

Full wwPDB X-ray Structure Validation Report (i)

Jun 16, 2024 - 09:04 PM EDT

PDB ID : 5MBU

Title: CeuE (H227A, Y288F variant) a periplasmic protein from Campylobacter je-

juni

Authors: Wilde, E.J.; Blagova, E.V.; Hughes, A.; Raines, D.J.; Moroz, O.V.; Turken-

burg, J.P.; Duhme-Klair, A.-K.; Wilson, K.S.

Deposited on : 2016-11-08

Resolution : 1.81 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
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 (1)) were used in the production of this report:

MolProbity: 4.02b-467 Xtriage (Phenix): 1.13

EDS : 2.37.1

Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)

 $Refmac \quad : \quad 5.8.0158$

CCP4 : 7.0.044 (Gargrove)

Ideal geometry (proteins) : Engh & Huber (2001) Ideal geometry (DNA, RNA) : Parkinson et al. (1996)

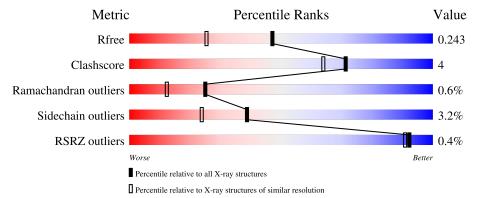
Validation Pipeline (wwPDB-VP) : 2.37.1

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 1.81 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 $(\# \mathrm{Entries})$	$\begin{array}{c} {\rm Similar\ resolution} \\ (\#{\rm Entries,\ resolution\ range(\mathring{A})}) \end{array}$
R_{free}	130704	7484 (1.84-1.80)
Clashscore	141614	8401 (1.84-1.80)
Ramachandran outliers	138981	8290 (1.84-1.80)
Sidechain outliers	138945	8290 (1.84-1.80)
RSRZ outliers	127900	7371 (1.84-1.80)

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	291	89%	9%
1	С	291	87%	9% • •
2	В	291	89%	9%



2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 6439 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.

\mathbf{Mol}	Chain	Residues	${f Atoms}$					ZeroOcc	AltConf	Trace
1	A 287	Total	С	N	О	S	0	2	0	
	201	2116	1376	337	400	3		2		
1	C	C 202	Total	С	N	Ο	S	0	9	1
	283	2099	1363	335	398	3	U	3		

There are 12 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
A	288	PHE	TYR	engineered mutation	UNP Q0P8Q4
С	20	GLY	-	expression tag	UNP Q0P8Q4
С	21	PRO	-	expression tag	UNP Q0P8Q4
С	22	ALA	-	expression tag	UNP Q0P8Q4
С	23	MET	-	expression tag	UNP Q0P8Q4
С	227	ALA	HIS	engineered mutation	UNP Q0P8Q4
С	288	PHE	TYR	engineered mutation	UNP Q0P8Q4

• Molecule 2 is a protein called Enterochelin uptake periplasmic binding protein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace	
2	В	287	Total 2115	C 1376	N 340	O 397	S	0	3	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
В	20	GLY	-	expression tag	UNP Q0P8Q4
В	21	PRO	-	expression tag	UNP Q0P8Q4



Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
В	22	ALA	-	expression tag	UNP Q0P8Q4
В	23	MET	-	expression tag	UNP Q0P8Q4
В	41	ILE	LEU	conflict	UNP Q0P8Q4
В	227	ALA	HIS	engineered mutation	UNP Q0P8Q4
В	288	PHE	TYR	engineered mutation	UNP Q0P8Q4

• Molecule 3 is water.

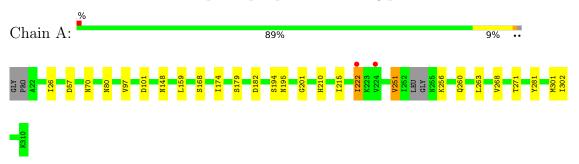
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	54	Total O 54 54	0	0
3	В	30	Total O 30 30	0	0
3	С	25	Total O 25 25	0	0



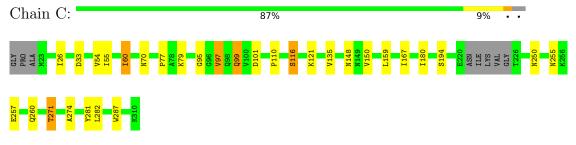
3 Residue-property plots (i)

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.

