



# Full wwPDB X-ray Structure Validation Report

May 23, 2020 – 01:18 am BST

PDB ID : 4YVM  
Title : X-ray structure of Helicobacter pylori CagL-K74  
Authors : Choi, J.M.; Choi, Y.H.; Cha, J.H.; Lee, S.H.  
Deposited on : 2015-03-20  
Resolution : 2.79 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

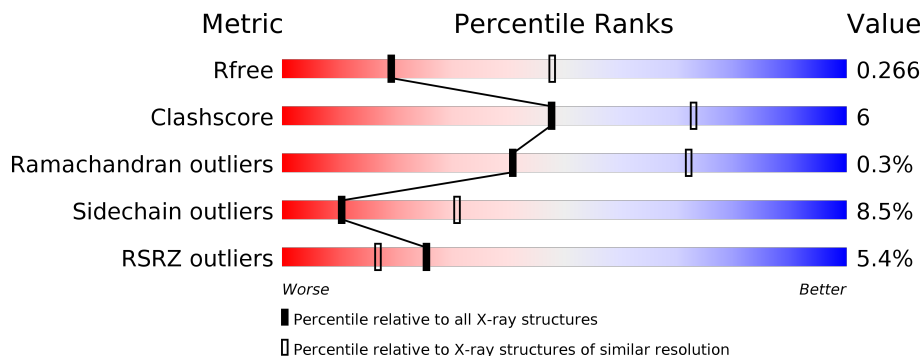
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.79 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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	221	 4% 73% 18% • 9%
1	B	221	 6% 73% 16% • 8%

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3325 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
			Total	C	N	O	S			
1	A	202	1614	1024	270	315	5	0	0	0
1	B	204	1629	1032	273	319	5	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	17	GLY	-	expression tag	UNP Q75XL3
A	18	SER	-	expression tag	UNP Q75XL3
A	19	HIS	-	expression tag	UNP Q75XL3
A	20	MET	-	expression tag	UNP Q75XL3
B	17	GLY	-	expression tag	UNP Q75XL3
B	18	SER	-	expression tag	UNP Q75XL3
B	19	HIS	-	expression tag	UNP Q75XL3
B	20	MET	-	expression tag	UNP Q75XL3

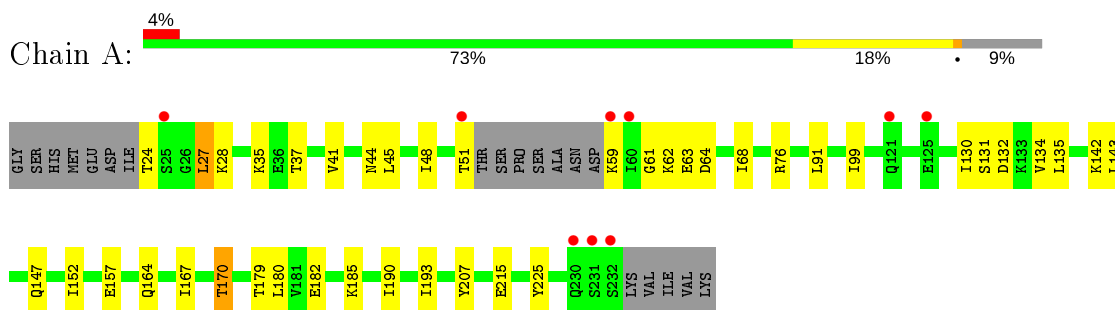
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	51	Total	O	0	0
			51	51		
2	B	31	Total	O	0	0
			31	31		

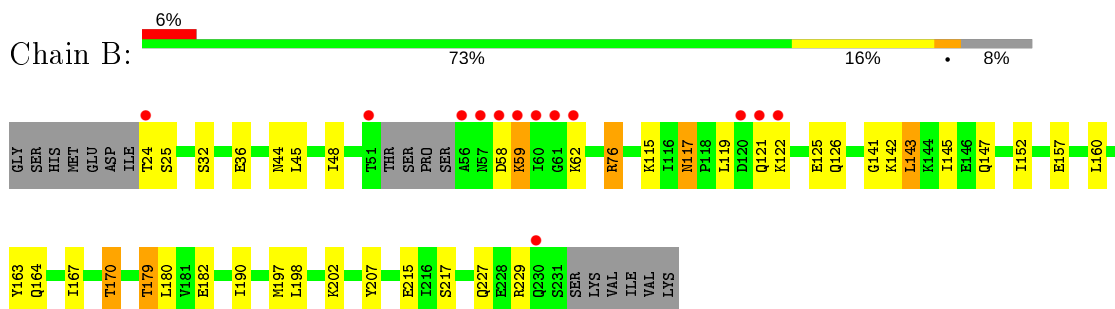
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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



- Molecule 1: Cag pathogenicity island protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	97.62Å 97.62Å 144.24Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	37.35 – 2.79 37.35 – 2.79	Depositor EDS
% Data completeness (in resolution range)	98.3 (37.35-2.79) 98.4 (37.35-2.79)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	0.11	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.61 (at 2.81Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, $R_{free}$	0.199 , 0.267 0.200 , 0.266	Depositor DCC
$R_{free}$ test set	905 reflections (5.12%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	38.6	Xtrriage
Anisotropy	0.136	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 40.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3325	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	31.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 45.30 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.3472e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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	0.48	0/1629	0.54	0/2182
1	B	0.46	0/1644	0.55	0/2203
All	All	0.47	0/3273	0.55	0/4385

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	1614	0	1674	20	0
1	B	1629	0	1684	21	0
2	A	51	0	0	2	0
2	B	31	0	0	0	0
All	All	3325	0	3358	40	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 6.

