WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 22, 2023 – 12:12 PM JST

PDB ID	:	8IEV
Title	:	Crystal structure of the DUF2891 family protein CJ0554 from Campylobacter jejuni in space group C2
Authors	:	Kim, S.Y.; Cho, H.Y.; Yoon, S.I.
Deposited on	:	2023-02-16
Resolution	:	2.08 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.33
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)	:	2.33

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 2.08 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.
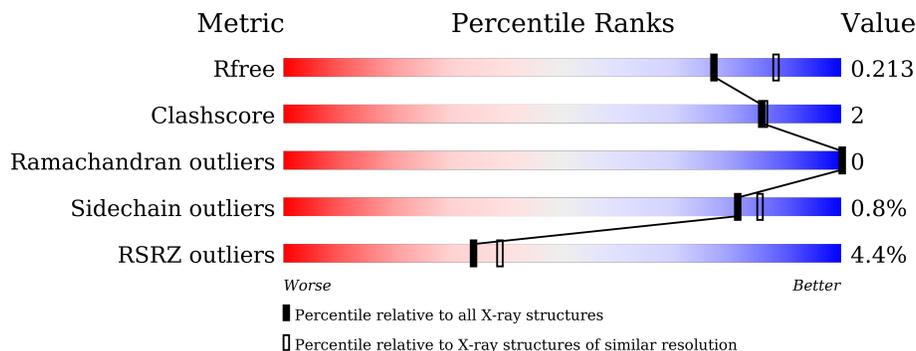
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	6189 (2.10-2.06)
Clashscore	141614	6738 (2.10-2.06)
Ramachandran outliers	138981	6663 (2.10-2.06)
Sidechain outliers	138945	6664 (2.10-2.06)
RSRZ outliers	127900	6057 (2.10-2.06)

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 of 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	A	340	94%
1	B	340	93% 5%
1	C	340	% 92% 6%
1	D	340	4% 91% 7%
1	E	340	3% 91% 7%
1	F	340	17% 91% 6%

2   Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 16668 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 DUF2891 domain-containing protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	335	Total 2713	C 1785	N 431	O 488	S 9	0	0	0
1	B	335	Total 2725	C 1792	N 436	O 488	S 9	0	0	0
1	C	335	Total 2691	C 1771	N 431	O 480	S 9	0	0	0
1	D	335	Total 2675	C 1759	N 429	O 478	S 9	0	0	0
1	E	335	Total 2685	C 1768	N 431	O 477	S 9	0	0	0
1	F	332	Total 2600	C 1715	N 416	O 461	S 8	0	0	0

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	GLY	-	expression tag	UNP A0A2U0QSR5
A	-4	SER	-	expression tag	UNP A0A2U0QSR5
A	-3	ALA	-	expression tag	UNP A0A2U0QSR5
A	-2	LYS	-	expression tag	UNP A0A2U0QSR5
A	-1	ASP	-	expression tag	UNP A0A2U0QSR5
A	0	PRO	-	expression tag	UNP A0A2U0QSR5
A	200	SER	ASN	engineered mutation	UNP A0A2U0QSR5
B	-5	GLY	-	expression tag	UNP A0A2U0QSR5
B	-4	SER	-	expression tag	UNP A0A2U0QSR5
B	-3	ALA	-	expression tag	UNP A0A2U0QSR5
B	-2	LYS	-	expression tag	UNP A0A2U0QSR5
B	-1	ASP	-	expression tag	UNP A0A2U0QSR5
B	0	PRO	-	expression tag	UNP A0A2U0QSR5
B	200	SER	ASN	engineered mutation	UNP A0A2U0QSR5
C	-5	GLY	-	expression tag	UNP A0A2U0QSR5
C	-4	SER	-	expression tag	UNP A0A2U0QSR5
C	-3	ALA	-	expression tag	UNP A0A2U0QSR5

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	-2	LYS	-	expression tag	UNP A0A2U0QSR5
C	-1	ASP	-	expression tag	UNP A0A2U0QSR5
C	0	PRO	-	expression tag	UNP A0A2U0QSR5
C	200	SER	ASN	engineered mutation	UNP A0A2U0QSR5
D	-5	GLY	-	expression tag	UNP A0A2U0QSR5
D	-4	SER	-	expression tag	UNP A0A2U0QSR5
D	-3	ALA	-	expression tag	UNP A0A2U0QSR5
D	-2	LYS	-	expression tag	UNP A0A2U0QSR5
D	-1	ASP	-	expression tag	UNP A0A2U0QSR5
D	0	PRO	-	expression tag	UNP A0A2U0QSR5
D	200	SER	ASN	engineered mutation	UNP A0A2U0QSR5
E	-5	GLY	-	expression tag	UNP A0A2U0QSR5
E	-4	SER	-	expression tag	UNP A0A2U0QSR5
E	-3	ALA	-	expression tag	UNP A0A2U0QSR5
E	-2	LYS	-	expression tag	UNP A0A2U0QSR5
E	-1	ASP	-	expression tag	UNP A0A2U0QSR5
E	0	PRO	-	expression tag	UNP A0A2U0QSR5
E	200	SER	ASN	engineered mutation	UNP A0A2U0QSR5
F	-5	GLY	-	expression tag	UNP A0A2U0QSR5
F	-4	SER	-	expression tag	UNP A0A2U0QSR5
F	-3	ALA	-	expression tag	UNP A0A2U0QSR5
F	-2	LYS	-	expression tag	UNP A0A2U0QSR5
F	-1	ASP	-	expression tag	UNP A0A2U0QSR5
F	0	PRO	-	expression tag	UNP A0A2U0QSR5
F	200	SER	ASN	engineered mutation	UNP A0A2U0QSR5

