wwPDB X-ray Structure Validation Summary Report (i)

Jan 29, 2024 – 08:32 pm GMT

PDB ID	:	8CB2
Title	:	A complex of cagX and cagY components of Helicobacter pylori type IV secretion system
Authors	:	Isupov, M.N.; Jiang, M.; Zhang, H.; Wu, Y.
Deposited on	:	2023-01-25
Resolution	:	2.70 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 (1)) were used in the production of this report:

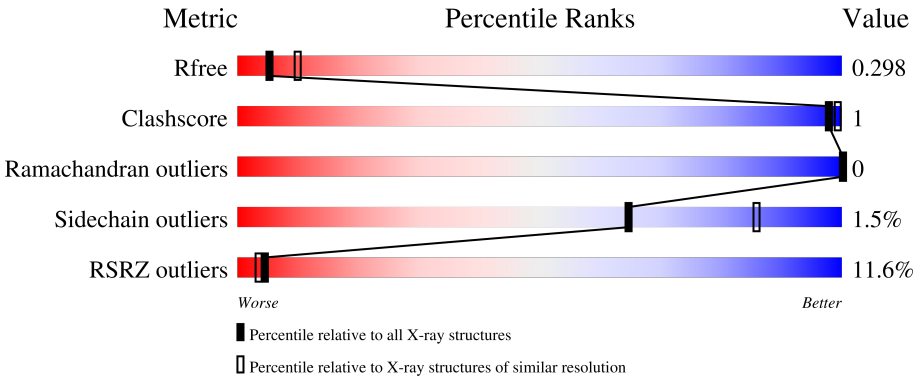
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.70 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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	AAA	164	15% 93% · 5%
1	CCC	164	13% 87% · 13%
2	BBB	264	6% 86% · 11%
2	DDD	264	11% 86% · 10%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 6032 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	156	Total	C	N	O	S	0	0	0
			1281	825	222	231	3			
1	CCC	143	Total	C	N	O	S	0	0	0
			1163	750	200	210	3			

- Molecule 2 is a protein called Type IV secretion system apparatus protein CagY (Fragment).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	BBB	236	Total	C	N	O	S	0	0	0
			1786	1142	292	343	9			
2	DDD	238	Total	C	N	O	S	0	0	0
			1800	1150	294	346	10			

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	CCC	1	Total	Ca	0	0
			1	1		

- Molecule 4 is water.

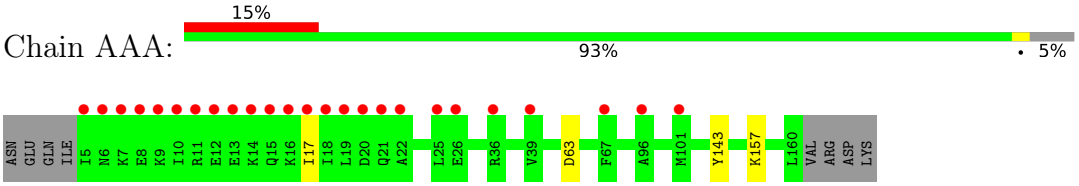
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	CCC	1	Total	O	0	0
			1	1		

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.

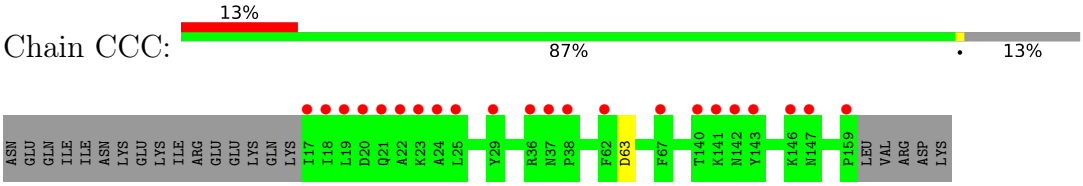
• Molecule 1: Cag pathogenicity island protein

Chain AAA:



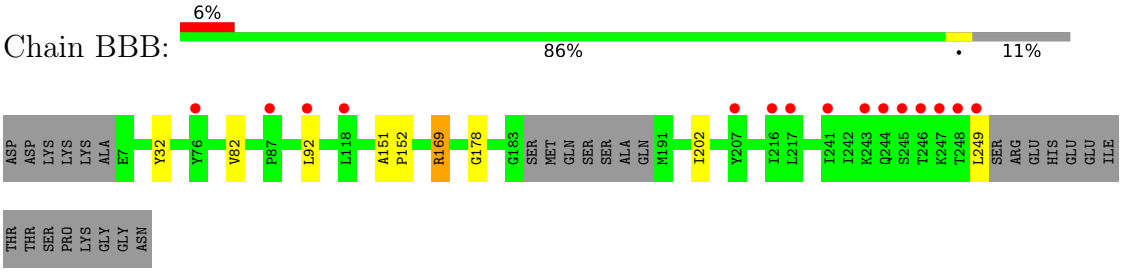
• Molecule 1: Cag pathogenicity island protein

