



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 04:45 PM GMT

PDB ID : 4G0H  
Title : Crystal structure of the N-terminal domain of Helicobacter pylori CagA protein  
Authors : Kaplan-Turkoz, B.; Remaut, H.; Dian, C.; Louche, A.; Terradot, L.  
Deposited on : 2012-07-09  
Resolution : 3.60 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](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  
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

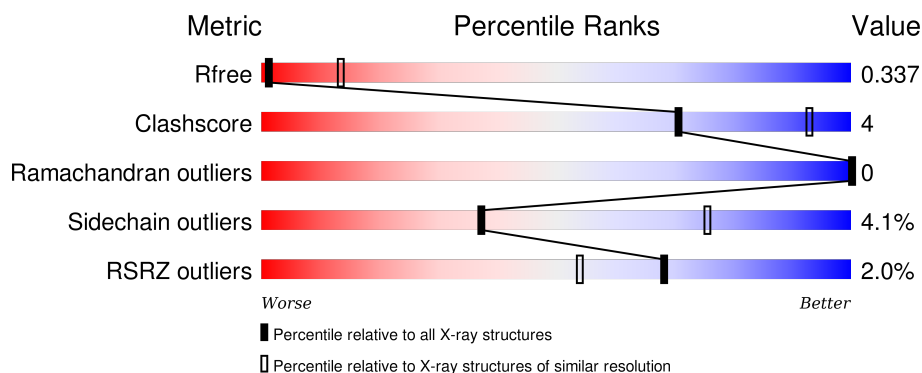
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.60 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	1408 (3.80-3.40)
Clashscore	102246	1010 (3.74-3.46)
Ramachandran outliers	100387	1007 (3.76-3.44)
Sidechain outliers	100360	1007 (3.76-3.44)
RSRZ outliers	91569	1003 (3.78-3.42)

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	A	916	<div> <div></div> <div>43%</div> <div>5% •</div> <div>51%</div> </div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 3275 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	A	450	Total	C	N	O	S	0	0	0
			3275	2040	572	658	5			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	885	LYS	-	EXPRESSION TAG	UNP P55980
A	886	GLY	-	EXPRESSION TAG	UNP P55980
A	887	GLU	-	EXPRESSION TAG	UNP P55980
A	888	LEU	-	EXPRESSION TAG	UNP P55980
A	889	ASN	-	EXPRESSION TAG	UNP P55980
A	890	SER	-	EXPRESSION TAG	UNP P55980
A	891	LYS	-	EXPRESSION TAG	UNP P55980
A	892	LEU	-	EXPRESSION TAG	UNP P55980
A	893	GLU	-	EXPRESSION TAG	UNP P55980
A	894	GLY	-	EXPRESSION TAG	UNP P55980
A	895	LYS	-	EXPRESSION TAG	UNP P55980
A	896	PRO	-	EXPRESSION TAG	UNP P55980
A	897	ILE	-	EXPRESSION TAG	UNP P55980
A	898	PRO	-	EXPRESSION TAG	UNP P55980
A	899	ASN	-	EXPRESSION TAG	UNP P55980
A	900	PRO	-	EXPRESSION TAG	UNP P55980
A	901	LEU	-	EXPRESSION TAG	UNP P55980
A	902	LEU	-	EXPRESSION TAG	UNP P55980
A	903	GLY	-	EXPRESSION TAG	UNP P55980
A	904	LEU	-	EXPRESSION TAG	UNP P55980
A	905	ASP	-	EXPRESSION TAG	UNP P55980
A	906	SER	-	EXPRESSION TAG	UNP P55980
A	907	THR	-	EXPRESSION TAG	UNP P55980
A	908	ARG	-	EXPRESSION TAG	UNP P55980
A	909	THR	-	EXPRESSION TAG	UNP P55980
A	910	GLY	-	EXPRESSION TAG	UNP P55980
A	911	HIS	-	EXPRESSION TAG	UNP P55980

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	912	HIS	-	EXPRESSION TAG	UNP P55980
A	913	HIS	-	EXPRESSION TAG	UNP P55980
A	914	HIS	-	EXPRESSION TAG	UNP P55980
A	915	HIS	-	EXPRESSION TAG	UNP P55980
A	916	HIS	-	EXPRESSION TAG	UNP P55980



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	97.83Å 97.83Å 245.39Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	62.75 – 3.60 62.75 – 3.60	Depositor EDS
% Data completeness (in resolution range)	91.8 (62.75-3.60) 91.8 (62.75-3.60)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.57 (at 3.57Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1056)	Depositor
R, $R_{free}$	0.329 , 0.341 0.324 , 0.337	Depositor DCC
$R_{free}$ test set	669 reflections (5.01%)	DCC
Wilson B-factor (Å <sup>2</sup> )	159.9	Xtriage
Anisotropy	0.064	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 160.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtriage
Outliers	0 of 13349 reflections	Xtriage
$F_o, F_c$ correlation	0.85	EDS
Total number of atoms	3275	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	158.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.45% 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, 0.375 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	0.22	0/3304	0.37	0/4458

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	3275	0	3025	28	0
All	All	3275	0	3025	28	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.

The worst 5 of 28 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:334:HIS:NE2	1:A:337:ASN:OD1	2.27	0.65
1:A:507:ASP:N	1:A:507:ASP:OD2	2.30	0.64
1:A:676:ALA:O	1:A:680:ASN:ND2	2.30	0.63
1:A:319:LYS:HB3	1:A:342:TYR:HB2	1.82	0.61
1:A:410:GLN:HG3	1:A:471:LEU:HD21	1.84	0.59

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	432/916 (47%)	414 (96%)	18 (4%)	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	A	316/801 (40%)	303 (96%)	13 (4%)	37	75

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	474	THR
1	A	507	ASP
1	A	687	GLU
1	A	399	ARG
1	A	684	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	467	ASN

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	A	810	GLN
1	A	468	ASN
1	A	446	ASN
1	A	669	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	450/916 (49%)	-0.01	9 (2%)	68 54	92, 156, 214, 255	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	569	GLY	3.4
1	A	633	VAL	2.9
1	A	630	GLU	2.8
1	A	565	LEU	2.6
1	A	812	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.