wwPDB X-ray Structure Validation Summary Report (i)

May 22, 2020 – 08:27 pm BST

PDB ID	:	5MBQ
Title	:	CeuE (H227A variant) a periplasmic protein from Campylobacter jejuni
Authors	:	Wilde, E.J.; Blagova, E.V.; Hughes, A.; Raines, D.J.; Moroz, O.V.; Turkenburg, J.P.; Duhme-Klair, A.-K.; Wilson, K.S.
Deposited on	:	2016-11-08
Resolution	:	1.33 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 following versions of software and data (see references (1)) were used in the production of this report:

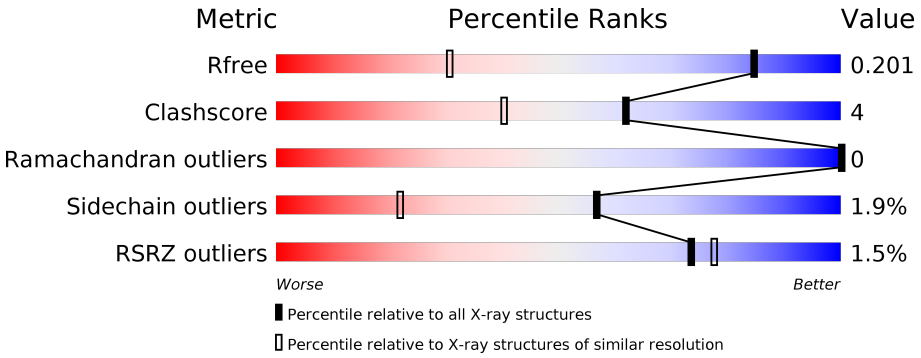
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 1.33 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	1385 (1.36-1.32)
Clashscore	141614	1417 (1.36-1.32)
Ramachandran outliers	138981	1397 (1.36-1.32)
Sidechain outliers	138945	1397 (1.36-1.32)
RSRZ outliers	127900	1369 (1.36-1.32)

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 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	291	2% 82% 14%
1	B	291	2% 81% 17%
1	C	291	78% 18%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 7062 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 Enterochelin uptake periplasmic binding protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	286	Total	C	N	O	S	0	10	0
			2189	1417	347	422	3			
1	B	287	Total	C	N	O	S	0	11	1
			2194	1423	346	422	3			
1	C	282	Total	C	N	O	S	0	5	0
			2156	1399	344	410	3			

There are 15 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	20	GLY	-	expression tag	UNP Q0P8Q4
A	21	PRO	-	expression tag	UNP Q0P8Q4
A	22	ALA	-	expression tag	UNP Q0P8Q4
A	23	MET	-	expression tag	UNP Q0P8Q4
A	227	ALA	HIS	engineered mutation	UNP Q0P8Q4
B	20	GLY	-	expression tag	UNP Q0P8Q4
B	21	PRO	-	expression tag	UNP Q0P8Q4
B	22	ALA	-	expression tag	UNP Q0P8Q4
B	23	MET	-	expression tag	UNP Q0P8Q4
B	227	ALA	HIS	engineered mutation	UNP Q0P8Q4
C	20	GLY	-	expression tag	UNP Q0P8Q4
C	21	PRO	-	expression tag	UNP Q0P8Q4
C	22	ALA	-	expression tag	UNP Q0P8Q4
C	23	MET	-	expression tag	UNP Q0P8Q4
C	227	ALA	HIS	engineered mutation	UNP Q0P8Q4

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	199	Total	O	0	0
			199	199		
2	B	181	Total	O	0	0
			181	181		

Continued on next page...

Continued from previous page...