Chain CCC:



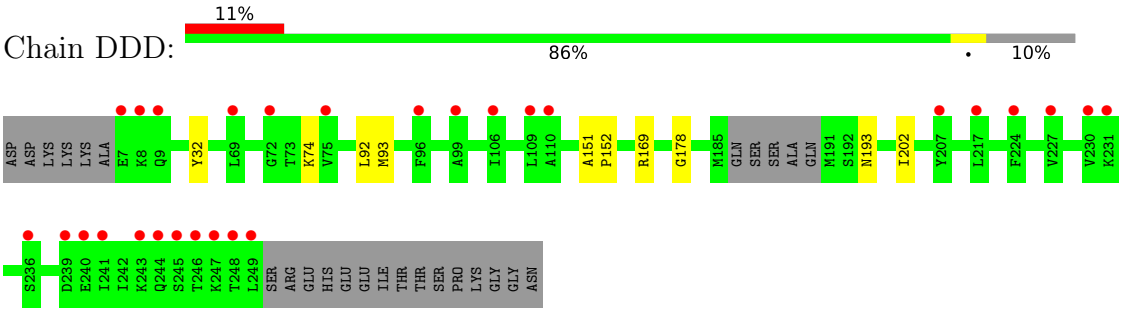
• Molecule 2: Type IV secretion system apparatus protein CagY (Fragment)

Chain BBB:



• Molecule 2: Type IV secretion system apparatus protein CagY (Fragment)

Chain DDD:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	55.06Å 196.12Å 200.13Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.08 – 2.70 49.03 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.9 (49.08-2.70) 99.9 (49.03-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.36 (at 2.69Å)	Xtriage
Refinement program	REFMAC 5.8.0267, BUSTER, ISOLDE	Depositor
R, R_{free}	0.252 , 0.298 0.252 , 0.298	Depositor DCC
R_{free} test set	1515 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	76.6	Xtriage
Anisotropy	0.245	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 70.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	6032	wwPDB-VP
Average B, all atoms (Å^2)	109.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 15.97% 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

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

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	AAA	0.37	0/1308	0.65	0/1763
1	CCC	0.39	0/1190	0.67	0/1609
2	BBB	0.36	0/1813	0.64	0/2454
2	DDD	0.37	0/1827	0.64	0/2472
All	All	0.37	0/6138	0.65	0/8298

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	AAA	1281	0	1319	1	0
1	CCC	1163	0	1179	0	0
2	BBB	1786	0	1838	6	0
2	DDD	1800	0	1852	4	0
3	CCC	1	0	0	0	0
4	CCC	1	0	0	0	0
All	All	6032	0	6188	8	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.

The worst 5 of 8 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)
2:BBB:92:LEU:H	2:BBB:202:ILE:HD13	1.60	0.67
2:DDD:92:LEU:H	2:DDD:202:ILE:HD13	1.65	0.60
1:AAA:17:ILE:HD13	2:BBB:249:LEU:HD11	1.98	0.46
2:DDD:151:ALA:N	2:DDD:152:PRO:HD2	2.33	0.44
2:BBB:169:ARG:HD3	2:BBB:169:ARG:C	2.39	0.43

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	AAA	154/164 (94%)	150 (97%)	4 (3%)	0	100	100
1	CCC	141/164 (86%)	133 (94%)	8 (6%)	0	100	100
2	BBB	232/264 (88%)	230 (99%)	2 (1%)	0	100	100
2	DDD	234/264 (89%)	229 (98%)	5 (2%)	0	100	100
All	All	761/856 (89%)	742 (98%)	19 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	AAA	140/148 (95%)	137 (98%)	3 (2%)	53 80
1	CCC	126/148 (85%)	125 (99%)	1 (1%)	81 93
2	BBB	198/222 (89%)	196 (99%)	2 (1%)	76 91
2	DDD	200/222 (90%)	196 (98%)	4 (2%)	55 81
All	All	664/740 (90%)	654 (98%)	10 (2%)	65 86

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	DDD	93	MET
2	DDD	169	ARG
2	DDD	193	ASN
2	BBB	82	VAL
2	BBB	169	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