All (40) 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:179:THR:HG22	1:B:182:GLU:H	1.47	0.78

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:163:TYR:CD1	1:B:197:MET:HG3	2.24	0.72
1:B:58:ASP:HB3	1:B:62:LYS:HB2	1.74	0.68
1:A:152:ILE:HD13	1:A:207:TYR:HB3	1.81	0.63
1:B:117:ASN:OD1	1:B:117:ASN:N	2.37	0.58
1:B:152:ILE:HD13	1:B:207:TYR:HB3	1.90	0.54
1:A:142:LYS:HE3	1:A:215:GLU:OE2	2.11	0.50
1:B:44:ASN:HB3	1:B:48:ILE:HG13	1.94	0.49
1:A:132:ASP:HB3	1:A:135:LEU:HB2	1.95	0.48
1:A:24:THR:HG23	1:A:27:LEU:HB2	1.96	0.48
1:B:121:GLN:O	1:B:125:GLU:HG2	2.14	0.47
1:B:45:LEU:HD23	1:B:76:ARG:HE	1.80	0.47
1:A:190:ILE:HA	1:A:193:ILE:HD12	1.97	0.46
1:B:45:LEU:HB3	1:B:76:ARG:HH21	1.79	0.46
1:A:61:GLY:HA2	2:A:340:HOH:O	2.16	0.45
1:B:45:LEU:HA	1:B:45:LEU:HD12	1.55	0.45
1:B:163:TYR:CE1	1:B:197:MET:HG3	2.52	0.44
1:A:143:LEU:HD12	1:A:143:LEU:HA	1.69	0.44
1:A:45:LEU:HB3	1:A:76:ARG:NH1	2.33	0.44
1:A:45:LEU:HA	1:A:45:LEU:HD12	1.77	0.44
1:B:141:GLY:O	1:B:145:ILE:HG13	2.18	0.44
1:B:142:LYS:HE2	1:B:215:GLU:OE2	2.18	0.44
1:A:35:LYS:HE3	2:A:303:HOH:O	2.18	0.43
1:A:44:ASN:O	1:A:48:ILE:HG13	2.17	0.43
1:B:143:LEU:O	1:B:147:GLN:HG2	2.19	0.43
1:A:182:GLU:OE2	1:A:185:LYS:HD2	2.19	0.43
1:A:37:THR:O	1:A:41:VAL:HG23	2.19	0.42
1:A:167:ILE:O	1:A:170:THR:HB	2.19	0.42
1:A:64:ASP:O	1:A:68:ILE:HG12	2.19	0.42
1:B:198:LEU:HG	1:B:202:LYS:HE2	1.99	0.42
1:B:170:THR:HG23	1:B:190:ILE:HD13	2.01	0.42
1:B:45:LEU:HD23	1:B:76:ARG:NE	2.35	0.41
1:A:157:GLU:HG2	1:B:157:GLU:HG2	2.02	0.41
1:A:91:LEU:HD12	1:A:91:LEU:HA	1.75	0.41
1:B:160:LEU:HD23	1:B:160:LEU:HA	1.88	0.41
1:A:62:LYS:HD3	1:A:62:LYS:HA	1.91	0.41
1:B:167:ILE:O	1:B:170:THR:HG22	2.21	0.41
1:A:134:VAL:HG11	1:A:225:TYR:CE1	2.56	0.40
1:B:179:THR:HG22	1:B:182:GLU:N	2.26	0.40
1:A:28:LYS:HD2	1:A:28:LYS:HA	1.81	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	198/221 (90%)	191 (96%)	7 (4%)	0	100	100
1	B	200/221 (90%)	191 (96%)	8 (4%)	1 (0%)	29	61
All	All	398/442 (90%)	382 (96%)	15 (4%)	1 (0%)	41	72

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	59	LYS

### 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	181/198 (91%)	169 (93%)	12 (7%)	16	44
1	B	182/198 (92%)	163 (90%)	19 (10%)	7	21
All	All	363/396 (92%)	332 (92%)	31 (8%)	10	31

All (31) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	27	LEU
1	A	51	THR
1	A	59	LYS
1	A	63	GLU
1	A	99	ILE

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	A	130	ILE
1	A	131	SER
1	A	147	GLN
1	A	164	GLN
1	A	170	THR
1	A	179	THR
1	A	180	LEU
1	B	24	THR
1	B	25	SER
1	B	32	SER
1	B	36	GLU
1	B	59	LYS
1	B	76	ARG
1	B	115	LYS
1	B	117	ASN
1	B	119	LEU
1	B	122	LYS
1	B	126	GLN
1	B	143	LEU
1	B	164	GLN
1	B	170	THR
1	B	179	THR
1	B	180	LEU
1	B	217	SER
1	B	227	GLN
1	B	229	ARG

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	A	202/221 (91%)	-0.05	9 (4%) 33 23	7, 25, 75, 99	0
1	B	204/221 (92%)	0.11	13 (6%) 19 12	8, 25, 80, 129	0
All	All	406/442 (91%)	0.03	22 (5%) 25 17	7, 25, 76, 129	0

All (22) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	56	ALA	9.2
1	B	57	ASN	8.2
1	B	51	THR	4.8
1	B	58	ASP	4.6
1	A	59	LYS	4.3
1	A	60	ILE	4.1
1	A	231	SER	4.1
1	A	51	THR	4.0
1	B	60	ILE	3.9
1	A	232	SER	3.3
1	B	230	GLN	3.1
1	B	122	LYS	2.9
1	B	59	LYS	2.8
1	B	120	ASP	2.8
1	B	61	GLY	2.5
1	A	121	GLN	2.4
1	B	121	GLN	2.4
1	A	125	GLU	2.4
1	B	62	LYS	2.3
1	A	230	GLN	2.2
1	A	25	SER	2.1
1	B	24	THR	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.