



wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 09:33 am BST

PDB ID : 6JHO
Title : The complex crystal structure of Cagbeta with CagZ revealed a novel regulatory mechanism for T4SS coupling ATPase in Helicobacter pylori
Authors : Wu, X.; Zhao, Y.; Sun, L.; Ye, X.; Jiang, M.; Wang, Q.; Wang, Q.; Wu, Y.
Deposited on : 2019-02-18
Resolution : 2.10 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 ⓘ symbol.

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

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

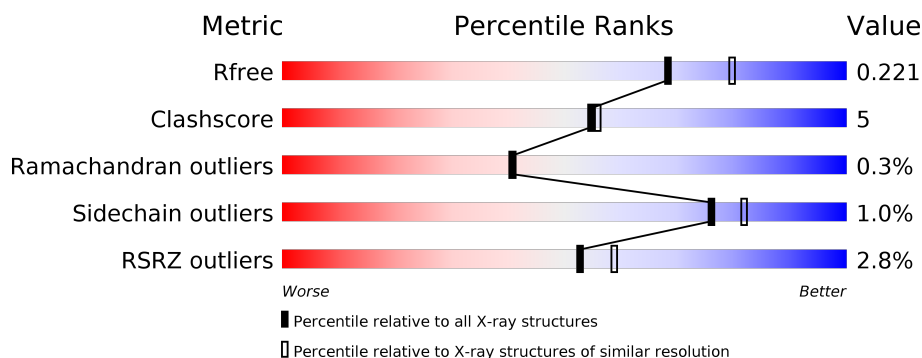
1 Overall quality at a glance

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))
R_{free}	130704	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (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 ≥ 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 $\leq 5\%$. 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	208	<div> <div>4%</div> <div> <div></div> <div>86%</div> <div>11%</div> <div></div> </div> </div>
1	B	208	<div> <div>%</div> <div> <div></div> <div>84%</div> <div>12%</div> <div></div> </div> </div>
2	C	210	<div> <div>2%</div> <div> <div></div> <div>75%</div> <div>7%</div> <div></div> </div> </div>
2	D	210	<div> <div>2%</div> <div> <div></div> <div>72%</div> <div>10%</div> <div></div> </div> </div>

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 6384 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 Cag pathogenicity island protein (Cag6).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	200	Total	C	N	O	S	0	0	0
			1619	1024	272	316	7			
1	B	200	Total	C	N	O	S	0	0	0
			1619	1024	272	316	7			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-8	GLY	-	expression tag	UNP O25261
A	-7	THR	-	expression tag	UNP O25261
A	-6	SER	-	expression tag	UNP O25261
A	-5	SER	-	expression tag	UNP O25261
A	-4	MET	-	expression tag	UNP O25261
A	-3	ALA	-	expression tag	UNP O25261
A	-2	ASP	-	expression tag	UNP O25261
A	-1	ILE	-	expression tag	UNP O25261
A	0	GLY	-	expression tag	UNP O25261
A	1	SER	-	expression tag	UNP O25261
B	-8	GLY	-	expression tag	UNP O25261
B	-7	THR	-	expression tag	UNP O25261
B	-6	SER	-	expression tag	UNP O25261
B	-5	SER	-	expression tag	UNP O25261
B	-4	MET	-	expression tag	UNP O25261
B	-3	ALA	-	expression tag	UNP O25261
B	-2	ASP	-	expression tag	UNP O25261
B	-1	ILE	-	expression tag	UNP O25261
B	0	GLY	-	expression tag	UNP O25261
B	1	SER	-	expression tag	UNP O25261