Of 1 ligands modelled in this entry, 1 is 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^{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	AAA	156/164 (95%)	1.09	25 (16%)	1	1	50, 91, 188, 218	0
1	CCC	143/164 (87%)	0.98	22 (15%)	2	1	50, 84, 176, 196	0
2	BBB	236/264 (89%)	0.60	15 (6%)	19	18	63, 103, 173, 223	0
2	DDD	238/264 (90%)	0.88	28 (11%)	4	3	60, 108, 202, 258	0
All	All	773/856 (90%)	0.85	90 (11%)	4	3	50, 102, 187, 258	0

The worst 5 of 90 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	10	ILE	13.5
2	BBB	249	LEU	13.3
1	CCC	38	PRO	12.1
1	CCC	17	ILE	10.9
2	DDD	248	THR	9.8

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

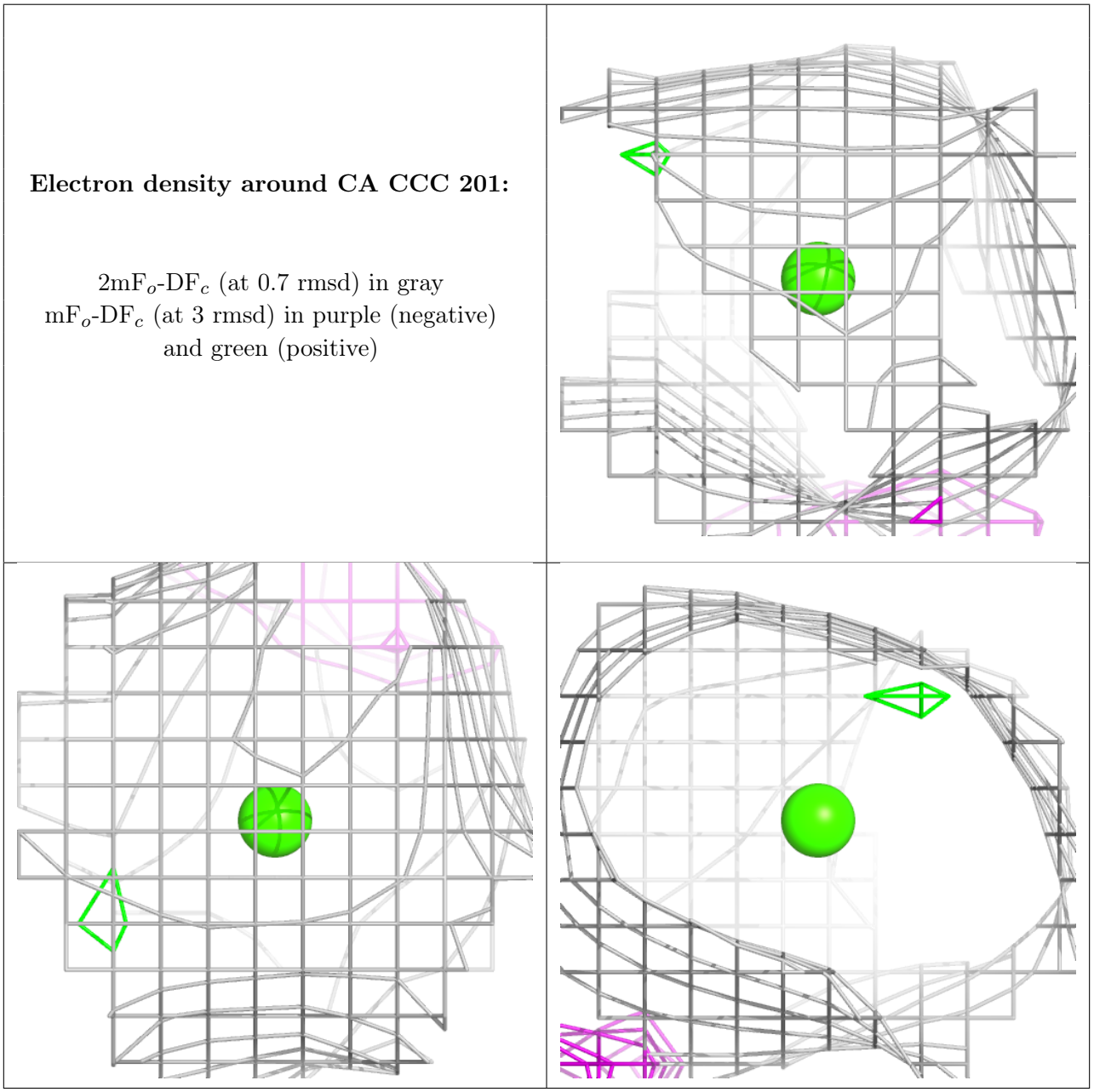
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	CA	CCC	201	1/1	0.97	0.20	66,66,66,66	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around CA CCC 201:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



6.5 Other polymers (i)

There are no such residues in this entry.