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	151	Total 151	O 151	0	0
2	B	149	Total 149	O 149	0	0
2	C	101	Total 101	O 101	0	0
2	D	71	Total 71	O 71	0	0
2	E	84	Total 84	O 84	0	0
2	F	23	Total 23	O 23	0	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: DUF2891 domain-containing protein

Chain A:
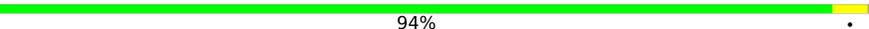
• Molecule 1: DUF2891 domain-containing protein

Chain B:

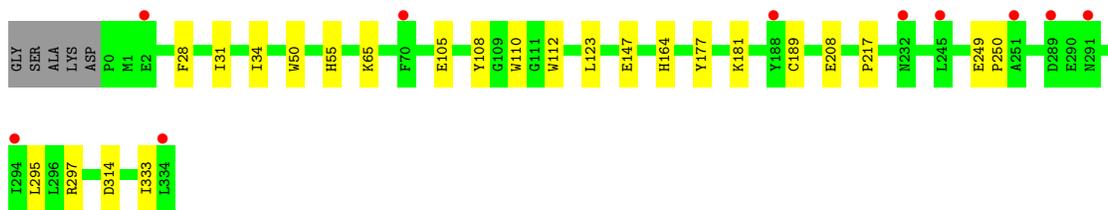• Molecule 1: DUF2891 domain-containing protein

Chain C:

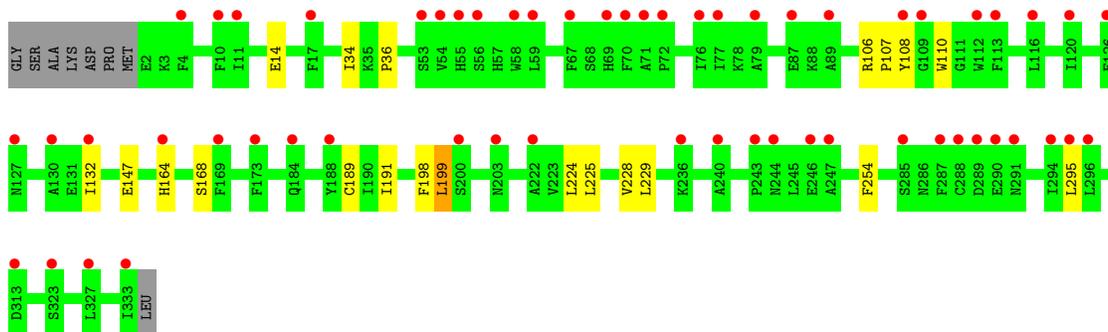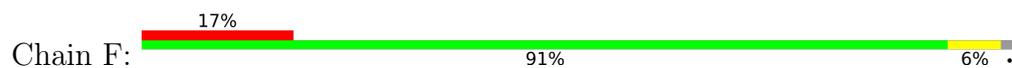• Molecule 1: DUF2891 domain-containing protein

Chain D:

• Molecule 1: DUF2891 domain-containing protein

Chain E:

GLY SER ALA LYS ASP P0 M1 E2 F28 I31 I34 W50 H55 K65 F70 E105 Y108 G109 W110 G111 W112 L123 E147 H164 Y177 K181 Y188 C189 E208 P217 N232 L245 E249 P250 A251 D289 E290 N291

I294 L295 L296 R297 D314 I333 L334

• Molecule 1: DUF2891 domain-containing protein

Chain F:

17%

91%   6%   •

GLY SER ALA LYS ASP PRO MET E2 K3 F4 F10 I11 E14 F17 I34 K35 P36 S53 V54 H55 S56 H57 W58 L59 F67 S68 H69 F70 A71 P72 I76 I77 K78 A79 E87 K88 A89 R106 P107 Y108 G109 W110 G111 W112 F113 L116 I120 E126

