



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 14, 2017 – 06:29 am GMT

PDB ID : 4DVY  
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 : 3.30 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

<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
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28949

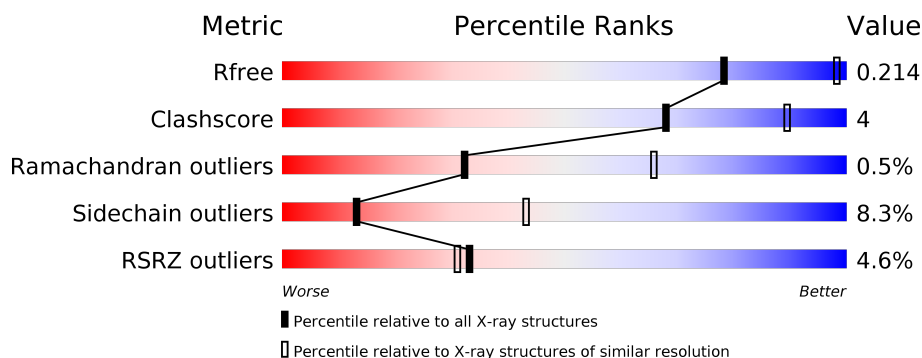
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.30 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))
$R_{free}$	100719	1034 (3.36-3.24)
Clashscore	112137	1100 (3.36-3.24)
Ramachandran outliers	110173	1081 (3.36-3.24)
Sidechain outliers	110143	1080 (3.36-3.24)
RSRZ outliers	101464	1039 (3.36-3.24)

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. 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	P	876	<div> <div>3%</div> <div>61%</div> <div>13%</div> <div>•</div> <div>25%</div> </div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 5229 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
1	P	656	Total	C	N	O	S	0	0	0
			5229	3288	902	1030	9			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
P	131	MET	HIS	ENGINEERED MUTATION	UNP P55980
P	274	MET	LEU	ENGINEERED MUTATION	UNP P55980



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	97.31Å 97.31Å 244.76Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	90.42 – 3.30 90.42 – 3.30	Depositor EDS
% Data completeness (in resolution range)	99.5 (90.42-3.30) 99.5 (90.42-3.30)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.69 (at 3.33Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, $R_{free}$	0.191 , 0.248 0.208 , 0.214	Depositor DCC
$R_{free}$ test set	921 reflections (5.27%)	DCC
Wilson B-factor (Å <sup>2</sup> )	119.0	Xtriage
Anisotropy	0.020	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 144.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	5229	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	145.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.85% 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	P	0.49	0/5297	0.71	0/7091

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	P	5229	0	5219	46	0
All	All	5229	0	5219	46	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 (46) 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:P:731:VAL:HG23	1:P:741:ILE:HG22	1.73	0.69
1:P:309:MET:HE1	1:P:381:TYR:HE1	1.57	0.69
1:P:60:ARG:HA	1:P:63:PHE:HD2	1.62	0.65
1:P:354:ILE:HD11	1:P:376:PRO:HB3	1.80	0.64
1:P:309:MET:HE1	1:P:381:TYR:CE1	2.33	0.64