- Molecule 2 is a protein called Cag pathogenicity island protein (Cag5).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	175	Total	C	N	O	S	0	0	0
			1392	892	228	262	10			
2	D	174	Total	C	N	O	S	0	0	0
			1384	888	226	260	10			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	279	GLY	-	expression tag	UNP O25260
C	280	THR	-	expression tag	UNP O25260
C	281	SER	-	expression tag	UNP O25260
C	282	SER	-	expression tag	UNP O25260
C	283	MET	-	expression tag	UNP O25260
C	284	ALA	-	expression tag	UNP O25260
C	285	ASP	-	expression tag	UNP O25260
C	286	ILE	-	expression tag	UNP O25260
C	287	GLY	-	expression tag	UNP O25260
C	288	SER	-	expression tag	UNP O25260
D	279	GLY	-	expression tag	UNP O25260
D	280	THR	-	expression tag	UNP O25260
D	281	SER	-	expression tag	UNP O25260
D	282	SER	-	expression tag	UNP O25260
D	283	MET	-	expression tag	UNP O25260
D	284	ALA	-	expression tag	UNP O25260
D	285	ASP	-	expression tag	UNP O25260
D	286	ILE	-	expression tag	UNP O25260
D	287	GLY	-	expression tag	UNP O25260
D	288	SER	-	expression tag	UNP O25260

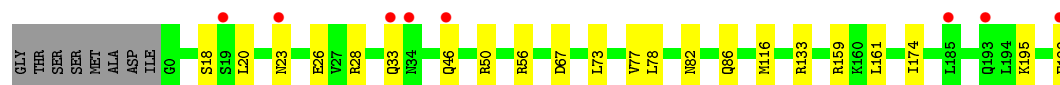
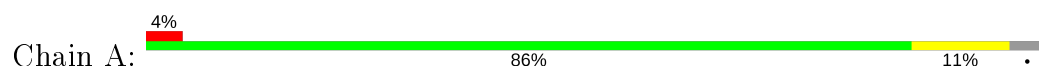
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	104	Total	O	0	0
			104	104		
3	B	110	Total	O	0	0
			110	110		
3	C	78	Total	O	0	0
			78	78		
3	D	78	Total	O	0	0
			78	78		

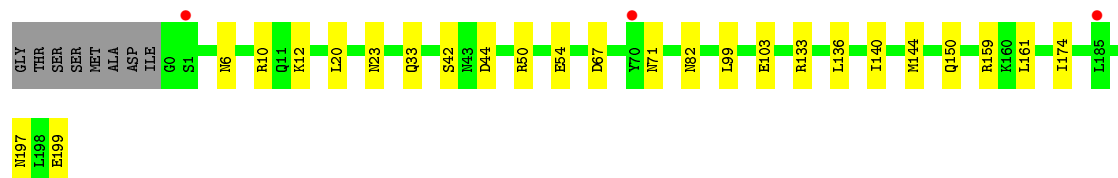
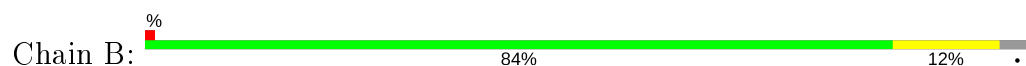
3 Residue-property plots [i](#)

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 ($RSRZ > 2$). 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.

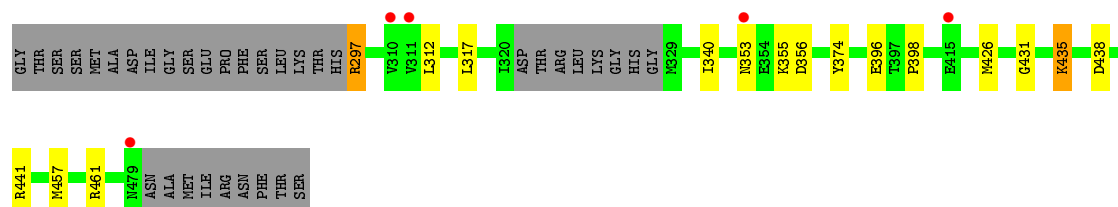
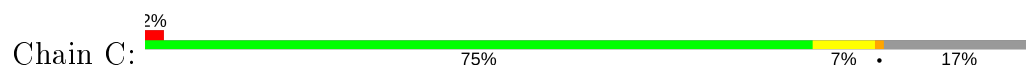
- Molecule 1: Cag pathogenicity island protein (Cag6)