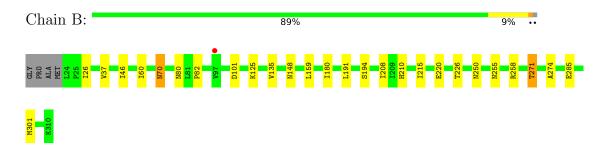
• Molecule 1: Enterochelin uptake periplasmic binding protein



• Molecule 1: Enterochelin uptake periplasmic binding protein



• Molecule 2: Enterochelin uptake periplasmic binding protein





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants	56.90Å 62.61Å 67.79Å	Donositor
a, b, c, α , β , γ	82.23° 76.96° 75.98°	Depositor
Resolution (Å)	65.80 - 1.81	Depositor
rtesolution (A)	65.80 - 1.81	EDS
% Data completeness	95.8 (65.80-1.81)	Depositor
(in resolution range)	95.8 (65.80-1.81)	EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	1.99 (at 1.81Å)	Xtriage
Refinement program	REFMAC 5.8.0155	Depositor
D D.	0.200 , 0.239	Depositor
R, R_{free}	0.207 , 0.243	DCC
R_{free} test set	3731 reflections (4.85%)	wwPDB-VP
Wilson B-factor (Å ²)	31.5	Xtriage
Anisotropy	0.038	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.30, 35.5	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	0.96	EDS
Total number of atoms	6439	wwPDB-VP
Average B, all atoms (Å ²)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 5.85% of the height of the origin peak. No significant pseudotranslation is detected.

²Theoretical 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.



¹Intensities estimated from amplitudes.

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	Bo	nd lengths	Bond angles		
WIOI		RMSZ	# Z > 5	RMSZ	# Z > 5	
1	A	0.94	0/2151	0.95	5/2908~(0.2%)	
1	С	0.96	$4/2138 \; (0.2\%)$	0.98	5/2891 (0.2%)	
2	В	0.89	0/2155	0.95	4/2914 (0.1%)	
All	All	0.93	4/6444 (0.1%)	0.96	14/8713 (0.2%)	

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$\operatorname{Observed}(\operatorname{\AA})$	$\operatorname{Ideal}(ext{\AA})$
1	С	116[B]	SER	CA-CB	11.50	1.70	1.52
1	С	116[C]	SER	CA-CB	11.50	1.70	1.52
1	С	116[B]	SER	N-CA	5.15	1.56	1.46
1	С	116[C]	SER	N-CA	5.15	1.56	1.46

All (14) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^o)$	$\operatorname{Ideal}({}^{o})$
1	A	301	MET	CG-SD-CE	-8.24	87.01	100.20
2	В	301	MET	CG-SD-CE	-8.03	87.35	100.20
1	С	116[B]	SER	N-CA-CB	6.42	120.12	110.50
1	С	116[C]	SER	N-CA-CB	6.42	120.12	110.50
1	С	101	ASP	CB-CG-OD1	6.34	124.00	118.30
1	С	33	ASP	N-CA-CB	-6.10	99.61	110.60
2	В	101	ASP	CB-CG-OD1	6.10	123.79	118.30
1	A	101	ASP	CB-CG-OD1	5.81	123.53	118.30
1	A	251	VAL	C-N-CA	5.72	136.00	121.70
1	С	101	ASP	CB-CG-OD2	-5.71	113.17	118.30
2	В	258	ARG	NE-CZ-NH2	-5.62	117.49	120.30
1	A	57	ASP	CB-CG-OD2	5.30	123.07	118.30
1	A	182	ASP	CB-CG-OD1	5.11	122.90	118.30
2	В	101	ASP	CB-CG-OD2	-5.02	113.78	118.30

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	A	2116	0	2079	12	0
1	С	2099	0	2076	18	0
2	В	2115	0	2079	13	0
3	A	54	0	0	2	0
3	В	30	0	0	0	0
3	С	25	0	0	1	0
All	All	6439	0	6234	44	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 4.

All (44) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	$\begin{array}{c} \text{Interatomic} \\ \text{distance } (\text{\AA}) \end{array}$	Clash overlap (Å)
1:A:195:ASN:CA	3:A:449:HOH:O	2.22	0.86
1:C:79:LYS:O	3:C:401:HOH:O	1.92	0.86
1:A:26:ILE:H	1:A:148:ASN:HD21	1.25	0.82
1:A:201:GLY:HA3	1:A:222:ILE:HD11	1.65	0.79
1:C:26:ILE:H	1:C:148:ASN:HD21	1.31	0.78
2:B:26:ILE:H	2:B:148:ASN:HD21	1.46	0.63
1:C:260:GLN:HG3	1:C:281:TYR:CZ	2.37	0.60
1:A:201:GLY:HA3	1:A:222:ILE:CD1	2.33	0.57
1:A:70:ASN:H	1:A:70:ASN:HD22	1.52	0.54
1:C:150[B]:VAL:CG2	1:C:167:ILE:HD11	2.37	0.53
1:A:210:HIS:HD2	1:A:215:ILE:O	1.92	0.53
1:C:70:ASN:HD22	1:C:70:ASN:H	1.55	0.53
2:B:70:ASN:H	2:B:70:ASN:HD22	1.56	0.53
1:A:80:ASN:N	1:A:80:ASN:O	2.44	0.51
2:B:250:ASN:HD22	2:B:255:ASN:HD22	1.59	0.51
1:C:250:ASN:HD22	1:C:255:ASN:HD22	1.58	0.50
2:B:37:VAL:HG21	2:B:46:ILE:HD12	1.94	0.50
1:C:194:SER:HA	1:C:257:GLU:HG2	1.93	0.49



Continued from previous page...

