
1:A:128:LEU:O	1:A:140:GLU:HA	2.18	0.43
1:A:197:ILE:HG23	1:A:231:ALA:HB1	2.00	0.43
1:A:438:ASN:HA	1:A:438:ASN:HD22	1.74	0.41

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	454/463 (98%)	436 (96%)	16 (4%)	2 (0%)	38	35

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	412	PRO
1	A	212	SER

### 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	368/379 (97%)	333 (90%)	35 (10%)	10	6

All (35) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	14	SER
1	A	39	ASN
1	A	44	LYS
1	A	64	GLN
1	A	123	SER
1	A	131	THR
1	A	134	SER
1	A	135	ASP
1	A	137	LYS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	176	LYS
1	A	202	ASP
1	A	242	SER
1	A	244	SER
1	A	251	SER
1	A	263	THR
1	A	285	SER
1	A	299	SER
1	A	319	LYS
1	A	346	LYS
1	A	353	ASP
1	A	361	LEU
1	A	381	THR
1	A	395	SER
1	A	413	LYS
1	A	414	THR
1	A	417	THR
1	A	420	THR
1	A	438	ASN
1	A	439	LYS
1	A	443	LYS
1	A	448	THR
1	A	450	SER
1	A	452	ASP
1	A	455	ILE
1	A	456	LEU

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

Mol	Chain	Res	Type
1	A	96	ASN
1	A	161	ASN
1	A	272	ASN
1	A	438	ASN

### 5.3.3 RNA ⓘ

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](#)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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 ⓘ

### 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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	456/463 (98%)	0.49	28 (6%)	22 27	27, 41, 52, 74	0

All (28) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	134	SER	5.7
1	A	284	GLY	5.1
1	A	360	VAL	4.6
1	A	427	VAL	4.6
1	A	449	ALA	4.6
1	A	369	ILE	4.6
1	A	365	VAL	4.0
1	A	413	LYS	3.8
1	A	411	ALA	3.6
1	A	424	TYR	3.5
1	A	378	LEU	3.1
1	A	395	SER	3.0
1	A	366	GLN	2.8
1	A	133	GLY	2.7
1	A	358	THR	2.6
1	A	137	LYS	2.5
1	A	361	LEU	2.5
1	A	371	VAL	2.4
1	A	375	GLY	2.4
1	A	1	MET	2.4
1	A	86	ALA	2.4
1	A	447	VAL	2.4
1	A	322	TYR	2.3
1	A	325	ALA	2.3
1	A	357	LEU	2.2
1	A	417	THR	2.2
1	A	455	ILE	2.0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	415	ALA	2.0

## 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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	MG	A	464	1/1	0.46	0.40	5.42	103,103,103,103	0
2	MG	A	467	1/1	0.97	0.29	-	59,59,59,59	0
2	MG	A	465	1/1	0.93	0.35	-	71,71,71,71	0
2	MG	A	466	1/1	0.94	0.20	-	77,77,77,77	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.