



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 02:12 AM GMT

PDB ID : 2G3V  
Title : Crystal structure of CagS (HP0534, Cag13) from Helicobacter pylori  
Authors : Cendron, L.; Tasca, E.; Angelini, A.; Seydel, A.; Battistutta, R.; Montecucco, C.; Zanotti, G.  
Deposited on : 2006-02-21  
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](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

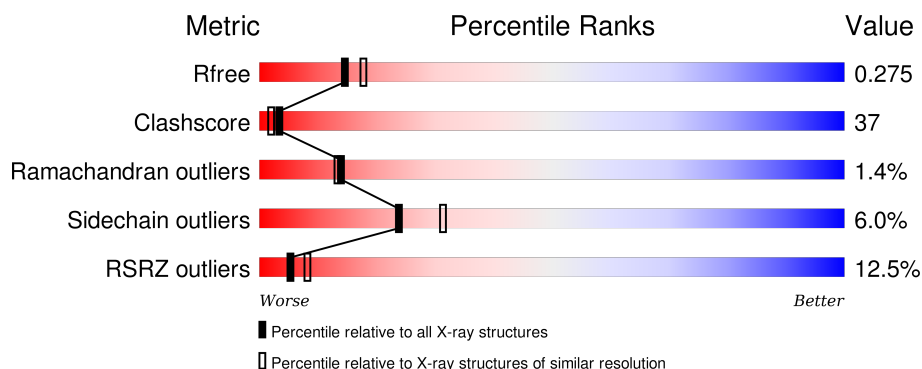
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 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  $\geq 3$ , 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	208	<div> <div>7%</div> <div>37%</div> <div>38%</div> <div>23%</div> </div>
1	B	208	<div> <div>8%</div> <div>37%</div> <div>37%</div> <div>23%</div> </div>
1	C	208	<div> <div>10%</div> <div>36%</div> <div>38%</div> <div>23%</div> </div>
1	D	208	<div> <div>12%</div> <div>35%</div> <div>39%</div> <div>23%</div> </div>
2	E	7	<div> <div>43%</div> <div>57%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	F	7	 43% 57%
2	G	7	 43% 57%
2	H	7	 29% 71%

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 5790 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 13.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	161	Total	C	N	O	Se	0	0	0
			1349	877	218	245	9			
1	B	161	Total	C	N	O	Se	0	0	0
			1349	877	218	245	9			
1	C	161	Total	C	N	O	Se	0	0	0
			1349	877	218	245	9			
1	D	161	Total	C	N	O	Se	0	0	0
			1349	877	218	245	9			

- Molecule 2 is a protein called (UNK)(UNK)(UNK)(UNK)(UNK)(MSE)(UNK).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	7	Total	C	N	O	Se	0	0	0
			39	23	7	8	1			
2	F	7	Total	C	N	O	Se	0	0	0
			39	23	7	8	1			
2	G	7	Total	C	N	O	Se	0	0	0
			39	23	7	8	1			
2	H	7	Total	C	N	O	Se	0	0	0
			39	23	7	8	1			

- Molecule 3 is water.

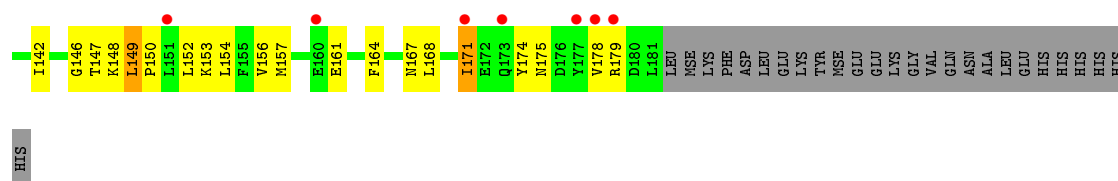
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	54	Total	O	0	0
			54	54		
3	B	48	Total	O	0	0
			48	48		
3	C	57	Total	O	0	0
			57	57		
3	D	62	Total	O	0	0
			62	62		

*Continued on next page...*

*Continued from previous page...*

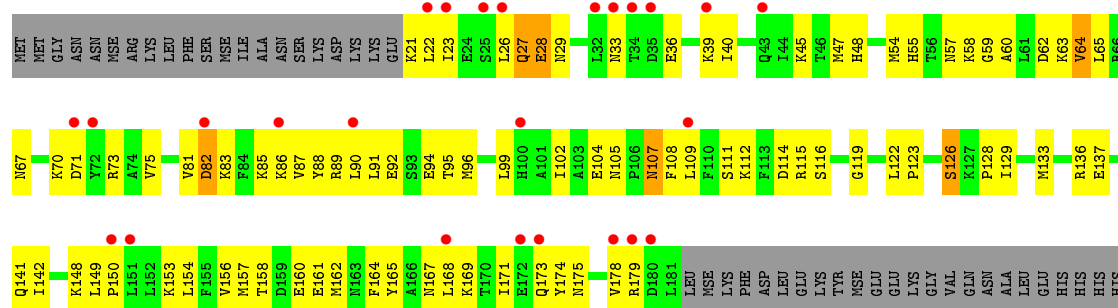
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	E	3	Total 3	O 3	0	0
3	F	5	Total 5	O 5	0	0
3	G	5	Total 5	O 5	0	0
3	H	4	Total 4	O 4	0	0





- Molecule 1: CAG pathogenicity island protein 13

Chain D: 12% 35% 39% 23%



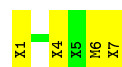
- Molecule 2: (UNK)(UNK)(UNK)(UNK)(UNK)(MSE)(UNK)