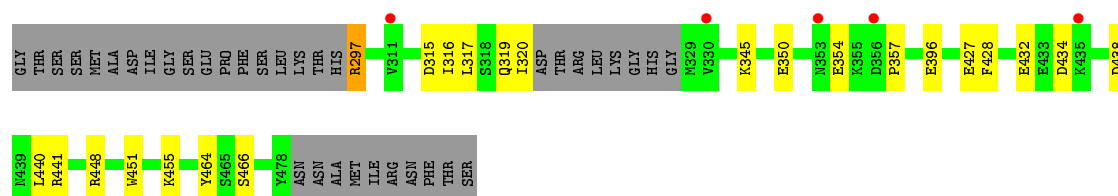
- Molecule 1: Cag pathogenicity island protein (Cag6)



- Molecule 2: Cag pathogenicity island protein (Cag5)



- Molecule 2: Cag pathogenicity island protein (Cag5)



4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	96.30 Å 96.30 Å 212.66 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.15 – 2.10 49.11 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.9 (48.15-2.10) 99.8 (49.11-2.00)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.32 (at 2.00 Å)	Xtriage
Refinement program	PHENIX 1.11.1 _2575	Depositor
R, R_{free}	0.191 , 0.221 0.191 , 0.221	Depositor DCC
R_{free} test set	3435 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å ²)	37.3	Xtriage
Anisotropy	0.010	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 48.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	6384	wwPDB-VP
Average B, all atoms (Å ²)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.19% 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 [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	Bond lengths		Bond angles	
		RMSZ	# $ Z > 5$	RMSZ	# $ Z > 5$
1	A	0.42	0/1643	0.52	0/2212
1	B	0.41	0/1643	0.52	0/2212
2	C	0.41	0/1423	0.53	0/1918
2	D	0.41	0/1415	0.53	0/1907
All	All	0.42	0/6124	0.52	0/8249

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 [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	1619	0	1607	19	0
1	B	1619	0	1607	20	0
2	C	1392	0	1357	14	0
2	D	1384	0	1351	18	0
3	A	104	0	0	6	0
3	B	110	0	0	7	0
3	C	78	0	0	5	0
3	D	78	0	0	5	0
All	All	6384	0	5922	62	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 62 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:78:LEU:HD13	1:B:71:ASN:HD21	1.28	0.98
1:A:56:ARG:NH1	3:A:202:HOH:O	2.01	0.92
2:D:464:TYR:OH	3:D:501:HOH:O	1.95	0.85
1:A:33:GLN:O	3:A:201:HOH:O	1.95	0.83
2:C:297:ARG:NH1	3:C:501:HOH:O	2.14	0.80

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	198/208 (95%)	196 (99%)	2 (1%)	0	100	100
1	B	198/208 (95%)	195 (98%)	2 (1%)	1 (0%)	29	26
2	C	171/210 (81%)	164 (96%)	6 (4%)	1 (1%)	25	21
2	D	170/210 (81%)	165 (97%)	5 (3%)	0	100	100
All	All	737/836 (88%)	720 (98%)	15 (2%)	2 (0%)	41	41

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	355	LYS
1	B	23	ASN

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	181/187 (97%)	178 (98%)	3 (2%)	60	67
1	B	181/187 (97%)	180 (99%)	1 (1%)	86	90
2	C	153/182 (84%)	151 (99%)	2 (1%)	69	75
2	D	152/182 (84%)	151 (99%)	1 (1%)	84	88
All	All	667/738 (90%)	660 (99%)	7 (1%)	76	82

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	44	ASP
2	D	297	ARG
2	C	297	ARG
1	A	28	ARG
2	C	435	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	71	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 [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	200/208 (96%)	0.37	8 (4%) 38 44	23, 41, 61, 69	0
1	B	200/208 (96%)	0.23	3 (1%) 73 77	25, 37, 61, 71	0
2	C	175/210 (83%)	0.21	5 (2%) 51 57	27, 41, 64, 88	0
2	D	174/210 (82%)	0.24	5 (2%) 51 57	28, 40, 65, 85	0
All	All	749/836 (89%)	0.26	21 (2%) 53 59	23, 40, 63, 88	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	356	ASP	2.9
2	D	353	ASN	2.9
1	B	1	SER	2.9
1	A	199	GLU	2.7
1	A	185	LEU	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 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.