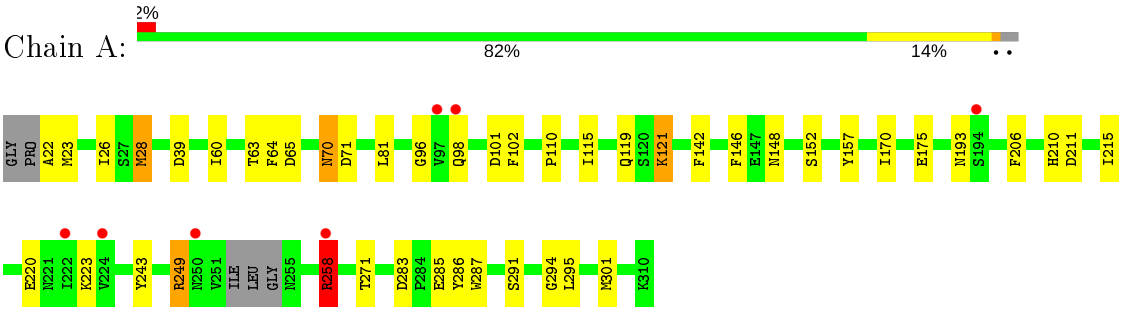
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	C	143	Total	O	0	0
			143	143		

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.

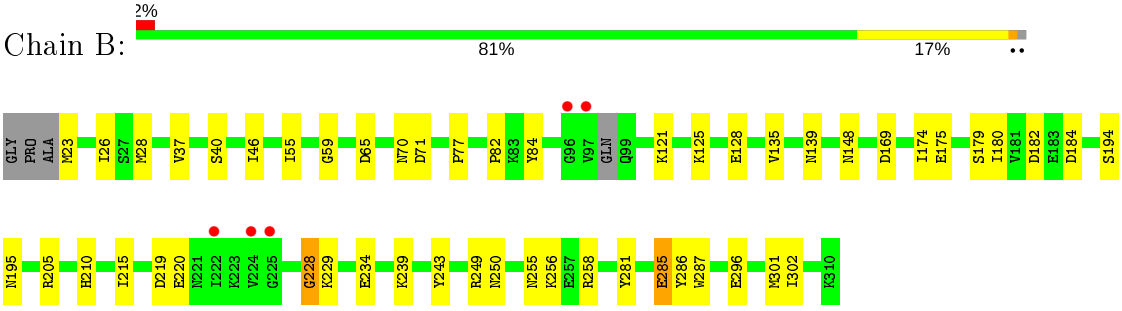
- Molecule 1: Enterochelin uptake periplasmic binding protein

Chain A:



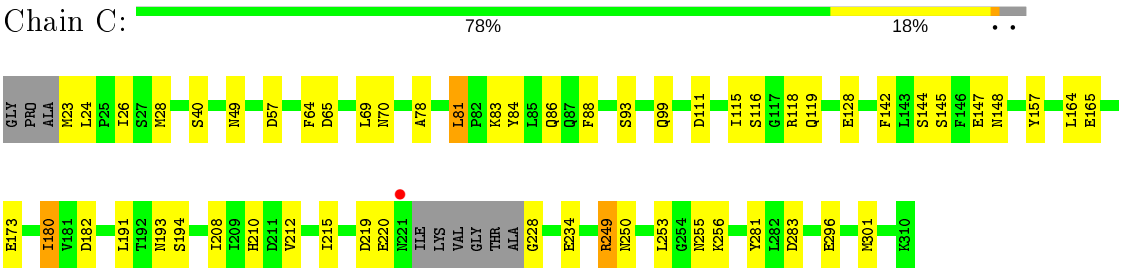
- Molecule 1: Enterochelin uptake periplasmic binding protein

Chain B:



- Molecule 1: Enterochelin uptake periplasmic binding protein

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	56.99Å 62.68Å 67.97Å 82.11° 77.21° 76.28°	Depositor
Resolution (Å)	66.02 – 1.33 66.02 – 1.33	Depositor EDS
% Data completeness (in resolution range)	87.2 (66.02-1.33) 87.2 (66.02-1.33)	Depositor EDS
R_{merge}	0.00	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.11 (at 1.33Å)	Xtriage
Refinement program	REFMAC 5.8.0155	Depositor
R, R_{free}	0.148 , 0.193 0.153 , 0.201	Depositor DCC
R_{free} test set	8932 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	13.5	Xtriage
Anisotropy	0.319	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 42.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	7062	wwPDB-VP
Average B, all atoms (Å^2)	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.62% 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