Chain E: 43% 57%



- Molecule 2: (UNK)(UNK)(UNK)(UNK)(UNK)(MSE)(UNK)

Chain F: 43% 57%



- Molecule 2: (UNK)(UNK)(UNK)(UNK)(UNK)(MSE)(UNK)

Chain G: 43% 57%



- Molecule 2: (UNK)(UNK)(UNK)(UNK)(UNK)(MSE)(UNK)

Chain H: 29% 71%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	124.36Å 124.36Å 55.86Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.41 – 2.30 39.41 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.8 (39.41-2.30) 99.5 (39.41-2.00)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.01	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.89 (at 2.00Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.219 , 0.279 0.220 , 0.275	Depositor DCC
$R_{free}$ test set	3691 reflections (10.68%)	DCC
Wilson B-factor (Å <sup>2</sup> )	9.3	Xtriage
Anisotropy	0.589	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 49.1	EDS
Estimated twinning fraction	0.236 for h,-k,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.28$ , $\langle L^2 \rangle = 0.11$	Xtriage
Outliers	0 of 57722 reflections	Xtriage
$F_o, F_c$ correlation	0.77	EDS
Total number of atoms	5790	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	10.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 31.65 % 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 1.0684e-03.*

<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	0.44	0/1365	0.63	0/1812
1	B	0.43	0/1365	0.64	0/1812
1	C	0.44	0/1365	0.65	2/1812 (0.1%)
1	D	0.43	0/1365	0.62	0/1812
2	E	0.43	0/7	1.04	0/7
2	F	0.62	0/7	1.14	0/7
2	G	0.32	0/7	1.04	0/7
2	H	0.54	0/7	1.15	0/7
All	All	0.43	0/5488	0.64	2/7276 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	118	LEU	CB-CA-C	-6.66	97.55	110.20
1	C	108	PHE	N-CA-C	-5.27	96.78	111.00

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	1349	0	1384	94	0
1	B	1349	0	1384	101	0
1	C	1349	0	1384	103	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	D	1349	0	1384	107	0
2	E	39	0	17	5	0
2	F	39	0	17	3	0
2	G	39	0	17	6	0
2	H	39	0	17	5	0
3	A	54	0	0	17	0
3	B	48	0	0	16	0
3	C	57	0	0	15	0
3	D	62	0	0	22	0
3	E	3	0	0	4	0
3	F	5	0	0	1	0
3	G	5	0	0	6	0
3	H	4	0	0	1	0
All	All	5790	0	5604	412	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 37.

The worst 5 of 412 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:C:153:LYS:HE3	1:C:156:VAL:HG21	1.35	1.09
1:A:139:ASP:HA	3:A:218:HOH:O	1.64	0.97
1:D:21:LYS:HG3	1:D:22:LEU:HG	1.46	0.95
1:C:71:ASP:O	1:C:75:VAL:HG23	1.66	0.94
1:A:137:GLU:HG3	1:C:137:GLU:HG3	1.50	0.92

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	159/208 (76%)	145 (91%)	11 (7%)	3 (2%)	10	8
1	B	159/208 (76%)	140 (88%)	16 (10%)	3 (2%)	10	8
1	C	159/208 (76%)	149 (94%)	9 (6%)	1 (1%)	30	36
1	D	159/208 (76%)	141 (89%)	16 (10%)	2 (1%)	15	15
2	E	1/7 (14%)	0	1 (100%)	0	100	100
2	F	1/7 (14%)	0	1 (100%)	0	100	100
2	G	1/7 (14%)	0	1 (100%)	0	100	100
2	H	1/7 (14%)	0	1 (100%)	0	100	100
All	All	640/860 (74%)	575 (90%)	56 (9%)	9 (1%)	14	13

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	109	LEU
1	A	109	LEU
1	D	64	VAL
1	B	57	ASN
1	B	150	PRO

### 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	152/182 (84%)	144 (95%)	8 (5%)	28	37
1	B	152/182 (84%)	142 (93%)	10 (7%)	21	27
1	C	152/182 (84%)	141 (93%)	11 (7%)	18	22
1	D	152/182 (84%)	144 (95%)	8 (5%)	28	37
2	E	-	1 (100%)	0	100	100
2	F	-	1 (100%)	0	100	100
2	G	-	1 (100%)	0	100	100
2	H	-	1 (100%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
All	All	612/728 (84%)	575 (94%)	37 (6%)	24	31

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	149	LEU
1	C	82	ASP
1	D	126	SER
1	B	171	ILE
1	C	28	GLU

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	48	HIS
1	B	105	ASN
1	D	67	ASN
1	B	29	ASN
1	D	48	HIS

### 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 ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

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	152/208 (73%)	0.84	14 (9%) 11 16	1, 7, 21, 36	0
1	B	152/208 (73%)	0.93	17 (11%) 7 10	1, 8, 23, 35	0
1	C	152/208 (73%)	0.97	20 (13%) 4 7	1, 8, 21, 37	0
1	D	152/208 (73%)	1.22	25 (16%) 2 4	1, 9, 22, 36	0
2	E	0/7	-	-	-	-
2	F	0/7	-	-	-	-
2	G	0/7	-	-	-	-
2	H	0/7	-	-	-	-
All	All	608/860 (70%)	0.99	76 (12%) 5 8	1, 8, 23, 37	0

The worst 5 of 76 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	35	ASP	8.0
1	C	23	ILE	6.3
1	A	26	LEU	5.4
1	D	25	SER	5.2
1	C	151	LEU	4.7

### 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

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.