



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:59 AM GMT

PDB ID : 3CWX
Title : Crystal structure of cagA from helicobacter pylori pathogenicity island
Authors : Cendron, L.; Zanotti, G.; Angelini A.; Barison N.; Couturier, M.; Stein, M.
Deposited on : 2008-04-23
Resolution : 2.30 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
<http://wwpdb.org/validation/2016/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
Mogul : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk26865

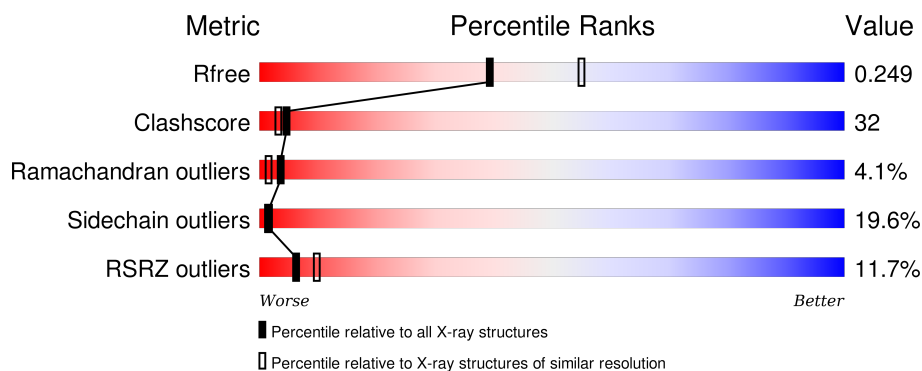
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.30 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}	91344	3852 (2.30-2.30)
Clashscore	102246	4452 (2.30-2.30)
Ramachandran outliers	100387	4410 (2.30-2.30)
Sidechain outliers	100360	4409 (2.30-2.30)
RSRZ outliers	91569	3857 (2.30-2.30)

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. 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	176	<div> <div>9%</div> <div>30%</div> <div>31%</div> <div>10%</div> <div>29%</div> </div>
1	B	176	<div> <div>4%</div> <div>37%</div> <div>26%</div> <div>7%</div> <div>29%</div> </div>
1	C	176	<div> <div>12%</div> <div>23%</div> <div>34%</div> <div>14%</div> <div>29%</div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3263 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 protein CagD.

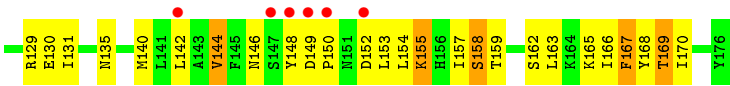
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	125	Total	C	N	O	S	Se	0	0	0
			1035	672	165	192	3	3			
1	B	125	Total	C	N	O	S	Se	0	0	0
			1035	672	165	192	3	3			
1	C	125	Total	C	N	O	S	Se	0	0	0
			1035	672	165	192	3	3			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	79	MSE	VAL	ENGINEERED	UNP P94837
A	97	ILE	LEU	ENGINEERED	UNP P94837
A	135	ASN	ASP	ENGINEERED	UNP P94837
A	140	MSE	VAL	ENGINEERED	UNP P94837
B	79	MSE	VAL	ENGINEERED	UNP P94837
B	97	ILE	LEU	ENGINEERED	UNP P94837
B	135	ASN	ASP	ENGINEERED	UNP P94837
B	140	MSE	VAL	ENGINEERED	UNP P94837
C	79	MSE	VAL	ENGINEERED	UNP P94837
C	97	ILE	LEU	ENGINEERED	UNP P94837
C	135	ASN	ASP	ENGINEERED	UNP P94837
C	140	MSE	VAL	ENGINEERED	UNP P94837

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	45	Total	O	0	0
			45	45		
2	B	80	Total	O	0	0
			80	80		
2	C	33	Total	O	0	0
			33	33		



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	78.18 Å 117.96 Å 65.16 Å 90.00° 110.26° 90.00°	Depositor
Resolution (Å)	62.26 – 2.30 42.44 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.6 (62.26-2.30) 96.1 (42.44-2.00)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.71 (at 2.00 Å)	Xtriage
Refinement program	SHELXL-97	Depositor
R, R_{free}	0.242 , 0.288 0.246 , 0.249	Depositor DCC
R_{free} test set	1226 reflections (5.26%)	DCC
Wilson B-factor (Å ²)	34.9	Xtriage
Anisotropy	0.199	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 100.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	1 of 35927 reflections (0.003%)	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	3263	wwPDB-VP
Average B, all atoms (Å ²)	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.75% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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	0.41	0/1050	0.88	1/1409 (0.1%)
1	B	0.43	0/1050	0.94	1/1409 (0.1%)
1	C	0.41	0/1050	0.85	1/1409 (0.1%)
All	All	0.42	0/3150	0.89	3/4227 (0.1%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	129	ARG	NE-CZ-NH1	-6.28	117.16	120.30
1	B	167	PHE	CB-CG-CD2	5.44	124.61	120.80
1	C	167	PHE	CB-CG-CD2	5.05	124.33	120.80

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	1035	0	1052	64	0
1	B	1035	0	1052	54	0
1	C	1035	0	1053	88	0
2	A	45	0	0	2	0
2	B	80	0	0	10	0
2	C	33	0	0	1	0
All	All	3263	0	3157	201	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 32.

The worst 5 of 201 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:B:53:ILE:HD13	1:B:80:ILE:HD13	1.52	0.92
1:B:98:GLN:HE22	1:B:105:ASP:H	1.20	0.89
1:B:98:GLN:NE2	1:B:105:ASP:H	1.84	0.76
1:C:155:LYS:O	1:C:159:THR:HG23	1.85	0.76
1:C:100:ASN:HB3	1:C:103:VAL:HG23	1.69	0.75

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	121/176 (69%)	102 (84%)	16 (13%)	3 (2%)	7	5
1	B	121/176 (69%)	110 (91%)	6 (5%)	5 (4%)	3	1
1	C	121/176 (69%)	107 (88%)	7 (6%)	7 (6%)	2	1
All	All	363/528 (69%)	319 (88%)	29 (8%)	15 (4%)	3	1

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	51	ASP
1	B	115	THR
1	C	50	PRO
1	C	53	ILE
1	C	115	THR

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	121/164 (74%)	98 (81%)	23 (19%)	2	1
1	B	121/164 (74%)	104 (86%)	17 (14%)	4	4
1	C	121/164 (74%)	90 (74%)	31 (26%)	0	0
All	All	363/492 (74%)	292 (80%)	71 (20%)	1	1

5 of 71 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	74	LYS
1	B	169	THR
1	C	154	LEU
1	B	90	GLU
1	B	123	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	98	GLN
1	B	116	ASN
1	C	98	GLN
1	A	146	ASN
1	C	64	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 [i](#)

There are no carbohydrates 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, 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	122/176 (69%)	0.99	15 (12%) 5 8	23, 47, 91, 115	0
1	B	122/176 (69%)	0.48	7 (5%) 27 36	14, 35, 76, 89	0
1	C	122/176 (69%)	1.19	21 (17%) 2 3	22, 55, 83, 112	0
All	All	366/528 (69%)	0.89	43 (11%) 6 10	14, 46, 84, 115	0

The worst 5 of 43 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	55	ASP	9.1
1	B	48	VAL	7.1
1	C	48	VAL	7.0
1	A	48	VAL	6.7
1	A	52	ILE	6.2

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.