



# Full wwPDB X-ray Structure Validation Report ⓘ

May 27, 2020 – 01:50 am BST

PDB ID : 4DVZ  
Title : Crystal structure of the Helicobacter pylori CagA oncoprotein  
Authors : Hayashi, T.; Senda, M.; Morohashi, H.; Higashi, H.; Horio, M.; Kashiba, Y.;  
Nagase, L.; Sasaya, D.; Shimizu, T.; Venugopalan, N.; Kumeta, H.; Noda, N.;  
Inagaki, F.; Senda, T.; Hatakeyama, M.  
Deposited on : 2012-02-23  
Resolution : 3.19 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](mailto: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

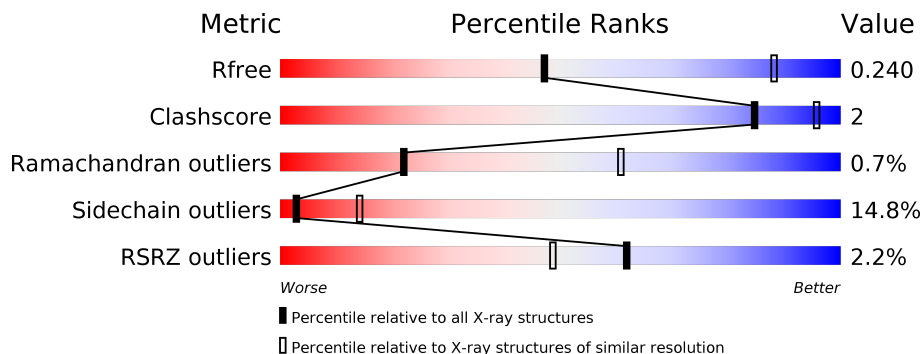
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.19 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	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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  $\geq 3$ , 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	569	

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 3635 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 Cytotoxicity-associated immunodominant antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	459	3635	2276	633	721	5	0	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	95.67Å 95.67Å 167.92Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	74.30 – 3.19 74.30 – 3.19	Depositor EDS
% Data completeness (in resolution range)	99.3 (74.30-3.19) 99.3 (74.30-3.19)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.29 (at 3.19Å)	Xtrriage
Refinement program	BUSTER 2.10.0	Depositor
R, $R_{free}$	0.222 , 0.279 0.240 , 0.240	Depositor DCC
$R_{free}$ test set	763 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	94.2	Xtrriage
Anisotropy	0.324	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 90.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.029 for -h,-k,l	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	3635	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	103.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.99% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup> Intensities estimated from amplitudes.

<sup>2</sup> Theoretical 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	0.47	0/3671	0.69	0/4900

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	3635	0	3656	17	0
All	All	3635	0	3656	17	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 2.

All (17) 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:354:ILE:HD11	1:A:376:PRO:HB3	1.50	0.93
1:A:319:LYS:HB2	1:A:342:TYR:HB2	1.68	0.76
1:A:558:ARG:HH12	1:A:580:LYS:HE2	1.64	0.62
1:A:474:THR:HB	1:A:493:THR:HB	1.81	0.62
1:A:406:GLU:HG3	1:A:471:LEU:HD11	1.82	0.61

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:335:ASP:HA	1:A:359:LYS:HE2	1.84	0.59
1:A:706:PHE:HE1	1:A:766:THR:HG22	1.70	0.57
1:A:618:ASP:HA	1:A:621:LYS:HD2	1.89	0.54
1:A:311:SER:HA	1:A:322:LEU:O	2.10	0.51
1:A:553:ILE:HG22	1:A:555:SER:HB3	1.93	0.50
1:A:666:LYS:HE3	1:A:670:ARG:HH12	1.77	0.49
1:A:581:ASP:HB3	1:A:629:LEU:HD11	1.95	0.48
1:A:590:ALA:O	1:A:594:LEU:HD12	2.13	0.48
1:A:691:LYS:HD3	1:A:734:LEU:HA	1.95	0.48
1:A:699:LEU:HD11	1:A:773:GLU:HA	1.98	0.45
1:A:334:HIS:NE2	1:A:337:ASN:OD1	2.45	0.44
1:A:475:LEU:HD11	1:A:609:TYR:HB3	2.03	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	441/569 (78%)	417 (95%)	21 (5%)	3 (1%)	22 61

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	332	ASP
1	A	642	GLY
1	A	416	ASP

### 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	398/487 (82%)	339 (85%)	59 (15%)	3 14

All (59) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	313	ASN
1	A	323	LEU
1	A	343	LYS
1	A	349	ASN
1	A	353	LEU
1	A	355	ASN
1	A	382	LYS
1	A	383	GLU
1	A	386	LEU
1	A	391	ARG
1	A	399	ARG
1	A	414	LYS
1	A	416	ASP
1	A	419	SER
1	A	420	GLU
1	A	424	GLU
1	A	429	GLU
1	A	434	GLN
1	A	437	SER
1	A	456	ASP
1	A	463	ILE
1	A	471	LEU
1	A	472	SER
1	A	491	ASN
1	A	493	THR
1	A	494	LEU
1	A	495	GLN
1	A	503	VAL
1	A	504	MET
1	A	506	VAL
1	A	545	LEU
1	A	561	LEU
1	A	571	SER
1	A	572	LEU

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	A	580	LYS
1	A	606	THR
1	A	632	GLU
1	A	636	LYS
1	A	637	LEU
1	A	641	SER
1	A	671	ASP
1	A	686	ARG
1	A	699	LEU
1	A	721	GLU
1	A	730	SER
1	A	734	LEU
1	A	736	ILE
1	A	739	GLU
1	A	744	VAL
1	A	745	GLU
1	A	770	SER
1	A	772	LEU
1	A	780	ILE
1	A	794	GLN
1	A	797	SER
1	A	807	ARG
1	A	808	VAL
1	A	812	LEU
1	A	822	GLN

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

Mol	Chain	Res	Type
1	A	680	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 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	459/569 (80%)	0.32	10 (2%) 62 48	62, 99, 147, 162	0

All (10) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	825	GLN	3.3
1	A	702	PHE	2.7
1	A	650	LYS	2.6
1	A	818	PHE	2.3
1	A	312	HIS	2.3
1	A	766	THR	2.2
1	A	324	TYR	2.2
1	A	325	ALA	2.1
1	A	815	LEU	2.1
1	A	332	ASP	2.1

### 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

There are no such residues in this entry.