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:P:135:PRO:HD2	1:P:427:GLN:HE22	1.66	0.60
1:P:113:SER:HA	1:P:117:PHE:HB2	1.86	0.58
1:P:42:ALA:HB2	1:P:49:LYS:HG3	1.87	0.56
1:P:212:TRP:HA	1:P:215:ILE:HD12	1.87	0.56
1:P:172:PHE:HA	1:P:175:ILE:HD12	1.88	0.56
1:P:394:SER:OG	1:P:397:GLU:HG3	2.06	0.55
1:P:632:GLU:HA	1:P:635:LYS:HD2	1.88	0.55
1:P:495:GLN:HG3	1:P:505:PHE:HB2	1.89	0.55
1:P:37:VAL:HA	1:P:40:VAL:HG22	1.88	0.54
1:P:181:ILE:HG22	1:P:188:MET:HE2	1.91	0.52
1:P:44:PHE:HB2	1:P:162:ALA:HB2	1.93	0.51
1:P:452:VAL:HG23	1:P:463:ILE:HG12	1.93	0.50
1:P:44:PHE:CD2	1:P:161:LYS:HB3	2.46	0.50
1:P:618:ASP:O	1:P:621:LYS:HB3	2.12	0.49
1:P:35:ILE:HG23	1:P:60:ARG:HH22	1.78	0.49
1:P:474:THR:HG23	1:P:493:THR:HB	1.94	0.48
1:P:494:LEU:HD23	1:P:494:LEU:N	2.28	0.48
1:P:666:LYS:HE2	1:P:670:ARG:HH12	1.77	0.48
1:P:592:LYS:O	1:P:595:ASN:HB2	2.16	0.46
1:P:660:ILE:HA	1:P:663:LEU:HD12	1.98	0.45
1:P:791:ASN:HB2	1:P:811:VAL:HG21	1.99	0.45
1:P:694:LYS:HA	1:P:697:LYS:HD2	1.99	0.45
1:P:134:ASP:HB2	1:P:427:GLN:OE1	2.18	0.43
1:P:786:THR:O	1:P:789:VAL:HG12	2.17	0.43
1:P:620:GLU:HA	1:P:623:LEU:HD12	2.00	0.43
1:P:724:LEU:HD13	1:P:751:LEU:HD12	2.01	0.43
1:P:750:ALA:HB1	1:P:765:VAL:HG22	2.01	0.43
1:P:738:PRO:O	1:P:741:ILE:HG12	2.19	0.43
1:P:158:ASP:HA	1:P:161:LYS:HB2	2.00	0.42
1:P:181:ILE:HG22	1:P:188:MET:CE	2.49	0.42
1:P:73:TYR:OH	1:P:94:LYS:HD3	2.18	0.42
1:P:404:PHE:CD2	1:P:429:GLU:HB3	2.53	0.42
1:P:669:ASN:HD21	1:P:673:ARG:HD2	1.85	0.42
1:P:111:THR:HA	1:P:114:PHE:HD2	1.86	0.41
1:P:35:ILE:HG23	1:P:60:ARG:NH2	2.35	0.41
1:P:393:LEU:HB3	1:P:397:GLU:HB2	2.02	0.41
1:P:105:ILE:HD11	1:P:177:ILE:HD12	2.02	0.41
1:P:320:VAL:HG22	1:P:341:GLY:HA2	2.03	0.41
1:P:503:VAL:HG12	1:P:504:MET:N	2.36	0.41
1:P:706:PHE:HD1	1:P:710:LYS:HE3	1.85	0.41
1:P:657:LYS:HA	1:P:660:ILE:HD12	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	P	638/876 (73%)	593 (93%)	42 (7%)	3 (0%)	32 66

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	P	115	GLN
1	P	154	PRO
1	P	570	LEU

### 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	P	577/765 (75%)	529 (92%)	48 (8%)	13 43

All (48) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	P	26	PHE
1	P	35	ILE
1	P	78	ILE
1	P	82	THR
1	P	115	GLN

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	P	119	ASP
1	P	121	ARG
1	P	150	ILE
1	P	151	ILE
1	P	170	GLN
1	P	180	GLN
1	P	196	LYS
1	P	312	HIS
1	P	319	LYS
1	P	322	LEU
1	P	355	ASN
1	P	386	LEU
1	P	453	SER
1	P	460	SER
1	P	467	ASN
1	P	471	LEU
1	P	472	SER
1	P	474	THR
1	P	493	THR
1	P	494	LEU
1	P	501	ASP
1	P	504	MET
1	P	545	LEU
1	P	554	THR
1	P	558	ARG
1	P	568	LYS
1	P	613	LYS
1	P	648	GLU
1	P	679	GLN
1	P	688	LEU
1	P	706	PHE
1	P	723	THR
1	P	727	LEU
1	P	741	ILE
1	P	751	LEU
1	P	769	LYS
1	P	770	SER
1	P	779	VAL
1	P	800	LYS
1	P	807	ARG
1	P	814	ASP
1	P	822	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	P	823	LEU

Some 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 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 ⓘ

### 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<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	P	656/876 (74%)	0.56	30 (4%) 33 31	80, 140, 217, 243	0

All (30) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	P	151	ILE	8.7
1	P	105	ILE	3.8
1	P	104	LEU	3.0
1	P	219	PHE	2.9
1	P	129	VAL	2.8
1	P	36	LYS	2.8
1	P	43	SER	2.8
1	P	138	ILE	2.8
1	P	727	LEU	2.7
1	P	413	THR	2.7
1	P	679	GLN	2.6
1	P	115	GLN	2.6
1	P	612	VAL	2.5
1	P	42	ALA	2.5
1	P	160	GLU	2.4
1	P	150	ILE	2.4
1	P	731	VAL	2.3
1	P	633	VAL	2.3
1	P	202	GLU	2.3
1	P	751	LEU	2.3
1	P	637	LEU	2.2
1	P	125	PHE	2.2
1	P	747	LEU	2.2
1	P	744	VAL	2.1
1	P	415	LEU	2.1
1	P	548	LEU	2.1
1	P	191	PHE	2.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	P	589	LEU	2.0
1	P	461	ALA	2.0
1	P	216	PHE	2.0

## 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.