N127 A130 E131 I132 E147 H164 S168 F169 F173 Q184 Y188 C189 I190 I191 F198 L199 S200 N203 A222 V223 L224 L225 V228 L229 K236 A240 P243 N244 L245 E246 A247 F254 S285 N286 F287 C288 D289 E290 N291 I294 L295 L296

D313 S323 L327 I333 LEU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	306.17Å 62.42Å 134.88Å 90.00° 112.45° 90.00°	Depositor
Resolution (Å)	29.67 – 2.08 29.67 – 2.08	Depositor EDS
% Data completeness (in resolution range)	99.6 (29.67-2.08) 99.2 (29.67-2.08)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.96 (at 2.08Å)	Xtriage
Refinement program	PHENIX 1.15.2_3472	Depositor
R, R_{free}	0.178 , 0.213 0.178 , 0.213	Depositor DCC
R_{free} test set	6962 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å^2)	25.0	Xtriage
Anisotropy	0.224	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 52.8	EDS
L-test for twinning[2]	$< \lvert L \rvert > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	16668	wwPDB-VP
Average B, all atoms (Å^2)	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 27.90 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.0265e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< \lvert L \rvert >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality ⓘ

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	0.46	0/2794	0.55	0/3787
1	B	0.47	0/2807	0.56	0/3802
1	C	0.40	0/2772	0.53	0/3761
1	D	0.37	0/2756	0.50	0/3742
1	E	0.39	0/2766	0.52	0/3754
1	F	0.33	0/2680	0.47	0/3652
All	All	0.41	0/16575	0.52	0/22498

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 ⓘ

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	2713	0	2613	8	0
1	B	2725	0	2641	13	0
1	C	2691	0	2585	13	0
1	D	2675	0	2545	14	0
1	E	2685	0	2576	14	0
1	F	2600	0	2421	13	0
2	A	151	0	0	0	0
2	B	149	0	0	1	0
2	C	101	0	0	2	0
2	D	71	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	84	0	0	0	0
2	F	23	0	0	0	0
All	All	16668	0	15381	68	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 2.

The worst 5 of 68 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:E:31:ILE:HD11	1:F:36:PRO:HB3	1.79	0.62
1:C:149:LYS:HE2	1:C:185:GLU:HG2	1.82	0.61
1:D:159:ILE:HD12	2:D:430:HOH:O	2.01	0.60
1:A:154:LYS:HE2	1:B:261:ASP:OD2	2.03	0.58
1:B:232:ASN:HD22	1:B:232:ASN:H	1.51	0.57

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	A	333/340 (98%)	325 (98%)	8 (2%)	0	100	100
1	B	333/340 (98%)	325 (98%)	8 (2%)	0	100	100
1	C	333/340 (98%)	326 (98%)	7 (2%)	0	100	100
1	D	333/340 (98%)	323 (97%)	10 (3%)	0	100	100
1	E	333/340 (98%)	325 (98%)	8 (2%)	0	100	100
1	F	330/340 (97%)	322 (98%)	8 (2%)	0	100	100
All	All	1995/2040 (98%)	1946 (98%)	49 (2%)	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	283/300 (94%)	281 (99%)	2 (1%)	84	87
1	B	287/300 (96%)	286 (100%)	1 (0%)	92	95
1	C	279/300 (93%)	278 (100%)	1 (0%)	91	93
1	D	274/300 (91%)	271 (99%)	3 (1%)	73	78
1	E	277/300 (92%)	274 (99%)	3 (1%)	73	78
1	F	257/300 (86%)	253 (98%)	4 (2%)	62	67
All	All	1657/1800 (92%)	1643 (99%)	14 (1%)	81	85

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	189	CYS
1	E	295	LEU
1	F	199	LEU
1	F	132	ILE
1	F	189	CYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	184	GLN
1	B	232	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 monosaccharides 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	A	335/340 (98%)	-0.39	1 (0%)	94	94	14, 24, 41, 63	0
1	B	335/340 (98%)	-0.36	1 (0%)	94	94	14, 25, 43, 63	0
1	C	335/340 (98%)	-0.17	4 (1%)	79	81	16, 32, 54, 77	0
1	D	335/340 (98%)	0.13	15 (4%)	33	37	21, 40, 64, 88	0
1	E	335/340 (98%)	-0.07	10 (2%)	50	55	18, 35, 58, 81	0
1	F	332/340 (97%)	0.92	57 (17%)	1	1	34, 60, 84, 106	0
All	All	2007/2040 (98%)	0.01	88 (4%)	34	39	14, 34, 69, 106	0

The worst 5 of 88 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	70	PHE	6.6
1	F	116	LEU	5.3
1	F	112	TRP	4.7
1	F	294	ILE	4.6
1	F	295	LEU	4.4

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 monosaccharides 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.