A	A. 0	Interatomic	Clash
Atom-1	Atom-2	${\rm distance}(\mathring{\rm A})$	overlap (Å)
1:C:116[B]:SER:HB2	1:C:135:VAL:O	2.13	0.48
1:A:174:ILE:HG13	1:A:302:ILE:HD13	1.97	0.47
1:C:116[B]:SER:HB2	1:C:135:VAL:HG23	1.97	0.46
1:C:194:SER:H	1:C:250:ASN:HD21	1.64	0.46
1:A:194:SER:OG	1:A:195:ASN:N	2.48	0.46
2:B:194:SER:H	2:B:250:ASN:HD21	1.62	0.46
1:C:54:VAL:HG23	1:C:110:PRO:HB3	1.98	0.45
1:C:194:SER:N	1:C:250:ASN:HD21	2.14	0.45
2:B:210:HIS:HD2	2:B:215:ILE:O	2.00	0.45
2:B:82:PRO:HB3	2:B:285:GLU:HB3	1.98	0.44
1:A:222:ILE:N	1:A:222:ILE:HD13	2.33	0.44
2:B:194:SER:N	2:B:250:ASN:HD21	2.15	0.44
1:C:60:ILE:HG13	1:C:135:VAL:HG21	2.01	0.43
1:A:263:LEU:HB3	1:A:268:VAL:HG11	2.02	0.42
2:B:250:ASN:ND2	2:B:255:ASN:HD22	2.16	0.42
1:C:55:ILE:O	1:C:77:PRO:HD3	2.20	0.42
2:B:60:ILE:HG13	2:B:135:VAL:HG21	2.00	0.42
1:C:271:THR:HG22	1:C:274:ALA:H	1.85	0.42
2:B:80:ASN:C	2:B:80:ASN:N	2.73	0.42
1:C:95:GLY:HA3	1:C:99:GLN:O	2.21	0.41
2:B:271:THR:HG22	2:B:274:ALA:H	1.85	0.41
2:B:271:THR:CG2	2:B:274:ALA:H	2.34	0.41
1:C:271:THR:CG2	1:C:274:ALA:H	2.34	0.40
1:C:282:LEU:HB3	1:C:287:TRP:CD1	2.56	0.40
1:A:260:GLN:HG3	1:A:281:TYR:CZ	2.57	0.40

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	283/291 (97%)	269 (95%)	11 (4%)	3 (1%)	14 4



Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentile	\mathbf{s}
1	С	281/291 (97%)	270 (96%)	10 (4%)	1 (0%)	34 21	
2	В	287/291 (99%)	276 (96%)	9 (3%)	2 (1%)	22 10	
All	All	851/873 (98%)	815 (96%)	30 (4%)	6 (1%)	25 10	

All (6) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	97[A]	VAL
1	A	97[B]	VAL
2	В	226	THR
2	В	208	ILE
1	A	256	LYS
1	С	97	VAL

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	210/251~(84%)	204 (97%)	6 (3%)	42 28
1	С	212/251 (84%)	205 (97%)	7 (3%)	38 23
2	В	208/251 (83%)	201 (97%)	7 (3%)	37 22
All	All	630/753~(84%)	610 (97%)	20 (3%)	39 24

All (20) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	159	LEU
1	A	168	SER
1	A	179	SER
1	A	222	ILE
1	A	251	VAL
1	A	271	THR
2	В	70	ASN
2	В	125	LYS



Continued from previous page...

Mol	Chain	Res	Type
2	В	159	LEU
2	В	180	ILE
2	В	191	LEU
2	В	220	GLU
2	В	271	THR
1	С	60	ILE
1	С	97	VAL
1	С	99	GLN
1	С	121	LYS
1	С	159	LEU
1	С	180	ILE
1	С	271	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (15) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	70	ASN
1	A	148	ASN
1	A	172	ASN
1	A	210	HIS
2	В	70	ASN
2	В	148	ASN
2	В	210	HIS
2	В	216	ASN
2	В	250	ASN
1	С	49	ASN
1	С	70	ASN
1	С	148	ASN
1	С	195	ASN
1	С	210	HIS
1	С	250	ASN

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 monosaccharides 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, 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	$\langle { m RSRZ} \rangle$	# RSRZ > 2	$OWAB(A^2)$	Q<0.9
1	A	287/291 (98%)	-0.43	2 (0%) 87 86	21, 33, 53, 77	2 (0%)
1	С	283/291 (97%)	-0.44	0 100 100	21, 38, 51, 76	6 (2%)
2	В	287/291 (98%)	-0.50	1 (0%) 94 92	23, 36, 54, 83	2 (0%)
All	All	857/873 (98%)	-0.46	3 (0%) 92 91	21, 36, 54, 83	10 (1%)

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	222	ILE	4.4
2	В	97	VAL	2.6
1	A	224	VAL	2.6

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 monosaccharides 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.