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	1.69	19/2250 (0.8%)	1.44	22/3045 (0.7%)
1	B	1.57	13/2256 (0.6%)	1.38	17/3049 (0.6%)
1	C	1.63	21/2202 (1.0%)	1.52	18/2984 (0.6%)
All	All	1.63	53/6708 (0.8%)	1.45	57/9078 (0.6%)

The worst 5 of 53 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	116	SER	CB-OG	12.50	1.58	1.42
1	A	220	GLU	CD-OE2	8.72	1.35	1.25
1	A	286	TYR	CE1-CZ	-7.83	1.28	1.38
1	B	128[A]	GLU	CD-OE1	7.59	1.34	1.25
1	B	128[B]	GLU	CD-OE1	7.59	1.34	1.25

The worst 5 of 57 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	28	MET	CG-SD-CE	-15.24	75.81	100.20
1	C	118	ARG	NE-CZ-NH1	13.59	127.10	120.30
1	C	118	ARG	NE-CZ-NH2	-12.62	113.99	120.30
1	A	28	MET	CG-SD-CE	-12.15	80.75	100.20
1	C	249	ARG	NE-CZ-NH2	-12.10	114.25	120.30

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	2189	0	2209	19	0
1	B	2194	0	2211	20	1
1	C	2156	0	2166	19	1
2	A	199	0	0	6	1
2	B	181	0	0	3	0
2	C	143	0	0	3	1
All	All	7062	0	6586	58	2

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 4.

The worst 5 of 58 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:23:MET:SD	1:A:23:MET:CE	2.02	1.47
1:B:195:ASN:HB2	2:B:495:HOH:O	1.42	1.20
1:C:234:GLU:CG	2:C:530:HOH:O	1.91	1.16
1:A:26:ILE:H	1:A:148:ASN:HD21	1.20	0.88
1:B:26:ILE:H	1:B:148:ASN:HD21	1.23	0.86

All (2) 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)
1:B:256:LYS:CG	2:C:523:HOH:O[1_456]	1.76	0.44
1:C:256:LYS:CG	2:A:570:HOH:O[1_565]	2.01	0.19

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	292/291 (100%)	283 (97%)	9 (3%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	293/291 (101%)	281 (96%)	12 (4%)	0	100	100
1	C	283/291 (97%)	272 (96%)	11 (4%)	0	100	100
All	All	868/873 (99%)	836 (96%)	32 (4%)	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	233/251 (93%)	227 (97%)	6 (3%)	46	12
1	B	233/251 (93%)	231 (99%)	2 (1%)	78	52
1	C	227/251 (90%)	221 (97%)	6 (3%)	46	12
All	All	693/753 (92%)	679 (98%)	14 (2%)	57	20

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

Mol	Chain	Res	Type
1	B	121	LYS
1	B	239	LYS
1	C	165	GLU
1	A	291[B]	SER
1	C	144	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	148	ASN
1	B	210	HIS
1	C	210	HIS
1	B	70	ASN
1	C	216	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, 95^{th} 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	286/291 (98%)	-0.11	7 (2%)	59	65	14, 23, 40, 49	5 (1%)
1	B	287/291 (98%)	-0.24	5 (1%)	70	74	16, 24, 38, 50	8 (2%)
1	C	282/291 (96%)	-0.19	1 (0%)	92	93	14, 25, 39, 62	3 (1%)
All	All	855/873 (97%)	-0.18	13 (1%)	73	77	14, 24, 39, 62	16 (1%)

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	221	ASN	11.0
1	B	97	VAL	7.3
1	A	97	VAL	5.6
1	B	222	ILE	4.1
1	B	224	VAL	3.4

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 carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.